BSI

20. Deutscher
IT-Sicherheitskongress

Cybernation Deutschland:
Kooperation gewinnt.

Die Themen des 20. Deutschen IT-Sicherheitskongresses

 | Artikelnummer: BSI-Kon24/001

Liebe Leserin, lieber Leser,

es freut mich sehr, dass wir mit dem diesjährigen Deutschen IT-Sicherheitskongress bereits zum 20. Mal Expertinnen und Experten zu den relevantesten Cybersicherheitsthemen der Zeit zusammenbringen. 30 Fachvorträge beleuchten aktuelle Themen, die die Cybersicherheitscommunity bewegen und zugleich von Relevanz für unsere digitale Gesellschaft sind.

Der Programmbeirat hat eine exzellente Auswahl getroffen. So ist eine der drängendsten Fragen, wie sich die Künstliche Intelligenz weiterentwickelt: Sie kann Angriffe gefährlicher machen, aber auch selbst Ziel von Attacken sein. Längst sollte klar sein, dass beispielsweise Deepfakes auf Social-Media-Plattformen Meinungen beeinflussen, Wahlen manipulieren oder Gesellschaften destabilisieren. KI kann jedoch auch dazu beitragen, Systeme resilienter zu machen.

Zur Erhöhung der Cybersicherheit tragen darüber hinaus in erheblichem Umfang Automatisierung und Skalierung bei. Ein standardisiertes Common Security Advisory Framework (CSAF) zum Beispiel macht Informationen zu Schwachstellen maschinenlesbar und reduziert so den manuellen Aufwand, Sicherheitsinformationen zu suchen und herauszufinden, ob Produkte betroffen sind oder nicht.

Um diese Potenziale für mehr Cybersicherheit zu heben, übernehmen wir im BSI viele Rollen: Wir treiben an und machen möglich, wir knüpfen partnerschaftliche Netzwerke und geben unser Wissen weiter.

In diesem Sinne wünsche ich eine anregende Lektüre.

Herzliche Grüße

Ihre

Claudia Plattner

Präsidentin des Bundesamts für Sicherheit in der Informationstechnik

Inhalt der Dokumentation

1. Kongresstag 7. Mai 2024

2. Kongresstag 8. Mai 2024

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Krypto-agile IoT-Kommunikation bei einem Schienennetz-Betreiber

Gunnar Preissler[1], Klaus Schmeh[2]

Kurzfassung:
Die Kommunikation zwischen den Komponenten eines Schienennetzes muss kryptografisch geschützt werden, da beispielsweise eine manipulierte Weiche oder ein manipuliertes Signal schnell zur Katastrophe führen kann. Der vorliegende Beitrag beschreibt eine Projektarbeit, in deren Rahmen bei einem überregionalen Schienennetz-Betreiber in Deutschland eine Vielzahl von Stelleinheiten (unter anderem Weichen und Signale) und Stellwerken digitalisiert werden, wozu auch die Ausstattung mit hardwarebasierten Krypto-Komponenten gehört. Da diese Krypto-Komponenten mehrere Jahrzehnte lang im Einsatz sein können, spielt die Krypto-Agilität eine wichtige Rolle.

Stichworte: Digitalisierung, IKEv2, Internet of Things, IoT, IPsec, Krypto-Agilität, Krypto-Netzwerk-Module, Kryptografie, PKI, Post-Quanten-Verfahren, Schienennetz, Stelleinheit

1 Einführung

Dieser Beitrag präsentiert eine Fallstudie zu einem Digitalisierungsprojekt, das derzeit bei einem deutschen Eisenbahnunternehmen durchgeführt wird, das ein bundesweites Schienennetz betreibt. Die Autoren dieser Arbeit sind Teil eines Entwicklungskonsortiums, das sich die Aufgabe gestellt hat, die hohen Anforderungen an die funktionale Sicherheit (Safety), die Langlebigkeit der Komponenten und die Dynamik der kryptografischen Verfahren in diesem Projekt in Einklang zu bringen. Sie berichten im Folgenden über ihre Erfahrungen.

Im Mittelpunkt des Projekts stehen digitale Stellwerke, also ortsfeste Anlagen zur Steuerung des Eisenbahn-Betriebs über sogenannte Stelleinheiten. Zu Letzteren gehören insbesondere Weichen, Signale und verschiedene Sensorsysteme im Gleisfeld. Ziel ist es, dass alle Stelleinheiten einer Region von einem zentralen Ort aus auf sichere Weise gesteuert werden können.

[1] Eviden Germany, Erfurt
[2] Eviden Germany, Gelsenkirchen

2 Istzustand

Allgemein gibt es in den europäischen Schienennetzen einen erheblichen Modernisierungsbedarf. Insbesondere streben momentan alle größeren Betreiber die Digitalisierung und Konzentration ihrer Netze an. In diesem Zusammenhang haben die europäischen Betreiber von Eisenbahn-Infrastruktur 2014 die Organisation EULYNX gegründet, die die Entwicklung einheitlicher Industriestandards für digitale Stellwerkstechnik zum Ziel hat [1].

Abbildung 1: In den europäischen Schienennetzen besteht ein erheblicher Modernisierungsbedarf. Quelle: Pixabay

In dem in dieser Arbeit betrachteten deutschen Schienennetz gibt es knapp 4000 Stellwerke [2]. Diese teilen sich wie folgt auf (Abbildung 2):

- *1369 elektronische Stellwerke*: Diese haben ein mittleres Alter von etwa 13 Jahren und steuern 78.478 Stelleinheiten.
- *1219 Drucktasten-Stellwerke*: Deren mittleres Alter beträgt etwa 44 Jahre. Sie steuern 92.072 Stelleinheiten. Drucktasten-Stellwerke sind der häufigste Stellwerk-Typ im betrachteten Schienennetz.
- *279 elektromechanische Stellwerke*: Elektromechanische Stellwerke, die ein mittleres Alter von etwa 64 Jahren aufweisen, steuern 8349 Stelleinheiten.
- *651 mechanische Stellwerke*: Mit einem mittleren Alter von etwa 77 Jahren steuern diese 9309 Stelleinheiten.
- *328 sonstige Stellwerke:* Diese steuern 5371 Stelleinheiten.

Diese Aufstellung zeigt, dass noch längst nicht alle Stellwerke und Stelleinheiten im hier betrachteten Schienennetz digitalisiert sind. Unter anderem

deshalb verfügt der Betreiber bisher nicht über ein digitales Kommunikationssystem, an das alle Stellwerke und Stelleinheiten angeschlossen sind. Es ist daher nicht möglich, von einer zentralen Stelle aus alle Weichen und Signale des Schienennetzes zu steuern. Es werden lediglich in verschiedenen Regionen jeweils eigenständige, proprietäre Kommunikationsnetze betrieben, mit denen zumindest ein Teil der Stelleinheiten erreicht wird.

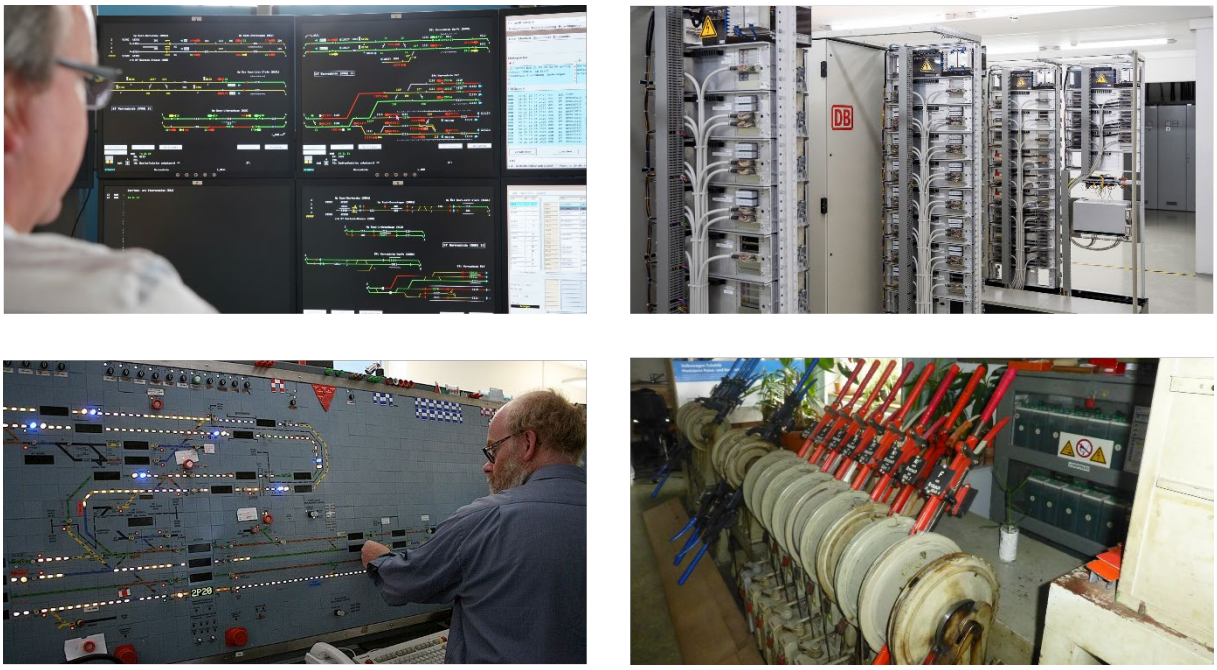


Abbildung 2: Im Schienennetz, das in dieser Arbeit betrachtet wird, gibt es elektronische Stellwerke (oben links) sowie Drucktasten-Stellwerke (oben rechts), elektromechanische Stellwerke (unten links) und mechanische Stellwerke. Es besteht ein erheblicher Modernisierungsbedarf. Quellen: DB188861 Deutsche Bahn AG Volker Emersleben / DB231750 Deutsche Bahn AG Sebastian Berger / Roger Carvell / Werner Jäckisch

Eine Digitalisierung der Steuerung ist deshalb geboten, weil die Politik für die nähere Zukunft eine Verdoppelung des Schienenverkehrs anstrebt. Dies erfordert eine dichtere Taktung der Züge, was nur mit effektiverer Steuerung erreichbar ist.

In der Vergangenheit wurde die Gefahr durch Cyber-Angriffe auf Bahnsteuerungsanlagen kaum beachtet, da es sich um geschlossene Netzwerke handelte, auf welche kein unerwünschter externer Zugriff möglich ist. Diese Betrachtung hat allerdings ihre Gültigkeit verloren, denn kein Netzwerk dieser Art ist zu 100 Prozent geschlossen, vor allem weil die gleichzeitige Forderung nach Wartungs- und Diagnosezugängen dies nicht zulässt.

Die bisherigen Sicherungsmethoden, die den Personenschutz (Safety) gewährleisten, bilden diverse Fehlermöglichkeiten im System per se ab, beispielsweise zufällige Defekte an Signal- oder Stellanlagen. Safety schützt also vor unabsichtlichen Fehlern. Security hingegen schützt vor absichtlichen Manipulationen – dieser Aspekt findet sich in den bisher verbauten Anlagen überhaupt nicht wieder.

Ein besonderes Augenmerk muss auf die Usability gelegt werden, vor allem im Zusammenhang mit der Handhabbarkeit der notwendigen Update- bzw. Patch-Prozesse. Funktionale Updates an den Objectcontrollern sind üblicherweise nur in mehrjährigen Zyklen erforderlich. Hingegen kann es bei Bekanntwerden von schwerwiegenden Security-Schwachstellen zukünftig notwendig sein, dass sich Update-Intervalle der Security-Funktionen durch kurzfristig durchgeführte Patch-Maßnahmen verkürzen. Da eine Rückwirkungsfreiheit auf die Safety-Funktionen gewährleistet bleiben muss, wird die Handhabbarkeit dieser Patch-Prozesse (Freigabe sowie Ausrollen) einen besonderen Stellenwert haben.

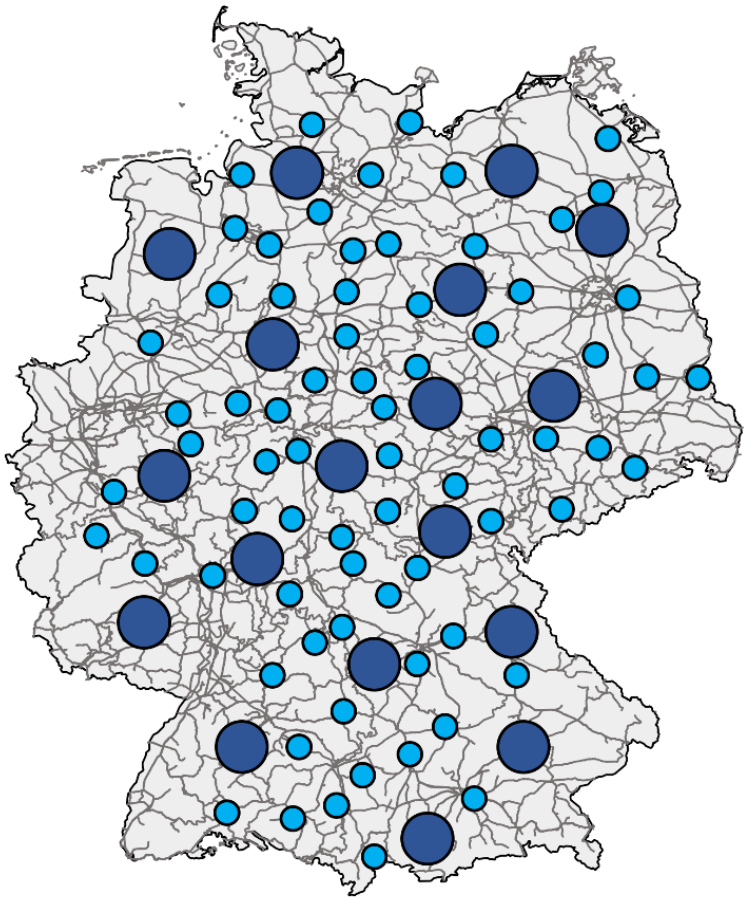


Abbildung 3: Geplant ist ein bundesweites IT-Netz, das eine zentrale Steuerung aller Stelleinheiten erlaubt

3 Das Projekt

Der in dieser Arbeit beschriebene Schienennetz-Betreiber plant die stufenweise Digitalisierung und Vernetzung seiner etwa 3000 Stellwerke und der angeschlossenen Stelleinheiten (Abbildung 2 und 3). Aus Sicht der Informationstechnik fällt dieses Projekt in die Bereiche Operational Technology (OT), Internet of Things (IoT) und Kritische Infrastrukturen. Ein erstes Digitalisierungsvorhaben wurde im Raum Stuttgart begonnen und soll 2024 in den Betrieb übergeben werden. Anschließend sollen im Laufe von mehreren Jahren bundesweit alle Stelleinheiten erfasst werden.

Zur digitalen Kommunikation innerhalb des Schienennetzes werden die im Rahmen von EULYNX festgelegten Standards SCI (Standard Communication Interface) und RaSTA (Rail Safe Transport Application) verwendet [3]. RaSTA kommt üblicherweise über das Internet Protocol (IP) zum Einsatz.

In dieser Arbeit geht es um einen wichtigen Teilaspekt dieses Vorhabens: die kryptografische Absicherung der Kommunikation zwischen den Stellwerken und den Stelleinheiten. Es ist offensichtlich, dass diesem Thema eine bedeutende Rolle zukommt, da ein Angreifer, der die Kommunikation innerhalb des Schienennetzes manipulieren kann, einen großen Schaden anrichten kann. Die kryptografische Absicherung hat in diesem Zusammenhang folgende Ziele:

- *Authentizität*: Es darf für einen Angreifer nicht möglich sein, falsche Anweisungen zu verschicken, da dies beispielsweise das Verstellen einer Weiche oder eines Signals zur Folge haben kann. Der Authentizität kommt daher eine hohe Bedeutung zu.
- *Integrität*: Ein Angreifer darf nicht die Möglichkeit haben, eine Nachricht zu manipulieren, da dies ebenfalls gefährliche Eingriffe in den Schienenverkehr erlauben würde. Auch die Integrität hat in diesem Zusammenhang daher eine hohe Bedeutung.
- *Vertraulichkeit*: Die Vertraulichkeit der Kommunikation zwischen Komponenten des Schienennetzes wird angestrebt, ist den vorstehenden Schutzzielen jedoch nachgeordnet.

4 Verwendete Krypto-Komponenten

Die kryptografische Absicherung der Kommunikation zwischen den Schienennetz-Komponenten soll mit geeigneten Krypto-Netzwerk-Modulen realisiert werden. Es wird angestrebt, dass handelsübliche Krypto-Hardware (Commercial of the Shelf) eingesetzt wird, während Sonderanfertigungen

und Insellösungen vermieden werden sollen. Die verwendeten Krypto-Netzwerk-Module müssen auf eine Laufzeit von mehreren Jahrzehnten ausgelegt sein, wobei ein unterbrechungsfreier Betrieb gewährleistet sein muss.

Die lange Laufzeit der Krypto-Netzwerk-Module macht Krypto-Agilität zu einem wichtigen Thema. Da sich die Sicherheitsanforderungen bezüglich der gängigen Krypto-Verfahren im Laufe der Jahrzehnte ändern können, muss es möglich sein, diese zu aktualisieren oder auszutauschen. Dies gilt auch und gerade für die Umstellung auf Post-Quanten-Verfahren, die in den nächsten zehn Jahren zu erwarten ist [4].

Abbildung 4: Das eingesetzte Krypto-Netzwerk-Modul (links) wird mit einem HSM im µSD-Format (rechts) betrieben. Das HSM ist austauschbar, wodurch unter anderem neue Krypto-Algorithmen eingespielt werden können.

Für die Umsetzung kommen Netzwerkmodule des Anbieters Westermo Network Technologies infrage, die bereits heute in verschiedenen Anwendungsszenarien bei elektronischen Stellwerken verwendet werden [5]. Die Westermo-Gruppe mit Hauptsitz in Västerås (Schweden) ist ein weltweit agierender Spezialist für industrielle Datenkommunikation. Unter anderem bietet das Unternehmen ein komplettes Sortiment an Netzwerkprodukten für Bahnunternehmen mit den entsprechenden Zulassungen, wie beispielsweise EN 50121 für die elektromagnetische Verträglichkeit für alle Komponenten im Eisenbahnbereich. Für die Weiterentwicklung der Komponenten zu Krypto-Netzwerk-Modulen, die für digitale Stellwerke verwendet werden können, wurde zwischen Westermo und dem Technologieunternehmen Eviden eine Entwicklungspartnerschaft initiiert.

Die bisher verwendeten Netzwerk-Module wurden mit austauschbaren Hardware Security Modulen (HSMs) in Form von µSD-Karten des Anbieters Eviden Germany ausgestattet und so zu Krypto-Netzwerk-Modulen erweitert (Abbildung 4). Diese HSMs werden als BSI-zertifizierte Lösung seit über drei Jahren im deutschen Fiskalmarkt genutzt [6] und stellen eine wertvolle Erweiterung der nach EN 62443 entwickelten Netzwerk-Module von Westermo dar.

Eviden Digital Identity ist ein Anbieter, der aus dem Atos-Konzern hervorgegangen ist. Eviden Digital Identity schützt elektronische Identitäten mit kryptografischen Lösungen und Anwendungen für Secure Elements, digitale Identitäten und benutzerfreundliche Verschlüsselung. Zu Eviden Digital Identity gehören insbesondere die Marken cryptovision und IDnomic.

Während die Plattform der Krypto-Netzwerk-Module über längere Zeit unverändert in Betrieb bleiben sollen, können die HSMs bei Bedarf ausgetauscht werden. Dadurch können die verwendeten Krypto-Algorithmen und Schlüssel auf einfache Weise geändert werden, außerdem kann bei Bedarf der Speicher vergrößert werden. Unter anderem wird dadurch die geforderte Krypto-Agilität realisiert. Der kryptografische Kern der verwendeten HSMs ist gemäß Common Criteria EAL 4+ zertifiziert. Sie bieten 8 GB Flash Memory, wovon 6 GB beispielsweise für gesicherte Log Files zur Verfügung stehen.

Ergänzend zu einem HSM ist auch der Einsatz eines Trusted Platform Modules (TPM) möglich. Ein TPM ist ein in zahlreichen PCs, Smartphones und anderen Geräten fest eingebauter Hardware-Chip, der etwa die Leistungsfähigkeit einer Chipkarte aufweist. Da ein TPM nicht ohne Weiteres ausgetauscht werden kann, bietet es keine Krypto-Agilität. Da Krypto-Agilität inzwischen ein entscheidender Baustein für langlebige OT-Komponenten ist, kann ein TPM ein HSM nicht ersetzen. Es kann jedoch anderweitig verwendet werden. So lässt sich ein TPM durch die Nutzung von manipulationssicheren Hashwerten (digitaler Fingerabdruck) für den Schutz vor Malware-Angriffen einsetzen. Außerdem kann es sichere Firmware-Updates durch digitale Signaturen absichern, wobei es sich empfiehlt, ein hashbasiertes Signaturverfahren (zum Beispiel XMSS) zu nutzen. Hashbasierte Signaturverfahren sind quantensicher, und nach Ansicht des BSI so gut verstanden, dass keine anderweitigen Sicherheitslücken zu erwarten sind [7]. Hashbasierte Signaturverfahren sind zwar für den Alltagsgebrauch nicht praktikabel, jedoch für ein Firmware-Update, das nur alle paar Jahre stattfindet, durchaus geeignet.

Die Krypto-Netzwerk-Module sind unter anderem für die Verwendung in sogenannten Feldelementanschlusskästen (FeAK) entlang der Schienen vorgesehen. In den FeAK sind auch die Objectcontroller untergebracht, welche die Stelleinheiten oder Signalanlagen ansteuern und damit die Sensor- und Aktor-Ebene des digitalen Stellwerks abbilden. Diese Kästen sind unmittelbar im Gleisfeld so nah wie möglich an den Stelleinheiten verbaut und müssen den herausfordernden Umgebungsbedingungen standhalten, ohne dass es zu Kommunikationsausfällen kommt. Sie sind nur in geringer Weise physisch vor unbefugten Zugriffen und Umwelteinflüssen geschützt. Die Module sind daher durch verschiedene Maßnahmen gegenüber physischen Angriffen abgesichert. Sie sind zudem auf extreme Außentemperaturen ausgelegt sowie unempfindlich gegenüber Vibrationen und elektromagnetischen Interferenzen. Alle elektrischen Ports sind galvanisch getrennt. Der Energieverbrauch ist besonders gering, was eine geringe Wärmeentwicklung zur Folge hat und Lüftungsschlitze unnötig macht.

Der Betreiber kann die Krypto-Netzwerk-Module über ein zentrales Management-System steuern. Auf diese Weise wird das Ausrollen von Konfigurationen, Konfigurationsänderungen und Software-Updates bewerkstelligt. Zudem ist es möglich, die Komponenten über SSH auf sichere Weise direkt anzusprechen. Software-Updates sind durch digitale Signaturen geschützt.

Die Rollout- und Replacement-Prozesse sind aufgrund ihrer Zeitkritikalität nicht zu unterschätzen, da die Montagezeiten im Gleisfeld so kurz wie möglich gehalten werden müssen.

Da sich auf den Geräten teils sensible Daten befinden, ist eine sichere Lieferkette (secure Supply Chain) vom Chiphersteller über den Krypto-Modul-Hersteller, den Schaltschrankbauer und Systemintegrator bis zum Wartungspersonal aufrechtzuerhalten. Dies ist prozesstechnisch möglich, erfordert aber neue Prozesse, die in der Sicherungstechnik bisher noch nicht verbreitet sind.

5 Kryptografische Verfahren und Protokolle

Die in diesem Projekt verwendeten Krypto-Netzwerk-Module unterstützen zur Verschlüsselung und Authentifizierung das IPsec-Protokoll mit IKEv2 für den Schlüsselaustausch. Es wird also auf eine bewährte und standardisierte Technik gesetzt. Die Verschlüsselungsparameter entsprechen der TR-02102-3 des BSI [8], die Empfehlungen für die Verwendung von kryptografischen Mechanismen in den Protokollen IPsec und IKE gibt und dabei auf die Technische Richtlinie TR-02102-1 (Kryptographische Verfahren: Empfehlungen und Schlüssellängen) verweist [9].

Die IPsec-Verbindungen werden mit einem digitalen Zertifikat aufgebaut, das von einer Zertifizierungsstelle (CA) bereitgestellt wird (siehe unten). Zertifikate können gesperrt werden, wobei die Sperrprüfung mittels des OCSP-Protokolls oder einer Sperrliste möglich ist.

6 Public-Key-Infrastruktur

Die aktuelle Absicherung der Kommunikationskomponenten von Zügen (On-Board Units, OBUs) und Gleisfeldkomponenten (Radio Block Centre, RBCs) im European Train Control System (ETCS) erfolgt derzeit auf der Basis symmetrischer Schlüssel mit dem Verfahren Triple-DES. Zur Verwaltung der symmetrischen Schlüssel betreiben die jeweiligen Verkehrsunternehmen ein Key Management Center (KMC). Die Verteilung der Schlüssel erfolgt derzeit vornehmlich noch über Datenträger. Dies gilt für das Enrollment, für den Schlüsselwechsel, für Cross-Border-Traffic sowie für die Integration von Komponenten anderer Verkehrsunternehmen, deren Züge absehbar im hier betrachteten Kommunikationsbereich fahren werden.

Mit der Einführung von Public-Key-Infrastrukturen erreichen die Infrastrukturbetreiber und Verkehrsunternehmen eine erhebliche Vereinfachung des manuellen Aufwandes bei der Gewährleistung eines angemessen Sicherheitsniveaus. Folgende Ziele werden angestrebt:

- Kurz- bis mittelfristig kann die Verteilung der symmetrischen Schlüssel zwischen den KMC zweier Verkehrsunternehmen beziehungsweise zwischen dem KMC und der Kommunikationskomponente eines Unternehmens TLS-gesichert und mittels zertifikatsbasierter Authentisierung erfolgen.
- Mittel- bis langfristig kann die Absicherung der Kommunikation von einfachen symmetrischen Schlüsseln um zertifikatsbasierte Authentisierung der Komponenten untereinander und der Verwendung symmetrischer Sitzungsschlüssel erweitert werden.

Es ist geplant, X.509-Zertifikate und -Sperrlisten einzusetzen sowie eine Validierung über OCSP zu ermöglichen. Das Enrollment der Zertifikate auf die Kommunikationskomponenten sowie die Validierung der Zertifikate und Zertifikatssperrlisten sind gegenwärtig Gegenstand von Evaluierungsprojekten. Gleiches gilt für die Etablierung der Vertrauensbeziehungen der teilnehmenden PKIs.

7 Fazit und Ausblick

Das in dieser Fallstudie beschriebene Vorhaben ist ein anspruchsvolles IoT-Security und OT-Security-Projekt im Bereich der Kritischen Infrastrukturen

mit einer langen Laufzeit. Zu den Herausforderungen, die sich in diesem Projekt stellen, gehören die große Zahl der Komponenten und deren jahrzehntelanger Einsatz. Die entwickelte Lösung soll einerseits die notwendige Sicherheit und andererseits einen möglichst effektiven Betrieb ohne größeren Wartungsaufwand gewährleisten. Ein wichtiges Ziel ist es, diese Anforderungen mit Standard-Technologien zu erreichen und Insellösungen zu vermeiden.

Wie in allen langfristig angelegten IT-Projekten, in denen die Sicherheit eine Rolle spielt, müssen die Belange der Krypto-Agilität und der Migration auf Post-Quanten-Kryptografie berücksichtigt werden.

Bisher steht das Projekt noch in einem frühen Stadium. Ein erstes Teilprojekt wurde im Raum Stuttgart begonnen und soll 2024 in den Betrieb übergeben werden. Anschließend sollen im Laufe von mehreren Jahren bundesweit alle Stelleinheiten erfasst werden. Ähnliche Projekte gibt es in zahlreichen anderen Ländern, in denen die Digitalisierung des Schienenverkehrs ebenfalls ein wichtiges Ziel ist. Es ist zu erwarten, dass sich die in diesem Zusammenhang gemachten Erfahrungen auf zahlreiche andere IoT- und OT-Vorhaben übertragen lassen.

Literaturhinweise

[1] Poschinger, Richard; Sen, Adem: EULYNX Security – Standardisierung für gemeinsamen Schutz. Signal + Draht (114) 4 / 2022. S. 6-13

[2] Deutscher Bundestag: Digitale Schiene – aktueller Stand der Modernisierung der Leit- und Sicherungstechnik. Antwort der Bundesregierung auf eine Kleine Anfrage. Drucksache 19/30653. 29.07.2021

[3] Heinrich, Markus: RaSTA im Safety und Security Kontext. Digital Rail Summer School 2021.

[4] PQC Migration Guide. The essentials. Eviden 2023. https://www.cryptovision.com/en/download-access-2/

[5] Westermo: Data network solutions for the rail industry. https://www.westermo.de/industries/rail (abgerufen am 29.1.2024)

[6] Eviden: Sichere und praxistaugliche Speicherung von Transaktionsdaten, konform zu gesetzlichen Vorschriften. https://www.cryptovision.com/de/produkte/security-token-hardware-solutions/cryptovision-tse-v2/ (abgerufen am 29.1.2024)

[7] Bashiri, Kaveh; Kousidis, Stavros: Migration zu einer quantensicheren Verwaltungs-PKI. BSI Forum 5/2023

[8] Technische Richtlinie TR-02102-3, Kryptographische Verfahren: Verwendung von Internet Protocol Security (IPsec) und Internet Key Exchange (IKEv2). Bundesamt für Sicherheit in der Informationstechnik, 2023.

[9] Technische Richtlinie TR-02102-1, Kryptographische Verfahren: Empfehlungen und Schlüssellängen. Bundesamt für Sicherheit in der Informationstechnik, 2023.

[10] Schmeh, Klaus: Kryptografie – Verfahren, Protokolle, Infrastrukturen. Heidelberg: DPunkt-Verlag, 2007.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Cyber-Resilienz im Stromnetz: OT-Schwachstellenmanagement in der Praxis

Andreas Klien[1]

Kurzfassung:
Aufgrund ihrer großen Anzahl und ihrer oft unzureichenden Schutzmechanismen stellen Umspannwerke und Kraftwerke eine große Angriffsfläche für Cyberangriffe auf das europäische Stromnetz dar. In diesen Anlagen befinden sich jeweils dutzende kritische Schutz- und Steuergeräte, bei denen sich das Ausrollen von Security-Updates schwierig gestaltet. Dadurch rückt das effektive Schwachstellenmanagement für diese Anlagen in den Fokus, um die Cyber-Resilienz des Stromnetzes zu stärken.

Im Zuge des EU Cyber Resilience Act (CRA) und der EU NIS2 Direktive sind sowohl Hersteller als auch Energieversorger gefordert, ein proaktives und systematisches Schwachstellenmanagement zu betreiben und sicherheitsrelevante Ereignisse zu melden. Trotz dieser zukünftigen gesetzlichen Anforderungen sind nur wenige Energieversorgungsunternehmen (EVU) aktuell in der Lage, ein effektives Schwachstellenmanagement zu betreiben.

Das Patchen von Schutz- und Leittechnik im Stromnetz stellt aufgrund von Netzfreischaltungen und umfangreichen Tests vor der Implementierung eine besondere Herausforderung dar. Häufig ist ein vollständiges Testen aller Funktionen nicht praktikabel, was Energieversorger vor die schwierige Frage stellt, ob das operative Risiko eines Patches das Cyber-Risiko überwiegt. Dies unterstreicht die Bedeutung eines fundierten Schwachstellenmanagements und der daraus resultierenden Risikobewertung. Der Beitrag zeigt auf, wie komplex das Patchen und das Schwachstellenmanagement für EVU ist, gestützt durch Statistiken zu Security Advisories. Ein Fokus liegt auf dem Abgleich dieser Advisories mit den vorhandenen Assets, was durch die nötige Detailgenauigkeit des Asset-Inventory erschwert wird. Dabei wird die Rolle des Common Security Advisory Frameworks (CSAF) für eine verbesserte Verarbeitung und Verwaltung von Sicherheitshinweisen betont.

Der Beitrag präsentiert eine automatisierte Lösung für das OT-Vulnerability-Management, einschließlich der Nutzung des CSAF-Formats und eines Web-Crawlers zur Erfassung neuer Advisories von Herstellerwebsites, und diskutiert, wie automatisierte Prozesse die Cyber-Resilienz im Stromnetz verbessern können.

Stichworte: CSAF, Cyber Resilience Act, Digitalisierung der Energiewende, Energieanlagen, ICS/OT Sicherheit, NIS2Resilienz Kritischer Infrastrukturen, Schwachstellenmanagement, Stromnetz

[1] OMICRON electronics GmbH, Österreich

1 Anforderung Schwachstellenmanagement in Energieanlagen

Netzbetreiber und Betreiber von Energieanlagen sind nach dem IT-Sicherheitsgesetz, beziehungsweise dem dadurch geänderten Energiewirtschaftsgesetz dazu verpflichtet, Maßnahmen aus entsprechenden IT-Sicherheitskatalogen umzusetzen. Damit soll sichergestellt werden, dass Cyber-Risiken fortlaufend identifiziert, minimiert und im Idealfall verhindert werden. Netzbetreiber müssen ihr Informationssicherheitsmanagementsystem (ISMS) zertifizieren lassen und dafür den „Stand der Technik“ umgesetzt haben. Nach §11 Abs. 1e EnWG müssen sie ferner seit Mai 2023 Systeme zur Angriffserkennung einsetzen. *„Sie sollten dazu in der Lage sein, fortwährend Bedrohungen zu identifizieren und zu vermeiden sowie für eingetretene Störungen geeignete Beseitigungsmaßnahmen vorsehen“* [1].

Der "Stand der Technik" im Bereich des Schwachstellenmanagements für KRITIS-Betreiber wird nicht explizit definiert, sondern ergibt sich durch die Einhaltung bestimmter Normen und Richtlinien. Im IT-Grundschutzkompendium des BSI wird der Stand der Technik mit der Umsetzung der Standardmaßnahmen definiert. Hier ist insbesondere *IND.1.A12, „Etablieren eines Schwachstellen-Managements“* [2] als SOLL-Kriterium zu nennen, von dem eine Abweichung zumindest begründet werden muss. In der BSI-Orientierungshilfe zum Einsatz von Systemen zur Angriffserkennung (SzA) wird als MUSS-Kriterium gefordert, dass *„fortlaufend Meldungen der Hersteller (Hard- und Software), von Behörden, den Medien und weiteren relevanten Stellen geprüft werden und in dokumentierte Prozesse des Schwachstellenmanagements einfließen“* [3]. Eine weitere etablierte Definition des Standes der Technik wurde durch TeleTrusT, einen Verband für IT-Sicherheit, erstellt. Dort wird aufgeführt, dass *„Hersteller-Webseiten, CERTs, CVSS-Datenbanken, Mailing-Listen von Software- und Hardware-Herstellern, Newsgruppen von Drittanbietern, usw.“* überwacht werden müssen, um *„Schwachstellen, Softwarekorrekturen und Bedrohungen zu identifizieren“* [4].

Schließlich stellen die internationalen Normen ISO 27001 [5], ISO 27002 [6] und ISO 27019 [7] spezielle Anforderungen an das Management technischer Schwachstellen als Teil des Risikomanagements. Auch dort wird gefordert, dass Organisationen Informationen über Schwachstellen beschaffen, dass sie die eigene Exposition gegenüber diesen Schwachstellen bewerten und geeignete Maßnahmen ergreifen. Die ISO 27019 erweitert spezifisch für den Energiebereich die Anforderungen der ISO 27002 und bietet detaillierte Anleitungen für das Management technischer Schwachstellen, einschließlich der Pflicht, nach jeder relevanten Softwareinstallation, nach Upgrade oder Änderung, ein aktuelles Softwareinventar zu erhalten.

Um dies zu ermöglichen, ist es natürlich auch notwendig, dass die Hersteller von kritischen Schutz- und Steuergeräten die nötigen Informationen zeitgerecht zur Verfügung stellen. Hier stellt der EU Cyber Resilience Act (CRA) eine Reihe von Anforderungen in Bezug auf das Schwachstellenmanagement [8]: Hersteller sind verpflichtet, Schwachstellen innerhalb von 24 Stunden zu melden und sie tragen die Verantwortung, Sicherheitslücken über mindestens fünf Jahre oder während der erwarteten Lebensdauer eines Produkts zu identifizieren, zu beheben und entsprechende Sicherheits-Updates bereitzustellen. Während des gesamten Produktlebenszyklus müssen Hersteller die Schwachstellen ihrer Produkte effektiv managen, regelmäßige Tests durchführen und ein umfassendes Patch-Management ihrerseits vorweisen.

2 Warum ist das Patchen von Geräten im Stromnetz so schwierig?

In Umspannwerken kommunizieren mehrere Dutzend IP-Adressen miteinander und tauschen Messwerte, Statusinformationen und Befehle aus. Diese Schutz- und Leittechnikgeräte sind vor allem dafür verantwortlich, dass Hochspannungsleitungen und Transformatoren vor Kurzschlüssen geschützt werden und dass die Steuerbefehle der Netzleitstellen im Umspannwerk ausgeführt werden. Insgesamt kommen so bei einem Energieversorger Tausende von Geräten zusammen, die für hochkritische Aufgaben eingesetzt werden [9] [10] [11]. Dennoch werden Sicherheitslücken in Schutzrelais und Leittechnik oft nur unzureichend erkannt, und die Betreiber sind sich der Schwachstellen ihrer Geräte oft nicht bewusst [12].

Abschaltungen

Wie sonst auch in der Automatisierungsbranche üblich, erfordern Firmware-Updates die Abschaltung des Steuergerätes und damit auch des operativen Prozesses, der mit diesem Gerät gesteuert oder überwacht wird. In Produktionsanlagen ist es bereits eine Herausforderung, Zeitfenster zu identifizieren, um Updates einzuspielen und Testläufe durchzuführen. Im Stromnetz müssen analog dazu Freischaltungen von Betriebsmitteln wie Leitungen, Transformatoren und Leistungsschaltern bei der Netzführung beantragt werden. Angesichts der derzeitigen Überlastung der Übertragungsstrecken aufgrund der veränderten Netznutzung durch erneuerbare Energieerzeugung kann die Genehmigung für solche Freischaltungen oft mehrere Monate dauern [13]. Während dieser Zeit stehen oftmals wieder neue Patches zur Verfügung. Zudem kann das Wartungspersonal die Firmware-Updates nur an jenen Anlagenteilen durchführen, an denen die Betriebsmittel abgeschaltet wurden. Eine vollständige Abschaltung aller Anlagenteile und Betriebsmittel, die mit einer Schaltanlage gesteuert und überwacht werden, ist selten umsetzbar, weshalb zur Aktualisierung aller Geräte einer Anlage oft mehrere

Wartungsfenster notwendig sind. Daraus ergibt sich, dass Energieversorger zeitweise mehrere verschiedene Firmwarestände des gleichen Gerätetyps im Einsatz haben, unter Umständen kann dies sogar innerhalb einer Anlage der Fall sein.

Testbarkeit

Durch die Komplexität dieser Geräte sind unterschiedliche Verhaltensweisen zwischen Firmware-Versionen und Fehler, die selbst bei kleineren Updates in kritischen Schutzfunktionen eingeführt wurden, möglich. Vor der Implementierung einer neuen Firmware-Version in kritischen Schutzrelais oder Leittechnikgeräten ist daher eine umfassende Testung erforderlich. Energieversorger richten hierfür spezialisierte Labore ein, in denen versucht wird, die reale Anlagenumgebung von möglichst vielen Umspannwerken nachzustellen, um sämtliche Gerätefunktionen realistisch zu überprüfen. Geräte mit kombinierten Schutz- und Steuerfunktionen verfügen über tausende Einstellparameter, und die Komplexität der programmierbaren Steuerung ist beträchtlich. Da jedes Gerät im Feld spezifische Funktionen erfüllt und wenig Standardisierung existiert, ist eine zusätzliche Überprüfung nach einem Update auch vor Ort notwendig [13]. Während die Geräte in den Laboren extensiv getestet werden können, besteht die Herausforderung darin, das notwendige Ausmaß der Tests in der Anlage zu bestimmen. Eine vollständige Prüfung aller Schutzfunktionen, Steuerungsabläufe, Schaltlogiken und der Kommunikation zur Leitstelle wäre zu zeitaufwendig und würde eine Freischaltung der gesamten Anlage erfordern. Folglich verbleibt oft ein Restrisiko nach dem Durchführen von Firmware-Updates.

Da oft nicht ausreichend getestet werden kann, kann das operative Risiko größer sein, den Patch einzuspielen, als ihn nicht einzuspielen.

3 Personalaufwand für Schwachstellenmanagement für EVU

Für die Etablierung eines effektiven OT-Schwachstellenmanagements ist es für Energieversorger erforderlich, zunächst die Sicherheitshinweise (Security Advisories) der Gerätehersteller zu beschaffen. Leider können diese Sicherheitswarnungen nicht gesammelt von einer Quelle bezogen werden, sondern müssen von jedem Hersteller einzeln abonniert werden. Es gibt Initiativen, wie die des ICS-CERT [14], diese Hinweise gesammelt anzubieten, jedoch bilden diese die Komponenten in der Energieautomatisierung noch nicht vollständig ab. Daher müssen die Sicherheitshinweise direkt bei den Herstellern abgerufen oder abonniert werden.

Auf diese Weise veröffentlichen alleine drei führende Hersteller von Schutz- und Leittechnik zusammen jährlich über 300 Sicherheitshinweise [15] [16] [17]. Jeder dieser Hinweise adressiert mehrere Schwachstellen, wobei im Durchschnitt jeweils etwa ein Dutzend Gerätetypen betroffen sind.

2022	2023	2024	Gesamt	
262	311	51*	ca. 1112*	*Stand: 21.2.2024

Abbildung 1: Anzahl der Security Advisories veröffentlicht durch [15] [16] [17].

Die Identifikation, ob ein Security Advisory auf die eigenen Assets zutrifft, ist jedoch komplexer als die bloße Kenntnis der Modellbezeichnung und Firmware-Version des Gerätes. Oftmals ist die Schwachstelle nicht direkt in der Firmware des Hauptgerätes zu finden, sondern in der Firmware von Erweiterungskarten, die im Gerät eingebaut sind. Ein typisches Beispiel hierfür sind Ethernet-Kommunikationskarten, die in unterschiedlichen Geräten desselben Herstellers verwendet werden und verschiedene Firmware-Versionen aufweisen können. Allerdings enthalten Inventarverzeichnisse der Betriebsmittel in der Regel nur Angaben zum Gerätetyp und zur Hauptfirmware-Version, nicht aber zu Firmware-Versionen von Modulen innerhalb des Gerätes. Ein detailliertes Verzeichnis der Betriebsmittel, das auch diese Informationen umfasst, ist bei vielen Energieversorgern nicht vorhanden. Stattdessen sind relevante Informationen über die Betriebsmittel häufig über diverse Projektdateien und Excel-Tabellen verteilt und selten auf dem neuesten Stand.

Der Aufwand für das Schwachstellenmanagement hängt also davon ab, wie homogen die Betriebsmittel des Netzbetreibers hinsichtlich der Hersteller und der eingesetzten Firmwareversionen ist. Durch die verschiedenen Komponenten wie Schutzrelais, Leittechnik, Netzwerkswitches, Protokollumsetzer und GPS-Zeitserver ergeben sich automatisch eine größere Zahl an Herstellern, deren Schwachstellenmeldungen beachtet werden müssen. Demzufolge müssen hunderte neue Security Advisories pro Jahr überprüft werden. Die Überprüfung kann zwischen 10 Minuten (betroffene Geräteserie wird nicht eingesetzt) und über 8 Stunden (Gesprächsbedarf, ob betroffene Gerätetypen in dieser Konfiguration eingesetzt werden) betragen. Daraus ergeben sich viele Arbeitstage Aufwand pro Jahr, die nur Fachexpertinnen und -Experten durchführen können – alleine um abzuschätzen, ob eine Schwachstelle zutrifft oder nicht. Dieses Personal wird jedoch typischerweise für den Ausbau und Erhalt des Netzes dringend benötigt.

Automatisierte Lösungen und einheitliche Standards könnten hier eine große Arbeitsentlastung darstellen. In den folgenden Abschnitten wird daher

ein Lösungsansatz vorgestellt, in dem die oben beschriebenen manuellen Schritte für das OT-Schwachstellenmanagement – Inventarisierung, Sammeln der Advisories und Abgleich der Geräte mit den Advisories – automatisiert werden.

4 Automatisierte Betriebsmittelinventarisierung

Ein solches Betriebsmittelverzeichnis für tausende von Geräten zu erstellen ist nicht nur aufwendig, sondern auch fehleranfällig. Dieses Verzeichnis muss, je nach Größe des Energieversorgers, von typischerweise 10-20 oder mehr Technikerinnen und Technikern durch manuelle Eingaben aktuell gehalten werden.

Im Idealfall wird ein solches Betriebsmittelverzeichnis automatisiert erstellt. Hierfür gibt es die Möglichkeit, ohne aktive Kommunikation mit der Anlage (passiv) oder mit einer aktiven Abfrage der Komponenten zu arbeiten. Besonders bei der aktiven Abfrage müssen Cyber-Risiken, aber auch operative Risiken beachtet werden: Durch die unüblichen Abfragen, bzw. unübliches Client-Verhalten können die teils veralteten Geräte über Softwarefehler in der Protokollimplementierung zum Absturz gebracht werden.

Passive Erfassung der Geräteinformationen

Für die passive Erfassung von Geräteinformationen gibt es in der Energiebranche einige Vorteile gegenüber der Industrieautomatisierung: Es gibt Projektdateien, die solche Anlagen und ihre Komponenten im Detail beschreiben. Gerade bei modernen Anlagen sind diese nicht nur maschinenlesbar, sondern auch durch den IEC-61850-6-Standard [18] standardisiert. Security-Tools können die System-Configuration-Description-(SCD)-Dateien einlesen und Informationen über die meisten Betriebsmittel entnehmen.

Die Methode, Geräteinformationen passiv über mitgeschnittenen Netzwerkverkehr abzurufen wird oft im Marketing propagiert, ist jedoch in der Praxis wenig zielführend. In der normalen Leittechnikkommunikation werden Typinformationen und Firmwareversionen nicht übertragen, somit können diese Informationen auch nicht mitgeschnitten werden. Man müsste also einen anderen Client in der Anlage konfigurieren, genau diese Abfragen durchzuführen, was im Endeffekt einer aktiven Abfrage entspricht.

Aktive Abfrage der Geräteinformationen

In sehr vielen Umspannwerken und Kraftwerken unterstützen die Geräte die Protokolle in der IEC 61850 Normenreihe. Damit können Typenschilder aktiv über das Manufacturing-Message-Specification-Protokoll (MMS) abgerufen werden. MMS ist vergleichbar mit dem Simple Network Management Protocol (SNMP), jedoch hat es den entscheidenden Vorteil, dass die Datenrepräsentation über die IEC 61850 ebenfalls genormt ist [19]. Damit können die Typ-, Modell- und Firmware-Informationen von allen Geräten einheitlich abgerufen werden, während bei SNMP für jeden Gerätetyp eine andere Management Information Base (MIB) geladen werden muss und z.B. das Feld für Firmwareversion in jeder MIB auf eine andere Art realisiert ist. Leider sind es pro Hersteller mehrere verschiedene MIBs. Eine Inventarisierung durch SNMP erfordert also, dass man alle Gerätetypen bereits kennt und die jeweils richtige MIB laden kann. Abbildung 2 zeigt ein Beispiel, welche Parameter über das MMS-Protokoll aktiv und über die IEC 61850-SCL-Projektdateien passiv abrufbar sind: Name und Beschreibung des Geräts, Hersteller und Modellangabe, Firmwareversion und teilweise auch die Fabrikationsnummer, bzw. Hardware-Version der Komponente.

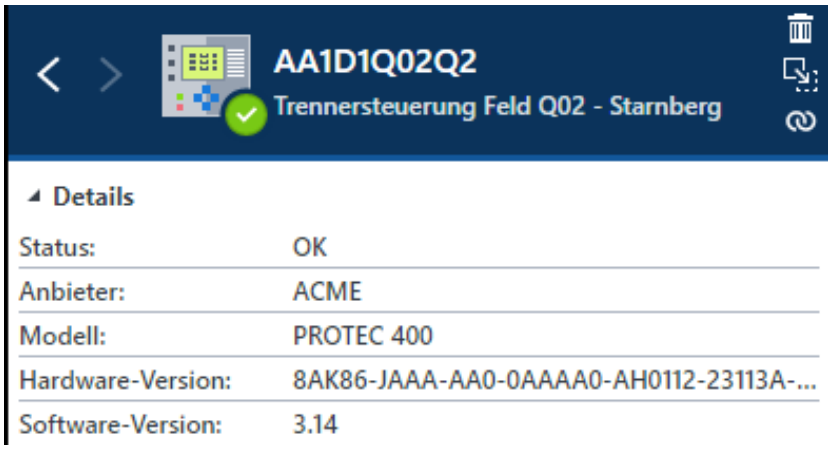


Abbildung 2: Beispiel für über MMS abrufbare Geräteinformationen

Unsere Forschung bestätigte, dass eine aktive Abfrage durch das MMS-Protokoll mit nicht-standardkonformen Client-Stacks durchaus die oben erwähnten Abstürze in Geräten auslösen kann. Obwohl MMS auf TCP/IP übertragen wird, bringt das Protokoll nochmal die OSI-Transportmechanismen mit, die eigentlich schon durch TCP erfüllt werden. Durch die Vielzahl an Handshakes mit jeweils ausgehandelten Verbindungsparametern und Window-Sizes ergibt sich eine große Komplexität, die häufig zu Softwarefehlern und Schwachstellen führt. Ein Indiz dafür ist die Anzahl der OT-Schwachstellen, die im Zusammenhang mit der MMS-Portnummer 102 bisher öffentlich gemacht worden und z.B. unter [20] zu finden sind. Es

empfiehlt sich daher, einen praxiserprobten und standardkonformen Client-Stack für diesen Zweck zu verwenden.

Ebenfalls können proprietäre Herstellerprotokolle verwendet werden, um Geräteinformationen abzurufen. Dies hat den Vorteil, dass auch Geräte abge rufen werden können, die den IEC-61850-Standard und SNMP-Standard nicht unterstützen. Hier sollten jedoch unbedingt die Herstellertools verwendet werden und keine durch „Reverse-Engineering" erstellten Implementierungen dieser Protokolle, wie z.B. NMAP-Scripts, wie in [21], da sonst Geräteabstürze möglich sind. In [22] wurden verschiedene aktive Abfragetechniken für OT-Geräte evaluiert und potenzelle Seiteneffekte erprobt.

Zusammenfassend gibt es folgende Möglichkeiten, Betriebsmittelinformationen automatisch zu erfassen [13]:

- *Passiv, über den Netzwerkverkehr:* Nur die Präsenz der Assets kann detektiert werden, meist keine Typ- oder Firmware-Informationen.
- *Import der Projektdatei als IEC 61850 SCD oder proprietäre Formate:* Liefert zuverlässige Ergebnisse für IEC-61850-Geräte, die dort angegebenen Firmwareversionen könnten jedoch bereits verändert worden sein.
- *Import der Anlagendokumentation und Inventarlisten:* Liefert Ergebnisse für alle angegebenen Geräte, die Firmwareversionen können jedoch veraltet sein.
- *Aktive Abfrage über das MMS-Protokoll:* Liefert genaue Ergebnisse, nicht immer verfügbar, unbedingt standardkonformen Client verwenden.
- *Aktive Abfrage über das SNMP-Protokoll:* Liefert genauer Ergebnisse, Implementierung gerätetypabhängig.
- *Aktive Abfrage über proprietäre Engineering-Protokolle:* Liefert genaue Ergebnisse, Implementierung gerätetypabhängig, unbedingt Herstellertools verwenden, daher oft nicht automatisierbar.

In unserer Lösung für das automatisierte OT-Schwachstellenmanagement setzen wir daher auf eine Kombination von mehreren Maßnahmen, um die bestmögliche Abdeckung zu erreichen [12]. Es werden IEC 61850- und CSV-Projektdateien importiert und es ist eine optionale aktive Abfrage über das MMS-Protokoll möglich. Wir arbeiten daran, dies mit weiteren aktiven Abfragemöglichkeiten gerätespezifisch zu erweitern.

5 Automatisch erstellte OT-Schwachstellendatenbank

Um ein automatisiertes Vulnerability-Management zu realisieren, werden die Schwachstelleninformationen in maschinenlesbarer Form benötigt. Wie oben beschrieben wurde, sollte man dafür mit Security Advisories der Hersteller arbeiten, da es ohne Software Bill of Materials (SBOM) nicht möglich ist, die Schwachstellen auf Komponentenebene den Geräten zuzuordnen. Da es für Security Advisories von OT-Herstellern leider noch keine zentrale Quelle gibt, setzen wir Web-Crawler ein, die die Herstellerwebsites auf neue oder geänderte Sicherheitshinweise scannen. Die gefundenen Informationen werden zeitgestempelt protokolliert und Änderungen an jedem Advisory werden über die Datenbank nachvollziehbar gespeichert. Nicht alle Hersteller veröffentlichen (alle) ihre Sicherheitshinweise auf ihrer Website. Daher wird die Datenbank zusätzlich manuell aus anderen Quellen erweitert.

Common Security Advisory Framework (CSAF)
Beim Aufbau unserer Schwachstellen-Datenbank spielt auch das Common Security Advisory Framework (CSAF) [23] eine wichtige Rolle. Das CSAF-Format ermöglicht es Herstellern, Sicherheitshinweise in einem maschinenlesbaren Format zu publizieren. Wir benutzen das CSAF auch als internes Speicherformat in der Datenbank.

Einige große Hersteller von Schutz- und Leittechnik veröffentlichten bereits Tausende Advisories in dem CSAF-Format. Dies erfolgt aber oft nur für die neuen Advisories, sodass bestimmte ältere Sicherheitshinweise nicht im CSAF-Format vorliegen. Für die Security Advisories, die nur als HTML- oder PDF- veröffentlicht wurden, erstellen unsere Web-Crawler teilweise automatisch die dazugehörigen CSAF-Dateien, die dann von Security Analysten überprüft werden. Abbildung 3 veranschaulicht die verschiedenen Quellformate, die vom Web-Crawler abgerufen und archiviert werden.

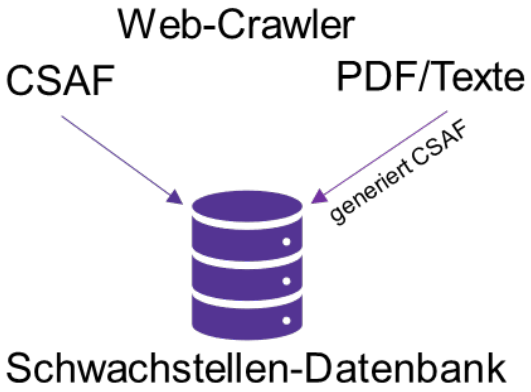


Abbildung 3: Schematische Darstellung der vom Web-Crawler eingesammelten Informationen

Geräte-Metainformationen

Für den Abgleich der Security Advisories mit dem Betriebsmittelinventar ist jedoch immer noch Expertenwissen notwendig, um zu beurteilen, ob der in der Sicherheitswarnung erwähnte Gerätetyp und die Modulkonfiguration mit entsprechender Firmware im Einsatz ist.

Um auch diese Herausforderung zu automatisieren, haben wir unser Expertenwissen in die Vulnerability-Datenbank als Metainformationen implementiert. Diese Metainformationen umfassen Gerätetypen, Module und deren Firmwareversionen, aber auch Regeln für das Zuordnen von Versionsnummern in Advisories, bis hin zur Marken-Historie von Gerätetypen, die unter verschiedenen Herstellerbezeichnungen erhältlich waren. Auf diesem Weg kann das System z.B. auch anhand der Fabrikationsnummer des Gerätes die Modulkonfiguration ermitteln und auch diese Informationen für die korrekte Zuordnung von Schwachstellen nutzen.

Abbildung 4 zeigt die Anzahl der Schwachstellen, die über diesen Prozess über die Security Advisories in unserer OT-Schwachstellendatenbank eingesammelt und mit Metainformationen versehen wurden, wobei Abbildung 5 die Verteilung nach Hersteller, bzw. Herausgeber der Security Advisories veranschaulicht.

Mit diesem Ansatz ist es uns schließlich möglich, automatisiert nur genau die Schwachstellen anzuzeigen, die für das jeweilige Gerät und seine Komponenten zutreffen.

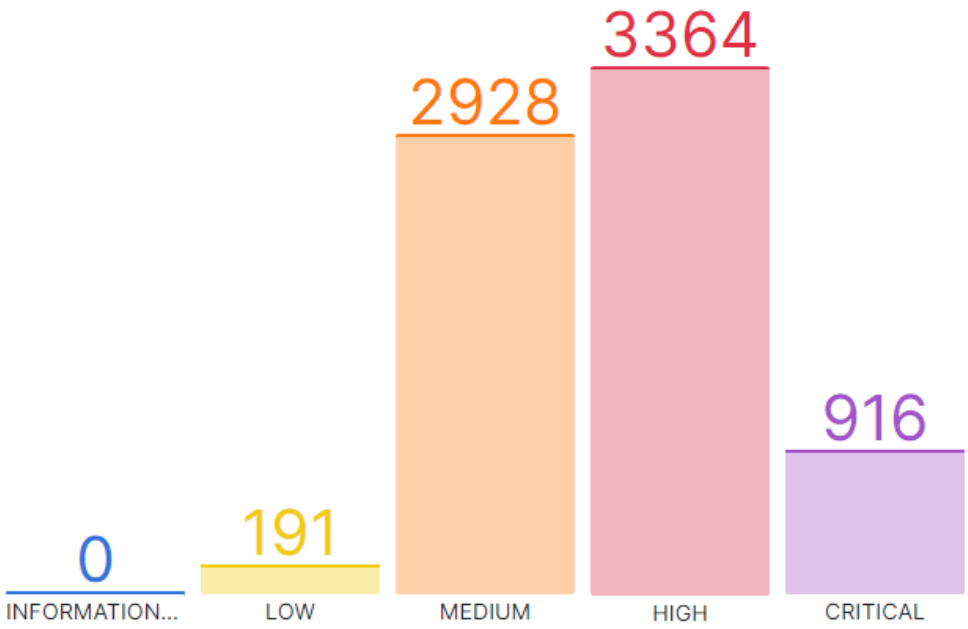


Abbildung 4: Anzahl der über Advisories eingesammelten Schwachstellen nach CVSSv3-Score

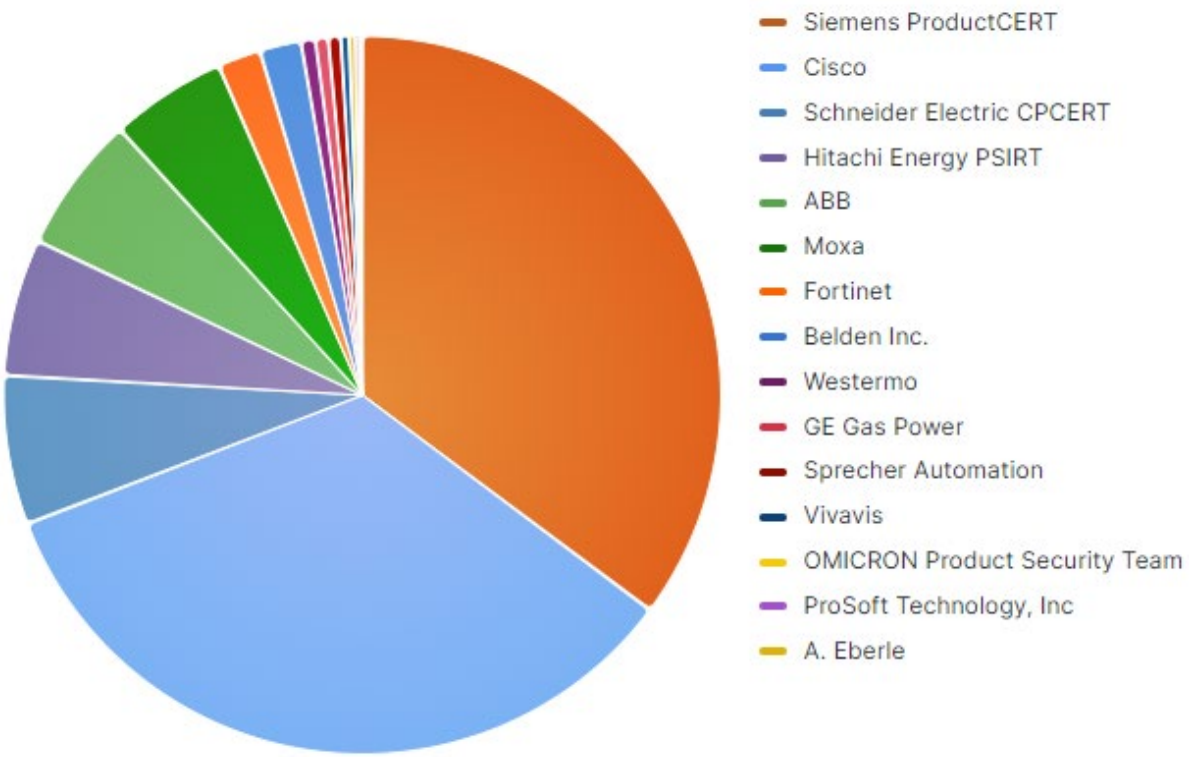


Abbildung 5: Verteilung der Schwachstellen nach Hersteller

6 Zusammenfassung und Ausblick

Das Management von Schwachstellen im Stromnetz, insbesondere bei Geräten der Schutz- und Leittechnik, stellt eine komplexe und anspruchsvolle Herausforderung dar. Die Automatisierung dieser Prozesse mittels maschinenlesbarer Anlagenbeschreibungen, des Common Security Advisory Framework (CSAF) und einer speziell angepassten OT-Schwachstellendatenbank erscheint als eine aussichtsreiche Methode, um Energieversorgungsunternehmen effektiv bei der Identifikation und Behebung von Sicherheitslücken zu unterstützen. Diese Ansätze ermöglichen ein effizientes Risikomanagement und unterstützen fundierte Entscheidungen hinsichtlich der Durchführung von Softwareaktualisierungen oder dem Verzicht darauf.

Aktuell basieren viele der Prozessschritte in der präsentierten automatisierten Lösung noch auf manueller Arbeit von Expertinnen und Experten, die die Daten von verschiedensten Quellen über Web-Crawler einsammeln und anschließend aufbereiten, was den Prozess verkompliziert. Die fortlaufende Verbesserung des Informationsaustauschs zwischen Energieversorgern, Herstellern und staatlichen Behörden wird daher eine Schlüsselrolle spielen,

um die Resilienz gegenüber Cyberbedrohungen zu erhöhen und die Sicherheit der Kritischen Infrastruktur nachhaltig zu gewährleisten.

Ein Schlüsselelement für die zukünftige Verbesserung dieser Prozesse könnte die Einführung und breite Akzeptanz des Software Bill of Materials (SBOM) sein. Der SBOM bietet eine detaillierte Auflistung der Komponenten, die in der Software enthalten sind, und spielt eine entscheidende Rolle bei der Transparenz und Sicherheit von Softwareprodukten. Für eine effektive Nutzung des SBOM im Rahmen des Schwachstellenmanagements ist es jedoch essenziell, dass Endbenutzer zusätzlich eine Vulnerability Exploitability eXchange (VEX)-Information erhalten. Diese Kombination ermöglicht es den Anwendern, sich gezielt auf relevante Sicherheitslücken zu konzentrieren, ohne sich mit potenziell Hunderten von Schwachstellen befassen zu müssen, die in ihrem spezifischen Kontext möglicherweise nicht ausnutzbar sind.

Literaturhinweise

[1] „Zweites Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme," Bundesgesetzblatt , Nr. Teil I Nr. 25, 2021.

[2] Bundesamt für Sicherheit in der Informationstechnik (BSI), „IT-Grundschutz-Kompendium (Edition 2023)," 2023.

[3] Bundesamt für Sicherheit in der Informationstechnik (BSI), „Orientierungshilfe zum Einsatz von Systemen zur Angriffserkennung (OH SzA)," 2022.

[4] Bundesverband IT-Sicherheit e.V. (TeleTrusT), „Handreichung zum Stand der Technik in der IT-Sicherheit," 2023.

[5] International Organization for Standardization (ISO) & International Electrotechnical Commission (IEC), "ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection - Information security management systems - Requirements," 2022.

[6] International Organization for Standardization (ISO) & International Electrotechnical Commission (IEC), "ISO/IEC 27002:2022, Information security, cybersecurity and privacy protection - Information security controls".

[7] International Organization for Standardization (ISO) & International Electrotechnical Commission (IEC), "ISO/IEC 27019:2017, Information technology - Security techniques - Information security controls for the energy utility industry," 2017.

[8] Europäische Kommission, „Vorschlag für eine Verordnung des Europäischen Parlamentes und des Rates über über horizontale Cybersicherheitsanforderungen für Produkte mit digitalen Elementen und zur Änderung der Verordnung (EU) 2019/1020," 2022.

[9] G. Wolf, "Managing the Power Grid's Vulnerabilities," T&D World, 10 December 2021. [Online]. Available: https://www.tdworld.com/smart-utility/article/21180723/managing-the-power-grids-vulnerabilities. [Accessed 21 February 2024].

[10] M. Faheem, S.B.H. Shah, et.al., "Smart grid communication and information technologies in the perspective of Industry 4.0: Opportunities and challenges," Computer Science Review, vol. 30, pp. 1-30, 2018.

[11] C.-C. Sun, A. Hahn and C.-C. Liu, "Cyber Security of a Power Grid: State-of-the-Art," International Journal of Electrical Power & Energy Systems, vol. 99, pp. 45-56, 2018.

[12] A. Klien, „Automatisiertes Schwachstellenmanagement: Effiziente Identifikation und Bewertung," atp magazin, Bd. 5, pp. 32-34, 2023.

[13] A. Klien, "Automated Vulnerability Management for the Power Grid," PAC World Magazine, no. 64, 2023.

[14] Cybersecurity & Infrastructure Security Agency, „Cybersecurity Alerts & Advisories," [Online]. Available: https://www.cisa.gov/news-events/cybersecurity-advisories. [Zugriff am 21 Februar 2024].

[15] Siemens, „Siemens Security Advisories," [Online]. Available: https://www.siemens.com/global/en/products/services/cert.html#SiemensSecurityAdvisories. [Zugriff am 21 Februar 2024].

[16] Hitachi Energy, „Cybersecurity Alerts and Notifications," [Online]. Available: https://www.hitachienergy.com/products-and-solutions/cybersecurity/alerts-and-notifications. [Zugriff am 21 Februar 2024].

[17] Schneider Electric, „Security Notifications Archive," [Online]. Available: https://www.se.com/ww/en/work/support/cybersecurity/security-notifications-archive.jsp. [Zugriff am 21 Februar 2024].

[18] International Electrotechnical Commission, IEC 61850-6:2009, Communication networks and systems for power utility automation - Part 6: Configuration description language for communication in electrical substations related to IEDs, 2009.

[19] F. Steinhauser und A. Klien, „How standards improve substation cybersecurity," PAC World Magazine, 2021.

[20] speedguide.net, „Known Port Assignments and Vulnerabilities," [Online]. Available: https://www.speedguide.net/port.php?port=102. [Zugriff am 21 Februar 2024].

[21] ICSecurity Project on github, „Nmap NSE Scripts," [Online]. Available: https://github.com/xl7dev/ICSecurity/blob/master/ICS_Nmap_Port.md. [Zugriff am 21 Februar 2024].

[22] J. Helms, B. Salazar, P. Scheibel and M. Engels, "Safe Active Scanning for Energy Delivery Systems Final Report," U.S. Department of Energy, 2017.

[23] OASIS Open Europe Foundation, „Common Security Advisory Framework," [Online]. Available: https://oasis-open.github.io/csaf-documentation/ [Zugriff am 21 Februar 2024].

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Cybersicherheit für die Schifffahrt – mit einer Schiffsbrücke als Test- und Entwicklungslabor

Philipp Sedlmeier[1], Dr. rer. nat. Jan Bauer[2], Dr.-Ing. Anisa Rizvanolli[1], Dr. rer. pol. Ole John[1]

Kurzfassung:
Angriffe auf Kritische Infrastrukturen, wie zum Beispiel auf die Energieversorgung, die Informations- und Kommunikationstechnologie, aber auch auf die Transport- und Logistikbranche finden nicht erst seit heute statt und verlagern sich mehr und mehr in den Cyberspace. Die geopolitischen Veränderungen der jüngsten Zeit haben die Verwundbarkeit dieser Infrastrukturen besonders deutlich gemacht. Dies gilt auch für die maritime Logistik, beziehungsweise Schiffe als deren unverzichtbares Rückgrat.
Die fortschreitende Digitalisierung, Automatisierung und Vernetzung der Systeme an Bord von Schiffen erfordern daher Maßnahmen, mit dem erhöhten Risiko für Cyberangriffe umzugehen. Sie umfassen nicht nur die Entwicklung geeigneter Schutzmaßnahmen, sondern auch Methoden zur Erkennung von Angriffen, zur fundierten Erprobung existierender Cybersicherheitskomponenten auf der Brücke und im Maschinenraum, und zu speziellem Training für die Schiffsbesatzung.
Für all diese Aspekte ist eine Testumgebung von Vorteil, welche die echten Bedingungen möglichst originalgetreu abbildet und es ermöglicht, Cyberangriffe einfach und benutzerfreundlich durchzuspielen. Trotz des akuten Bedarfs an einer solche Umgebung sind "Trainingsplätze", die solche Anforderungen erfüllen, nur spärlich gesät. In diesem Artikel wird deshalb ein Labor präsentiert, das reale Schiffsbrücken-Hardware mitsamt dazugehöriger Antennenplattform mit digitalen Hilfsmitteln kombiniert, um Test- und Trainingsszenarien für die maritime Cybersicherheit zu entwickeln. In diesem Labor wird untersucht, welche konkreten Auswirkungen Cyberangriffe nach sich ziehen, wie eine Schiffsbesatzung auf mögliche Angriffe reagieren sollte und wie integrierte Schiffsbrücken cybersicher gestaltet werden können.

Stichworte: Brückenangriffe (Schifffahrt), GPS-Störungen, Maritime Cybersicherheit, Navigationssysteme, RADAR Attacken, Schiffsbrücke

1 Cybersicherheit im maritimen Umfeld

Die maritime Logistik bildet das Rückgrat der globalen Wirtschaft – der mit Abstand größte Teil des Welthandels findet über den Seeweg statt [1]. Da fast zwei Drittel der weltweiten Versorgung mit Erdöl und anderen flüssigen Energieträgern per Schiff transportiert werden, sind die meisten globalen Lieferketten vom maritimen Sektor abhängig.

[1] Fraunhofer CML – Ship and Information Management, Hamburg
[2] Fraunhofer FKIE – Cyber Analysis & Defense, Wachtberg

Seitdem in Deutschland Erdgasimporte über den Seeweg die Lieferungen aus Russland ersetzen sollen oder müssen, wird diese nationale Abhängigkeit in Zukunft eher noch stärker. Seit etwa einem Jahr sind daher auch LNG-Terminals als Kritische Infrastruktur eingestuft[3]. Aber auch die LNG-Tanker sind für die Versorgung kritisch – insbesondere unter dem Gesichtspunkt, dass die weltweite Flotte im Moment lediglich um die 700 Schiffe umfasst und selbst die beauftragten Neubauten den Bedarf absehbar nicht vollständig decken werden [2].

Die insgesamt fortschreitende Vernetzung und Abhängigkeit von digitalen Technologien stellt jedoch eine große Angriffsfläche für Cyberattacken dar, nicht nur für Unternehmen an Land, sondern auch für alle möglichen Schiffstypen. Solche Angriffe können direkt auf Brücken- oder Navigationssysteme abzielen, oder diese zufällig treffen. Beide Fälle können zu schweren Konsequenzen für die Schiffsbesatzung, die Umwelt oder sogar für die globale Wirtschaft führen – beispielsweise im Falle einer Blockade des Suezkanals wie durch die *Ever Given* im März 2021.

Schwachstellen zu identifizieren, präventive Sicherheitsvorkehrungen zu treffen sowie Cybervorfälle zu erkennen und angemessen darauf zu reagieren ist angesichts der wachsenden Bedrohungslage und sich kontinuierlich weiterentwickelnder Cyberangriffe eine herausfordernde Aufgabe, jedoch wesentlich für einen sicheren und ungestörten Schiffsbetrieb.

Um diesen Herausforderungen zu begegnen, hat die *Internationale Seeschifffahrts-Organisation* (IMO) die Reedereien aufgefordert, Cyberrisiken seit spätestens 2021 in ihren *Safety Management Systems* (SMSs) zu behandeln [3]. Durch den *International Safety Management Code* (ISM-Code) sind Reedereien von allen Fahrgast- und von größeren Frachtschiffen verpflichtet, darin Verfahrensweisen für einen sicheren Schiffsbetrieb einzuführen, Risiken für Schiffe, Personal und Umwelt zu identifizieren und Sicherheitsmaßnahmen dagegen zu ergreifen [4]. Durch die Resolution der IMO werden dabei nun Cyberrisiken mit “klassischen” Sicherheitsrisiken, wie dem Austritt giftiger Gase bei der Tankreinigung oder dem Reißen eines Ankertaus, gleichgestellt. Dies macht die Entwicklung von Vorgehensweisen und Maßnahmen für wirkungsvolle Reaktionen auf Cyberangriffe auch auf operationeller Ebene erforderlich.

Mit den beiden *Unified Requirements* E26 und E27 [5, 6] hat die IACS umfangreiche Vorschriften veröffentlicht, die ab Juli 2024 für die Klassifizierung von

[3] Dritte Verordnung zur Änderung der BSI-Kritisverordnung (BGBl. 2023 I Nr. 53 vom 01.03.2023)

Schiffsneubauten verpflichtend sein werden. Diese Anforderungen verlangen die Einbeziehung von Aspekten der Cybersicherheit praktisch über die gesamte Lebensdauer des Schiffes, das heißt bereits ab der Entwicklungsphase.

Dabei werden durch Cyberattacken verursachte Schäden üblicherweise nicht von den gängigen Versicherungen in der Schifffahrt abgedeckt. Erst in jüngster Zeit wurden erste Cyberversicherungen für den maritimen Bereich aufgelegt [7]. Um solch eine Versicherung abschließen oder in Anspruch nehmen zu können, müssen Schiffseigner allerdings ebenfalls präventive Maßnahmen zur Erhöhung der Cybersicherheit ergreifen.

Indes haben Schiffe eine sehr lange Lebensdauer. Momentan ist die globale Handelsflotte im Schnitt über 22 Jahre alt [8] und viele dieser Schiffe werden mindestens noch über das nächste Jahrzehnt hinweg in Betrieb sein. Zur Bauzeit dieser Schiffe fand deren Cybersicherheit häufig jedoch noch keine Beachtung. Um also eine Strategie für die maritime Cybersicherheit [9] erfolgreich und wirkungsvoll gestalten zu können, muss diese jeweils auf das Alter eines Schiffes und seiner Systeme abgestimmt sein.

Der Trend zu autonomen Fahrzeugen hat auch den maritimen Sektor ergriffen; auch größere Schiffe wie die japanische Autofähre *Soleil* oder das norwegische Containerschiff *Yara Birkeland* können bereits jetzt oder in nächster Zukunft autonom fahren [10]. Es gibt auch bereits erste Richtlinien für den Einsatz automatisierter Prozesse, zum Beispiel in der Navigation; für autonome Schiffe. Im Allgemeinen sind diese jedoch erst in der Erarbeitung. Gerade die Sicherheit von Antennen, Sensoren, IT-Systemen und Datenverbindungen ist in diesem Zusammenhang besonders wichtig [10].

Dennoch werden Menschen auch für den Betrieb autonomer Schiffe notwendig sein, die bemannte Schifffahrt wird also in den nächsten Jahrzehnten weiterhin die vorherrschende Rolle spielen [11, 12]. Aber auch in diesen Fällen spielt Cybersicherheit eine große Rolle – zum Beispiel, wenn Lotsen nicht mehr an Bord kommen, sondern ihre Aufgaben aus der Ferne mittels geeigneter Telesysteme übernehmen [13]. Der Ausbildung von Seefahrern mangelt es jedoch bisher an Aufmerksamkeit für die Cybersicherheit [14]. Inwiefern solche Aspekte in das STCW-Übereinkommen, also in die Ausbildung von Seeleuten integriert werden sollten, wird momentan diskutiert [15].

2 Angriffsszenarien

Es ist nur in wenigen Fällen vorgeschrieben, Cybervorfälle an zuständige Behörden zu melden (zum Beispiel in Norwegen [16]); eine universelle Ver-

pflichtung dazu besteht nicht. Die Dunkelziffer, d.h. die Anzahl der ungemeldeten Vorfälle, ist also vermutlich sehr hoch. Doch auch auf anderen Wegen lässt sich ein Zuwachs an Cybervorfällen in den letzten Jahren bestätigen – beispielsweise durch Berichte von Mitarbeitern und Beratungsfirmen aus der Branche, der Anzahl von Stellenausschreibungen maritimer Firmen im Bereich Cybersicherheit oder wissenschaftlichen Beiträgen auf diesem Gebiet, wie beispielsweise [17, 10].

Betrachtet man die Typen der bekannt gewordenen Cybervorfälle im maritimen Umfeld, dann fällt auf, dass knapp die Hälfte der Vorfälle (47 %) auf Infektionen mit Viren oder Malware zurückzuführen ist [18]. Weitere 39 % sind "konventionelle" Cyberangriffe, wie z.B. *Denial of Service* (DoS)-Attacken. Es lässt sich annehmen, dass ein Großteil dieser Angriffe ungeschützte Dienste oder Schnittstellen betrifft, die eher zufällig durch ungezielte Malware "getroffen" werden, und nicht speziell gegen konkrete maritime Systeme, Geräte oder digitale Infrastrukturen gerichtet ist.

Funk-basierte Angriffe zur Störung und Manipulation des *Globalen Navigations-Satelliten-Systems* (GNSS) oder dem *Automatic Identification System* (AIS) dagegen machten in der Vergangenheit nur etwa 4 % der Vorfälle aus. Diese beinhalten in der Regel zwar eher aufwendigere, aber durchaus gefährliche Angriffe aus dem elektromagnetischen Spektrum gegen die Sensoren eines Schiffes oder das Integrierte Brückensystem (IBS) selbst, die keinesfalls vernachlässigt werden dürfen. Die Wissenschaft im Themenfeld der maritimen Cybersicherheit hat bereits vor einigen Jahren auf diese Gefahren aufmerksam gemacht und in Praxisexperimenten mögliche Auswirkungen solcher Angriffe demonstriert, siehe beispielsweise [19, 20].

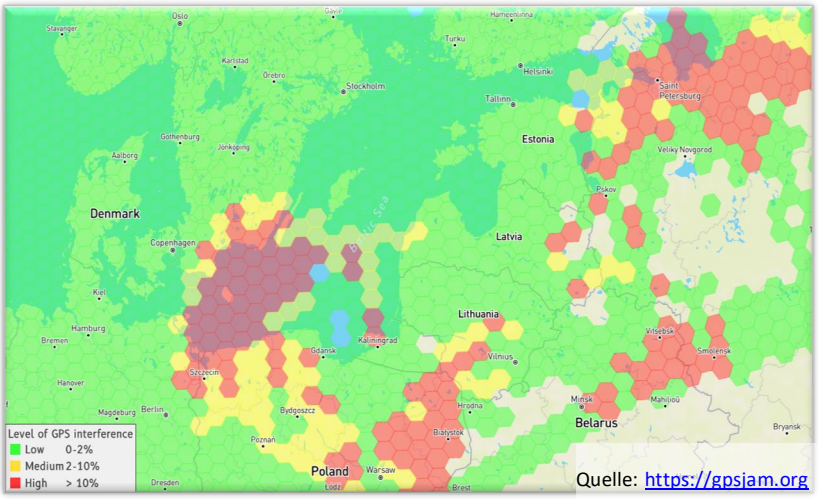


Abbildung 1: Großflächige Störungen von GPS im Ostseeraum (Stand Dezember 2023) [22].

Im Ostseeraum treten zudem in jüngster Vergangenheit Störungen der Navigationssatellitensysteme wie GPS gehäuft auf, wie Untersuchungen des Open Source Intelligence (OSINT)-Projekts GPSJAM.org [21] zeigen, das basierend auf ADS-B-Signalen aus dem Luftfahrtbereich weltweit Unregelmäßigkeiten bei der Satellitennavigation erfasst und die gewonnenen Informationen auf der Weltkarte tagesaktuell visualisiert, wie z.B. in Abbildung 1 für das Baltikum gezeigt. Die auftretenden Unregelmäßigkeiten, deren Ursachen und Hintergründe noch unklar sind, mutmaßlich aber in Zusammenhang mit der russischen Invasion in der Ukraine stehen, haben teilweise empfindliche Störungen unter anderem für den Fährverkehr zur Folge [22, 23].

Für die Untersuchung solcher Angriffe sind bisher weltweit nur sehr wenige Testumgebungen vorhanden. Was in existierenden Testumgebungen ebenso wenig Beachtung findet, sind Angriffe, die sich den menschlichen Faktor direkt zunutze machen. Dies umfasst die Gewinnung von persönlichen und sensiblen Informationen wie Passwörtern oder Zugangsdaten durch *Social Engineering*, insbesondere *Phishing*, aber auch Bedienungsfehler oder sogar vorsätzliche Handlungen mit böswilliger Absicht.

Meist sind derartige Techniken und Vorkommnisse jedoch nur der erste Schritt zur Vorbereitung im Rahmen einer sogenannten *Kill-Chain*, also eines größeren, komplexer angelegten Angriffs, sodass viele dieser Vorfälle bei der Erfassung in Datenbanken letztendlich anderen Kategorien zugeordnet werden. Dies wäre zum Beispiel der Fall, wenn ein argloser oder böswilliger Benutzer ein mit Schadsoftware infiziertes Speichermedium an ein Brückensystem anschließt und dadurch die Schadsoftware auf dem System verbreitet, sodass diese im Anschluss für eine DoS-Attacke ausgenutzt werden kann. Und solche Fälle sind keine Seltenheit, zum Beispiel weil Privatgeräte mit einem USB-Anschluss der Brückensysteme verbunden werden, um die Akkus der Geräte aufzuladen.

Folglich dürfen durch menschliche Fehler ausgelöste Cybervorfälle, auch wenn sie lediglich einen geringen Teil der erfassten Angriffe (beispielsweise 6 % in der ADMIRAL-Datenbank [18]) ausmachen, keinesfalls in Testumgebungen außer Acht gelassen werden. Gartner schätzt, dass nächstes Jahr mehr als die Hälfte der bedeutenden Cybervorfälle durch mangelnde Begabung oder menschliche Fehler („lack of talent or human failure“) verursacht werden wird [24].

In der Schifffahrt, und insbesondere auf der Brücke, kommt als menschlicher Faktor noch der Stress in kritischen Situationen sowie häufig damit einhergehend die Übermüdung der Seeleute hinzu [25]. Dabei lassen viele Situationen nur wenig Raum für Fehler. Für die Havarie der *Ever Given* im Suezkanal

waren Verständigungsprobleme und widersprüchliche Angaben der Lotsen mitursächlich [26] – und nur wenige Minuten Verwirrung waren ausreichend, den Welthandel über mehrere Tage zu blockieren. Es ist nicht schwer, sich vorzustellen, dass diese Verwirrung beim nächsten Mal nicht durch schwierige Wetterbedingungen, sondern durch fehlerhafte und widersprüchliche Geräteanzeigen aufgrund eines Cyberangriffs ausgelöst werden könnte. Deshalb ist es essenziell, integrierte Brückensysteme einerseits gegen Cyberangriffe zu härten, andererseits aber auch potenzielle Cyberangriffe möglichst frühzeitig zu erkennen, sodass die Schiffsbesatzung angemessen darauf reagieren kann.

3 Test- und Entwicklungsumgebung

Die Praktikabilität und der tatsächliche Nutzen der neuen Vorschriften zur Cybersicherheit aus Kapitel 1 muss sich erst noch erweisen. Zunächst bieten die Vorschriften lediglich ein Rahmenwerk – sie schreiben vor, *was* getan werden muss, aber nicht, *wie*. Wie sich die konkrete Ausgestaltung in Prozeduren möglichst effektiv umsetzen lässt, muss untersucht werden. Dazu wiederum schreibt die Untersuchungsstelle für maritime Unfälle Dänemarks: „Safety procedures should be acknowledged as bureaucratic tools that do not create safety by their mere existence. Procedures manage risk, but do not necessarily create safety" [27]. Deshalb ist es von Vorteil, wenn man die Wirksamkeit dieser Prozeduren bereits vor dem Eintritt des Ernstfalls untersuchen kann. Die gezielte und kontrollierte Ausführung oder Simulation von Cyberangriffen bedarf einer speziellen Umgebung, um deren konkrete Auswirkungen zu erforschen und den Umgang mit und die Abwehr von solchen Angriffen auf technischer und menschlicher Ebene zu untersuchen und zu trainieren. Um diese Problematik anzugehen, wird am Fraunhofer CML in Hamburg ein einzigartiges *Maritimes Cybersicherheits-Labor* entwickelt (vgl. auch [28]).

3.1 Laboreinrichtung

Einerseits ist dieses Labor mit denselben Hardware-Komponenten eines IBS ausgestattet wie ein reguläres Schiff (vgl. Abbildung 2); die Komponenten

wurden dabei gemäß den Vorgaben der Klassifizierungsgesellschaften eingebaut. Andererseits umfasst das Labor ein Portfolio an Software-Lösungen, die es ermöglichen, Cyberangriffe auf die Hardware tatsächlich durchzuführen.

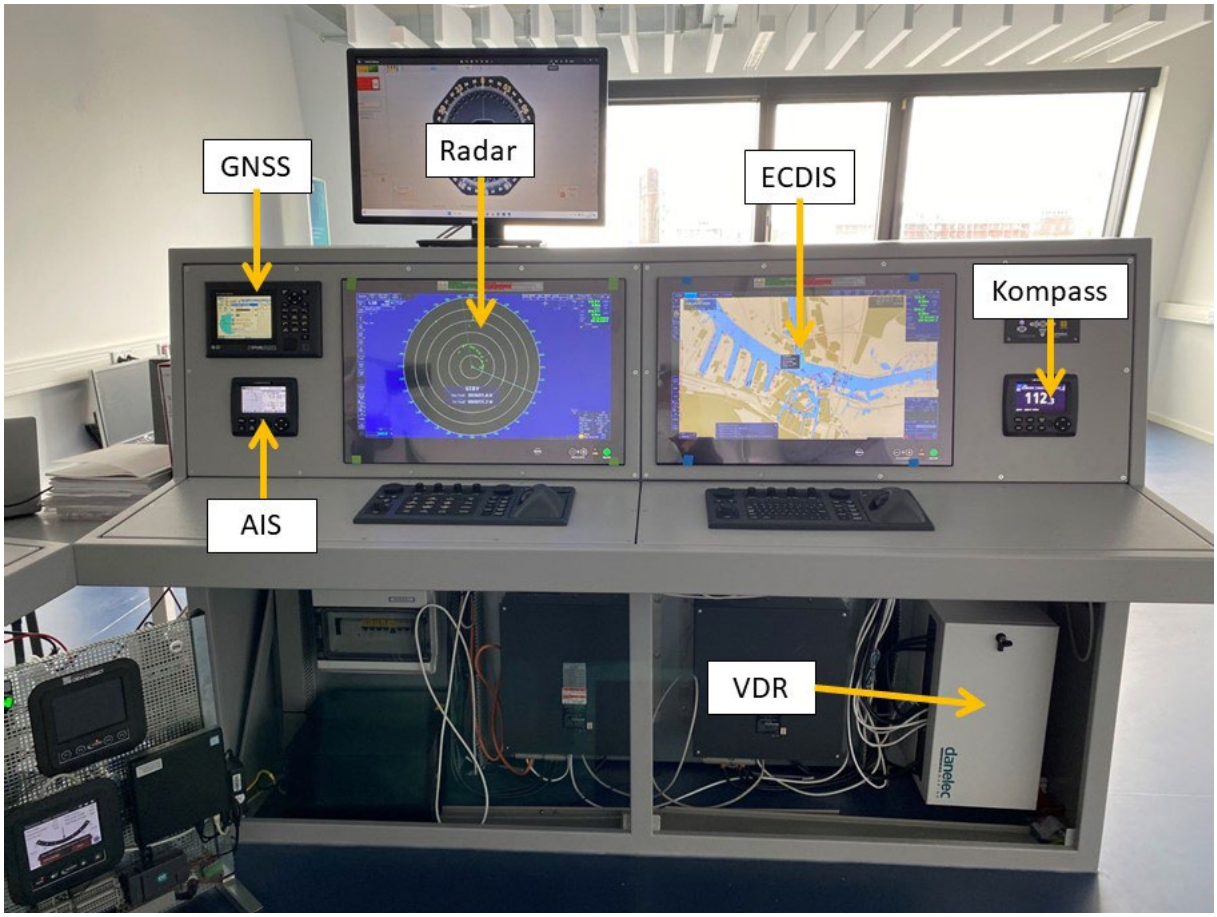


Abbildung 2: Labor-Schiffsbrücke mit Radar, ECDIS, AIS, GNSS, VDR und Satelliten-Kompass.

Bei den in Abbildung 2 gezeigten Geräten handelt es sich um

- eine Radarkonsole (Furuno Chart Radar FAR-3220 NXT),
- eine elektronische Seekarte (Furuno ECDIS FMD-3200BB),
- ein GPS-Navigationsgerät (Furuno GP-170),
- einen AIS-Transponder (Furuno FA-170),
- einen Satellitenkompass (Furuno SC-70),
- und einen Voyage Data Recorder (Danelec DM 100 VDR).

Die Verwendung echter Hardware erlaubt es ForscherInnen, EntwicklerInnen und AnwenderInnen aus dem maritimen Bereich gleichermaßen, zugeschnittene Cyberangriffe kontrolliert und replizierbar durchzuführen, zu untersuchen und auszuwerten – und das, ohne ein Schiff dafür in die Werft schicken zu müssen.

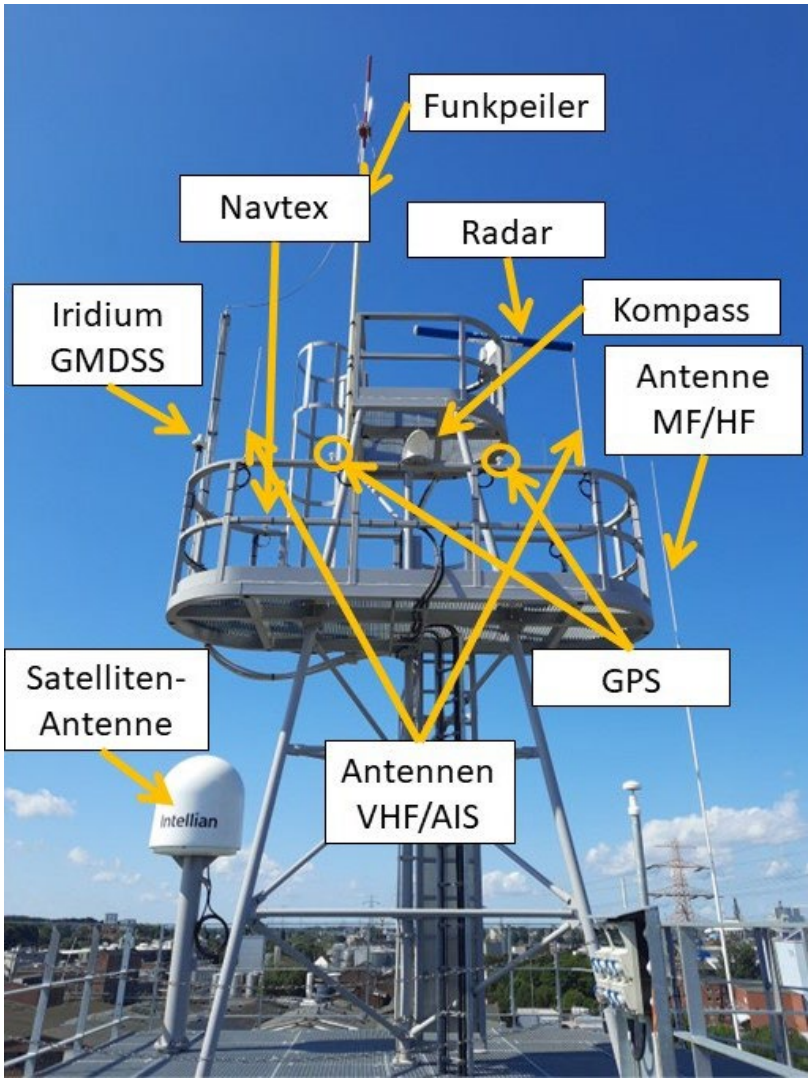


Abbildung 3: Plattform mit den zum Brückenlabor gehörigen Antennen.

Zusätzlich verfügt das Brückenlabor auch über ein MF/HF-Funkgerät (Furuno FS-1575), ein VHF-Funkgerät (Furuno FM-8900S), einen Navtex-Empfänger (Furuno NX-700B), ein Terminal für Iridium-Satellitenfunk (LT-3100S) und einen Funkpeiler mit zwei Kanälen. Weiterhin sind eine NMEA Interface Matrix der Firma Veinland und ein Furuno Intelligent Hub-3000 verbaut, sodass auch auf serielle Daten im Netzwerk einfach zugegriffen und diese untersucht werden können.

Diese Geräte sind – bis auf VDR und ECDIS – jeweils mit einer Antenne auf dem Dach des CML-Gebäudes verbunden (vgl. Abbildung 3). Die Verbindungen zwischen diesen Antennen und dem IBS, sowie der Geräte untereinander sind schematisch in Abbildung 4 dargestellt. Diese Anlagen können, wie bei einem richtigen Schiff, von der Brücke aus bedient werden und rund um die Uhr Realdaten aus der Hafenumgebung liefern. Dadurch verfügt das Labor

ständig über aktuelle und echte Daten, ohne sich diese von extern beschaffen oder simulieren zu müssen. Das Labor ermöglicht daher auch Anwendungen wie marFM®, ein System zur Spracherkennung und zur Sprachtranskription maritimer Funksprüche [29, 30].

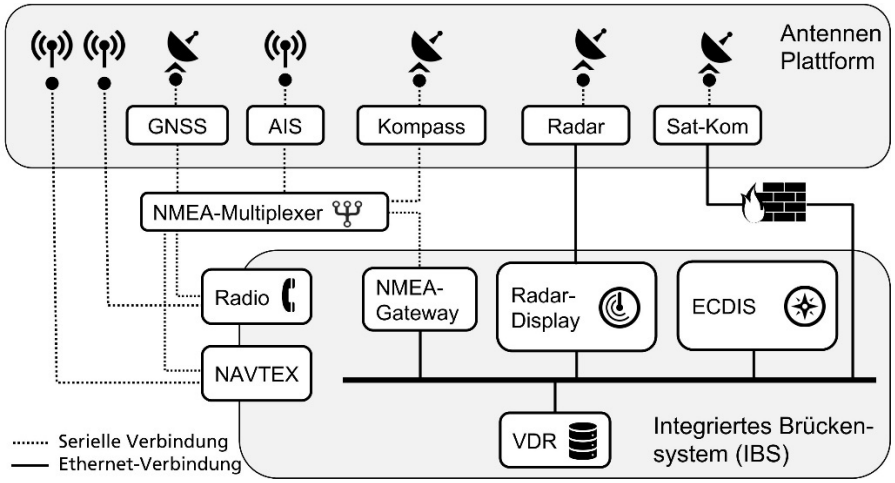


Abbildung 4: Verfügbare Hardwarekomponenten und Netzwerkarchitektur der Verbindung von Antennenplattform mit IBS.

3.2 Softwarelösungen

Die entwickelten Softwarelösungen dienen sowohl einer einfach bedienbaren Durchführung von Angriffen als auch der Entwicklung reproduzierbarer Testszenarien. Dadurch soll die Aufmerksamkeit des maritimen Personals gegenüber den Auswirkungen von Cyberrisiken geweckt werden. So können gemeinsam Reaktionsstrategien entwickelt und das erarbeitete Verhalten im Falle eines Cyberangriffs eingeübt werden.

Als software-seitige Grundlage der neu aufgebauten Test- und Entwicklungsumgebung dient ein am Fraunhofer FKIE entwickeltes, virtualisiertes Labor, welches die Simulation und Analyse maritimer Cyberattacken und deren Auswirkungen ermöglicht. Wesentliche Komponenten dieses Labors werden im Folgenden beschrieben.

3.2.1 Brückenangriffs-Werkzeuge

Das *BRidge Attack Tool* (BRAT) [31] ist der eigentliche Kern des Sicherheitslabors und bietet verschiedene Implementierungen von netzwerk-basierten Cyberangriffen gegen maritime Systeme. Es umfasst eine zweckmäßige grafische Benutzeroberfläche, die es ermöglicht, zahlreiche Angriffe auf das zu testende maritime System interaktiv auszuwählen, zu konfigurieren, zu kombinieren und zu planen.

Weil die nautische Kommunikation mit der Standardisierung nach IEC 61162-450 und NMEA 1803 weder authentifiziert noch verschlüsselt ist, ermöglicht dies eine Analyse der Netzwerktopologie und das Scannen nach aktiven Geräten auf der Grundlage von passivem Netzwerk-Sniffing. Die durch das Abhören gewonnenen Informationen können dann genutzt werden, um aktiv einfache, komplexe oder subtile Cyberangriffe gegen typische Einrichtungen innerhalb des IBS auszuführen, beispielsweise durch die in Kapitel 2 angesprochene Manipulation von GNSS-Daten.

Erweitert wird BRAT durch das *Radar Attack Tool* (RAT) [32], das ähnlich wie BRAT auf die digitale Kommunikation zwischen der Radarantenne und dem Radardisplay abzielt und in der Lage ist, kommunizierte Informationen zu modifizieren und künstliche Falschdaten einzuschleusen. Das Werkzeug ermöglich unterschiedliche Formen von Radar-Angriffen, die von einfachen DoS-Angriffen durch offensichtliches Stören des angezeigten Radarbildes bis hin zu komplexeren, schwer erkennbaren Manipulationen reichen, wie beispielsweise dem gezielten Entfernen von Objekten vom Bildschirm oder dem plausiblen Erzeugen von „Geisterschiffen“. Eine Taxonomie über derartige Angriffe findet sich in [32]. Ebenfalls denkbar sind simulierte elektromagnetische Angriffe auf Radar-Systeme, die unter anderem für sogenannte False-Flag-Attacks eingesetzt werden könnten (vgl. [33]).

3.2.2 Netzwerk-Analysewerkzeug

Der *Maritime System Analyzer* (MARISTA) erfasst und verarbeitet systematisch den Netzwerkverkehr, um die Architektur der maritimen Systeme, ihre Netzwerktopologie und die bestehenden Kommunikationsmuster automatisch abzuleiten und zu analysieren. Darüber hinaus bietet MARISTA dem Benutzer eine grafische Visualisierung des Netzwerks und liefert statistische Informationen an andere Komponenten des Labors, die aus der Netzwerkverkehrsanalyse abgeleitet werden.

3.2.3 Ansätze zur präventiven Cybersicherheit

Zu den Laborkomponenten zählen auch zwei speziell entwickelte und prototypisch implementierte präventive Ansätze zur Härtung der IT-Kommunikation innerhalb maritimer Systeme. *MARitime Message Authentication Codes* (MARMAC) [34] und *digital SIGnatures for MARitime Systems* (SIGMAR) [35] zeigen Möglichkeiten auf, wie Nachrichtenintegrität und -authentizität effektiv und zugleich kosteneffizient in bestehende Brückensysteme nachgerüstet werden könnten. Beide am Fraunhofer FKIE entwickelten Verfahren basieren auf Kryptografie. MARMAC baut auf ein symmetrisches kryptografisches Verfahren und erreicht einen geringeren Kommunikations-Overhead als SIGMAR, welches auf komplexerer asymmetrischer Kryptografie basiert,

ist dafür aber weniger invasiv bei der Nachrüstung. Diese und ähnliche Ansätze können in der oben beschriebenen Test- und Entwicklungsumgebung unter realen Bedingungen erprobt, validiert und weiterentwickelt werden.

3.2.4 Cyber Incident Monitoring als Schnittstelle zum Nautiker

Weil präventive Cybersicherheitsmaßnahmen jedoch niemals eine vollständige Sicherheit garantieren können, ist es in der Praxis unerlässlich, zusätzlich stets auch detektive Verfahren zu implementieren. In diesem Zusammenhang dient der *Cyber Incident Monitor* (CIM) [36] des Sicherheitslabors dem Schutz der nautischen Kommunikation durch eine frühzeitige, auf Anomalie-Erkennung gestützte Detektion von möglichen Cyberangriffen und basiert dazu auf einem maritim-spezifischen *Network Intrusion Detection System* (NIDS). Auch CIM ist speziell auf nautische Brückensysteme zugeschnitten und als domänenspezifische NIDS in der Lage, auch sogenannte *Person-on-the-Side* (PotS)-Angriffe innerhalb dieser Systeme zu detektieren. Für die Erkennung von Anomalien und Missbrauch auf der Anwendungsschicht werden drei grundlegende Methoden verwendet:

- protokollbasierte Ansätze, die Denial-of-Service (DoS)-Angriffe (z.B. Flooding von Nachrichten) und ungewöhnliche Häufigkeit der Nachrichtenübermittlung erkennen,
- inhalts- bzw. prozess-basierte Ansätze, welche die Plausibilität von Werten und Abweichungen zwischen ihren Nutzdaten und physischen Prozessen an Bord prüfen,
- und strukturbasierte Ansätze, die eine definierte oder gelernte Topologie verwenden und als ein Set von Regeln ausgedrückt werden. Abweichungen von derartigen Regeln werden erkannt und schließlich als potenzielle Cyberangriffe eingestuft.

Ein weiteres einzigartiges Merkmal von CIM ist das ergonomische *Human-Machine-Interface* (HMI), das für nautisches Personal entwickelt wurde. Alarme und mögliche Detektionen werden dem Bediener auf vertraute Weise präsentiert, kombiniert mit spezifischen Anweisungen und von Experten erarbeitete, individuelle Handlungsempfehlungen, wie auf einzelne Cybervorfälle angemessen reagiert werden kann.

4 Auswertung der Ergebnisse

Ziel ist es nun, die beiden Komponenten aus Kapitel 3 miteinander zu kombinieren. Die Verbindung der Softwaretools mit der Brückenhardware er-

möglicht insbesondere die Untersuchung der schiffsspezifischen Angriffsszenarien aus Kapitel 2, wie dies nur in wenigen bereits existierenden Testumgebungen möglich ist.

Der Prozess der Zusammenführung aller Hard- und Softwarekomponenten steht gegenwärtig im Fokus des Projekts. Dabei fanden bereits erste Tests statt. Beispielsweise ist es erfolgreich gelungen, mit BRAT mehrere Cyberangriffe auf das reale Radarsystem durchzuführen. Dazu wurde im Rahmen eines Proof-of-Concepts ein kleiner Einplatinencomputer (Raspberry Pi), auf dem eine BRAT-Instanz installiert ist, als bösartiges Hardware-Modul in das Brückennetzwerk (Ethernet, s. Abbildung 4) eingeschleust. Über eine WLAN-Verbindung zu einem Tablet mit einer grafischen BRAT-Benutzerschnittstelle ist es nun möglich, verschiedene in [31] beschriebene Angriffe gegen das Brückensystem und das von diesem bereitgestellte maritime Lagebild auszuführen.

An dieser Stelle ist einerseits anzumerken, dass in der Praxis BRAT-Angriffe nicht notwendigerweise lokal von einem sich in WLAN-Reichweite befindenden Nutzergerät wie dem Tablet konfiguriert und ausgelöst werden müssen. Zu diesem Zweck könnte auch externe, funkbasierte Langstreckenkommunikation genutzt werden, wie z.B. AIS oder Radar, vgl. [37]. Andererseits ist zu betonen, dass mittels BRAT auf die beschriebene Weise nicht nur eine Vielzahl netzwerk-basierter Cyberangriffe durchgeführt werden kann, sondern sich gleichzeitig auch elektromagnetische Angriffe effektbasiert simulieren lassen [31].

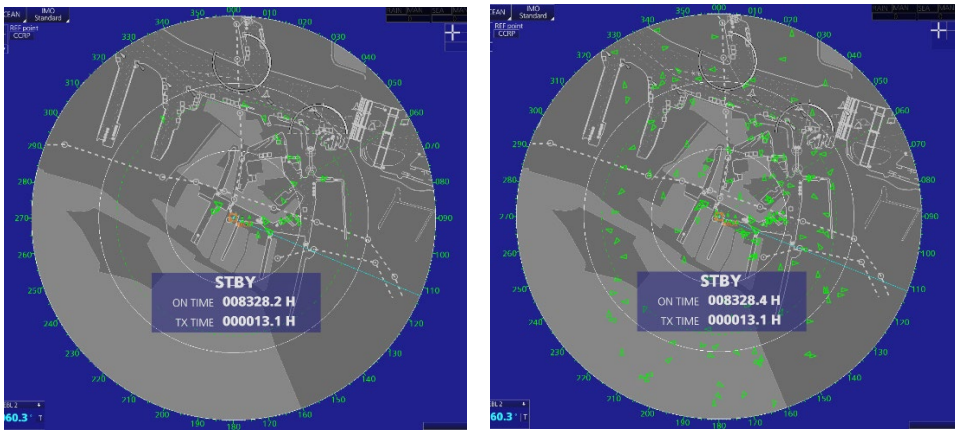


Abbildung 5: Darstellung der AIS-Ziele im Harburger Binnenhafen auf dem Radarschirm. Tatsächliche AIS-Daten (links) und Darstellung nach AIS-Flooding-Angriff mit BRAT (rechts).

In Abbildung 5 ist exemplarisch ein Cyberangriff auf den AIS-Empfang dargestellt. Gezeigt ist das Radardisplay bei ausgeschalteter Radarantenne. Die

linke Abbildung stellt dabei die tatsächlich vorhandenen AIS-Ziele im Harburger Binnenhafen in der Umgebung des Fraunhofer CML dar. Die rechte Abbildung zeigt das Ergebnis eines AIS-Flooding-Angriffs. Dabei konnte über den Datenaustausch zwischen Radardisplay und ECDIS zusätzliche AIS-Ziele einschleust werden. Welche der auf dem Display angezeigten Schiffe tatsächlich vorhanden sind und bei welchen es sich um künstlich durch den Angreifer generierte Ziele handelt, ist nicht zu erkennen. Über denselben Angriffsvektor war es auch möglich, lediglich gezielt einzelne AIS-Ziele hinzuzufügen oder beispielsweise auch die Darstellung vorhandener Ziele auf dem Display erfolgreich zu unterdrücken.

Die Durchführung dieser speziellen Tests in einer so realistischen Umgebung bietet die Möglichkeit, direktes Feedback aus der maritimen Branche zu erhalten. Damit können einerseits sämtliche digitale Werkzeuge für Cyberangriffe, zur Prävention und Härtung von Systemen sowie zur automatisierten Detektion von Angriffen angepasst, weiterentwickelt und verbessert werden. Zudem können gemeinsam mit den Benutzern Handlungsempfehlungen für die Reaktion auf einen Cyberangriff erarbeitet und erprobt werden. Andererseits lassen sich dadurch die unterschiedlichen Auswirkungen verschiedener Angriffe auf die Schiffsbesatzung identifizieren und weitreichendere Konsequenzen untersuchen. Schließlich lässt sich durch praktische Demonstrationen in dem neuartig entwickelten Labor das Bewusstsein über die zunehmende Gefahrenlage erhöhen.

Der in Abbildung 5 gezeigte AIS-Flooding-Angriff beispielsweise eignet sich sehr gut zu Demonstrationszwecken, da gleich auf den ersten Blick eine Manipulation erkennbar ist, auch, weil die künstlichen AIS-Ziele erst bei aufwendiger, deckungsgleicher Manipulation des Radarbilds (das in Abbildung 5 zur besseren Übersicht deaktiviert ist) realistisch erscheinen. Aus demselben Grund wird dieser Angriff auf einer Schiffsbrücke nur geringe Konsequenzen nach sich ziehen – in so einem Fall kann die Offizierin auf der Brücke einfach nach draußen und auf andere Instrumente schauen. Bei subtileren Angriffen fällt die Einschätzung der praktischen Folgen für die Schiffsführung jedoch schwerer; hier ermöglicht die einfache Simulation der Angriffe einen Austausch mit Reedereien und NautikerInnen.

Zudem werden sich weiterführende und vertiefende Erkenntnisse davon erhofft, weitere Geräte in das in Abschnitt 3.1 beschriebene Brückenlabor einzubauen – insbesondere auch die Geräte von alternativen, in der Praxis weit verbreiteten Herstellern. Basieren diese zum Beispiel auf anderen Betriebssystemen oder Kommunikationsschnittstellen, so werden dafür andere An-

griffsvektoren möglich, aber auch nötig sein. Auch gilt es in zukünftigen Arbeiten zu erforschen, für welche Angriffe spezifische Anpassungen an einen bestimmten Gerätetyp vorgenommen werden müssen, und welche Angriffstypen dagegen herstellerübergreifend möglich sind.

5 Zusammenfassung

In diesem Beitrag wurde ein neuartiges maritimes Cybersicherheitslabor für integrierte Schiffsbrückensysteme präsentiert, das gemeinschaftlich von Fraunhofer FKIE und Fraunhofer CML am Standort Hamburg entwickelt und realisiert wurde und fortwährend weiter ausgebaut wird. Hinsichtlich der stetig wachsenden Bedrohungen aus dem Cyberinformationsraum und dem elektromagnetischen Spektrum bietet dieses Labor das Potenzial, adäquate Konzepte zum Schutz maritimer IT-Systeme zu entwickeln, durch prototypische Implementierungen deren Effektivität zu demonstrieren und damit die in der Schifffahrt zu lange vernachlässigte Cybersicherheit voranzutreiben.

Mit seiner einzigartigen Kombination aus echter Brückenhardware mit einem software-seitigen Portfolio an brückenspezifischen Cyberangriffs-Werkzeugen ermöglicht dieses Labor als Test- und Entwicklungsumgebung praktische Sicherheitsanalysen von maritimen Systemen, die Identifikation von Schwachstellen und Verwundbarkeiten auf systemischer, aber auch operationeller Ebene, sowie eine fundierte Untersuchung möglicher Auswirkungen von Cyberangriffen. Damit ermöglicht es die Entwicklung von passgenauen Sicherheitslösungen zu deren Prävention, Detektion und Reaktion.

Literaturverzeichnis

[1] Umweltbundesamt, „Fakten zur Seeschifffahrt und zu ihren Auswirkungen auf die Umwelt," März 2023. [Online]. Available: https://www.umweltbundesamt.de/themen/wasser/gewaesser/meere/nutzung-belastungen/schifffahrt\#fakten-zur-see-schifffahrt-und-zu-ihren-auswirkungen-auf-die-umwelt.

[2] P. Mitrou, „Huge expansion of LNG tanker fleet needed to power energy transition," LR Horizons, 1 Juni 2023.

[3] International Maritime Organization (IMO), Resolution MSC.428(98) Maritime Cyber Risk Management in Safety Management Systems, 2017.

[4] Deutsche Flagge, „Sicherer Schiffsbetrieb (ISM)," [Online]. Available: https://www.deutsche-flagge.de/de/sicherheit/ism-code. [Zugriff am 23 Februar 2024].

[5] International Association of Classification Societies (IACS), UR E26 -- Cyber resilience of ships, 2022.

[6] International Association of Classification Societies (IACS), UR E27 -- Cyber resilience of on-board systems and equipment, 2022.

[7] The Swedish Club, „Cyber Insurance," 2024. [Online]. Available: https://www.swedishclub.com/insurance/cyber-insurance/. [Zugriff am 19 Februar 2024].

[8] United Nations Conference on Trade and Development (UNCTAD), Review of Maritime Transport 2023, New York, NY, USA, 2023.

[9] A. Hahn, M. Steidel, S. Berner, A. Nies, G. S. Schaal und A. Weiß, „Roadmap Sichere Digitale Küste 2030," 2021.

[10] R. R. Negenborn, F. Goerlandt, T. A. Johansen, P. Slaets, O. A. Valdez Banda, T. Vanelslander und N. P. Ventikos, „Autonomous ships are on the horizon: here's what we need to know," Nature, Bd. 615, pp. 30-33, 2023.

[11] H. Bosch, „False economy: remote-controlled autonomous ships?," Baird Maritime, 7 Januar 2021.

[12] C. Wienberg, „Maersk's CEO Can't Imagine Self-Sailing Box Ships in His Lifetime," Bloomberg, 15 Februar 2018.

[13] R. Grundmann, A. Ujkani, J. Weisheit, J. Seppänen, M. Salokorpi und H.-C. Burmeister, „Use Case Remote Pilotage – Technology Overview," Journal of Physics: Conference Series, Bd. 2618, Nr. 1, Oktober 2023.

[14] D. Heering, O. M. Maennel und A. N. Venables, „Shortcomings in Cybersecurity Education for Seafarers," in Proc. of MARTECH, Lissabon, Portugal, 2020.

[15] BIMCO, IMO to start comprehensive review of STCW Convention and Code, 2022.

[16] Sjøfartsdirektoratet (Norwegian Maritime Authority), RSV 18-2022 - Reporting cyber incidents, 2022.

[17] M. Schwarz, M. Marx und H. Federrath, „A Structured Analysis of Information Security Incidents in the Maritime Sector," arXiv:2112.06545, 2021.

[18] Maritime Computer Emergency Response Team (M-CERT), „ADMIRAL," 2023. [Online]. Available: https://gitlab.com/m-cert/admiral. [Zugriff am 11 4 2023].

[19] J. Bhatti und T. E. Humphreys, „Hostile Control of Ships via False GPS Signals: Demonstration and Detection," ION Navigation, Bd. 64, Nr. 1, pp. 51-66, 2017.

[20] G. C. Kessler und J. P. Craiger, „A Taxonomy Framework for Maritime Cybersecurity: A Demonstration Using the Automatic Identification System," Int. Journal on Maritime Navigation and Safety of Sea Transportation (TransNav), Bd. 12, Nr. 3, pp. 429-437, 2018.

[21] J. Wiseman, „GPSJAM," 2022. [Online]. Available: https://gpsjam.org. [Zugriff am 26 02 2024].

[22] M. Kirchner, „GPS-Jamming: Ostseeraum erlebt großflächige Ausfälle," Heise Online, 31 Januar 2024.

[23] M. Möller und H. Strüber, „Ostsee: Rätselhafte GPS-Störungen behindern Schiffs- und Flugverkehr," NDR, 4 Februar 2024.

[24] Gartner, Inc., „Gartner Predicts Nearly Half of Cybersecurity Leaders Will Change Jobs by 2025," Gartner Newsroom, 22 Februar 2023.

[25] National Transportation Safety Board, „Marine Accident Report - Grounding of the U.S. Tankship EXXON VALDEZ on Bligh Reef, Prince William Sound, near Valdez, Alaska, March 24, 1989," Washington, D.C., USA, 1990.

[26] Maritime Cyprus, „Panama Issues Report Critical of SCA Pilots in Ever Given Grounding," The Maritime Executive, 11 Juli 2023.

[27] The Danish Maritime Accident Investigation Board, „Safety Report June 2016. Proceduralizing marine safety," Valby, Denmark, 2016.

[28] J. Kutzner, P. Sedlmeier, A. Rizvanolli und J. Bauer, „Phish & Ships and Other Delicacies from the Cuisine of Maritime Cyber Attacks," in European Workshop on Maritime Systems Resilience and Security MARESEC, virtual, 2023.

[29] Fraunhofer CML, „marFM: Automatische Spracherkennung von maritimen Funksprüchen," [Online]. Available: https://www.cml.fraunhofer.de/de/forschungsprojekte1/marFM.html. [Zugriff am 23 Februar 2024].

[30] E. C. Nakilcioglu, M. Reimann und O. John, „Adaptation and Optimization of Automatic Speech Recognition (ASR) for the Maritime Domain in the Field of VHF Communication," arXiv:2306.00614, 2023.

[31] C. Hemminghaus, J. Bauer und E. Padilla, „BRAT: A BRidge Attack Tool for Cyber Security Assessments of Maritime Systems," International Journal on Marine Navigation and Safety of Sea Transportation (TransNav), Bd. 15, Nr. 1, März 2021.

[32] K. Wolsing, A. Saillard, J. Bauer, E. Wagner, C. van Sloun, I. B. Fink, M. Schmidt, K. Wehrle und M. Henze, „Network Attacks Against Marine Radar Systems: A Taxonomy, Simulation Environment, and Dataset," in Proc. of the 47th IEEE Conference on Local Computer Networks, LCN, Edmonton, Canada, 2022.

[33] G. Longo, A. Merlo, A. Armando und E. Russo, „Electronic Attacks as a Cyber False Flag Against Maritime Radars Systems," in Proc. of the 1st IEEE LCN Workshop on Maritime Communication and Security (MarCaS), in conjuction with the 48th IEEE International Conference on Local Computer Networks (LCN), Daytona Beach, FL, USA, 2023.

[34] L. Ruhland, M. Schmidt, J. Bauer und E. Padilla, „Keeping the Baddies Out and the Bridge Calm: Embedded Authentication for Maritime Networks," in Proc. of the International Symposium on Networks, Computers and Communications - Trust, Security, and Privacy, ISNCC-TSP, Shenzhen, China, 2022.

[35] C. Hemminghaus, J. Bauer und K. Wolsing, „SIGMAR: Ensuring Integrity and Authenticity of Maritime Systems using Digital Signatures," in Proc. of the International Symposium on Networks, Computers and Communications - Trust, Security, and Privacy, ISNCC-TSP, Dubai, UAE, 2021.

[36] M. von Rechenberg, N. Rößler, M. Schmidt, K. Wolsing, F. Motz, M. Bergmann, E. Padilla und J. Bauer, „Guiding Ship Navigators through the Heavy Seas of Cyberattacks," in Proc. of the European Workshop on Maritime Systems Resilience and Security (MARESEC), Bremerhaven, Germany, 2022.

[37] W. C. Leite Junior, C. C. de Moraes, C. E. P. de Albuquerque, R. C. S. Machado und A. O. de Sá, „A Triggering Mechanism for Cyber-Attacks in Naval Sensors and Systems," MDPI sensors, Bd. 21, Nr. 9, 2021.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Verwendung von Internet-Scans und passiven Messungen zur Analyse russischer Angriffe und ihrer Auswirkungen in der Ukraine

Johannes Klick[1], Stephan Lau[1], Daniel Marzin[1]

Kurzfassung:
Die Rolle der Cyber-Domäne im ukrainischen Krieg ist noch nicht vollständig erforscht. Derzeit werden primär Cyberangriffe analysiert, aber es wird stark vernachlässigt, dass mit Hilfe der Cyberdomäne messbare und überprüfbare Aussagen über die reale Welt gemacht werden können. So kann eine gute Cyber-Aufklärung nicht nur die digitale Angriffsfläche des Gegners bestimmen, sondern auch Erfolgsaussagen von physischen kinetischen Großangriffen ermitteln. Durch globale Internet-Scans und unser passives Blackhole-Sensornetzwerk können wir sowohl Cyber- als auch konventionelle Angriffe und deren Auswirkungen auf das Territorium der Ukraine identifizieren. Zusätzlich korrelieren wir frei verfügbare Radardaten der Europäischen Weltraumorganisation und Open-Source-Intelligence-Daten zu russischen Angriffen. Auf dieser Grundlage zeigen wir, wie Internet-Scan-Daten eine gute Quelle für die Messung des Erfolgs von Angriffen in einer bestimmten Region sein können. Darüber hinaus können sie auch zur Verifizierung von Angriffsberichten in sozialen Medien verwendet werden, um Fake News zu bekämpfen, die in modernen Konflikten eine wichtige Rolle spielen. Unser Ansatz kann auch Nachrichtenagenturen dabei helfen, die Authentizität von Berichten schnell zu bestimmen. Dieser starke Zusammenhang zwischen unseren Internet-Scans und den tatsächlichen Ereignissen wird durch einen Korrelationskoeffizienten von 0,74 belegt.

Stichworte: Blackhole-Sensornetzwerk, Cyberdomäne, DDoS, digitale Angriffsfläche, Fake News Detection, Internet-Scans, IP-Adressen, KRITIS, Radardaten, Russland-Ukraine-Krieg, Ukraine

1 Einleitung

Am 24. Februar 2022 begann der russische Angriffskrieg gegen die Ukraine. Russlands Vorgehen war konventionell, was einige Außenstehende überraschte, welche einen deutlich stärkeren Einsatz der Cyber-Domäne erwartet hatten. Dass dies nicht der Fall war, ist jedoch auf die Rahmenbedingungen der operativen Planung zurückzuführen, welche unter strengster Geheimhaltung erfolgte. Bis heute ist die Rolle der Cyber-Domäne im Ukraine-Krieg noch nicht vollständig berücksichtigt worden. Derzeit werden primär Cyberangriffe analysiert, aber es wird stark vernachlässigt, dass mit Hilfe der

[1] Alpha Strike Labs GmbH, Berlin

Cyberdomäne messbare und überprüfbare Aussagen über die reale Welt gemacht werden können.

So kann eine gute Cyber-Aufklärung nicht nur die digitale Angriffsfläche des Gegners bestimmen, sondern auch Aussagen über die Auswirkung von physisch-kinetischen Großangriffen ermitteln.

Darüber hinaus ist auch die Erkennung von Fake News besonders wichtig, vor allem wenn sie auch für die militärische Aufklärung ein immer wichtigeres Element ist. Während immer mehr Informationsquellen zur Verfügung stehen, nimmt deren Manipulation sowie die Qualität der Fälschungen zu. Insbesondere eine zeitnahe Beurteilung der Echtheit von Informationen wird zunehmend schwieriger. Um diese Unzulänglichkeiten zu überwinden, stellen wir einen neuen Ansatz vor, der Internet-Scandaten nutzt, um kinetische Angriffe frühzeitig zu erkennen und ihre Auswirkungen genauer abzuschätzen sowie bisher ungenutzte Cyber-Domain-Informationen zur Bewertung der Authentizität von Nachrichten zu nutzen und die Erkennung von Fake News zu unterstützen.

2 Erläuterung der Netzwerk-Scan-Methodik

Die im Internet sichtbare Oberfläche eines Landes genau zu erfassen und zu bewerten ist eine Herausforderung. Aus unserer Sicht besteht diese Oberfläche zunächst aus Servern, die öffentlich zugängliche Dienste über das Internet darstellen und sich im Zielland, in diesem Fall der Ukraine, befinden. Den Kern unserer Analyse bilden die Daten, welche wir mit dem Distributed Cyber Recon System (DCS) gesammelt haben. Dabei handelt es sich um eine Suchmaschine, welche das gesamte IPv4-Internet, das derzeit aus rund 3 Milliarden routingfähigen IP-Adressen besteht, innerhalb weniger Stunden scannen kann. Ein Scan erfasst immer einen bestimmten Dienst auf einem festen Port, der den Zustand des Internets möglichst genau in einem so genannten Snapshot erfasst. Zusätzlich zu den allgemeinen Erreichbarkeitsdaten und den dienstspezifischen Daten werden die Aufzeichnungen mit weiteren Daten angereichert. Informationen über das AS, in dem sich die Server befinden, Daten über den Eigentümer und die Verwalter des zugehörigen IP-Adressbereichs (WHOIS) sowie Analysedaten über die Dynamik des Netzbereichs und Geolokalisierungsdaten (GeoIP). Die Kombination und kontinuierliche Sammlung dieser Daten ermöglicht eine genaue Kartierung des Internetraums. Mithilfe von Big-Data-Analysen kann ein fachkundiger Analyst die Daten dann nach verschiedenen Gesichtspunkten auswerten und so einerseits einen genauen Überblick über die Cyberlandschaft einzelner Netzwerke, Länder und Organisationen gewinnen und andererseits vielschichtige Analysen vornehmen [1].

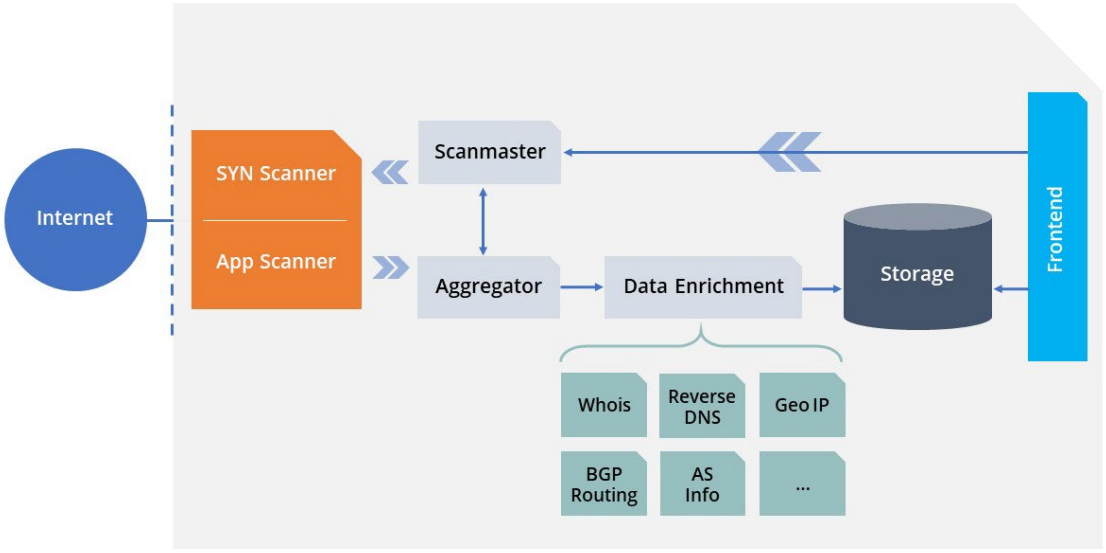


Abbildung 1: Schematische Darstellung der DCS-Architektur

Eine schematische Darstellung zum Aufbau unseres DCS findet sich in Abbildung 1. Im Frontend werden die Scans konfiguriert und die gesammelten Daten ausgewertet. Üblicherweise wird das gesamte Internet auf einem Port mit einem bestimmten Dienst gescannt. Die Flexibilität des DCS erlaubt es aber beispielsweise auch, die einzelnen Suchknoten geografisch zu verteilen und die zu scannenden IP-Bereiche von den nächstgelegenen Scan-Knoten aus zu scannen oder das Scannen auf vielversprechende Netzwerkbereiche zu konzentrieren [2]. Sobald die Scan-Parameter definiert sind, teilt der Scanmaster diese je nach Scan-Methode in einzelne Arbeitspakete auf und weist sie den vorhandenen Scan-Knoten zu. Fällt ein Suchknoten während eines Scans aus, werden seine Arbeitspakete auf andere Scan-Knoten verteilt. Die Scan-Knoten senden pseudo-zufällig TCP SYN-Pakete (Transmission Control Protocol Synchronize) an alle im Arbeitspaket definierten Zieladressen. Sobald eine gescannte IP-Adresse antwortet, wird ein anwendungsspezifischer Scanner (App Scanner) gestartet und sammelt protokollspezifische Daten, wie das Software-Banner des Hosts. Alle gesammelten Daten werden an den Aggregator weitergeleitet, welcher eventuell vorhandene Duplikate aussortiert und zusätzliche Metadaten hinzufügt. Genaue Ergebnisse sind ein wichtiger Aspekt des Internet-Scanings. Allerdings blockieren Filtersysteme automatisch intensiven Suchverkehr. Suchmaschinenbetreiber wie Shodan und Censys scannen daher kontinuierlich alle Protokolle mit geringer Geschwindigkeit und aktualisieren ihre Datenbanken permanent. Das DCS hingegen versucht, den aktuellen Zustand des Internets innerhalb eines kleinen Zeitfensters möglichst genau zu erfassen. Dies hat den Vorteil, dass sich ständig ändernde Netze, wie z. B. Endkundennetze oder

Cloud-Anbieter, das Scan-Ergebnis weniger verfälschen. Dies wird durch einen Scan mit 1022 verschiedenen Quelladressen erreicht, wobei ein einzelner Scan 6 Stunden dauert. Die pseudo-zufällige Verteilung aller 3 Milliarden zu scannenden IP-Adressen ergibt somit eine durchschnittliche Scan-Last von etwa 42 Paketen pro Stunde für ein /24-Netzwerk (256 IP-Adressen). Für ein /16-Netz mit 65536 IP-Adressen beträgt die Scan-Last etwa 10923 Scan-Pakete pro Stunde, wobei ein Zielnetz dieser Größe im Durchschnitt nur 10 Pakete pro Stunde von derselben Quell-IP-Adresse empfängt. Nach unserer Erfahrung liegen diese Raten weit unter den Schwellenwerten automatischer Filtersysteme. Außerdem führt eine geografische Verteilung der Scan-Knoten zu weniger Paketverlusten und kann regionale Filter vermeiden. Für die Zwecke dieser Studie beziehen wir uns nur auf die Erreichbarkeit von Hosts, die sich geografisch in der Ukraine befinden. Wir definieren Hosts als erreichbar, wenn sie auf ein TCP SYN-Paket reagieren, gefolgt von einem gültigen Handshake auf der Anwendungsebene. Für die Geolokalisierung verwenden wir die GeoLite2 City-Datenbank von Maxmind[2]. Nach eigenen Angaben beträgt die Genauigkeit der korrekten Lokalisierung in einem Radius von 250 km 84 %, wobei 10 % nicht korrekt und 6 % überhaupt nicht aufgelöst werden. Die Lokalisierung auf Stadtebene gelingt zu 64 %, wobei 31 % der Ergebnisse falsch und 5 % gar nicht aufgelöst werden [3]. Diese Zahlen gelten nicht für IPs von mobilen Geräten, die jedoch standardmäßig keine Dienste im Internet anbieten und daher für unsere Analyse nicht relevant sind.

Das DCS scannt weit verbreitete Protokolle wie HTTP-Port 80, HTTPS-Port 443 und SSH-Port 22 alle zwei Wochen, während weniger verbreitete Protokoll-Port-Kombinationen nur alle vier bis sechs Wochen gescannt werden. Dies reicht zwar für die Erfassung von Konektivitätstrends aus, aber diese niedrige Frequenz erlaubt es uns nicht, Kriegsereignisse mit den Scandaten zu korrelieren. Zu diesem Zweck haben wir so genannte Hochfrequenz-Scans entwickelt, die die gewünschten Hosts mit einer hohen Frequenz von 4 Stunden scannen. Wir wählen die IP-Adressen, die sich aktuell im Aufklärungsgebiet befinden, anhand der Geodaten des letzten Scans für jede Protokoll/Port-Kombination aus. Darüber hinaus werden dynamische Adressbereiche herausgefiltert, da diese das Ergebnis ständig verfälschen. Alle gesammelten Adressen werden erneut mit dem Appscanner gescannt, mit Metadaten angereichert und zur weiteren Analyse gespeichert.

[2] Dieses Produkt enthält GeoLite2-Daten, die von MaxMind erstellt wurden und unter https://www.maxmind.com erhältlich sind.

2.1 Aufbau und Widerstandsfähigkeit des ukrainischen Internets

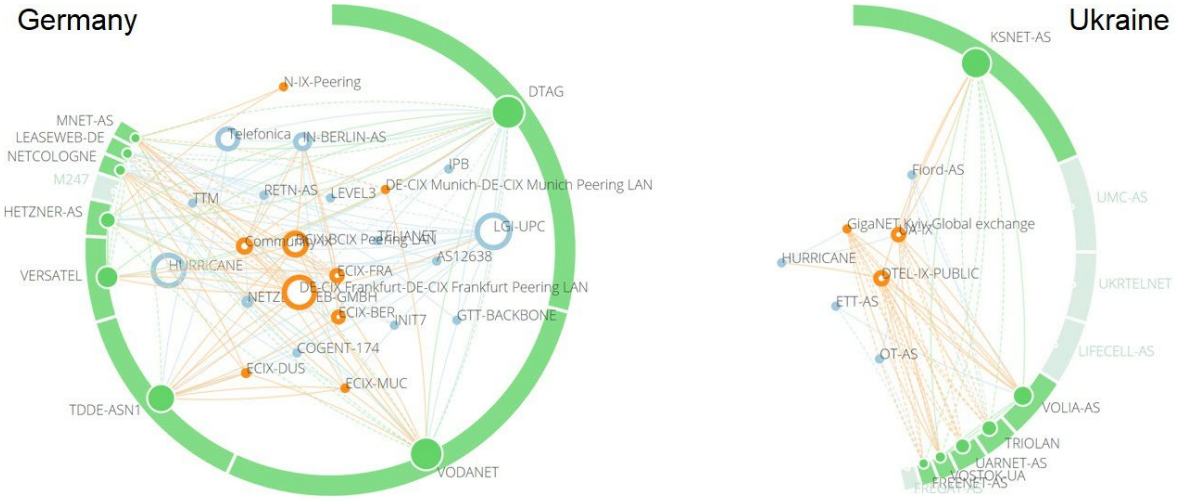


Abbildung 2: Darstellung der proportionalen Endkundenverteilung pro ISP. Der gesamte imaginäre Kreis stellt 100 Prozent aller Endkunden dar. Die grünen Kreiselemente stellen den prozentualen Anteil dar. Dargestellt sind nur ISPs, die mindestens 1 Prozent der Endkunden versorgen. Quelle: RIPE Atlas

Das ukrainische Internet ist wesentlich widerstandsfähiger gegen physische Angriffe, als die meisten westlichen Internet-Infrastrukturen. Dies liegt daran, dass das ukrainische Internet aus wesentlich mehr kleinen ISPs besteht als die westlichen Gegenstücke. Abbildung 2 zeigt das ukrainische Internet im Vergleich zum deutschen Internet. Es ist leicht zu erkennen, dass etwa 2/3 der deutschen Endkunden von 3 großen zentralen ISPs bedient werden. Weniger als 15 Prozent der Endnutzer in Deutschland werden von ISPs bedient, die jeweils weniger als 1 Prozent der Endnutzer bedienen. Wenn wir die Ukraine zum Vergleich heranziehen, sehen wir sofort, dass die drei größten ISPs fast ein Drittel der Endkunden bedienen und mehr als 50 Prozent aller Endkunden von ISPs bedient werden, die jeweils weniger als 1 Prozent der Endkunden bedienen. Folglich besteht das ukrainische Internet aus viel mehr kleinen ISPs und weniger zentral wichtigen ISPs. So führt der Ausfall eines ISP, z. B. durch militärische Angriffe, zu wesentlich geringeren Beeinträchtigungen, da der Netzverkehr leichter über einen anderen ISP geleitet werden kann. Einschränkungen bei der Verfügbarkeit von Netzdiensten und Servern sind also eher auf den Ausfall der Server selbst oder des direkt angeschlossenen ISP zurückzuführen als auf eine unzureichende Anzahl von Transit-ISPs.

3 Ermittlung des Zerstörungsgrades von Mariupol anhand von Internet-Scandaten und ESA Radardaten

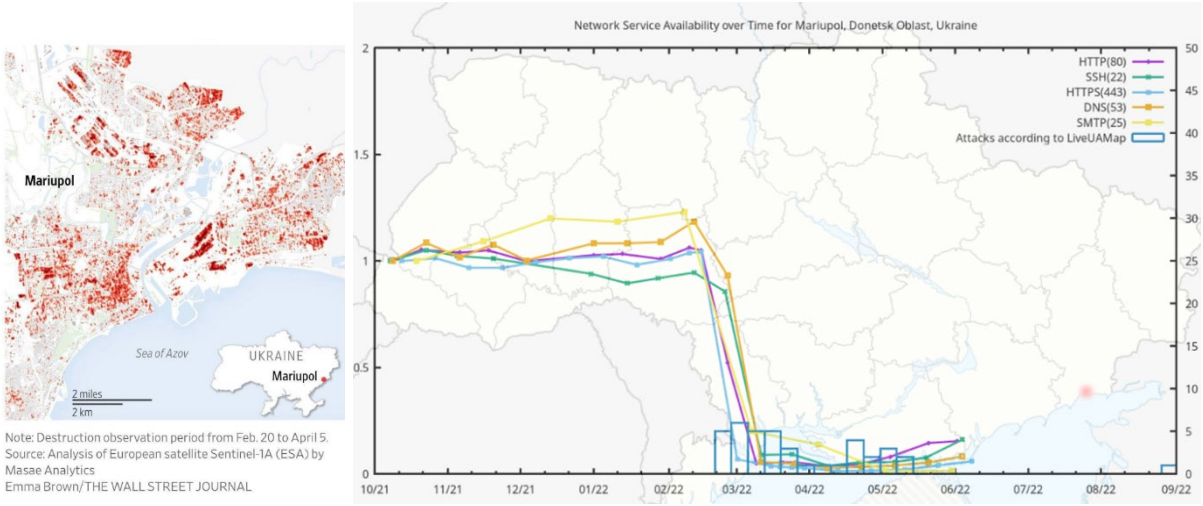


Abbildung 3: Unsere Internet-Scans zeigen eine Verringerung der Verfügbarkeit von Netzdiensten um bis zu 90 % in der Stadt Mariupol, kurz nach dem Beginn der russischen Invasion Ende Februar 2022. Diese Daten werden zusätzlich von der European Space Administration (ESA) und deren Radar-Satellitenbildern bestätigt, welche ebenfalls einen Zerstörungsgrad von ca. 90 % anzeigen.

Als Ausgangspunkt unserer Betrachtung wurde der 01. Oktober 2021 gewählt, welcher ungefähr 6 Monate vor dem russischen Angriffskrieg in der Ukraine liegt. Abbildung 3 zeigt, wie sich die Netzwerkdienste und Server in Mariupol während der russischen Invasion verhalten. Es ist deutlich zu erkennen, dass seit Kriegsbeginn am 24. Februar 2022, die Verfügbarkeit der Dienste über alle Protokolle hinweg, um etwa 90 Prozent zurückging. Darüber hinaus ist zu erkennen, dass auch die DNS-Dienste stark betroffen sind und nicht mehr zur Verfügung stehen. Diese Nichtverfügbarkeit hält über einen längeren Zeitraum an, was darauf hindeutet, dass die Infrastruktur von Mariupol weitgehend zerstört ist: Dies deckt sich mit Medienberichten und offiziellen Angaben, wonach Mariupol zu 70-90 Prozent zerstört ist [4,5].

Daten werden von der Europäischen Weltraumorganisation (ESA) im Rahmen des Sentinel-1-Programms bereitgestellt. Das Sentinel-1-Programm verwendet ein Radar mit synthetischer Apertur (SAR), welches den Vorteil hat, bei Wellenlängen zu arbeiten, die nicht durch Wolken oder mangelnde Beleuchtung beeinträchtigt werden. Seine Aufgabe ist die kontinuierliche Radarkartierung der Erde[3]. Diese Radardaten können auch zur Messung der Höhenunterschiede von Gebäuden verwendet werden. So zeigt die linke Seite von Abbildung 3 den Grad der Zerstörung Mariupols. Je intensiver der

[3] https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview

rote Farbton ist, desto höher ist der Grad der Zerstörung. Anhand dieser Radarkarten lässt sich schnell erkennen, dass Mariupol w e i t g e h e n d zerstört ist, was einen weiteren Beweis für die Korrelation zwischen dem Grad der Zerstörung in Mariupol und den Ergebnissen aus unseren Internet-Scans liefert. Die in diesem Abschnitt vorgestellte Beispielanalyse, welche wir ebenfalls für 23 andere Oblaste der Ukraine durchgeführt haben, zeigt, dass Internet-Scandaten als zusätzliche Quelle hilfreich sein können, um Angriffe, Gebietseroberungen oder den Grad der Zerstörung bestimmter umkämpfter Regionen zu ermitteln oder andere Informationsquellen zu bestätigen.

4 Russische Angriffe auf die Energieinfrastruktur und deren Auswirkung im Oktober 2022 und März 2023

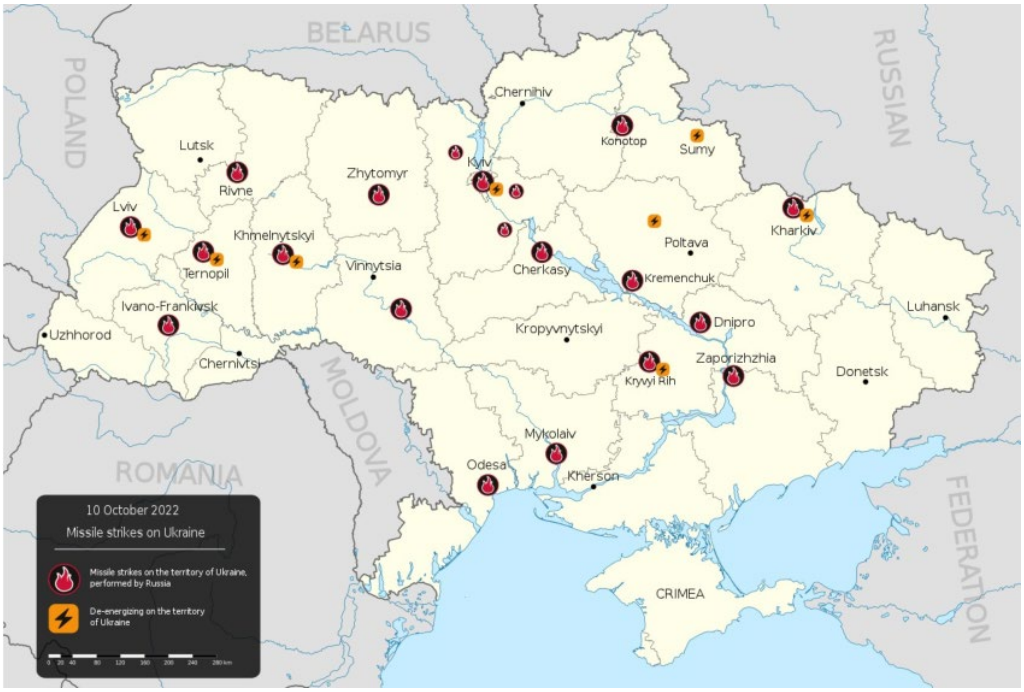


Abbildung 4: Illustration der russischen Raketenangriffe vom 10. Oktober 2022 und der damit verbundenen Stromausfälle in den jeweiligen Regionen. Quelle: Alex Kozur, Wikimedia [6]

Am 10. Oktober 2022 führte Russland massive und großflächige Angriffe gegen die Ukraine durch. Dabei wurden explizit Einrichtungen der Energieversorgung angegriffen [7-9]. Abbildung 4 basiert auf OSINT-Informationen und zeigt die verschiedenen Angriffsorte und zusätzlich, welche Oblaste in der Folge von Stromausfällen betroffen waren. Da Stromausfälle auch zum Ausfall von IT-Systemen führen, wenn diese nicht durch eine unterbrechungsfreie Stromversorgung (USV) gesichert sind, sollte sich dies auch in der Verfügbarkeit von IT-Systemen in der jeweiligen Region widerspiegeln. Seit September 2022 scannen wir die Ukraine im 4h-Hochfrequenztakt, um mögliche

Stromausfälle und damit verbundene Verfügbarkeitsverluste zu erkennen. Im Folgenden wird gezeigt, dass wir die OSINT-Informationen auf Basis unserer Scandaten verifizieren können.

Hierzu betrachten wir beispielhaft das Oblast Sumy im Norden der Ukraine, wo laut OSINT-Daten (siehe Abbildung 4) Stromausfälle zu verzeichnen waren. Abbildung 5 zeigt, dass unseren Scandaten zufolge vor dem 10. Oktober 2022 kein größerer Verlust der Verfügbarkeit von Internetdiensten auftrat. Unsere Scans zeigen jedoch, dass es zwischen 3:00 und 9:00 Uhr am 10. Oktober 2022 zu einem schwerwiegenden Verlust an erreichbaren Systemen kam. In dieser Zeit waren bis zu 90 Prozent der Systeme über alle Protokolle hinweg nicht mehr erreichbar. Wir führen dieses Verhalten auf den Ausfall der Stromversorgung zurück, was auch durch OSINT-Daten entsprechend bestätigt wird. Anhand unserer Scans lässt sich ableiten, dass der Stromaus fall etwa 12 Stunden dauerte. Wir definieren einen Stromausfall, wenn plötzlich weniger als 80 Prozent der Dienste über einen längeren Zeitraum verfügbar sind. Es ist leicht zu erkennen, dass die Stromversorgung in der Zeit zwischen dem 11. und 12. Oktober nicht stabil war, da es immer wieder zu leichten Schwankungen in der Verfügbarkeit der Dienste kommt. Erst am 13. Oktober wird eine ähnliche Stabilität in der Verfügbarkeit der Systeme erreicht wie vor dem 10. Oktober 2022.

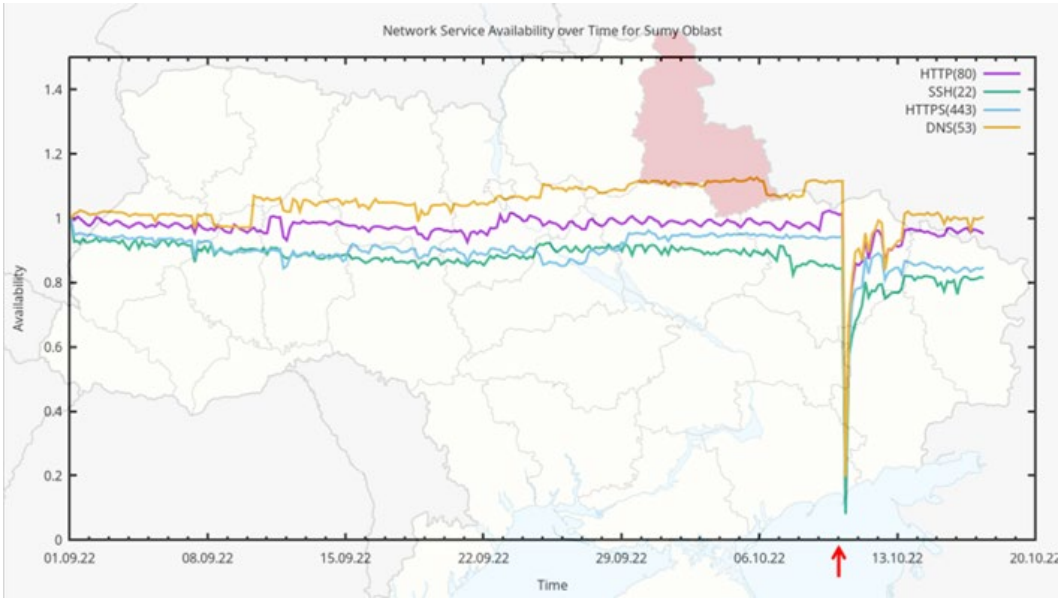


Abbildung 5: Messung und Darstellung der Dienstverfügbarkeit von via Internet erreichbaren Systemen im Oblast Sumy.

	Oblast																							
	Kharkiv	Khmelnytskyi	Kiev City + Oblast	Lviv	Poltava	Sumy	Ternopil	Dnipropetrovsk	Cherkasy	Chernihiv	Chernivtsi	Donetsk (RU-controlled)	Ivano-Frankivsk	Kherson (RU-controlled)	Kirovohrad	Luhansk (RU-controlled)	Mykolayiv	Odessa	Rivne	Transcarpathia	Vinnytsia	Volyn	Zaporizhzhya (RU-controlled)	Zhytomir
Power Outage OSINT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Drop visible in scan results	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Slightly	No	No	No	Slightly	No	No	No	No	No	No	No	Slightly	No	Yes

Abbildung 6: Datenreihe zur Ermittlung des Korrelationskoeffizienten (no drop: 0 ; slightly drop: 0…0,25 ; drop: 0,26…1)

Wie am Beispiel des Oblast Sumy geschildert, haben wir diese Analysen auch für 23 weitere Oblaste durchgeführt. Die Ergebnisse zeigen, dass es Korrelationen zwischen OSINT-Informationen und unseren Internet-Scan-Daten gibt. Dabei haben wir die Korrelation zwischen der Anzeige eines Stromausfalls in den OSINT-Daten und einem großen Verlust der Verfügbarkeit von Netzwerkdiensten für alle Oblasts bestimmt (siehe Abbildung 6).

Der von uns ermittelte Korrelationskoeffizient nach Pearson beträgt 0,74, was einen starken Zusammenhang indiziert. Folglich können unsere Scan-Daten verwendet werden, um z. B. die Wirksamkeit fremder Angriffe zu messen. So ließe sich beispielsweise messen, ob ein Angriff auf ein Kraftwerk, das zum Beispiel militärische Liegenschaften versorgt, erfolgreich war oder nicht. Außerdem ließe sich damit messen, wann die Reparaturen abgeschlossen sind, wann die Stromversorgung wiederhergestellt ist und wann die Netzdienste wieder online sind.

Diese Informationen könnten aus feindlicher Sicht auch genutzt werden, um die entsprechende Anlage kurz nach erfolgreicher Reparatur erneut anzugreifen. Außerdem können Meldungen aus sozialen Netzwerken auf ihre Richtigkeit überprüft werden. Wenn zum Beispiel in den sozialen Medien über großflächige Bombardierungen berichtet wird, können diese schnell durch Internet-Scans verifiziert werden. Sobald die Infrastruktur im Bereich der Telekommunikation oder der Stromversorgung angegriffen oder beeinträchtigt wird, hat dies immer auch Auswirkungen auf die Verfügbarkeit von Netzdiensten im Internet.

5 Mutmaßliche russische Cyberangriffe

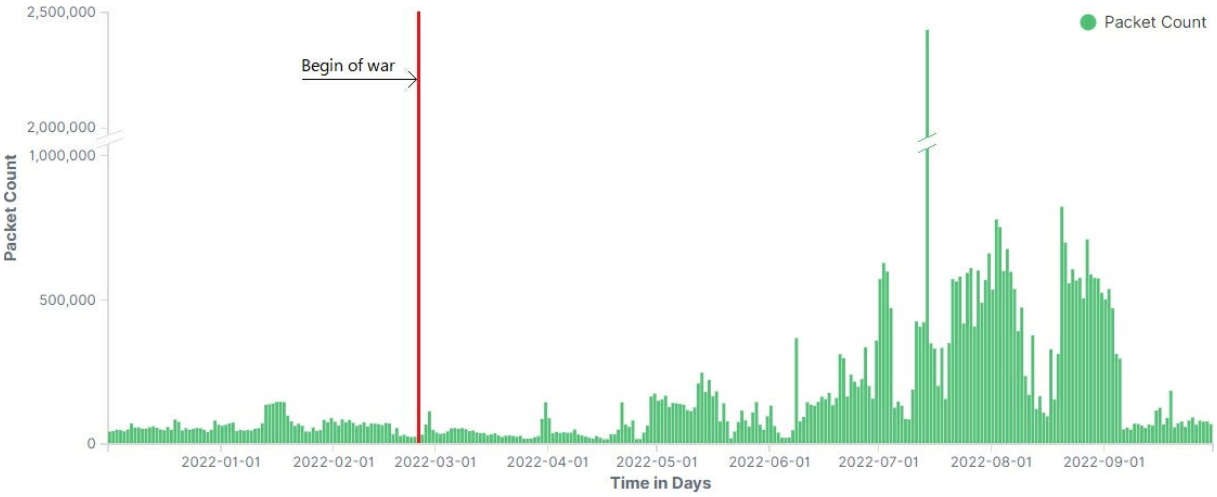


Abbildung 7: Alle IP-Pakete aus der Ukraine zwischen dem 01.12.2021 und dem 01.10.2022

Seit 2018 betreiben wir ein Blackhole-Sensornetz mit 1024 öffentlichen IPv4-Adressen. Ein Blackhole-Sensornetzwerk überwacht ständig den gesamten eingehenden Verkehr des Internets und fängt jedes empfangene IP-Paket auf. Eine maßgeschneiderte Software analysiert jedes IP-Paket, indem sie alle Header- und Datenfelder analysiert. Ausgehend von der Quell-IP-Adresse reichert die Software den Datensatz mit Informationen aus Reverse Domain Name System (rDNS), WHOIS, geografischen Standort- und AS-Datenbanken an. Die rDNS-, WHOIS- und AS-Einträge identifizieren den aktuellen Besitzer, und die geografischen Standortinformationen ermöglichen es, die IP-Adresse auf einer Karte einzuzeichnen und sie einem Land zuzuordnen. Anschließend werden die Ergebnisse in eine Hochleistungsdatenbank übertragen, die Such- und Aggregationsfunktionen für jeden Datenpunkt bietet. Die Software ist in der Lage, IPv4- und IPv6-Pakete unabhängig von den darüber gesprochenen Protokollen zu analysieren, z. B. Transmission Control Protocol (TCP), User Datagram Protocol (UDP) oder Internet Control Message Protocol (ICMP). Ein Teil des eingehenden Datenverkehrs wird von Scannern, Angreifern und falsch konfigurierten Geräten verursacht und wird als Grundrauschen bezeichnet. Da keine Antwort auf eingehende IP-Pakete erfolgt, erhalten die Blackhole-Sensornetzwerke meist Anwendungsdaten aus UDP-Paketen, wie DNS-Anfragen oder TCP-Implementierungen, die bereits im ersten IP-Paket Daten senden. Die meisten IP-Pakete versuchen, eine Verbindung über den Austausch von IP-Paketen aufzubauen, z. B. bei TCP mit einem 3-Wege-Handshake. Das Internet leitet IP-Pakete über eine vorher definierte Routing-Policy weiter. Das bedeutet, dass zwei IP-Pakete mit dem

gleichen Ziel, die von der gleichen Quelle gesendet werden, nicht den gleichen Weg durch das Internet nehmen müssen und auch nicht in der richtigen Reihenfolge auf der Empfängerseite ankommen müssen. Die TCP-Implementierung löst dieses Problem, indem sie die IP-Pakete in der richtigen Reihenfolge sortiert und fehlende Pakete behandelt. TCP muss jedoch eine Verbindung in drei IP-Paketen aushandeln: Ein erstes TCP-Synchronize-Paket (SYN) von einem Client zeigt einen Verbindungsversuch an. Der Server antwortet mit einem TCP Synchronize Acknowledgement (SYN-ACK), um eine Verbindung herzustellen und zur Bestätigung der hergestellten Verbindung sendet der Client ein abschließendes TCP ACK-Paket zurück an den Server. Danach können die Daten bidirektional gesendet werden und die Implementierung kümmert sich um verworfene IP-Pakete und die richtige Reihenfolge der IP-Pakete. Abbildung 7 zeigt alle aus der Ukraine empfangenen IP-Pakete in der Zeit zwischen dem 01.12.2021 und dem 01.10.2022. Ab Mai 2022, also 2,5 Monate nach Kriegsbeginn, ist eine erhöhte Menge an empfangenen Paketen zu erkennen. Bei genauerer Betrachtung dieser Pakete heben sie sich vom Grundrauschen ab. Viele Pakete sind TCP SYN-ACK-Pakete, die nur dann empfangen werden sollten, wenn das Blackhole-Sensornetzwerk versucht, eine Verbindung aufzubauen. Da wir keine Daten senden, müssen die empfangenen Pakete eine Antwort auf TCP SYN-Pakete sein, welche die IP-Adressen unseres Blackhole-Sensornetzes fälschen. Dieses Verhalten wird durch Angriffe auf die Quell-IP-Adresse erkannt.

5.1 Analyse von DDoS-Angriffen anhand von IPv4-Blackhole-Sensornetzwerk-Daten

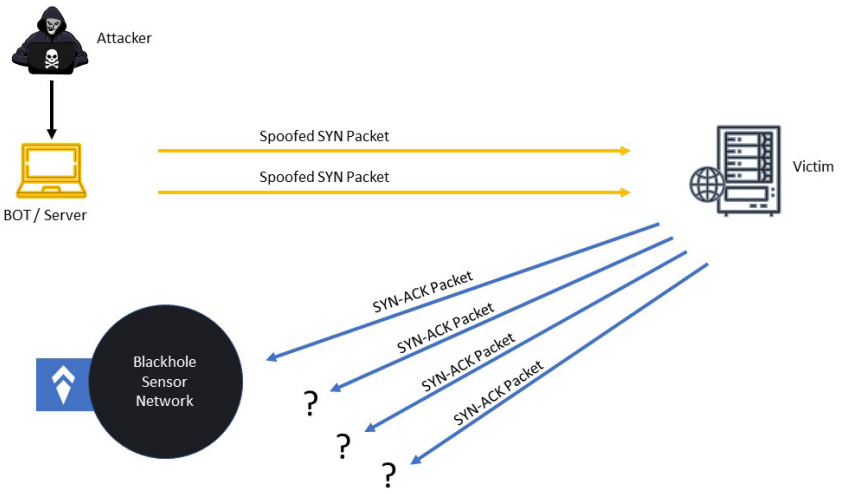


Abbildung 8: Schematische Darstellung eines DDoS-Angriffs mit TCP SYN Flooding

Bei einem DDoS-Angriff (Distributed Denial-of-Service) wird versucht, so viele Ressourcen wie möglich auf dem Opferrechner zu binden, um einen Betriebsausfall zu verursachen. Auf Webservern wird dies sichtbar, wenn die gehostete Website nicht mehr erreichbar ist. Wie in Abbildung 8 dargestellt, kontrolliert ein Angreifer ein Bot-Netz im Internet. Ein Bot-Netz besteht aus vielen mit dem Internet verbundenen Geräten (Bots) wie Computern, Laptops, Smartphones und IoT-Geräten, bei denen der Angreifer eine Sicherheitslücke ausnutzt, um die Kontrolle zu erlangen.

Um einen DDoS-Angriff durchzuführen, installiert ein Angreifer eine Software auf jedem Bot, um so viele TCP SYN-Pakete wie möglich an das Opfer zu senden, ohne die TCP SYN-ACK-Antwort zu verarbeiten. Dies wird als TCP SYN-Flood-Angriff bezeichnet. Jedes Paket reserviert Speicher auf dem Rechner des Opfers, um den TCP-Status zu verarbeiten, bis die Ressourcen oder die Verbindungskapazität erschöpft sind und das Opfer nicht mehr antwortet. Zur Tarnung wählen sie zufällige IP-Adressen aus dem Internet als ihre eigene Quell-IP-Adresse im Paketkopf. Dies führt dazu, dass das TCP SYN-ACK-Paket an die gefälschte IP-Adresse weitergeleitet wird, bei der es sich um einen anderen Host im Internet handelt. Diese Rückstreuung wird von unserem Blackhole-Sensornetzwerk aufgefangen. Wir haben nach Beginn des Krieges begonnen, TCP SYN-ACK-Pakete aus der Ukraine zu erfassen. Anhand der Quell-IP-Adressen, der Quell-Ports der TCP SYN-Pakete und der Datenanreicherung lassen sich die DDoS-Ziele ermitteln. Abbildung 9 zeigt drei verschiedene DDoS-Angriffe. Der erste beginnt mit einem hohen Peak an empfangenen TCP SYN-ACK-Paketen von www.president.gov.ua auf Port 443 um den 1. April 2022. Wir gehen davon aus, dass der Angreifer die Quell-IP-Adressen für seinen SYN-Flood-Angriff zufällig ausgewählt hat, so dass wir die Auswirkungen des Angriffs extrapolieren und abschätzen können. 42.000 IP-Pakete innerhalb von 24 Stunden wurden von unseren 1024 IP-Adressen empfangen. Das Internet besteht aus rund 3 Milliarden routbaren IP-Adressen. Hochgerechnet würde der DDoS-Angriff 1.453.586 TCP SYN-Pakete pro Sekunde auf dem Ziel generieren, was einer Datenrate von rund 697,7 Mbps entspricht. Nach dem Angriff wurde die Website des Präsidenten durch einen bekannten DDoS-Schutzanbieter geschützt. Außerdem wurde die Buchungswebsite der ukrainischen Eisenbahnen (abgekürzt UZ) booking.uz.gov.ua über HTTPS-Port 443 angegriffen und wir erhielten rund 4.500 TCP SYN-ACK-Pakete pro 24 Stunden. Durch Extrapolation dieser Rate gehen wir von einem DDoS-Angriff mit einer Datenrate von etwa 74,8 Mbps aus, der bis zum 1. Juli 2022 andauerte. Die ukrainischen Eisenbahnen änderten die IP-Adresse auf eine andere Adresse innerhalb desselben Netzes, welches von den ukrainischen Eisenbahnen selbst betrieben wird. Der

Beginn und die Dauer der beiden DDoS-Angriffe scheinen zu gezielten Angriffskampagnen zu passen, welche am Beginn eines Monats starten. Eine weitere und langanhaltende DDoS-Attacke richtete sich gegen den Web-Media-Streaming-Server radio.ukr.radio, welcher mehrere verschiedene Radiostreams hostet. Der Dienst ist normalerweise über Port 443 erreichbar, aber der DDoS-Angriff fand über Port 8000 statt und begann um den 20. Juni 2022. Zu Beginn erhielten wir eine Paketrate von 15.000 Paketen pro 24 Stunden (Datenrate rund 249 Mbit/s), die drei Tage lang anhielt und bis zum 6. August 2022 auf ein Niveau von 5.000 TCP SYN-ACK-Paketen pro 24 Stunden (Datenrate rund 83,1 Mbit/s) sank.

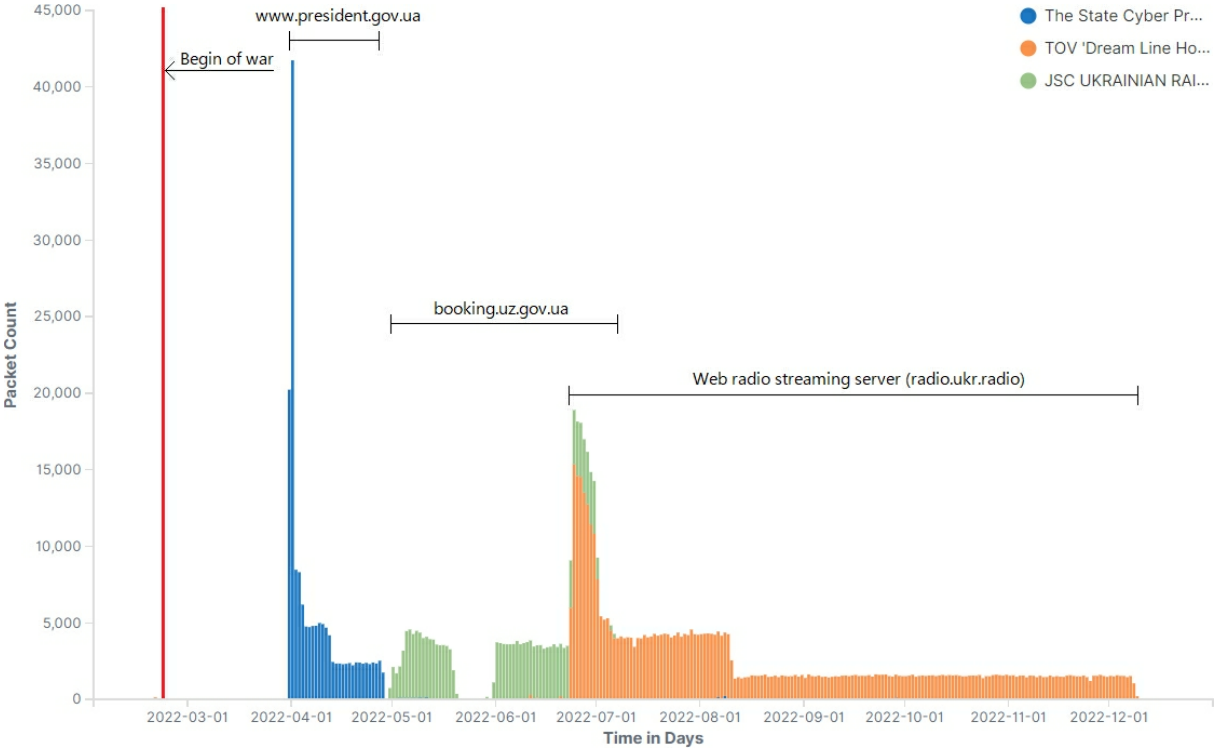


Abbildung 9: Drei ausgewählte DDoS-Ziele und ihre erfassten Pakete: https://www.president.gov.ua, https://booking.uz.gov.ua, https://radio.ukr.radio

6 Zusammenfassung und Ausblick

Die Ergebnisse von Internet-Scans in Kombination mit OSINT-Daten, und Satellitenbildern, zeigen eine starke Korrelation zwischen den Ereignissen wie z. B. Stromausfällen. Intensivere Internet-Scans können verwendet werden, um die Auswirkungen kinetischer Angriffe zu messen, die zu Ereignissen wie z. B. Stromausfällen oder "Battle Damage Assessment" (BDA) führen. Darüber hinaus können Internet-Scans auch dazu verwendet werden, Berichte in sozialen Medien über weit verbreitete schwere Angriffe daraufhin zu überprüfen, ob sie der Realität entsprechen oder eine gegnerische INFOOPS-Kampagne darstellen (Stichwort: Fake News Detection).

Die Einbeziehung einer Informationsquelle aus einer anderen Ebene, z. B. Internet-Scans, kann die Überprüfung von Medieninhalten verbessern und beschleunigen. Anhand unserer eigenen Scandaten konnten wir nachweisen, dass sich groß angelegte Kriegsereignisse stark auf den Cyberbereich auswirken. Als unabhängiger Betreiber einer modularen Suchmaschine haben wir einen eigenen Scan-Algorithmus implementiert, der an die Situation in der Ukraine angepasst ist. Je nach Aufklärungsgebiet genügt es, einige hunderttausend bekannte Hosts zu kennen und deren Erreichbarkeit regelmäßig zu erfassen, um tiefe Einblicke in das Geschehen zu erhalten - von jedem Punkt der Welt aus. Ob die angewandte Methode für diesen Anwendungsfall noch Optimierungspotenzial bietet, ist Gegenstand weiterer Arbeiten.

Weiterhin liefert die Analyse der gesammelten Daten aus einem Blackhole-Sensornetz wertvolle Informationen. Durch Anreicherung jedes Pakets mit zusätzlichen Attributen, Aggregation und Filterung ist es möglich, Übersichten über den Verkehr aus einer bestimmten Region oder einem bestimmten Netzwerk zu erstellen. Insbesondere das Analysieren von TCP SYN-ACK-Paketen bringt eine neue Perspektive auf die Daten. Die Aggregation nach der Quell-IP-Adresse und die Darstellung in einem Diagramm über die Zeit ermöglicht es, DDoS-Angriffe zu verfolgen, wenn zufällige Absender-IP-Adressen aus dem Internet gefälscht werden. So könnten wir verschiedene DDoS-Angriffe auf Ukrainische Institutionen identifizieren und darlegen, dass diese ein sehr spezifisches temporales Muster hatten.

Das hier vorgestellte System der Internet-Scans und die Methodik, welche sich dahinter verbirgt, hat bereits Anerkennung vor internationalem Publikum erhalten [10, 11].

Zusammenfassend haben wir in dieser Arbeit gezeigt, dass es möglich ist, durch Messungen innerhalb des Cyberspace, Rückschlüsse und Aussagen auf reale Geschehnisse in der Ukraine zu ziehen bzw. darzustellen. Globale Internet-Scans sind eine neue Ebene der Aufklärung bzw. Lagedarstellung, zusätzlich zu Satelliten und Radar-Daten, mit dem Unterschied, dass für die Messungen keine Systeme in oder über das Zielland gebracht werden müssen.

Literaturhinweise

[1] Johannes Klick. Globales Internetscanning - Herausforderungen und neue Ansätze oder wie man sein eigener ISP wird, 2019. Accessed 2023-01-06.

[2] Johannes Klick, Stephan Lau, Matthias Wählisch, und Volker Roth. Towards better internet citizenship. In Proceedings of the 2016 Internet Measurement Conference. ACM, November 2016.

[3] Auflistung: MaxMind. Geolite2 city accuracy for broadband networks in the ukraine, 2022. Accessed 2022-12-20.

[4] Ukraine Government. Twitter report; mariupol had lived in peace before russia invaded, ruined 95 percent of the city by heavy bombardment. Available online: https://twitter.com/Ukraine/status/1515783801456181259, April 2022.

[5] United Nation OHCR. 50th session of the human rights council. Available online: https://www.ohchr.org/en/statements/2022/06/high-commissioner-updates-human-rights-council-mariupol-ukraine, Juni 2022.

[6] AlexKozu. 10 october 2022 missile strikes on ukraine. Available online: https://commons.wikimedia.org/wiki/File:10_October_2022_missile_strikes_on_Ukraine.svg, October 2022.

[7] Alex Hardie. Energy minister: About 30 percent of Ukraine's energy infrastructure has been hit by Russian missiles since Monday. Available online: https://edition.cnn.com/europe/live-news/russia-ukraine-war-news-10-11-22/h_2093c5424cc0e8f0c2b366dd9e147957, 2022.

[8] Interfax-Ukraine. Ukrenergo uses backup power schemes to restore power supply. Available online: https://en.interfax.com.ua/news/general/864364.html, 2022.

[9] Witali Kropmann Anatoly Arni. Russia strikes critical infrastructure in ukraine. Available online: https://www.dw.com/ru/rossia-nanesla-raketnye-udary-po-kriticeskoj-infrastrukture-ukrainy/a-63386482, 2022.

[10] Johannes Klick. Fast Global Internet Scanning - Challenges and new Approaches. Chaos Communication Camp (CCC), 2019

[11] Johannes Klick, Robert Koch, und Thomas Brandstetter. Epidemie? Die Angriffsfläche deutscher Krankenhäuser während der Covid-19-Pandemie. In „2021 13th International Conference on Cyber Conflict (CyCon)", Seiten 73-94, 2021.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Striking the Balance: Proactive Detection of Malicious Infrastructure Amidst the Rise of Legitimate Internet Services (LIS) Abuse

Julian-Ferdinand Vögele[1]

Abstract:
Identifying malicious infrastructure proactively aids in detecting associated network traffic, estimating the volume of unreported intrusions, imposing costs on threat actors, uncovering changes in infrastructure patterns, and forecasting trends, among others. While most trends varied between 2022 and 2023, threat actors in both years have increasingly relied on legitimate internet services (LIS) such as Telegram or GitHub as part of their infrastructure, blending their activities within benign network traffic. The increased adoption of LIS, also known as "living off trusted sites" (LOTS), is likely driven by the improved ability of organizations and tools to detect network traffic anomalies. Based on a unique dataset, this paper systematically outlines the abuse of LIS for malicious infrastructure. It explores its advantages and limitations, identifies four primary infrastructure schemes, and highlights key trends. While there is no one-size-fits-all solution suitable for all organizations, the paper offers various mitigation and detection strategies applicable depending on factors such as organizational structure, maturity, availability of logs and resources, and overall risk tolerance. The paper concludes by evaluating the likely evolution of LIS abuse trends and emphasizes the crucial role that LIS themselves play in effective abuse prevention.

Keywords: command-and-control (C2), detection engineering, malicious infrastructure, malware, reverse engineering, threat hunting

1 Motivations for Proactively Identifying Malicious Infrastructure

Malicious infrastructure includes various categories, such as command-and-control (C2) servers, payload and staging servers, exploit servers, management servers, phishing servers, and others. From a more strategic perspective, proactive detections guide updates to risk assessments, influence security control decisions, and enhance the overall understanding of the cyber threat landscape. From a technical perspective, it helps flagging or blocking of related network traffic, but also serves other purposes, including:

- Estimated volume of still unreported intrusions
- Forecast trends by measuring server creation tempo and volume
- Discover changes in infrastructure patterns
- Impose costs on threat actor by “burning” infrastructure
- Identify victims by combining detections with network traffic data

[1] Recorded Future, Berlin

- Attribute malicious activities to threat actors through technical links

2 Behind the Scenes: Methods for Detecting Malicious Infrastructure

Proactively detecting malicious infrastructure can be achieved through diverse methods, with feasibility largely contingent on factors such as data availability (e.g. passive DNS), infrastructure (e.g. scanning capacity), tooling (e.g. probing scripts), skill sets, and other resources (e.g. time). All methods rely on the implicit notion that malicious infrastructure exhibits observable patterns that can be linked to a specific malware (family) or threat actor, ideally before the infrastructure is used in attacks.

2.1 Before the Hack: C2 Weaponization Lifecycle

Detecting malicious infrastructure before its so-called weaponization is often feasible because there are typically multiple steps involved before a malicious server becomes operational[2]. Recorded Future's C2 Weaponization Lifecycle outlines the five essential steps (see **Figure 1**).

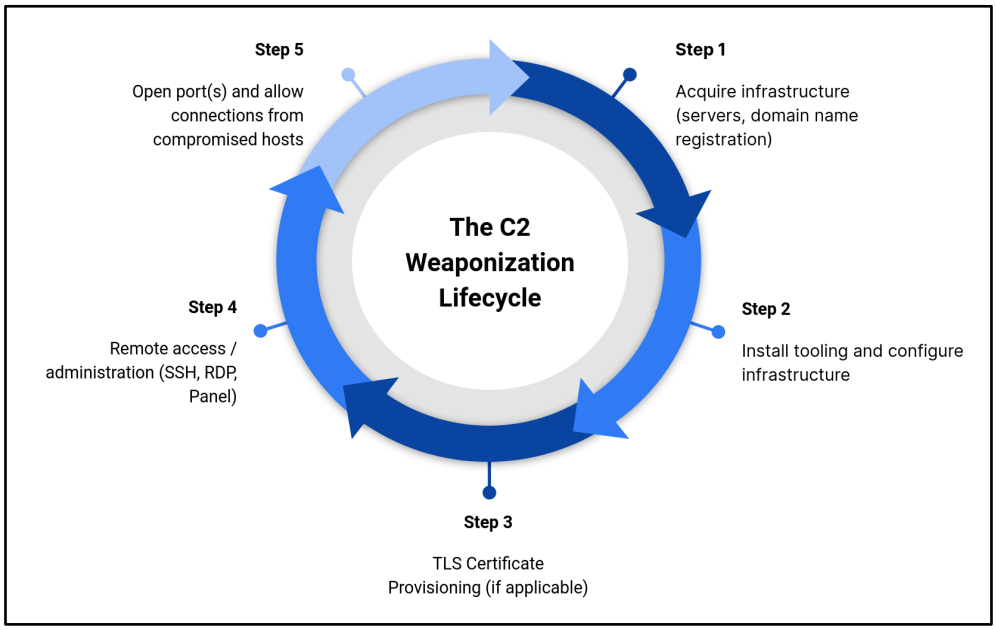


Figure 1: C2 Weaponization Lifecycle (Source: Recorded Future)

However, during this multi-step process, threat actors inadvertently leave behind observable artifacts. These artifacts present opportunities for detection by defenders, including aspects such as software versions, TLS certificate patterns, HTTP responses, ASNs, or even payloads. The ultimate goal of proactive detection is to minimize the time lag between the creation of malicious infrastructure and its identification and classification as such.

[2] https://www.recordedfuture.com/2023-adversary-infrastructure-report

2.2 Limitations in Traditional Detection Methods

Traditional datasets employed for identifying malicious infrastructure include malware repositories (e.g. VirusTotal), passive DNS (e.g. FarSight), and domain registration data (e.g. DomainTools), among others. While these data sources are crucial, they do have two shortcomings:

- Data gaps (e.g. malware repositories depend on the sample being uploaded)
- Inherent latency (e.g. passive DNS depends on the request being made)

These shortcomings raise the question of whether these data sources truly enable proactive detection of malicious infrastructure, highlighting the crucial role played by network scan-based data and its associated benefits.

2.3 Passive and Active Network-Based Detection Methods

Utilizing network scan-based data alongside the aforementioned traditional data sources provides three main advantages: It eliminates latency by allowing scans at any time, is (currently still mostly) confined to the IPv4 space in contrast to the limitless variability of domains, and offers a level of data independence.

Detections based on network scan data are typically based on artifacts like TLS certificates, HTTP responses, payloads, port combinations, and more, and can be broadly categorized into passive and active detections. Passive detections involve querying existing databases (e.g. Shodan) for specific patterns. **Figure 2** provides an example of passive detection for a suspected MuddyC2Go infrastructure[3].

```
HTTP/1.1 404 Not Found
Content-Type: text/html; charset=utf-8
Date: Sun, 07 Jan 2024 19:36:02 GMT
Server: web.go
Content-Length: 14

Page not found
```

Figure 2: HTTP responses observed on 94.131.109[.]65:443 (Source: Recorded Future)

[3] https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/iran-apt-seedworm-africa-telecoms

Active detections, including probe-based methods (e.g. specific set of bytes), require a scan and can be employed to validate whether malicious infrastructure is up at a particular moment in time. **Figure 3** provides an example of a specific endpoint that can be queried to validate instances of Raccoon Stealer C2 servers by checking for the return of the DLL file.

```
http://IP/aN7jD0qO[...]xP7hL2vK/mozglue.dll
```

Figure 3: Active detection to validate Raccoon Stealer C2 (Source: Recorded Future)

3 Key Trends in Malicious Infrastructure in 2023

Based on network data based detections for hundreds of malware families and dozens of threat actors, including financially motivated groups and Advanced Persistent Threats (APTs), Recorded Future gains a comprehensive overview of the state of malicious infrastructure. This extends far beyond publicly reported information and includes not only the volumes of infrastructure per malware family and threat actors but also offers insights into the adoption of new tools, operational security decisions, impact of law enforcement activities, hosting patterns (e.g. by geolocation or hosting provider), malware development pace, and various other trends. For 2023, we identified[4] the following core themes:

- **Open-source and commodity malware** C2 servers maintain their dominance in terms of infrastructure numbers.
- Despite numerous alternatives (e.g. Sliver[5]), **Cobalt Strike** continues to be the predominant C2 framework with a substantial lead.
- **RedLine Stealer** and **Raccoon Stealer** stand out as the leading infostealers in terms of the number of C2 servers detected in 2023.
- **Takedowns of malicious infrastructure**, such as the recent dismantling of the QakBot network, prove to be effective methods for introducing hurdles to malicious operations.
- **Russian state-sponsored actors** often use better operational security for infrastructure (e.g. 1:1 relationship between victim and C2 server) and continue to discover and abuse new LIS.

[4] https://www.recordedfuture.com/2023-adversary-infrastructure-report
[5] https://github.com/BishopFox/sliver

- **Chinese state-sponsored actors** are increasingly using anonymization networks of compromised IoT systems, routers, and other devices[6].

While themes in malicious infrastructure varied between 2022 and 2023, both years witnessed an increased appetite for LIS abuse. This trend was observed not only in association with Russian state-sponsored activities but also among other malware families and threat actors. The lack of a systematic overview of how LIS are abused for malicious infrastructure, both qualitatively and quantitatively, led to an in-depth analysis of the current state of LIS abuse.

4 Under the Radar: The Rise of Legitimate Internet Services (LIS) Abuse

In both 2022 and 2023, threat actors, both cybercriminals and APTs (e.g. BlueBravo[7]), have increasingly been leveraging LIS such as Telegram, OneDrive, Notion, and GitHub as part of their infrastructure, blending their malicious activities within benign enterprise network traffic. One driving factor behind the rising adoption of LIS, also known as "living off trusted sites" (LOTS)[8], is the enhanced capability of organizations and tools to detect anomalies in network traffic (e.g. irregularities caused by custom protocols, non-standard ports, and known malicious or suspicious IP addresses and domains).

4.1 Advantages and Limitations

However, using LIS for malicious infrastructure also offers other advantages. These advantages include:

- No blocking of LIS domains in most corporate networks given that many organizations rely on them for legitimate activities (e.g. Trello).
- Lower operational overhead and decreased susceptibility to errors given the streamlined C2 server installation process.
- Reduced infrastructure costs by avoiding standard hosting or registration fees[9].
- High uptime due to the design of many LIS, prioritizing high availability through redundancy and failover mechanisms.

6 https://www.recordedfuture.com/charting-chinas-climb-leading-global-cyber-power
7 https://www.recordedfuture.com/bluebravo-uses-ambassador-lure-deploy-graphicalneutrino-malware
8 https://lots-project.com/
9 https://blog.avast.com/greedy-cybercriminals-host-malware-on-github

- Minimal vetting for LIS accounts, such as the absence of a credit card requirement, represents significant cost savings for APTs, eliminating the need for time-consuming and complex creation of untraceable payment methods.
- Limited detection possibilities for LIS providers, particularly concerning human-controlled accounts.
- Tracking threat actors upstream or identifying victims becomes increasingly challenging, particularly, if tracking relies on network traffic analysis, encountering an LIS poses a significant obstacle, often leading to a virtual dead end during an investigation.

While these advantages may paint LIS as an ideal choice, leveraging it for malicious infrastructure comes with limitations, including:

- LIS functionality limitations hinder full flexibility (e.g. GitHub cannot host PHP-based tools like phishing kits due to the absence of PHP backend services[10]).
- LIS domains as well as other related static artifacts can in theory be easily blocked within the victim network.
- LIS likely have enhanced visibility into hosted infrastructure compared to scenarios where threat actors manage it independently on dedicated servers, possibly leading to improved understanding of threat actor operations and the identification of victims.
- The presence of dedicated teams within LIS tasked with detecting and countering system abuses causes frictions for threat actors.
- Technical constraints may arise from limitations on file sizes for free accounts or usage and bandwidth restrictions (e.g. Git Large File Storage and GitHub Pages sites have bandwidth limits of 1 GB and 100 GB per month, respectively[11]).
- The lack of privacy in public repositories allows unrestricted access to hosted code and its Git history, potentially enabling researchers to acquire operational insights (e.g. an operational security failure by the North Korean state-sponsored APT37 abusing GitHub in 2023 revealed previously unknown targets and attack vectors[12]).

10 https://www.bleepingcomputer.com/news/security/github-service-abused-by-attackers-to-host-phishing-kits/

11 https://docs.github.com/en/repositories/working-with-files/managing-large-files/

12 https://www.zscaler.com/blogs/security-research/unintentional-leak-glimpse-attack-vectors-apt37

4.2 Types of LIS-Based Infrastructure Schemes

The abuse of LIS for malicious infrastructure purposes can be broadly classified into four main schemes, in addition to a range of other schemes. These infrastructure schemes are not mutually exclusive and may be employed in combination with each other: Full C2, dead drop resolving (DDR), payload delivery, and exfiltration.

4.2.1 Full C2

Full C2 based on LIS occurs when the threat actor and the implant do not directly communicate but instead rely on an "abstraction layer" for exchanging communications such as issuing commands (e.g. Kimsuky abused GitHub for hosting command files in the FlowerPower campaign[13]) (see **Figure 4**). Any service with a public API, enabling programmatic data reading and writing, can theoretically function as such an abstraction layer.

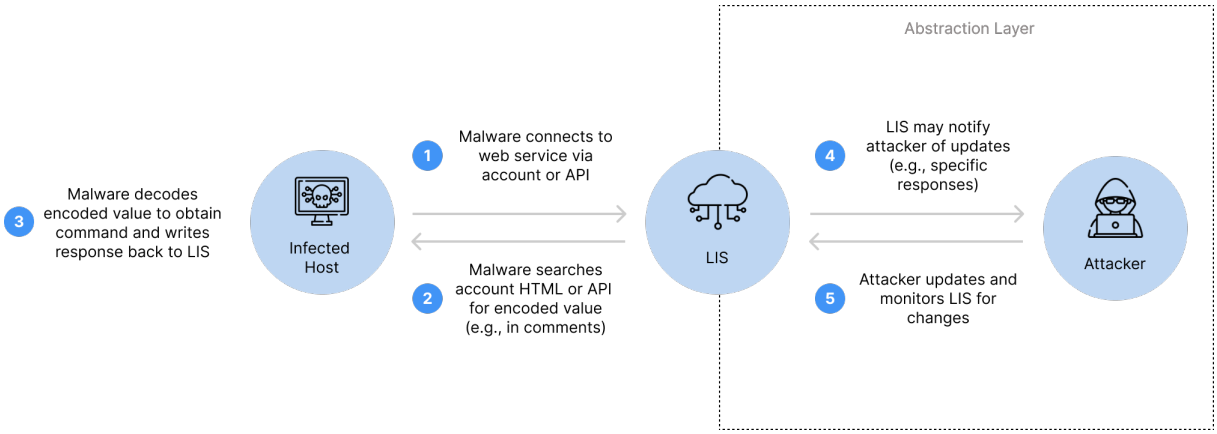


Figure 4: Overview of a full C2 infrastructure setup using LIS (Source: Recorded Future)

A recent noteworthy instance of the full C2 scheme is a campaign by Lazarus Group targeting job seekers, being the first occurrence of the threat actor abusing GitHub for full C2 purposes[14]. The attack starts by executing malicious macros embedded in a decoy Word document. The absence of string encoding simplifies the analysis, as it includes hard-coded details like the username, repo name, directory, and token for GitHub HTTP requests under the username "DanielManwarningRep". The malware fetches files from the image repository containing malicious modules, runs them, and uploads the outcomes of the executed commands to the metafiles directory via an HTTP PUT request.

13 https://www.genians.co.kr/blog/flowerpower

14 https://www.recordedfuture.com/flying-under-the-radar-abusing-gitHub-malicious-infrastructure

4.2.2 Dead Drop Resolving (DDR)

DDR is a technique where malware is configured to obtain its actual C2 server from another location, such as a web service (see **Figure 5**). The term DDR is inspired by conventional intelligence methods, describing a scenario in which an agent discreetly deposits valuable information in an inconspicuous location, commonly known as a "dead drop". While there are cases where the IP addresses or domains of the actual C2 servers are openly listed in plain text (e.g. LimeRAT using Pastebin[15]), threat actors often use encryption and encoding techniques to increase the difficulty of detection (e.g. Astaroth hides encoded C2 information in YouTube comments[16]) as well as other forms of redirection.

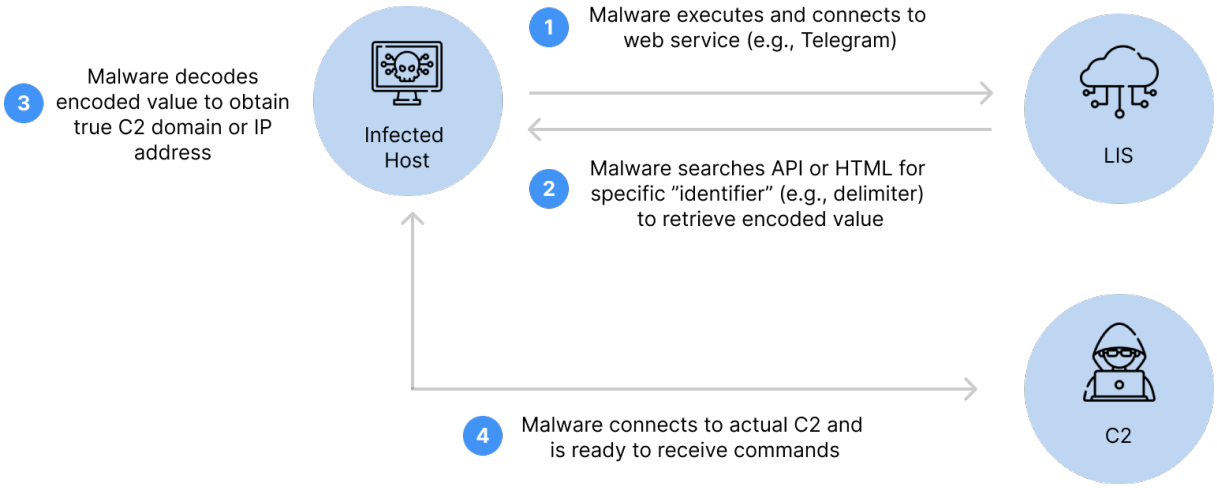


Figure 5: Overview of DDR infrastructure setup (Source: Recorded Future)

Unlike full C2 setups, with DDR, the malware initiates direct communication with the C2 server after obtaining the address from the web service. In essence, any web service that permits data reading can serve for DDR with relevant accounts being either created or stolen by the threat actors.

In September 2023, for example, Recorded Future discovered[17] a campaign by BlueDelta, overlapping with Russia's APT28/Fancy Bear[18], employing a multi-step attack chain in order to target specific locations and evade detection. The campaign relied on GitHub Pages hosting an HTTP redirect script, directing traffic to the free mock API services on mockbin[.]org (see **Figure 6**). BlueDelta also mimicked Windows Update identifiers and Microsoft domains, a tactic previously associated with the threat actor.

15 https://any.run/cybersecurity-blog/limerat-malware-analysis/
16 https://securelist.com/the-tetrade-brazilian-banking-malware/97779/
17 https://www.recordedfuture.com/flying-under-the-radar-abusing-gitHub-malicious-infrastructure
18 https://www.recordedfuture.com/bluedelta-exploits-ukrainian-government-roundcube-mail-servers

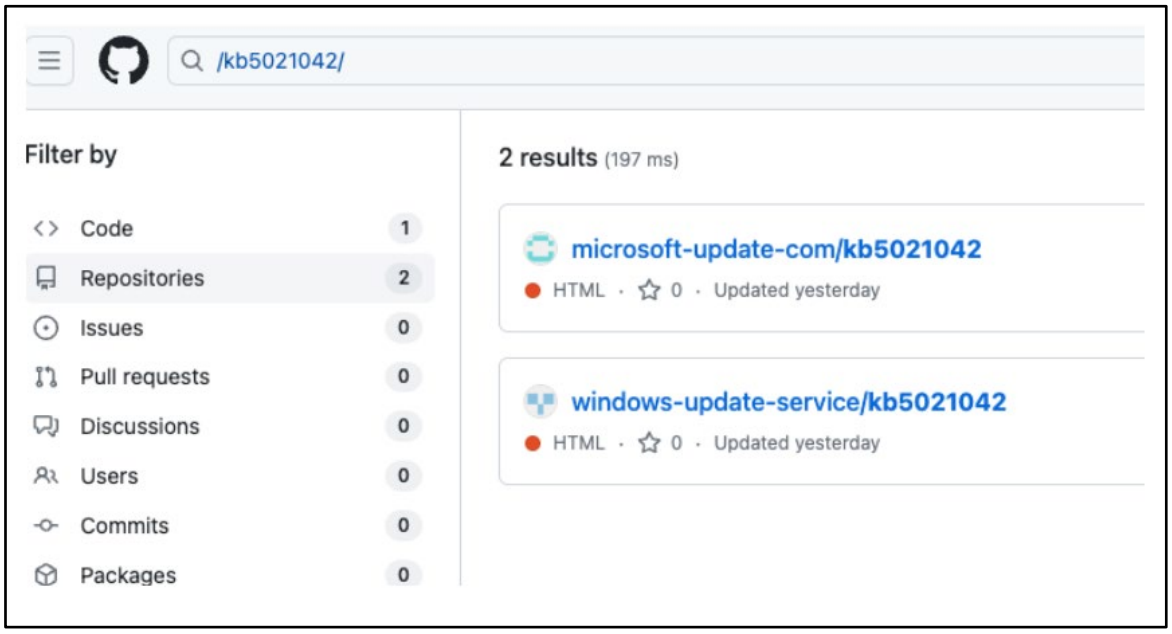


Figure 6: BlueDelta mimicking Windows Update identifiers (Source: Recorded Future)

Another interesting instance involves Mustang Panda[19], which employed a GitHub repository in the exfiltration process to acquire an encrypted token, which is subsequently used to exfiltrate data to Google Drive[20].

4.2.3 Payload Delivery

LIS are frequently used by threat actors for delivering payloads, offering users a way to share and store diverse information (e.g. binaries or text-based data). Their attractiveness for abuse stems from their accessibility and widespread usage, resulting in detection difficulties. Much like DDR, any web service allowing data reading can, in principle, be utilized for payload delivery and threat actors rely on both created and stolen accounts (e.g. Lumma has been spread through breached YouTube accounts[21]). Instances of abuse within the payload delivery infrastructure scheme can be broadly classified into staging and infection-focused scenarios[22].

Staging through LIS is a method of delivering malicious code to a target system, irrespective of the initial infection, which might be achieved through (spear)phishing or exploiting vulnerabilities, among others. Examples include Agent Tesla's loader retrieving base64-encoded, obfuscated code in a

19 https://malpedia.caad.fkie.fraunhofer.de/actor/mustang_panda
20 https://decoded.avast.io/threatresearch/avast-q4-2022-threat-report/
21 https://www.fortinet.com/blog/threat-research/lumma-variant-on-youtube
22 https://www.recordedfuture.com/flying-under-the-radar-abusing-gitHub-malicious-infrastructure

multi-stage process from Pastebin[23], GuLoader loading its encrypted payloads from Google Drive[24], or malware using GitHub Gists for staging[25]. Employing a less commonly observed technique, ClearFake operators host a portion of their malicious code on Binance's Smart Chain contracts[26].

On the other hand, in infection-focused scenarios, the LIS plays a crucial role in the initial infection process itself, for example, through fake repositories or techniques such as repository poisoning. For example, the Leiden Institute of Advanced Computer Science found[27] thousands of GitHub repositories on GitHub hosting fake proof-of-concept (PoC) exploits for a range of vulnerabilities, some of which contained malware such as Cobalt Strike or Houdini RAT.

4.2.4 Exfiltration

LIS are abused for exfiltration, with any web service allowing data writing or sending being theoretically usable for this purpose. This encompasses publicly accessible APIs (e.g. the Telegram Bot API is frequently abused[28] by numerous infostealers, including Snake Keylogger) or email services (e.g. Iranian threat actor OilRig uses email-based exfiltration techniques[29]). While Telegram Bot API and Discord are frequently used for exfiltration, likely due to their user-friendly APIs, cost-free nature, and general popularity in the cybercriminal ecosystem, other LIS such as GitHub are less commonly used for this purpose, possibly due to file-size limitations, lower efficiency, or detectability concerns[30].

4.2.5 Other Schemes

Beyond the four main infrastructure schemes outlined earlier, there are other infrastructure-related purposes for which LIS are abused. For example, GitHub Pages (*.github.io) are frequently used as phishing hosts due to their ease of setup, subdomain customization, and the absence of blocking in victim environments (see **Figure 7**).

---

23 https://news.sophos.com/en-us/2021/02/02/agent-tesla-amps-up-information-stealing-attacks/

24 https://www.crowdstrike.com/blog/guloader-malware-analysis/

25 https://www.reversinglabs.com/blog/malware-leveraging-public-infrastructure-like-github-on-the-rise

26 https://labs.guard.io/etherhiding-hiding-web2-malicious-code-in-web3-smart-contracts-65ea78efad16

27 https://arxiv.org/abs/2210.08374#

28 https://www.scmagazine.com/analysis/abuse-of-telegram-bots-for-credential-phishing-increased-800-in-2022

29 https://therecord.media/oilrig-apt34-iran-linked-hackers-new-downloaders-israel

30 https://www.recordedfuture.com/flying-under-the-radar-abusing-gitHub-malicious-infrastructure

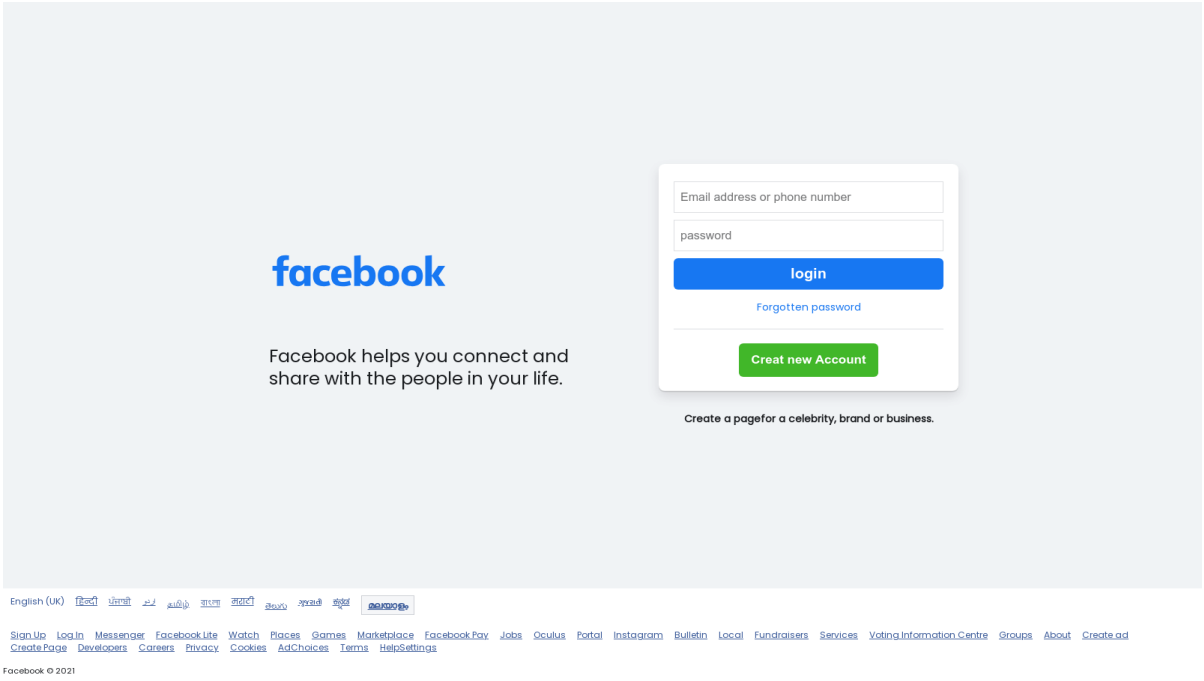


Figure 7: Suspected phishing page hosted on github.io (Source: URLScan[31])

In other cases, LIS may also be used as fallback communication channels in case the primary channel is detected or disrupted (e.g. BlackEnergy was able to communicate over a fallback channel via the now deprecated Google+[32]). Threat actors may opt for particular LIS with the expectation of encountering a lower risk of account takedown compared to traditional server setups. Although still a PoC, DuckDuckGo's image proxy, designed with a focus on user privacy, has been repurposed to function as a C2 channel[33]. Lastly, though not primarily for infrastructure purposes, threat actors abuse LIS to enhance credibility in social engineering, as exemplified by the Lazarus Group's fake GitHub operation[34].

4.3 Key Trends in LIS Abuse in 2023

To get an overview of the current state of LIS abuse, we gathered a dataset primarily comprising malware families identified by Recorded Future Triage, supplemented with additional malware families detected and monitored by Recorded Future through alternative sources. We first categorized the malware families by category (e.g. infostealer) and then evaluated whether, how (i.e., which infrastructure scheme), and which LIS (category) (e.g. social me-

31 https://urlscan.io/result/206169ac-4f53-42d0-8181-29f381535adb/
32 https://securelist.com/be2-custom-plugins-router-abuse-and-target-profiles/67353/
33 https://nopcorn.github.io/2023/09/25/duckduckgo-as-c2
34 https://twitter.com/blackorbird/status/1738026685688434745?s=43

dia) they abused. This classification process involved a combination of manual and automated methods, leading to the creation of an enriched dataset. This dataset enabled the identification of a range of core themes:

- Infostealers, as key innovators in the evolving cybercrime ecosystem, are the most common malware category abusing LIS, benefiting from lower infrastructure demands as opposed to other categories (e.g. RATs) and catering to operators with lower technical expertise.
- There is a likely increase in LIS abuse for malicious infrastructure, evident through the gradual inclusion of LIS abuse support by long-standing malware families (e.g. Agent Tesla), the prevalence among recent commodity infostealers, and the rapid innovation (i.e., more services and more sophistication within services).
- The infrastructure scheme used is heavily dependent on the malware category (see **Figure 8**). For example, while exfiltration is the most prevalent infrastructure scheme, constituting 47 % across all malware families, the share is significantly higher among infostealers (72 %) compared to other malware categories.

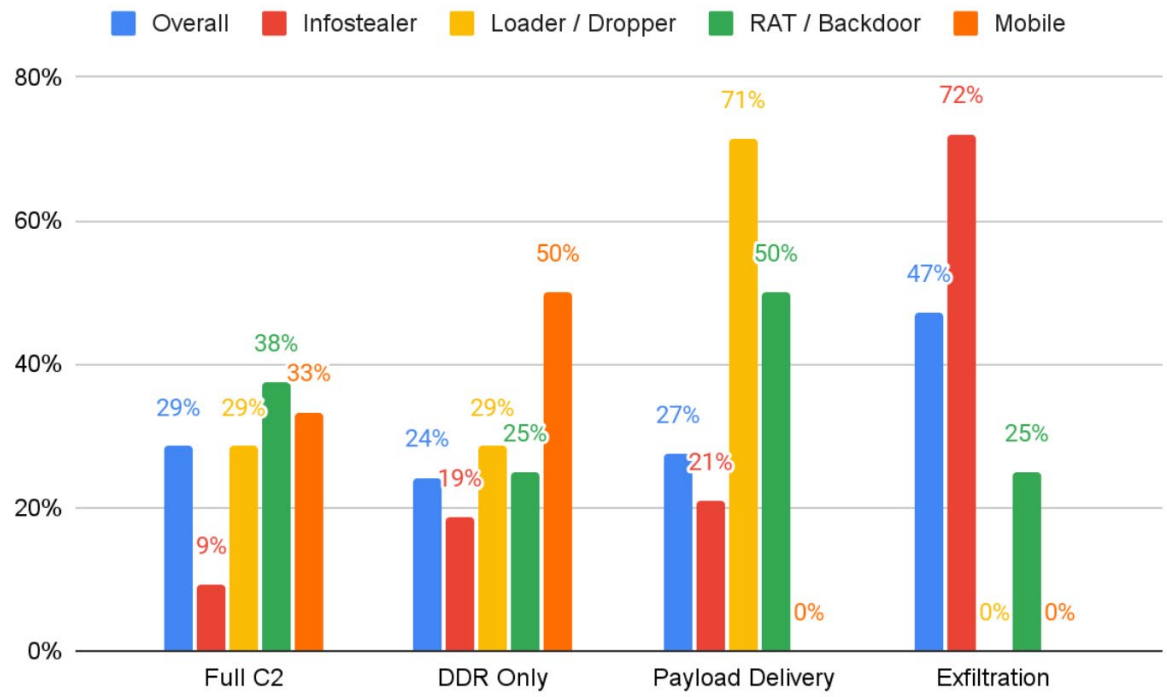


Figure 8: Share of malware families abusing LIS by scheme (Source: Recorded Future)

- Cloud platforms (e.g. Pastebin) are the most commonly abused LIS, likely due to the diverse array of services available, the ability to

blend in, and the ease of implementation. Threat actors predominantly utilize these platforms for payload delivery.

- Telegram is the most commonly abused messaging application (predominantly by infostealers). This is likely due to Telegram's free accessibility, widespread usage making it challenging to block, a user-friendly API, and its established presence in the cybercriminal ecosystem (e.g. for sale[35] or "customer support"[36]).
- APT groups often take the lead in discovering and abusing new LIS (categories), likely resulting in trickle-down effects influencing less-sophisticated groups over time. For example, APT29 was observed using the chat application Zulip for C2[37], quickly following instances of abusing various other LIS in previous campaigns.

5 Striking a Balance: Mitigations and Detection Strategies

Abusing LIS for malicious infrastructure is primarily driven by the desire to evade detection and complicate mitigations. Although there is a range of mitigation and detection strategies, there is no one-size-fits-all solution that is universally applicable to all organizations. Instead, the feasibility of implementing specific mitigation and detection strategies varies based on factors such as organizational structure, maturity, and purpose (e.g. industry), the availability of logs and resources (e.g. detection engineering), and overall risk tolerance. Typical high-level mitigations include:

- Keep an updated and thorough asset inventory that specifies authorized users, including employees, internal IP addresses, or VLANs, permitted to use specific LIS for improved anomaly detection.
- Consider engaging with LIS vendors to proactively and effectively address known malicious activities on their platforms. This also helps painting a precise assessment of LIS abuse in the medium term.
- Incorporate instances of LIS abuse (e.g. red teaming exercises) into regular attack simulations to continuously evaluate your infrastructure's detection capabilities.
- Deploy TLS network interception, taking privacy and compliance implications into account, in order to enhance visibility in response to the widespread adoption of encryption within LIS.

35 https://www.stealthmole.com/blog/telegram-channel-for-fraud-and-cybercrime
36 https://www.secureworks.com/research/the-growing-threat-from-infostealers
37 https://thehackernews.com/2023/08/russian-hackers-use-zulip-chat-app-for.html

- Create flexible security policies that can be adjusted to evolving threat landscapes and requirements, and, if possible, deploy application allowlisting for increased control.
- Secure your employees' LIS accounts, taking into account that threat actors may target them not only for information theft but also to employ them as infrastructure for follow-on attacks, leveraging their ideal nature to blend in.
- Conduct proactive threat hunting to find new instances of LIS abuse or LIS with the potential for being abused. Hunters can search for such instances using a multitude of tools, including website scanners and LIS-specific tools (e.g. to analyze commit histories on GitHub).

While blocking the use of specific LIS is generally an option, it is often impractical due to an arguably increasing number of legitimate use cases, some of which may not be immediately apparent. In this context, organizations need to find a balance between thwarting malicious communication via LIS and refraining from overly restrictive measures. One approach involves defining and combining more nuanced detection strategies beyond file-based methods (e.g. Yara) that rely on understanding employees' LIS usage patterns and requirements, knowledge of the network, and threat intelligence. These detection strategies can be broadly classified into the following categories:

- Context-based detections operate on the principle that if only specific network segments need to interact with a particular LIS (e.g. GitHub API access is only needed by developers), any network traffic directed towards these services from other segments is deemed suspicious.
- Service-based detections operate on the principle that only specific (sub-)services of a LIS are needed in a corporate environment (e.g. only a specific cloud platform is internally supported for file transfer).
- Log-based detections operate on the principle that interactions between a system and a LIS, as evident in both proxy and audit logs, can serve as indicators for malicious or suspicious activity (e.g. LIS connection is made through specific “LOLbins”[38]).
- Detection mechanisms may also take advantage of specific combinations of LIS observed in threat activity, given that the abuse of LIS often entails the simultaneous abuse of multiple of such LIS[39].

[38] Living Off The Land Binaries

[39] https://go.recordedfuture.com/hubfs/reports/cta-2023-0816.pdf

- Network-based detections operate on the principle that malware, especially when using LIS for payload delivery and DDR, connects to malicious infrastructure (e.g. C2) at a later stage. A limitation is that detection may happen after the infection has already occurred.

Defining, implementing, and integrating these detection strategies is a complex undertaking, demanding thorough testing and astute decision-making regarding which detections should be deployed in production, which ones are reserved for hunting, and the criteria for labeling their hits as outright malicious or suspicious. In addition, considering threat actors' reconnaissance activities, it is probable that particularly sophisticated threat actors may also determine the specific LIS use patterns of their targets.

6 What's Next?

Proactive detection of malicious infrastructure results in a more comprehensive, systematic, and timely understanding of the state of malicious infrastructure and, consequently, the cyber threat landscape. This increases the visibility and minimizes the time gap between the infrastructure creation and its identification by (e.g. average detection lead time of 33 days in 2022[40]), thereby preventing intrusions, tracking long-term developments, imposing costs on threat actors, identifying victims, and facilitating attribution. While the landscape of malicious infrastructure is constantly evolving, marked by the frequent emergence, disappearance, or alteration of malware families and their techniques, we have witnessed a notable upswing in LIS abuse over the past years.

This surge, in both quality and quantity, introduces added complexity for security defenders, who now can not simply trust top level domains of common LIS. To get an idea of where LIS abuse is heading to, one must consider the dimensions of the attacker, defender, and LIS itself. On the attackers' side, we anticipate an increase in the proportion of malware families and threat actors abusing LIS, a growth in the number and diversity of abused LIS categories, and an increase in sophistication, likely spearheaded by APT groups. In this context, defenders must navigate the challenge of balancing security enhancement without imposing excessive restrictions achieved through a deeper understanding of their environments and the use of threat intelligence, among others.

However, defenders' influence is limited, and the LIS themselves play a pivotal role in effective abuse prevention (e.g. to apply an allow-list of domains

[40] https://www.recordedfuture.com/blog/2022-adversary-infrastructure-report

to API like DuckDuckGo's image proxy). This involves dedicated teams for detection and counteraction, and more structural shifts in the approach to LIS abuse, including policy adjustments (e.g. Dropbox limiting business storage plans[41]), technical modifications (e.g. Discord using temporary file links for malware prevention[42]), and potential service shutdowns (e.g. AnonFiles[43]). One thing, among others, that remains to be seen in the months to come is how the adoption of large language models (LLMs), already playing a considerable role in diverse software development and security domains[44], will also aid in mitigating LIS abuse from detection and prevention.

41 https://blog.dropbox.com/topics/product/updates-to-our-storage-policy-on-dropbox-advanced

42 https://www.bitdefender.com/blog/hotforsecurity/discord-tightens-security-with-temporary-file-links/

43 https://therecord.media/anonfiles-shuts-down-extreme-abuse-by-users

44 https://github.blog/2023-11-08-ai-powered-appsec/

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

MANTRA – Graphen-basierte Methoden und Modelle zum Austausch, zur Analyse und zur Wissensmodellierung in der Cyber Threat Intelligence

Mirko Ross[1], Rohit Bohara[1]

Kurzfassung:
Die MANTRA-Forschungsgruppe untersucht im Auftrag der Agentur für Innovation in der Cybersicherheit (Cyberagentur) neue Methoden und Modelle zum Austausch von Cyberangriffsmustern (Cyber Threat Itelligence) und deren Analyse. Schwachstellen werden über Software-Lieferketten und Beziehungssysteme vererbt und verbreiten somit ihre Gefährdungspotenziale an alle verbundenen Akteure. Haben Angreifer eine Organisation erfolgreich kompromittiert, können sie das Beziehungsnetzwerk analysieren und die gewonnenen Erkenntnisse zur Maximierung der Angriffswirkung verwenden. Dagegen unterliegen Organisationen in der Abwehr zahlreichen Restriktionen, die es erschweren, detaillierte Informationen zu einem Angriff, beispielsweise Indicators of Compromise (IoC) im eigenen Beziehungsnetzwerk weiterzugeben. Die Gründe hierfür sind vielfältig: Datenschutz, Schutz von Betriebsgeheimnissen, rechtliche Hürden bei laufenden Ermittlungen, Quellenschutz oder Auflagen durch die Mitarbeit von verbundenen Nachrichtendiensten und Behörden. Die Folge sind bei der Abwehr von Cyberangriffen zeitliche Verzögerungen in der Weitergabe von Informationen und mehrfache Transformationen der Informationen. Im Gegenzug können Angreifer nahezu ungehindert Informationen austauschen und Täterwissen zur Optimierung von Angriffen nutzen. Hier besteht eine Asymmetrie zwischen Angreifern und Verteidigern, die Angreifern deutlich Vorteile verschafft. Eine Verschiebung dieser Asymmetrie zugunsten der Verteidiger erfordert ein tiefgreifendes Verständnis des Beziehungsgeflechts der Akteure in der Cybersicherheit sowie die Entwicklung innovativer Lösungen, um die bestehenden Hürden zu überwinden. Ziel muss es sein, die Resilienz Kritischer Infrastrukturen zu stärken. Zur Lösung dieses Problems schlägt MANTRA neue Wege im Teilen, Klassifizieren und Bewerten von Cyberbedrohungen vor. Anstatt Cybersicherheit von reinen technologischen Gesichtspunkten, wie Schwachstellen-Management zu betrachten, versteht MANTRA Cybersicherheit als eine Herausforderung von sozio-ökonomischen Netzwerken. Instrumente zur Verbesserung der Cybersicherheit sollen daher die Aspekte solcher Netzwerke und Ökosysteme berücksichtigen: sicheres Teilen von Informationen direkt zwischen vernetzten Akteuren und die Abbildung von Netzwerkbeziehungen zur Identifikation und Bewertung von Risiken. Dazu müssen Akteure in die Lage versetzt werden, im Kontext der Netzwerkbeziehungen Informationen zu Wissen verarbeiten zu können. MANTRA legt hier einen Fokus auf dezentrale Lösungsansätze mittels Federated Learning in Peer-to-Peer-Netzwerken. Die semantische Abbildung von Kontextbezügen und deren Bewertung soll über Methoden der Graphen-Theorie deutliche Verbesserung zum Status Quo in der Cyberabwehr erzielen.

Stichworte: CTI, Federated Learning, Graphen, KRITIS, MANTRA, Peer2Peer

[1] asvin GmbH, Stuttgart

1 Einleitung

Mit zunehmender Vernetzung der öffentlichen, privaten und wirtschaftlichen Infrastruktur werden Schwachstellen über Software-Lieferketten und über Beziehungssysteme vererbt und verbreiten somit ihre Gefährdungspotenziale an alle verbundenen Akteure. Dieses komplexe und dynamische Bedrohungsumfeld erfordert ein tiefgreifendes Verständnis der Cybersicherheit sowie die Entwicklung innovativer Lösungen, um die Resilienz Kritischer Infrastrukturen zu stärken. Neben dem ansteigenden Ausmaß der Angriffsoberfläche zeigt sich, dass bedrohungsrelevante Informationen vorwiegend zentralisiert gemeldet werden. Dies geschieht aber mit zeitlichem Verzug, mit stark abweichender Qualität und oft ohne ausreichende Kontextinformationen, da Organisationen befürchten, dadurch vertrauliche Informationen preiszugeben oder weil rechtliche Rahmenbedingungen die Weitergabe von Informationen limitieren. Beispielsweise können im Rahmen eines Cyberangriffs sehr konkrete Informationen zum Indicator of Compromise (IoC) [1] nur sehr eingeschränkt oder mit limitiertem Informationsgehalt geteilt werden: Organisationen schützen Informationen zur Infrastruktur, in laufenden Ermittlungsverfahren gibt es rechtliche Hürden für Polizeibehörden, Akteure in Kritischen Infrastrukturen können dem Geheimschutz unterliegen, verbundene Dienste limitieren die Weitergabe oder es besteht kein ausreichendes Vertrauen zwischen Parteien in der Cyberabwehr. Daten und Informationen zu Cyberangriffen werden dadurch nur auf das gesetzlich vorgeschriebene Mindestmaß eingeschränkt geteilt oder können nur mit einer limitierten Gruppe von Akteuren in der Cyberabwehr geteilt werden. In Folge entstehen bei der Abwehr Informationssilos, die ein zeitnahes und wirksames Handeln innerhalb von Akteuren der Kritischen Infrastruktur erschweren. Zudem führen diese Mängel im Informationsaustausch zu einer fragmentierten Wissenslandschaft, in der die Erkenntnisse und Erfahrungen in der Cyber Threat Intelligence (CTI) ungenutzt bleiben können. Wissenschaftliche Arbeiten (vgl. R. Rohan et al. [2]) deuten darauf hin, dass die gegenwärtigen Methoden und Vorgehensweisen im Bereich der Cybersicherheit nicht adäquat auf die erhöhte Skalierung und die komplexen, intransparenten Beziehungsstrukturen reagieren. Innerhalb des MANTRA-Projektes werden über Workshop-Formate und Interviews die Hürden beim Austausch von CTI-Informationen bei Akteuren der Kritischen Infrastruktur erhoben und wissenschaftlich ausgewertet. Diese sind:

- **Unzureichend automatisierte Prozesse**, die durch Medienbrüche eine manuelle Nachbearbeitung von Cyber Threat Intelligence (CTI) benötigen und an die limitierte Ressource Mensch gebunden sind.

- **Ein Mangel an Vertrauen zwischen Akteuren** beim Informationsaustausch, der zu einem unvollständigen Lagebild führt und damit die Umsetzung von Maßnahmen verzögert und erschwert.
- **Silodenken und eine limitierte Sichtweise** von Akteuren auf Lagebilder, welche die Komplexität der bestehenden Beziehungen zu stark vereinfachen oder lückenhaft abbilden und damit Schwachstellen oder schädliche Akteure übersehen.
- **Fehlende Information zum Kontext von Schwachstellen**, deren Angriffspfaden und den Beziehungen von Akteuren, was eine Priorisierung von Maßnahmen erschwert.
- **Fokussierung auf bekannte Schwachstellen (CVE)**, die die Gefahren durch unbekannte Schwachstellen in der Risikobewertung unberücksichtigt lassen.
- **Latenzen im Informationsaustausch** durch zentralisierte Meldeverfahren und Informationsketten, welche Zeitfenster für erfolgreiche Angriffe öffnen.
- **Strafrechtliche Hürden,** die es Akteuren nicht erlauben, Daten und Informationen zu Cyberangriffen, sowie forensische Erkenntnisse außerhalb von Ermittlungsbehörden zu teilen.
- **Rechtliche Hürden im Informationsschutz,** die es nicht erlauben, Daten und Informationen mit Akteuren zu teilen, die einen niedrigeren oder keine Einstufung zur Einsicht von Informationen haben, die als vertraulich und besonders schützenswert klassifiziert sind.
- **Hürden bei Datenschutz und Compliance**, die es nicht erlauben, Daten und Informationen zu teilen, die personenbezogene Informationen oder für betriebliche Abläufe, Produkte und für den Eigenschutz sensible Informationen enthalten.
- **Anwendungshindernisse**, wenn maschinenlesbare Informationen (z.B. Stix) durch Menschen bearbeitet werden müssen und wenn menschenlesbare Informationen (z.B. Texte in PDF, Excel, E-Mail) in semantische Informationen für Maschinenverarbeitung transformiert werden.

2 Lösung

Eine Verbesserung der Situation beim Teilen von CTI-Informationen muss die Komplexität der bestehenden Gegebenheiten berücksichtigen. Diese Komplexität beruht auf rechtlichen, organisatorischen und sozio-ökonomischen Rahmenbedingungen, die nicht oder nur mit erheblichen Anstrengungen von Politik, Gesellschaft und Wirtschaft geändert werden können. Diese Anstrengungen potenzieren sich noch zusätzlich, wenn der Spielraum von

einer nationalen auf eine internationale Ebene erweitert wird. Daher wählt MANTRA[2] einen Lösungsansatz, der es ermöglicht, wesentliche Hürden beim Teilen von Daten mit Bezug zur Cybersicherheit zu senken und Anreizinstrumente für die Mitwirkung zu etablieren.

Ein wesentlicher Baustein in MANTRA ist die Anwendung von Methoden des Federated Learning über Maschine-Learning-Modelle, die in dezentralen Netzwerken trainiert werden. Dabei teilen Akteure nicht direkt Daten zu Cybersicherheitsinformationen, sondern Modelle, die mit diesen Informationen lokal beim Akteur trainiert werden. Der Vorteil: Sensitive Daten verlassen nicht den Sicherheits- und Datenschutz-kontrollierten Bereich der Akteure. Vielmehr werden diese Daten über Maschine Learning in einem Modell als Wissen abstrahiert und aggregiert. Dieses aggregierte Wissen aus den einzelnen Modellen der Akteure, kann dann in einem generellen Modell zusammengeführt werden. Dieses Modell enthält damit das kollektive Wissen aller Akteure, die durch Training ihr spezifisches Wissen hinzugefügt haben. Dieses Wissen wird über Cybersecurity Knowledge Graphs (CSKG) [3] abstrahiert im Model repräsentiert. Diese Abstraktion ermöglicht drei Spielräume:

1. zum einen ermöglicht der CSKG die Abbildung von Kontextbeziehungen auf Basis von CTI-Informationen.
2. Auf diese Kontextbeziehungen können Methoden der Graphen-Mathematik angewendet werden, um neue Erkenntnisse über Zusammenhänge von Vorkommnissen in komplexen Ökosystemen zu gewinnen. Insbesondere für die cybersicherheitsrelevanten Felder der Prävention, Detektion, Reaktion und Attribution.
3. Die Abstraktion über CSKG ermöglicht es Wissen zu teilen, ohne die konkreten Daten preisgeben zu müssen, auf deren Basis der CSKG aus dem Federated Learning Model erzeugt wurde.

Die aus dem Federated Learning Model erzeugten CSKG ermöglichen es damit Akteuren auf ein kollektives Wissen zuzugreifen, ohne dass die Teilnehmer am Lernprozess gegen Policies wie Datenschutz, Recht oder IP verstoßen. Damit das Model-Training im Federated Learning hochwertige Ergebnisse erzielen kann, ist es notwendig, für die teilnehmenden Organisationen eine Governance Struktur zu definieren:

[2] Die Bezeichnung MANTRA ist kein Akronym, sondern nimmt Bezug auf einen Begriff aus dem Sanskrit. Wörtlich übersetzt bedeutet Mantra: „Schutz des Geistes“ und beschreibt eine Silbe, ein Wort, einen Satz, welcher denjenigen beschützt, der das Mantra erhalten hat.

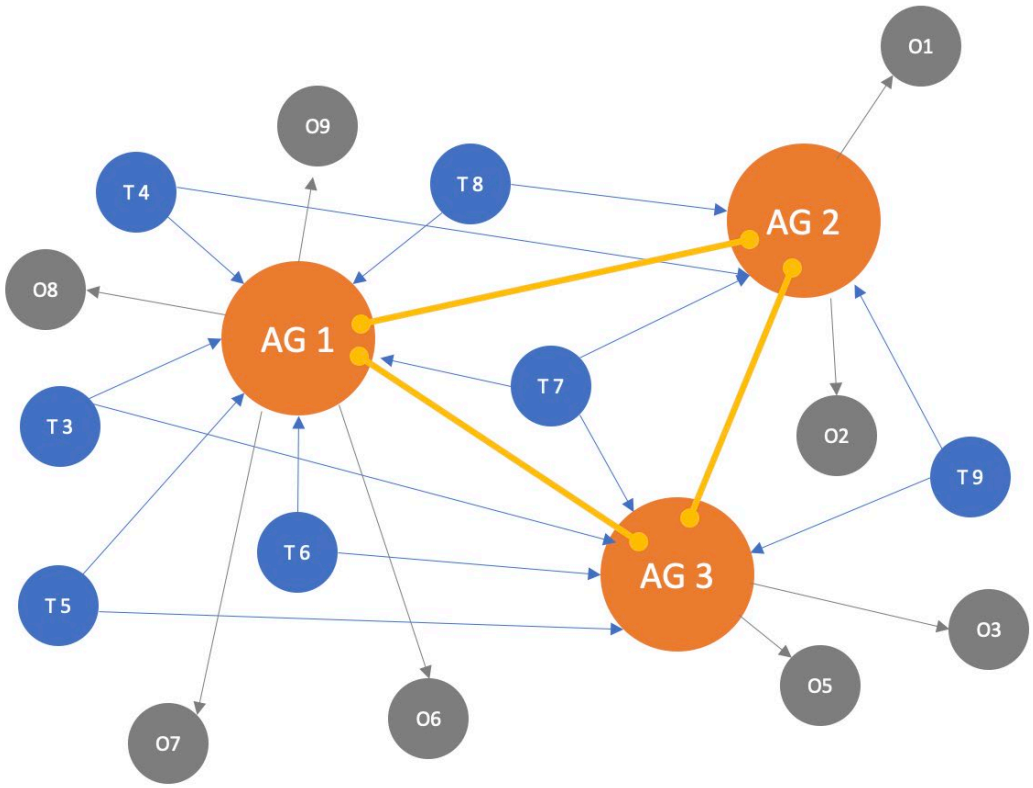


Abbildung 1: Peer-to-Peer Struktur des MANTRA-Federated-Learning-Netzwerk

Die Struktur des Federated-Learning-Netzwerks in MANTRA basiert auf Peer-to-Peer-Prinzipien, wobei es 3 unterschiedliche Klassen von Peers im Netzwerk gibt:

1. **Operational-Peers (O)**: diese können Trainingsmodelle aus dem MANTRA-Netzwerk erhalten und nutzen. So kann ein operationaler Peer das MANTRA-maschine-learning-Modell als Grundlage von eigenen Cybersecurity-Anwendungen verwenden: beispielsweise das Erzeugen von organisationsspezifischen CSKGs und Analysen auf deren Basis. Ebenso können Operational-Peers die Modelle von MANTRA als Basis (Foundation-Modelle) für ein weiteres organisationsspezifisches Training verwenden. Ein Operational-Peer kann allerdings solche antrainierten Modelle nicht wieder in den Pool des Federated Learning zurückspielen. Ein Operational-Peer beschränkt sich somit auf die Anwendung der MANTRA-Basismodelle, beziehungsweise der organisationsspezifischen Erweiterung eines solchen Modells aus dem MANTRA-Netzwerk.
2. **Trainings-Peers (T):** diese dürfen eigene Modelle auf Basis des MANTRA-Basismodel trainieren und diese eigenen Modelle im Prozess des Federated Learnings zur Erstellung einer neuen Version eines Basismodells einbringen. Ein Trainings-Peer hat also die Möglichkeit,

durch sein eigenes Training die nächste Version eines MANTRA-Basismodells zu beeinflussen. Aus diesem Grund tragen Trainings-Peers eine besondere Verantwortung im Training-Prozess. Mit dieser Verantwortung geht einher, dass Trainings-Peers für diese Aufgabe besonders qualifiziert sein müssen.

3. **Aggregations-Peers (AG):** Diese Peers haben wiederum 3 wichtige Aufgaben im MANTRA-Netzwerk: Zum einen sind sie eine vertrauenswürdige Quelle, die die Versionen des Basismodells zur Verfügung stellen. Diese Version kann von Operation-Peers und Trainings-Peers geladen werden. Zum Zweiten entscheiden diese, ob ein Trainings-Peer vertrauenswürdig ist und eine eigene trainierte Version eines Basismodells in den Federated-Learning-Prozess als Commit einsteuern darf. Schlussendlich im Federated-Learning-Prozess muss aus allen Commits eine neue Version des Basismodells erzeugt werden. Dazu verhandeln die Aggregations-Peers im Building-Prozess, welche Commits mit welcher Gewichtung in die neue Modellversion einfließen. Ebenso können Aggregations-Peers entscheiden, dass Commits nicht berücksichtigt werden – beispielsweise im Fall von mangelndem Vertrauen in die Trainingsprozesse oder im Falle einer Kompromittierung eines Trainings-Peers. Die Aggregations-Peers tragen somit die größte Verantwortung im MANTRA-System zur Bereitstellung eines qualitativ hochwertigen Basismodells. Zudem müssen die Aggregations-Peers die Integrität des Basismodells und des Federated-Learnung-Prozesses schützen.

Modell Training im Federated Learning und Deployment in MANTRA

Damit aussagekräftige Cybersecurity Knowlede Graphs (CSKG) durch das MANTRA-Basismodell erzeugt werden können, ist es wichtig, dass der Modell-Training- und Modell-Deployment-Prozess mit einer hohen Qualität an Trainingsdaten durchgeführt wird. Zudem muss die Integrität der Modelle im gesamten Prozess des Federated Learnings gewahrt und nachvollziehbar sein, damit das System gegen Angriffe auf Modelle oder Trainingsdaten, beispielsweise durch Data Poisoning, geschützt ist. Ein Trainings-Peer hat vier Funktionsebenen, die geschützt werden müssen:

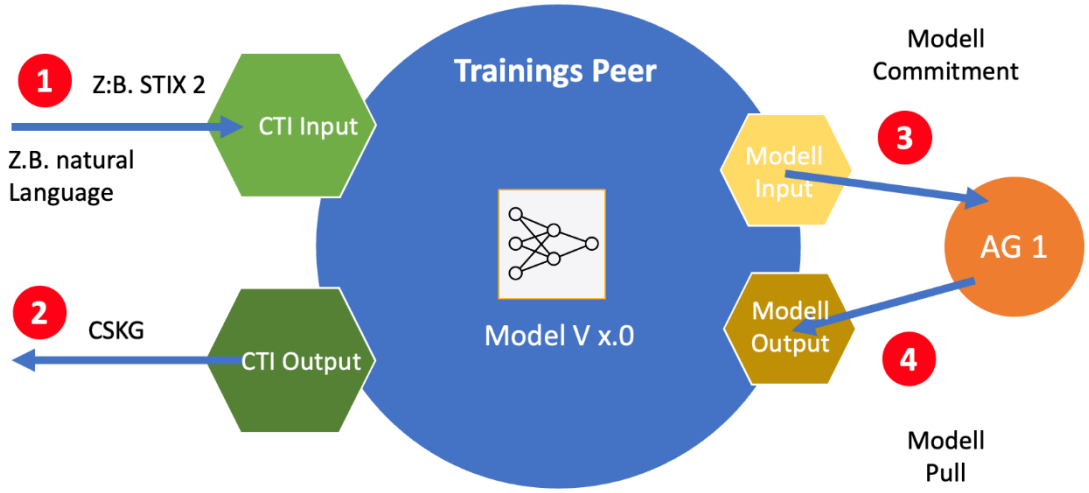


Abbildung 2: Federated-Learning-Prozesse im Trainings-Peer

1. Die Eingabe von Daten in das Modell-Training: Hier können CTI-Daten für das Trainieren eines MANTRA-Modells verwendet werden. Bevorzugt werden dazu semantisch strukturierte CTI-Daten, beispielsweise im STIX 2 [4] Format. Sind CTI-Informationen unstrukturiert, beispielsweise als Textinformation im PDF-Format vorhanden, können diese mit Hilfsmitteln wie Large-Language-Modellen (LLMs) in STIX 2 konvertiert werden. Hierzu kann beispielsweise ein LLM auf Basis von MISTRAL [5] genutzt werden, um Informationen in natürlicher Sprache in semantisch strukturierte STIX 2 Daten zu wandeln.
2. Das Modell beantwortet Fragestellungen mit der Ausgabe eines Cybersecurity Knowledge Graphs (CSKG). Der CSKG bildet die Grundlage für Anwendungsschichten, die auf dem MANTRA-System aufbauen, insbesondere in den Bereichen: Prävention, Detektion, Reaktion und Attribution. Zudem können auch weitere Aspekte zur Resilienz von Infrastrukturen über CSKG in Simulationen ermittelt werden [6].
3. Zu einem von Aggregations-Peers definierten Zeitpunkt werden die Modell-Trainingsstände von den Trainings-Peers angefordert. Die Modell-Trainingsperiode auf dem Trainings-Peer ist damit beendet. Das antrainierte Modell wird an den Aggregations-Peer übermittelt.
4. Nach der Aggregation liefern die Aggregations-Peers eine neue Version des Basis-Modells an die Trainings-Peers. Der Trainings-Peer übernimmt diese neue Version des Basis-Modells für die aktuelle Trainingsperiode.

Im Federated-Learning-Prozess haben Trainings-Peers (T) und Aggregations-Peers (AG) eine herausragende Stellung und Beziehung. Ziel ist es mit

jeder Trainingsperiode die Qualität des MANTRA-Basismodells zu verbessern. Dies setzt voraus,

- dass Trainings-Peers (T) ein Modell mit qualitativ hochwertigen Daten trainieren, da die Qualität der Daten einen direkten Einfluss auf die Modell-Qualität hat. Ebenso müssen auf dem Trainings-Peer IT-Sicherheitsmaßnahmen implementiert sein, um sowohl Daten, Training und das Modell vor Angriffen und Manipulationen zu schützen.
- dass die Peer-to-Peer-Kommunikation zwischen Trainings-Peers und Aggregations-Peers geschützt ist, damit keine Manipulationen an Modellen, beispielsweise durch man-in-the-middle-Angriffe erfolgen können. Ebenso muss es für beide Parteien möglich sein, den Integrationsschutz eines Modells prüfen zu können, beispielsweise über die Bindung von Modellen an digitale Zertifikate. Denkt man diese Anforderung deutlich in die Zukunft, sollten die eingesetzten Sicherheitsmethoden und Verfahren die Anforderung an Post-Quantum-Encryption erfüllen.
- dass eine ausreichende Anzahl von Trainings-Peers in ausgewogener Verteilung die Gewichte zur Erstellung eines neuen Basismodells beisteuern, damit die Aggregations-Peers die Commits der Trainings-Peers verhandeln können. Aus dem Ergebnis dieser Verhandlung wird ein neues Basismodell durch die Aggregations-Peers erstellt und vom Trainings-Peer für den nächsten Trainingszyklus verwendet.

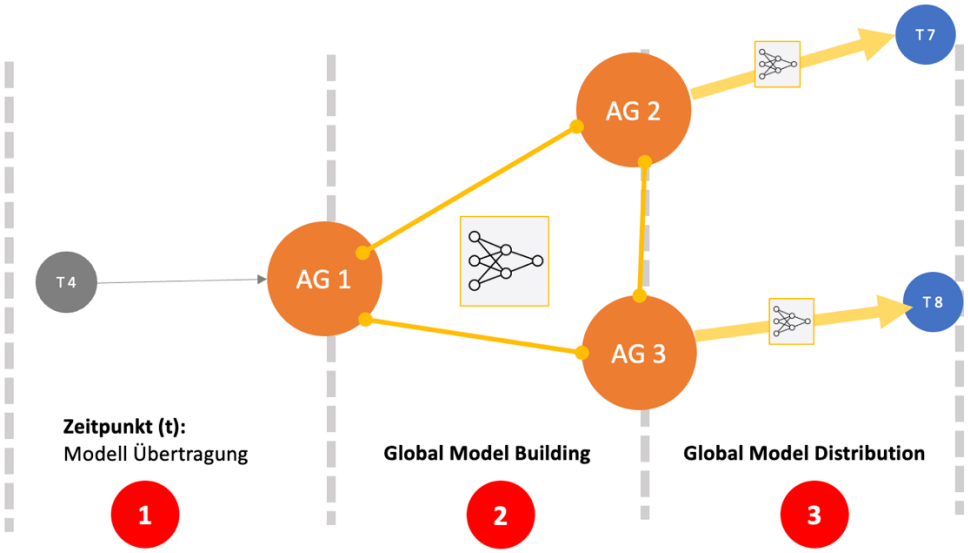


Abbildung 3: Der Modell-Deployment-Prozess in MANTRA in 3 Schritten

Basismodelle im Operational-Peer (O) für CSKG-Anwendungen

Die MANTRA-Basismodelle aus dem Federated Learning bilden die Wissensbasis, auf der CSKG für Cybersecurity-Anwendungen bereitgestellt werden. Im MANTRA-Projekt sind diese für den Anwendungsfall für Akteure in Kritischen Infrastrukturen ausgelegt. Dabei können für die Anwendungsfälle spezifische Basismodelle ausgelegt werden, die durch Akteure mit spezifischen für den Anwendungsfall ausgelegten CTI-Daten trainiert werden können. Ein Operational-Peer (O) kann das für den Akteur jeweils notwendige spezifische Anwendungsmodell im Peer-to-Peer-Netz laden und für seine fachspezifische Anwendung betreiben. Dabei können eines oder mehrere spezifische Basismodelle auf einem Operational-Peer (O) in Anwendungen integriert werden:

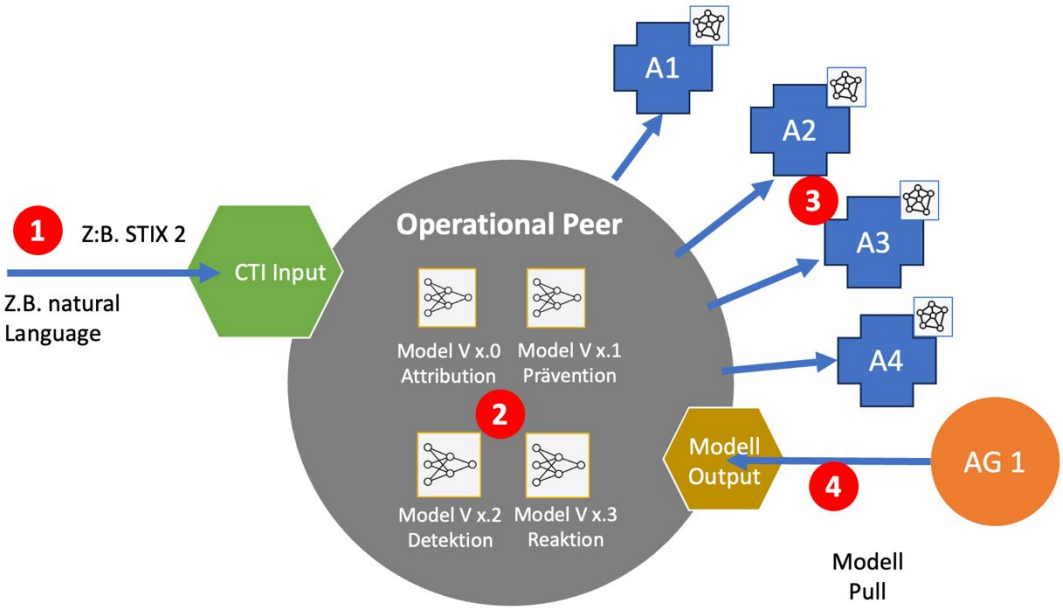


Abbildung 4: MANTRA-Basismodelle in der CSKG-Anwendung auf Operational-Peers

Anwendung 1 (A1) Prävention: Dieses Modell erfasst Beziehungen zwischen potenziellen Schwachstellen und ermöglicht die Abbildung von Angriffspfaden im CSKG. Das Modell wird dazu ausgelegt, das Ausbreitungspotenzial von Angriffen innerhalb und zwischen Organisationseinheiten analysieren zu können. Damit soll es ermöglicht werden, priorisierte Cybersicherheits-Entscheidungen und -Maßnahmen zu treffen. Eine darauf basierende Anwendung kann beispielsweise Simulationen zur Cyber-Resilienz einer Infrastruktur durchführen und Aussagen zur Resilienz gegenüber Cyberangrif-

fen treffen (vgl. R. Pal et al. 2023 [6]). Mit den Ergebnissen können Investitionen in Schutzmaßnahmen sowie die Ressourcenplanung besser priorisiert werden, um die Resilienz gegenüber Cyberangriffen zu erhöhen.

Anwendung 2 (A2) Detektion: Eine auf CSKG basierte Analyse von Angriffspfaden bietet Kontextinformationen für eine bessere Erkennung von Anomalien auf der Ebene der Mustererkennung. Detektierte Anomalien werden über den CSKG in funktionale sowie organisatorische Beziehung gesetzt. Dieser Kontext kann beispielsweise zur Früherkennung von Cyberangriffen genutzt werden.

Anwendung 3 (A3) Reaktion: Im Fall eines Cyberangriffs ist die Auswahl der richtigen Gegenmaßnahmen eine komplexe Herausforderung. Deren Wirksamkeit hängt von verschiedensten Einflussfaktoren und Beziehungen ab. Zudem müssen Entscheidungen oftmals unter hohem Zeitdruck gefällt werden. Manuelle Prozesse sind zeitintensiv und können deshalb dazu führen, dass Zeit zur Bewältigung eines Angriffs verloren geht. Die Bereitstellung von Kontext-Information und deren Aufbereitungen von Beziehungen in der semantischen Struktur eines CSKG unterstützt eine automatisierte Verarbeitung der Erkenntnisse. Im Fall eines Cyberangriffs können durch MANTRA über die automatisierte Verarbeitung auf Grundlage der Kontext-Informationen aus dem CSKG eine bessere Priorisierung von Gegenmaßnahmen ermöglicht und Latenzen in der Abwehr reduziert werden.

Anwendung 4 (A4) Attribution: Das im MANTRA-Modell antrainierte Wissen zu Angriffspfaden und Beziehungen kann zur Erkennung von Verhaltensmustern zu Tätern und Tätergruppierungen in der Attribution genutzt werden. MANTRA ermöglicht eine organisationsübergreifende Analyse auf Grundlage von CSKG. Im Modell trainierte Erkenntnisse über Attributionen können dabei über Mustererkennung im CSKG zum Vergleich neuer Angriffspfade genutzt werden. Der Deckungsgrad von übereinstimmenden Mustern im Graph liefert Indizien für weitere forensische Untersuchungen bei der Attribution von Angriffen

3 Zusammenfassung

MANTRA beschreibt einen Weg des Austausches von Wissen zur Cybersicherheit (CTI) in komplexen Ökosystemen verschiedener Akteure mittels Methoden des Federated Learnings. Dieser Ansatz ermöglicht die Aggregation von Wissen unter den bestehenden rechtlichen und organisatorischen Restriktionen bei der Behandlung von sensitiven Daten, insbesondere bei Akteuren in Kritischen Infrastrukturen und Behörden. Dieses Wissen kann aus den trainierten Basismodellen über Cybersecurity Knowledge Graphs

(CSKG) extrahiert werden, um in der Auswertung über Kontext-basierte Informationen Mehrwerte für Anwendungsbereiche der Cybersicherheit zu schaffen: Prävention, Detektion, Reaktion und Attribution. Damit MANTRA sein Potential ausschöpfen kann, ist die Akzeptanz und das Vertrauen von Akteuren in der Cybersicherheit und in Kritischen Infrastrukturen notwendig. Die Wirksamkeit des in MANTRA beschriebenen Lösungsansatzes steigt mit der Partizipation von Akteuren am Federated-Learning-Prozess. Dazu ist es im weiteren Verlauf der Forschung notwendig, Methoden zu identifizieren, die die Hürden einer Partizipation senken. Dazu zählen sozio-ökonomische Anreize, Governance-Strukturen, Sicherheit und technische Werkzeuge im Peer-to-Peer-Verfahren. Dazu gibt es in MANTRA-Forschungsgruppen, die sich mit Teilaspekten zur Lösung dieser Aufgaben beschäftigen.

Diese Studie/Forschungsarbeit wurde durch die bundeseigene Agentur für Innovation in der Cybersicherheit GmbH beauftragt und finanziert. Eine Einflussnahme der Agentur für Innovation in der Cybersicherheit GmbH auf die Ergebnisse fand nicht statt.

Literaturverzeichnis

[1] P. Funken et al., MANTRA: A Graph-based Unified Information Aggregation Foundation for Enhancing Cybersecurity Management in Critical Infrastructures, 2023, Open Identity Summit 2023, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2023 123, https://publica-rest.fraunhofer.de/server/api/core/bitstreams/79bb0aa4-0c1d-477d-9434-1d9497f84336/content

[2] R. Rohan et al.: A systematic literature review of cybersecurity scales assessing information security awareness, 2023, DOI:https://doi.org/10.101 6/j.heliyon.2023.e142342023.

[3] L. Sikos.: Cybersecurity Knowledge Graphs, 2023, 65(9):1-21DOI:10.1007/s10115-023-01860-3

[4] A. Jiang et al.: Mistral 7B assessing information security awareness, 2023, arXiv:2310.06825 [cs.CL]

[5] F. Marchiori et al.: STIXnet: A Novel and Modular Solution for Extracting all STIX Objects in CTI Reports, 2023, rXiv:2303.09999v2 [cs.IR]

[6] R.Pal et al.: A Mathematical Theory to Quantify Cyber-Resilience in IT/OT Networks, 2023, INFORMS Winter Simulation Conference

[7] K. Liu, et. al, Recent Progress of Using Knowledge Graph for Cybersecurity, 2022, *Electronics*, *11*(15), 2287; https://doi.org/10.3390/electronics11152287

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Governance und Security bei der End-To-End-Automatisierung in der Cloud am Beispiel der Microsoft Power Platform

Michael Arns[1]

Kurzfassung:
In Deutschland hat sich Microsoft Office 365 als integraler Bestandteil von täglichen Geschäftsprozessen vieler Unternehmen etabliert. Insbesondere Programme wie MS Word, PowerPoint und Excel sind in zahlreichen Unternehmensprozessen nicht mehr wegzudenken. Zudem führt die verstärkte Umstellung auf Remote-Arbeit dazu, dass MS Teams in vielen Unternehmen rasch an Bedeutung gewinnt. Neben den bekannten Tools haben Mitarbeiterinnen und Mitarbeiter durch die Office-365-Konten zudem Zugriff auf weitere, weniger bekannte Anwendungen von Microsoft, wie der Power Platform. Obwohl viele Unternehmen großen Wert auf die Einhaltung der Informationssicherheit legen und dafür spezielle Abteilungen eingeführt haben, die Mitarbeitende schulen, Prüfungen durchführen und potenzielle Sicherheitslücken identifizieren, übersehen sie die Tatsache, dass diese Microsoft Tools oft uneingeschränkt von den Mitarbeiterinnen und Mitarbeitern genutzt werden können. Mit der Power Platform haben User dank des Low-Code/No-Code-Ansatzes die Möglichkeit, ohne vorherige Programmierkenntnisse, als sogenannte Citizen Developer, Automatisierungen vorzunehmen. Es klingt zunächst äußerst vielversprechend, dass Abteilungen in der Lage sind, spezifische Workflows und Apps zu erstellen, um sich damit die alltägliche Arbeit zu erleichtern. Da jedoch viele Unternehmen die Microsoft Power Platform nicht als Teil ihrer strategischen Planung berücksichtigen und daher kein Rollen- und Berechtigungskonzept oder standardisiertes Application Lifecycle Management (ALM) bereitstellen, können Citizen Developer kritische Sicherheitslücken im Unternehmen verursachen. Workflows und Apps entstehen, ohne dass überhaupt jemand in der IT-Abteilung etwas davon mitbekommt. Es entsteht eine Schatten-IT.

Um dieses Szenario zu vermeiden, möchte der Vortrag „Governance und Security bei der End-to-End-Automatisierung in der Cloud am Beispiel der Microsoft Power Platform“ zum einen ein Bewusstsein für die Thematik schaffen, da der Trend zu Low Code/No Code-Tools stetig zunimmt und auch in Zukunft von weiteren Anbietern zur Anwendung kommen wird. Zum anderen werden die Möglichkeiten der Prävention sowie der strategische Umgang mit Citizen Developern vorgestellt. In einem mehrstufigen Governance-Modell werden die entscheidenden Schritte erläutert, die die Informationssicherheit eines Unternehmens sicherstellen. Dieses Modell behandelt fünf Schlüsselbereiche, darunter Mandantenzugriffe, die sich mit den Zugriffen auf Microsoft 365 in der Azure Cloud befassen, durch deren Einschränkungen die IT-Abteilung die Vertraulichkeit der Zugriffe kontrollieren kann. Die Umgebungsstrategie legt klare Umgebungen fest, um erstellte Apps mit Zugriff auf interne Informationen zu verwalten und die Sicherheit dieser Daten

[1] Almato AG - Ein Unternehmen der Datagroup SE

in diesem Kontext zu gewährleisten. Bei der Festlegung und Verwaltung des Ressourcenzugriffs geht es um die Gewährleistung der Integrität von Workflows und den Schutz vor unautorisierten Änderungen. Ein weiterer bedeutender Faktor bei der Einhaltung der Governance besteht darin, sich der potenziellen Sicherheitsrisiken bewusst zu sein, die von Konnektoren ausgehen und Datenleaks zur Folge haben könnten. Beim letzten Schlüsselbereich wird das On-Premises Data Gateway behandelt, das für eine schnelle, sichere Datenübertragung zwischen lokalen Daten und den Diensten der Microsoft Power Platform fungiert. Unternehmen können sämtliche Maßnahmen individuell implementieren, um maßgeschneiderte unternehmensweite Richtlinien zu etablieren und einzuhalten. So kann auch in Zukunft die Informationssicherheit bei der End-To-End Automatisierung in der Cloud sichergestellt werden.

Stichworte: Citizen Developer, End-to-End-Automatisierung, Low Code/No Code-Tools, Power Platform, Robotic Process Automation (RPA), RPA

1 Einführung in das Thema

In der modernen Geschäftswelt ist die zunehmende Verlagerung von Prozessen in die Cloud unübersehbar. Diese Entwicklung hat zu einer Revolution in der Art und Weise geführt, wie Unternehmen ihre Abläufe automatisieren und optimieren. Die Microsoft Power Platform ist ein weit verbreitetes Tool bei dieser Transformation zur End-To-End-Automatisierung in der Cloud. Als integraler Bestandteil des weit verbreiteten Microsoft Office 365 Ökosystems, bietet die Power Platform Unternehmen eine flexible Plattform für die Entwicklung von Geschäftsanwendungen und die Automatisierung von Prozessen durch Low-Code/No-Code-Lösungen. Dieser Ansatz ermöglicht es, dass auch Nicht-Programmierer, sogenannte Citizen Developer, aktiv an der Digitalisierung und Automatisierung von Geschäftsprozessen teilnehmen können.

1.1 Informationssicherheit bei der RPA-Entwicklung

Die Integration von Robotic Process Automation (RPA) in die Geschäftsprozesse, insbesondere in Verbindung mit der Power Platform, wirft wichtige Fragen in Bezug auf die Informationssicherheit auf. Die effiziente Handhabung sensibler Daten und die Einhaltung von Compliance-Richtlinien sind dabei von zentraler Bedeutung. Angesichts der zunehmenden Automatisierung von Abläufen müssen Unternehmen sicherstellen, dass ihre Sicherheitsstrategien mit diesen technologischen Entwicklungen Schritt halten.

1.2 Kurze Vorstellung der Microsoft Power Platform

Die Microsoft Power Platform ist ein integriertes Anwendungsentwicklungsset, das die Erstellung und Bereitstellung von geschäftsorientierten Applikationen ermöglicht, ohne dass dafür tiefgreifende Programmierkenntnisse erforderlich sind. Sie setzt sich aus fünf Hauptkomponenten zusammen: Power

BI, Power Apps, Power Automate, Power Virtual Agents (Copilot Studio) und Power Pages, die alle auf der gemeinsamen Datenservice-Plattform von Microsoft, bekannt als Dataverse, aufbauen.

Power BI ist ein Tool zur Geschäftsanalyse, das es Nutzern ermöglicht, Daten aus einer Vielzahl von Quellen zu visualisieren und zu analysieren. Es bietet eine breite Palette von Datenverbindungsoptionen, Berichterstellungsfunktionen und interaktiven Dashboards, die es Entscheidungsträgern erleichtern, datengesteuerte Entscheidungen zu treffen.

Power Apps ist eine Software, die es Benutzern ermöglicht, sowohl einfache als auch komplexe Anwendungen mit wenig oder keinem Code zu entwerfen. Mit vorgefertigten Vorlagen können Benutzer Anwendungen erstellen, die auf verschiedenen Geräten funktionieren und sich in bestehende Datenquellen und externe Anwendungen integrieren lassen.

Power Automate, früher bekannt als Microsoft Flow, ist ein Tool zur Automatisierung von Workflows zwischen verschiedenen Anwendungen und Diensten. Es ermöglicht Benutzern, automatisierte Prozesse zu erstellen, die repetitive Aufgaben eliminieren und die Effizienz verbessern. Die Benutzer können aus vorgefertigten Vorlagen wählen oder eigene Flows von Grund auf neu erstellen.

Power Virtual Agents (Copilot Studio) ermöglicht es, intelligente Chatbots zu erstellen, die mit den Benutzern interagieren und ihnen helfen, Fragen zu beantworten oder Aufgaben zu erledigen. Diese Bots können auf Websites, in Diensten wie Microsoft Teams oder anderen Kanälen eingesetzt werden und erfordern keine Kenntnisse in der Programmierung von KI oder maschinellem Lernen.

Power Pages zur Erstellung von interaktiven datengesteuerten Geschäftswebsites.

Die Power Platform ist eng mit anderen Microsoft-Diensten und Anwendungen wie Office 365, Dynamics 365 und Azure integriert. Dabei ist das Konzept der Konnektoren wesentlicher Bestandteil dieser Integration. Das Dataverse spielt eine zentrale Rolle in der Power Platform, indem es eine einheitliche Datenbank bietet, die eine konsistente Datennutzung über alle Komponenten der Plattform hinweg ermöglicht. Es unterstützt komplexe Datenmodelle, Sicherheitsmodelle und Logiken, die das Fundament für Anwendungen und Analysen bilden, die auf der Power Platform entwickelt werden.

Ein entscheidendes Merkmal der Power Platform ist die Einfachheit und Benutzerfreundlichkeit. Die Low-Code- bzw. No-Code-Funktionalität ermög-

licht es Geschäftsanwendern, sogenannten Citizen Developers, eigene Lösungen zu erstellen und anzupassen, ohne auf die IT-Abteilung angewiesen zu sein. Dies fördert die Demokratisierung der Anwendungsentwicklung innerhalb von Organisationen und ermöglicht es Fachabteilungen, schnell auf sich ändernde Geschäftsanforderungen zu reagieren, erfordert aber auch ein durchdachtes Governance Konzept.

1.3 Chancen bei der Nutzung und Aufbau von Citizen-Developern

Die Nutzung und der Aufbau von Citizen-Developern bieten Unternehmen eine Reihe von Chancen, die digitale Transformation voranzutreiben und die IT-Entwicklung zu demokratisieren. Citizen-Developer sind Endbenutzer, die unternehmensspezifische Anwendungen mit wenig oder keinem Coding erstellen, indem sie Plattformen wie die Microsoft Power Platform nutzen. Diese Entwicklung spiegelt einen wachsenden Trend wider, der es Mitarbeitern ohne technischen Hintergrund ermöglicht, digitale Werkzeuge zu gestalten und anzupassen, die ihre tägliche Arbeit unterstützen.

Eine der größten Chancen, die sich aus der Nutzung von Citizen-Developern ergibt, ist die Beschleunigung der Anwendungsentwicklung. Indem Fachkräfte, die direkt mit Geschäftsprozessen vertraut sind, eigene Lösungen erstellen können, wird die Abhängigkeit von IT-Abteilungen reduziert und die Umsetzungsgeschwindigkeit erhöht. Dies fördert eine agile Arbeitsweise, da Anwendungen schnell adaptiert und iterativ verbessert werden können, um den sich wandelnden Anforderungen des Geschäftsalltags gerecht zu werden.

2 Herausforderungen der Informationssicherheit bei der Nutzung von Microsoft Power Platform

Die Implementierung von Cloud-basierten Automatisierungslösungen wie der Microsoft Power Platform birgt eine Vielzahl von Herausforderungen im Bereich der Informationssicherheit. Diese Herausforderungen sind komplex und erfordern eine gründliche Analyse sowie die Entwicklung einer durchdachten Strategie, um die Integrität und Sicherheit der Unternehmensdaten zu gewährleisten. Zu den Herausforderungen gehören die Sicherstellung der Datenintegrität und -vertraulichkeit in einer Umgebung, in der möglicherweise sensible Unternehmensdaten in der Cloud gespeichert und verarbeitet werden. Die Integration der Power Platform in vorhandene IT-Infrastrukturen und Datenbanken muss sicher erfolgen, um potenzielle Sicherheitslücken zu minimieren. Die Gewährleistung der Zugriffskontrolle und die Verhinderung unbefugter Zugriffe auf sensible Daten sind ebenfalls entschei-

dend, insbesondere vor dem Hintergrund, dass Mitarbeiter eigenständig Lösungen innerhalb der Power Platform entwickeln können. Die Einhaltung von Datenschutzvorschriften und -richtlinien stellt eine weitere wichtige Herausforderung dar, insbesondere in Branchen mit strengen Compliance-Anforderungen wie dem Gesundheitswesen, dem Finanzwesen und der Regierung.

2.1 Schatten IT vs. Citizen Developer

Das Befähigen von Citizen-Developern durch Low-Code/No-Code-Plattformen birgt das Risiko der Schatten-IT: unkontrollierte und oft unsichtbare IT-Projekte innerhalb einer Organisation. Solche Projekte entstehen außerhalb der offiziellen IT-Abteilung und entziehen sich damit häufig den etablierten Sicherheits- und Compliance-Standards. Die Herausforderung besteht darin, die Innovationskraft der Citizen-Developer zu nutzen, während gleichzeitig Sicherheitsrisiken minimiert und Governance-Standards eingehalten werden.

2.2 Anforderungen an Datenschutz und Compliance

Mit der wachsenden Verbreitung von Cloud-Diensten wie der Power Platform steigen auch die Anforderungen an Datenschutz und Compliance. Unternehmen müssen sicherstellen, dass ihre Nutzung dieser Technologien mit lokalen und internationalen Datenschutzgesetzen wie der DSGVO in Einklang steht. Dies umfasst die Gewährleistung der Datensicherheit, die Einhaltung von Speicher- und Verarbeitungsvorschriften sowie die Implementierung von Maßnahmen zur Überwachung und Berichterstattung.

2.3 Risiken der Low-Code/No-Code Entwicklung

Die Low-Code/No-Code Entwicklung ermöglicht eine schnelle und effiziente Erstellung von Anwendungen, birgt aber auch Risiken. Dazu zählen die potenzielle Vernachlässigung von Sicherheitsaspekten, da die Entwickler oft nicht über das notwendige Sicherheitsbewusstsein verfügen, sowie die Schwierigkeit, den Überblick über alle erstellten Anwendungen und deren Zugriffsrechte zu behalten. Eine proaktive Sicherheitsstrategie, die sowohl technische als auch organisatorische Maßnahmen umfasst, ist daher unerlässlich.

2.4 Gefahren durch Zugriff und Datenlecks

Die Möglichkeit, dass unautorisierte Personen Zugang zu sensiblen Unternehmensdaten erlangen, ist eine ständige Bedrohung. Datenlecks können durch menschliche Fehler, unzureichend gesicherte Schnittstellen oder durch die unsachgemäße Verwendung von Cloud-Diensten entstehen. Um

solche Gefahren zu minimieren, ist eine umfassende Strategie erforderlich, die sowohl technische Sicherheitsmaßnahmen als auch Schulungen und Richtlinien für die Benutzer umfasst.

3 Best Practices für die Einhaltung der Informationssicherheit - Mehrstufiges Governance-Modell

Um die Herausforderungen der Informationssicherheit bei der Nutzung von Cloud-Plattformen wie der Microsoft Power Platform zu bewältigen, ist die Implementierung von Best Practices unerlässlich. Diese Praktiken sollen nicht nur die Sicherheit stärken, sondern auch die produktive Nutzung der Plattform unterstützen. Ein effektives Governance-Modell ist entscheidend, um die Sicherheit in der Cloud zu gewährleisten. Ein mehrstufiges Governance-Modell bietet einen strukturierten Ansatz, um sicherzustellen, dass die Nutzung der Power Platform den Unternehmensrichtlinien entspricht. Dies umfasst die Definition von Rollen und Verantwortlichkeiten, die Einrichtung von Richtlinien für die Anwendungsentwicklung und -nutzung sowie die Überwachung und Kontrolle von Datenzugriffen und -flüssen. Eine solche Herangehensweise ermöglicht es, die Vorteile der Citizen-Developer zu nutzen, während gleichzeitig Sicherheitsstandards aufrechterhalten werden. Das Governance Model zur Microsoft Power Platform umfasst 5 Stufen:

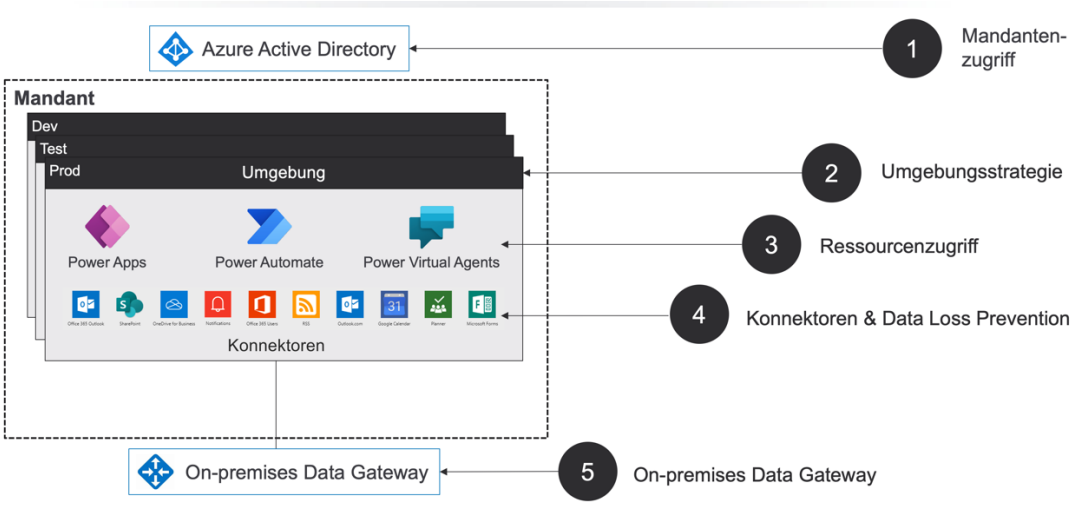


Abbildung 1: Fünfstufiges Governance-Modell zur Microsoft Power Platform

3.1 Mandantenzugriff (Zugang und Isolation auf Tenant-Ebene)

Die Sicherheit beginnt auf der obersten Ebene des Tenants. Hier werden Richtlinien festgelegt, die für alle Umgebungen und Ressourcen innerhalb des Tenants gelten. Diese Ebene ist entscheidend, um sicherzustellen, dass der Datenfluss zwischen verschiedenen Organisationen, die Geschäfte miteinander machen, sicher ist. In der Power Platform ist ein Mandant eine dedi-

zierte Instanz der Plattform, die für eine bestimmte Organisation oder Abteilung eingerichtet wird. Jeder Mandant ist von den anderen isoliert und verfügt über eigene Ressourcen wie Datenbanken, Benutzerkonten und Berechtigungen. Der Zugriff auf einen Mandanten wird durch Azure Active Directory (Entra ID) gesteuert. Benutzer müssen über ein Entra-ID-Konto verfügen, um sich bei der Power Platform anmelden zu können. Berechtigungen steuern, welche Aktionen Benutzer in der Power Platform ausführen können. Berechtigungen können auf Mandantenebene oder auf Umgebungsebene zugewiesen werden. Die Power Platform bietet verschiedene Mechanismen, um die Isolation auf Mandantenebene zu gewährleisten:

- **Datenbankisolation**: Die Daten jedes Mandanten werden in einer separaten Datenbank gespeichert. Dies verhindert, dass Benutzer eines Mandanten auf die Daten eines anderen Mandanten zugreifen können.
- **Berechtigungsgrenzen**: Berechtigungen können auf Mandantenebene zugewiesen werden. Dies bedeutet, dass Benutzer nur Zugriff auf die Ressourcen des Mandanten haben, für den sie Berechtigungen besitzen.
- **Sicherheitskonfiguration**: Die Sicherheitskonfiguration der Power Platform kann auf Mandantenebene angepasst werden. Dies ermöglicht es Organisationen, ihre Sicherheitsanforderungen zu erfüllen.

3.2 Umgebungsstrategie (Environment Access & Strategy)

Umgebungen sind Container, die Administratoren nutzen können, um Apps, Workflows, Verbindungen und andere Assets zu verwalten, zusammen mit den Berechtigungen, die es Organisationsnutzern erlauben, diese Ressourcen zu verwenden. Umgebungen sind an einen geografischen Standort gebunden und können für verschiedene Zielgruppen oder Zwecke wie Entwicklung, Test und Produktion verwendet werden. Es gibt 7 verschiedene Arten von Umgebungen:

Umgebungsart	Beschreibung
1. Default	Eine für jeden Mandanten automatisch erstellte Standardumgebung, die von allen Nutzern gemeinsam genutzt wird.
2. Produktionsumgebung	Eine Umgebung, die für die permanente Arbeit in einem Unternehmen gedacht ist
3. Sandbox-Umgebung	Abgetrennte Umgebung, für Entwicklungs- und Testzwecke ohne Auswirkungen auf andere Umgebungen

4. Testumgebung	Kurzfristige Testumgebung, die auf einen Benutzer beschränkt ist und nach 30 Tagen gelöscht wird
5. Entwicklungsumgebung	Umgebung, die speziell zur Entwicklung bestimmt ist
6. Teams-Umgebung	Teams-Umgebungen werden automatisch für ein Team in MS Teams erstellt, wenn zum ersten Mal eine App in Teams mit Power Apps erstellt wird oder eine Power App aus dem Katalog installiert wird
7. Supportumgebung	Eine Umgebung, die von Microsoft auf direkte Anweisung geschaffen wird, um Probleme, die den Onlinedienst betreffen, zu beheben

Der Zugriff auf Umgebungen wird durch Berechtigungen gesteuert, die Benutzern und Teams zugewiesen werden können. Es gibt verschiedene vordefinierte Rollen mit unterschiedlichen Berechtigungen, z. B. Administrator, Entwickler, Tester und Benutzer. Bei der Festlegung sollte das Prinzip des geringsten Privilegs angewendet werden, um Nutzern nur die Berechtigung zuzuweisen, die sie für ihre Aufgabe benötigen.

3.3 Ressourcenzugriff (Resource Permissions)

Diese Ebene bezieht sich auf die Berechtigungen, die Benutzern für den Zugriff auf bestimmte Ressourcen innerhalb einer Umgebung gewährt werden. Es gibt verschiedene Arten von Berechtigungen, die für Ressourcen zugewiesen werden können, z. B. Lesen, Schreiben, Löschen, Teilen und Verwalten. Die richtige Zuweisung von Ressourcenberechtigungen ist entscheidend, um eine sichere und effektive Governance zu gewährleisten. Auch hier ist das Prinzip des geringsten Privilegs ratsam.

3.4 Konnektoren und Datenverlustrichtlinien (Connector Access and Data Loss Policies).

Auf dieser Ebene werden Richtlinien definiert, die den Zugriff auf Konnektoren steuern und Datenverlust verhindern. Konnektoren sind die Schnittstellen, die die Kommunikation zwischen den Power-Platform-Diensten und externen Diensten ermöglichen. Es ist wichtig, Konnektoren in unterschiedliche Risikogruppen zu kategorisieren, um ungewünschten Datenverlust zu vermeiden. Für jede Kategorie sollten entsprechende DLP-Richtlinien gelten.

Abbildung 2: Konnektoren ermöglichen die Kommunikation zwischen den Power-Platform-Diensten und externen Diensten. Für den Zugriff auf die Konnektoren sollten DLP-Richtlinien definiert werden, um unerwünschte Datenabflüsse zu verhindern.

3.5 On-premises Data Gateway

Diese Komponente fungiert als Brücke für die sichere Datenübertragung zwischen lokalen Datenquellen und Cloud-Diensten der Power Platform wie Power BI, Power Automate, Logic Apps und Power Apps. Updates werden für das lokale Daten-Gateway nicht automatisch installiert. Die Version des Gateways sollte immer aktuell sein.

4 Zukunftstrends

Die Landschaft der Informationstechnologie, insbesondere im Bereich der Cloud-Plattformen und der Automatisierung, entwickelt sich ständig weiter. Diese Dynamik bringt sowohl neue Herausforderungen als auch Chancen mit sich, die es im Kontext der Informationssicherheit zu betrachten gilt.

Die Cloud-Technologie wird voraussichtlich weiter an Bedeutung gewinnen, mit einem zunehmenden Fokus auf Sicherheit, Skalierbarkeit und Flexibilität. Technologische Fortschritte, wie verbesserte künstliche Intelligenz und fortschrittlichere Automatisierungsfunktionen, werden neue Möglichkeiten für Unternehmen eröffnen, ihre Prozesse zu optimieren und zu personalisieren. Gleichzeitig werden diese Entwicklungen Unternehmen vor neue Sicherheitsherausforderungen stellen, insbesondere in Bezug auf Datenmanagement und -schutz.

Künstliche Intelligenz wird eine zunehmend wichtigere Rolle in der Informationssicherheit spielen. KI-basierte Systeme können dabei helfen, Sicherheitsbedrohungen proaktiv zu erkennen und darauf zu reagieren, indem sie

Muster in Datenverkehr und Nutzerverhalten analysieren. Diese Technologien könnten dazu beitragen, die Effizienz von Sicherheitsmaßnahmen zu steigern und neue Sicherheitsstandards zu setzen.

In einer sich schnell verändernden IT-Welt ist die kontinuierliche Anpassungsfähigkeit von Unternehmen von entscheidender Bedeutung. Dies umfasst die regelmäßige Aktualisierung von Sicherheitssystemen, die Anpassung von Geschäftsprozessen an neue Technologien und die fortlaufende Schulung der Mitarbeiter, insbesondere im Bereich der Informationssicherheit.

Die zukünftige Landschaft der Informationstechnologie und Automatisierung wird zweifellos von Innovationen und Weiterentwicklungen geprägt sein, die sowohl Chancen als auch Herausforderungen für die Informationssicherheit darstellen. Unternehmen, die proaktiv auf diese Trends reagieren und ihre Strategien entsprechend anpassen, werden am besten positioniert sein, um von diesen Entwicklungen zu profitieren.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Die Entwicklung des CSAFversums: 17½ Monate nach dem Big Bang

Dr. Dina C. Truxius[1]

Kurzfassung:
Die Anzahl an veröffentlichten Schwachstellen steigt stetig. Da diese potenziell zu jeder Zeit in Soft- und Hardware auftreten können, ist das Wissen von und über sie und ihre Behebungsmaßnahmen für die Risikoab- und einschätzung elementar. Schwachstellen- und Mitigationsinformationen werden üblicherweise in Form von Security Advisories (deutsch: Sicherheitsinformationen) veröffentlicht. Hierbei ist die Herausforderung, dass es viele unterschiedliche Formate, Dokumentstrukturen und Kommunikationswege zur Übermittlung dieser Informationen gibt. Daher ist der Prozess der Informationsbeschaffung und Risikobewertung mit einem enormen personellen und zeitlichen Aufwand verbunden. Darüber hinaus sind Sicherheitsinformationen oft nur in menschenlesbarer und nicht in maschinenlesbarer Form vorhanden. Risikobasiertes Schwachstellenmanagement, das eine gewisse Automatisierung von Prozessen erlaubt, ist zwingend notwendig, um der heutigen Informationsfülle adäquat begegnen zu können und auch für die Zukunft gewappnet zu sein. Das Common Security Advisory Framework (CSAF) 2.0, das im November 2022 als offener Standard veröffentlicht wurde, erfüllt ebendiese Voraussetzungen und revolutioniert so das risikobasierte Schwachstellenmanagement.

Stichworte: Automatisierung, Behebungsmaßnahmen, Common Security Advisory Framework, CSAF, Mitigationsmaßnahmen, Risikobewertung, Schwachstellen, Security Advisory, Sicherheitsinformationen, VEX, Vulnerability Exploitability eXchange

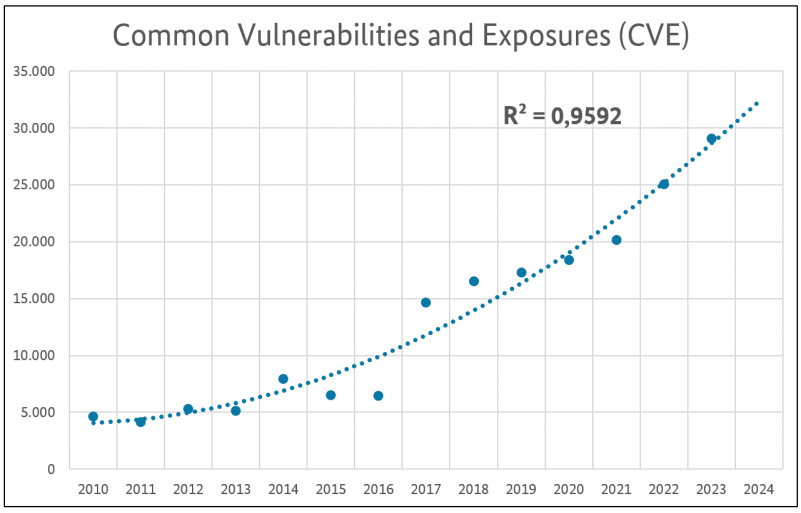
1 Am Anfang war das Chaos…

Nahezu täglich werden weltweit neue Sterne, Galaxien, Sonnen oder auch Tierarten entdeckt. Ebenso verhält es sich mit IT-Sicherheitslücken in Systemen und Komponenten. Die Anzahl sicherheitsrelevanter Schwachstellen, die veröffentlicht wurden [1], stieg in den letzten Jahren so erheblich an (Abbildung 1), dass die Handhabung dieser Informationen nicht mehr sinnvoll oder nur mit erheblichem personellem Aufwand manuell bearbeitbar ist [2].

Im Zuge einer immer stärker vernetzten, automatisierten und dadurch auch komplexeren Welt wird die Anzahl an Schwachstellen perspektivisch nicht geringer werden. Es ist zu erwarten, dass sie signifikant wachsen wird, da sich statistisch gesehen in mehr als 75 % aller Applikationen mindestens eine Sicherheitslücke befindet [3, 4]. Insofern ist jedwede Soft- und Hardware ab einem gewissen Komplexitätsgrad als fehlerbehaftet anzusehen und sind IT-Sicherheitslücken omnipräsent. Im Falle von sicherheitsrelevanten

[1] Bundesamt für Sicherheit in der Informationstechnik (BSI), Bonn

Schwachstellen, die ausgenutzt werden, können diese als Eintrittsvektoren für direkte oder auch weitergehende Angriffe verwendet werden. Sicherheitslücken zu ignorieren ist daher keine Option. Es ist auch nicht sichergestellt, dass eine Schwachstelle, die aktuell keine Sicherheitsrelevanz für ein bestimmtes Produkt oder eine Komponente hat, nicht in Zukunft doch gefährlich werden könnte. Der Reputations- und Vertrauensverlust in Produkte oder ganze Organisationen, der durch ausgenutzte Sicherheitslücken entstehen kann, ist schlichtweg nicht in Zahlen messbar. Es wird also langfristig nicht mehr um die Frage gehen, ob ein System oder eine Komponente von sicherheitsrelevanten Schwachstellen betroffen ist, sondern wie schnell bekannte Sicherheitslücken geschlossen werden. Daher ist zeitgemäßes Schwachstellen- und Risikomanagement über alle Unternehmensgrößen und Bereiche hinweg nicht mehr wegzudenken.



Jahr	Anzahl der CVE
2010	4.639
2011	4.150
2012	5.288
2013	5.142
2014	7.948
2015	6.494
2016	6.457
2017	14.645
2018	16.512
2019	17.308
2020	18.375
2021	20.161
2022	25.059
2023	29.065
2024	~ 33.000

Abbildung 1: Grafische Darstellung und tabellarische Auflistung aller sicherheitsrelevanten und veröffentlichten Schwachstellen (CVE, Common Vulnerabilites and Exposures) in den Jahren 2010-2023 (Quelle: https://www.cve.org/, [1]). Der Kurvenfit zeigt für das Jahr 2024 eine zu erwartende Anzahl von ungefähr 33.000 Schwachstellen.

Um Informationen zu Schwachstellen in Produkten oder in einer Infrastruktur und die zugehörigen Behebungsmaßnahmen zu kommunizieren, werden Sicherheitsinformationen, sogenannte "Security Advisories", veröffentlicht. Von Unternehmen zu Unternehmen und von Advisory zu Advisory können sich die Struktur, der Inhalt, die Qualität und das Format (z. B. HTML, PDF, TXT, etc.) zur Kommunikation von Schwachstellen und den zugehörigen Behebungsmaßnahmen erheblich unterscheiden. Darüber hinaus ist nicht abzusehen, wie sich die Gefährdungslage entwickelt und wann es neu entdeckte Schwachstellen oder weitergehende Informationen zu bereits bekannten Sicherheitslücken gibt und auf welchen Wegen (Mail, Download von einer Webseite, Feed, etc.) und

in welchem Zeitraum die Konsumierenden diese erhalten. Aufgrund der unterschiedlichen Strukturen sind die Advisories nur menschenlesbar. Eine automatisierte Verarbeitung der Informationen ist nur mit Vorarbeit möglich und fehleranfällig bei Veränderungen an Struktur oder Format. Zusammengefasst ist es schlichtweg nicht möglich, die Flut an Informationen zu beherrschen und wenn, dann nur mit enormem personellem Aufwand. IT-Sicherheitsfachkräfte, an denen es weltweit mangelt, werden durch die Suche und die Evaluation gebunden. Sie stehen damit nicht für die eigentliche Risikobewertung von Informationen über Schwachstellen, deren Behebung und andere Fachaufgaben zur Verfügung. Die Ressourcen, die Unternehmen also für die Suche, Sichtung und Bewertung von Security Advisories aufbringen müssen und zukünftig müssten, wachsen exorbitant.

2 Ein Standard bringt Ordnung ins System

Um dieser Vielzahl an neu entdeckten Schwachstellen, den zugehörigen Informationen, den Security-Advisory-Formaten, den unterschiedlichen Informationswegen und der Bewertung überhaupt gerecht werden zu können, braucht es dringend eine Lösung. Ein Teil dieser Lösung besteht darin, dass die Prozesse zur Kommunikation und zum Erhalt von Sicherheitsinformationen strukturiert erfolgen und automatisierbar werden. IT-Sicherheitsfachkräfte aus Wirtschaft, Verwaltung und Gesellschaft haben daher ihre Expertise auf internationaler Ebene vereint und gemeinsam das Common Security Advisory Framework (CSAF) 2.0 entwickelt. Am 18.11.2022 wurde CSAF als OASIS Open Standard veröffentlicht [5] und ist damit weltweit kostenfrei zugänglich. Das Framework spezifiziert einerseits das Format für maschinenverarbeitbare Sicherheitsinformationen und andererseits deren automatisierte Verteilung in Form von Bereitstellung, Auffindung und Abruf.

CSAF-Dokumente werden im JSON-Format erstellt und sind immer gleich aufgebaut. Sie enthalten Metadaten, einen Produktbaum und Informationen zu Schwachstellen, bezogen auf die einzelnen Produkte und sind bis auf den Versionslevel aufgeschlüsselt. Die Verwendung von Dateien im JSON-Format erlaubt die gewünschte Maschinenlesbarkeit und die daraus resultierende Skalierbarkeit für die Verarbeitung von CSAF-Dokumenten. Dadurch kann schnell und effizient bewertet, priorisiert und auf bekannt gewordene Gefährdungen reagiert werden. Es werden zudem nur solche Security Advisories, die eine Relevanz für die Konsumierenden haben, beispielsweise, weil das betroffene Produkt in einer Anlage eingesetzt wird, automatisiert abgerufen. Hierzu ist ein umfassendes Assetmanagement erforderlich, das alle Produkte und Komponenten, die ein Unternehmen einsetzt, auflistet.

Für die Bereitstellung, das Auffinden und den Abruf von CSAF-Dokumenten spezifiziert der CSAF-Standard 2.0 insgesamt drei Rollen. Den CSAF publis-

her, den CSAF provider und den CSAF trusted provider. Nur der CSAF provider und der CSAF trusted provider, der die Sicherheitsinformationen mit einer Signatur, einem OpenPGP-Schlüssel zur Überprüfung der Integrität und einem Hash-Wert versieht, erlauben ein automatisierbares Auffinden und den automatisierbaren Abruf von Sicherheitsinformationen, gemäß CSAF-Standard 2.0. Um den Bezug weiter zu vereinfachen, gibt es sogenannte CSAF Lister, bzw. Aggregatoren, die wie ein Telefonbuch agieren und die Quellen aufführen, die CSAF-Dokumente bereitstellen. Die Security Advisories von CSAF publishers können von einer Stelle gesammelt werden, um so ihren automatisierbaren Abruf für andere konsumierende Stellen zu gewährleisten. Das BSI betreibt einen solchen kostenfreien Aggregator [6]. Unternehmen, die Advisories im CSAF-Format bereitstellen, egal, ob sie bereits trusted provider sind oder nicht, können sich zum einen selbst beim BSI melden oder von Dritten vorgeschlagen werden. Das BSI führt dann eine Plausibilitätsprüfung der veröffentlichenden Stelle hinsichtlich der Bereitstellung von standardkonformen Sicherheitsinformationen durch. Durch die Nutzung des CSAF-Standards werden also sowohl die manuellen Aufwände als auch die Anzahl der relevanten Advisory-Informationen erheblich reduziert und wird ein automatisierter Abruf von Sicherheitsinformationen ermöglicht.

3 Das CSAFversum wurde geboren

Um die Vorteile der Automatisierung von strukturierten Informationen zu nutzen, muss der CSAF-Standard eine gewisse Verbreitung finden. Nur mit seiner Verbreitung kann überhaupt der Makrokosmos, das sogenannte CSAFversum, entstehen. Ein wichtiger Faktor für die Verbreitung ist zum einen die Kenntnis darüber, dass es den CSAF-Standard gibt und welche Vorteile er bringt. Zum anderen sind die Kosten und der Aufwand, die hinter der (Weiter-) Entwicklung des CSAFversum stecken und von Organisationen aufgewendet werden müssen, ein wichtiges Kriterium für die Nutzung und Verbreitung von CSAF-Dokumenten. Um die Akzeptanz zu erhöhen und Kosten einsparen zu können, werden BSI-seitig konsequent alle Entwicklungen und künftigen Entwicklungsvorhaben als kostenfreie Open Source-Software, mit entsprechenden Implementierungsanleitungen, zur Verfügung gestellt. Die einzelnen Werkzeuge können von Organisationen genutzt und weiterentwickelt werden. Ebenso ermöglichen Rückmeldungen und Verbesserungsvorschläge aus der Community die stetige Weiterentwicklung des CSAFversums. Das führt dazu, dass bereits heute Advisories, welche die Anforderungen dieses Standards erfüllen, von vielen, auch großen Unternehmen, angeboten werden und deren Anzahl kontinuierlich ansteigt. Um das CSAFversum weiter wachsen zu lassen, müssen künftig mehr und mehr Organisationen ihre

Advisories im CSAF-Format veröffentlichen und bereitstellen. Anreize für die Bereitstellung von Sicherheitsinformationen bieten insbesondere Betreiber, Integratoren, Behörden und Hersteller, die innerhalb von Lieferketten CSAF-Advisories einfordern und dies sogar, beispielsweise bei Ausschreibungen, vertraglich festlegen.

Bisher gibt es diverse Open Source-Tools, um die Nutzung des CSAF-Standards weiter zu fördern und Organisationen bei der Erstellung, Verteilung, Darstellung und Bewertung von Advisories zu unterstützen [7, 8, 9, 10, 11]. Weitere werden in den kommenden Monaten und Jahren folgen. Die grundlegende Infrastruktur des CSAFversums ist somit nach 17 ½ Monaten geschaffen, sodass das CSAFversum weiter expandieren kann. Wo die Reise im CSAFversum noch hingehen kann, wird sich oftmals aus aktuellen und kommenden Bedarfen ergeben, wie beispielsweise gesetzlichen Regulierungen. Diese Bedarfe können sehr agil abgebildet werden, denn sie ermöglichen die kontinuierliche Erweiterung, Verbesserung sowie die Individualisierung der bestehenden Werkzeuge. Die Zusammenarbeit durch wertvolle Community-Beiträge ist mit einem Teleskop vergleichbar, das neue und bisher unbekannte Bereiche des CSAFversums beleuchten und deren Erschließung ermöglichen wird.

4 Hinzu kam die Regulierung

Das Entstehen neuer Himmelskörper oder Tier- und Pflanzenarten wird durch besondere Rahmenbedingungen ermöglicht. Ebenso verhält es sich im Kontext der internationalen, paneuropäischen und nationalen Regulierung. Durch einen vorgegebenen Rechtsrahmen müssen Dinge oder Prozesse entsprechend verändert oder angepasst werden oder erhalten überhaupt erst Einzug in die Regulierung, wie im Fall der IT-Sicherheit. Durch aktuelle Gesetze auf nationaler und europäischer Ebene, wie das IT-Sicherheitsgesetz 2.0 [12] oder die NIS-2-Richtlinie [13], wird IT-Sicherheit inzwischen mitgedacht und mehr IT-Sicherheitsbewusstsein von Organisationen eingefordert, wie beispielsweise mittels eines Warn- und Meldewesens zur Kommunikation von IT-Sicherheitsproblemen. Die immer digitaler werdende und zunehmend komplexere Welt bietet, wie bereits zu Anfang erwähnt, neben den vielen Vorteilen auch Risiken. Schwachstellen in Soft- und Hardware sind omnipräsent und die Schäden, die durch Angriffe aus dem Cyberraum verursacht werden, sind enorm. Dennoch müssen sich heutzutage alle Organisationen auf ihre IT-Lösungen verlassen, um marktfähig zu sein und zu bleiben. Die Digitalisierung lässt sich schlichtweg nicht mehr aufhalten oder gar verhindern. Bedrohungen aus dem Cyberraum werden immer professioneller und

adaptieren sich an aktuelle geopolitische Lagen oder politische Entscheidungen. Sie machen selbst vor sensiblen Kritischen Infrastrukturen oder Einzelpersonen nicht Halt, so lange sie erfolgreich sind und wirtschaftlich lohnenswert, wie das Beispiel Ransomware zeigt [14].

Das Ziel resiliente Infrastrukturen und sichere Produkte einzusetzen rückt also mehr und mehr in den Fokus der nationalen, paneuropäischen und internationalen Gesetzgebung und Standardisierung. Durch künftige Regulierungsvorhaben auf europäischer Ebene, wie den Cyber Resilience Act (CRA) [15], werden weitere Vorgaben an die IT-Sicherheit und die Prozesse innerhalb und außerhalb von Organisationen gestellt. Der CRA ist eine Marktzugangsregelung für Produkte mit digitalen Elementen, die im Europäischen Wirtschaftsraum (EWR) vertrieben werden. Er wird in den nächsten Jahren in Kraft treten und fordert, dass es transparente Schwachstellenmanagementprozesse und eine Software Bill of Materials (SBOM) [16] seitens der Organisationen gibt. Des Weiteren ist im CRA festgeschrieben, dass es einen IT-Sicherheitskontakt in die entsprechende Organisation geben muss, was beispielsweise durch die Implementierung der security.txt (RFC 9116) [17, 18] realisiert werden kann. Der CRA wird Marktaufsichtsbehörden in den Mitgliedsstaaten ermächtigen, Produkte mit IT-Sicherheitsbedenken zu untersuchen und gegebenenfalls vom Markt nehmen zu dürfen.

Allen zuvor genannten Regulierungen gemein ist, dass es strukturierte (Schwachstellenmanagement-) Prozesse sowie eine Sammelstelle für Schwachstelleninformationen geben muss. Dies ist notwendig, um resilienter zu werden und sich beispielsweise vor Angriffen oder unsicheren Produkten zu schützen und entsprechende Schutzmaßnahmen länderübergreifend kommunizieren zu können. Das wird dazu führen, dass letztlich mehr über Schwachstellen gesprochen wird und als logische Konsequenz mehr Advisories publiziert werden müssen, um Organisationen und auch die Marktaufsicht zu informieren. Hier könnte sich die Nutzung des CSAF-Standards als unabdingbar erweisen, sodass das CSAFversum in unsere Welt, mit all ihren aktuellen und zukünftigen Prozessen und Regulierungen, vollständig integriert würde. Die Adaption von CSAF erleichtert die Erfüllung der regulatorischen Anforderungen. Hersteller erfüllen ihre Pflichten zur Bereitstellung von Advisories. Anwender erhalten eine Möglichkeit, mit geringem Aufwand über Schwachstellen in ihren Systemen informiert zu werden und können schneller handeln. Durch die maschinenverarbeitbaren Sicherheitsinformationen, den standardisierten Abruf und die Verteilung ermöglicht das Framework nicht nur die gewünschte Skalierbarkeit und Delegation von Aufgaben, sondern auch die Einhaltung der aktuellen und kommenden gesetzlichen Vorgaben (Compliance).

5 Das CSAFversum expandiert - ein Blick in die Sterne

Nach nur 17½ Monaten ist klar, dass das CSAFversum keine Einbildung ist, sondern tatsächlich existiert. Es ist nicht nur gewachsen, sondern expandiert stetig weiter– und das zu Recht. Behördliche Ausschreibungen, die Kenntnisse des CSAF-Standards erfordern, sind mittlerweile auf der Tagesordnung. Es wird gefordert, dass sämtliche Schwachstellen und ihre Behebungsmaßnahmen in Form von CSAF-Advisories kommuniziert werden. Das führte dazu, dass BSI-seitig eine Technische Richtlinie zu CSAF [19] erstellt wurde und auch BSI-seitig Security Advisories im CSAF-Format bereitgestellt werden. Auch die amerikanische Cybersecurity and Infrastructure Security Agency (CISA) publiziert mittlerweile standardkonforme CSAF-Advisories für den Bereich der OT (Operational Technology) [20].

Es ist ein Trend erkennbar, dass mehr und mehr Organisationen sich für den CSAF-Standard interessieren und um Unterstützung bei der Implementierung bitten, um künftig ihre Sicherheitsinformationen auch im CSAF-Format zur Verfügung stellen zu können. Die Notwendigkeit der Automatisierung, aber auch die aktuellen und künftigen Anforderungen durch neue Gesetzgebungen, sind den meisten Organisationen bewusst. Denn gerade bei kritischen Schwachstellen ist es enorm wichtig, zeitnah eine Aussage hinsichtlich der Betroffenheit oder eben Nicht-Betroffenheit zu treffen. Auch dies lässt sich mit dem CSAF-Standard, in Form eines VEX-Advisories (Vulnerability Exploitability eXchange) [21], abbilden, da VEX ein Profil in CSAF ist.

Auch die Anzahl an unterschiedlichen CSAF-Werkzeugen, die auf verschiedenen Github-Repositories [7, 8, 9, 10, 11] zur Verfügung stehen und durch die Community genutzt und stetig weiterentwickelt werden, belegen, die Expansion des CSAFversums. Ein weiterer Bedarf zeigt sich in der gestiegenen Nachfrage für praxisnahe Workshops, die sowohl das Veröffentlichen und Erstellen von CSAF-Advisories, als auch die Verteilung von Sicherheitsinformationen und deren Abruf trainieren. Durch das „train-the-Trainer-Prinzip“ werden CSAF-Kenntnisse in die Fläche gebracht Jede Person und Organisation kann also einen aktiven Beitrag leisten, egal ob durch die Veröffentlichung von Security Advisories, durch die Forderung nach solchen, durch das Erstellen neuer Werkzeuge oder durch die Weitergabe von Wissen. Durch diese Art der vertrauensvollen Zusammenarbeit erlangt das CSAF eine große Reichweite und das CSAFversums ist unterwegs in die nächste Dimension. Der CSAF-Standard 2.1 wird viele Anregungen zur Verbesserung aus der Community berücksichtigen und es ist geplant, dass er noch 2024 das Licht der Welt erblicken wird.

Um das CSAFversum zu sehen, braucht es 2024 weder eine Lupe noch ein Teleskop. Der CSAF-Standard ist weltweit angekommen und hat gute Chancen zu bleiben, um sich weiterzuentwickeln und nicht als Supernova oder aussterbende Art zu enden.

Literaturhinweise

[1] https://cve.org/ (abgerufen am 23.02.2024)

2] https://opensourcesecurity.io/2021/03/30/its-time-to-fix-cve/ (abgerufen am 23.02.2024)

[3] https://www.veracode.com/state-of-software-security-report (abgerufen am 23.02.2024)

[4] https://www.comparitech.com/blog/information-security/cybersecurity-vulnerability-statistics/ (abgerufen am 23.02.2024)

[5] https://oasis-open.github.io/csaf-documentation/ (abgerufen am 23.02.2024)

[6] https://wid.cert-bund.de/.well-known/csaf-aggregator/aggregator.json (abgerufen am 23.02.2024)

[7] https://github.com/csaf-poc/csaf_webview (abgerufen am 23.02.2024)

[8] https://github.com/csaf-poc/csaf_distribution (abgerufen am 23.02.2024)

[9] https://github.com/csaf-poc/csaf_backend (abgerufen am 23.02.2024)

[10] https://github.com/ISDuBA/ISDuBA/ (abgerufen am 23.02.2024)

[11] https://github.com/secvisogram (abgerufen am 23.02.2024)

[12] https://www.gesetze-im-internet.de/bsig_2009/ (abgerufen am 23.02.2024)

[13] https://eur-lex.europa.eu/eli/dir/2022/2555/oj (abgerufen am 23.02.2024)

[14] https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2023.html?nn=129410 (abgerufen am 23.02.2024)

[15] https://www.european-cyber-resilience-act.com/ (abgerufen am 23.02.2024)

[16] https://www.bsi.bund.de/dok/TR-03183 (abgerufen am 23.02.2024)

[17] CS-E 149: https://www.allianz-fuer-cybersicherheit.de/dok/1109660 (planned publication)

[18] https://www.rfc-editor.org/info/rfc9116 (abgerufen am 23.02.2024)

[19] TR-03191 Common Security Advisory Framework https://www.bsi.bund.de/dok/1109942 (planned publication)

[20] https://www.cisa.gov/news-events/news/transforming-vulnerability-management-cisa-adds-oasis-csaf-20-standard-ics-advisories (abgerufen am 23.02.2024)

[21] https://www.cisa.gov/sites/default/files/2023-04/minimum-requirements-for-vex-508c.pdf (abgerufen am 23.02.2024)

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Die Erforderlichkeit zur Prüfung von Dienstleistern aus Sicht der Informationssicherheit

Alexander Weidenhammer[1], Max Just[1]

Kurzfassung:
Zunehmend werden Tätigkeiten oder Geschäftsprozesse von Organisationen ganz oder teilweise an einen oder mehrere Dienstleister ausgelagert (Outsourcing). Oftmals werden hierbei die (branchenspezifischen) Anforderungen aus dem Bereich der Informationssicherheit, die an den jeweiligen Dienstleister zu stellen wären, nur unzureichend berücksichtigt. Für den Auftraggeber können hieraus unterschiedlichste Risiken resultieren, die im schlimmsten Fall zur Einschränkung der Gewährleistungsziele oder zum Erliegen ganzer Geschäftsprozesse führen. Insofern können beispielsweise auch bereits Normen des Datenschutz- sowie des Gesellschaftsrechts eine nähere Prüfung von Dienstleistern erforderlich machen. Aus den benannten Gründen ist es für Organisationen essenziell, die bestehenden Anforderungen zu kennen und vorab sowie sodann regelmäßig zu prüfen. Vor diesem Hintergrund wird es jedoch in aller Regel nicht als ausreichend zu erachten sein, lediglich vorliegende Testate und Zertifizierungen einzusehen. Ebenso ist die Ausgestaltung des zugrundeliegenden Vertrages und hierin die Verankerung entsprechender Rechte und Pflichten entscheidend. Für die Praxis ergeben sich zudem Schwierigkeiten, da sich die für die Umsetzung der Prüfpflichten konkret aufzuwendenden Ressourcen nicht grundsätzlich definieren lassen werden. Eine entscheidende Rolle spielen hierbei vielmehr die Größe der Organisation sowie deren Tätigkeitsfeld. Ein funktionierendes Auslagerungsmanagement kann zudem dabei helfen, die bestehenden Prüfpflichten praktikabel umzusetzen.

Stichworte: Auslagerungsmanagement, Dienstleister, Outsourcing, Prüfpflichten, Vertragsgestaltung

1 Bedeutung des Outsourcings für die Informationssicherheit

Die zunehmende Komplexität von Prozessen, Systemen und Anwendungen sowie die unter anderem hieraus resultierende Spezialisierung von Unternehmen machen immer öfter eine Einbeziehung Dritter in die eigenen Geschäftsprozesse erforderlich. Gleichzeitig sorgen steigende regulatorische Anforderungen für eine zunehmende Verrechtlichung der Informationssicherheit. So können je nach Art und Umfang des Dienstleistereinsatzes zahlreiche Rechtsgebiete betroffen sein. Hierunter ist allen voran das Datenschutzrecht zu fassen. Umfasst werden außerdem eine Vielzahl spezialgesetzlicher sowie sektoren- und branchenspezifischer Regelungen wie z. B. aus dem Finanzsektor oder dem Versicherungsbereich.

[1] Dresdner Institut für Datenschutz (DID), Dresden.

Insbesondere der europäische Gesetzgeber verfolgt hierbei zunehmend seine Regulierungsstrategie und erlässt auf dem Gebiet der Cybersicherheit zahlreiche Rechtsakte. Hierunter fallen u. a. die Richtlinie über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der gesamten Union (*NIS-2*-Richtlinie)[2] sowie der Cyber Resilience Act (CRA).[3] Die Auflistung wird in Zukunft fortzusetzen sein, ohne dass bislang ein Ende absehbar ist.

Darüber hinaus bedürfen die Vorgaben der Informationssicherheit – insbesondere die Aspekte aus dem technischen Bereich der IT-Sicherheitsstandards – einer stetigen Aufrechterhaltung, um die Effektivität der Maßnahmen sicherstellen zu können. Damit soll auf die fortschreitenden technischen Anforderungen durch die immer kürzeren Innovationszyklen sowie die stetig zunehmende und sich verändernde Bedrohungslage reagiert werden.

Probleme ergeben sich ergänzend aus dem Umstand, dass der Bereich des IT-Outsourcing – und hier insbesondere die Nutzung von Cloud-Diensten - zunehmend ein Massengeschäft darstellt, wobei die Umsetzung individueller vertraglicher oder sicherheitsrechtlicher (und regulatorisch bedingter) Anforderungen nicht oder nur selten möglich ist.[4] Organisationen sollten daher bei der Auslagerung von Prozessen, Systemen und Anwendungen grundsätzlich ein großes Interesse an der Erfüllung von Informationssicherheitsvorgaben durch den Outsourcing-Partner hegen, der mitunter Zugriff auf sensible Informationen erhält.[5] Dienstleister in diesem Bereich setzen hingegen oftmals in hohem Maße auf standardisierte Prozesse und Vertragsbedingungen, um ihre Leistungen besser und effektiver skalieren zu können.[6]

---

2 Richtlinie (EU) 2022/2555 des Europäischen Parlamentes und des Rates vom 14. Dezember 2022 über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der Union, zur Änderung der Verordnung (EU) Nr. 910/2014 und der Richtlinie (EU) 2018/1972 sowie zur Aufhebung der Richtlinie (EU) 2016/1148 (NIS-2-Richtlinie).

3 Vorschlag für eine Verordnung des Europäischen Parlamentes und des Rates über horizontale Cybersicherheitsanforderungen für Produkte mit digitalen Elementen und zur Änderung der Verordnung (EU) 2019/1020, https://data.consilium.europa.eu/doc/document/ST-12429-2022-INIT/de/pdf, zuletzt abgerufen am 23.2.2024.

4 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 141.

5 *Küchler*, Technische und wirtschaftliche Grundlagen, in Bräutigam: IT-Outsourcing und Cloud-Computing, Teil 1, Rn. 1.

6 *Thalhofer*, Vertragspraxis und IT-Sicherheit – aktuelle Herausforderungen und mögliche Lösungen, in Bundesamt für Sicherheit in der Informationstechnik: Tagungsband zum 16. IT-Sicherheitskongress 2019, S. 465 (469).

Auftraggeber sind regelmäßig aufgrund beschränkter Möglichkeiten zur Einflussnahme auf eine ausreichende technische und organisatorische Gewährleistung von Informationssicherheit durch den jeweiligen Dienstleister angewiesen. Die vertragliche Vereinbarung von Kontroll- und Steuerungsrechten ist aus diesem Grund für Auftraggeber zur Überwachung des Auftragnehmers essenziell. Konflikte können sich aber bereits auf der vertraglichen Gestaltungsebene ergeben, denn der Auftraggeber wird in der Regel weitreichende Rechte verankern lassen, der Auftragnehmer hingegen allen voran einen Zugang zu seinen betriebsinternen sensiblen Bereichen weitestgehend ausschließen wollen.[7] Die meisten Anbieter versuchen stattdessen, durch die Vorlage von Prüfungsergebnissen und Zertifikaten Dritter, die Einhaltung der Informationssicherheitsvorgaben nachzuweisen. Zur Wahrheit gehört jedoch ebenfalls, dass außerhalb regulierter Bereiche Überprüfungen von Auftragnehmern im Kontext der Organisation, des Informationsverbundes sowie der Kritikalität der betroffenen Informationen oftmals zu selten oder in einem zu geringen Umfang durchgeführt werden.

Zwar kann mittels Überprüfungen von Auftragnehmern kein Ausschluss sämtlicher Risiken, jedoch eine Reduktion auf einen für die Organisation beherrschbaren Umfang erreicht werden. Nachstehend sollen daher mögliche Prüfpflichten näher betrachtet werden.

2 Mögliche Prüfpflichten im Detail

Mit einem Blick auf die aktuelle Normenlandschaft wird deutlich, dass bereits zum gegenwärtigen Zeitpunkt auch außerhalb gängiger Standards der Informationssicherheit umfassende Prüfpflichten bestehen und Leitungspersonen die Wirksamkeit von Prüfprozessen zu gewährleisten haben. Allen voran ergibt sich dies beispielsweise bereits aus den nationalen Normen des Gesellschaftsrechts.

2.1 Normen des Gesellschaftsrechts

Die Verpflichtungen zur Umsetzung und Überwachung geeigneter Informationssicherheitsvorgaben ist in der jüngeren Vergangenheit stetig angewachsen. Bereits aus den gesellschaftsrechtlichen Normen lässt sich eine Überwa-

[7] Pour Rafsendjani/Bomhard, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 141.

chungspflicht der Leitungsebene herleiten. Allen voran ist hierbei insbesondere die Regelung des § 93 AktG[8] zu benennen.[9] Die Norm legt unter Berücksichtigung verschiedener Kriterien, z. B. die Art und Größe des Unternehmens, das wirtschaftliche Umfeld sowie die Art der Geschäftsführungsmaßnahmen, einen Sorgfaltsmaßstab zugrunde.[10] Zwar steht der Geschäftsleitung auch ein erheblicher Ermessenspielraum in Form der sogenannten Business Judgement Rule zu, jedoch sind unternehmerische Entscheidungen stets auf Grundlage angemessener Informationen zu treffen.[11] Insoweit kann argumentiert werden, dass die Auslagerung von (Teil-)Prozessen an Dienstleister eine vorherige und fortlaufende Prüfung voraussetzt. Nur so kann die Zuverlässigkeit des Dienstleisters bzw. das Bestehen etwaiger Risiken sowie die Erforderlichkeit zum Ergreifen risikominimierender Maßnahmen unter Berücksichtigung der jeweiligen Risikostrategie (Risikomanagement) wirksam festgestellt werden.[12]

Die vorbenannte Norm bezieht sich zwar zunächst ausschließlich auf Aktiengesellschaften, jedoch lässt sich über § 43 GmbHG[13] eine solche Verpflichtung ebenfalls für die Geschäftsführungen von Gesellschaften mit beschränkter Haftung sowie nach § 347 HGB[14] für Leitungspersonen von Handelsgesellschaften herleiten.[15] Aus den allgemeinen Verantwortlichkeits- und Haftungsnormen lassen sich auch für den Bereich der Informationssicherheit entsprechende Legalitäts- und Organisationspflichten ableiten. Flankierend ist weiterhin § 91 Abs. 2 AktG zu nennen, wonach beispielsweise Aufsichtsmaßnahmen ergriffen werden müssen. Insbesondere gilt es existenzgefährdende Risiken rechtzeitig zu erkennen und diese abzuwenden bzw. vorzubeugen. Hierbei sind selbstverständlich auch die wachsenden Risiken aufgrund der stetig ansteigenden Anzahl von (erfolgreichen) Cyber-Angriffen und der regelmäßig damit verbundene Verlust vertrauenswürdiger Informationen (z. B. in Form von Kundendatenbanken, Forschungsergebnissen oder

8 Aktiengesetz vom 6. September 1965 (BGBl. I S. 1089), das zuletzt durch Artikel 13 des Gesetzes vom 11. Dezember 2023 (BGBl. 2023 I Nr. 354) geändert worden ist.

9 *Schultz/Sarre*, CR 2022, 281 (282 f.).

10 *Voigt*, in: Voigt, IT-Sicherheitsrecht, IT-Sicherheit in der Unternehmensorganisation, Rn. 57 f.

11 *Voigt*, in: Voigt, IT-Sicherheitsrecht, IT-Sicherheit in der Unternehmensorganisation, Rn. 63.

12 *Schultz/Sarre*, CR 2022, 281 (282).

13 Gesetz betreffend die Gesellschaften mit beschränkter Haftung in der im Bundesgesetzblatt Teil III, Gliederungsnummer 4123-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 9 des Gesetzes vom 22. Februar 2023 (BGBl. 2023 I Nr. 51) geändert worden ist.

14 Handelsgesetzbuch in der im Bundesgesetzblatt Teil III, Gliederungsnummer 4100-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 34 Absatz 1 des Gesetzes vom 22. Dezember 2023 (BGBl. 2023 I Nr. 411) geändert worden ist.

15 *Schultz/Sarre*, CR 2022, 281 (281 f.).

Patenten) sowie die damit einhergehenden finanziellen Verluste[16] zu berücksichtigen. Vergleichbare Organisationspflichten finden sich zudem in Spezialgesetzen wie z.B. § 25a KWG[17] oder §§ 23 ff. VAG[18] wieder. Auch eine Prüfpflicht für Vereine kann unter analoger Anwendung des § 91 Abs. 2 AktG auf den Vereinsvorstand argumentiert werden.[19]

2.2 Anforderungen aus dem Informationssicherheitsrecht

Darüber hinaus ergeben sich weitere zahlreiche Verpflichtungen aus den Vorgaben des Rechtes der Informationssicherheit. So lassen sich speziell aus dem Datenschutzrecht unter Heranziehung von Art. 5 Abs. 2 und Art. 24 DSGVO entsprechende Nachweispflichten zur ordnungsgemäßen Datenverarbeitung herleiten. Im Rahmen einer sogenannten Auftragsverarbeitung kann sich eine Prüfpflicht zudem bereits aus Art. 28 Abs. 1 DSGVO (Geeignetheit des Auftragsverarbeiters) und einem bestehenden Haftungsrisiko des für die Datenverarbeitung Verantwortlichen für die Handlungen des Auftragsverarbeiters ergeben.[20] Ebenso kann Art. 32 DSGVO mit seiner Verpflichtung des Verantwortlichen und des Auftragsverarbeiters, für eine angemessene Sicherheit der Verarbeitung zu sorgen, angeführt werden. Besondere Aufmerksamkeit erlangte an dieser Stelle jüngst ein Urteil des Europäischen Gerichtshof in dem der EuGH konstatiert, dass grundsätzlich eine Haftung gemäß Art. 82 DSGVO aufgrund der Verletzung der Art. 24 und Art. 32 DSGVO denkbar ist.[21] Durch die Entscheidung wird u.a. deutlich, dass die für die Verarbeitung Verantwortliche durchaus weitreichende Nachweispflichten treffen können, dass die umgesetzten Schutzmaßnahmen auch tatsächlich geeignet waren, um einen ausreichenden Schutz sicherzustellen.

Für den Bereich der Kritischen Infrastrukturen gilt zudem das BSIG[22] und insbesondere § 8a BSIG mit der Verpflichtung, angemessene organisatori-

16 Der Branchenverband Bitkom e. V. beziffert den Schaden durch Angriffe auf deutsche Unternehmen im Jahr 2022 auf insgesamt rund 203 Milliarden Euro, https://www.bitkom.org/Presse/Presseinformation/Wirtschaftsschutz-2022, zuletzt abgerufen am 23.2.2024.

17 Kreditwesengesetz in der Fassung der Bekanntmachung vom 9. September 1998 (BGBl. I S. 2776), das zuletzt durch Artikel 6 des Gesetzes vom 22. Dezember 2023 (BGBl. 2023 I Nr. 411) geändert worden ist.

18 Versicherungsaufsichtsgesetz vom 1. April 2015 (BGBl. I S. 434), das zuletzt durch Artikel 14 des Gesetzes vom 22. Dezember 2023 (BGBl. 2023 I Nr. 411) geändert worden ist.

19 *Heermann*, NJW 2016, 1687 (1692).

20 EuGH, Urt. v. 5.12.2023, Rs. C-683/21.

21 EuGH, Urt. v. 14.12.2023, Rs. C-340/21, GRUR-RS 2023, 35786.

22 BSI-Gesetz vom 14. August 2009 (BGBl. I S. 2821), das zuletzt durch Artikel 12 des Gesetzes vom 23. Juni 2021(BGBl. I S. 1982) geändert worden ist.

sche und technische Vorkehrungen zur Vermeidung von Störungen der Verfügbarkeit, Integrität, Authentizität und Vertraulichkeit ihrer informationstechnischen Systeme, Komponenten oder Prozesse zu treffen. Vergleichbare Anforderungen gelten nach § 8c BSIG für die Anbieter digitaler Dienste. Beide Regelungen des BSIG gehen auf die NIS-Richtlinie[23] zurück.[24] Mit Spannung wird daher zu beobachten sein, welche Änderungen sich durch die Umsetzung der NIS-2-RL durch den nationalen Gesetzgeber ergeben werden. Insbesondere Art. 21 Abs. 2 NIS-2-RL enthält eine Vielzahl von Maßnahmen, welche u.a. die Sicherheit in der Lieferkette umfassen. Die Umsetzung der NIS-2-Richtlinie muss in den Mitgliedsstaaten bis zum 17. Oktober 2024 erfolgen. In Deutschland ist die Verabschiedung des NIS-2-Umsetzungs- und Cybersicherheitsstärkungsgesetz (NIS2UmsuCG) vorgesehen.[25]

Gleichfalls gibt es für den öffentlichen Bereich landespezifische Regelungen – welche ebenfalls durch die Umsetzung der NIS-2-RL in nationales Recht deutlich ausgeweitet werden sollen - wie den § 4 SächsISichG[26], welcher angemessene organisatorische und technische Vorkehrungen sowie sonstige Maßnahmen zur Gewährleistung der Informationssicherheit fordert. Bemerkenswert in Bezug auf die spezielle Vorschrift ist, dass es eine konkrete Inbezugnahme zum BSI-IT-Grundschutz gibt, da die nach dem SächsISichG verpflichteten Stellen zur Erreichung und Aufrechterhaltung eines angemessenen Informationssicherheitsniveaus die jeweils geltenden Standards und das jeweils geltende IT-Grundschutz-Kompendium des BSI zu berücksichtigen haben bzw. dies zur Anwendung empfohlen wird.

Indes kann in Anbetracht der zahlreichen Normen zum Recht der Informationssicherheit die Auflistung auch an dieser Stelle beliebig fortgeschrieben werden. Auch wenn aus den meisten der vorbenannten Normen keine Prüfpflicht aus dem direkten Wortlaut hervorgeht, ist eine Prüfung von Dienstleistern zur vollumfänglichen Erfüllung der rechtlichen Anforderungen wohl unumgänglich.

---

23 Richtlinie (EU) 2016/1148 des Europäischen Parlaments und des Rates vom 6. Juli 2016 über Maßnahmen zur Gewährleistung eines hohen gemeinsamen Sicherheitsniveaus von Netz- und Informationssystemen in der Union.

24 *Ritter*, in: Kipker/Reusch/Ritter, Recht der Informationssicherheit, § 8a BSIG, Rn. 1.; *Hessel/Potel*, in Kipker/Reusch/Ritter, Recht der Informationssicherheit, § 8c BSIG, Rn. 1.

25 Der Referentenentwurf liegt seit dem 3. Juli 2023 vor, abrufbar unter: www.ag.kritis.info/2023/07/19/referentenentwurf-des-bmi-nis-2-umsetzungs-und-cybersicherheitsstaerkungsgesetz-nis2umsucg/, zuletzt abgerufen am 27.2.2024.

26 Gesetz zur Gewährleistung der Informationssicherheit im Freistaat Sachsen (Sächsisches Informationssicherheitsgesetz – SächsISichG) erlassen als Artikel 1 des Gesetzes zur Neuordnung der Informationssicherheit im Freistaat Sachsen vom 2. August 2019.

2.3 Ableitung aus dem Informationssicherheitsmanagementsystem

Ergänzend ist anzuführen, dass bereits aus der regelmäßigen Funktionsweise eines Managementsystems sich zwangsläufig das Bestehen regelmäßiger Prüfpflichten ergibt.[27] Dies lässt sich insbesondere bereits aus dem PDCA-Zyklus, welches jedem Informationssicherheits-Managementsystem (ISMS) zugrunde liegt, ableiten.[28] Die Organisation hat demnach nicht nur ein bestimmtes Vorgehen zu planen und vorgabenkonform danach zu handeln, sondern die hieraus resultierenden Ergebnisse stets zu prüfen und die betreffenden Abläufe gegebenenfalls anzupassen. Freilich gilt dies auch dann, wenn einzelne (Teil-)Prozesse oder gar Geschäftsbereiche an Dienstleister ausgegliedert werden und das Ziel in einer möglichst umfassenden Gewährleistung von Informationssicherheit liegt. Von diesem Ziel ist grundsätzlich auszugehen, da die Etablierung eines ISMS in der Regel aus gesetzlichen Anforderungen, aufgrund von Branchenstandards oder zur transparenten Darstellung eines Qualitätsmerkmals einschließlich hieraus potenziell resultierender Wettbewerbsvorteile erfolgt.[29]

Weiterhin spricht auch die mögliche Reduzierung etwaiger Haftungsrisiken für eine ganzheitliche Betrachtung und die hieraus resultierenden weitreichenden Prüfungspflichten gegenüber Dienstleistern.[30]

Selbstverständlich können sich hierbei zum Teil deutliche Unterschiede in der konkreten Umsetzung etwaiger Prüfpflichten ergeben. Grundsätzlich sind dabei die spezifischen Anforderungen und die zur Verfügung stehenden Ressourcen einer Organisation angemessen zu berücksichtigen. Es gibt demnach keine einheitlichen Empfehlungen, wie und mit welcher Intensität es den Prüfpflichten zu begegnen gilt, solange die Wirksamkeit und Funktionalität des ISMS nachgewiesen werden kann. Dass hierbei unterschiedliche Abstufungen zwischen kleinen und großen Organisationen bestehen (dürfen), liegt in der Natur der Sache.

3 Maßnahmen nach BSI IT-Grundschutz

Mit Blick auf die Praxis stellt sich anschließend die Frage der Umsetzung. Die Betrachtung von Sicherheitsaspekten in einem Outsourcing-Vorhaben ist,

---

27 *Bundesamt für Sicherheit in der Informationstechnik*, BSI-Standard-200-1, Kap. 3.2.1 Der Lebenszyklus in der Informationssicherheit.

28 *Bundesamt für Sicherheit in der Informationstechnik*, BSI-Standard 200-1, Kap. 3.2.2 Beschreibung des Prozesses Informationssicherheit; *Voigt*, in: Voigt, IT-Sicherheitsrecht, Einrichtung eines Informationssicherheitsmanagementsystems, Rn. 650 ff.

29 *Voigt*, in: Voigt, IT-Sicherheitsrecht, Einrichtung eines Informationssicherheitsmanagementsystems, Rn. 657.

30 Ebenda.

wie im BSI-Grundschutz-Kompendium im Baustein *OPS.2.3 Nutzung von Outsourcing* explizit vorgeschlagen, typischerweise direkt im Outsourcing-Vertrag zu regeln. Teilweise bestehen sogar gesetzliche Verpflichtungen zur vertraglichen Fixierung diverser Aspekte, wie mit Blick auf Art. 28 DSGVO deutlich wird. Es empfiehlt sich zudem, die Aspekte der Informationssicherheit bereits ab der Entscheidung für das jeweilige Outsourcing-Projekt einzubeziehen. Ziel sollte der Aufbau einer möglichst umfassenden Schutzsphäre sein. [31]

3.1 Festlegung von Anforderungen an Dienstleister und Verträge

Vorab kann trefflich die Frage aufgeworfen werden, ob es angesichts der Vielzahl an gesetzlichen Regularien und Anforderungen durch Standards sowie die technische Normung überhaupt vertraglicher Regelungen zur Informationssicherheit bedarf.[32] Dies dürfte abseits der ausdrücklichen Verpflichtungen, beispielsweise im Rahmen einer Auftragsverarbeitung gemäß Art. 28 DSGVO, dennoch zu bejahen sein, da es sich bei den meisten Normierungen zum einen um generalklauselartige Verpflichtungen handelt, die weder zur Einhaltung bestimmter Standards noch zur Umsetzung konkreter Maßnahmen verpflichten.[33] Zum anderen erlangen vertragliche Regelungen im Bereich der Haftung aufgrund der Nichteinhaltung bestimmter Informationssicherheitspflichten insbesondere vor dem Hintergrund der oben erwähnten Entwicklung der Rechtsprechung des Europäischen Gerichtshof im Bereich Schadensersatzhaftung zunehmend an Bedeutung.[34] Für die Praxis empfiehlt sich hieraus und unter Berücksichtigung des Bausteins *OPS.2.3 Nutzung von Outsourcing* nunmehr allem voran die Berücksichtigung folgender Aspekte:

Zunächst bietet sich eine vertragliche Spiegelung der verschiedenen regulatorischen Anforderungen zur Informationssicherheit an. Hierbei erlegt der Auftraggeber den jeweiligen Vertragspartnern die jeweils einschlägigen Vorgaben zur Informationssicherheit auf. Zu beachten ist in diesem Zusammenhang jedoch, dass selbst im Fall derartiger Auslagerung von Aufgaben und Pflichten an Dritte dies nicht gleichbedeutend mit einer Entbindung der Informationssicherheitspflicht, mithin von der generellen Verantwortlichkeit,

31 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 118.

32 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 52.

33 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 52 ff.

34 Exemplarisch zu nennen sind hier die Urteile in den Rechtssachen vom 14. Dezember 2023 (C-340/21) und vom 25. Januar 2024 (C-687/21).

seitens der auslagernden Stelle ist.[35] Zwar erfolgt die Umsetzung der jeweiligen Vorgaben durch den Vertragspartner, aber die Verantwortung gegenüber Dritten wie z.B. Kunden, Geschäftspartnern, den eigene Beschäftigten verbleibt beim Auftraggeber. Aus diesem Grund erfolgt vielfach die vertragliche Vereinbarung zur ordnungsgemäßen Leistungserbringung sowie zur Überwachung derselben.[36]

Aus den vorstehenden Überlegungen geht hervor, dass bei vertraglichen Absicherungen zunächst die Vereinbarung der konkret zu erbringenden Leistungen das Herzstück bildet.[37] Aufgrund der Inbezugnahme zahlreicher gesetzlicher Anforderungen, technische und organisatorische Maßnahmen (TOM) umzusetzen, besteht regelmäßig ein Interesse seitens der gesetzlichen Adressaten, ebenso die Vertragspartner zu konkreten technischen und organisatorischen Maßnahmen zu verpflichten.[38] An dieser Stelle prallen häufig die Interessen der Auftragnehmer an einer größtmöglichen Flexibilität in Bezug auf die Umsetzung der TOM – z.B. durch einseitige Änderungsrechte – und die Interessen der Auftraggeber an der Vereinbarung weiterreichender Kontroll- und Nachweisrechten aufeinander.[39] In der Vertragsgestaltung werden derartige Konflikte häufig durch die Vereinbarung von flexiblen Änderungsrechten seitens der Auftragnehmer unter Einhaltung bestimmter Mindeststandards, entsprechender Dokumentationspflichten und Einräumung von Sonderkündigungsrechten seitens der Auftraggeber bei wesentlichen Änderungen gelöst.[40] Im Rahmen der konkreten Umsetzung erfolgt nicht selten zunächst im Hauptvertragswerk eine abstrakte Verpflichtung der Auftragnehmer zur Umsetzung bestimmter TOM in Übereinstimmung mit dem aktuellen Stand der Technik– so beispielsweise bekannt aus der Gestaltung von Auftragsverarbeitungsverträgen nach Art. 28 DSGVO - und hieran anknüpfend eine konkrete Festlegung von Einzelmaßnahmen,

35 *Voigt*, in: Voigt, IT-Sicherheitsrecht, IT-Sicherheit als vertragliche Pflicht, Rn. 101.

36 Ebenda.

37 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 145.

38 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 148.

39 *Thalhofer*, Vertragspraxis und IT-Sicherheit – aktuelle Herausforderungen und mögliche Lösungen, in Bundesamt für Sicherheit in der Informationstechnik: Tagungsband zum 16. IT-Sicherheitskongress 2019, S. 465 (469).

40 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 148.

meist in Form einer gesonderten Anlage zum Hauptvertrag.[41] Ein in der Praxis nicht selten anzutreffendes Problem besteht darin, dass zahlreiche Auftragnehmer mit generischen TOM-Aufstellungen arbeiten und sich somit zuweilen nur auf sehr abstrakte Weise zu konkreten Maßnahmen verpflichten. Zudem bergen derartige Auflistungen wiederrum die Gefahr, dass die Sicherheit der im Einzelfall zu erbringenden Leistungen nicht oder nicht vollständig abgebildet wird. Hier sollten Auftraggeber besondere Sorgfalt walten lassen. Ein "prüfender Blick", ob die zu erbringende Leistung und die dargestellten Maßnahmen übereinstimmen, ist stets zu empfehlen.

Weiterhin werden die übertragenen Anforderungen durch entsprechende vertragliche Nachweisregelungen abgesichert. Sinn und Zweck liegt darin, dass der Auftraggeber zunächst einmal in die Lage versetzt wird, seinen gesetzlichen Nachweispflichten – wie z.B. aus Art. 5 Abs. 2 DSGVO – nachkommen zu können. Der Auftraggeber muss bestrebt sein, die einzuhaltenden Informationssicherheitsanforderungen detailliert darzulegen und für den Nachweis Konzepte sowie unabhängige Zertifikate und Prüfberichte seitens des Auftragnehmers einzufordern. Schwierigkeiten bei der Überprüfung ergeben sich allerdings regelmäßig dann, wenn die Wirksamkeit sowie die Aussagekraft hinter derartigen Nachweisen zu hinterfragen ist. Abzugrenzen sind an dieser Stelle zunächst Normen und Standards, die ein Informationssicherheitsmanagementsystem innerhalb einer Organisation beschreiben, wie z.B. die ISO 27000-Reihe oder der BSI-Grundschutz von Zertifizierungen zur IT-Sicherheit, bei denen z.B. ein Produkt oder ein IT-Dienst durch eine zuständige Stelle geprüft und zertifiziert wird.[42] Nur bei der zweiten Gruppe an Zertifizierungen, Normen und Standards werden tatsächlich technische Anforderungen und mithin der Stand der Technik beschrieben.[43]

Dies soll keinesfalls die Bedeutung oder Wertigkeit der Zertifizierungen eines ISMS schmälern. Es ist jedoch hervorzuheben, dass die Unterscheidung in der Praxis nicht immer trennscharf durchgeführt wird und somit beispielsweise eine fehlerhafte Bewertungsgrundlage für die Zuverlässigkeit eines Auftragnehmers geschaffen werden kann. Bestandteil vertraglicher Regelungen können ferner die Festlegung von Zugangsrechten sowie die Durchführung von Audits und anderen Kontrollmaßnahmen sein. Hinsicht-

41 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 150 ff.

42 *Deusch/Eggendorfer,* Messbarkeit von IT-Sicherheit, in: Bernzen/Fritzsche/Heinze/Thomsen: Das IT-Recht vor der (europäischen) Zeitenwende, S. 323 (325 f.).

43 *Deusch/Eggendorfer*, Messbarkeit von IT-Sicherheit, in: Bernzen/Fritzsche/Heinze/Thomsen: Das IT-Recht vor der (europäischen) Zeitenwende, S. 323 (325).

lich des konkreten Umfangs der vertraglich zu vereinbarenden Pflichten können insbesondere Orientierungshilfen und Positionspapiere der deutschen und europäischen Aufsichtsbehörden, Best Practices von Interessenverbänden sowie einschlägige technische Normen, Standards und Regelwerke als Maßstab herangezogen werden.[44] Ferner kann die Implementierung von Berichts- und Informationspflichten vereinbart werden.

Gleichwohl können sich die Überprüfungspflichten seitens des Auftraggebers im Bereich der Haftungsfrage ergeben. Zunächst ist der relevante Anknüpfungspunkt unstreitig, welche Leistung und mithin auch welche Informationssicherheitspflichten geschuldet sind. Möglicherweise fällt so in diesem Bereich der Nachweis einer Pflichtverletzung leichter, wenn die vertraglich in Bezug genommenen Vorgaben nicht oder nicht vollständig umgesetzt wurden. Jedoch kann sich im Rahmen des Mitverschuldens nach § 254 BGB das Vernachlässigen bzw. die Nichteinhaltung der Überprüfungspflichten niederschlagen.

3.2 Etablierung eines Auslagerungsmanagements

Der im BSI IT-Grundschutz enthaltene Baustein *OPS.2.3 Nutzung von Outsourcing* enthält unter dem Begriff des Auslagerungsmanagements bereits eine Reihe von Maßnahmen, mit denen die zuvor genannten Problematiken adressiert sowie bestehende Prüfpflichten strukturiert und dokumentiert umgesetzt werden können. Insbesondere werden auf verschiedensten Ebenen zahlreiche Möglichkeiten zur Prüfung in Form eines Soll-Ist-Abgleichs aufgezeigt:

Die Auslagerung von (Teil-)Prozessen und Geschäftsbereichen bedarf zunächst stets einer kontextbasierten Betrachtung potenzieller Risiken. Zu berücksichtigen sind hierbei insbesondere die Kritikalität und die Abhängigkeit der betroffenen Prozesse sowie die Art und die Kategorien der in diesem Zusammenhang betroffenen Informationen.[45] So muss zunächst geklärt werden, ob überhaupt eine Auslagerung an einen Dienstleister erfolgen kann oder ob die Gefährdungslage eine ausschließlich interne Umsetzung gebietet.[46] Dies stellt grundsätzlich zwar keine Prüfung eines konkreten Auftragnehmers, in jedem Falle jedoch eine essenzielle Vorprüfung dar.

---

44 *Pour Rafsendjani/Bomhard*, IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung/Schallbruch, IT-Sicherheitsrecht, § 9, Rn. 155 f.

45 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A1.

46 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A2.

Ist der Einsatz eines Auftragnehmers grundsätzlich möglich, sind weiterhin konkrete Anforderungen an diesen zu definieren. Hierzu muss bereits im Vorfeld festgelegt werden, welche Kompetenzen für die Erbringung der Leistung aus Sicht der Informationssicherheit als erforderlich angesehen werden und welchen Grad an Vertrauenswürdigkeit sowie Zuverlässigkeit der Auftragnehmer leisten muss.[47] Möglicherweise sind derartige Aspekte im öffentlichen Sektor bereits in einem Vergabeverfahren zu berücksichtigen. Gegen diese definierten Kriterien ist ein Dienstleister möglichst noch vor Aufnahme vertraglicher Gespräche zu prüfen. In diesem Zusammenhang sollten ebenfalls gegebenenfalls bestehende Interessenkonflikte ausgeschlossen werden.[48]

Im nächsten Schritt sind konkrete Anforderungen an die vertraglichen Regelungen mit dem jeweiligen Dienstleister unter Berücksichtigung der oben aufgeführten Aspekte festzulegen und zu prüfen. So müssen Verträge zumindest ein Recht auf Überprüfung, einen Zustimmungsvorbehalt zur weiteren Verlagerung der Tätigkeit auf Unterauftragnehmer sowie weitere wesentliche Aspekte der Informationssicherheit wie beispielsweise eine Verpflichtung auf Vertraulichkeit und die Gewährleistung angemessener Sicherheitsmaßnahmen nach IT-Grundschutz oder vergleichbare Maßnahmen umfassen.[49] Auch die Bereitstellung eines auf den jeweiligen Prozess angepassten Sicherheitskonzeptes durch den Dienstleister ist zu vereinbaren.[50] Sinnvoll ist in diesem Zusammenhang ebenfalls die Etablierung eines Vertragsmusters, welches die seitens der Organisation als essenziell angesehenen Anforderungen dienstleisterunabhängig auch für zukünftige bzw. weitere Vertragsverhältnisse einheitlich darstellt.[51]

Ausgehend von diesen Mindestanforderungen können sich auch weitergehende Maßnahmen als sinnvoll erweisen. Hierzu können beispielsweise die Erstellung einer organisationsweiten Strategie sowie die Verabschiedung

47 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A3.

48 Ebenda.

49 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A4.

50 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A6.

51 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A4.

von Richtlinien zu den Voraussetzungen für die Auslagerungen von Tätigkeiten gehören.[52] Je umfangreicher die Anforderungen einer Organisation zur Inanspruchnahme derartiger Dienstleister gestaltet werden, umso eher empfiehlt sich die Etablierung eines umfassenden Auslagerungsmanagements, einschließlich der Benennung einer zuständigen Person, welche beispielsweise auch die zuvor genannten Prüfungen im Rahmen vorvertraglicher Maßnahmen durchführen kann.[53]

3.3 Durchführung fortlaufender Prüfungen

Jedoch erschöpft sich die Pflicht zur Prüfung von Dienstleistern nicht in einer einmaligen Vorab-Prüfung. Schließlich vermag eine solche ausschließlich eine Momentaufnahme abzubilden, wobei nicht sichergestellt werden kann, dass der Dienstleister fortlaufend ein ISMS oder entsprechende Maßnahmen aufrechterhält und weiterentwickelt. Insofern ist auch die ausschließliche Vorlage einer Zertifizierung regelmäßig als unzureichend zu erachten.

Dementsprechend obliegt der Organisation die Pflicht, die Einhaltung der auferlegten Kriterien fortlaufend zu prüfen. Eine solche Prüfung ist einerseits anlassbezogen, also beispielsweise bei Vorliegen von rechtlichen Änderungen oder bei Eintritt von Sicherheitsvorfällen bzw. -ereignissen, andererseits regelmäßig - anlassunabhängig - durchzuführen.[54] Hinsichtlich des Begriffs der Regelmäßigkeit wird keine allgemeingültige Definition möglich sein. Auch hierbei ist Bezug auf die Kritikalität der jeweiligen Prozesse und der hieraus resultierenden Gefährdungslage zu nehmen. Demnach kann es durchaus sinnvoll sein, innerhalb einer Organisation unterschiedliche Prüfungsintervalle zu etablieren. Auch hierbei kann je nach Komplexität der Anforderungen die Erstellung einer entsprechenden Richtlinie oder zumindest die Festlegung entsprechender Kennzahlen Erleichterungen in der praktischen Umsetzung ermöglichen.[55] Gegenstand der fortlaufenden Prüfungen sollten stets die vertraglich festgelegten Sicherheitsanforderungen, ergänzt

52 *Bundesamt für Sicherheit in der Informationstechnik*, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A8, OPS.2.3.A9.

53 Bundesamt für Sicherheit in der Informationstechnik, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A10.

54 Bundesamt für Sicherheit in der Informationstechnik, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A16.

55 Bundesamt für Sicherheit in der Informationstechnik, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A14.

um zwischenzeitlich gegebenenfalls hinzugetretene gesetzliche Anforderungen sowie Inhalte des vorliegenden prozessspezifischen Sicherheitskonzeptes sein.[56]

Sämtliche der aufgeführten Prüfungen sollten in einer dokumentierten und nachvollziehbaren Form erfolgen. Derartige Prüfberichte ermöglichen sodann unter anderem auch der Leitungsebene die Nachweisbarkeit eingerichteter Prüf- und Überwachungsprozesse zur Abwendung existenzgefährdender Risiken. Auch der zugrundeliegende PDCA-Zyklus kann durch die Prüfergebnisse genährt und die Gewährleistung der Informationssicherheit somit insgesamt (besser) sichergestellt werden.

4 Fazit

Für den Aufbau einer ganzheitlichen Informationssicherheit ist es für Organisationen unumgänglich, sich mit der (Über-)Prüfung der eingesetzten Auftragnehmer auseinanderzusetzen. Im Detail sind die bestehenden Prüfpflichten und deren Zersplitterung über unterschiedliche Rechtsbereiche ein Paradebeispiel für die Vielschichtigkeit des Rechts der Informationssicherheit. Der Pflichtenkanon reicht von den nationalen Vorschriften des Gesellschaftsrechts über das Datenschutzrecht bis hin zu einer eindeutigen Erforderlichkeit aus dem Funktionieren des Informationssicherheitsmanagementsystems selbst. Ein Vernachlässigen dieser Anforderungen kann weitreichende Folgen haben. Dies gilt nicht bereits wegen des steigenden Haftungsrisikos aufgrund konkretisierender höchstrichterlicher Rechtsprechung. Bereits die Risiken, welche sich grundsätzlich beim Outsourcing ergeben, gepaart mit einem unzureichenden Auslagerungsmanagement, sind Grund genug. Organisationen sind deshalb gut beraten ihren Prüfpflichten nachzukommen.

Mit Blick auf den Baustein *OPS.2.3 Nutzung von Outsourcing* des BSI IT-Grundschutzes sind Organisationen gut beraten, ein entsprechendes Auslagerungsmanagement zu betreiben. Vielfach bildet die vertragliche Gestaltung die Basis für den Start einer Outsourcing-Beziehung. Organisationen sehen sich hierbei nicht selten vorgegebenen Vertragsdokumenten der Auftragnehmer ausgesetzt. Ein schlechter Ratgeber ist das einfache Durchsehen und das „Abnicken“ vorgelegter TOM-Auflistungen und/oder bestimmter Zertifizierungen, ohne deren konkrete Bedeutung und Auswirkung für den

[56] Bundesamt für Sicherheit in der Informationstechnik, IT-Grundschutz-Bausteine, Edition 2023, OPS.2.3.A18.

zu betrachtenden Einzelfall zu würdigen. Das Auslagerungsmanagement erschöpft sich jedoch keinesfalls in der Beauftragung, sondern erfordert eine laufende Überprüfung der eingesetzten Auftragnehmer.

Literaturhinweise

[1] Deusch, Florian/Eggendorfer, Tobias, Messbarkeit von IT-Sicherheit, in: Bernzen, Anna K./Fritzsche, Jörg/Heinze, Christian/Thomsen, Oliver (Hrsg.): Das IT-Recht vor der (europäischen) Zeitenwende, S. 323 (325 f.)., Edewecht 2023, S. 323 – 337.

[2] Heermann, Peter W., Haftung des Vereinsvorstandes bei Ressortaufteilung sowie für unternehmerische Entscheidungen, in: Neue Juristische Wochenschrift (NJW), 2016, S. 1687 – 1692.

[3] Kipker, Dennis-Kenji/Reusch, Philipp/Ritter, Steve (Hrsg.): Kommentar Recht der Informationssicherheit, München 2023.

[4] Küchler, Peter, Technische und wirtschaftliche Grundlagen, in: Bräutigam, Peter (Hrsg.): IT-Outsourcing und Cloud-Computing – Eine Darstellung aus rechtlicher, technischer, wirtschaftlicher und vertraglicher Sicht, 4. Auflage, Berlin 2019, Teil 1.

[5] Pour Rafsendjani, Mansur/ Bomhard, David: IT-Sicherheit im Zivilrecht und in der Vertragsgestaltung, in: Hornung, Gerrit/Schallbruch, Martin (Hrsg.): IT-Sicherheitsrecht, Praxishandbuch, Baden-Baden 2021, § 9.

[6] Schultz, Marion/Sarre, Frank, Nutzung von Cloud Services im Unternehmen – Verantwortlichkeiten für die IT-Sicherheit, in: Computer und Recht (CR), 2022, S. 281 – 291.

[7] Thalhofer, Thomas: Vertragspraxis und IT-Sicherheit – aktuelle Herausforderungen und mögliche Lösungen, in: Bundesamt für Sicherheit in der Informationstechnik – BSI (Hrsg.): IT-Sicherheit als Voraussetzung für eine erfolgreiche Digitalisierung, Tagungsband zum 16. Deutschen IT-Sicherheitskongress, Gau-Algesheim 2019, S. 465 – 474.

[8] Voigt, Paul, IT-Sicherheitsrecht, 2. Auflage, Köln 2022.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Exploring EAP-TLS as authentication mechanism for private 5G networks

Julius Röttger[1]

Abstract:
The fifth generation of wireless technology, referred to as 5G, introduces enhancements including high data rates, ultra-low latency, massive device connectivity, and network slicing for various use cases. In addition to mobile operator networks, 5G is also expected to be used for private networks with applications in industrial IoT, enterprise networking and critical communications. A private 5G network is a dedicated mobile network that is managed by an organization for its exclusive use and offers more control and customization options compared to public networks. Although the 5G standard defaults to UICC-based authentication, UICCs are not universally available across all device classes, e.g. for devices used in industrial IoT, creating a need for alternative authentication methods. As there is no open-source 5G modem firmware, implementing experimental 5G private networks to explore alternative authentication mechanisms is currently a challenge.
In this paper, we introduce a first prototype of an EAP-TLS implementation in a 5G core network over a non-3GPP access, which enables to explore EAP-TLS authentication in 5G private networks. Furthermore, we present our research results regarding vulnerabilities of EAP-TLS to e.g. replay attacks and entity-in-the-middle attacks as well as other relevant performance indicators. Our results demonstrate the potential of EAP-TLS in a private 5G network and provide a framework for further research and development.

Keywords: 5G, Authentication, EAP-TLS, Private Network, UICC

1 Introduction

The fifth generation of wireless technology, referred to as 5G, introduces enhancements including high data rates, ultra-low latency, massive device connectivity, and network slicing for various use cases. The 5G system consists of the three main components User Equipment (UE), Radio Access Network (RAN) and the 5G core. Besides radio-related features such as enhanced broadband or low latency communication, 5G introduces a new type of core network design [1]. Based on virtual network functions, this new design offers high scalability and enables the development of large-scale systems based on commercial off-the-shelf (COTS) hardware. Combined with the efforts of regulatory authorities to deploy the latest generation of mobile networks beyond mobile operators. This offers promising opportunities for industrial use cases in which, e.g., reliable wireless communication based on a private 5G network is a necessary technical precondition for autonoumous

[1] Institute for Cyber Security & Privacy, Bonn-Rhein-Sieg University of Applied Sciences

production scenarios. In such use cases, the wireless devices may be fundamentally different from the user handsets in more common public 5G networks. The 5G standard already considers the use of the core network independently of the radio access network. However, current 5G implementations are mostly focused on use in conjunction with 5G-capable clients. This also applies to the authentication of subscribers in the core network, which is carried out via the Universal Subscriber Identity Module (USIM) application on the Universal Integrated Circuit Card (UICC), which is currently the de facto standard in mobile communications. The 5G security architecture [2], which describes the security architecture and procedures for 5G systems in detail, requires the USIM application on a UICC for the authentication mechanisms EAP-AKA' [2] and 5G-AKA [2]. However, it can no longer simply be assumed – especially in the context of private networks – that every mobile endpoint has an UICC, making alternative mechanisms for authenticating communication endpoints necessary. For EAP-TLS the 5G standard does not require an UICC. EAP-TLS is a certificate-based authentication mechanism based on the EAP framework [14], enabling mutual authentication via the TLS handshake [15].

In an initial attempt to implement EAP-TLS, we aimed to extend a 5G modem enabling authentication through 3GPP access. After an intense market analysis, in which we could find neither a programmable 5G modem nor an open source 5G modem firmware, we had to dismiss this approach. Given these current limitations, we then pursued EAP-TLS authentication through a non-3GPP access method [3]. As especially devices, which do not support 3GPP access, e.g. industrial IoT (IIoT) devices, do not have an UICC, require an alternative authentication mechanism, authentication via a non-3GPP access emerges as a promising option. Of course, IoT devices can be equipped with built-in UICCs (eSims) which can be provisioned over the network with appropriate software. But this would limit the number of devices to only those with eSim, making it more expensive and limiting the choice. The possibility to use other authentication mechanisms from the EAP framework are only foreseen for private 5G networks as the 5G security specification states. The scope of this paper is therefore limited to private networks. Without a non-3GPP access, devices which don’t support 5G can’t be part of the 5G network. If there is already a Wifi network in place for IIoT devices, those devices could not be used. With the non-3GPP access you can integrate devices, that are already in use, or devices only capable of Wifi. Making the migration from an existing Wifi infrastructure to a private 5G network more appealing.

In this paper, we introduce a first prototype implementation of EAP-TLS in a 5G core network and evaluate it over a non-3GPP access to gain first insights about EAP-TLS as an authentication mechanism in a 5G system.

2 Goals and Research Questions

The goal is to design and implement an EAP-TLS authentication component for use in a 5G system and explore its properties within an experimental setup with an untrusted non-3GPP access point. The latter involves vulnerability analysis, targeted attacks, and performance measurement. To facilitate this evaluation, EAP-TLS has been added to the open-source 5G core Open5GS [5], which currently lacks this feature. The prototype is evaluated along the following research questions in order to gain initial insights into the practical use of EAP-TLS in 5G systems:

- RQ1a: What are the characteristics of EAP-TLS that differ to 5G-AKA?
- RQ1b: To what extent can differences between variables be effectively measured in practice using appropriate Key Performance Indicators (KPIs)? What are the potential pitfalls and challenges associated with the implementation of EAP-TLS?
- RQ2: What are the possible advantages of using EAP-TLS over 5G-AKA from an end-user perspective and operator perspective regarding security and manageability?
- RQ3: What is the impact of a new access point and authentication mechanism on the attack surface in 5G systems, and how does the attack surface change compared to the previous access point and authentication mechanism?

3 Private 5G Network with a non-3GPP access

The Proof of Concept aims to integrate EAP-TLS into a 5G core and evaluate its performance in a system with an untrusted non-3GPP access point. Open5GS [5], an open source 5G core lacking EAP-TLS, was extended for this purpose with the help of the BearSSL library [6]. The modifications include adding EAP-TLS to key network functions and interfaces. The N3IWF from the free5GCore project [7] was adapted for use in Open5GS. This PoC builds on previous work, incorporating EAP-TLS into the open source project my5G-non3GPP-access [8] for authentication and system evaluation on the UE side. To be able to add an EAP-TLS authentication component to the 5G core, the architecture and implementation of the 5G core had to be studied based on the 5G standard and the Open5GS core. The Open5GS core was reverse engineered to understand the source code and to pinpoint where EAP-

TLS components have to be added. After finding out in which network functions and their respective APIs code had to be added, the implementation could be realized. The goal was to design an EAP-TLS implementation that can decode the Base64 encoded EAP-messages received from the AMF and handle incoming fragmented TLS Records as outgoing fragmented TLS Records. The main challenge encountered during the implementation was effectively handling fragmentation in the context of TLS records. New API parts had to be added to enable the communication between the two NFs AMF and AUSF for the EAP-TLS authentication. For the implementation an EAP library was implemented to wrap the TLS records in EAP messages. Besides the EAP-TLS implementation in the core EAP-TLS had to be implemented in the UE for the PoC to be able to evaluate it afterwards.

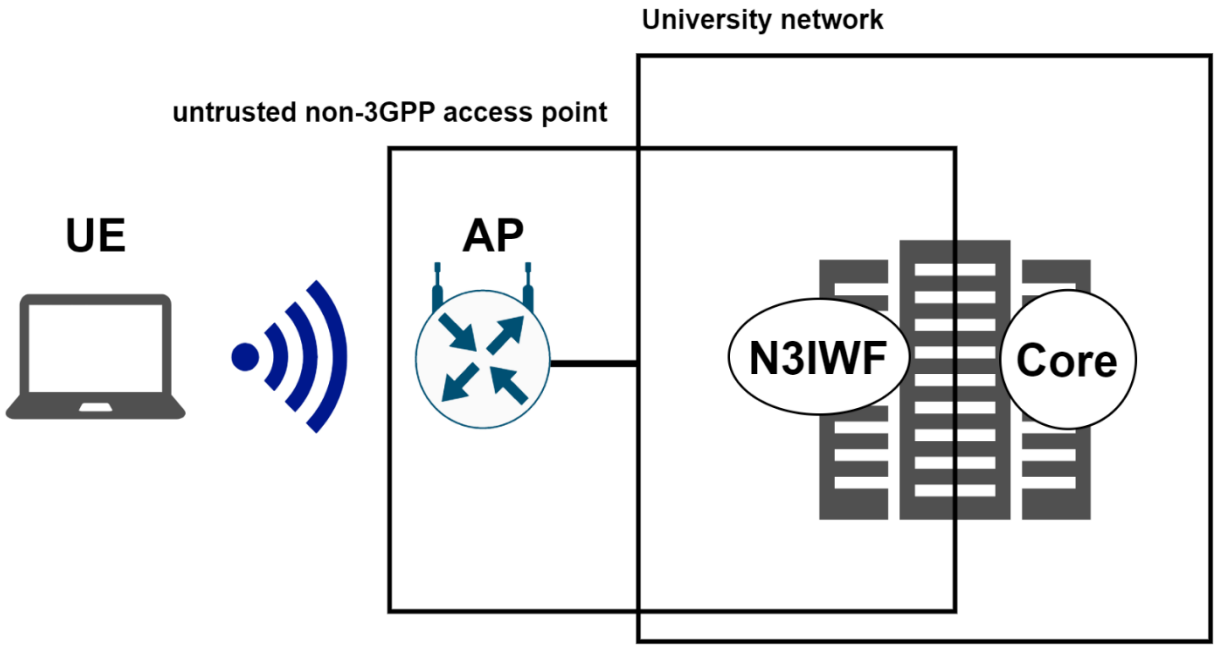


Figure 1: Developed private 5G testbed with a non-3GPP access

For the evaluation of EAP-TLS as an alternative primary authentication for private networks the testbed pictured in Figure 1 is used. The 5G core, the Open5GS core with the new EAP-TLS implementation, and the N3IWF, of the free5GC Open Source project, are operated in a virtual machine based on an Ubuntu system. Each 5G core Network Function runs as a Docker Container, only the N3IWF is run as a process. An UE has to be in the same network to be able to connect to the N3IWF. Therefore as the Access Point of the Non-3GPP Access Network a wireless router is used. As the UE a Laptop is used, which connects itself to the WLAN and is then capable of reaching the N3IWF in the university network. For the connection and authentication procedure

over the non-3GPP access the modified my5G-non3GPP-access runs on the laptop.

4 Evaluation

Based on the novel architecture of the 5G core, enabling the use of Software-Defined Networking (SDN) and Network Function Virtualization (NFV), and the OpenRAN technology, resulting in opening up the interfaces between the components, every part of the system can be from different vendors. Hence, neither the RAN nor the backend, e.g. 5G core, can be seen as a black box for attackers, as both are standardized. From all this information publicly available and possible different vendors involved, emerges a complicated system with many possible attack vectors and weaknesses.

4.1 Differences and Advantages of EAP-TLS

4.1.1 Mutual Authentication

A key advantage of EAP-TLS, distinguishing it from 5G-AKA, is the mutual authentication through the exchange and validation of certificates. In contrast to 5G-AKA, where the 5G network does not authenticate itself to the UE during the authentication process, EAP-TLS requires the network to authenticate itself with a certificate. This provides assurance to the UE that it is connecting to the intended network. In the context of private networks, ensuring the UE's connection to the correct network holds greater significance compared to public networks, as these scenarios often involve device control and management. If an attacker were able to spoof or fake the private network, they could gain unauthorized access to devices and potentially compromise sensitive data.

4.1.2 Messages exchanged

A disadvantage of EAP-TLS is that it needs more messages to achieve the authentication compared to 5G-AKA. Especially when combined with a fragmentation of the TLS records more messages are needed. The maximum size of an EAP message allowed by the 5G-Standard [9] is 1503 bytes long, but normal certificates exceed this limitation. The TLS records need to be fragmented and sent in different EAP messages. Resulting in 2-4 times more messages sent, when using EAP-TLS instead of 5G-AKA.

4.1.3 Key Cryptography

Another difference to 5G-AKA is that EAP-TLS is based on asymmetric key cryptography, whereas 5G-AKA uses symmetric key cryptography. Therefore, the UE does not need to be provisioned with a private symmetric key,

but requires a user-defined or existing certificate. The downside is that the UEs have to be provisioned with certificates and operating a Public-Key-Infrastructure (PKI) is more elaborate and complex.

4.1.4 UICC

For 5G-AKA and EAP-AKA' an UICC has to be used for storing the subscriber credentials, as specified in ETSI 133 501 [2]. For any other authentication mechanism, this obligation does not apply. Thus with the implementation of EAP-TLS devices without an UICC are enabled to connect to a 5G network.

4.1.5 Summary

In summary, for private networks that prefer to avoid the key management associated with symmetric mechanisms, and may already have an existing PKI in place, opting for EAP-TLS offers a more practical solution. EAP-TLS provides the added security benefit of mutual authentication, ensuring that devices, such as IoT devices, can confidently verify the authenticity of the network they are connecting to. Additionally, devices without UICCs are enabled through EAP-TLS to use the 5G-network.

4.2 Attacks realised

In the following the attacks, that are either realized in practice or mapped on the operated system with its configuration, are presented.

4.2.1 Denial-of-Service

During a Denial-of-Service (DoS) attack, an aggressor attempts to disrupt a service by overwhelming it with a flood of messages. During the test, an UE overloaded the 5G core with either a flood of messages or manipulated messages, targeting the Network Functions of the 5G core AMF and AUSF. With 10 UEs on the same device, 1000 requests were sent, gradually increasing to 10000 requests. Despite the service quality being impaired, the 5G core and N3IWF did not crash.

The 5G core was also flooded with 1000 and 10000 requests within the same time span using InitialUEMessages. In the first attempt, the 5G core operated normally afterward, but the service was impaired during the attack. In the second attempt with 10000 requests, the 5G core network functions worked as intended, except for impaired quality of service and a MongoDB database crash. However, the database restarted without data loss.

The 5G core demonstrated resilience in a smaller-scale DoS attack, restarting the crashed database when operated in a docker container. For larger private networks, using container orchestration software like Kubernetes is recom-

mended to manage load, scale, and automatic restarts. The system should detect and block DoS attacks, potentially by monitoring the traffic from specific IDs and rejecting excessive requests. In the KPI measurement 4.3, the 5G core's vulnerability to DoS attacks is further discussed.

4.2.2 Replay attacks

In a replay attack, messages exchanged between parties are intercepted and replayed or manipulated to breach security. Various security measures combat these attacks. If NAS messages are replayed during authentication, the AMF rejects the attacker's ClientHello due to an outdated AMF-UE-NGAP-ID. If the ClientHello includes the updated ID, it proceeds to authentication, but the AUSF rejects the EAP-Message due to incorrect EAP-ID, returning an error. Even if anti-replay measures fail, the BearSSL library detects replayed Client Certificate messages, resulting in a TLS Handshake failure. Additionally, the EAP-ID's predictability raises security concerns, therefore randomizing it could enhance security. Private networks with private PKI prevent unauthorized certificate validation. With public CA, the 5G core should ensure certificates match the Subscription Permanent Identifier (SUPI) to counter attackers.

4.2.3 Entity-in-the-Middle

The following entity-in-the-middle attacks are the results of a formal analysis of EAP-TLS in 5G by Zhang et al. [10]. For both attacks it will be discussed how the theoretical security weakness works on a real 5G system. Both attacks assume that the attacker sits between the N3IWF and AMF. Zhang et al. setup different security properties for statements made in the 5G standard regarding the security architecture [2]. Two security properties are not fulfilled, as found by J. Zhang et al.

- Both the home network (AUSF) and the subscriber (UE) should agree on the identity of each other after successful termination.
- Both the home network (AUSF) and the subscriber (UE) should agree on the premaster key (Rprekey) after successful termination

Zhang et al. suggest the following adjustment to prevent both attacks. The UE sends a random number in the InitialUEMessage and the encrypted SUPI with the public key of the AUSF to the AUSF. Only the AUSF can decrypt this random number with its private key and sends this random number in future messages. This random number has to be sent encrypted by the AUSF. The AUSF could encrypt the random number with the public key of the UE as it can map the certificate to the SUPI. For that all certificates of the UEs would

need to be stored or at least characteristics that can be used to uniquely identify the certificates, accessible by the 5G core.

The 5G system has been operated without any encryption of the SUPI as privacy concerns were not part of the research efforts and Open5GS at the start of the research only supported the Null Scheme encryption. The latest Open5GS version supports Profiles A and B for encryption of the SUPI. Therefore, the fix suggested by the paper would not work for the used version. For the violation of the first property, a private PKI with its own operated CA is the more adequate solution for a private network as these then can be operated completely independent. Operating its own PKI makes the system more secure as no attacker can get a signed certificate and the operator has full control of any certificate in circulation. For the second property, the attack would fail with RSA or ECDH used. ECDHE is recommended to use as it provides forward secrecy. RSA for key exchange has a history of known security vulnerabilities and should not be used.

4.3 Key Performance Indicator Measurements

In continuation of the formal discussion in section 4.1 regarding the variances and benefits of EAP-TLS over 5G-AKA, practical assessments were conducted to compare their authentication procedures using key performance indicators (KPIs) under research question RQ1b. The KPIs measured include Authentication Procedure Duration, RTT Average, Packet Counter, and Error Counter. The Authentication Procedure Duration starts from the initiation of the IKEv2 protocol between UE and N3IWF and ends with UE authentication. The RTT Average measures the round-trip time of packets between UE and 5G core, while the Packet Counter tallies the exchanged packets. The Error Counter indicates whether an error occurred during authentication (1 for error, 0 for success). These measurements were taken on the UE side across various scenarios for both EAP-TLS and 5G-AKA to discern practical differences in authentication mechanisms and draw conclusive insights for the research question.

4.3.1 Authentication procedure with one UE

One UE started approximately 350 consecutive authentication procedures with 10 different SUPIs for each protocol 5G-AKA, EAP-TLS with EC and EAP-TLS with RSA. The following graphs display the results. Using EAP-TLS as an authentication mechanism takes nearly twice as much time as 5G-AKA (see Figure 2). Depending on the certificate signature algorithm used (EC or RSA), there is a slight measurable difference in duration during the authentication process. Although EAP-TLS takes approximately double the time compared to 5G-AKA, it still completes within approximately 200 ms, while 5G-AKA

takes around 100 ms on average. When authenticating one UE at a time towards the 5G core, the duration does not exhibit significant deviations, as indicated by the low standard deviation values 18.3 ms, 21.6 ms and 28.8 ms. However, it is worth noting that the standard deviation for EAP-TLS with EC is slightly higher at 28.8 ms compared to 5G-AKA or EAP-TLS with RSA. Additionally, the 75%-Quantile for EAP-TLS with EC (190.8 ms) is lower than that of EAP-TLS with RSA (107.6 ms). Nevertheless, there are two noticeable spikes in the duration distribution, as depicted in Figure 2. These spikes can be considered outliers, which explain the higher standard deviation. In terms of statistics, EAP-TLS with RSA and EAP-TLS with EC are quite similar, with EAP-TLS with EC even slightly faster based on the mean (188.9 ms/179.3 ms) and median (50%-Quantile) (184.7 ms/173.7 ms) values. EAP-TLS requires 8 packets to be sent by UE and 5G core in total, whereas 5G-AKA only needs 4, one reason for longer Authentication procedure duration with EAP-TLS. The measured RTT difference between EAP-TLS and 5G-AKA is relatively small but becomes more significant when considering the additional packets involved in the extended Authentication Procedure of EAP-TLS. 5G-AKA has the smallest RTT followed by EAP-TLS with RSA and then EAP-TLS with EC and the values from the statistics mean (17.2 ms/18.8 ms/24.6 ms), standard deviation (4.3 ms/2.3 ms/6.0 ms), median (16.0 ms/18.5 ms/24.0 ms) draw a similar picture. Notably, the EAP-TLS with RSA version sends 12 packets, as the 8 packet version was unstable, and therefore the Authentication Procedure duration is a bit higher than with EAP-TLS with EC. The error rate in the authentication process includes errors that are received from the 5G core, N3IWF, or the UE itself, indicating the inability to process the messages. Among the three protocols, EAP-TLS with RSA has the lowest error rate at 6.8%, while 5G-AKA has an error rate of 7.6%. It is worth noting that most of the errors in both protocols are attributed to the UE's performance. However, EAP-TLS with EC exhibited an unusually high error rate when communicating with the N3IWF during this measurement, resulting in an overall error rate of 26%. Even without considering the N3IWF errors, the error rate for EAP-TLS with EC remains at 18%. It should be mentioned that the measurements for EAP-TLS with EC were conducted using a version where only 4 messages were sent on the UE side for the TLS authentication, whereas the

version used for RSA involved 6 packets. This version appears to be more unstable than the one with 6 packets which was used for RSA.

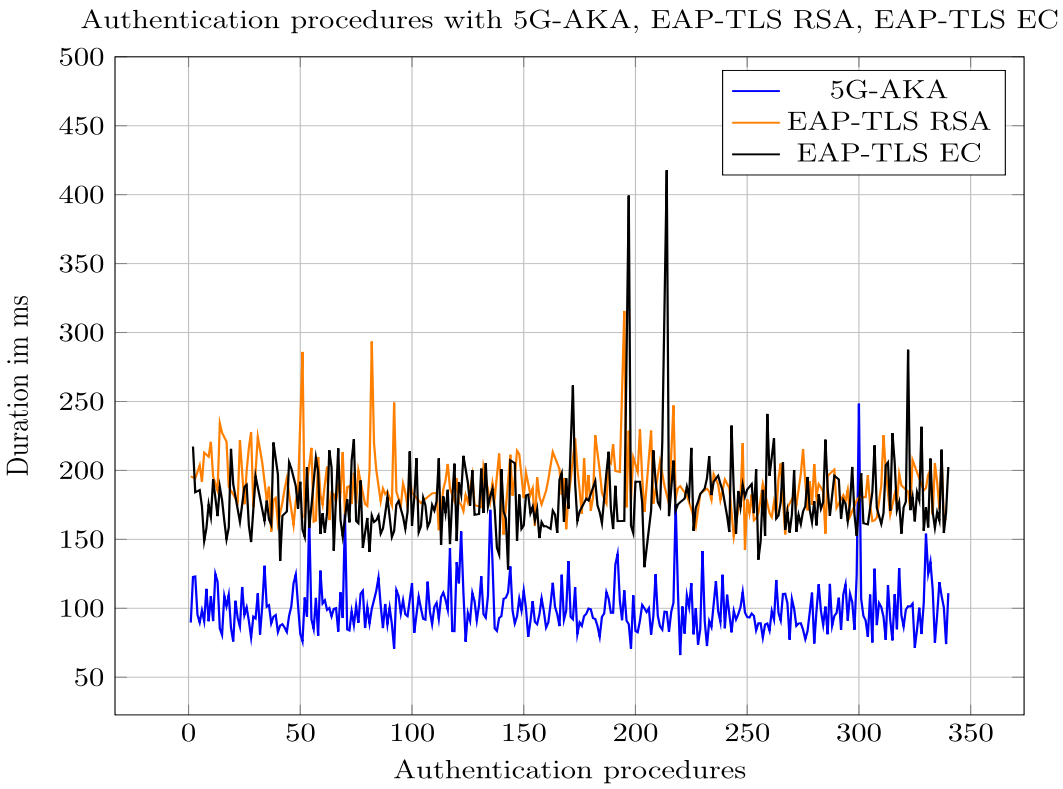


Figure 2: One UE Authentication Procedure Duration

4.3.2 Authentication procedures with ten UEs

Ten UEs each with one SUPI started approximately 250 authentication procedures together. The following graphs display the results. The graph for the Authentication procedure duration for ten UE (see Figure 3) exhibits the single UE (see Figure 2). The graph for 5G-AKA for ten UEs looks very similar to the one for one UE and the numbers in the statistics mean (103.1 ms/232.7 ms/231.6 ms), standard deviation (20.0 ms/396.5 ms /610.9 ms) and median (99.0 ms/222.7 ms/166.1 ms) are in the same range. EAP-TLS has more outliers, which make the standard deviation and the mean larger. The quantiles are very similar to the ones from the data with one UE. Hence, with 10 UEs the deviation increases, but most procedure durations stay the same. In the Round-Trip-Times graphs, it can be observed that there are more outliers compared to the scenario with one UE. Similar to the duration measurements, the Round-Trip-Times for 5G-AKA do not significantly differ from those of EAP-TLS. However, both EAP-TLS with RSA and EAP-TLS with EC exhibit higher standard deviations and means. It is important to note that the majority of authentication procedures fall within a close range, as indicated

by the quantiles. The presence of outliers is responsible for distorting the mean and standard deviation values. For ten UEs the error rates for 5G-AKA and EAP-TLS with RSA were actually smaller than for one UE with 2% and 0.4%. For EAP-TLS with EC the error rate skyrocketed to 61.8%, where 94% of the errors are attributable to an unstable UE. If only the errors from N3IWF and the 5G core are considered the error rate drops to 3.2%.

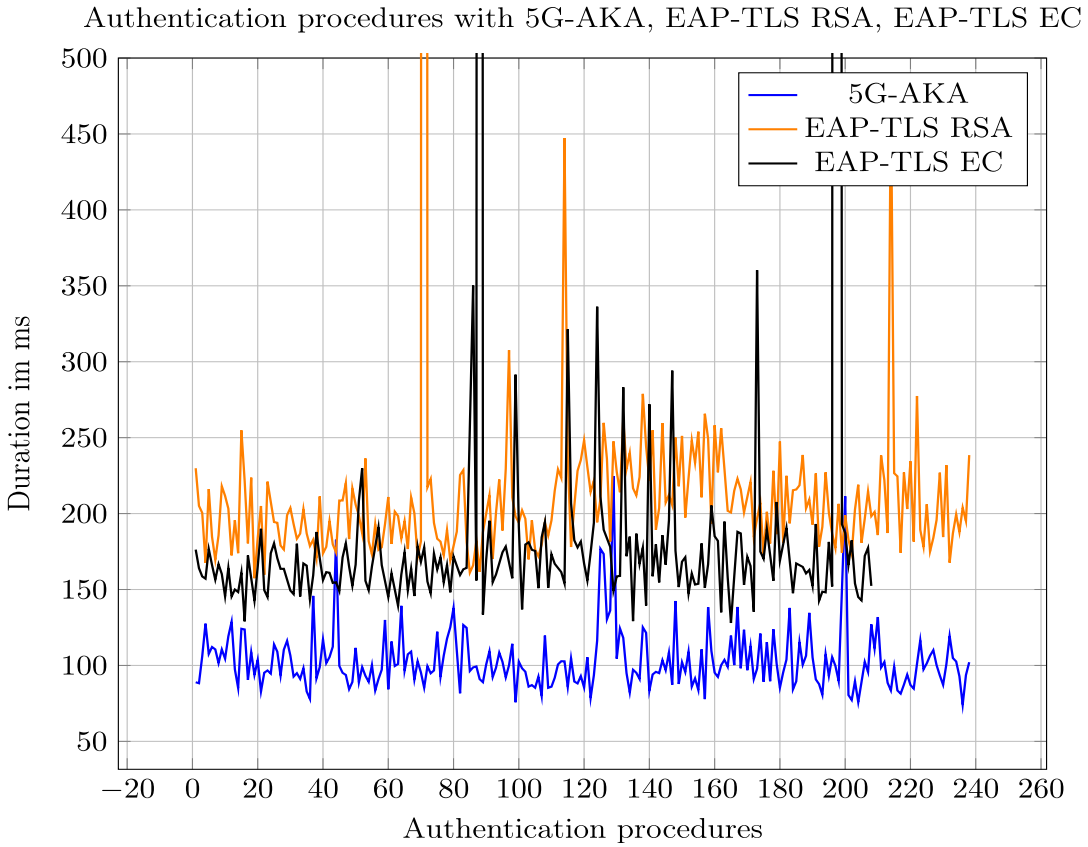


Figure 3: Ten UE Authentication Procedure Duration

4.3.3 Throttled Network

5G-AKA and EAP-TLS were both tested in a simulated unstable network with a latency of 200 ms and a packet loss of 10% to see how this impacts the KPIs. The authentication procedure duration is for both as expected a lot longer because of the latency of 200 ms. The graph in Figure 4 shows that both don't have a big fluctuation despite the 200 ms latency and 10% packet loss. This is reflected in the statistics mean (952.7 ms/1852.8 ms/1386.0 ms) and standard derivation (521.7 ms/39.6 ms /15.7 ms) for all of the three. The one outlier, which can be seen in the graph, for 5G-AKA distorts the mean and the standard deviation. The quantiles

- 25%-Quantile (897.2 ms/1835.6 ms/1375.5 ms)

- 50%-Quantile (904.1 ms/1849.9 ms/1386.5 ms)
- 75%-Quantile (912.2 ms/1867.3 ms/1392.5 ms)

reveal that the majority of authentication procedures exhibit minimal deviation, indicating a consistent performance across most cases. The RTT draws the same picture as the authentication procedure duration in a throttled network. The numbers mean (244.1 ms/225.0 ms/226.1 ms) and median (220.5 ms/224.0 ms/226.0 ms) are increased by the latency, but the deviation (260.5 ms/4.7 ms/3.7 ms) is still very similar to the execution in the unthrottled network. The latency of 200 ms and the 10% packet loss result in an error rate of 31% for 5G-AKA, 70% for EAP-TLS with RSA and EAP-TLS with EC of 83%. EAP-TLS with RSA had an unusual high error rate with the N3IWF, but even if this is neglected the error rate is at 50%. An increase in failure in an unstable network is expected, as the transport layer protocol between the UE and N3IWF, UDP is unreliable. Between N3IWF and AMF a reliable transport layer protocol with SCTP is used and therefore packet loss should not occur.

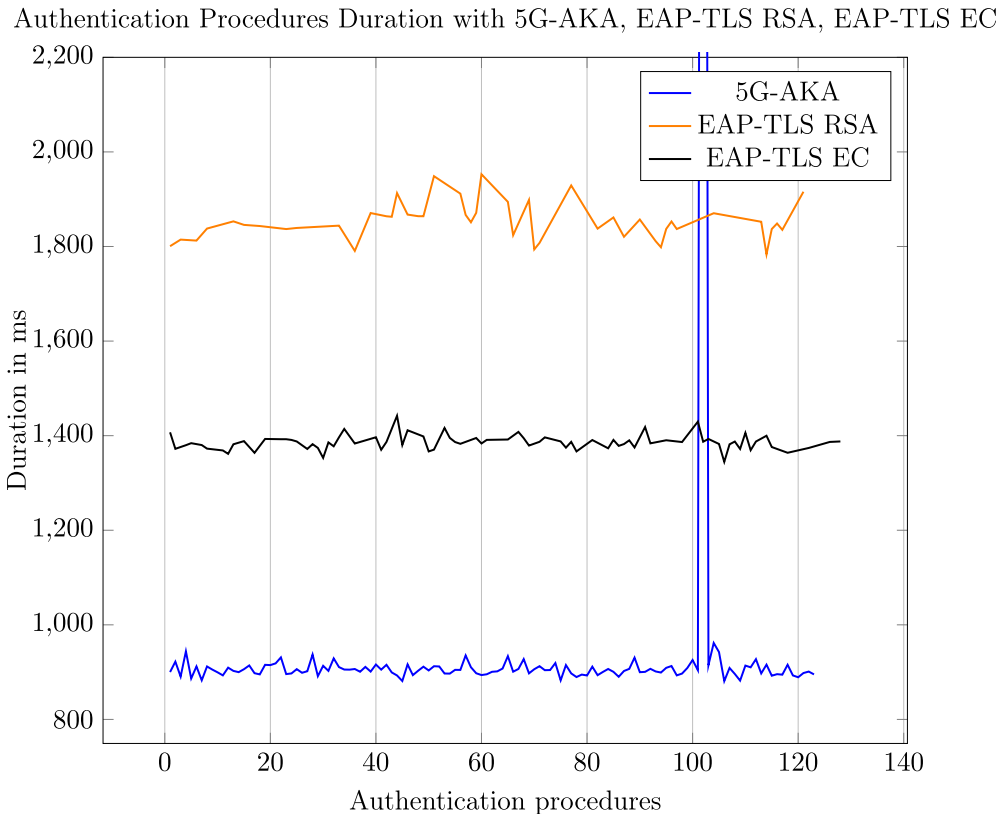


Figure 4: One UE Authentication Procedure Duration in Throttled Network

5 Conclusion

Implementing the EAP-TLS authentication protocol in a 5G system extends secure authentication to non-UICC-equipped UEs, enhancing device compatibility and system flexibility. This integration was achieved through a prototype 5G system, facilitating authentication via N3IWF for devices lacking 5G capability. Realizing this implementation involved analyzing the service-based architecture specified in the 5G standard and employing Network Function Virtualization and Software Defined Networking technologies. Adherence to the Service Based Interface and NGAP/NAS specifications ensured seamless communication between network functions and between UE and the 5G core.

The proof of concept underwent testing in various practical scenarios to provide insights into usage, configuration, and vulnerabilities. In the context of RQ1, emphasizing asymmetric key management and possessing a functional Public Key Infrastructure, EAP-TLS emerges as a more practical choice for private networks. Despite requiring more time and packets compared to 5G-AKA, EAP-TLS ensures mutual authentication, enhancing trust towards the network for devices like IoT devices. However, its implementation exhibited a higher error rate, especially in unstable networks, making it more susceptible to DoS attacks. Additionally, the complexity of EAP-TLS implementation elevates the risk of encountering issues compared to 5G-AKA.

Turning our attention to RQ2, an evident benefit for the end-user and their UE is the requirement for the 5G Core to authenticate itself using its certificate, thereby achieving mutual authentication. An operator of a private 5G-network should operate its own Public-Key-Infrastructure with his own Certificate Authority, gaining full control over certificates that are signed and therefore minimizing the risk that the wrong person has a valid certificate. A huge advantage for both end-user and operator is that a UICC is no longer mandatory. An operator should not use TLS below version 1.2 for EAP-TLS and at best use ECDHE for the key exchange.

Regarding RQ3, the inclusion of a Wi-Fi access point and N3IWF, along with the adoption of EAP-TLS for authentication, shifts the attack surface, introducing new vectors compared to gNodeB and 5G-AKA. While EAP-TLS brings a distinct authentication method, few vulnerabilities were found, mainly related to potential DoS attacks from retransmission issues. Thus, it's recommended to use TLS 1.2 with ECDHE cipher suites and a private PKI in private networks. The results of this work are focused on completely independently operated private 5G-networks. 5G also offers the possibility to operate dif-

ferent deployment models of nonpublic networks, e.g. a public network operator provides a private 5G network enabled by network slicing within his public network. The necessity of a PKI in such networks operated by public network operators may be debated, but the benefits of having complete control of the signed certificates remain.

6 Outlook

Based on the findings of this paper, there are promising fields for future research. One such field is the implementation of EAP-TLS in an UE with a modem that is 5G capable and utilizing the 3GPP-access, such as a gNodeB. This next step would involve assessing the behavior of EAP-TLS when executed over a 3GPP-access network and drawing a comparison to 5G-AKA in this particular scenario. This would provide valuable insights into the performance and security implications of EAP-TLS in different access network environments. Building on this work, the next course of action involves exploring ways to leverage the key generated during authentication in conjunction with a modem. This approach facilitates UE authentication via non-3GPP access, eliminating the need for a UICC. Consequently, the UE would still have the capability to access 3GPP and utilize a mobile network. The recommendation for a self-hosted PKI needs to be re-evaluated in the context of private networks operating within a public network. Future research should take into account factors such as the types of devices utilized and the accessibility considerations, along with the operational effort required for maintaining a PKI.

Acknowledgement

I would like to to thank Prof. Dr. Luigi Lo Iacono for his great support over the course of my work and thanks to my colleagues from the Data and Application Security Group, especially Florian Dehling and Benedikt Malchow.

My work was part of the research project ODEA.5G (2008gif018) which was funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia.

Thank you to my girlfriend Lena, for all her love and support.

References

[1] ETSI. 5G; System architecture for the 5G System (5GS) (3GPP TS 23.501 version 16.6.0 Release 16). Standard, ETSI, Oct. 2020.

[2] ETSI. 5G; Security architecture and procedures for 5G System (3GPP TS 33.501 version 16.6.0 Release 16). Standard, ETSI, Apr. 2021.

[3] ETSI. 5G; Access to the 3GPP 5G Core Network (5GCN) via non-3GPP access networks (3GPP TS 24.502 version 16.4.0 Release 16). Standard, ETSI, Aug. 2020.

[4] ETSI. 5G; NR; NR and NG-RAN Overall description; Stage-2 (3GPP TS 38.300 version 16.4.0 Release 16). Standard, ETSI, Jan. 2021.

[5] S. Lee. Open5GS is a C-language Open Source implementation for 5G Core and EPC, i.e. the core network of LTE/NR network (Release-17). https://open5gs.org/.

[6] T. Pornin. BearSSL. https://www.bearssl.org/, 2018.

[7] free5Gc. The free5GC is an open-source project for 5th generation (5G) mobile core networks. The ultimate goal of this project is to implement the 5G core network (5GC) defined in 3GPP Release 15 (R15) and beyond. . https://www.free5gc.org/.

[8] M. T. Lemes, A. M. Alberti, C. B. Both, A. C. De Oliveira Junior, and K. V. Cardoso. A tutorial on trusted and untrusted non-3gpp accesses in 5g systems—first steps toward a unified communications infrastructure. IEEE Access, 10:116662–116685, 2022. doi: 10.1109/ACCESS.2022.3219829.

[9] ETSI. 5G; Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3 (3GPP TS 24.501 version 16.5.1 Release 16). Standard, ETSI, Aug. 2020.

[10] J. Zhang, L. Yang, W. Cao, and Q. Wang. Formal analysis of 5g eaptls authentication protocol using proverif. IEEE Access, 8:23674–23688, 2020. doi: 10.1109/ACCESS.2020.2969474.

[10] ETSI. 5G; 5G System; Authentication Server Services; Stage 3 (3GPP TS 29.509 version 16.4.0 Release 16). Standard, ETSI, July 2020.

[11] ETSI. 5G; Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3 (3GPP TS 24.501 version 16.5.1 Release 16). Standard, ETSI, Aug. 2020.

[12] ETSI. 5G; Access to the 3GPP 5G Core Network (5GCN) via non-3GPP access networks (3GPP TS 24.502 version 16.4.0 Release 16). Standard, ETSI, Aug. 2020.

[13] S. Rommer, P. Hedman, M. O. andLars Frid, S. Sultana, and C. Mulligan. 5G Core Networks - Powering Digitalization, publisher = Elsevir. 2019.

[14] Vollbrecht, J., Carlson, J. D., Blunk, L., Aboba, Dr. B. D., & Levkowetz, H. (2004). Extensible Authentication Protocol (EAP). RFC Editor. https://doi.org/10.17487/RFC3748

[15] E. Rescorla and T. Dierks. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246, Aug. 2008. URL https://www.rfc-editor.org/info/rfc5246.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Kombinatorisches Testen von TLS-Bibliotheken

Marcel Maehren[1], Dr. Robert Merget[2], Niklas Niere[3], Dr. Simon Oberthür[3], Conrad Schmidt[4], Prof. Dr. Jörg Schwenk[1], Prof. Dr. Juraj Somorovsky[3], Ovidiu Ursachi[5]

Kurzfassung:
TLS (Transport Layer Security) ist der wichtigste praktisch eingesetzte Sicherheitsstandard – mit TLS werden die Authentizität, Integrität und Vertraulichkeit privater und geschäftlicher Kommunikation sichergestellt, Datenschutz gewährleistet und komplexe IT-Systeme abgesichert.
Während die theoretische Sicherheit von TLS gut untersucht und verstanden ist, entstehen durch Implementierungsfehler immer wieder gravierende Sicherheitslücken, die für Angriffe ausgenutzt werden können (HeartBleed [1], POODLE [2], ROBOT [3], DROWN [4], ...). Während die ersten Implementierungsfehler noch leicht manuell gefunden werden konnten, nutzen neuere Angriffe ein komplexes Zusammenspiel mehrerer TLS-Versionen und zahlreicher TLS-Features.
Um die Sicherheit solch komplexer kryptografischer Implementierungen sicherstellen zu können, ist die Entwicklung von Methoden zum automatisierten Testen aller möglichen Kombinationen dieser Features auf allen Ebenen unbedingt erforderlich. In diesem Beitrag präsentieren wir unsere TLS-Testsuite, welche auf der TLS-Analysebibliothek „TLS-Attacker“ [5] basiert und große Testabdeckung durch den Einsatz von kombinatorischem Testen erzielt. Die Testabdeckung ist konfigurierbar und somit ist unsere Testsuite für unterschiedliche Zwecke für den Einsatz durch Entwickler, Integratoren, Betreibern sowie Prüfinstitute und Aufsichtsbehörden geeignet, um TLS-Implementierungen im Hinblick auf Sicherheit und Interoperabilität zu testen.

Stichworte: Automatisiertes Testen, Kombinatorisches Testen, Testsuite, TLS

1 Einführung

TLS ist mit Abstand die wichtigste kryptografische Sicherheitskomponente: Neben seinem bekannten Einsatz in HTTPS wird TLS auch in vielen anderen Protokollen wie beispielsweise SMTPS, POP3S, oder FTPS genutzt. Auch mobile Apps setzen TLS als Baustein ein, um Softwareupdates abzusichern und Daten vertraulich und integer zu übertragen. Die Sicherheit von TLS als Komponente muss daher für alle möglichen Einsatzgebiete garantiert werden. Ohne die Sicherheitsgarantien, die diese Komponente bietet, können

[1] Ruhr-Universität Bochum
[2] Technology Innovation Institute
[3] Universität Paderborn / SICP
[4] Hackmanit GmbH, Bochum
[5] InnoZent OWL e.V., Paderborn

sensible Daten abgehört werden oder Schadsoftware kann in die Datenübertragung eingeschleust werden.

Dabei ist die Aufgabe, welche die TLS-Bibliothek übernehmen muss, keine einfache. Das TLS-Ökosystem, in dem eine TLS-Bibliothek agieren muss, ist hochkomplex und in einem ständigen Wandel. Neue Standards wie beispielsweise TLS 1.3 oder DTLS müssen integriert werden, während auch ältere Standards weiterhin unterstützt werden müssen. Zudem ist die eingesetzte Kryptografie hochsensibel gegenüber Implementierungsfehlern und immer komplexeren Seitenkanalangriffen, wodurch auch schon kleine Fehler in der Implementierung zu einem kompletten Verlust der Sicherheitseigenschaften führen können.

Die Komplexität von TLS und DTLS, aber auch die ständige (Weiter-) Entwicklung von neuartigen Angriffen machen es den Entwicklern sehr schwer, alle Schwachstellen für Angriffe schon während der Implementierung zu beseitigen. Dieses Testen gestaltet sich in der Praxis allerdings äußerst schwierig: Während funktionale Basistests (d.h. Tests, die überprüfen, ob TLS im normalen Betrieb funktioniert) relativ leicht zu implementieren sind, gibt es noch keine befriedigenden Lösungen für andere Bereiche des Testens.

Die von uns entwickelte quelloffene TLS-Testsuite adressiert diese Probleme, um zukünftig vollautomatisch und umfassend die Entwicklung und den Einsatz von TLS-Bibliotheken überwachen und analysieren zu können. So können TLS-Entwickler schon während des Entwicklungsprozesses Fehler erkennen und Produkthersteller können ihren Integrationsprozess überwachen.

Unsere TLS-Testsuite baut auf dem TLS-Analyse-Werkzeug TLS-Attacker [5] auf. TLS-Attacker ist eine TLS-Analysebibliothek, welche seit 2015 entwickelt wird, um TLS-Bibliotheken auf Schwachstellen zu untersuchen. TLS-Attacker enthält eine eigens zur Analyse von TLS-Bibliotheken entwickelte TLS-Implementierung, welche einem Sicherheitsanalysten einfachen Zugriff zu allen Aspekten einer Verbindung gibt.

2 Grundlagen

2.1 TLS – Funktionsweise

Das TLS-Protokoll ermöglicht es zwei Endpunkten, einen sicheren Kommunikationskanal aufzubauen. Zu diesem Zweck führen die Kommunikationspartner einen TLS-Handshake durch, um kryptografische Para-

meter auszuhandeln und TLS-Sitzungsschlüssel abzuleiten. Die ausgehandelten Parameter und Schlüssel werden dann innerhalb des Kanals verwendet, um die Vertraulichkeit, Integrität und Authentizität der ausgetauschten Nachrichten zu schützen.

Ein TLS-Handshake enthält viele kryptografische Parameter und kann durch optionale Erweiterungen, sogenannte TLS-Extensions, ergänzt werden. Handshakes unterscheiden sich außerdem nach TLS-Version. Die derzeit aktuellen TLS-Versionen sind TLS 1.2 und TLS 1.3. Ein Parameter, welcher ausgehandelt wird, ist die TLS Cipher-Suite. Sie legt die Art des Schlüsselaustausches, die Art der Authentifizierung, die Verschlüsselungsmethode sowie einen zu verwendenden Hash-Algorithmus fest. Eine Cipher-Suite könnte beispielsweise so aussehen:

TLS_DHE_RSA_WITH_AES_128_CBC_SHA256.

TLS-Extension werden im Handshake mit ausgehandelt und können Veränderungen des Handshake-Ablaufes bewirken. Beispielsweise können sie verändern, wie Nachrichten authentifiziert werden (Encrypt-Than-Mac), oder welche zusätzlichen Nachrichten geschickt werden (Heartbeat) und vieles mehr. Insgesamt ist die Anzahl der veränderbaren Parameter in einem TLS-Handshake zu groß, um jede Kombination durch einen naiven Ansatz systematisch zu testen.

2.2 TLS – Spezifikation und Komplexität

TLS ist in mehreren Dokumenten, sogenannten RFCs, spezifiziert. Da TLS iterativ gewachsen ist, alte Dokumente neue ergänzen oder invalidieren und es eine Vielzahl an verschiedenen Versionen und Erweiterungen gibt, ist die Anzahl der TLS-spezifischen RFCs sehr groß.

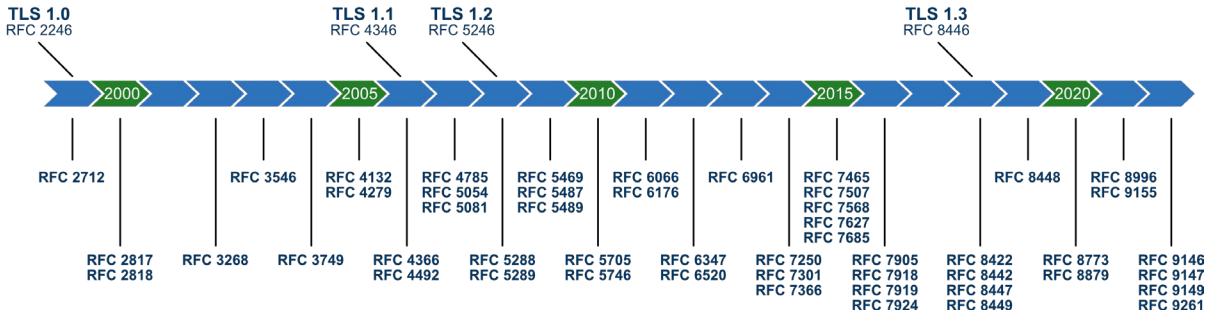


Abbildung 1: Zeitlinie zur Veröffentlichung von TLS-relevanten RFCs

Die große Anzahl von Protokollversionen, Erweiterungen und kryptografischen Algorithmen erhöht die Komplexität von TLS und macht die Implementierung einer sicheren TLS-Bibliothek sehr schwierig. Aus Gründen der Abwärtskompatibilität müssen Standard-TLS-Bibliotheken mehrere TLS-

Versionen ab TLS 1.0 und auch veraltete kryptografische Algorithmen wie 3DES unterstützen. Die Komplexität von TLS und die Anforderungen an die Abwärtskompatibilität führten in der Vergangenheit zu mehreren kritischen Angriffen.

Den Vorgaben in den RFCs zu folgen ist also von großer Bedeutung für die Sicherheit der Implementierung. Ein Beispiel für so eine sicherheitsrelevante Anforderung aus RFC 5246 [7] ist die folgende:
*The receiver **MUST** check this padding and **MUST** use the **bad_record_mac alert** to indicate padding errors.*

Die Vorgabe ist eine Gegenmaßnahme gegen Padding Oracle Angriffe und muss von der Implementierung unabhängig von anderen gewählten Parametern erfüllt werden, um die Sicherheit der Verbindung sicherzustellen.

2.3 TLS – Parameter und Sicherheitsverhalten

Während frühere Versionen von TLS noch die wichtigsten kryptografischen Parameter ausschließlich über die Cipher-Suite ausgehandelt haben, gibt es inzwischen eine Reihe von weiteren Mechanismen, um der gewachsenen Zahl von kryptografischen Primitiven gerecht zu werden. Beispielsweise kann über die Cipher-Suite zunächst der Schlüsselaustausch über Elliptische Kurven vereinbart werden. Die Auswahl der elliptischen Kurve, einem zentralen Pfeiler der Sicherheit des aufzubauenden Kanals, läuft dann zusätzlich über die Supported Groups Extension ab. Außerdem definiert die Cipher-Suite zwar einen Schlüsseltyp für die Authentifizierung, lässt Details dazu, wie etwa die Frage nach der verwendeten Hashfunktion, aber offen. Auch hier erfolgt das Aushandeln über eine TLS-Extension. Bei der Interaktion von diesen drei beispielhaften Parametern gilt es zunächst zu beachten, dass jeder einzelne das Sicherheitsniveau der etablierten TLS-Verbindung beeinflussen kann. Beispielsweise kann der Sicherheitseffekt einer starken Elliptische Kurve durch die Wahl einer schwachen Hashfunktion für die Signatur negiert werden. Außerdem gibt es für jeden der drei Parameter Auswahlmöglichkeiten, die die zulässigen Optionen für die jeweils beiden anderen einschränken. In der Vergangenheit haben invalide Kombinationen von Parametern immer wieder Corner-Cases in Bibliotheken ausgelöst, durch die Sicherheitsüberprüfungen umgangen werden konnten. Sicherheitstests sollten daher Parameter nicht isoliert betrachten, sondern besonderen Fokus auf die Interaktion von Parametern untereinander legen.

2.4 Kombinatorisches Testen

Die Komplexität von Schwachstellen in Softwaresystemen kann stark variieren. Um manche Schwachstellen zu finden, reicht es aus einfache Softwaretests zu schreiben, da die Schwachstellen unabhängig von weiteren Parametern auftreten. Andere komplexe Schwachstellen tauchen nur auf, wenn eine bestimmte Kombination von Parametern auftritt, wie in 2.3 beschrieben. Um solche Fehler finden zu können, sollten Tests parametrisiert werden, wobei gleiche Tests mit unterschiedlichen Parametern durchgeführt werden. Jedoch kann die steigende Anzahl an Parametern zu einer kombinatorischen Explosion führen, wodurch nicht mehr alle Testkombinationen durchgetestet werden können. So kann es beispielsweise bei einem Test mit 10 Parametern, bei dem jeder Parameter 10 unterschiedliche Werte annehmen kann, zu 10000000000 Parameterkombinationen kommen, welche die Testausführung nicht praktikabel machen. Es ist zudem unwahrscheinlich, dass alle 10 Parameter einen bestimmten Wert annehmen müssen, damit sich eine Schwachstelle zeigt. Es ist jedoch wahrscheinlich, dass lediglich eine Untermenge der zu testenden Parameter gewisse Werte annehmen muss, damit sich eine Schwachstelle zeigt.

An dieser Stelle setzt kombinatorisches Testen an. Beim kombinatorischen Testen wird eine Teststärke k definiert, um dann gezielt alle Teilparameterkombinationen zu testen, welche aus insgesamt k Parametern bestehen. Dabei wird versucht die Parameter so auszuwählen, dass möglichst wenige Tests ausgeführt werden müssen, um alle Teilparameterkombinationen zu testen.

Bisherige Arbeiten im Bereich von Analysen von TLS und DTLS haben sich vor allem auf einzelne Sicherheitslücken konzentriert. Das volle Potenzial von kombinatorischem Testen von umfangreichen Bibliotheken und dessen Performanz wurde noch nicht erforscht. Damit legen wir einen neuen Ansatz für die Analysen von kryptografischen Protokollen.

3 Design

3.1 Testsuite

Die in diesem Projekt entwickelte Testsuite “TLS-Anvil” [8] baut auf dem TLS-Analyse Werkzeug TLS-Attacker auf. TLS-Attacker ist eine in Kooperation mit der Ruhr-Universität Bochum, der Universität Paderborn und der Hackmanit GmbH entwickelte TLS-Analysebibliothek, welche seit 2015 entwickelt wird, um TLS-Bibliotheken auf Schwachstellen zu untersuchen.

TLS-Attacker enthält eine eigens zur Analyse von TLS-Bibliotheken entwickelte TLS-Implementierung, welche einem Sicherheitsanalysten einfachen Zugriff zu allen Aspekten einer Verbindung gibt.

Weiterhin setzen wir in unserem Projekt das von der RWTH Aachen entwickelte Testframework coffee4j [9] ein, welches es ermöglicht, kombinatorische Tests mit Hilfe von JUnit durchzuführen. Coffee4j kann für die Inputgenerierung von kombinatorischen Test-Parametern und zur Fehlercharakterisierung genutzt werden. Die Nutzung von JUnit, dem am meisten verwendeten Testing-Framework in Java, erleichtert das Schreiben von Tests, da eine gewohnte und effiziente Umgebung genutzt werden kann.

TLS-Anvil geht bei einer Testausführung in 3 Phasen vor:

1. Feature Extraction: In dieser Phase wird die zu untersuchende Software zunächst auf ihre Funktionalität gescannt. Dazu zählen unter anderem die unterstützten Cipher-Suites und TLS-Versionen. Nach der Feature Extraction können Tests ausgeschlossen werden, welche Funktionen voraussetzen, die nicht unterstützt werden.
2. Test-Phase: Im nächsten Schritt werden hintereinander alle Tests mit Hilfe von TLS-Attacker ausgeführt. Dabei wird jedes Test-Template nach dem Prinzip des kombinatorischen Testens mehrmals mit unterschiedlichen Parametern gestartet. Bei einem Test werden ein oder mehrere Handshakes mit der zu testenden Software durchgeführt und die Antworten gespeichert.
3. Testauswertung: Im letzten Schritt wird ausgewertet, inwieweit die geprüfte Software die Tests bestanden hat. Dabei wird für jedes Test-Template in 4 Kategorien unterteilt: erfolgreich bestanden, konzeptuell bestanden, teilweise fehlgeschlagen oder komplett fehlgeschlagen. Eine Auswertung kann anschließend über Einsicht in den generierten Report erfolgen.

3.2 Test-Templates

Ein Test-Template ist eine Vorlage, basierend auf einer Anforderung an die zu testende Software. Es definiert das gewünschte Ergebnis aller von ihm abgeleiteten Testfälle. Test-Templates definieren welche TLS-Nachrichten gesendet werden und welche Antworten von der Testsuite erwartet werden. Außerdem definiert es, wann ein Testfall erfolgreich ist oder fehlschlägt.

In TLS-Anvil haben wir verschiedene Test-Templates implementiert, welche hauptsächlich auf den Anforderungen aus 13 verschiedenen TLS RFCs basieren. Der Fokus lag dabei auf Anforderungen, die durch die Schlüsselwörter „MUST“, „SHALL“ und „REQUIRED“ sowie „MUST NOT“, „SHALL NOT“

als besonders wichtig gekennzeichnet sind [10]. Weiterhin wurden Tests zu impliziten Anforderungen geschrieben, welche zwar aus dem Kontext des RFCs hervorgehen, jedoch nicht explizit genannt werden, sowie Tests zur Einhaltung der im RFC definierten State Machine und aus der Literatur bekannten Fehlern.

Da wir für die Testausführung JUnit nutzen, ist es leicht einen neuen Test zu schreiben. Im Folgenden wird ein Beispiel gezeigt, welches prüft, ob die richtige TLS-Alert-Nachricht geschickt wird, wenn die Verify-Data der Finished-Nachricht nicht stimmt. Die Anforderung stammt aus RFC 5246 und lautet wie folgt:

*„Recipients of Finished messages **MUST** verify that the contents are correct.“*

In TLS-Anvil sieht der Test dann wie folgt aus:

- In Zeile 1 annotieren wir die Testfunktion mit @AnvilTest und vergeben eine TestId.
- In Zeile 3 wird eine Konfiguration für die Ausführung durch TLS-Attacker erstellt.
- Zeile 4 generiert eine Bitmaske, die zur Veränderung der Verify-Data verwendet wird.
- In Zeile 5 und 6 wird die Finished-Nachricht mit Hilfe der Bitmaske verändert.
- Zeile 7 generiert einen Handshake, jedoch ohne Finished-Nachricht.
- In Zeile 8 wird unsere modifizierte Finished-Nachricht angehangen.
- Zeile 9 führt den Handshake aus.
- In Zeile 10 wird anschließend getestet, ob der korrekte fatal alert empfangen wurde.

```
1 @AnvilTest(id = "5246-mEQLrje2mh")
2 public void verifyFinishedMessageCorrect(ArgumentsAccessor
                            argumentAccessor, WorkflowRunner runner) {
3    Config c = getPreparedConfig(argumentAccessor, runner);
4    byte[] modificationBitmask = parameterCombination.buildBitmask();
5    FinishedMessage finishedMessage = new FinishedMessage();
6    finishedMessage.setVerifyData(
         Modifiable.xor(modificationBitmask, 0));
7    WorkflowTrace workflowTrace = runner
        .generateWorkflowTraceUntilSendingMessage(
          WorkflowTraceType.HANDSHAKE, HandshakeMessageType.FINISHED);
8    workflowTrace.addTlsActions(new SendAction(finishedMessage),
                             new ReceiveAction(new AlertMessage()));
9    runner.execute(workflowTrace, c)
10      .validateFinal(Validator::receivedFatalAlert);
  }
```

3.3 Funktionalität

TLS-Anvil kann sowohl über die Kommandozeile, als auch über eine Webbenutzeroberfläche gestartet werden. Es ist möglich TLS oder DTLS Server und Clients zu testen. Dabei kann vorher ausgewählt werden, ob alle Tests oder nur ein bestimmtes Subset ausgeführt werden soll. Weiterhin kann die Teststärke gewählt werden, welche den k-Wert des kombinatorischen Testens bestimmt.

Wird die Testsuite über die Kommandozeile gestartet, generiert das Programm einen Ordner, welcher den Testreport sowie alle detaillierten Ergebnisse enthält. Dieser kann nachträglich in der webbasierten Benutzeroberfläche (siehe Abbildung 2) importiert werden. Wird der Test direkt aus der Oberfläche gestartet, werden die Ergebnisse automatisch importiert und können auch schon während der Testausführung eingesehen werden.

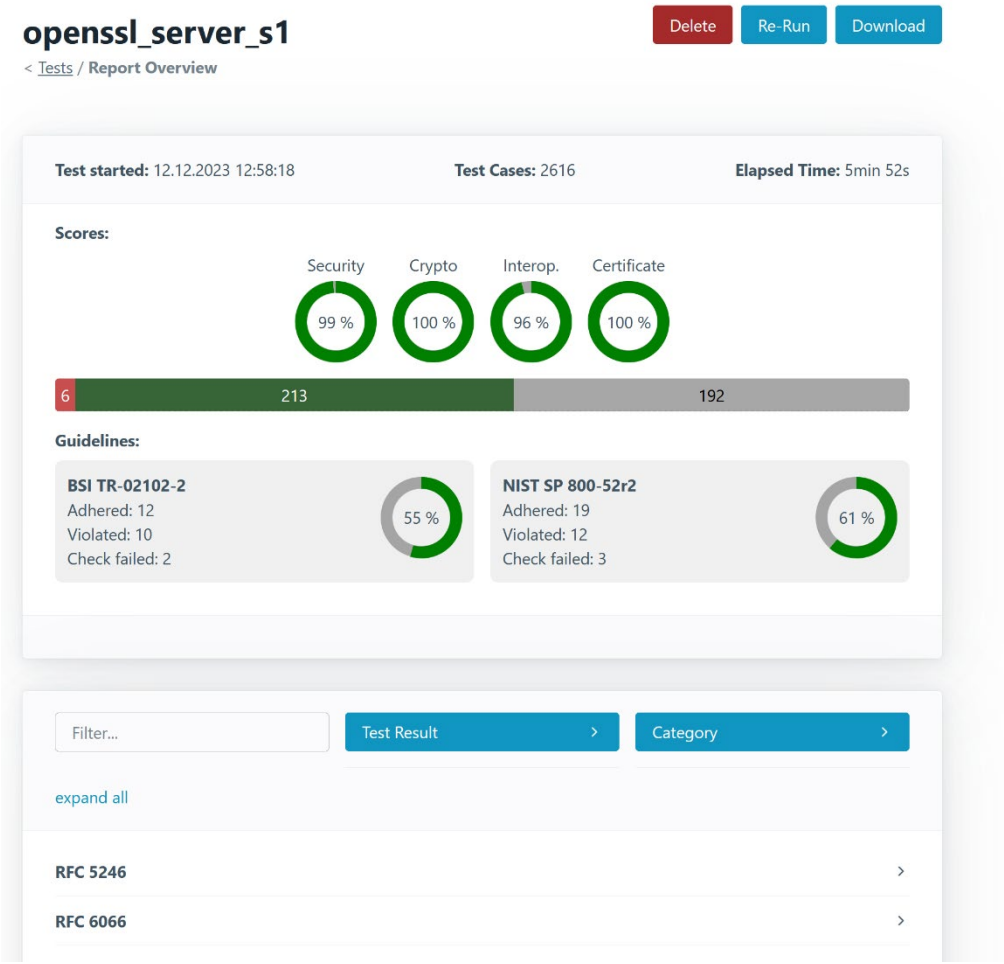


Abbildung 2: Screenshot der Benutzeroberfläche von TLS-Anvil. Übersicht eines Testreports: oben eine generelle Übersicht sowie Verweise auf Richtlinien, unten einzelne Testfälle geordnet nach RFC.

Wie bereits beschrieben, wird bei einer Testausführung zunächst eine Feature-Extraction ausgeführt, welche das zu testende System auf unterstützte Funktionen prüft. Anhand dieses Vorab-Scans kann die Testsuite zusätzlich herausfinden, ob die entprechende TLS-Implementierung sich an Sicherheitsrichtlinien des BSI oder der NIST hält. Die Einhaltung dieser Richtlinien kann in der Benutzeroberfläche nachgeschlagen werden.

Die Auswertung gibt einen Überblick über die Testergebnisse in verschiedenen Kategorien, wie „Sicherheit“, „Kryptografie“ oder „Interoperabilität“. Möchte man Details sehen, so können für jedes Test-Template das Testergebnis und die einzelnen Testcases angezeigt werden. Wurde ein Test nicht bestanden, kann man hier nachschlagen, welcher Testcase ausschlaggebend war. Man kann weiterhin den aufgezeichneten Netzwerkverkehr betrachten um genauere Details herauszufinden. Die Testsuite gibt außerdem eine Vermutung ab, welcher Parameter oder welche Parameterkombination für den Fehler verantwortlich sein könnte. Schlägt ein Test-Template bei allen Testcases fehl, so handelt es sich um einen generellen Implementierungsfehler in dem getesteten System.

4 Evaluierung von TLS-Bibliotheken

In einer ersten Evaluation [6] des kombinatorischen Testansatzes für RFC-Anforderungen konnte TLS-Anvil bereits erfolgreich eingesetzt werden. Die automatisierte Analyse von 13 bekannten quelloffenen TLS-Bibliotheken zeigte dabei neben einer Vielzahl von unkritischen Abweichungen vom Standard auch Interoperabilitätsprobleme und Sicherheitslücken auf. Für die kommerziell eingesetzte Bibliothek MatrixSSL konnte beispielsweise gezeigt werden, dass für bestimmte Cipher-Suites eine Padding Oracle Schwachstelle [15] ausgelöst werden kann, über die die Vertraulichkeit der TLS-Verbindung gebrochen werden kann. Außerdem konnten durch manipulierte Längenfelder einzelner Nachrichten Parsingfehler ausgelöst werden, die die CPU-Auslastung deutlich erhöhten und so für Denial-of-Service Angriffe genutzt werden könnten. Die Evaluation bestätigte außerdem, dass es gerade für TLS sinnvoll ist, Tests aus den Standards abzuleiten, da diese insbesondere für die neueste Version (TLS 1.3) die Erfahrungen aus Implementierungsfehlern der Vergangenheit widerspiegeln. So konnte bei der Bibliothek wolfSSL festgestellt werden, dass durch das Versenden invalider Zertifikatnachrichten die Authentifikation des Servers vollständig umgangen werden konnte. Dadurch wären Verbindungen eines Clients anfällig für Man-in-the-Middle Angriffe, die alle Sicherheitsziele von TLS brechen. Der RFC enthält genau für diesen Fehlerfall eine Anforderung, die die Detektion solcher invaliden Nachrichten sicherstellen

soll. Diese gravierende Sicherheitslücke hätte daher durch RFC basierte Tests, wie sie in TLS-Anvil implementiert sind, vor der Veröffentlichung der Bibliothek festgestellt werden können.

Als eines von 15 festgestellten Interoperabilitätsproblemen hat sich gezeigt, dass MatrixSSL bei bestimmten Kombinationen aus Cipher-Suite und elliptischen Kurven falsche kryptografische Parameter in TLS 1.3 Verbindungen eingesetzt hat. Für einen Kommunikationspartner wäre dieses Fehlverhalten nicht von einem Angriff auf die Verbindung zu unterscheiden, sodass die Verbindung sofort beendet werden müsste.

Diese Ergebnisse zeigen, dass über RFC-Konformitätstests sicherheitskritische Fehler in TLS-Bibliotheken erkannt werden können. Außerdem hat sich gezeigt, dass kombinatorisches Testen sinnvoll ist, da manche Fehler nur für bestimmte Parameterkombination aufgetreten sind. Daher wäre es sinnvoll, kombinatorische Konformitätstest in den Entwicklungszyklus von TLS-Bibliotheken aufzunehmen, um Schwachstellen und Kompatibilitätsprobleme vor der Veröffentlichung zu vermeiden.

5 Zusammenfassung und Ausblick

Mit TLS-Anvil haben wir eine TLS-Testsuite für Softwarebibliotheken entwickelt, welche im Gegensatz zu bisher verwendeten Ansätzen parametrisiert testen kann, und so auch komplexe Schwachstellen automatisiert aufdeckt. Unsere Testsuite deckt dabei die wichtigsten Anforderungen aus TLS-relevanten RFCs ab und kann diese über einen Blackbox-Ansatz prüfen. Dabei nutzen wir kombinatorische Testalgorithmen, um die Anzahl der Testfälle praktisch klein zu halten, während die Testabdeckung ausreichend groß bleibt.

Die von uns entwickelte Testsuite eignet sich aus diesen Gründen perfekt, um bereits während der Entwicklung von TLS-Produkten auf Fehler zu testen. So kann dafür gesorgt werden, dass Schwachstellen gefunden und beseitigt werden, bevor ein Produkt veröffentlich wird. Um dies zu erreichen, planen wir unsere Testsuite kompatibel mit gängigen Continuous Testing Plattformen zu machen. Wie in Abbildung 2 zu sehen, könnte TLS-Anvil sowohl die Entwicklung von TLS-Bibliotheken als auch den Entwicklungsprozess von Anwendungen, welche TLS nutzen möchten, unterstützen.

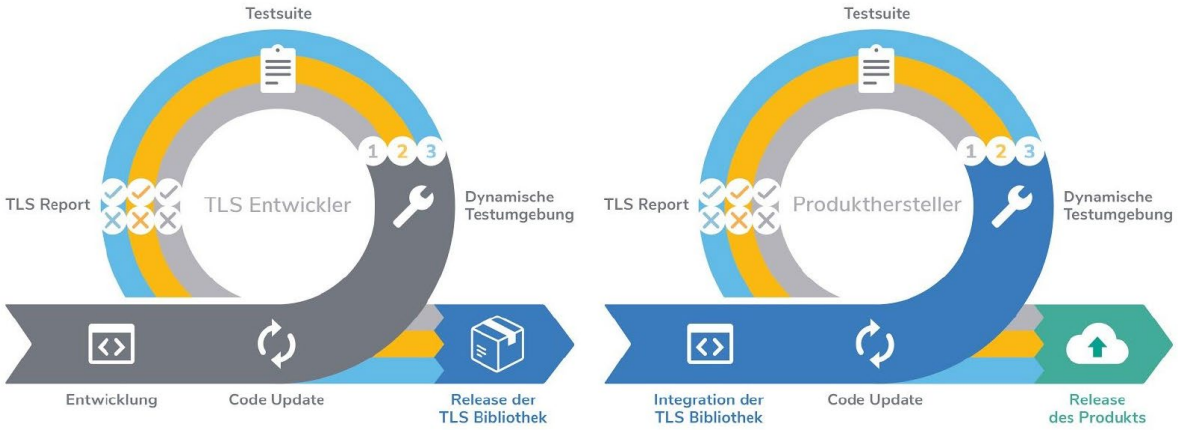


Abbildung 3: Einsatz von TLS-Anvil in der Entwicklung durch Continuous Testing. Links bei der Entwicklung von TLS-Bibliotheken, rechts bei der Verwendung von TLS in Soft- und Hardware-Produkten.

Auch Penetration-Tester und Prüfinstitute können unsere Testsuite nutzen, um TLS-Integrationen auf Konformität und Sicherheit zu testen. Dazu bietet sie unter anderem die Möglichkeit, verschiedene Richtlinien wie BSI-TR-02102-2 [11] oder NIST SP 800-52r2 [12] direkt automatisiert zu prüfen. Wir planen außerdem mittels verschiedener Ansichten in der Benutzeroberfläche, für jeden Anwendungsfall eine zugeschnittene Nutzererfahrung zu bieten. Dabei werden relevante Informationen zuerst gezeigt und Details zusammengefasst. Sicherheitsforscher und Experten können mittels einer Detailansicht alle aufgezeichneten Informationen durchsuchen.

Durch TLS-Anvil haben wir ein Framework zum kombinatorischen Testen und Auswerten von Netzwerkprotokollen und Sicherheitsstandards entwickelt, welches auch für andere Zwecke verwendet werden kann. So planen wir in Zukunft auch X.509-Zertifikate auf Konformität und Sicherheit zu testen. Dies kann getrennt oder in Kombination mit TLS-Tests stattfinden.

Um unsere Testsuite weiterzuentwickeln und an die Wünsche und Bedürfnisse von potenziellen Nutzern anpassen zu können, sind wir auf der Suche nach weiteren Projektpartnern. Interessierte Firmen können uns mit ihren Erfahrungen und Fachkenntnissen dabei helfen, die Bedürfnisse der Nutzerbasis und Zielgruppe zu erkennen und somit das Projekt für Endnutzer signifikant zu verbessern. Gleichzeitig profitieren Sie von den Ergebnissen, Innovationen und Technologien des KoTeBi-Projektes. Bei Interesse schreiben Sie uns gern eine Mail oder besuchen Sie unsere Webseite unter www.kotebi.de.

6 Danksagungen

Das Projekt KoTeBi wird gefördert durch das Bundesministerium für Bildung und Forschung. Als assoziierte Partner unterstützen uns die TÜV Informationstechnik GmbH, sowie die Utimaco IS GmbH. Weitere Beiträge am Projekt haben geleistet: Fabian Albert, Philipp Nieting, Henrik Schäfer, Karsten Meyer zu Selhausen, Jonas Thiele und David Ziemann.

Vielen Dank für die Zusammenarbeit und Unterstützung.

Literaturverzeichnis

[1] Durumeric, Z., Li, F., Kasten, J., Amann, J., Beekman, J., Payer, M., ... & Halderman, J. A. (2014, November). The matter of heartbleed. In Proceedings of the 2014 conference on internet measurement conference (pp. 475-488).

[2] Möller, B., Duong, T., & Kotowicz, K. (2014). This POODLE bites: exploiting the SSL 3.0 fallback. Security Advisory, 21, 34-58.

[3] Böck, H., Somorovsky, J., & Young, C. (2018). Return Of Bleichenbacher's Oracle Threat (ROBOT). In 27th USENIX Security Symposium (USENIX Security 18) (pp. 817-849).

[4] Aviram, N., Schinzel, S., Somorovsky, J., Heninger, N., Dankel, M., Steube, J., ... & Shavitt, Y. (2016). DROWN: Breaking TLS Using SSLv2. In 25th USENIX Security Symposium (USENIX Security 16) (pp. 689-706).

[5] https://github.com/tls-attacker/TLS-Attacker, abgerufen am 09.01.2024

[6] Maehren, M., Nieting, P., Hebrok, S. N., Merget, R., Somorovsky, J., & Schwenk, J. (2022). TLS-Anvil: Adapting Combinatorial Testing for TLS Libraries. 31st USENIX Security Symposium (USENIX Security '22).

[7] T. Dierks and E. Rescorla. The Transport Layer Security (TLS), Protocol Version 1.2. RFC 5246 (Proposed Standard), August 2008.

[8] https://github.com/tls-attacker/TLS-Anvil, abgerufen am 15.02.2024

[9] https://coffee4j.github.io/, abgerufen am 15.02.2024

[10] S. Bradner. Key words for use in RFCs to Indicate Requirement Levels (Internet Best Current Practices), March 1997

[11] BSI TR-02102-2. "Kryptographische Verfahren: Verwendung von Transport Layer Security (TLS). Version: 2023-1"

[12] NIST SP 800-52r2, "Guidelines for the Selection, Configuration, and Use of Transport Layer Security (TLS) Implementations: NIST SP 800-52 Rev. 2"

[13] E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446 (Proposed Standard), August 2018

[14] E. Rescorla and N. Modadugu. Datagram Transport Layer Security Version 1.2. RFC 6347 (Proposed Standard), January 2012

[15] Merget, R., Somorovsky, J., Aviram, N., Young, C., Fliegenschmidt, J., Schwenk, J., & Shavitt, Y. (2019). Scalable Scanning and Automatic Classification of TLS Padding Oracle Vulnerabilities. USENIX Security Symposium.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Secure C-ITS Communication – Ein Beitrag zur Absicherung der digitalen Verkehrsinfrastruktur

Maximilian Wahner[1], Sandro Berndt-Tolzmann[2], Markus Wagner[1]

Kurzfassung:
Kooperative Intelligente Transport Systeme (C-ITS) ermöglichen die direkte Vernetzung von Fahrzeugen und Verkehrsinfrastruktur. Ziel ist die Steigerung der Verkehrssicherheit und des Verkehrsflusses durch Dienste wie Warnungen vor Baustellen oder Stauenden. Die IT-Sicherheit dieser Systeme ist unerlässlich, auch da sie zukünftig direkte und sicherheitsrelevante Entscheidungen automatisierter Systeme potenziell beeinflussen werden. Dieser Beitrag beleuchtet die Entwicklung eines gemeinsamen europäischen IT-Sicherheitsrahmens für C-ITS sowie die Schaffung eines Common Criteria Schutzprofils, das als Blaupause für die Sicherheit der kooperativen Straßeninfrastruktur dient.

Stichworte: Ampeln, Automotive, Baustellen, C-ITS, Common Criteria, Intelligente Verkehrssysteme, IVS, Smart City, Straßenverkehr, Verkehrssysteme

1 Einleitung und Kontext – Kooperative Intelligente Verkehrssysteme

1.1 C-ITS (Cooperative Intelligent Transport Systems)

Unter kooperativen intelligenten Verkehrssystemen (englisch „cooperative intelligent transport systems“ – C-ITS) versteht man die direkte Vernetzung von Fahrzeugen untereinander und zwischen Fahrzeugen und Verkehrsinfrastruktur. Die so ermöglichten Dienste sollen die Verkehrssicherheit erhöhen und den Verkehrsfluss verbessern. Beispielsweise kommen Dienste wie die Warnungen vor Tagesbaustellen, die derzeit von der Autobahn des Bundes GmbH bundesweit ausgerollt werden, und Warnungen vor Stauenden in Betracht, aber auch (vorwiegend urbane) Anwendungen im Kontext von Lichtsignalanlagen (ÖPNV-Priorisierung, Vorrangschaltung für Einsatzfahrzeuge) sowie viele weitere.

Erste Serienfahrzeuge unterstützen die in Europa verwendete Funkkommunikation „ETSI ITS G5“ auf Basis des WLAN-Protokolls IEEE 802.11p bereits seit 2019, die entsprechenden Infrastrukturkomponenten werden seit 2022 ausgerollt [1].

Für die verschiedenartigen C-ITS Komponenten werden, je nach Einsatzgebiet bzw. Charakteristik, oft Begriffe wie „R-ITS-S“ (Roadside ITS Station,

[1] TÜV Informationstechnik GmbH (TÜVIT), Essen
[2] Bundesanstalt für Straßenwesen (BASt), Bergisch Gladbach

Verkehrsinfrastruktur) oder „V-ITS-S“ (Vehicle ITS Station, Fahrzeugseite) verwendet.

Die Absicherung der Nachrichten, die zwischen den C-ITS Stationen ausgetauscht werden, erfolgt Signatur-basiert, unter Nutzung von speziellen Zertifikaten, welche auf Basis asymmetrischer Kryptografie und elliptischen Kurven arbeiten. Dies setzt die Schaffung eines europäischen Vertrauensraumes in Form einer dezidierten Public Key Infrastructure voraus.

1.2 Der (aktuelle) europäische Rechtsrahmen

Kooperative Systeme einzuführen ist aus den vorgenannten Gründen auf nationaler und internationaler Ebene ein erklärtes politisches Ziel. C-ITS zählt zu den Intelligenten Verkehrssystemen (IVS), den maßgeblichen Rechtsrahmen bildet seit 2010 die europäische IVS-Richtlinie, welche in Deutschland durch das IVS-Gesetz umgesetzt wurde [2,3]. Erst Ende 2023 verabschiedete man auf europäischer Ebene die Fortschreibung der IVS-Richtlinie [4]. Die in Artikel 2 definierten „vorrangigen Bereiche“ wurden aktualisiert, C-ITS gehört zum Bereich IV „IVS-Dienste für kooperative, vernetzte und automatisierte Mobilität“. Die IVS-Richtline konkretisiert nun erstmals auch den Begriff C-ITS, was in der zuvor gültigen Fassung noch nicht gegeben war. So versteht man unter C-ITS gemäß Artikel 4 (e) lit. 14 Systeme „*die es IVS-Nutzern ermöglichen, durch den Austausch gesicherter und vertrauenswürdiger Nachrichten zu interagieren und zu kooperieren, ohne einander bereits zu kennen*“. Die Sicherheit der Nachrichten in diesen ad-hoc gebildeten Netzwerken ist somit essenzieller Bestandteil der C-ITS-Definition.

Die IVS-Richtlinie beinhaltet, sowohl in alter als auch neuer Fassung, die Kompetenzen, auf europäischer Ebene weitere delegierte Verordnungen für spezifische Bereiche zu erlassen, u.a. für C-ITS.

Im Jahr 2019 legte die Europäische Kommission einen Entwurf einer solchen delegierten Verordnung vor [5], worin Rahmenbedingungen für den Betrieb von C-ITS-Diensten in Europa geschaffen werden sollten. Die inhaltlichen Vorarbeiten waren in den Jahren 2014 bis 2017 in entsprechenden Expertengruppen unter dem Dach der Europäischen Kommission, unter Einbeziehung aller relevanten Stakeholdergruppen, erarbeitet worden. Der Entwurf der delegierten Verordnung umfasste in den Anhängen neben den Profilen, die für die Spezifikation der Nachrichteninhalte erforderlich sind, auch Vorgaben zur IT-Sicherheit [5]. Nach einer Vielzahl politischer Debatten in Parlament und Rat wurde der Entwurf der Europäischen Kommission jedoch letztlich nicht angenommen, die rechtliche Rahmensetzung erfolgte nicht wie angestrebt.

Für die Umsetzung der aktualisierten IVS-Richtline wird es bis spätestens 21.12.2024 auch ein Arbeitsprogramm der Europäischen Kommission geben. Darin sind beabsichtige Aktivitäten, Spezifikationen und Rechtsakte darzulegen sowie Zeiträume für die Realisierung anzugeben. In aktuellen Entwürfen des Arbeitsprogrammes steht C-ITS bereits für 2024/2025 auf der Agenda, was auch die Möglichkeit eines erneuten Anlaufs hin zu einer delegierten Verordnung im Bereich C-ITS umfasst.

1.3 Entwicklung gemeinsamer IT-Sicherheitsvorgaben (Policies) für C-ITS

Trotz Ablehnung des Entwurfes der delegierten Verordnung in 2019 dienten die von allen Stakeholdern gemeinsam erarbeiteten Anhänge auch weiterhin als Arbeitsgrundlage. Verschiedene Akteure bekannten sich bereits 2019 zur gemeinsamen Anwendung der Rahmenbedingungen aus dem Entwurf der delegierten Verordnung, insbesondere der IT-Sicherheitsanforderungen [6]. Dabei wurden sie seit 2020 auch durch die Europäische Kommission durch die Schaffung einer dedizierten Expertengruppe unterstützt [7]. Die in diesem Rahmen fortgeschriebenen Vorgaben bezüglich der IT-Sicherheit, die ursprünglich als Anhänge zum Entwurf der delegierten Verordnung entstanden waren, unterteilen sich in zwei spezifische Dokumente [8]:

- Die „*Certificate Policy*“ (CP), die die Vorgaben für den Betrieb der verteilten Public Key Infrastructure (PKI) beinhaltet, regelt Details zum Bezug von digitalen Zertifikaten. Die CP fordert auch die Verwendung eines nach Common Criteria zertifizierten kryptografischen Moduls, um eine sichere Handhabung des privaten Schlüsselmaterials sicherzustellen.
- Die „*Security Policy*“ (SP) gibt Anforderungen an die Betreiber von C-ITS-Stationen und deren Prozesse vor (z.B. Betrieb im Rahmen eines zertifizierten ISMS) und macht darüber hinaus Vorgaben bezüglich der Sicherheit der Stationen selbst. Hier ist die Anforderung zu finden, dass für den Regelbetrieb ausschließlich nach Common Criteria zertifizierte C-ITS-Stationen zum Einsatz kommen [9].

Zusätzlich zu diesen europäischen Regelwerken gibt es mit den beiden Teilen der TR-03164 bereits konkretisierte Umsetzungsvorgaben des BSI, die auf den europäischen Policies aufsetzen [10].

2 C-ITS-Dienste in Europa – harmonisiert und sicher

2.1 Die C-Roads-Projektfamilie und die C-Roads-Plattform

Unter dem Namen „C-Roads“ agieren seit 2016 eine Vielzahl europäischer Straßenbetreiber, um die infrastrukturbezogenen Dienste zu harmonisieren

und die Einführung der Technologie bestmöglich voranzubringen [11]. Dafür arbeiten alle nationalen C-Roads-(Förder-)Projekte im Rahmen der C-Roads-Plattform in entsprechenden Arbeitsgruppen zusammen und koordinieren sich mit anderen Akteuren wie der Europäischen Kommission, Fahrzeugherstellern und Anbietern von Verkehrsdiensten. Das Ziel ist es, einheitliche Dienstespezifikationen und Kommunikationsprofile zu erstellen, sodass Fahrzeuge und Infrastruktur in Europa von Finnland bis Portugal über alle Hersteller hinweg interoperabel miteinander kommunizieren können.

Um die C-ITS-Aktivitäten finanziell zu unterstützen und auch organisatorisch gut aufzustellen, fördert die Europäische Kommission die C-Roads-Plattform im Rahmen des CEF-Programmes (Connecting Europe Facility) im Zuge verschiedener Förderaufrufe [12]. So wurde C-Roads sowohl vom geografischen Umfang als auch vom inhaltlichen Zuschnitt immer mehr erweitert, von zunächst 8 auf mittlerweile 18 Mitgliedsstaaten, die von Autobahnanwendungen hin zu urbanen und interurbanen Diensten und Szenarien ein breites Spektrum abdecken. Auch Deutschland beteiligte sich von Beginn an mit mehreren Pilotregionen [13, 14].

2.2 Sichere C-ITS-Straßeninfrastruktur

Über die inhaltlichen Abstimmungen der Straßenbetreiber hinaus spielen auch organisatorische, rechtliche und andere nicht-funktionale, querschnittliche Aspekte bei der Einführung von C-ITS in Europa eine Rolle. So hat die C-Roads-Plattform, neben einer Arbeitsgruppe zu organisatorischen Fragestellungen, die IT-Sicherheit seit ihrem Start in einer dedizierten Task Force im Arbeitsprogramm. Neben den technischen Anforderungen und Spezifikationen, z.B. Details der Berechtigungen innerhalb der spezifischen C-ITS-Zertifikate oder bezüglich der erforderlichen Zertifikatsprüfung der C-ITS-Kommunikation, ist es diese Gruppe, die auch die Schaffung eines gemeinsamen Mindestniveaus an IT-Sicherheitsanforderungen sicherstellen soll. Unter Einbeziehung einer Vielzahl an nationalen Experten und unter Berücksichtigung vieler unterschiedlicher Architekturen von Verkehrssystemen wurde der Input gesammelt, der die Grundlage für die Erstellung eines Schutzprofils bilden sollte, auf welches in Kapitel 3 näher eingegangen wird.

Darüber hinaus ist diese Gruppe auch ein wichtiges Bindeglied zu den Diskussionen der europäischen Policies mit anderen relevanten Stakeholdern im Rahmen der europäischen Expertengruppe.

2.3 (Zukünftige) Anwendungen und Möglichkeiten

Kooperative Straßeninfrastruktur kommt in unterschiedlichen Ausprägungen und Kontexten zum Einsatz. Darunter fallen zudem vielfältige Anlagenkategorien, wie man der nachfolgenden beispielhaften Auswahl entnehmen kann:

- Lichtsignalanlagen (besser bekannt als Ampeln), die ihren Schaltzustand kommunizieren können und/oder in der Lage sind, Priorisierungsanfragen berechtigter Fahrzeuge zu berücksichtigen: Dies umfasst Einsatzfahrten von Krankenwagen und Polizei, aber auch Bevorrechtigungen von Bussen und (Straßen-)Bahnen. Insbesondere für letztere ist die Erneuerung der bereits seit 40 Jahren in Betrieb befindlichen Analogfunk-Technologie zwingend erforderlich, zum einen aus IT-Sicherheitserwägungen [15], aber auch weil der dafür genutzte Frequenzbereich umgewidmet wird und somit nicht mehr für die ÖPNV-Priorisierung zur Verfügung stehen wird [16].
- Schilderbrücken, die zumeist auf Autobahnen anzutreffen sind und die sogenannte Wechselverkehrszeichenanlagen oder andere Beschilderungen tragen: Die Art der übermittelten Informationen deckt eine große Bandbreite ab, von dynamischen Geschwindigkeitsbegrenzungen, Spursperrungen, temporären Seitenstreifenfreigaben (insbesondere im Berufs- oder Veranstaltungsverkehr wie Messen, Konzerten oder bei Fußballspielen) bis hin zu intelligentem und bedarfsgerechtem Routing und Auflagen bzw. Begrenzungen für Schwerlastverkehre.
- Auch der eingangs in Abschnitt 1.1 bereits erwähnte Baustellenwarner, d.h. die Ausstattung der fahrbaren Absperrtafeln, die zur Absicherung von (Tages-)Baustellen, Mäh- und Markierungsarbeiten sowie Unfällen eingesetzt werden, fällt in den Anwendungsbereich dieser Gerätekategorie.

In technischer Hinsicht bedeutet dies, dass die R-ITS-S C-ITS[3]-Nachrichten im Kontext der Kommunikation von Infrastruktur zu Fahrzeugen (Infrastructure-to-Vehicle (I2V)) und Fahrzeugen zu Infrastruktur (Vehicle-to-Infrastructure (V2I)) mit anderen C-ITS-Stationen austauscht. Bei den zu versendenden Nachrichten handelt es sich um Ereignisse, Warnungen und Informationen, die den Straßenverkehr betreffen. Ein Beispiel für einen Service

[3] Je nach Standard findet man den Begriff „ITS Nachricht“ oder „C-ITS Nachricht“. In den für das Schutzprofil relevanten Policy-Dokumenten wird der Begriff C-ITS Nachricht verwendet und daher auch in diesem Paper.

in der I2V-Kommunikation ist die rechtzeitige Information von Verkehrsteilnehmern (Fahrzeugen / V-ITS-S) innerhalb der Kommunikationsreichweite der R-ITS-S über die aktuelle Verkehrssituation. Die zu versendenden Informationen können entweder direkt von der R-ITS-S oder aber auch aus einer Verkehrsrechenzentrale (englisch Traffic Control Center (TCC)) im Hintergrund ausgelöst werden. Vereinfacht betrachtet kann man bei der Kommunikation von der R-ITS-S zu anderen V-ITS-S von einem Komplement bzw. einer Digitalisierung bis hin zur Virtualisierung der physischen Straßenschilder und Lichtsignalanlagen sprechen. Zusätzlich ist es der R-ITS-S möglich die Daten der von Fahrzeugen (V-ITS-S) empfangenen C-ITS-Nachrichten zu aggregieren, um Informationen für ein verbessertes Verkehrsmanagement bereitzustellen. Diese Verarbeitung kann rein lokal oder auch teilweise bzw. vollständig im TCC erfolgen und/oder in einem anderen Dienst des Straßenbetreibers genutzt und eventuell von anderen Dienstleistern weiterverwendet werden. Die verbesserte Datengrundlage ermöglicht es den Straßenbetreibern wiederum, neue Dienste zu entwickeln und bestehende Dienste in Sachen Genauigkeit und Aktualität zu verbessern.

Über die reinen Vernetzungsaspekte hinaus, die heute primär dem Verkehrsmanagement und der Information und Warnung von Fahrzeugführenden dienen, wird die Automatisierung für zukünftige Mobilitätsformen immer mehr an Bedeutung gewinnen. Auf europäischer Ebene wird diese Weiterentwicklung unter dem Begriff CCAM (Cooperative, Connected, Automated Mobility) geführt [17]. So sind beispielsweise direkte und z.T. sicherheitsrelevante Entscheidungen durch automatisierte Systeme vorstellbar. Die Vernetzung der Verkehrsteilnehmer untereinander und mit der Verkehrsinfrastruktur trägt auch als zusätzlicher Sensor in Fahrzeugen zur erhöhten Sicherheit bei, denn sie kann im Gegensatz zu bisher gängigen Radar- und Kamerasystemen auch „um die Ecke“ schauen und somit in bisher schwer adressierbaren Gefahrensituationen dabei helfen, Unfälle zu vermeiden. Ohne nachgewiesene IT-Sicherheitsmaßnahmen wird es jedoch nur schwerlich Vertrauen in die zugrundeliegenden Informationen und Schnittstellen geben können. IT-Sicherheit wird somit zur zwingenden Voraussetzung, um die skizzierten Vorteile zukünftig überhaupt realisieren zu können.

3 IT Security – Herausforderungen im Kontext von C-ITS

3.1 Zielvorgaben organisatorischer und technischer Natur

Durch die Vernetzung verschiedener C-ITS-Komponenten untereinander sowie mit dem Internet eröffnen sich zahlreiche neue Möglichkeiten und innovative Dienste. Dabei ist es von entscheidender Bedeutung, sicherzustellen,

dass die IT-Sicherheit einen höchstmöglichen Stellenwert einnimmt. In diesem Zusammenhang spielen die in Abschnitt 1.3 erläuterten Regeln der Security Policy und der Certificate Policy eine tragende Rolle [8]. Diese Richtlinien stellen verbindliche Vorgaben dar, welche sich in sich grob in zwei Bereiche unterteilen lassen:

- **Organisatorische Maßnahmen zum Schutz der Prozesse beim Betreiben von C-ITS-Stationen**
  Die Umsetzung der organisatorischen Maßnahmen umfasst das Betreiben eines Informationssicherheitsmanagementsystems gemäß ISO-27001 [18]. Alternativ kann diese Anforderung auch durch Betreiben eines ISMS gemäß einem nationalen Standard wie dem BSI-Grundschutz [19] erfüllt werden bzw. den jeweiligen nationalen Umsetzungen der europäischen NIS- und NIS2-Richtlinie [20, 21]. Für Automobilhersteller, die Betreiber von C-ITS-Stationen in Fahrzeugen sind (V-ITS-S), ist es zudem möglich ein Cyber Security Management System (CSMS) zu betreiben und damit die Konformitätsanforderungen zu erfüllen [22, 23], solange die gesamten Systeme sowie die zugehörige Backend-Infrastruktur dadurch abgedeckt sind.
- **Anforderungen an den Schutz der Komponenten wie beispielsweise Sicherheitsmodule und Stationen**
  Der Schutz der Komponenten muss durch ein standardisiertes IT-Sicherheitsprüfungsverfahren bzw. -evaluierungsverfahren sichergestellt werden. Dies wird durch die Forderung nach einer Common Criteria (CC) Zertifizierung konkretisiert, die sich für die C-ITS-Stationen in der Security Policy findet, sowie die Forderung nach der Verwendung zertifizierter Sicherheitsmodule in der Certificate Policy, die sowohl für die C-ITS-Stationen als auch für alle Entitäten im europäischen PKI-Verbund gelten [8, 9]. Um möglichst einheitliche (minimale) Sicherheitsanforderungen gewährleisten zu können, besteht die Möglichkeit, Schutzprofile (englisch Protection Profile) für Produktgruppen zu definieren. Diese dienen als systematische Vorlagen für Entwickler und Prüfer, da sie erforderliche Sicherheitsmechanismen sowie Methoden zur Validierung dieser Mechanismen vorgeben. Die Definition der Minimalanforderungen ermöglicht auch eine gewisse Vergleichbarkeit der angebotenen Produkte unterschiedlicher Hersteller.

Insbesondere mit Blick auf die Anforderungen, die den Schutz von Komponenten betreffen, und die maßgeblichen Randbedingungen, erläutert dieser Beitrag die Entwicklungsschritte hin zu einem Common Criteria Schutzprofil

für kooperative Verkehrsinfrastruktur, welches im Folgenden genauer vorgestellt wird.

3.2 Erste Schritte – Schaffung der Grundlagen

Der erste Schritt hin zu zertifizierten Geräten im Bereich C-ITS erfolgte bereits im Jahr 2019 mit der Zertifizierung des Schutzprofils für Baustellenwarner nach Common Criteria [24]. Aufgrund verschiedener Faktoren wurde bisher allerdings noch kein Produkt mit dem Schutzprofil zertifiziert:

- Das Schutzprofil wurde speziell für den deutschen Kontext entwickelt, zum damaligen Zeitpunkt noch unabhängig von den parallel im europäischen Rahmen im Entstehen befindlichen Policies.
- Es gab nie eine (beispielsweise gesetzliche) Verpflichtung, Produkte nach diesem Schutzprofil zu zertifizieren.
- Einige der in diesem Schutzprofil herangezogenen Standards und Referenzen wurden mehrfach und zu unterschiedlichen Zeitpunkten aktualisiert, was erhebliche Nacharbeiten am Schutzprofil erforderlich gemacht hätte.

Dennoch konnten durch die Entwicklung dieses Schutzprofils wichtige Erkenntnisse und Erfahrungen gewonnen werden, die bei der Definition von Schutzmechanismen im europäischen Kontext hilfreich waren.

Im Rahmen der (Weiter-)Entwicklung der europäischen Richtlinien in der Expertengruppe [7, 8] wurde das Schutzprofil für Baustellenwarner als Input für ein Übergangsverfahren herangezogen.

Ende 2021 wurde eine erste Version dieses Übergangsverfahren als Anhang im sog. CPOC-Protokoll veröffentlicht und umfasst drei verschiedene Stufen, auch Level genannt [8]:

- **Level 0:** Testphase für C-ITS-Stationen während der Entwicklung und Erprobung: Abstriche gegenüber allen Policy-Anforderungen sind möglich, z.B. für Projekte und Piloten noch (weit) vor dem regulären Betrieb.
- **Level 1:** Betrieb der Stationen in einer ersten, vom Umfang noch reduzierten Produktivumgebung: Es existiert nur ein klar definiertes Delta gegenüber den Policy-Anforderungen, z.B. die Zertifizierung nach Common Criteria betreffend, dies ermöglicht einen Start in den Regelbetrieb in Abgrenzung von Tests und Piloten sowie die Migration hin zu einem vollem Regelbetrieb.

- **Level 2:** Uneingeschränkter Regelbetrieb der C-ITS Stationen: Volle Policy-Konformität und die ausschließliche Verwendung von zertifizierten Komponenten, sowohl auf Seite der Stationen als auch bei der Nutzung des europäischen PKI-Verbundes.

Für jedes Level wird ein eigenes verteiltes PKI-System betrieben, inklusive der entsprechenden zentralen Instanzen und der erforderlichen TLM-Zertifikate, sowie der ECTLs und RCA-Zertifikate[4].

In Abbildung 1 ist der zeitliche Ablauf schematisch abgebildet:

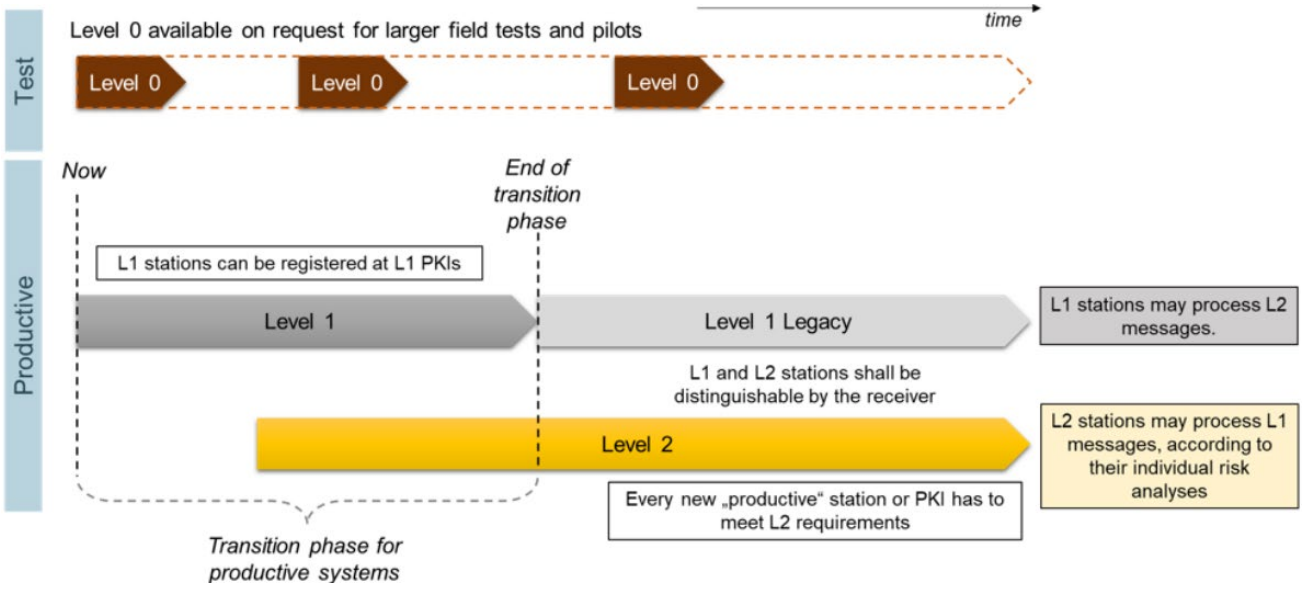


Abbildung 1: Überblick und Zeitlinie der verschiedenen Level (Quelle: [8], Figure 8)

Bei der Einführung von C-ITS-Diensten und Systemen werden zunächst vorbereitende und im weiteren Verlauf begleitende Tests in einer Level 0 Umgebung durchgeführt. Nach erfolgreichem Abschluss dieser Tests und vorausgesetzt, dass die Level 1 Prüfung bzw. später die für Level 2 erforderliche Zertifizierung erfolgreich war, können die C-ITS-Stationen dann in den Level 1 bzw. Level 2 überführt werden. Level 1 fungiert dabei lediglich als zeitlich befristete Übergangsphase/-umgebung während der Initialisierungsphase der C-ITS-Einführung. Zum Ende dieser Übergangsperiode ist vorgesehen, dass die zunächst Level 1 konformen C-ITS-Stationen auch in das europäische C-ITS-System auf Level 2 integriert werden, sofern sie vollständig alle

[4] Die Rollen des zentralen TLM (Trust List Manager), der durch die Europäische Kommission gestellt wird, der ECTL (European Certificate Trust List) als Basis des gemeinsamen Vertrauensnetzwerkes und der RCAs (Root Certificate Authorities) sind im Detail in den Policy-Dokumenten beschrieben.

Anforderungen erfüllen können. Stationen, die zwar den Anforderungen des Level 1 genügen, aber Level 2 nicht erreichen, können in einer „Legacy"-Umgebung weiter mitgeführt werden, unterliegen jedoch gegebenenfalls funktionalen Einschränkungen, z.B. was sicherheitsrelevante Funktionen betrifft.

Mit der Veröffentlichung des Release 3.0 des CPOC-Protokolls wurde mit Ende 2025 ein Ablauf dieser Übergangsphase festgelegt [8].

Durch diese Level soll eine sichere Einführung von C-ITS-Diensten ermöglicht und ein reibungsloser Übergang zu einem vollständig Policy-konformen EU CCMS[5] gewährleistet werden.

Obwohl es keine formelle Verpflichtung dazu gibt, liegt es zumindest für die „Frontrunner", wie die Autobahn GmbH des Bundes und die österreichische ASFINAG [25, 26], die C-ITS als erste in den Regelbetrieb übernehmen, nahe, dass C-ITS-Stationen in einer sequenziellen Reihenfolge ausgerollt werden.

3.3 Auf dem Weg zur Zertifizierung nach Common Criteria

Für Level 1 wurde als Zwischenschritt ein vereinfachtes Prüfverfahren entwickelt, welches u.a. Anforderungen aus dem Baustellenwarner-Schutzprofil berücksichtigt. Dieses vereinfachte Übergangsverfahren basiert im Grundsatz auf den „Assurance Components" der Common Criteria (CC). Allerdings wurden die Anforderungen reduziert, um den Aufwand beim Einstieg für Produktentwickler von C-ITS-Stationen zu reduzieren. Dies soll eine beschleunigte Implementierung ermöglichen und den Herstellern, die bisher keine CC-Erfahrung haben, den Einstieg in die CC-Welt erleichtern.

Dabei wurden u.a. aus diesem ersten Schutzprofil für Baustellenwarner Sicherheitsziele für die C-ITS-Stationen übernommen, welche im Rahmen des vereinfachten Verfahrens von den zu prüfenden Stationen erfüllt werden müssen. Für diese Prüfung sind 49 Evaluationsaufgaben definiert worden, welche von einer von SO-GIS anerkannten Prüfstelle durchzuführen sind [27]. Als Unterstützung für den Produktentwickler solcher C-ITS-Stationen sind Mindestanforderungen bereitgestellt, welche Informationen als Input für die einzelnen Evaluationstaufgaben benötigt werden. Erste Hersteller haben für einzelne Produkte bzw. Stationen bereits den Prozess für Level 1 erfolgreich durchlaufen.

Ein weiterer Schritt ist ebenfalls im Jahr 2021 durch die Zertifizierung des Schutzprofils des Car 2 Car Communication Consortium für V2X-Sicherheits-

[5] Mit dem EU CCMS (European Union C-ITS Security Credential Management System) wird das gesamte verteilte PKI-System für den Betrieb des C-ITS-Ökosystems in Europa beschrieben

module durch das BSI erfolgt [28, 29]. Solche zertifizierten Sicherheitsmodule sind in jeder C-ITS-Station erforderlich und verpflichtend. Sie übernehmen wichtige Sicherheitsfunktionen wie die Signaturerzeugung der einzelnen C-ITS-Nachrichten und die Absicherung des Schlüsselaustausches bei der Kommunikation mit der PKI. Um diese konkrete Anforderung aus der Certificate Policy umsetzen zu können, wurden bereits entsprechende Produkte zertifiziert. Aktuell werden diese zertifizierten Sicherheitsmodule auch schon in den C-ITS-Stationen verwendet, die nach den Kriterien für Level 1 geprüft sind.

3.4 Ein Schutzprofil als Blaupause für die kooperative Straßeninfrastruktur

Für das Betreiben einer C-ITS-Station in Level 2 bzw. konform zu allen Anforderungen des europäischen PKI-Gesamtsystems, benötigen diese eine Zertifizierung nach Common Criteria. Um die gleichen grundlegenden Sicherheitsmechanismen für verschiedene C-ITS-Stationen zu gewährleisten, wurde innerhalb der C-Roads-Arbeitsgruppe beschlossen, Schutzprofile für verschiedene Arten von C-ITS-Stationen zu entwickeln. Ein Typ dieser Stationen ist die Roadside-ITS-Station (R-ITS-S). Dies umfasst alle in Abschnitt 2.3 erläuterten C-ITS-Stationen, die Aufgaben innerhalb der Straßeninfrastruktur wahrnehmen.

Durch die Zertifizierung des Schutzprofils für Baustellenwarner, siehe Abschnitt 3.2, wurden bereits minimale Sicherheitsfunktionalitäten für genau einen Produkttypen dieser Kategorie der R-ITS-S definiert.

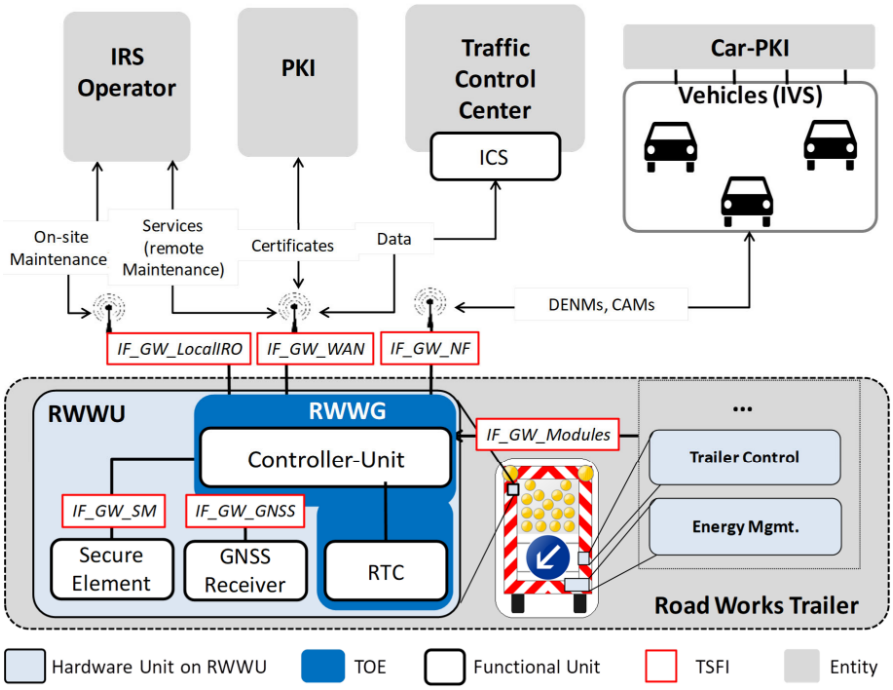


Abbildung 2: Scope des spezifischen Schutzprofils für Baustellenwarner

Dabei deckt das Schutzprofil hauptsächlich den Teil der R-ITS-S ab, welcher für den Versand und Empfang der C-ITS-Nachrichten sowie den Teil, welcher für die sichere Verbindung zu verschiedenen Kommunikationspartnern verantwortlich ist, wie in Abbildung 2 blau dargestellt. Insbesondere ist für die Kommunikation mit dem Administrator und dem TCC eine TLS-Verbindung verpflichtend vorgeschrieben [30]. Dies ist problematisch, da einige europäische Straßenbetreiber für die Absicherung ihrer internen Netze beispielweise auf IPsec-basierte VPN-Netze setzen [31]. An dieser Stelle sind im Schutzprofil somit nicht alle in Frage kommenden kryptografischen Netzwerkprotokolle möglich, ohne gegen die Anforderungen aus dem Schutzprofil zu verstoßen.

Basierend auf den Anforderungen des zuvor zertifizierten Baustellenwarner-Schutzprofils wurden diese Anforderungen generalisiert, überarbeitet und in ein neues Schutzprofil aufgenommen, welches durch den generischeren Ansatz auf verschiedene Einsatzszenarien und Architekturen abgebildet werden kann und muss, wodurch eine breitere Anwendbarkeit angestrebt wird.

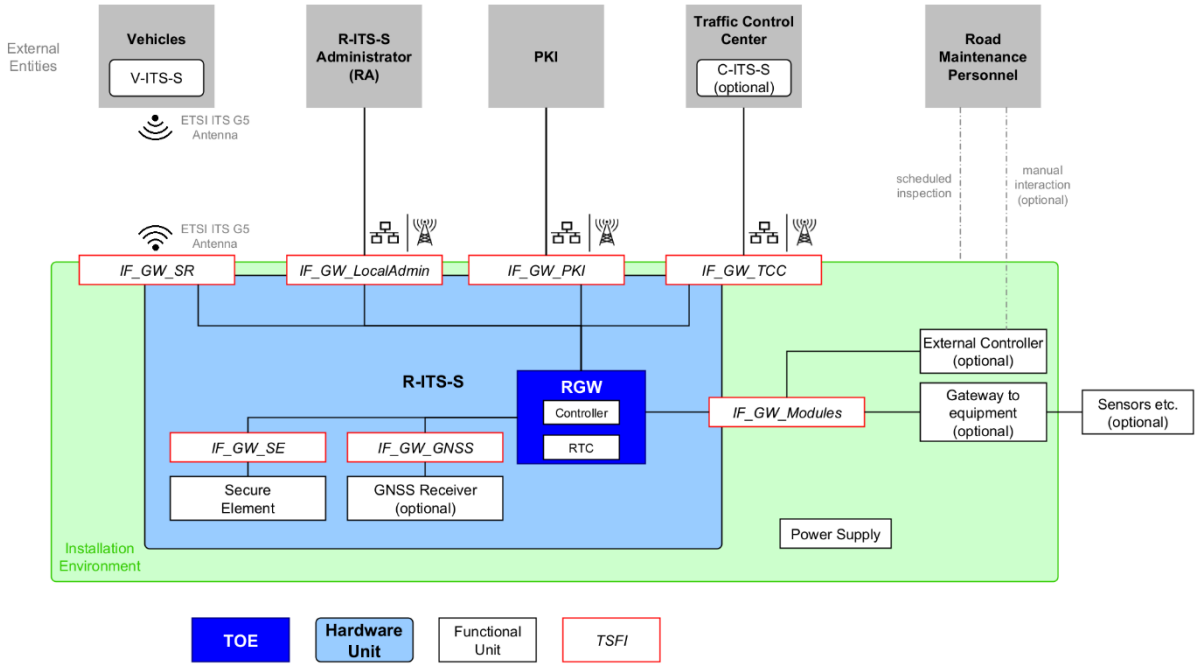


Abbildung 3: Scope des generalisierten Schutzprofils für alle Arten von R-ITS-S

In Abbildung 3 ist der Scope dargestellt, wie bereits zuvor ist nur der Software-Teil der R-ITS-S als Evaluationsziel (im Folgenden Target of Evaluation (TOE) genannt) festgelegt, welcher die nachfolgend definierten Sicherheitsfunktionalitäten umsetzt. Der TOE ist so konzipiert, dass dieser über defi-

nierte logische Schnittstellen mit verschiedenen externen Entitäten kommunizieren kann und umgekehrt. Für jede Interaktion mit einer dieser Entitäten sind Sicherheitsfunktionalitäten im Schutzprofil festgelegt, hier auszugsweise kurz zusammengefasst:

- **Fahrzeuge bzw. V-ITS-S in Reichweite:** Das Überprüfen der Signatur der empfangenen C-ITS-Nachrichten mittels Elliptischer-Kurven-Kryptografie (ECDSA, [30]). Das Signieren der C-ITS-Nachrichten erfolgt im Sicherheitsmodul.
- **Administrator:** Der Administrator hat die Möglichkeit, über eine lokale und/oder eine remote Schnittstelle auf den TOE zuzugreifen. Dabei stehen dem Administrator verschiedene Managementfunktionen zur Verfügung, wie beispielsweise das Auslesen von Logeinträgen, das Starten eines Selbsttestes zum Überprüfen der Integrität, oder das Einspielen eines Firmware-Updates. Diese Verbindung bzw. der Zugriff wird mittels eines kryptografisch sicheren Protokolls geschützt. Das genaue Protokoll wurde im Schutzprofil nicht festgelegt. Mögliche Optionen könnten beispielsweise die Nutzung von TLS, IPsec oder SSH sein [31,32,33].
- **Verkehrszentrale (Traffic Control Center):** Diese Verbindung mit der Verkehrszentrale wird ebenfalls über ein kryptografisch sicheres Protokoll abgesichert. Dieses Protokoll ist wie bei der Verbindung zum Administrator im Schutzprofil nicht festgelegt.
- **PKI:** Diese Verbindung wird über ein standardisiertes Protokoll abgesichert [34], welches auf einen Schlüsselaustausch basierend auf Elliptischer-Kurven-Kryptografie (ECIES, [35,36]) sowie eine symmetrische Verschlüsselung (AES-CCM, [37,38]) der Daten setzt. Dabei wird der symmetrische Schlüssel vom Sicherheitsmodul mittels ECIES verschlüsselt. Die eigentliche symmetrische Ver- und Entschlüsselung mittels AES-CCM wird im TOE umgesetzt.

Zusätzlich zu den Sicherheitsfunktionalitäten, die die Schnittstellen des TOEs direkt schützen, gibt es noch interne Sicherheitsfunktionalitäten, die u.a. zum Selbstschutz des TOEs beitragen. Dazu gehören die Generierung und Speicherung von sicherheitsrelevanten Log-Einträgen, den Rückfall in einen sicheren Zustand bei bestimmten Fehlerzuständen, das sichere Löschen nicht mehr verwendeter kryptografischer Schlüssel, die Validierung von Zertifikaten sowie Zertifikats(sperr)listen, die Generierung von verlässlichen Zeitstempeln mittels der integrierten Echtzeituhr (RTC, englisch Real Time Clock) und eine Nutzerauthentifizierung/-autorisierung.

Darüber hinaus sind auch die sichere Einbindung des zertifizierten Sicherheitsmoduls in den TOE und der Betrieb der Station im Rahmen klar definierter Verantwortlichkeiten und Prozesse sicherzustellen.

Das hier vorgestellte Schutzprofil trägt die Zertifizierungs-ID BSI-CC-PP-0122 [39].

4 Zusammenfassung und Ausblick

Das Schutzprofil für die C-ITS-Komponenten der Straßeninfrastruktur stellt einen wichtigen Baustein für den Rollout sicherer Dienste dar. Diese „Blaupause“ dient den europäischen Straßenbetreibern als wichtiger Teil zukünftiger Ausschreibungsverfahren, um neben rein funktionalen Aspekten auch der IT-Sicherheit der Systeme gebührend Rechnung zu tragen. Neben der öffentlichen Hand erarbeitet parallel auch die Automobilindustrie im Car 2 Car Communication Consortium ein entsprechendes Schutzprofil für C-ITS-Stationen der Fahrzeugseite (V-ITS-S). Mit einer Fertigstellung wird im Laufe der nächsten Monate gerechnet.

Auf das in diesem Beitrag beschriebene Schutzprofil wurde bereits in Ausschreibungen der Autobahn GmbH Bezug genommen, wodurch Hersteller/Anbieter dies für zukünftige Verkehrsinfrastrukturkomponenten berücksichtigen müssen. Die Autobahn in Deutschland ist als Kritische Infrastruktur eingestuft, der Betrieb der C-ITS-Systeme erfolgt demnach im Rahmen eines durch BSI-Grundschutz zertifizierten ISMS.

Auch die österreichische ASFINAG verfolgt den Rollout ihres C-ITS-Systems mit großem Engagement, inkl. Verweis auf das in C-Roads gemeinsam erarbeitete Schutzprofil. Weitere europäische Straßenbetreiber bereiten die Einführung ebenfalls vor, wobei auch für diese die zuvor genannten Anforderungen aus den europäischen Policies gelten, sodass auch diese immer den Betrieb zertifizierter C-ITS-Stationen fordern werden. Sofern neue bzw. sehr spezifische Bedarfe, die durch das hier beschriebene Schutzprofil (noch) nicht abgedeckt sind und auch nicht abgedeckt werden können, auftreten, so könnte dies die Erstellung weiterer Schutzprofile für Infrastrukturkomponenten auslösen. Solange dieser Fall nicht eintritt und/oder das erläuterte Schutzprofil im Zuge von Maintenance-Verfahren und (re-)zertifizierten Aktualisierungen mit allen neuen Anforderungen Schritt halten können (neue Standards, geänderte Policies, weitere Anwendungskontexte), so wird das Schutzprofil für Straßeninfrastrukturkomponenten bzw. R-ITS-S de facto *die* Referenz in diesem Kontext darstellen.

Literaturhinweise

[1] IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments, https://standards.ieee.org/ieee/802.11p/3953/

[2] Richtlinie 2010/40/EU des Europäischen Parlaments und des Rates vom 7. Juli 2010 zum Rahmen für die Einführung intelligenter Verkehrssysteme im Straßenverkehr und für deren Schnittstellen zu anderen Verkehrsträgern (IVS-Richtlinie) https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32010L0040

[3] Gesetz über Intelligente Verkehrssysteme im Straßenverkehr und deren Schnittstellen zu anderen Verkehrsträgern https://www.gesetze-im-internet.de/ivsg/index.html

[4] Richtlinie (EU) 2023/2661 des Europäischen Parlaments und des Rates vom 22. November 2023 zur Änderung der Richtlinie 2010/40/EU zum Rahmen für die Einführung intelligenter Verkehrssysteme im Straßenverkehr und für deren Schnittstellen zu anderen Verkehrsträgern, https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202302661

[5] Commission Delegated Regulation supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the deployment and operational use of cooperative intelligent transport, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=PI_COM%3AC%282019%291789

[6] C-ITS Deployment Group, https://c-its-deployment-group.eu/mission/statements/december-2019-constitutive-act-of-c-its-deployment-group/

[7] Cooperative Intelligent Transport Systems (C-ITS) (E01941/3), https://ec.europa.eu/transparency/expert-groups-register/screen/expert-groups/consult?do=groupDetail.groupDetail&groupID=1941&Lang=EN

[8] "Documentation – C-ITS relevant documents for the deployment within the European Union C-ITS Security Credential Management System (EU CCMS), https://cpoc.jrc.ec.europa.eu/Documentation.html

[9] ISO/IEC 15408 - Information security, cybersecurity and privacy protection — Evaluation criteria for IT security, siehe auch https://commoncriteriaportal.org/cc/

[10] BSI TR-03164 Guidance for Cooperative Intelligent Transport Systems (C-ITS), https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03164/TR-03164_node.html

[11] C-Roads - The Platform of harmonised C-ITS Deployment in Europe, https://www.c-roads.eu/platform.html

[12] Förderprogramm „Connecting Europe Facility", https://transport.ec.europa.eu/transport-themes/infrastructure-and-investment/connecting-europe-facility_en

[13] Ursprüngliche Förderung C-Roads Germany, https://ec.europa.eu/assets/cinea/project_fiches/cef/cef_transport/2015-DE-TM-0431-S.pdf

[14] Gemeinsamer Webauftritt aller deutschen Piloten und Aktivitäten, https://www.c-roads-germany.de/

[15] heise Investigativ: Viele Ampeln sind per Funk einfach manipulierbar, https://www.heise.de/hintergrund/heise-Investigativ-Viele-Ampeln-sind-per-Funk-einfach-manipulierbar-7367885.html

[16] VDV-Mitteilung 4021 Abkündigung analoger Betriebsfunk-frequenzen im 20 kHz-Raster – Migrations-möglichkeit für den analogen LSA-Datenfunk, https://www.beka-verlag.info/VDV-Mitteilungen/Informationstechnik-Informationsverarbeitung/Nachrichtentechnik/VDV-Mitteilung-4021-Abkuendigung-analoger-Betriebsfunk-frequenzen-im-20-kHz-Raster-Migrationsmoeglichkeit-fuer-den-analogen-PDF::1190.html

[17] Cooperative, connected and automated mobility – CCAM, https://transport.ec.europa.eu/transport-themes/intelligent-transport-systems/cooperative-connected-and-automated-mobility-ccam_en

[18] ISO 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements, https://www.iso.org/standard/27001

[19] IT-Grundschutz des Bundesamtes für Sicherheit in der Informationstechnik (BSI), https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/it-grundschutz_node.html

[20] Richtlinie (EU) 2016/1148 des Europäischen Parlaments und des Rates vom 6. Juli 2016 über Maßnahmen zur Gewährleistung eines hohen gemeinsamen Sicherheitsniveaus von Netz- und Informationssystemen in der Union, https://eur-lex.europa.eu/eli/dir/2016/1148/oj

[21] Richtlinie (EU) 2022/2555 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der Union, zur Änderung der Verordnung (EU) Nr. 910/2014 und der Richtlinie (EU) 2018/1972 sowie zur Aufhebung der Richtlinie (EU) 2016/1148 (NIS-2-Richtlinie), https://eur-lex.europa.eu/eli/dir/2022/2555/oj?locale=de

[22] ISO/SAE 21434:2021 Road vehicles – Cybersecurity engineering, https://www.iso.org/standard/70918.html

[23] UN-Regelung Nr. 155 — Einheitliche Bedingungen für die Genehmigung von Fahrzeugen hinsichtlich der Cybersicherheit und des Cybersicherheitsmanagementsystems [2021/387], https://eur-lex.europa.eu/eli/reg/2021/387/oj

[24] Protection Profile for a Roadworks Warning Unit Version 1.1, https://www.bsi.bund.de/SharedDocs/Zertifikate_CC/PP/aktuell/PP_0106.html

[25] Autobahn des Bundes GmbH, Pressemitteilung 29. April 2021, https://www.autobahn.de/fileadmin/user_upload/Pressemitteilung_C_ITS_1.pdf

[26] ASFINAG „C-ITS oder wie Kommunikation zwischen Fahrzeug und Straße die Verkehrssicherheit erhöht?“, https://blog.asfinag.at/innovationsgeist/c-its-oder-wie-kommunikation-zwischen-fahrzeug-und-strasse-die-verkehrssicherheit-erhoeht/

[27] Senior Officials Group – Information Systems Security, https://www.sogis.eu/

[28] Protection Profile V2X Hardware Security Module by CAR 2 CAR Communication Consortium Version 1.0.1, https://www.bsi.bund.de/SharedDocs/Zertifikate_CC/PP/aktuell/PP_0114.html

[29] Car 2 Car Communication Consortium – Protection Profile als Teil des “Basic System Profile”, https://www.car-2-car.org/documents/basic-system-profile/

[30] FIPS 186-4 – Digital Signature Standard (DSS), https://csrc.nist.gov/pubs/fips/186-4/final

[31] RFC 8446 – The Transport Layer Security (TLS) Protocol Version 1.3, https://www.rfc-editor.org/rfc/rfc8446

[32] RFC 4301 – Security Architecture for the Internet Protocol, https://www.rfc-editor.org/rfc/rfc4301

[33] RFC 4253 – The Secure Shell (SSH) Transport Layer Protocol, https://www.rfc-editor.org/rfc/rfc4253

[34] ETSI TS 102 941 - Intelligent Transport Systems (ITS); Security; Trust and Privacy Management, https://www.etsi.org/

[35] IEEE Standard Specifications for Public-Key Cryptography - Amendment 1: Additional Techniques, https://standards.ieee.org/ieee/1363a/2050/

[36] 1609.2b-2019 - IEEE Standard for Wireless Access in Vehicular Environments--Security Services for Applications and Management Messages - Amendment 2--PDU Functional Types and Encryption Key Management, https://ieeexplore.ieee.org/document/8734860, FIPS 197, Advanced Encryption Standard (AES), https://csrc.nist.gov/pubs/fips/197/final

[37] FIPS 197, Advanced Encryption Standard (AES), https://csrc.nist.gov/pubs/fips/197/final

[38] NIST SP 800-38C, Recommendation for Block Cipher Modes of Operation: the CCM Mode for Authentication and Confidentiality, https://csrc.nist.gov/pubs/sp/800/38/c/upd1/final

[39] Schutzprofil "Roadside ITS Station Gateway PP", https://www.bsi.bund.de/SharedDocs/Zertifikate_CC/PP/aktuell/PP_0122.html

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

KI-basierte Lernplattform der nächsten Generation für mehr Cybersicherheit und IT-Awareness: Können adaptive Lernumgebungen und mehr Personalisierung die Lernerfolge und die User Experience im IT-Awareness-Training steigern?

Prof. Dr. Michael Massoth[1]

Kurzfassung: KI-basierte Lernplattformen können die Cybersicherheitskompetenz und IT-Awareness signifikant verbessern. Sie bieten personalisierte, adaptive Lernumgebungen und setzen Gamification ein, um auf individuelle Lernbedürfnisse und Spielertypen einzugehen. Dadurch können die Motivation, das Engagement, die User Experience und der Lernerfolg der Nutzer signifikant erhöht werden.

Stichworte: Adaptive Lernumgebungen, Deepfake-Erkennung, Gamification, KI-basierte Lernplattform, KI-Chatbots, Personalisierung, postquantensichere Verschlüsselung

1 Bedeutung von Cybersicherheit und IT-Awareness

In der digitalen Ära ist Cybersicherheit und IT-Awareness essenziell. Dies wird verstärkt durch geopolitische Spannungen und die fortschreitende Digitalisierung. Vorschriften wie das NIST Cybersecurity Framework 2.0 und die NIS-2-Richtlinie der EU betonen die Notwendigkeit für Organisationen, ihre Abwehr gegen Cyberangriffe zu verstärken und Sicherheitsstandards zu vereinheitlichen. Die Integration von regelmäßigen IT-Awareness-Schulungen ist entscheidend für den Schutz ihrer Daten, Informationen und Geschäftsgeheimnisse und die Pflege einer Sicherheitskultur.

2 Der Faktor Mensch in der Cybersicherheit und IT-Awareness

Unsere Zielgruppe sind alle Menschen, Unternehmen und Organisationen mit essenziellen IT-Sicherheitsrisiken. Das konstante Grundbedürfnis unserer Zielgruppe besteht darin, ihre Assets (Daten, Informationen, Geschäftsgeheimnisse) zu schützen und zu verhindern, dass Angreifende ihre Assets ausspähen, manipulieren, kopieren, zerstören oder durch Ransomware unbenutzbar machen können.

[1] Hochschule Darmstadt / Member of European University of Technology (EUT+), and ATHENE – Nationales Forschungszentrum für angewandte Cybersicherheit

Menschliches Fehlverhalten, wie Nachlässigkeit und mangelndes Bewusstsein, bildet eine Hauptgefahr für IT-Sicherheit, wobei bis zu 99% der Cyberattacken auf solche Fehler zurückzuführen sind [1]. Diese Angriffe verlangen oft menschliche Interaktion, was Mitarbeiter zu einem kritischen und unverzichtbaren Faktor der Verteidigung macht. Eine effektive Abwehr gegen Cyberbedrohungen erfordert daher gut geschultes Personal. Umfassende Cybersicherheitsmaßnahmen sind unverzichtbar, um Risiken zu minimieren. Dazu gehört insbesondere die Schulung und Sensibilisierung der Mitarbeiter als essenzieller Abwehrschirm gegen Cyber-Angriffe.

3 Probleme und Mängel traditioneller Lernmethoden in der Cybersicherheitsausbildung und die Notwendigkeit für Innovation

In der heutigen schnelllebigen und technologiegetriebenen Welt stehen IT-Fachkräfte, Unternehmen, Behörden und Organisationen ständig vor neuen und sich entwickelnden Cybersicherheitsbedrohungen. Um diesen Herausforderungen wirksam begegnen zu können, ist eine fundierte Ausbildung und kontinuierliche Weiterbildung im Bereich der Cybersicherheit unerlässlich. Jedoch stoßen traditionelle Lernmethoden, die in der IT-Awareness-Ausbildung eingesetzt werden, zunehmend an ihre Grenzen, was die Notwendigkeit für innovative Bildungslösungen unterstreicht.

3.1 Probleme traditioneller Lernansätze

Die herkömmlichen Lehrmethoden in der Cybersicherheitsausbildung, wie etwa PowerPoint-Schulungen, Lernvideos, standardisierte Tests, Multiple-Choice-Quizze und passiv rezeptive Lernformate, erweisen sich als unzureichend für die effektive Vermittlung des notwendigen Wissens und der Fähigkeiten, die zur Bewältigung aktueller und zukünftiger Sicherheitsherausforderungen benötigt werden. Diese Ansätze berücksichtigen nicht die individuellen Lernstile und Lerngeschwindigkeiten der Lernenden und bieten kaum Möglichkeiten für praktische, interaktive und kontextbezogene Lernerfahrungen. Darüber hinaus mangelt es ihnen an der Flexibilität, sich an die dynamisch ändernden Technologielandschaften und Bedrohungsszenarien anzupassen. Dies führt zu einer Diskrepanz zwischen den Lerninhalten und den realen, praktischen Anforderungen, die an IT- und Cybersicherheitsfachkräfte gestellt werden.

3.2 Mängel und deren Auswirkungen

Diese Mängel traditioneller Lernmethoden haben weitreichende Auswirkungen. Sie beeinträchtigen nicht nur die Effektivität der Cybersicherheitsausbildung, sondern auch die Fähigkeit von Organisationen und Unternehmen,

auf Cybersicherheitsbedrohungen angemessen zu reagieren. Die Kluft zwischen dem, was gelehrt wird, und dem, was in der Praxis benötigt wird, vergrößert das Risiko von Sicherheitsverletzungen, da Mitarbeiter möglicherweise nicht ausreichend darauf vorbereitet sind, mit realen Bedrohungen umzugehen.

3.3 Notwendige Innovationen

Die Überwindung der Probleme und Mängel traditioneller Lernmethoden in der Cybersicherheitsausbildung durch innovative Ansätze ist entscheidend, um die Lücke zwischen theoretischem Wissen und praktischen Fähigkeiten zu schließen. Durch die Implementierung adaptiver, personalisierter und praxisorientierter Lernmethoden mit Gamification-Elementen können wir eine neue Generation von IT- und Cybersicherheitsfachkräften ausbilden, die besser auf die Herausforderungen der modernen Cybersicherheitslandschaft vorbereitet sind.

4 Einführung in KI-basierte Lernplattformen der nächsten Generation und deren Potenzial

4.1 Adaptive Lernumgebungen und Personalisierung

KI-basierte Lernplattformen der nächsten Generation revolutionieren die Cybersicherheits- und IT-Awareness-Ausbildung durch adaptive Lernumgebungen und personalisiertes Lernen, das sich besser auf die individuellen Lernbedürfnisse, -stile und -geschwindigkeiten der Nutzer einstellen kann. Durch personalisiertes Lernen in adaptiven digitalen Lernumgebungen haben die Anwender signifikant bessere Lernerlebnisse (Motivation/UX/Spaßfaktor) und höhere Lernerfolge. Die digitale Lernumgebung kann sowohl auf Makro-Ebene (z.B. Lern-/Spielertyp) als auch auf Mikro-Ebene (Lernaufgaben, Challenges, Gamification-Element) individuell an den Nutzer angepasst werden und ermöglicht so maximale Lernerfolge. Supervised-Machine-Learning-Algorithmen errechnen Vorhersagen (Empfehlungen) über die erfolgreichsten Lernpfade einer bestimmten Nutzergruppe. Diese Empfehlungen gehen anschließend in die Gestaltung der Lernaufgaben, Gamification-Elemente und möglichen Lernpfade ein. Diese Personalisierung führt zu verbesserten Lernergebnissen, erhöhter Motivation und einem tieferen Verständnis des Lehrmaterials. Ihre Benutzerfreundlichkeit und der modulare Aufbau machen sie ideal für vielbeschäftigte Berufstätige, auch für unterwegs.

4.2 Anwendung von interaktiven Mindmaps zur Steigerung der Personalisierung

Die Anwendung von interaktiven Mindmaps in adaptiven Lernumgebungen bietet eine sehr gute Möglichkeit, die Personalisierung des Lernprozesses zu verstärken. Durch die Visualisierung von Wissensstrukturen in einer interaktiven Form ermöglichen Mindmaps den Lernenden, Konzepte und deren Zusammenhänge intuitiv zu erfassen. Interaktive Mindmaps dienen nicht nur als Lernhilfe, sondern auch als Werkzeug zur Selbstreflexion, indem sie den Lernenden ermöglichen, ihr eigenes Wissen zu strukturieren und Lücken zu identifizieren. Außerdem bereiten sie die Nutzer auf die Interaktion mit den KI-Chatbots vor, mit denen sie auf derselben Webseite angezeigt werden. Durch die Mindmaps haben insbesondere Anfänger und Laien viele Stichworte, die sie sich als Fragende zusammen mit dem KI-Lernchatbot erarbeiten können.

4.3 Anwendung von KI-Lernchatbots zur Steigerung der Personalisierung

Die Integration von KI-Lernchatbots in adaptive Lernumgebungen stellt eine innovative Methode zur Steigerung der Personalisierung im Lernprozess dar. Diese Chatbots nutzen künstliche Intelligenz, um in natürlicher Sprache mit den Lernenden zu interagieren, Fragen zu beantworten, Feedback zu geben oder Trainingsgespräche zur Überprüfung und Festigung des Wissens zu führen. Damit fördern die KI-Lernchatbots durch effizientes dialogbasiertes Lernen, niedrigschwellig, mobil, durch modulare Microcontents, und auch mal „für zwischendurch“, ein tieferes Verständnis des Lehrstoffs, erhöhen die Motivation und unterstützen ein effektiveres und engagierteres Lernen.

4.4 KI-Chatbots im Einsatz für mehr Cybersicherheit und IT-Awareness

Die Forschungsgruppe Tihanyi et al. haben 02/2024 in „CyberMetric“ [2] gezeigt, dass generative Large Language Models, wie zum Beispiel GPT-4, GPT-3.5, GEMINI-pro, Falcon-180B und Zephyr-7B, menschliches Expertenwissen in Benchmark-Vergleichstests durchaus übertreffen können. Dennoch ist beim Einsatz von KI-Chatbots in der Praxis auch das Haftungsrecht zu beachten. Fehlentscheidungen und Falschinformationen von KI-Chatbots aufgrund unzureichender Datenqualität oder algorithmischer Fehler können zu Haftungsansprüchen führen. Dies gilt vor allem für schwerwiegende Folgen wie Personenschäden, Schäden an Infrastruktur sowie Ausfälle von Dienstleistungen und Produktionen.

Deshalb nutzt unser Demonstrator einen KI-gestützten Intent-Klassifizierungschatbot, um Nutzerfragen zu verstehen und sie auf vordefinierte Mus-

terantworten abzubilden. Hierfür verwenden wir das KI-Framework PyTorch und neuronale Netze. Der Prozess erfolgt in zwei Schritten: **(1) Intent-Erkennung:** Das Modell erkennt die Intention hinter einer Anfrage durch Analyse von Textmustern und Kontext. **(2) Antwortzuordnung:** Nachdem die Absicht identifiziert wurde, ordnet der Chatbot die Frage einer vordefinierten Musterantwort zu. Diese Antworten sind im Voraus von Experten erstellt worden und decken ein breites Spektrum möglicher Nutzeranfragen ab. Diese Technik ermöglicht es dem Chatbot, komplexe Anfragen effektiv zu bearbeiten, indem er über die Schlüsselwortsuche hinausgeht und den tieferen Sinn hinter einer Frage erfasst. PyTorch bietet durch seine Flexibilität und Anpassungsfähigkeit eine ideale Plattform für die Entwicklung solcher fortschrittlichen KI-Modelle.

5 Gamification und Engagement

5.1 Spielen - ein Grundbedürfnis des Menschen

Das menschliche Bedürfnis zu spielen ist ein tief verwurzelter und integraler Bestandteil unserer Entwicklung und unseres psychischen Wohlbefindens. Die Gründe, warum Menschen spielen, sind vielschichtig und spiegeln eine Kombination aus biologischen, psychischen und sozialen Faktoren wider. Spielen unterstützt Entwicklung und Lernen, soziale Verbindung, emotionale Ausdrucksmöglichkeiten und Stressabbau, mentale Stimulation und Neugier, ein Gefühl der Autonomie und Kontrolle, sowie das Bedürfnis nach Anerkennung, Belohnung und Erfolg. Das Bedürfnis zu spielen ist somit ein fundamentaler Aspekt der menschlichen Natur, der über alle Altersgruppen hinweg zur Erfüllung grundlegender psychischer Bedürfnisse beiträgt.

Als Wissenschaftler im Bereich der Cybersicherheit erkennen wir, dass Spiele nicht nur Unterhaltung bieten, sondern auch wertvolle Werkzeuge für Bildung, soziale Vernetzung und persönliche Entwicklung darstellen können, solange sie sicher und verantwortungsbewusst genutzt werden.

5.2 Anwendung von HEXAD-Spielertypen zur Steigerung der Personalisierung

Die Anwendung der HEXAD-Spielertypen [3] in adaptiven Lernumgebungen nutzt spielbasierte Motivationspsychologie, um die Lernerfahrung zu personalisieren. Durch Anpassung der Lerninhalte und -methoden an diese Spielertypen können KI-basierte Systeme individuell zugeschnittene Lernpfade erstellen, die die intrinsische Motivation und das Engagement der Lernenden erhöhen. Dies führt zu einer verbesserten User Experience und höheren Lernerfolgen, indem es ein tieferes Eintauchen in den Lernstoff ermöglicht und die Interaktion mit dem Material anregt.

Die HEXAD-Spielertypen nach Marczewski und Gamification-Elemente

Marczewskis HEXAD-Typologie identifiziert sechs Haupt-Spielertypen in Gamification: **Achiever** sind leistungsorientiert und suchen Herausforderungen, um sich stetig zu verbessern. **Socialiser** ziehen ihre Motivation aus sozialen Interaktionen und dem Bedürfnis nach Zugehörigkeit. **Philanthropen** agieren altruistisch, mit dem Ziel, anderen zu helfen und einen Unterschied zu machen. **Free Spirits** sind durch das Streben nach Autonomie und Selbstausdruck angetrieben und wollen ihre eigenen Wege gehen. **Player** konzentrieren sich auf extrinsische Belohnungen wie Punkte oder Abzeichen. **Disruptoren** streben nach Veränderung und können sowohl konstruktiv als auch destruktiv agieren, um Neuerungen anzustoßen [3, 4, 5, 6].

Geeignete Gamification-Elemente je Haupt-Spielertyp	
Achiever	Ziele setzen und verfolgen, Fortschrittsanzeigen, Level/Stufen, Bestenlisten/Leaderboards, Statistiken, Zertifikate, Auszeichnungen und Anerkennung.
Socialiser	Team- und Gruppenarbeit, soziale Funktionen und Entdeckungen, sozialer Status, Diskussionsforen, sowie Peer-Feedback.
Philanthropist	Kooperative Herausforderungen, Aufgaben im Team, ein Forum zum Austausch von Erfahrungen/Wissen, Schenken, Wissensspenden und administrative Rollen.
Free Spirits	Nicht-lineare Lernpfade, kreative Herausforderungen, Personalisierungsoptionen und Anpassungsmöglichkeiten, Erkundungsaufgaben und Kreativitäts-Tools
Player	Punktesysteme, Ranglisten, Leaderboards, (virtuelle) Belohnungen, Preise und Abzeichen
Disruptor	Ihm gezielt neue Ideen/Aufgaben vorstellen, testen und evaluieren lassen, Erkundungsaufgaben, regelmäßige Umfragen, Abstimmungen, Feedback und Entwicklungswerkzeuge.

Tabelle 1: Geeignete Gamification-Elemente für jeden HEXAD-Haupt-Spielertyp

Diese Erkenntnisse ermöglichen die Entwicklung ansprechender und motivierender Gamification-Lösungen für ein breites Benutzerspektrum [3].

5.3 Fallstudie: Deepfake-Erkennungsspiel als Best-Practice-Beispiel für effektive Gamification

In unserer Studie wurde ein Deepfake-Erkennungsspiel als Best-Practice-Beispiel für effektive Gamification durchgeführt [7]. Dabei wurde eine Kontrollgruppe direkt zum Spiel geführt, während eine Experimentgruppe vorab eine zusätzliche Online-Schulung absolvierte. Die Ergebnisse zeigten, dass

die vorgeschulte Gruppe beim Erkennen von Deepfakes signifikant sicherer war. Dies unterstreicht die Effektivität von Gamification, kombiniert mit gezielter Bildung, um die Kompetenzen im Umgang mit spezifischen digitalen Herausforderungen wie Deepfakes zu verbessern.

Spielflow: Es gibt zwei unterschiedliche Spielpfade je nach Gruppenzugehörigkeit. Eine Testgruppe mit Vorab-Schulung (N=50), sowie eine Kontrollgruppe ohne Schulung (N=50).

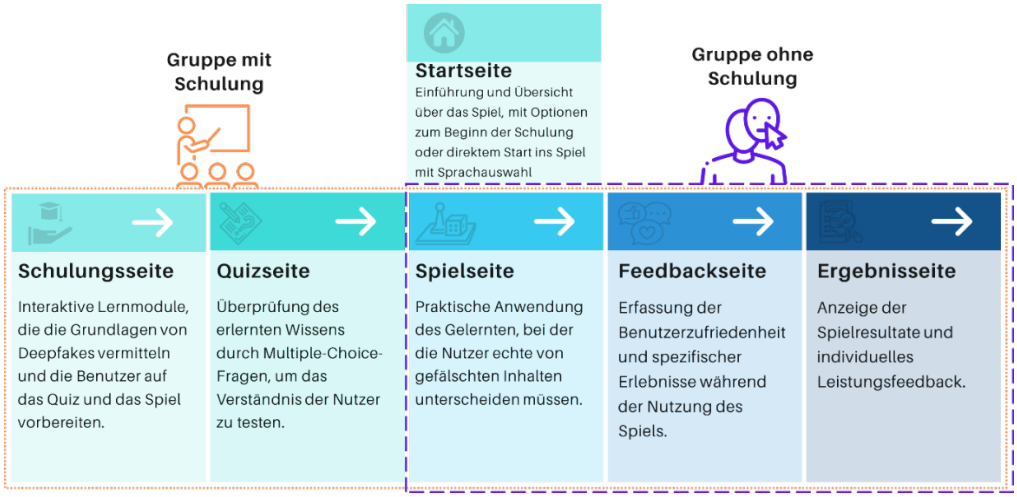


Abbildung 1: Der Spielflow für die Testgruppe mit Vorab-Schulung und für die Kontrollgruppe ohne Schulung [7]

Die Startseite:

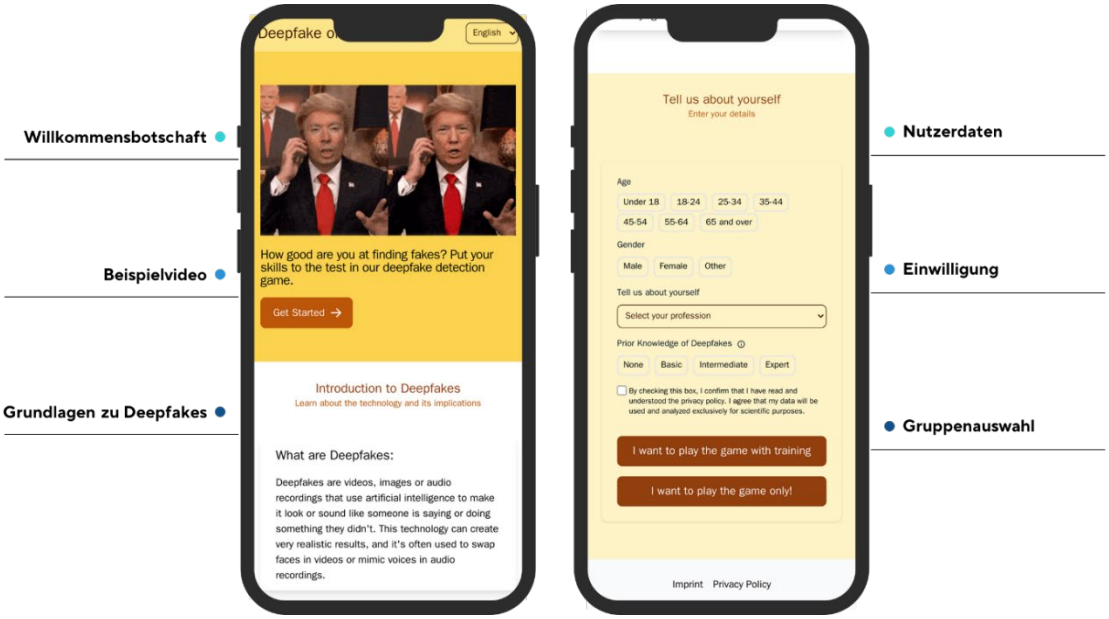


Abbildung 2: Startseite mit Willkommensbotschaft, Beispielvideo, einigen Grundlagen zu Deepfakes, Erhebung der Nutzerdaten, der Einwilligung zur Datenverarbeitung und der Gruppenauswahl (mit/ohne Vorab-Schulung) [7]

Vorab-Schulung mit Training und Quizseite: (Nur 1 Testgruppe, N=50)

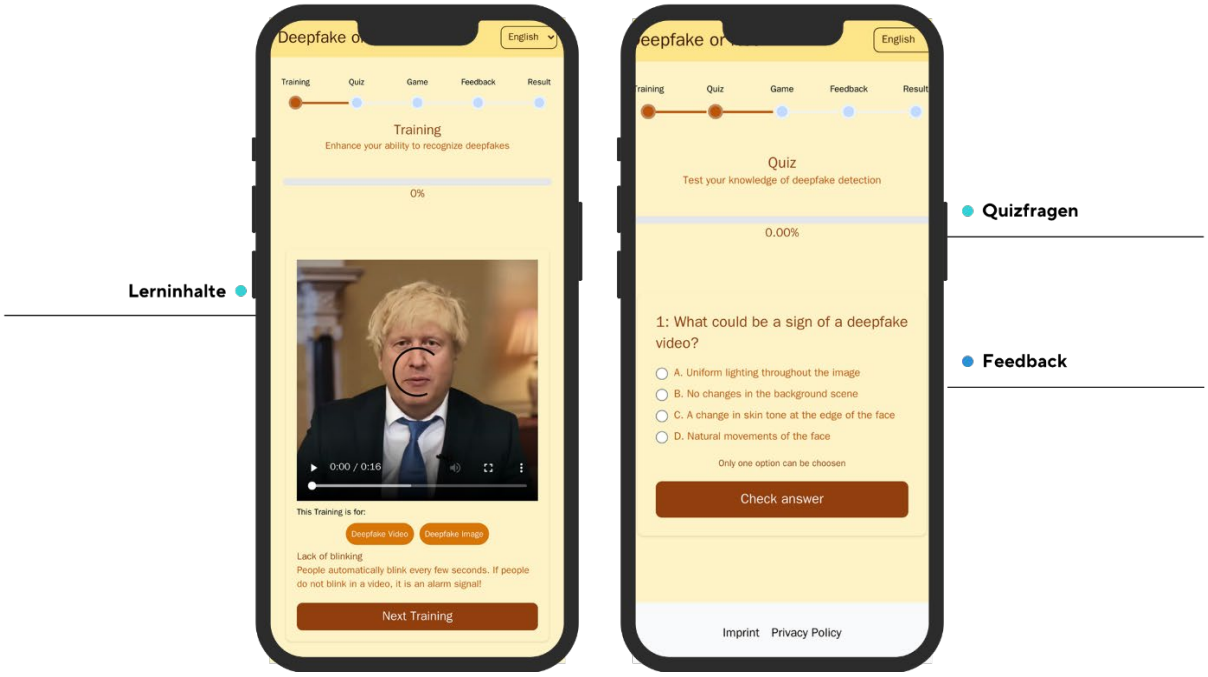


Abbildung 3: Vorab-Schulung mit Lerninhalten zu Audio- und Video-Deepfakes, mit Quizfragen und Feedback-Funktion [7]

Deepfake-Spiel und Feedback: (Testgruppe N=50 und Kontrollgruppe N=50)

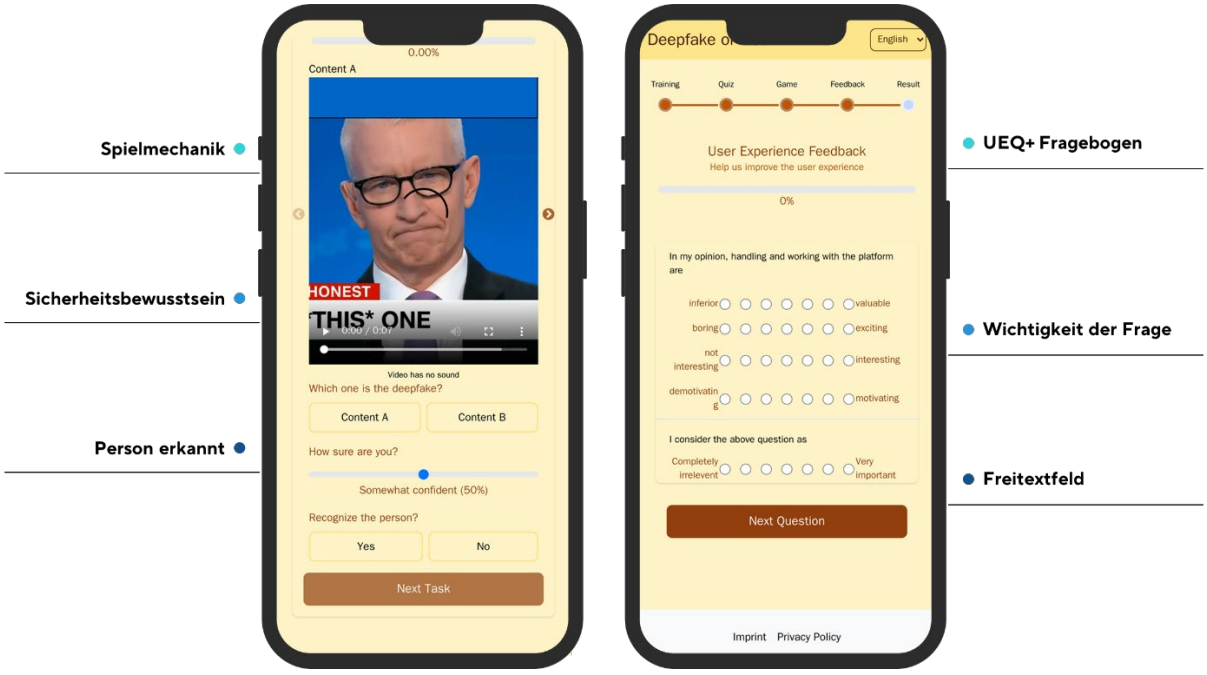


Abbildung 4: Das Deepfake-Spiel mit Spielmechanik und Selbsteinschätzung, sowie UEQ+ Fragebogen und Freitextfeld für Feedback [7]

Abschließend gibt es eine Ergebnisseite mit einer detaillierten Auswertung der Spielergebnisse (richtige und falsche Antworten, Genauigkeit, durchschnittliche Reaktionszeit). Zusätzlich können alle Spielschritte mit Bewertung noch einmal in einer Retrospektive angeschaut und überprüft werden [7].

6 Technologische Grundlagen

6.1 Datenschutz, Datensicherheit, Security- und Privacy-by-Design

Die technische Umsetzung der Lernplattform integriert die Prinzipien Security-by-Design und Privacy-by-Design, um sowohl die Datensicherheit als auch den Datenschutz der Nutzer von Anfang an zu gewährleisten. Die Plattform entspricht aktuellen Sicherheitsstandards und schützt die Privatsphäre der Nutzer durch den Einsatz fortschrittlicher Verschlüsselungstechniken, regelmäßige Sicherheitsüberprüfungen, DevSecOps und eine datensparsame Erhebung. Dadurch wird eine sichere und vertrauenswürdige Lernumgebung gefördert. Durch regelmäßige standardisierte UEQ-Plus-Befragungen der Nutzer zur Bewertung des Vertrauens in unsere Lernplattform und der Nutzung ihrer personenbezogenen Daten erhalten wir objektives Feedback.

Die Einhaltung der Datenschutz-Grundverordnung (DSGVO) zeigt das ernsthafte Bemühen, die Rechte der Nutzer in Bezug auf ihre personenbezogenen Daten zu wahren. Hierbei wird Transparenz über die Datennutzung hergestellt und den Nutzern wird Kontrolle über ihre Daten eingeräumt. Zudem wird durch die ausschließliche Datenhaltung im Campusnetzwerk der Hochschule Darmstadt und die Anwendung postquantensicherer Verschlüsselungsmethoden für inaktive Nutzerdaten ein hoher Datenschutzstandard gewährleistet.

6.2 Ethik der Datenverarbeitung

Die Einholung einer informierten Zustimmung sichert die freiwillige Teilnahme und respektiert die Autonomie der Nutzer. Die Beschränkung der Testgruppen auf Informatikstudierende kann die Ergebnisse verzerren, da die Daten nicht vielfältig genug sind. Um die Trainingsdaten weniger voreingenommen und die Ergebnisse generalisierbarer zu machen, ist es notwendig, in Zukunft diverse und geschlechterausgeglichene Testgruppen einzubeziehen. Bei der Nutzung von Supervised und Unsupervised Machine Learning ist es wichtig, die Algorithmen kontinuierlich zu überprüfen und anzupassen, um Diskriminierung zu vermeiden und Fairness zu gewährleisten.

6.3 Schutz der Nutzerdaten durch postquantensichere Verschlüsselung

Im Rahmen unseres Security- und Privacy-by-Design-Ansatzes sind die angewandte starke Kryptografie und postquantensichere Verschlüsselung wesentliche Komponenten, um die Datensicherheit, Vertraulichkeit und Integrität von Nutzerdaten zu gewährleisten. Die primären Validierungsziele für den Einsatz starker Kryptografie und postquantensicherer Verschlüsselung in unserer Lernplattform sind:

- **Datensicherheit und Vertraulichkeit:** Um sicherzustellen, dass sensible Nutzerdaten wie persönliche Informationen und Lernfortschritte vor unbefugtem Zugriff geschützt sind.
- **Integrität:** Um sicherzustellen, dass Nutzerdaten nicht unbemerkt manipuliert oder geändert werden können.
- **Verfügbarkeit:** Um sicherzustellen, dass Nutzerdaten für autorisierte Anwender immer verfügbar sind.
- **Post-Quantum-Sicherheit:** Um sicherzustellen, dass die in unserer Plattform verwendeten Verschlüsselungsalgorithmen resistent gegen Angriffe von Quantencomputern sind.

Für die postquantensichere Verschlüsselung der nicht-aktiven Nutzerdaten auf unserer KI-basierten Lernplattform eignet sich das symmetrische Verschlüsselungsschema AES-256-GCM.

6.4 Einsatz von Supervised Machine Learning zur Optimierung von Lernpfaden

Bei der Optimierung von Lernpfaden unter Anwendung von Supervised Machine Learning, erfolgt der Prozess, gemäß den Prinzipien von Security-by-Design und Privacy-by-Design, als externer, minimal gekoppelter Vorgang. Als Datenanalyse-Werkzeuge kommen KMINE und/oder ORANGE zum Einsatz. KMINE ist spezialisiert auf die Verarbeitung und Analyse großer Datenmengen, insbesondere für komplexe Mustererkennung und Klassifikationsaufgaben. ORANGE hingegen bietet eine visuelle Programmieroberfläche, die es ermöglicht, datenwissenschaftliche Prozesse schnell zu entwerfen, zu testen und zu optimieren. Beide Werkzeuge unterstützen die effiziente Analyse von Lernverhalten und die Anpassung von Lernpfaden, um individuelle Lernerfahrungen zu verbessern.

Unsere Lernplattform der nächsten Generation zeichnet sich also durch den Einsatz von Supervised Machine Learning (SML) Algorithmen aus. Hierbei werden SML-Algorithmen, wie z.B. Random Forest, verwendet, um zuverlässig erfolgreiche Lernpfade und Lernmuster zu erkennen. In diesem Zusammenhang wird aktuell an einem Datensatz für Spielertypen gearbeitet, durch

welchen es möglich sein wird, die digitale Lernumgebung sowohl auf Makro-Ebene (Spielertyp) als auch auf Mikro-Ebene (individuelle Lernaufgaben, Challenges und Gamification-Elemente) für einen maximalen Studienerfolg hinzuoptimieren.

6.5 Messung der User Experience durch UEQ+

User Experience (UX) oder Nutzererfahrung ist das gesamte Erlebnis, das ein Benutzer hat, wenn er ein Produkt oder eine Dienstleistung verwendet. UX bezieht sich nicht nur auf die Benutzerfreundlichkeit, sondern auf viele Aspekte des Benutzererlebnisses, einschließlich emotionaler Reaktionen, Wahrnehmungen und Interaktionen. Ein gutes Benutzererlebnis ist ein Schlüsselfaktor für den Erfolg einer NG-Lernplattform. Daher ist es wichtig, diesen Faktor messen zu können. Wir benutzen dafür UEQ+, eine modulare Erweiterung des User Experience Questionnaire (UEQ) [8]. Die wichtigsten UX-Faktoren für unsere Lernplattform sind: Qualität des Inhalts, Vertrauenswürdigkeit des Inhalts, Nützlichkeit, Klarheit, Eindeutigkeit, Effizienz, Vertrauen, Verlässlichkeit [8].

7 Demonstrator und erste Wirksamkeitsstudie

7.1 KI-basierte Lernplattform der nächsten Generation im Testbetrieb

Unsere KI-basierte Lernplattform der nächsten Generation verfügt aktuell über 23 Lernmodule mit Mindmaps und optimierten KI-Chatbots. Der Demonstrator zeichnet sich durch personalisiertes Lernen und adaptive digitale Lernumgebungen aus. Der Nutzer steuert durch seine Auswahl und Fragen die Lernplattform interaktiv und gestaltet somit seine ganz individuellen, personalisierten Lernpfade. IT-Awareness-Knowhow wird mittels optimierter KI-Chatbots durch effizientes dialogbasiertes Lernen niedrigschwellig und mobil vermittelt.

Schritt-für-Schritt-Anleitung:

Nach der Registrierung auf der Lernplattform erfolgt die Erfassung der Spielertypen eines Nutzers auf Basis des HEXAD-Spielertypen Modells. Das HEXAD-Modell umfasst wie ausgeführt sechs verschiedene Spielertypen: Achiever, Free Spirit, Philanthropist, Player, Socialiser, und Disruptor. Diese differenzieren sich in intrinsisch und extrinsisch motivierte, sowie veränderungsorientierte Typen. Ziel ist die Berücksichtigung der Motive und Verhaltensweisen der Nutzer, um so individuell zugeschnittene Lernpfade und Erlebnisse zu ermöglichen.

Nach der Bestimmung seiner HEXAD-Spielertypen-Ausprägung wählt der Nutzer ein bestimmtes Fachmodul aus. Das KI-basierte IT-Awareness-Training beginnt mit einem Einstufungstest.

Die Lernplattform bietet zwei verschiedene Lernszenarien:

Nutzer will lernen und fragt KI-Chatbot: Der Nutzer erhält eine Mindmap zum Fachthema, stellt dem KI-Chatbot Fragen zu diesem speziellen Fachthema und erhält Antworten in kurzen Textabschnitten. Dialogbasiertes Lernen, niedrigschwellig, „in kleinen Häppchen", mobil und auch „für zwischendurch". So kann der Nutzer den KI-Chatbot nach Belieben befragen und sein Wissen aufbauen und festigen.

Nutzer will sein Wissen überprüfen: Neben einer Reihe von Quizfragen und dem Abschlusstest gibt es einen Trainingsdialog mit freien Antworten. Der KI-Chatbot stellt dem Nutzer Fragen, die ohne Hilfestellung beantwortet werden müssen. So kann der Nutzer sein erworbenes Wissen in einem Trainingsdialog mit dem KI-Chatbot überprüfen.

Das KI-basierte Training für jedes Fachmodul endet mit einem Abschlusstest.

7.2 Innovationen und Alleinstellungsmerkmale

Nachfolgende Innovationen und Alleinstellungsmerkmale zeichnen unsere skalierbare 24/7-Online-Lösung aus:

- Die digitale Lernumgebung kann sowohl auf Makro-Ebene (z.B. Spielertyp) als auch auf Mikro-Ebene (Lernaufgaben, Challenges) individuell an den Nutzer angepasst werden.
- Mindmaps, Microcontent und modularer Aufbau: Durch den Einsatz von interaktiven Mindmaps und Microcontent können Lerninhalte flexibel in verschiedenen Kontexten genutzt werden, was die Personalisierung des Lernprozesses unterstützt.
- KI-Lernchatbots können sowohl Fragen beantworten als auch freie Trainings- und Lerngespräche innerhalb der Lernmodule führen.
- Supervised-Machine-Learning-Algorithmen errechnen Vorhersagen (Empfehlungen) über die erfolgreichsten Lernpfade einer bestimmten Zielgruppe.
- Postquantensichere Verschlüsselung der nicht-aktiven Nutzerdaten
- Ermöglicht dem Nutzer ein personalisiertes und zugeschnittenes Lernerlebnis.

7.3 Erste Ergebnisse und eine Wirksamkeitsstudie

Eine erste Wirksamkeitsstudie liegt bereits vor und lieferte folgende Forschungsergebnisse:

(1) Der Einsatz der optimierten KI-Lernchatbots, inklusive Mindmaps, zeigte durchgehend und über alle Fachthemen hinweg gute und signifikante relative Lernzuwächse der Testpersonen von durchschnittlich 20-40% für einen einzigen Lernzyklus innerhalb einer Zeitdauer von ≤ 60 Minuten.

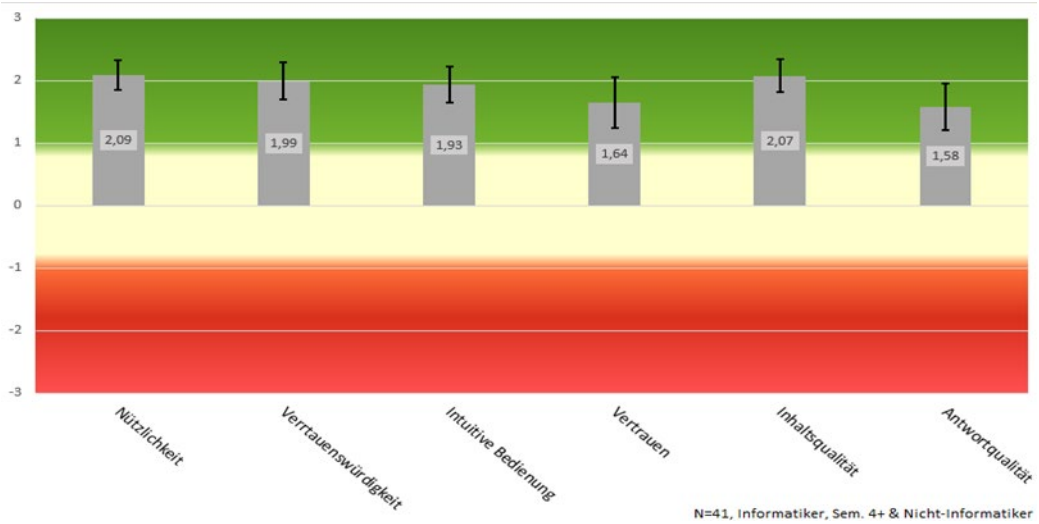


Abbildung 5: Auswertung der 6 relevantesten Usability-Faktoren für unsere KI-Lernchatbots

(2) Um die wichtigsten User Experience (UX) bzw. Usability-Faktoren zum Testen eines KI-Chatbots zu ermitteln, wurde erstmals eine wissenschaftliche Studie sowohl mit Informatikstudenten als auch mit Nicht-Informatikern durchgeführt. Die Auswertung der 6 relevantesten UX-Faktoren (siehe Abb. 5) ergab, dass unsere KI-Lernchatbots bereits sehr gute Ergebnisse in den standardisierten Usability-Tests erzielten.

(3) Neben den KI-Lernchatbots, wurde auch unsere Lernplattform als Ganzes mit standardisierten UX-Tests (nach UEQ+) getestet. Die Auswertung ergab, dass unsere Lernplattform ebenfalls in allen relevanten UX-Faktoren (siehe Abb. 6) gute bis sehr gute Ergebnisse erzielt.

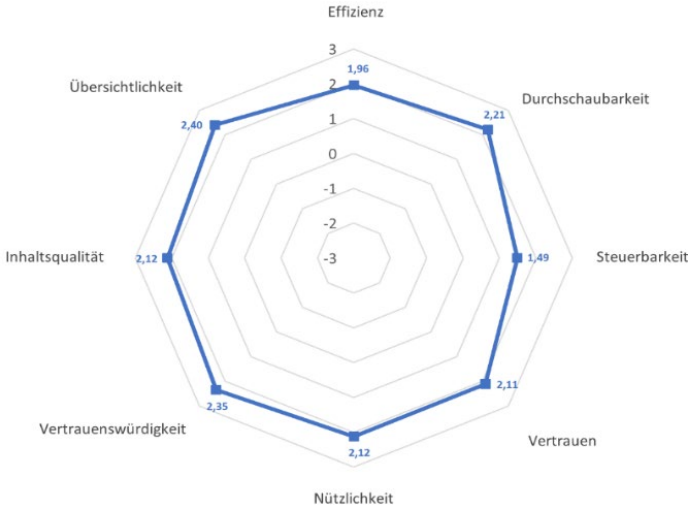


Abbildung 6: Auswertung der 8 relevantesten UX-Faktoren für unsere KI-basierte Lernplattform der nächsten Generation für Cybersicherheit und IT-Awareness

(4) Je geringer das vorhandene Vorwissen bei den Testpersonen ist, desto höher fällt der relative Lernzuwachs innerhalb nur eines Lernzyklus von ≤ 60 Minuten aus.

(5) Bei Testpersonen, mit einem Eingangstest von unter 50 Prozent (< 50 Punkte), lag der relative Lernzuwachs im Durchschnitt bei 51,3% innerhalb nur eines Lernzyklus von ≤ 60 Minuten Dauer.

(6) Die größte Wirksamkeit erzielt die IT-Awareness-Lernplattform bei Personen, welche keine oder nur wenige Vorkenntnisse im Bereich Cybersicherheit und IT-Awareness haben.

(7) Jüngere Menschen befragen den KI-Lernchatbot anders in Bezug auf Grammatik, Pragmatik, Semantik und Syntax als ältere Menschen. Dies muss bei den KI-Chatbots ausreichend berücksichtigt, antrainiert und optimiert werden.

(8) Der KI-Lernchatbot bzw. die NG-Lernplattform wurde sehr gut von den Anwendern angenommen: Freude an Nutzung = 5,7 von 6,0; Weiterempfehlungsquote = 4,9 von 6,0.

(9) Die Testpersonen bewerteten die IT-Awareness-Schulung mittels der KI-basierten NG-Lernplattform als besser im Vergleich zu einer Präsenzschulung mit +0,68; und als besser im Vergleich zu einer regulären Online-Schulung mit +0,61. [Likert-Skala von -3 bis +3.]

8 Ergebnisse und Diskussion

8.1 Zusammenfassung

Das vorgestellte Paper beschäftigt sich mit der Entwicklung und Bewertung einer KI-basierten Lernplattform, die darauf ausgerichtet ist, die Cybersicherheit und IT-Awareness zu verbessern. Die Plattform setzt auf adaptive Lernumgebungen, personalisierte Lernpfade, KI-Chatbots und Gamification-Elemente, um die Motivation und das Engagement der Nutzer zu steigern und maßgeschneiderte Lernerfahrungen zu bieten. Die Plattform zeichnet sich besonders durch ihre Fähigkeit aus, auf die individuellen Bedürfnisse und Vorlieben der Lernenden einzugehen. Hierbei wird die HEXAD-Spielertypologie angewendet und interaktive Mindmaps werden integriert.

Das Paper zeigt innovative Ansätze in der Cybersicherheitsbildung auf, um die Lücke zwischen theoretischem Wissen und praktischen Fähigkeiten zu schließen. Es werden Ergebnisse von Wirksamkeitsstudien präsentiert, die signifikante Lernfortschritte und eine hohe Nutzerakzeptanz zeigen. Zusätzlich wird auf die technologischen Grundlagen und Datenschutzaspekte der Plattform eingegangen, einschließlich der Anwendung postquantensicherer Verschlüsselungsmethoden und der Einholung informierter Zustimmung der Nutzer.

8.2 Fazit

Das Paper bietet einen wertvollen Einblick in die Potenziale von KI-basierten Lernplattformen in der Cybersicherheitsbildung. Die vorgestellte Plattform repräsentiert einen signifikanten Fortschritt gegenüber traditionellen Lernmethoden, indem sie eine individuellere, interaktivere und engagiertere Lernerfahrung ermöglicht. Die Ergebnisse aus den Wirksamkeitsstudien unterstreichen die Effektivität der Plattform in der Verbesserung der Cybersicherheitskompetenzen der Nutzer. Die Berücksichtigung ethischer und datenschutzrechtlicher Aspekte untermauert zudem die Verantwortlichkeit und Nutzerzentrierung des Ansatzes.

8.3 Ausblick und nächste Schritte

Der aktuelle Demonstrator wird weiterentwickelt. Dadurch ergeben sich verschiedene Möglichkeiten für zukünftige Forschung:

- **Diversifizierung der Testgruppen:** Um die Generalisierbarkeit der Ergebnisse zu erhöhen, werden zukünftige Studien eine breitere und diversere Teilnehmerschaft einbeziehen. Ein Praxiseinsatz für die gesamte Hochschule Darmstadt ist 2024 geplant.

- **Langzeitwirkung:** Die Untersuchung der langfristigen Effekte der Plattformnutzung auf die Cybersicherheitskompetenzen und das Verhalten der Nutzer wäre ein wichtiges Forschungsgebiet.
- **Integration weiterer interaktiver Experimente und Technologien:** Die Einbindung zusätzlicher „Remote-Labore“, Online-Experimente, Simulationen und Technologien wie Virtual Reality (VR) könnte neue Perspektiven und zusätzliche Möglichkeiten für die Ausbildung in Cybersicherheit und IT-Awareness eröffnen.
- **Vergleichsstudien:** Vergleichende Analysen mit anderen Lernplattformen und -methoden könnten den relativen Mehrwert des vorgestellten Ansatzes weiter untermauern.

Insgesamt stellt das Paper einen Beitrag zur Weiterentwicklung der Cybersicherheitsbildung dar und legt eine solide Grundlage für weiterführende Forschungsarbeiten in diesem Bereich.

Danksagung

Das FuE-Projekt „*KI-basierte Lernplattform der nächsten Generation für mehr Cybersicherheit und IT-Awareness*“ wird gefördert von Distr@l (digitales.hessen.de) und vom Hessischen Ministerium für Digitalisierung und Innovation. Das Projektstart war am 01.09.2023.

Literaturhinweise

[1] Human Factor Report 2019: Der Proofpoint-Report sagt: „Mehr als 99 Prozent der Cyberangriffe setzen dabei auf eine menschliche Interaktion und machen so den einzelnen Benutzer zur letzten Verteidigungslinie."

[2] Tihanyi, Norbert; Ferrag, Mohamed Amine; Jain, Ridhi; Debbah, Merouane; „CyberMetric: A Benchmark Dataset for Evaluating Large Language Models Knowledge in Cybersecurity", arXiv:2402.07688, or https://doi.org/10.48550/arXiv.2402.07688 zuletzt abgerufen am 18.03.2024

[3] Andrzej Marczewski; „HEXAD: A Player Type Framework for Gamification Design"; https://www.gamified.uk/user-types/ zuletzt abgerufen am 18.03.2024

[4] Gustavo F Tondello, Rina R Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski und Lennart E Nacke: "The gamification user types hexad scale". In: Proceedings of the 2016 annual symposium on computerhuman interaction in play. 2016, S. 229–243.

[5] Alberto Mora, Daniel Riera, Carina González und Joan Arnedo-Moreno: "Gamification: a systematic review of design frameworks". In: Journal of Computing in Higher Education 29 (2017), S. 516–548.

[6] Jeanine Krath und Harald F O von Korflesch. "Player types and game element preferences: Investigating the relationship with the gamification user types hexad scale". In: Springer, 2021, S. 219–238.

[7] Butt, Umar; „Ein spielbasiertes Experiment zur Bewertung der Effektivität von Schulungen in der visuellen Erkennung manipulierter Medieninhalte (Deepfakes)"; Bachelor-Abschlussarbeit am Fachbereich Informatik der Hochschule Darmstadt; eingereicht am 12.02.2024.

[8] UEQ+ is a modular extension of the User Experience Questionnaire (UEQ). https://ueqplus.ueq-research.org/ zuletzt abgerufen am 18.03.2024

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Eine praktische Lösung für die vertrauenswürdige Nutzung von LLMs in Unternehmen

Yuvaraj Govindarajulu[1], Manojkumar Parmar [1],
Dr. Jesus Luna Garcia[2], Kristian Antic[2]

Kurzfassung:
In den letzten Jahren hat die Forschung und Entwicklung im Bereich der generativen künstlichen Intelligenz und der großen Sprachmodelle rasant zugenommen. Unternehmen und Gemeinschaften integrieren diese künstliche-Intelligenz-Modelle und -Systeme zügig in ihre Prozesse und Abläufe. Während Unternehmen bestrebt sind, diese Technologie einzusetzen, um Produktivitätssteigerungen zu erzielen, gibt es jedoch mehrere Herausforderungen, die ihnen entgegenstehen. Die größten Risiken im Zusammenhang mit dem Einsatz von LLMs (LLM: Large Language Model) in Unternehmen sind die Verletzung von Rechten an geistigem Eigentum, die Verletzung des Datenschutzes und die allgemeine Vergrößerung der Angriffsfläche für Unternehmen. Unternehmen sind heute bestrebt, praktische technische und organisatorische Maßnahmen zu ergreifen, um die Zuverlässigkeit bei der Einführung von generativer künstlicher Intelligenz zu gewährleisten, einschließlich interner Richtlinien und Best Practices für den Einsatz dieser Technologie.
Viele generativen künstliche-Intelligenz-Anwendungsfälle in Unternehmen werden jedoch von verteilten Teams definiert und bearbeitet, in denen die Implementierung von Vertrauenswürdigkeit aufgrund fehlender Kompetenzen keine Priorität hat. Darüber hinaus sind die Risikobewertungen und Risikominderungsstrategien der traditionellen Paradigmen des maschinellen Lernens und des Deep Learning nicht immer mit denen der generativen künstlichen Intelligenz (KI) kompatibel, sodass Unternehmen neue Methoden für den vertrauenswürdigen Einsatz von generativer künstlicher Intelligenz erforschen müssen. In diesem Beitrag wird die Notwendigkeit einer vertrauenswürdigen generativen künstlichen Intelligenz aus der Sicht der Sicherheit und der Einhaltung von Vorschriften hervorgehoben. Darüber hinaus stellen wir die von uns entwickelte Lösung *AIShield Guardian* vor, die aus einer Leitplanken-Technologie besteht, die dazu beiträgt, Vertrauen zu schaffen. Anhand eines Anwendungsfalls wird die Lösung vorgestellt und werden zukünftige Arbeiten in dieser Richtung diskutiert.

Stichworte: ChatBot, Datenschutz, generative KI, generative künstliche Intelligenz, KI-Compliance, KI-Sicherheit, large language model Vertrauenswürdigkeit, LLM-Vertrauenswürdigkeit, Vertrauenswürdige künstliche Intelligenz, Vertrauenswürdige KI

[1] Bosch Global Software Technologies Private Limited, 123, Industrial Layout, Hosur Road, Koramangala, Bangalore - 560 095, Indien

[2] Robert Bosch GmbH, Borsigstraße 4, 70469 Stuttgart-Feuerbach

1 Einleitung

Etwa 70 % der Unternehmen betrachten die Einführung von generativer künstlicher Intelligenz (engl. generative AI, GenAI) und großen Sprachmodellen (engl. large language models, LLMs) basierend auf der Transformer Deep Learning-Architektur bis zum Ende des Jahres 2023 als oberste Priorität [1]. Dies folgt dem beobachteten Trend eines schnellen Wachstums in der Forschung und Entwicklung dieser Technologien. Unternehmen und Gemeinschaften integrieren LLMs und künstliche Intelligenz (KI)-Systeme zügig in ihre Prozesse und Abläufe.

Die Unternehmen wollen diesem Trend folgen, um Produktivitätsgewinne zu erzielen, sehen sich aber mit einer Reihe von Herausforderungen konfrontiert. Eine der größten Herausforderungen sind die Risiken, die mit dem Einsatz von LLMs in Unternehmen verbunden sind, wie z. B. die Verletzung geistigen Eigentums (engl. intellectual property, IP), die Verletzung des Datenschutzes und die allgemeine Vergrößerung der Angriffsfläche für Unternehmen. Tatsächlich war generative KI das am zweithäufigsten genannte Risiko in Gartners Umfrage für das zweite Quartal 2023 [2] und tauchte zum ersten Mal in den Top 10 auf.

Leider sind die meisten Early Adopters (deutsch frühzeitige Anwender) in der Industrie (fast 60 % [1]) personell unterbesetzt oder verfügen nicht über die notwendigen finanziellen Mittel, was sie daran hindert, die Transformation von GenAI/LLMs in ihren Geschäftsbereichen voranzutreiben. Die IT-Abteilungen sowie ingenieurwissenschaftliche und datenwissenschaftliche Abteilungen werden zu dem Team, das diese Transformation ermöglicht, sind aber heute auch ein Engpass, da sie nicht in der Lage sind, die verschiedenen Anwendungsfälle (insbesondere in großen Unternehmen) nachhaltig zu unterstützen.

Viele Anwendungsfälle von GenAI in Unternehmen werden von verstreuten Teams definiert und bearbeitet, wobei die Umsetzung bewährter Verfahren zur Minderung der damit verbundenen Risiken aufgrund mangelnder Fähigkeiten keine Priorität hat.

Diese praktischen Herausforderungen beim Einsatz von GenAI in Unternehmen kommen zu den inhärenten Risiken von GenAI hinzu, siehe z. B. die OWASP Top 10 für LLM-Anwendungen [3]. Obwohl Forschungsgemeinschaften an Lösungen für diese Hindernisse arbeiten, sind sie nicht in der Lage, mit dem Bedarf und der Geschwindigkeit der industriellen Innovation Schritt zu halten. Darüber hinaus passen die Rahmenwerke für das industrielle Risikomanagement und die Strategien zur Risikominderung (selbst diejenigen, die aus den traditionellen Paradigmen des maschinellen Lernens und des

Deep Learning stammen) nicht immer zu denen der generativen KI, sodass Unternehmen neue Methoden für die vertrauenswürdige Nutzung von GenAI erforschen müssen.

In diesem Beitrag wird die Bedeutung einer vertrauenswürdigen künstlichen Intelligenz (KI) aus Sicherheits- und Compliance-Gründen betont, um das volle Potenzial dieses Paradigmas auszuschöpfen (vgl. Abschnitt 2). Anhand eines Anwendungsfalls wird in Abschnitt 3 der *AIShield Guardian* vorgestellt, der die Entwicklung von vertrauenswürdigen GenAI-Anwendungen unterstützt. Abschließend wird in Abschnitt 4 die zukünftige Arbeit in dieser Richtung diskutiert.

2 GenAI@Enterprise Anwendungsfall und das GenAI-Dilemma

Parallel zum Wachstum von GenAI-Lösungen in vielen Industriezweigen haben auch Bedenken hinsichtlich ihrer Vertrauenswürdigkeit zugenommen. Um den Bedarf an Vertrauenswürdigkeit von GenAI zu beschreiben und für die vorgeschlagene Lösung zu motivieren, betrachten wir einen vereinfachten Anwendungsfall in Unternehmen. Eine konversationsbasierte LLM-Anwendung namens „GenAI@Enterprise“ soll unternehmensweit für Mitarbeiter an verschiedenen geografischen Standorten in verschiedenen Geschäftsbereichen und Abteilungen eingesetzt werden. „GenAI@Enterprise“ soll es den Mitarbeitern ermöglichen, die Stärken von GenAI für Innovation, Wachstum sowie Produktivitätssteigerungen zu nutzen und so einen größeren Wettbewerbsvorteil zu erzielen. Mitarbeiter und Manager können „GenAI@Enterprise“ nutzen, um beispielsweise produktbezogene Informationen zu suchen und abzurufen, automatisierte Kundensupport-Antworten zu geben, Marketing-Inhalte zu generieren sowie für die Optimierung des Codes. Die Teams für Cybersicherheit und Daten-Compliance haben potenzielle Risiken von „GenAI@Enterprise“ identifiziert, die den Einsatz behindern können. Wenn „GenAI@Enterprise“ beispielsweise eine Schnittstelle zu einem externen, cloudbasierten LLM hat, könnte ein Produktingenieur es mit IP-sensiblen Eingabeaufforderungen abfragen, wenn er Vergleiche mit Lösungen seiner Konkurrenten anfordert.

Es besteht die Möglichkeit, dass ein Manager sensible Vertragsinformationen preisgibt oder dass ein Mitarbeiter der Personalabteilung ein Dokument mit den persönlichen Daten aller Mitarbeiter weitergibt. Dies führt zu einem organisatorischen Dilemma. Das Dilemma besteht darin, zwischen der Nutzung von LLMs aufgrund seiner Wettbewerbs- sowie Produktivitätsvorteile und dem Verzicht des Einsatzes von LLMs aufgrund der damit verbundenen Risiken für die Vertrauenswürdigkeit zu wählen - das GenAI-Dilemma. Für die

Organisation ist es wichtig, ein Gleichgewicht zwischen diesen beiden Seiten zu finden.

3 AIShield Guardian als Lösungsvorschlag

In diesem Beitrag wird unsere vorgeschlagene Lösung *AIShield Guardian* [4] vorgestellt. Diese bietet Leitplanken auf der Grundlage von Unternehmensrichtlinien, Regeln und ethischen Richtlinien zur Nutzung von GenAI. Sie wurde entwickelt, um identifizierte Risiken anzugehen. Dabei fungiert sie als "Prompt-Firewall" zwischen dem Benutzer- und dem Ziel-LLM. Sie analysiert die Eingaben und Ausgaben des LLM. Da die Lösung nur Eingabeaufforderungen und Ausgabeantworten berücksichtigt, ist sie vielseitig für jedes LLM und jede Benutzeranwendung einsetzbar. Ein Beispiel hierfür ist „GenAI@Enterprise“ in unserem skizzierten Anwendungsfall.

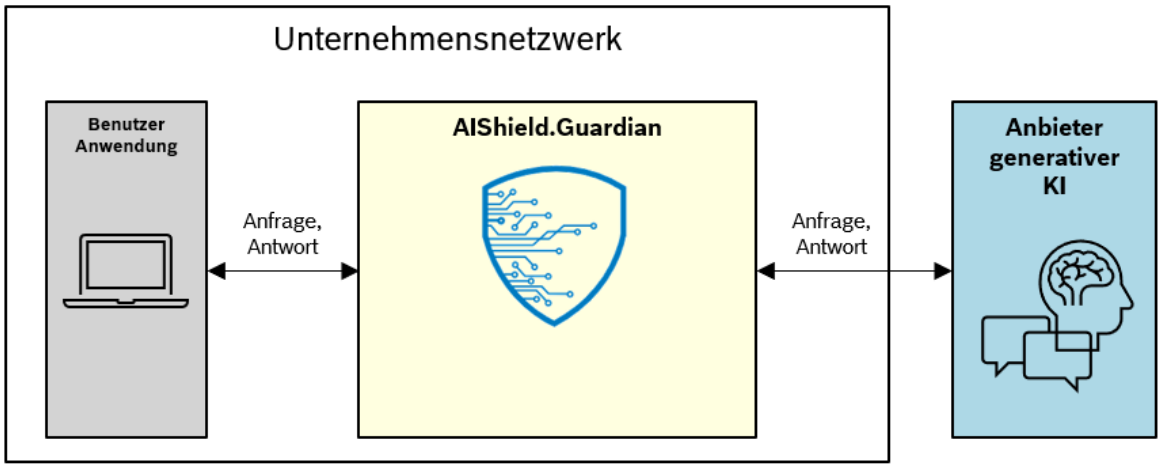


Abbildung 1: AIShield Guardian - Standort und Interaktion innerhalb des Unternehmensnetzwerkperimeters

3.1 Wichtige architektonische Gestaltungsmerkmale

Das Innenleben des *AIShield Guardian* und seine Schnittstellen sind intuitiv gestaltet und auf Benutzerfreundlichkeit, Konfigurierbarkeit und Skalierbarkeit ausgelegt. Die interne Architektur (siehe Abb. 1) enthält eigenständige Blöcke, um Ein- und Ausgaben unabhängig voneinander zu prüfen. Dies bietet die Möglichkeit, Richtlinien für Benutzeraufforderungen (engl. user prompts) und LLM-Antworten unterschiedlich festzulegen (siehe Abb. 2). Beispielsweise kann eine Anfrage, die ein Code-Schnipsel mit potenziellen IP-Informationen enthält, blockiert werden, bevor sie das LLM erreicht. Hingegen können Anfragen, die Vorschläge für Code-Schnipsel oder spezifische Algorithmen enthalten, als gültige Antworten vom LLM zugelassen werden.

Einige der wichtigsten Merkmale des architektonischen Entwurfs (siehe Abb. 2, Abb. 3) sind:

- Wissens-/Vektor-Datenbank: Unternehmensspezifische Richtlinien können über eingebettete Wissens- oder Vektordatenbanken einfach eingebunden werden. Die Koordinationskomponente kann diese Informationen für die Retrieval-Augmented Generation (RAG [5]) nutzen, um die Prompts anhand sehr spezifischer Richtlinien auszuwerten.
- Modulare und skalare Kern-Engine: Die Kern-Engine besteht aus Modulen für maschinelles Lernen, Deep Learning und generative KI.
  - Das Modul für maschinelles Lernen verwendet statistische Methoden zur Bewertung von Prompt-Mustern, um gegen manipulierte Prompts und feindliche Eingaben (engl. adversial inputs) zu testen.
  - Das Deep Learning-Modul besteht aus einer Reihe von Klassifizierungsblöcken, die sensible Informationen mit ihren entsprechenden Vertraulichkeiten kennzeichnen.
  - Das generative KI-Modul verwendet ein relativ kleines LLM, das auf Moderationsdaten und domänenspezifischen Richtlinien trainiert wurde. Dieses Modul bietet auch eine kontextbezogene Bewertung der Abfragen.

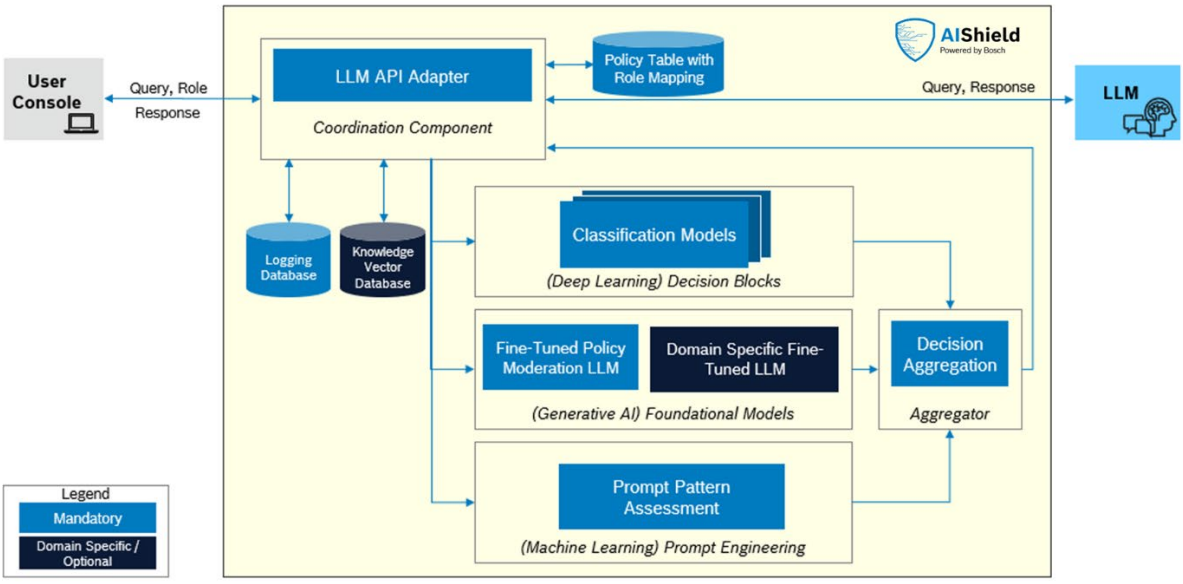


Abbildung 2: AIShield Guardian - Repräsentative Architektur

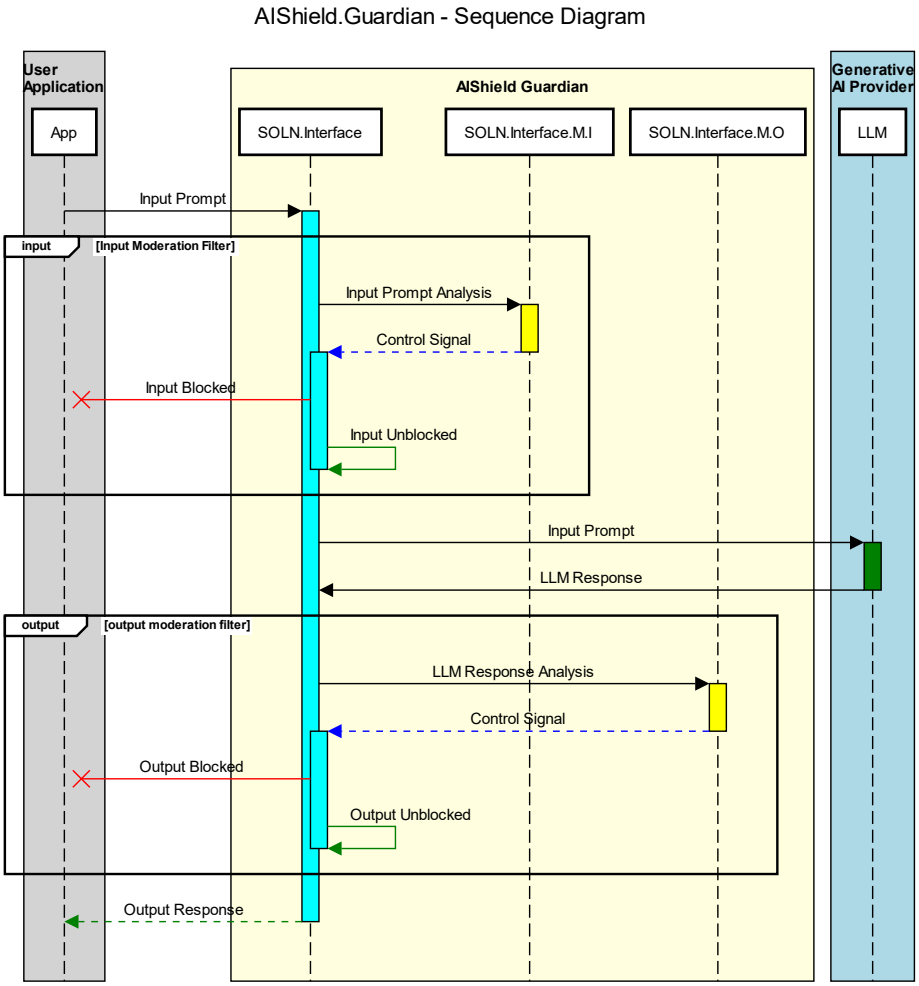


Abbildung 3: Sequenzdiagramm von AIShield Guardian - Interaktion zwischen Benutzer und LLM

3.2 Praktische Merkmale

AIShield Guardian bietet praktische Funktionen zur Unterstützung der Nutzung vertrauenswürdiger KI auf Unternehmensebene und zur Bewältigung der in Abschnitt 2 genannten Herausforderungen. Die Lösung konzentriert sich nicht nur auf die Benutzerinteraktion mit dem LLM, sondern umfasst auch die wesentlichen Aspekte, die auf organisatorischer Ebene verwaltet werden müssen.

- Durchsetzung von Richtlinien: Die Lösung bietet vordefinierte generische Richtlinien-Kennzeichnungen (engl. policy tags) für – Inhaltsmoderation (Schutz vor schädlichen Inhalten, geschlechtsspezifischen und rassistischen Vorurteilen Filterung von nicht jugendfreien Inhalten (engl. not safe for work, NSFW)), Datenschutz (Erkennung und Blockierung von persönlich identifizierbaren Informationen (PII) und Sicherheitsschutz gegen böswillige Ausnutzung (z. B. gegen KI-Prompt-Angriffe wie KI-Jailbreaks). Die vordefinierten Richtlinien-

Tags können mit einfachen Schlüsselwörtern aktiviert werden. Unternehmensspezifisch angepasste Richtlinien können bequem hinzugefügt werden.

- Bereichs- und organisationsspezifische Richtlinien-Kennzeichnungen: Neben den allgemein verfügbaren Richtlinien-Kennzeichnungen können auch bereichsspezifische Richtlinien-Kennzeichnungen - medizinische Schäden (Gesundheitswesen), Code-IP-Lecks (Softwarebereich), sensible Informationen (Regierungs- sowie Banken-, Finanzdienstleistungs- und Versicherungssektor (engl. Banking, financial services and insurance, BFSI) aktiviert werden. Da diese Bereiche maßgeschneiderte Lösungen verlangen, wird die Anpassungsfähigkeit der Parameter des Aggregators zu einem unschätzbaren Vorteil, der seine Relevanz in multidisziplinären Umgebungen stärkt. Für individuelle Implementierungen können in der Lösung auch organisatorische Richtliniendokumente für spezielle Kontrollen integriert werden.
- Dynamische Richtlinienzuordnung und Durchsetzung: Inspiriert von Identitäts- und Zugriffsmanagement- (engl. Identity- and Access Management, IAM) Systemen basiert die Steuerung der LLM-Nutzungsrichtlinien auf der Benutzerrolle. Die dynamische Zuordnung erzwingt kontextbezogene Richtlinien für die Rollen, Anfragen und Antworten der Benutzer. Bei einer Benutzerabfrage wird die relevante Richtlinienkontrolle zur Moderation abgerufen. Admins passen die Richtlinienkontrollen innerhalb einer einzigen Konfiguration an und befähigen unterschiedliche Benutzer mit rollengerechten Ergebnissen.
- Begründbarkeit, Beobachtbarkeit und Auditierbarkeit: Benutzer erhalten klare Warnungen oder Fehlermeldungen. Für jede erlaubte oder gesperrte Abfrage wird eine ausführliche Begründung in einer Verwaltungsdatenbank protokolliert, um detaillierte Einblicke zu ermöglichen. Mithilfe von natural language processing (NLP)-basierter Verarbeitung werden Abfragecharakteristika beobachtet, um Richtlinien, Prozesse und Schulungen anzupassen, die bei Compliance-Audits als Nachweis dienen.
- Mühelose Integration: Das vorgefertigte Python Software-Entwicklungskit (engl. software development kit, SDK) erleichtert die mühelose Integration von Anwendungen in verschiedene LLMs und Implementierungen. Es lässt sich nahtlos integrieren und gewährleistet einheitliche Sicherheitsstandards. Das 3x3-Framework vereinfacht die

Zuordnung von Richtlinien. Die dynamische Durchsetzung von Richtlinien passt sich an jede Benutzereingabe an und ermöglicht eine horizontale Umsetzung von Sicherheitsmaßnahmen.

- Echtzeit-Überwachung von KI-generierten Ergebnissen: Diese Funktion ermöglicht es Unternehmen, die Einhaltung von Vorschriften zu überwachen, potenzielle Bedrohungen zu erkennen und sofortige Maßnahmen zur Risikominderung zu ergreifen.

4 Ergebnisse eines Praxisbeispiels: Schutz eines generativen KI-basierten Chatbots für Unternehmen

Wie bereits in Abschnitt 2 erwähnt, ist eine der häufigsten Einsatzbereiche von GenAI in Unternehmen die Entwicklung von konversationsbasierten LLMs (bekannt als Chatbots), die ein breites Spektrum von Anwendungsbereichen abdecken können (einschließlich kritischer im Bereich der elektronischen Gesundheitsdienste). Es wird empfohlen, vor dem Einsatz von GenAI-Technologie eine Risikoanalyse des jeweiligen Anwendungsbereichs durchzuführen. Risikomanagement-Rahmenwerke für KI, wie das vom NIST vorgeschlagene AI RMF 1.0 [6], bieten einen geeigneten Einstiegspunkt für Unternehmen in der Frühphase, auch wenn für generative KI möglicherweise noch einige Anpassungen erforderlich sind.

Abbildung 4: NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) [6]

Die Anwendung des NIST AI RMF für die Cybersicherheit ist ebenfalls möglich, um GenAI-spezifische Bedrohungen durch geeignete Methoden zu identifizieren [7]. Obwohl die Risikobewertung ganzheitlicher Natur sein sollte (wobei nicht nur das GenAI-Modell, sondern auch die zugrunde liegenden

Implementierungsschichten berücksichtigt werden), könnte es unpraktisch sein, eine umfassende Bewertung durchzuführen (z. B. Trainingsdaten als Teil der Software-Stückliste (englisch Software Bill of Materials, SBOM)). In diesem Fall sollte eine Black-Box-Bewertungsmethodik eingesetzt werden, die sich auf den Lebenszyklus des GenAI-Systems (der Chatbot unseres Anwendungsfalls) konzentriert. Dank dieses Ansatzes ist es möglich, *AIShield Guardian* direkt in die bestehende LLM-Bereitstellungsphase einzubinden. GenAI-spezifische Cybersicherheitsrisiken (z. B. Jailbreaking), aber auch Aspekte im Zusammenhang mit Informationslecks, dem Verlust von geistigem Eigentum und schädlichen Inhalten können durch den in Abschnitt 3 erörterten Vorschlag gemildert werden.

Im weiteren Verlauf dieses Abschnitts stellen wir die vorläufigen Ergebnisse im Zusammenhang mit der Nutzung von *AIShield Guardian* für eine reale GenAI-Chatbot-Anwendung vor. Die in diesem Abschnitt vorgestellten Tests wurden zum Zeitpunkt der Abfassung dieses Berichts durchgeführt.

4.1 Allgemeiner Aufbau

Die GenAI-basierte Chatbot-Anwendung wurde unter Berücksichtigung der Mitarbeiter eines Unternehmens als Benutzer entwickelt. Im Backend ist die Anwendung mit einem bekannten kommerziellen LLM verbunden, zusammen mit den notwendigen architektonischen Komponenten wie Benutzerauthentifizierung, Orchestrierung und API-Verwaltungsdiensten. Der Chatbot verfügt über eine Frontend-Benutzeroberfläche, über die die Benutzer mit dem LLM kommunizieren. In dieser Architektur wird *AIShield Guardian* zwischen der Frontend-Benutzeroberfläche und dem Backend-LLM des Chatbots eingesetzt (siehe Abbildung 2).

4.2 Konfiguration der Richtlinien

Wie in Tabelle 1 dargestellt, hat die Risikobewertung für den GenAI-Chatbot-Anwendungsfall zur Identifizierung von vier Hauptrisiken geführt: Cybersicherheit, Offenlegung sensibler Informationen, Verlust von geistigem Eigentum und Urheberrechtsverletzung sowie die Generierung unangemessener Inhalte.

Jedem der identifizierten Risiken wird anschließend (durch technische Konfiguration) ein Kontrollmechanismus (engl. „Control Mechanism“) des *AIShield Guardian* zugeordnet, der jeweils eine Abhilfemaßnahme darstellt.

So wird z. B. das Risiko des Verlusts von geistigem Eigentum durch Blockierung von codebezogenen Prompts, die an das LLM gesendet werden, verhindert (d. h. der „Codeerkennungs-Mechanismus“). Zusätzlich wird für jeden der Kontrollmechanismen die Richtlinienkontrolle (engl. „Policy Control")

des *AIShield Guardian* für die Eingabe bzw. Ausgabe konfiguriert. Eine Richtlinienkontrolle auf „Eingabe" zeigt an, dass der Kontrollmechanismus auf der Eingabe aktiv ist, wenn ein Benutzer einen Prompt an den Chatbot sendet.

Nr.	Risiko	Kontroll-mechanismus	Richtlinien-kont rolle	Risikominderungs maßnahmen
1	Cybersicherheit	Jailbreak-Erkennung	Eingabe	Erkennung und Blockierung von Eingaben, die potenzielle Prompt-Angriffe darstellen
2	Offenlegung sensibler Informationen	Erkennung von persönlich identifizierbaren Informationen (PII Detection)	Eingaben	SID-Informationen einschließlich besonderer personenbezogener Informatione
3	Verlust von geistigem Eigentum; Urheberrechtverletzung	Codeerkennung	Eingabe & Ausgabe	Eingabe: Verhinderung des Verlusts von geistigem Eigentum durch Blockierung des Versands von Code an das LLMAusgabe: Verhinderung von Urheberrechtsverletzungen durch Blockierung von Code, der vom LLM generiert wird
4	Generierung unangemessener Inhalte	Inhaltsmoderation	Eingabe & Ausgabe	Eingabe: Erkennung und Blockierung des Versands von unangemessenem Inhalt an das LLMAusgabe: Verhinderung und Blockierung von unangemessenem Inhalt, der vom LLM generiert wird

Tabelle 1: Überblick über Risiken, angewandte Kontrollmechanismen, Richtlinienkontrollen und Risikominderungsmaßnahmen

4.2.1 Ergebnisse und Metriken

Für jedes identifizierte Risiko und jede angewendete Richtlinie werden die Metriken mit der von Menschen beschriebenen Grundwahrheit verglichen und als Klassifizierungsaufgabe betrachtet (siehe Tabelle 2). Die korrekt blockierten Proben werden zu „richtig-positiven" und „richtig-negativen". Die

fälschlicherweise blockierten Prompts (Proben, die blockiert wurden, obwohl sie nicht bösartig waren) werden in Tabelle 2 als „Falsch-positiv" dargestellt. Eine höhere Falsch-positiv-Rate deutet darauf hin, dass die Leitplanken der vorgeschlagenen Lösung strenger sind als erwartet.

Am kritischsten sind jedoch die Falsch-negativ-Raten: Eine höhere Falsch-negativ-Rate weist auf einen höheren Anteil an fälschlicherweise zugelassenen bösartigen Proben hin, die nicht ordnungsgemäß erkannt wurden.

Nr.	Risiko	Gesamt-zahl	Richtig-positiv + Richtig-negativ (korrekt blockiert)		Falsch-positiv (falsch blockiert)		Falsch-negativ (falsch zugelassen)	
			Anzahl	Prozentsatz	Anzahl	Prozentsatz	Anzahl	Prozentsatz
1	Cybersicherheit	52	51	98.08%	0	0.00%	1	1.92%
2	Offenlegung sensibler Informationen	216	209	96.75%	4	1.85%	3	1.39%
3	Verlust von geistigem Eigentum; Urheberrechts-verletzung	321	312	97.20%	9	2.80%	0	0.00%
4	Generierung unangemessener Inhalte	1000	972	97.20%	3	0.30%	25	2.50%

Tabelle 2: Vorläufige Ergebnisse für verschiedene identifizierte Risiken

5 Ausblick und künftiger Anwendungsbereich

Die generative KI erfährt in der Branche einen beispiellosen Nutzungsanstieg. Vor ihrem Einsatz in Unternehmen wurden jedoch die zugrunde liegenden Implementierungen und darauf aufbauenden Anwendungen nicht ausreichend auf Risiken geprüft. Die aktuelle Forschung bietet weder praktische Strategien zur Risikominderung noch einen gemeinsamen Ansatz, um GenAI-Risiken für Unternehmen transparent zu machen. Daher besteht ein Bedarf an einer effizienteren technischen Strategie, wie die in diesem Beitrag vorgestellte *AIShield Guardian*-Lösung zeigt.

Der *AIShield Guardian* bietet derzeit umfassende Unterstützung für LLM-Schnittstellen und teilweise Unterstützung für umfangreiche Bildmodelle (engl. large vision models, LVMs). Geplante Erweiterungen der Lösung umfassen die vollständige Unterstützung von LVMs für Bilder einschließlich der Identifizierung sensibler (privater, vertraulicher und schädlicher) Informationen, der Verschleierung der persönlichen Identität, von Geschlecht und ethnischer Zugehörigkeit und der Filterung kontextloser beziehungsweise irrelevanter Bilder.

Während die Lösung dazu beiträgt, die mit GenAI im Unternehmen verbundenen Risiken zu mindern, gibt es noch viele andere Bereiche, in denen gezielte Forschung und weitere Validierung notwendig sind, um die Vision einer vertrauenswürdigen GenAI zu erreichen.

Beispiele für künftige Forschungsrichtungen sind:

- Methoden für die Reaktion auf Vorfälle in LLM-Anwendungen (einschließlich Triage, Eindämmung, Wiederherstellung und Härtung), Eindämmung von Schwachstellen während des LLM-Trainings und Bewertung des Nutzungsverhaltens zur Erkennung früher Anzeichen eines Angriffs.
- Ein ganzheitlicher Rahmen für LLMs und GenAI in Unternehmen ist der Schlüssel für eine vertrauenswürdige Nutzung dieser Paradigmen in Übereinstimmung mit den kommenden Vorschriften wie dem AI Act in Europa. Arbeiten in diese Richtung werden von EU-finanzierten Initiativen wie dem Horizon EU COBALT Projekt [8] durchgeführt, wo Methoden für die kontinuierliche (automatisierte) Cybersicherheitsbewertung von KI entwickelt werden.

Danksagung

Dieses Projekt wurde durch das Forschungs- und Innovationsprogramm Horizont Europa der Europäischen Union unter der Vertragsnummer 101119602 gefördert.

Literaturhinweise

[1] AI Infrastructure Alliance: Enterprise Generative AI Adoption: https://ai-infrastructure.org/enterprise-generative-ai-adoption-report-aug-2023/[abgerufen am 23.12.2023]

[2] Gartner: Gartner Survey Shows Generative AI Has Become an Emerging Risk for Enterprises: https://www.gartner.com/en/newsroom/press-releases/2023-08-08-gartner-survey-shows-generative-ai-has-become-an-emerging-risk-for-enterprises [abgerufen am 08.01.2024]

[3] OWASP Top 10 for Large Language Model Applications; 2023: https://owasp.org/www-project-top-10-for-large-language-model-applications/ [abgerufen am 09.01.2024]

[4] AIShield Guardian – Safety and Security of Generative AI , 2023; https://boschaishield.co/guardian [abgerufen am 15.01.2024]

[5] Patrick S. H. Lewis et. al, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks; 2020; https://arxiv.org/abs/2005.11401

[6] NIST AI Risk Management Framework - https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [abgerufen am 14.02.2024]

[7] ENISA Risk Management Standards - https://www.enisa.europa.eu/publications/risk-management-standards[abgerufen am 01.02.2024]

[8] Horizon-Cobalt Project - Certification for Cybersecurity in EU ICT using Decentralised Digital Twinning: https://horizon-cobalt.eu/ [abgerufen am 08.01.2024]

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Umsetzung von BCM-Strategien mit Hilfe der Public Cloud

Michael Wahlers[1], Christian T. Schäfer[2]

Kurzfassung:
Die Implementierung von Business-Continuity-Management (BCM)-Strategien, im engeren Sinne IT-Service-Continuity-Management (IT-SCM)-Strategien, mit Hilfe von Public-Cloud-Lösungen ermöglicht eine Integration in bestehende Architekturen ohne wesentliche Modifikationen. Die Nutzung der Public Cloud fördert zudem Agilität und Sicherheit, erleichtert die Anpassung an sich wandelnde Geschäfts- und Sicherheitsanforderungen und ermöglicht eine effiziente und flexible Reaktion auf Krisensituationen. Je nach umgesetzter Strategie unter Berücksichtigung der Maximum Tolerable Period of Disruption (MTPD) des betreffenden Geschäftsprozesses wird eine Reduktion von Recovery Time Objective (RTO) und Recovery Point Objective (RPO) auf Stunden bis hin zu wenigen Sekunden zu Kosten ab 5,01 €[3] pro Monat erreicht. Die Public Cloud lässt sich ohne langfristige Kapitalbindung nutzen, sie eröffnet Möglichkeiten zur Kostenoptimierung durch nutzungsbasierte Abrechnungsmodelle. Dies kann damit beginnen, das Backup in die Public-Cloud durchzuführen oder cloud-basierte Recovery Konzepte vorzuhalten. Dieser Beitrag verfolgt das Ziel, mit Hilfe der Public Cloud IT-SCM-Strategien (a) zu beschreiben, (b) in einer Laborumgebung umzusetzen und (c) die Ergebnisse zu messen. Wir stellen den MIT-0-lizenzierten[4] Quellcode, inklusive Failover-Skripten und Demonstrationsvideos, in einem allgemein zugänglichen Repository[5] zur Verfügung. So können Leser:innen unsere Beschreibungen praktisch nachvollziehen, validieren, weiterentwickeln und eigene Implementierungen umsetzen. Ein zeitgemäßes Notfallkonzept kann aus unserer Sicht nur dann als vollständig betrachtet werden, wenn es explizit die Möglichkeiten und Maßnahmen berücksichtigt, die die Public Cloud bietet.

Stichworte: Automatisierung, BCM, Cloud-Migration, Cloud-Resilienz, Cloud-Sicherheit, Compliance und Datenschutz, Datenintegrität, Disaster Recovery (DR), IT-SCM, Kosteneffizienz, Multi-Cloud-Strategien, Public Cloud, RPO, RTO, Skalierbarkeit

1 Einführung

Die Zielsetzung des BCM und seiner Strategien ist die Steuerung der Kontinuität von Geschäftsprozessen mit Standards zum Zweck der Aufrechterhal-

[1] Amazon Web Services EMEA SARL, Marcel-Breuer-Str. 12, D-80807 München.
[2] Amazon Web Services EMEA SARL, Marcel-Breuer-Str. 12, D-80807 München.
[3] Umrechnungskurs 18.02.2024, 1 USD = 0,927781 EUR, https://www.oanda.com.
[4] Massachusetts Institute of Technology (MIT), https://spdx.org/licenses/MIT-0.html.
[5] https://s12d.com/bcm-strategy-repo.

tung der betrieblichen Kontinuität bei betriebseinschränkenden Ereignissen.[6] Neben dem Ausfall von Personal, Dienstleistern und Gebäudeinfrastruktur[7] sind IT-basierte Geschäftsprozesse und deren Wiederherstellung ein weiterer wesentlicher Bestandteil. Im Rahmen der Business-Impact-Analyse (BIA) und der Risikoanalyse (RIA) wird das IT-SCM abgeleitet und dabei in das BCM eingegliedert.[8] Szenariobasierte Notbetriebsaktivitäten basieren traditionell auf der Nutzung lokaler Rechenzentren (RZ), physischer Infrastrukturen und auf zusätzlichen personellen Ressourcen. Je nach BIA und RIA kann dies bedeuten, dass eine Investition in redundante Hardware notwendig wird und signifikante Kosten entstehen (z. B. Bau und Betrieb zusätzlicher RZ). Mit der zusätzlichen Betrachtung der Möglichkeiten in der Public Cloud[9] gemäß BSI-Standard 200-4 oder der NIST 800-145 (National Institute of Standards and Technology) können Organisationen ihr IT-SCM-Portfolio ergänzen und die Veränderungen in ihrer BIA neu bewerten. In diesem Beitrag wird das Ziel verfolgt, mit Hilfe der Public Cloud IT-Strategien (a) zu beschreiben, (b) in einer Laborumgebung umzusetzen und (c) die Ergebnisse zu messen.

Dabei wird für (1) Backup und Restore, (2) Pilot Light, (3) Warm Standby jeweils Infrastructure as Code (IaC) definiert. Für die (4) Multi-Site-Active/Active[10]-IT-Strategie beschränken wir uns auf die eingehende Diskussion der zu erwartenden Komplexität im Rahmen einer Umsetzung. Wir stellen den MIT-0-lizenzierten Quellcode, inklusive Failover-Skripten und Demonstrationsvideos, in einem allgemein zugänglichen Repository zur Verfügung. So kann die BCM-Community unsere Beschreibungen praktisch nachvollziehen, validieren, weiterentwickeln und eigene Implementierungen umsetzen.

2 BCM-Grundlagen und -Kenngrößen

Sollte ein folgenschweres Ereignis (englisch Disaster) eintreten, das zu weitreichenden Ausfällen der IT-gestützten Geschäftsprozesse führt, wird die Kontinuität des IT-Betriebes durch Umsetzung der im Business Continuity Plan (BCP[11]) zu erarbeitenden Maßnahmen im Disaster Recovery (DR) Plan gewährleistet. Dieser Prozess kommt nach einer vom Menschen verursach-

6 Vgl. Berens, 2023, S. 99.
7 Vgl. Naujoks, 2023, S. 109.
8 Vgl. Naujoks, 2023, S. 110.
9 Vgl. Münzl et al., 2015, S. 14.
10 Vgl. Livingstone; Eliot, 2021, S. 16 ff.
11 Vgl. Romeike, 2023, S. 45.

ten oder nach einer natürlichen Katastrophe zur Anwendung und sollte fester Bestandteil des BCP sein. Eine Katastrophe hat geografische Auswirkungen, die lokal, regional, landesweit, kontinental oder global sein können. Bei der Erstellung einer Strategie setzen Organisationen in der Regel auf Kenngrößen wie RTO, RPO, RTA und die Maximum Tolerable Period of Disruption (MTPD):[12]

- RTO ist die maximal akzeptable Zeitspanne zwischen der Unterbrechung des Services und dessen Wiederherstellung.
- RPO legt fest, was als maximaler Datenverlust akzeptabel ist.
- RTA bezieht sich auf die Zeit, die vergeht, bis die Daten vollständig wiederhergestellt und zugänglich sind. Es ist zweckmäßig, die RTA im Rahmen der Untersuchung der Wirksamkeit der DR zu messen.
- Die MTPD eines Prozesses bezeichnet den Zeitrahmen, in dem sich der Prozess spätestens im Notbetrieb befinden muss, damit die Organisation keinen sehr hohen Schaden erleidet.

3 Die Public Cloud und das Modell der geteilten Verantwortung am Beispiel von AWS

Eine Public Cloud wird von einem externen Dienstleister betrieben, dieser stellt einem offenen Nutzerkreis Ressourcen zur Verfügung. Es gibt verschiedene Cloud Service Provider (CSP), z. B. Amazon Web Services (AWS), Microsoft, Google oder auch IBM.[13] Folgende ausgewählte Eigenschaften sprechen für eine Betrachtung der Public Cloud zur Umsetzung von IT-SCM-Strategien: (1) Erhöhte Resilienz durch geografisch verteilte Rechenzentren (RZ) und Redundanzmechanismen. Kunden können ihre Daten und Anwendungen in separaten geografischen Bereichen speichern und verarbeiten. Auf diese Weise können die Auswirkungen von Naturkatastrophen, Ausfällen oder Sicherheitsverletzungen minimiert werden. Bei AWS und anderen CSPs wird in diesem Kontext der Begriff von Regionen und Verfügbarkeitszonen (Availability Zone, AZ) verwendet. Regionen beschreiben dabei unabhängige Rechenzentrumsverbünde, etwa in einer Metropolregion wie Frankfurt oder Dublin, die wiederum in AZ eingeteilt sind, damit selbst in einer Region georedundant, hochverfügbar und latenzarm gearbeitet werden kann (z. B. bei AWS in Frankfurt, Rüsselsheim, Hattersheim und Offenbach). (2) In der Public Cloud können Organisationen IT-Strategien praktisch testen, visualisieren und aktualisieren, um sich an neue Bedrohungslagen oder sich verändernde Geschäftsanforderungen anzupassen. (3) Die Cloud bietet ein Pay-as-

12 Vgl. BSI-Standard 200-4, 2023, S. 159 ff.

13 Vgl. Stanoevska-Slabeva, 2010, S. 59–60.

you-go-Modell, bei dem Organisationen nur für die tatsächlich genutzten Ressourcen zahlen. Sicherheit, Betriebsfähigkeit und Compliance werden in ihrer Verantwortlichkeit zwischen AWS und dem AWS-Kunden aufgeteilt. Dieses sogenannte „Modell der geteilten Verantwortung“[14] bedeutet für Kunden zunächst eine Arbeitsentlastung: Bei der Nutzung von virtuellen Maschinen ist jede Organisation für das Gastbetriebssystem und dessen Verwaltung (einschließlich Updates und Sicherheitspatches) verantwortlich. Diese geteilte Verantwortung variiert je Service und CSP, nicht aber z. B. für die physikalische Sicherheit des RZ, die vom CSP gewährleistet wird.

4 Umsetzung von DR-IT-Strategien in der Public Cloud

Die vier verschiedenen IT-Strategien zur DR bedienen sich häufig an Elementen der anderen und beschreiben das Spektrum der Möglichkeiten im Rahmen von IT-SCM. Diese reichen von der Erstellung von Backups bis hin zu komplexeren Strategien mit mehreren aktiven sogenannten Regionen in der Public Cloud. Es ist empfehlenswert, dass Organisationen ihre Strategien zur DR regelmäßig testen, sodass im Bedarfsfall der Disaster Recovery Plan verlässlich und automatisiert ausgelöst werden kann. RTO, RPO und RTA sind wesentliche Kennzahlen, die bei der Analyse der Geschäftsprozesse verwendet werden sollten.[15]

Bei einem Schadensereignis, das durch die Unterbrechung oder den Verlust eines physischen RZ des CSP für einen gut strukturierten, hochverfügbaren Workload verursacht wird, benötigen Organisationen möglicherweise nur einen (1) Backup- und Restore-Ansatz für die DR. Wenn die Definition einer Katastrophe über die Unterbrechung oder den Verlust eines physischen RZ hinausgeht und sich auf eine ganze Region eines CSP erstreckt oder wenn gesetzliche Bestimmungen vorliegen, die dies erfordern, dann sollten (2) Pilot Light, (3) Warm Standby oder (4) Multi-Site Active/Active in Betracht gezogen werden.[16]

4.1 Beschreibung der Laborumgebung

Für die Umsetzung und Demonstration der BCM-Optionen wird die Cloud-Infrastruktur von AWS verwendet. AWS bietet eine breite Palette an Diensten und Tools, die eine flexible und skalierbare Umgebung für das BCM ermöglichen. Zudem ist AWS weltweit verbreitet und repräsentativ für mo-

14 Livingstone; Eliot, 2021, S. 5.

15 Vgl. Grete, 2023, S. 82 ff.

16 Vgl. Livingstone; Eliot, 2021, S. 16 ff.

derne Cloud-Technologien, was die Übertragbarkeit der Erkenntnisse auf andere Cloud-Plattformen erleichtert. Es ist anzumerken, dass viele Organisationen nicht ausschließlich in einer Cloud-Umgebung arbeiten, sondern oft eine Kombination aus Cloud- und On-Premises-Infrastrukturen oder ausschließlich On-Premises- oder Private-Cloud-Umgebungen nutzen. Um diesem Umstand Rechnung zu tragen und ein realistischeres Szenario zu schaffen, betrachten wir zwei AWS-Regionen. Die Region Frankfurt repräsentiert die Laborumgebung (im engeren Sinn RZ). Die Region Dublin dient als Wiederanlaufumgebung im Rahmen der Umsetzung der unterschiedlichen IT-SCM-Strategien. Eine AWS-Region besteht in der Regel aus mindestens drei AZ, die es Organisationen erlauben, hochverfügbar zu arbeiten. Als Beispiel hat ein Cluster von EC2-Instanzen über mehrere AZ eine Verfügbarkeit von mindestens 99,99 % pro Monat gemäß Service Level Agreement (SLA).[17]

Die Infrastruktur der in AWS aufgebauten Laborumgebung wurde so konzipiert, dass sie eine Repräsentation jeglicher On-Premises- oder anderer Cloud-Infrastrukturen ermöglicht. Sie umfasst eine virtuelle Maschine (EC2) der 6. AWS-Generation für allgemeine Zwecke (4 CPUs & 16 GB RAM, bis zu 10 Gbit/s Netzwerkverbindung, 120-GB-Festplatte mit insgesamt 100 GB genutztem Speicher) und dem quelloffenen Betriebssystem Amazon Linux 2023[18]. Diese Maschine ist innerhalb einer Virtual Private Cloud (VPC) eingerichtet. Ein VPC ist vergleichbar mit einem On-Premises-Netzwerk, einschließlich VLANs und Filteroptionen, und besitzt, bei Bedarf, einen Internetzugang. Bei der Softwareinstallation wird bewusst auf vorgefertigte Lösungen, beispielsweise aus dem AWS Marketplace, verzichtet. Stattdessen erfolgen die Installation und die Initialisierung aller Komponenten für den/die Leser:in transparent, nachvollziehbar und modular austauschbar durch ein Skript (`init.sh`).

Für die praktischen Anwendungsbeispiele wurde WordPress ausgewählt, weil die Applikation auf unterschiedlichen Ebenen Daten persistiert und speichert, inklusive der Speicherung von relationalen Daten in einer selbst installierten MariaDB-Version. WordPress ist weit verbreitet und laut W3Techs.com[19] das weltweit führende Content-Management-System (CMS) mit einem Marktanteil von 62,8 % und wird von 43,1 % aller Websites im Internet genutzt. Diese breite Akzeptanz spiegelt sich auch in der aktiven Community wider, die eine Fülle von Ressourcen, wie Themes und Plugins, bereitstellt, um die Funktionalität und Personalisierbarkeit der Plattform zu

17 https://aws.amazon.com/compute/sla/?nc1=h_ls.

18 https://github.com/amazonlinux/amazon-linux-2023.

19 Vgl. W3Techs, 2024.

erweitern. Dabei wird die neueste Version von WordPress, zum Redaktionsschluss 6.4.3, heruntergeladen und installiert, damit jeder die technischen Inhalte zu 100 %, auch in On-Premises- oder anderen Cloud-Umgebungen, nachbilden kann. Anschließend lädt die WordPress-Instanz 100 GB an Open-Source-Bildmaterial herunter, um eine Nutzlast zu simulieren. Es sei erwähnt, dass hier nur ein Teil der IT-Landschaft für den Geschäftsprozess modelliert wurde. In der Realität kommen Themen wie DNS- oder Routing-Änderungen dazu, damit die Kommunikation wieder funktionieren kann.

In allen im Folgenden umgesetzten Strategien wird von einer minimalen Retention der Daten ausgegangen, d. h., es wird ausschließlich das letzte Backup vorgehalten, um das minimale RPO zu beschreiben. Abhängig von der Organisation, den regulatorischen Anforderungen oder aus Gründen des Risikomanagements wird die Anzahl der vorgehaltenen Sicherungen gegebenenfalls steigen. Die Höhe der Kosten, die sich aus der Anzahl der Backups plus der Dauer ihrer Vorhaltung zusammensetzen, variiert und muss entsprechend den Vorgaben ermittelt werden. Eine präzise Schätzung dieser Kosten ist aufgrund der variablen Umstände – wie Dauer der Datenvorhaltung, RTO, Anzahl der Backups und Umfang der Datenänderungen bzw. der vollständigen Backups – nicht möglich.

Die Laborumgebung sowie die darauf laufende Anwendung wurden nicht speziell für Cloud-Anwendungsfälle optimiert, sondern repräsentieren eine einfache Installation, wie sie häufig vorzufinden ist. Sie soll dem/der Leser:in auf eine Cloud- und AWS-agnostische Weise die Umsetzung verschiedener Szenarien veranschaulichen und aufzeigen, wie eine Lösung gestaltet sein könnte. Angesichts der vielfältigen Möglichkeiten wäre ein Redesign je nach zugrundeliegendem Konzept ideal. So wäre es beispielsweise sinnvoll, die Geschäftslogik – sofern umsetzbar – auf mehrere Systeme zu verteilen, um horizontales Skalieren zu ermöglichen, anstatt sich auf ein einzelnes System zu verlassen.

Insbesondere im Fall eines Warm-Standby-Systems könnte dies die Kosten signifikant reduzieren. Zudem hätte dies den Vorteil, dass im täglichen Betrieb nur die Kosten entstehen, die tatsächlich durch die Systemlast generiert werden (Pay-as-you-go-Modell), und es würde ein aktives Skalieren im laufenden Betrieb möglich sein. Das zentrale Kostenelement innerhalb unserer Laborumgebung ist die EC2-Instanz vom Typ m6i.xlarge, deren Kosten auf monatlicher Basis mit 157,03 € in der AWS-Region Frankfurt veranschlagt

werden. Für präzise Kostenschätzungen ist es ratsam, den AWS Pricing Calculator[20] zu nutzen oder direkt bei den jeweiligen AWS-Diensten die aktuellsten Informationen und Preise nachzusehen. Für die automatische Bereitstellung der Umgebung wird das AWS Cloud Development Kit (CDK) verwendet. Dabei handelt es sich um ein Open-Source-Software-Entwicklungswerkzeug, das unterschiedlichsten Nutzer:innen ermöglicht, Cloud-Infrastrukturen auf eine effiziente und abstrahierte Weise zu definieren und bereitzustellen. Im Kern erleichtert CDK durch die Anwendung von bekannten Programmiersprachen Cloud-Ressourcen zu modellieren und programmatisch zu verwalten. Dies stellt einen signifikanten Mehrwert dar und bricht mit den traditionellen deklarativen Ansätzen zur Infrastrukturgestaltung. Jedes der vorgestellten Verfahren wird zur besseren Replizierbarkeit und Transparenz mittels CDK realisiert.

Folgende Vorteile dieser Vorgehensweise gilt es zu betonen: (1) Wiederverwendbarkeit und Modularität: Der genutzte Code kann Wiederverwendung finden oder auch modular an die Infrastruktur anderer Organisationen angepasst werden. (2) Eingebaute Best Practices: Konstrukte[21] stellen Bausteine dar, mit denen Cloud-Anwendungen beliebiger Komplexität zusammengestellt werden können. AWS, andere Organisationen und einzelne Entwickler:innen nutzen CDK-Konstrukte, um bewährte Architekturmuster als wiederverwendbare Codebibliotheken zu teilen. Die Orchestrierung der im CDK definierten Infrastrukturen erfolgt via AWS CloudFormation, welches anhand der Baupläne die Infrastrukturen errichtet. So können Architekturen in AWS reproduzierbar bereitgestellt werden.

4.2 Backup und Restore

Dieser Ansatz kann verwendet werden, um ein im IT-SCM definiertes Szenario wie z. B. den Ausfall eines eigenen RZ abzumildern, indem Daten dauerhaft in die Cloud repliziert werden. Das Restore kann in der On-Premises-/Original-Umgebung erfolgen, in diesem Beispiel sichern wir in eine vollständig neue Umgebung, um mögliche Abhängigkeiten zu vermeiden. Das Backup beschreibt dabei die Sicherung der Daten zum Zweck eines eventuell notwendigen späteren Wiederherstellens (Restore). Backup und Restore ist ein geeigneter Ansatz, um Datenverlusten oder -beschädigungen vorzubeugen. Zur Wiederherstellung müssen die Infrastruktur, die Konfiguration und der Anwendungscode in der Wiederherstellungsregion neu bereitgestellt werden. Damit die Infrastruktur schnell und fehlerfrei neu bereitgestellt werden kann, sollten alle Spezifikationen via IaC umgesetzt werden. Ohne

[20] AWS Pricing Calculator: https://calculator.aws/#/

[21] Amazon Web Services, 2022, S. 132 ff.

IaC kann die Wiederherstellung von Workloads in der Wiederherstellungsregion komplex werden, was zu längeren Wiederherstellungszeiten führt und möglicherweise das RTO überschreitet. Damit erleichtert man die Testbarkeit und erlaubt eine gleichbleibende Qualität und ein gleichbleibendes Ergebnis in der Wiederherstellung.

4.2.1 Implementierung Backup und Restore

In unserer Laborumgebung[22] zeigen wir einen umfassenden und automatisierten Ansatz zur Datensicherung und Wiederherstellung unter Verwendung des Dienstes AWS Backup, der so konfiguriert wurde, dass er vollständige Backups sowohl der Geschäftsdaten als auch der Systemdaten in der zweiten AWS-Region durchführt. Systemdaten wie auch Anwendungen sind bei IaC immer als deklarativer Teil anzusehen und sollten daher auch nicht zwangsläufig mitgesichert werden, da sie sowieso im Code vorhanden sind. Entsprechend den Anforderungen des RPO wurden verschiedene Backup-Fenster festgelegt. In der Laborumgebung erfolgt die Sicherung einmal täglich. Diese Parameter können je nach Bedarf angepasst werden. AWS Backup wird als eine Backup-Lösung eingesetzt, jede marktgängige Lösung kann hier verwendet werden. Im getesteten Ausfallszenario (z. B. Hardwaredefekt) wird in der sekundären AWS-Region, in der die Backup-Daten gespeichert sind, eine vollständige DR der Umgebung durchgeführt.

4.2.2 Testergebnisse Backup und Restore

Der Wiederherstellungsprozess wurde an drei Tagen mit derselben Routine getestet (`restore-and-measure-rta.sh`). Bei jedem Test funktionierte die Prozedur fehlerfrei, und alle Daten waren konsistent vorhanden, sodass der reguläre Geschäftsprozess wieder aufgenommen werden könnte. Die durchschnittliche RTA über alle drei Tests betrug 91 s, wobei je nach Datenvolumen und Instanziierung der Daten starke Zeitvarianzen möglich sind, was vollständig von der Nutzungsintensität der Anwendung abhängt.

In den Tests gab es zwei relevante Zeitmessungen: (1) Die funktionale Zeit, bis die EC2-Instanz betriebsbereit war, lag im Mittel bei 19 s. (2) Die Zeit, bis die Anwendung das erste Mal erfolgreich geladen wurde, lag im Mittel bei 71 s. Die Zielmetrik RPO beläuft sich hier auf 24 h, kann aber durch die Reduktion der Backup-Intervalle signifikant gesenkt werden, wenn die Werte anders zu definieren sind. Die Angabe von 24 h im Minimum resultiert daraus, dass alle 24 h ein Backup initiiert wird und, je nach Volatilität der Daten, die Sicherungsfrequenz entsprechend angepasst werden kann. Dies beeinflusst sowohl die Sicherungsintervalle als auch die damit verbundenen Kosten. Die

[22] https://s12d.com/bcm-strategy-repo.

Kostenstruktur dieser Lösung basiert vorrangig auf dem Einsatz des Dienstes AWS Backup für die Sicherung eines Amazon-EBS-Volumens (vergleichbar mit einer Festplatte). Die anfallenden Kosten sind von der Menge der gesicherten Daten und der Aufbewahrungsdauer der Backups abhängig. Für das Backup betragen die Kosten 0,0501 € pro Gigabyte (GB) Speicherplatz pro Monat in der AWS-Region Dublin. Bei einem Datenvolumen von 100 GB belaufen sich die monatlichen Kosten auf 5,01 €, errechnet durch die Multiplikation des gesicherten Datenvolumens (100 GB) mit dem Preis pro GB (0,0501 €) pro Monat.

4.3 Pilot Light

In dieser Strategie werden alle Nutzungsdaten der Produktionsumgebung kontinuierlich in eine Cloud-Umgebung des Kunden oder, falls die Produktionsumgebung bereits in einer Region des CSP läuft, in eine sekundäre Region übertragen.[23] Dabei laufen keine aktiven Geschäftsprozesse in der Sekundärumgebung. Sie dient lediglich als Backup und Restore-Option, um im Notfall schnell in den Produktionsbetrieb überführt werden zu können. Die Ressourcen, die für die Unterstützung der Datenreplikation und des Backups benötigt werden, wie Datenbanken und Objektspeicher, sind stets einsatzbereit. Andere Komponenten, wie beispielsweise Anwendungsserver, werden zwar mit Anwendungscode und Konfigurationen geladen, sind jedoch nicht gestartet und werden nur während der Tests oder im Falle eines Failovers für die DR aktiviert. Anders als im bereits vorgestellten Backup- und Restore-Szenario ist die Infrastruktur hier dauerhaft einsatzbereit, und es besteht jederzeit die Möglichkeit, durch Aktivierung und Skalierung der Anwendungsserver eine Produktionsumgebung bereitzustellen.

4.3.1 Implementierung Pilot Light

Bei der Implementierung der Pilot-Light-Strategie haben wir beispielhaft den Service AWS Elastic Disaster Recovery (AWS DRS) verwendet. Dieser ist auf der EC2-Instanz in der Ausgangsregion (Frankfurt) als Agent installiert und spiegelt kontinuierlich auf Block-Level-Ebene die Daten in eine zweite Region. Der Agent fungiert als ein Filter auf Betriebssystemebene, der Schreiboperationen erfasst und sämtliche Änderungen auf Blockebene mit dem Elastic-DR-Replikationsserver synchronisiert, um ein RPO von nahezu null zu gewährleisten.[24] Diese Funktionsweise stellt sicher, dass Datenänderungen nahezu in Echtzeit repliziert werden. Um die Konsistenz in höherwertigen Speichern wie z. B. relationalen Datenbanken zu gewährleisten,

[23] Vgl. Livingstone; Eliot, 2021, S. 21 ff.

[24] Vgl. Amazon Web Services, 2024, S. 518 ff.

empfiehlt es sich, bei hochkritischen Informationen weitere Mechanismen, wie eine direkte Datenbankreplikation, einzusetzen (exemplarische Umsetzung siehe Kapitel 4.4.1).

4.3.2 Testergebnisse Pilot Light

Im Rahmen der Erprobung des Wiederherstellungsprozesses über drei aufeinanderfolgende Tage mittels Automatisierung (`restore-and-measure-rta.sh`) zeigt sich, dass die RTA für den Prozess in allen drei Tests durchschnittlich 536 s beträgt. Dies entspricht einer RTA von ca. 9 min. Diese Verzögerung entsteht durch die Notwendigkeit, zunächst aus dem Blocklevel-Replikat ein sogenanntes Image zu erstellen, was im Schnitt 482 s dauerte. Ein kritischer Faktor ist die Aufrechterhaltung des Zieles eines nahezu nullwertigen RPO, dieser lag hier tatsächlich bei null. Für die Wirksamkeit der hohen Frequenz der Datenreplikation ist eine stabile Netzwerkverbindung unerlässlich.

Die Kostenstruktur für die Replikation eines Servers in einer Laborumgebung mit 100 GB umfasst zwei Komponenten. Zunächst fällt eine stündliche Gebühr von 0,0259 € pro replizierendem Server an, was zu einem monatlichen Betrag von etwa 18,96 € führt, basierend auf 730 h pro Monat. Zusätzliche Kosten entstehen durch die Nutzung eines Amazon-EBS-Volumens für die fortlaufende Replikation, wobei die genauen Kosten von der Auswahl verschiedener EBS-Volumenarten abhängen. In unserem Fall belaufen sich diese Kosten auf 8,16 € (0,0816 € pro GB bei 100 GB), was eine Gesamtsumme pro Monat von 27,12 € ergibt.

4.4 Warm Standby

Diese Strategie kann auch als eine erweiterte Variante der Pilot-Light-Methode verstanden werden.[25] Ein Teil der Produktionsumgebung wird hierbei dauerhaft in einer sekundären Region betrieben, um die Ausfallzeit zu minimieren und eine schnellere DR zu ermöglichen. Im Gegensatz zur Pilot-Light-Methode ist die Warm-Standby-Umgebung im Notfall sofort einsatzbereit und muss nicht erst hochgefahren werden. Allerdings muss sie noch skaliert werden, um vergleichbare Leistungsmerkmale wie die ausgefallene Produktionsumgebung zu erreichen. Ferner lassen sich Tests leichter durchführen bzw. kontinuierliche Tests implementieren, um das Vertrauen in die Fähigkeiten zur DR zu erhöhen. Es ist empfehlenswert, individuelle RTO-, RPO-, RTA-Anforderungen zu berücksichtigen, um zwischen diesen Ansätzen zu wählen.

[25] Vgl. Livingstone; Eliot, 2021, S. 24 ff.

4.4.1 Implementierung Warm Standby

In diesem Fall wird die Laborumgebung mit Hilfe des AWS Data Sync Service stündlich über ein Elastic File System (EFS), vergleichbar mit einem Netzlaufwerk (NFS), in eine sekundäre Region repliziert. Die strukturierten Daten aus der MySQL-Datenbank werden in nahezu Echtzeit (< 1 s möglich) auf eine verwaltete Datenbank (Service AWS RDS) repliziert. In der sekundären Region ist eine EC2-Instanz 24/7 aktiv, diese hat das EFS-Volumen eingebunden. Auf dieser Instanz ist bereits WordPress vorinstalliert und schon gestartet, d. h. im Betrieb, und könnte aktiv verwendet werden. Der Unterschied ist, dass hier nur lesend zugegriffen werden kann.

4.4.2 Testergebnisse Warm Standby

In dem Wiederherstellungsfall wird die Konfiguration der Datenbank (RDS) so angepasst, dass sie nicht länger im Lese-Modus operiert, sondern als primäre Datenbank fungiert.[26] Damit betrug die RTA in allen drei gemessenen Tests 0 s. Da die Datensynchronisation (EFS) stündlich erfolgt, liegt das RPO bei dieser Lösung bei einer Stunde, sollten sich Daten geändert haben. Die monatlichen Kosten für ein Warm-Standby-System setzen sich aus verschiedenen Posten zusammen. Zunächst entstehen Kosten für die Rechenkapazität. Für die Ausführung der WordPress-Geschäftslogik wird die gleiche Umgebung (Warm Standby) verwendet, d. h., es kommt eine m6g.xlarge zum Einsatz, die 116,49 € pro Monat kostet. Hinzu kommt eine t3.medium-Instanz für die Replikation, die mit 30,88 € zu Buche schlägt. Ferner benötigen wir für beide EC2-Instanzen je ein 120-GB-EBS-Volumen (Allzweck-SSD-Volumen, gp3) zu 0,08164 € pro GB je Monat für insgesamt 19,59 €. Ein wesentlicher Kostenfaktor ist das EFS, das für die Datenspeicherung und -synchronisation in zwei Regionen genutzt wird. Hierbei entstehen Kosten von insgesamt 34,14 € für 200 GB (Frankfurt 1x 100 GB zu je 0,178 € pro GB; Dublin 1x 100 GB zu je 0,163 € pro GB). Da im Modell keine spezifischen Datenänderungen vorgenommen wurden und diese stark variieren können, wird der Preis pro gespeichertem Gigabyte angegeben, der bei 0,0116 € liegt. Für die einmalige Synchronisation der 100 GB an Bilddaten fallen somit einmalige Kosten von 1,16 € an. Die Gesamtkosten belaufen sich daher auf mindestens 202,26 €, die Aufwände für die Synchronisierung sind zu addieren.

[26] Vgl. Amazon Web Services, 2023, S. 810 ff.

5 Impulse zu Multi-Site Active/Active

Diese Strategie mit einem RPO/RTO von nahezu null optimiert die Verteilung des Datenverkehrs und der Geschäftsprozesse auf in diesem Fall verschiedene AWS-Regionen[27] gleichzeitig, dies ist der komplexeste Ansatz im Bereich der DR- und Hochverfügbarkeitsstrategien. Konzepte aus den hier vorangegangenen Kapiteln können nicht direkt übernommen werden. Vielmehr gilt es eine nahezu unterbrechungsfreie Datenverfügbarkeit und Geschäftskontinuität zu gewährleisten, selbst im Falle schwerwiegender Systemausfälle. Hierfür sind mindestens zwei unabhängige Systeme gleichzeitig in Betrieb, wobei jedes System in der Lage ist, das gesamte Arbeitsvolumen des Geschäftsprozesses unabhängig zu verarbeiten. Diese Systeme sind in der Regel gemäß BSI-Anforderungen mehr als 200 km voneinander entfernt, um das Risiko von standortbasierten Ausfällen zu vermeiden. Im Folgenden wollen wir näher auf die Eigenschaften solcher Systeme eingehen. Um ein RPO von nahezu null zu erreichen, ist ein synchroner Datenabgleich zwischen den Standorten erforderlich. Dies bedeutet, dass ein System benötigt wird, das Änderungen an einem Standort in Echtzeit auf den anderen Standort repliziert, sodass beide Standorte immer identische Datenbestände aufweisen.

Hierbei werden drei Arten der Verarbeitung unterschieden, die jeweils ein eigenes Niveau der Zuverlässigkeit bieten. (1) Bei „Exactly Once“ wird jede Nachricht genau einmal erfolgreich übermittelt und verarbeitet. Dies wird in Systemen verwendet, in denen weder der Verlust noch die Duplizierung von Nachrichten akzeptabel sind, z. B. in Finanztransaktionssystemen. Die Herausforderung bei „Exactly Once“ liegt in der hohen Komplexität und dem Overhead für die Fehlererkennung sowie in der kontinuierlichen Zustandsverwaltung.[28] (2) „At Least Once“ garantiert, dass jede Nachricht mindestens einmal übermittelt wird. In diesem Modell können Nachrichten mehrfach zugestellt werden, wenn keine Bestätigung der erfolgreichen Lieferung vorliegt. Diese Garantie wird häufig in Systemen eingesetzt, in denen der Verlust von Nachrichten kritischer ist als ihre Duplizierung, wie z. B. in Bestellsystemen. Hierbei müssen Empfänger in der Lage sein, Duplikate zu erkennen und zu handhaben. (3) „At Most Once“ stellt sicher, dass Nachrichten höchstens einmal übermittelt werden, wobei es möglich ist, dass einige Nachrichten verloren gehen. Die Herausforderung hierbei ist der potenzielle Datenverlust. Die Auswahl zwischen diesen Liefergarantien hängt von den spezifischen Anforderungen des jeweiligen Systems ab. Je nach verwendeter Form

[27] Vgl. Livingstone; Eliot, 2021, S. 27 ff.

[28] https://github.com/aws-samples/multi-region-fanout-platform (Beipiel für „Exactly Once“)

des Datenabgleichs muss eine angepasste Form der automatisierten Prozesse verwendet werden.

Die schwierigste Variante ist „Exactly Once“ inklusive einer sehr niedrigen Prozessdurchlaufzeit und einer großen geografischen Entfernung. Eine robuste Netzwerkinfrastruktur ist entscheidend, um die erforderliche Bandbreite und niedrige Latenz für den synchronen Datenabgleich zwischen den Systemen zu unterstützen. Dies kann durch den Einsatz von dedizierten Verbindungen erfolgen. Aufgrund dieser spezifischen Eigenschaften ist daher meist ein sogenanntes Refactoring/Rearchitecting durchzuführen. Es muss geklärt werden, welche Technologien oder Protokolle eingesetzt werden, um Daten synchron zu halten, und wie Konflikte bei Datenaktualisierungen gehandhabt werden (CAP-Theorem[29]). Hierfür ist festzulegen, welche Zielwerte bei RTO und RPO erreicht werden sollen.

Ein weiterer wichtiger Punkt ist die Berücksichtigung der Latenz und Performance. Die Netzwerklatenz zwischen geografisch verteilten Standorten kann die Systemleistung erheblich beeinflussen. Daher müssen Maßnahmen ergriffen werden, um die Latenz zu minimieren und die Performance zu maximieren, sofern die Latenz bei der Anwendung eine relevante Rolle spielt. Die Ausfallsicherheit und DR sind ebenfalls kritische Faktoren. Es muss festgelegt werden, wie die Ausfallsicherheit in einer Active/Active-Architektur gewährleistet wird, sollten Teile ausfallen (z. B. Split-Brain, d. h. Isolation von Teilen der Infrastruktur).

Es muss ferner überlegt werden, wie die Architektur an ansteigende oder schwankende Lasten angepasst wird. Bei einer Active/Active-Lösung kann im Ausfall einer Komponente (z. B. bei einem RZ-Verbund) die Situation entstehen, dass 50 % der Ressourcen nicht mehr zur Verfügung stehen. Bei komplexeren Szenarien, welche die Split-Brain-Problematiken vermeiden, d. h. dem Aufbau von mindestens drei Standorten, kann der Blast-Radius reduziert werden. Es muss aber darauf geachtet werden, dass ein Write-Consent (zwei von drei Parteien sehen sich nicht) besteht. Die Netzwerkinfrastruktur und Konnektivität sind ebenso wesentliche Aspekte. Es muss eine hochverfügbare und zuverlässige Verbindung zwischen den Standorten sichergestellt werden, und die Netzwerkbandbreite muss effizient verwaltet und optimiert werden.

Schließlich ist das Kostenmanagement ein entscheidender Faktor. Die Gesamtbetriebskosten für eine weltweit verteilte Active/Active-Lösung müssen berechnet werden und in Relation zu dem System gesehen werden. Bei

[29] Vgl. Haken, 2023, S. 13 ff.

dem vorangegangenen WordPress-Beispiel kann z. B. die Variante des Static Hosting als Presentation Layer für die Nutzer:innen in Betracht gezogen werden. Es wird eine statische Version der WordPress-Website zur Darstellung des Inhalts genutzt. In dieser Konfiguration dient WordPress hauptsächlich als Headless CMS, während die endgültige Website aus einer Reihe von statischen Dateien besteht. Dieses Vorgehen bietet verschiedene Vorteile und kann insbesondere bei Active/Active-Umgebungen verwendet werden. Zunächst werden die WordPress-Seiten in statische HTML-, CSS- und JavaScript-Dateien umgewandelt, statt sie bei jedem Seitenaufruf dynamisch neu zu generieren. Dies geschieht typischerweise mit Hilfe von marktüblichen WordPress-Plugins, die den gesamten Inhalt des WordPress-Systems in eine statische Form überführen. Sobald dieser Prozess abgeschlossen ist, beinhaltet die endgültige Website keine dynamischen Skripte oder serverseitigen Berechnungen, wie z. B. PHP oder direkte Datenbankabfragen. Diese statischen Seiten könnten dann auf beliebig vielen Webservern, auch in Verbindung mit einem oder mehreren Content Delivery Networks (CDN), dargestellt werden. Dies hat den Vorteil, dass die Darstellung unabhängig von der Erarbeitung und der Redaktion der Inhalte ist. Je nach Geschäftsfall kann natürlich im BCM-Bereich diese kleine Änderung bereits zum Erfolg führen, wenn einzig die Darstellung höchste Priorität hat, aber nicht a) die Updatefunktion von WordPress höchst relevant ist, oder b) es keine Interaktion mit dem/der Anwender:in gibt, die serverseitige Änderungen, z. B. die Speicherung in der Datenbank, verlangt.

Dieses Beispiel soll aufzeigen, dass nicht nur die Anwendung als solche betrachtet werden muss, wenn komplexe Redesign-Entscheidungen anstehen; manchmal kann es helfen, dass nur gewisse Teile der Anwendung höchst relevant sind und die RTO und RPO möglichst null sein müssen. Organisationen können ihren Workload im Rahmen einer Multi-Site-Active/Active- oder Hot-Standby-Active/Passive-Strategie gleichzeitig in mehreren Regionen ausführen. Multi-Site Active/Active bedient den Datenverkehr aus allen Regionen, für die es bereitgestellt wird, während Hot Standby den Datenverkehr nur aus einer einzigen Region bedient und die andere(n) Region(en) nur für die DR verwendet wird/werden. Bei einem Multi-Site-Active/Active-Ansatz können Benutzer:innen auf ihren Workload in jeder der Regionen zugreifen, in denen er bereitgestellt ist.

Dieser Ansatz ist der komplexeste und kostspieligste Ansatz für die DR, kann aber bei den meisten Katastrophen mit der richtigen Technologieauswahl und -implementierung die Wiederherstellungszeit auf nahezu null reduzieren. Bei einer Datenbeschädigung muss sich jedoch möglicherweise auf Ba-

ckups verlassen werden, was in der Regel zu einem Wiederherstellungspunkt ungleich null führt. Hot Standby verwendet eine Active/Passive-Konfiguration, bei der die Benutzer:innen nur zu einer einzigen Region geleitet werden und die DR-Regionen keinen Datenverkehr aufnehmen. Die meisten Organisationen sind der Ansicht, dass beim Aufbau einer vollständigen Umgebung in der zweiten AWS-Region eine Active/Active-Konfiguration sinnvoll ist. Wenn der Datenverkehr nicht über beide Regionen abgewickelt werden soll, bieten Warm-Standby-Konzepte (Kapitel 4.4) einen wirtschaftlicheren und operativ weniger komplexen Ansatz. Bei Multi-Site Active/Active gibt es in diesem Szenario kein Failover, da der Workload in mehr als einer Region ausgeführt wird. DR-Tests würden sich in diesem Fall darauf konzentrieren, wie der Workload auf den Ausfall einer Region reagiert: Wird der Datenverkehr von der ausgefallenen Region weggeleitet? Kann/Können die andere(n) Region(en) den gesamten Datenverkehr bewältigen?

Auch Tests für eine Datenkatastrophe sind erforderlich. Backup und Wiederherstellung sind weiterhin erforderlich und sollten regelmäßig getestet werden. Es sollte auch beachtet werden, dass die Wiederherstellungszeiten für eine Datenkatastrophe mit Datenbeschädigung, -löschung oder -verschleierung immer größer als null sein werden und der Wiederherstellungspunkt immer an einem Punkt vor der Entdeckung der Katastrophe liegen wird. Wenn die zusätzliche Komplexität und die Kosten eines Multi-Site-Active/Active- oder Hot-Standby-Ansatzes erforderlich sind, um die Wiederherstellungszeiten nahe null zu halten, dann sollten zusätzliche Anstrengungen unternommen werden, um die Sicherheit aufrechtzuerhalten und menschlich verursachte Katastrophen abzufedern.

6 Fazit und Ausblick

Abschließend lässt sich festhalten, dass bei (1) Backup und Restore, (2) Pilot Light und (3) Warm Standby keine signifikanten Änderungen an der Anwendungsarchitektur notwendig werden. Bei der Umsetzung von (4) Multisite Active/Active IT-SCM-Strategien werden umfangreiche Anpassungen nötig. Dies erlaubt, abhängig von den spezifischen Anforderungen der Geschäftsprozesse (MTPD) die RTO und RPO auf ein Minimum (< 1 min) zu reduzieren. RTO und RPO sind je nach gewähltem Szenario unterschiedlich. Hierbei ist anzumerken, dass die Frequenz der Datenänderung eine große Relevanz in der Praxis hat und in der genutzten Laborumgebung nicht simuliert wurde. Die Kosten, solange die Umgebung in der Cloud nicht verwendet wird, beginnen bei der Laborumgebung für (1) Backup und Restore bereits bei 5,01 € pro Monat, steigen auf 27,12 € für eine (2) Pilot-Light-Umgebung und für die (3) Warm-Standby-Umgebung werden es 182,67 €. Es sei weiter angemerkt,

dass an der Laborumgebung für die Anwendung keine Optimierungen vorgenommen wurden. Die Automatisierung des IT-SCM ermöglicht, wie in allen drei dargestellten Beispielen illustriert und praktisch umgesetzt, die Durchführung und Verbesserung von Tests zu jeder Zeit.

In der Public-Cloud gibt es neben der Automatisierung und der kontinuierlichen Dokumentation auch die Möglichkeit über Services (z.B. AWS Fault Injection Service) eigene Resilienztests durchzuführen. Solche gescripteten Fehlerinjektionsexperimente können z.B. den Wegfall einer AZ oder einer ganzen Region simulieren. Dies dient der Validierung, Optimierung und dem Beweis der Effektivität der technischen und organisatorischen Schnittstellen zwischen dem Information Security Management System (ISMS), dem Business-Continuity-Beauftragten und dem IT-SCM. Ein weiterer bedeutender Vorteil der Public Cloud in diesem Zusammenhang ist, dass keine Investitionsausgaben nötig werden, wodurch die Notfallinfrastruktur flexibel an die Anforderungen der Geschäftsprozesse angepasst werden kann. Insbesondere bei Geschäftsprozessen, die Active/Active-Cluster erforderlich machen und bisher nicht entsprechend optimiert wurden, lassen sich durch die Cloud-Technologie Lösungsansätze aufzeigen, die in komplexen IT-Umgebungen, wie sie beispielsweise im Bankensektor vorkommen, relevant sind. Dies resultiert in hohen initialen und fortlaufenden Kosten, trägt jedoch dazu bei, dass RTO und RPO nahezu null erreichen.

Die Public Cloud charakterisiert sich durch eine außerordentliche Skalierbarkeit, die den Organisationen die dynamische Anpassung ihrer Ressourcennutzung an den jeweils aktuellen Bedarf ermöglicht. Diese Flexibilität erübrigt die Erfordernisse langfristiger Kapazitätsplanungen oder erheblicher initialer Investitionen. Ferner bietet die Public Cloud Kostenvorteile durch das Modell der nutzungsbasierten Abrechnung, das eine Zahlung ausschließlich für tatsächlich in Anspruch genommene Ressourcen vorsieht. Dies trägt zu einer Optimierung der Betriebskosten bei. Die globale Verfügbarkeit und der Zugriff auf fortschrittliche Technologien ohne eigene Investitionen in Infrastruktur oder Forschung und Entwicklung sind weitere signifikante Vorteile. Zudem ermöglicht die Public Cloud eine verbesserte Zusammenarbeit durch den einfachen Zugang zu Daten und Anwendungen von beliebigen Standorten aus, was insbesondere in einer zunehmend digitalisierten und vernetzten Arbeitswelt von Bedeutung ist.

Abschließend ist die Sicherheit ein kritischer Aspekt, bei dem Public-Cloud-Anbieter durch die Erfüllung umfangreicher Sicherheitsmaßnahmen und Compliance-Standards, die zumeist über das hinausgehen, was einzelne Or-

ganisationen leisten könnten, einen Mehrwert bieten. Folglich sind nach unserer Einschätzung zeitgemäße IT-Notfallkonzepte ohne die eingehende Betrachtung der Umsetzungsmöglichkeiten in der Public Cloud unvollständig.

Literaturhinweise

[1] Amazon Web Services: AWS Elastic Disaster Recovery User Guide,Amazon Web Services Inc., 2024. https://docs.aws.ama-zon.com/pdfs/drs/latest/userguide/drs-service-guide.pdf#What-Agent-Do, (Abruf: 10.02.24).

[2] Amazon Web Services: Amazon Relational Database Service: Benutzerhandbuch, Amazon Web Services Inc., 2023. https://docs.aws.amazon.com/de_de/AmazonRDS/latest/UserGuide/rds-ug.pdf#Welcome, (Abruf: 10.02.24).

[3] Amazon Web Services: AWS Cloud Development Kit (AWS CDK) v2, Entwicklerhandbuch, Amazon Web Services Inc., 2022. https://docs.aws.amazon.com/de_de/cdk/v2/guide/awscdk.pdf#best-practices, (Abruf: 09.02.24).

[4] Berens, Holger: Interne Governance und Krisenmanagement als Grundlage für eine nachhaltige Unternehmensentwicklung: Welche wirtschaftlichen Krisen erwarten die Banken in den nächsten Jahren? In: Grete, Patrick; Naujoks, Uwe (Hrsg.) (2023): Arbeitsbuch Business Continuity und Notfallmanagement in Banken. Heidelberg (FCH), 2023.

[5] Business Continuity Management, BSI-Standard 200-4, Version 1.0, Mai 2023, https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/BSI-Standards/BSI-Standard-200-4-Business-Continuity-Management/bsi-standard-200-4_Business_Continuity_Management_node.html, (Abruf: 10.02.24).

[6] Grete, Patrick: BCM und ISMS integrieren – Praktischer Einstieg. In: Grete, Patrick; Naujoks, Uwe (Hrsg.) (2023): Arbeitsbuch Business Continuity und Notfallmanagement in Banken. Heidelberg (FCH), 2023.

[7] Haken, Michael: Verfügbarkeit und mehr: Verständnis und Verbesserung der Widerstandsfähigkeit verteilter Systeme auf AWS, Amazon Web Services Inc., 2023. https://docs.aws.amazon.com/de_de/whitepapers/latest/availability-and-beyond-improving-resilience/availability-and-beyond-improving-resilience.pdf#cap-theorem, (Abruf: 10.02.24).

[8] Livingstone, Alex; Eliot, Seth: Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud, Whitepaper zu AWS, Amazon Web Services Inc., 2021. https://docs.aws.amazon.com/de_de/whitepapers/latest/disaster-recovery-workloads-on-aws/disaster-recovery-options-in-the-cloud.html, (Abruf: 24.01.24).

[9] Münzl, Gerald; Pauly, Michael; Reti, Martin: Cloud Computing als neue Herausforderung für Management und IT. Berlin (Springer Vieweg), 2015.

[10] Naujoks, Uwe: Das BCMS in Theorie und Praxis. In: Grete, Patrick; Naujoks, Uwe (Hrsg.) (2023): Arbeitsbuch Business Continuity und Notfallmanagement in Banken. Heidelberg (FCH), 2023.

[11] Romeike, Frank: Sicherung globaler Wertschöpfungsnetze durch wirksames Risikomanagement und BCM. In: Grete, Patrick; Naujoks, Uwe (Hrsg.) (2023): Arbeitsbuch Business Continuity und Notfallmanagement in Banken. Heidelberg (FCH), 2023.

[12] Stanoevska-Slabeva, Katarina: Grid and Cloud Computing. A business perspective on technology and applications. Heidelberg (Springer), 2010.

[13] W3Techs: Ranking der Top 10 Content-Management-Systeme (CMS) weltweit nach Marktanteil im Februar 2024, In: Statista. https://de.statista.com/statistik/daten/studie/320670/umfrage/marktanteile-der-content-management-systeme-cms-weltweit/, (Abruf: 11.02.2024).

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Angriffserkennung mit IDS in Energieanlagen – Praxiseinsatz und Penetrationstests (ein Erfahrungsbericht)

Eric Heindl[1], Benjamin Teudeloff[2]

Kurzfassung:
In den vergangenen Jahren haben sich vermehrt Cyberangriffe auf Kritische Infrastrukturen, darunter auch auf Energieerzeuger und Umspannwerke, ereignet. Der Mythos, dass Angreifer wenig Verständnis von Steuerungssystemen in Energieanlagen haben, wurde spätestens nach den Angriffen auf das ukrainische Stromnetz in den Jahren 2015, 2016 und 2022 sowie den jüngsten Attacken auf mehrere dänische Energieversorger widerlegt. Aufgrund dieser Entwicklungen sind zusätzliche Schutzmechanismen erforderlich, um die Sicherheit dieser kritischen Anlagen und des Stromnetzes zu gewährleisten. Dieser Beitrag beleuchtet, wie Systeme zur Angriffserkennung (IDS) in kritischen Anlagen wie Leitstellen und Umspannwerken optimal eingesetzt werden können. Dabei werden sowohl die rechtlichen Rahmenbedingungen als auch die Anwendungsbereiche von IDS-Systemen betrachtet. Im Anschluss werden diese Erkenntnisse anhand eines Praxiseinsatzes und eines Penetrationstests veranschaulicht.

Stichworte: Energieanlagen, ICS/OT-Sicherheit, IDS, Incident Response, Intrusion Detection System, IoT-Sicherheit, Kritische Infrastrukturen, Resilienz, eingebettete Systeme, Stromnetz

1 Rechtliche Rahmenbedingungen

Zur Sicherstellung der Versorgungssicherheit müssen Betreiber von Energieversorgungsnetzen und Energieanlagen gemäß §11 Abs. 1e des Gesetzes über die Elektrizitäts- und Gasversorgung (EnWG), in angemessener Weise Systeme zur Angriffserkennung einsetzen. Zu schützen sind hierbei die informationstechnischen Systeme, Komponenten oder Prozesse, die für die Funktionsfähigkeit der betriebenen Energieversorgungsnetze oder Energieanlagen **maßgeblich** sind. [3]

Maßgeblich für das Kerngeschäft der Energieversorgung ist vor allem die Primär- und Sekundärtechnik eines Energieversorgungsunternehmens (EVU). Also alle Betriebsmittel, die für Energieerzeugung, -verteilung, -transport und deren Steuerung sowie Überwachung erforderlich sind. Diese gilt es gegen Cyberangriffe und Störungen abzusichern.

---

[1] OMICRON electronics GmbH, 6833 Klaus, AUSTRIA
[2] OMICRON electronics GmbH, Berlin
[3] https://www.gesetze-im-internet.de/enwg_2005/__11.html

Dazu zählen im Bereich der relevanten Komponenten beispielsweise Automatisierungs- und Leittechniksysteme, Schaltanlagen, Schutz- und Steuergeräte und RTUs. Solche Systeme sind beispielsweise Leitstellen-Clients, Admin-Clients, Wartungs- und Prüf-Clients sowie dem OT-Netzwerk vorgelagerte Router, Switche und Firewalls.

Um den besonderen Anforderungen eines OT-Netzwerks gerecht zu werden, sollte ein System zur Angriffserkennung alle oben genannten Komponenten und Systeme, deren Kommunikation und die hierzu genutzten OT-Protokolle detailliert überwachen können.

Neben der technischen Implementierung eines Systems zur Angriffserkennung muss ebenso die organisatorische und prozessuale Einbindung erfolgen. Dies umfasst insbesondere fachlich qualifiziertes Personal und die (Weiterentwicklung der) Prozesse des Incident-, Risiko- und Business-Continuity-Managements.

Zur Erreichung und Aufrechterhaltung der teils komplexen Anforderungen eines Systems zur Angriffserkennung kann die Orientierungshilfe des BSI[4] sowie deren Umsetzungsgradstufen genutzt werden.

2 Angriffserkennung im Stromnetz - Einsatzgebiete von IDS-Systemen

Vor der Implementierung eines IDS sollten einige wichtige Fragen geklärt werden:

- Was soll konkret überwacht werden?
- Welches IDS bzw. welcher Detektionsmechanismus soll verwendet werden?
- Was sind die Vor- und Nachteile eines IDS?

2.1 Unterscheidung IT und OT

Ein verbreiteter Irrtum besteht darin, dass für das OT-Netzwerk, in dem Schaltbefehle und Kommandos übertragen werden, dieselbe Angriffserkennung verwendet werden könne wie für das IT-Netzwerk eines Unternehmens. OT- (Operational Technology) und IT- (Information Technology) Netzwerke unterscheiden sich jedoch sehr stark in ihrer Funktionalität und in ihren Anforderungen.

IT-Netzwerke sind in erster Linie auf die Verarbeitung, Speicherung und Übertragung von Informationen ausgerichtet. Sie unterstützen Büroanwendungen, E-Mail-Kommunikation und andere datenbezogene Aufgaben. Im

[4] https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KRITIS/oh-sza.pdf?__blob=publicationFile&v=15

Gegensatz dazu sind OT-Netzwerke speziell darauf ausgerichtet, industrielle Prozesse und physische Abläufe zu steuern. Diese Netzwerke sind häufig in Kritischen Infrastrukturen wie Energieerzeugung, Fertigung und Transport zu finden. OT-Systeme überwachen und steuern Hardwarekomponenten wie Transformatoren, Freileitungen und Leistungsschalter.

In Kritischen Infrastrukturen müssen die meisten Systeme und Komponenten stets verfügbar sein, was die Sicherheitsanforderungen und die Anforderung an den Schutz dieser Anlagen deutlich erhöht.

Die folgende Abbildung zeigt den Unterschied der Priorisierung der grundsätzlichen Anforderungen zwischen OT und IT.

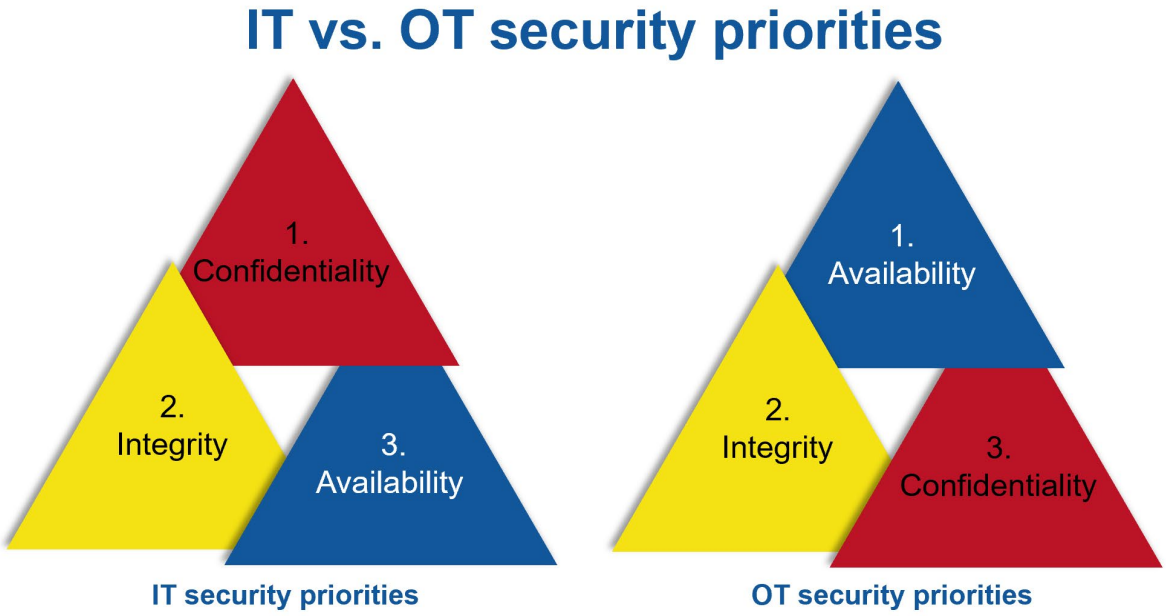


Abbildung 1: Unterschied der Prioritäten zwischen IT und OT

OT-Netzwerke müssen oft besonders robust und sicher sein, da ein Ausfall oder eine Störung schwerwiegende physische Konsequenzen haben kann. Aus diesem Grund ist in diesen Netzwerken die Verfügbarkeit (Availability) an höchster Priorität, gefolgt von Integrität (Integrity), welche die Korrektheit der Daten voraussetzt. Die Vertraulichkeit (Confidentiality) wird in einem OT-Netzwerk häufig am geringsten priorisiert. Folglich kommunizieren Schalt- oder Steuerbefehle meist unverschlüsselt und im Klartext im Netzwerk.

In der IT liegt der Fokus auf dem Schutz von Informationen und informationsverarbeitenden Systemen. Demzufolge ist die Vertraulichkeit von höchster Bedeutung, denn ein Datenabfluss ist eine der häufigsten Bedrohungen in der IT. Während Systemausfälle in der IT vor allem zu finanziellen Schäden

für das Unternehmen führen, können Störungen oder Ausfälle von OT-Komponenten zu weitrechenden Folgen für die Versorgungssicherheit führen.

2.2 Methoden der Angriffserkennung eines IDS

Die Kernkomponente eines IDS ist dessen Detektionsmechanismus, also die Komponente zur Erkennung von Angriffen. Dazu gibt es verschiedene Ansätze, um Attacken bestmöglich erkennen zu können:

Baseline-/Lernbasierter Ansatz: In der Lernphase werden bestimmte Protokollfelder beobachtet und anhand dieser Beobachtungen lernt das System, was innerhalb dieses Netzwerks als übliches Verhaltensmuster gilt. Dies hat den Vorteil, dass potenziell auch unbekannte Angriffsmuster durch diesen Mechanismus erkannt werden können. Es bedeutet aber auch, dass alle Aktionen, die in der Lernphase nicht vorgekommen sind, zum Beispiel Schaltvorgänge oder Wartungsaktivitäten, einen Alarm auslösen werden. Darüber hinaus besteht das Risiko, dass auch unerwünschte Kommunikation als Baseline erlernt werden kann.

Signaturbasierter Ansatz (Denylist): Bei diesem Ansatz, der oftmals auch von Virenscannern genutzt wird, sucht das IDS nach Mustern, welche es bereits von anderen Angriffen kennt. Systeme dieser Art haben eine geringere Fehlalarmquote als Systeme auf der Basis eines lernbasierten Ansatzes. Allerdings ist es Angreifern leicht möglich, die Muster jedes Mal geringfügig zu ändern und damit einer Erkennung auszuweichen. Um dem entgegenzuwirken, müssen die Signaturen entsprechend "offener" gestaltet werden, was wiederum zu mehr Fehlalarmen führt.

Allowlist-Ansatz von StationGuard: In OT-Netzwerken, insbesondere im Stromnetz, ist das Verhalten aller Geräte im Netzwerk klar definiert. Das IDS kann deswegen ein Live-Modell der zu überwachenden Betreiberanlage erstellen. Innerhalb des OT-Netzwerks und der Kommunikation der Netzwerkteilnehmer (zum Beispiel Router/Switche, Clients, Schalt- und Steuergeräte) werden alle Netzwerkpakete entsprechend des Allowlist-Ansatzes verglichen. Damit werden Abweichungen von zulässigem Netzwerkverhalten erkannt und Alarme ausgelöst. Dieser Allowlist-Ansatz überwacht auch Signalwerte in den Nachrichten, was eine präzise Erkennung von Cyberbedrohungen und Problemen bei den Funktionen ermöglicht. Der Allowlist-Ansatz wird darüber hinaus durch eine Anomalieerkennung sowie Signaturerkennung ergänzt.

2.3 Installation von IDS im Netzwerk

Nachdem die passende Methode zum Erkennen der Angriffe ermittelt wurde, stellt sich die nächste entscheidende Frage: Wo im Netzwerk soll das System installiert werden? Diese Überlegung ist von zentraler Bedeutung, da sie beeinflusst, welche Netzwerke und Komponenten überwacht werden können. Ein häufiges Szenario ist, dass die Leitstelle von den Umspannwerken physisch und logisch getrennt ist. Diese räumliche Trennung erschwert das netzwerkübergreifende Monitoring und erfordert daher eine sorgfältige Planung.

Um den optimalen Einsatzort für das IDS-System zu bestimmen, sollten folgende Aspekte berücksichtigt werden:

- Netzwerksegmente: Welche Bereiche des Netzwerks sollen überwacht werden? Sind es nur bestimmte Subnetze oder das gesamte Netzwerk?
- Verbindungspunkte: Wo sind die geeigneten Verbindungspunkte im Netzwerk? Dies kann an Router- oder Switch-Schnittstellen, an Firewalls oder anderen kritischen Systemen sein.
- Zugriffsrechte: Welche Zugriffsrechte sind erforderlich, um das IDS-System zu installieren und zu konfigurieren? Dies sollte im Einklang mit den Sicherheitsrichtlinien des Unternehmens stehen.
- Skalierbarkeit: Ist das Netzwerk skalierbar? Wenn ja, muss das IDS-System in der Lage sein, mit dem Wachstum des Netzwerks Schritt zu halten.

Die Wahl des richtigen Einsatzorts ist entscheidend für die Effektivität des IDS und die Sicherheit des gesamten Netzwerks. Eine sorgfältige Planung und Abstimmung mit den beteiligten Teams sind daher unerlässlich.

2.4 Vorteile und Herausforderungen der Nutzung eines IDS in OT-Netzwerken

Vorteile:

- Anomalieerkennung: Ein IDS kann ungewöhnliche Aktivitäten oder verdächtige Muster im Netzwerk erkennen, bevor sie zu ernsthaften Sicherheitsvorfällen führen.
- Echtzeitüberwachung: IDS bieten Echtzeitüberwachung, um auf Bedrohungen sofort reagieren zu können. Dies ist entscheidend für die Sicherheit von OT-Netzwerken.
- Inventarisierung: Ein IDS unterstützt die automatische Identifikation, Überwachung und Dokumentation des im OT-Netzwerk vorhandenen Anlageninventars.

- Funktionsüberwachung: Das OT-Netzwerk sowie deren Betriebsmittel und Assets werden bezüglich ihrer sicheren Funktionsweise überwacht. Netzwerkfehler, Fehlkonfiguration von Geräten sowie fehlerhafte Protokollübertragungen können unmittelbar festgestellt werden.
- Visualisierung: Übersichtliche Dashboards, die Visualisierung des Netzwerks und der Netzwerkkommunikation, sowie die für IT- und OT-Mitarbeiter verständliche Darstellung von Ereignis- sowie Alarmmeldungen.

Herausforderungen:

- Komplexität der Netzwerke: OT-Netzwerke sind oft komplex und heterogen. Sie bestehen aus verschiedenen Geräten, Protokollen und Technologien. Die Implementierung eines IDS erfordert daher eine genaue Kenntnis der Netzwerktopologie.
- Fehlalarme: IDS können Fehlalarme auslösen, wenn sie normale Aktivitäten fälschlicherweise als Angriffe interpretieren. Dies erfordert eine sorgfältige Konfiguration und Feinabstimmung.
- Performance-Auswirkungen: Ein IDS kann die Netzwerkperformance beeinträchtigen, insbesondere wenn es auf älteren oder ressourcenbeschränkten Systemen läuft.

Insgesamt ist festzustellen, dass ein OT-spezialisiertes IDS für Energiebetreiber unerlässlich ist, um die Verfügbarkeit in Kritischen Infrastrukturen zu gewährleisten. Durch die gezielte Überwachung des OT-Netzwerks und die frühzeitige Detektion von Anomalien können Cyberangriffe verhindert oder deren Auswirkungen minimiert werden.

3 Praxisbeispiel: Penetrationstests und Funktionsweise eines IDS

Im Zuge eines Proof of Concept bei einem norwegischen Verteilnetzbetreiber wurden mehrere IDS installiert, konfiguriert und getestet. In diesem Abschnitt wird das zuvor erläuterte Szenario anhand eines konkreten Praxisbeispiels beleuchtet.

Das Ziel des Kunden bestand darin, mehrere OT-IDS hinsichtlich ihres Implementierungsaufwands, ihrer Benutzerfreundlichkeit und ihrer Effektivität eingehend zu prüfen. Hierzu wurden gezielte Cyberangriffe für verschiedene Angriffsvektoren durchgeführt. Der Energieversorger erhielt dabei Unterstützung von Experten einer externen Firma, die den Penetrationstest in Zusammenarbeit mit dem DSO durchführten.

Einige grundlegende Eckdaten: Unser IDS war im Zeitraum von Ende 2022 bis Mitte 2023 bei dem genannten norwegischen DSO im Rahmen einer Proof-of-Concept-Installation aktiv. Dabei wurden sowohl die technische Umsetzung als auch die praktische Wirksamkeit des IDS intensiv geprüft.

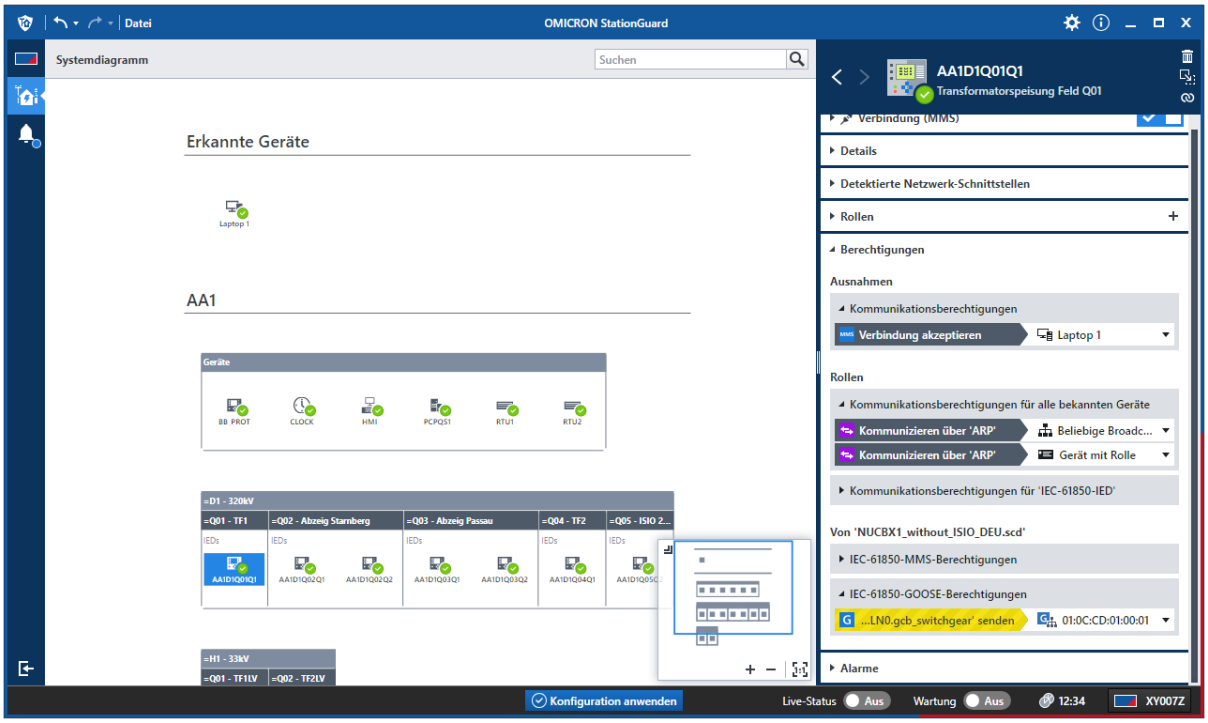


Abbildung 2: Systemdiagramm der betroffenen Anlage

Der Penetrationstest fand zwischen dem 18. und 19. April 2023 statt. Vor Beginn des Tests wurden seitens des DSO einige Konfigurationen an Routern und Switchen durchgeführt, um einen reibungslosen Ablauf zu gewährleisten. Die beteiligten IDS-Anbieter wurden nicht über den Zeitpunkt des Pentests oder über die Art und den Umfang der Angriffsszenarien informiert.

Phasen der Detektion des Penetrationstests: Eine deutliche Indikation für den Start des Tests war das Detektieren **neuer IP-Adressen im Netzwerk**.

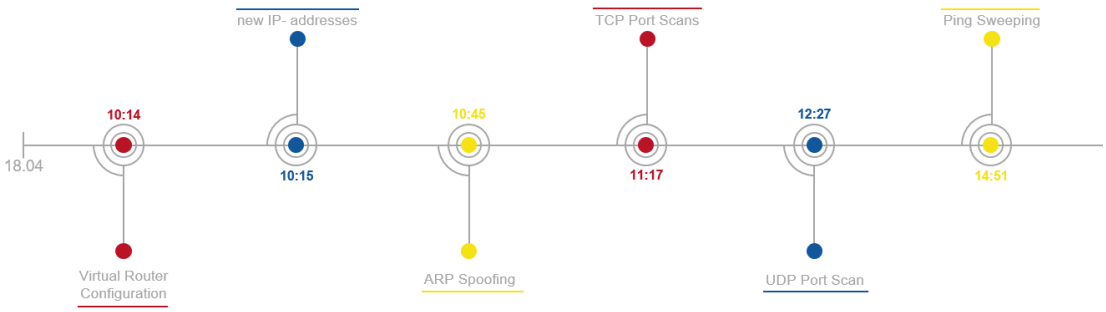


Abbildung 3: Pentest Timeline

Im Gegensatz zur IT stellen unbekannte IP-Adressen in der OT eine potenzielle Bedrohung dar. Diese sind einer der besten Indikatoren für die frühzeitige Erkennung bestimmter Angriffe.

In der nächsten Phase des Angriffs wurden verschiedene **Scans** durchgeführt. Dies ist ein übliches Verhalten, wenn der Angreifer das Netzwerk genauer verstehen und sich einen Überblick über den Aufbau des Netzwerks und dessen Netzwerkteilnehmer verschaffen will.

Zuerst wurden mehrere TCP Port Scans ausgeführt. Dabei versuchen Angreifer zum Beispiel offene Ports, verwendete Dienste oder Betriebssysteme sowie technische Schwachstellen zu identifizieren. Port Scans sind gut über die Auslastung des Netzwerkes erkennbar. Durch den Allowlist-Ansatz im IDS sind solche Verbindungen ebenfalls verboten und führen zur Alarmierung. Die folgende Abbildung zeigt eine Timeline der Alarme, welche durch diese Attacken ausgelöst worden sind.

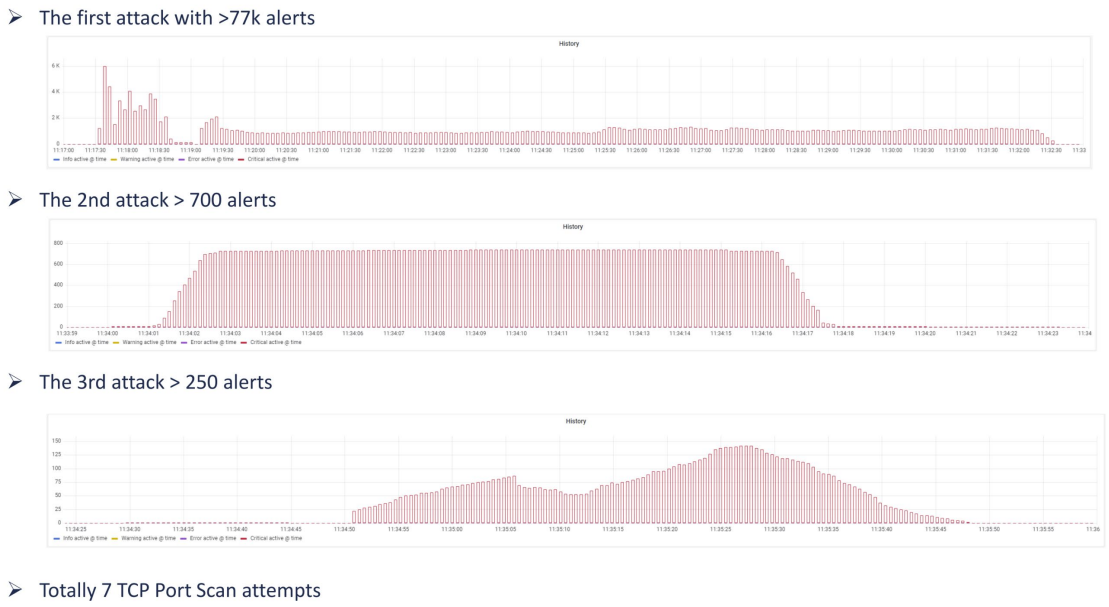


Abbildung 4: Timeline der TCP Port Scans

Im weiteren Verlauf wurden ein **UDP Port Scan** und einige **Ping-Sweeps** von dem IDS detektiert.

Durch einen UDP Port Scan versuchen Angreifer Rückschlüsse auf den Gerätetyp der Netzwerkteilnehmer zu ziehen. Also ob es sich beispielsweise um einen Client oder Server, um ein Schutz- oder Steuergerät handelt. Mit einem Ping-Sweep werden alle IP- Adressen oder bestimmte IP-Adressbereiche des Netzwerks kontaktiert. Antwortet eine IP-Adresse, ist dieses Gerät aktiv und eingeschaltet.

Mit Hilfe dieser Informationen können Angreifer ein Mapping der Ports, genutzten IP-Adressen, Dienste, Betriebssysteme und Netzwerkteilnehmertypen vornehmen.

In der nächsten Phase des Angriffs wurde eine **ARP-Spoofing** Attacke ausgeführt. Hierbei nutzen die Angreifer die fehlende Authentifizierung zwischen den Netzwerkteilnehmern im ARP-Protokoll aus.

Das ARP-Protokoll dient der Zuordnung von IP- zu MAC-Adressen im lokalen Netzwerk. Hierbei kommt ein Request-Response-Schema zum Einsatz, bei dem die erste Antwort (ARP-Response) akzeptiert und in einer ARP-Tabelle gespeichert wird. Dieser Mechanismus birgt jedoch das Risiko, dass Angreifer die Antwort fälschen und sich als der gesuchte Netzwerkteilnehmer ausgeben. Auf diese Weise können Angreifer ihre MAC-Adresse mit der IP-Adresse eines Zielhosts verknüpfen. Dies ermöglicht es ihnen, den Datenverkehr, der eigentlich für den Zielhost bestimmt ist, auf das kompromittierte System umzuleiten und zu manipulieren.

Somit entsteht ein manipuliertes IP-zu-MAC-Adressen-Mapping, welches dem Angreifer erlaubt eine **Man-in-the-Middle** Attacke zu initialisieren. Die folgende Abbildung zeigt die Detektion und Alarmierung dieses Verhaltens im Zeroline Diagramm:

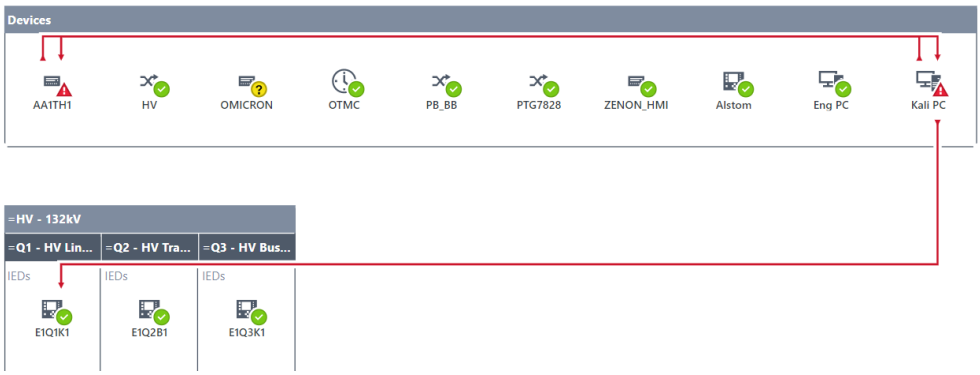


Abbildung 5: ARP Spoofing im Zeroline Diagramm

Als nächstes wurde eine **UDP-Traceroute**-Attacke ausgeführt. Ähnlich wie bei den vorherigen Scan-Attacken wird diese Methode genutzt, um den Netzwerkaufbau genauer zu untersuchen. Dabei wird versucht, durch gezieltes Setzen der TTL (Time-to-live) Variable, den Netzwerkpfad und die Netzwerkteilnehmer samt IP-Adresse und Hostnamen zu identifizieren.

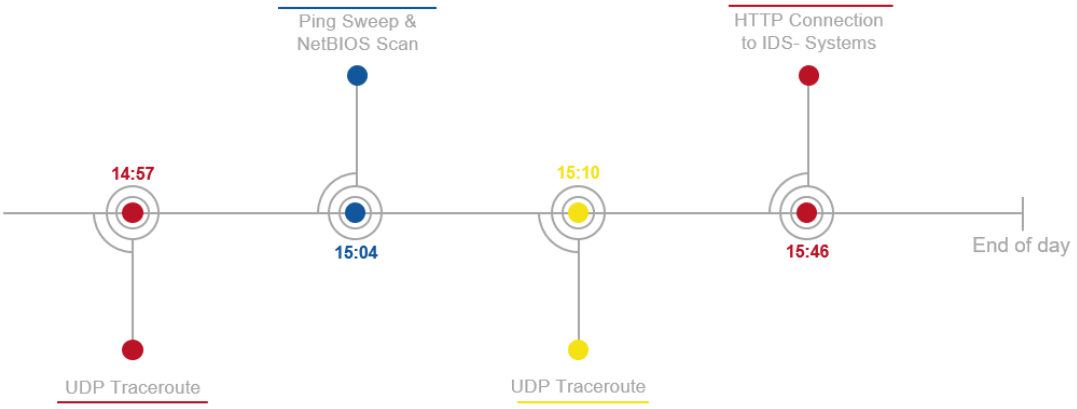


Abbildung 6: Pentest Timeline Teil 2

Mit den Erkenntnissen aus den Scans erfolgte anschließend ein gezielter Angriff auf die Windows-Geräte im Netzwerk. Hierbei wurden ein Ping-Sweep mit einem **NetBIOS-Scan** und Webserver-Verbindungen eingesetzt. Diese Attacke ist besonders interessant, da sie verschiedene Angriffsmethoden kombiniert. Zunächst wurden die Aktivitäten einiger Adressen im Netzwerk per ICMP (Ping) überprüft. Bei erfolgreicher Rückmeldung erfolgten NetBIOS-Anfragen, um die NetBIOS-Informationen der Geräte abzufragen. NetBIOS ist ein Windows Protokoll welches für den Austausch von Daten, zum Beispiel Dateifreigabe oder Drucker-Informationen, verwendet wird. Wenn eine Antwort auf die gesendeten NetBIOS Anfragen empfangen wurde, erfolgten zusätzliche Verbindungsversuche mit möglichen Webservern. Hierbei wurden die Ports 80, 443 und 8080 mithilfe spezieller TCP-Pakete kontaktiert. Anhand der im IDS protokollierten PCAP-Daten konnten die Angriffsmethoden nachvollzogen werden.

Protocol	Info
ICMP	Echo (ping) request id=0x0100, seq=610/25090, ttl=255 (reply in 175)
ICMP	Echo (ping) reply id=0x0100, seq=610/25090, ttl=64 (request in 174)
ICMP	Echo (ping) request id=0x0100, seq=681/43266, ttl=255 (reply in 304)
ICMP	Echo (ping) reply id=0x0100, seq=681/43266, ttl=64 (request in 303)
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
NBNS	Name query NBSTAT *<00><00><00><00><00><00><00><00><00><00><00><00><00><
TCP	61142 → 80 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
TCP	[TCP Retransmission] [TCP Port numbers reused] 61142 → 80 [SYN] Seq=0 W
TCP	61192 → 443 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
TCP	[TCP Retransmission] [TCP Port numbers reused] 61192 → 443 [SYN] Seq=0
TCP	61224 → 8080 [SYN] Seq=0 Win=32 Len=0 MSS=1460 WS=1 SACK_PERM
TCP	[TCP Retransmission] [TCP Port numbers reused] 61224 → 8080 [SYN] Seq=0

Abbildung 7: NetBIOS Attacke in Wireshark

Am zweiten Tag des Pentests wurden diverse OT-Attacken festgestellt.

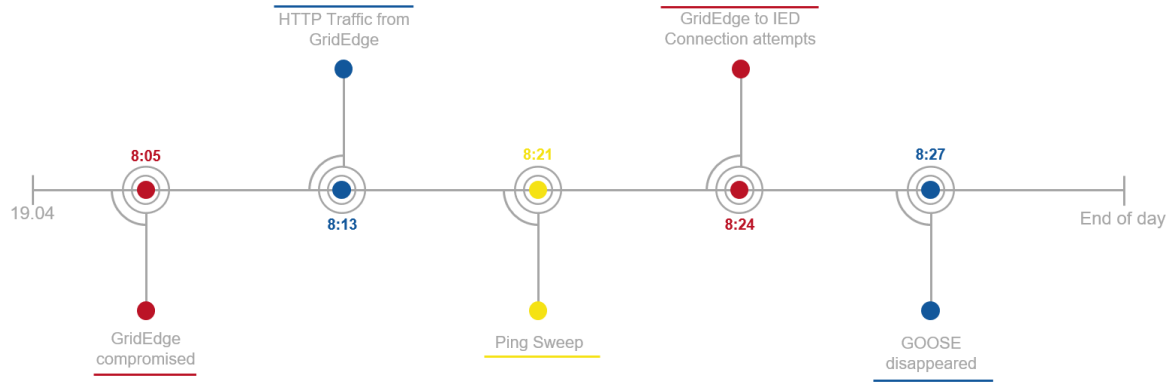


Abbildung 8: Pentest Timeline Tag 2

Zunächst wurde eine **physische Attacke** durchgeführt, bei der eine Hardwarekomponente mit der beim DSO vorhandenen GridEdge funktional kopiert und ausgetauscht wurde. Dabei wurden sowohl die MAC-Adresse als auch die IP-Adresse des GridEdge Gerätes übernommen, inklusive sämtlicher operativer Funktionen.

Das GridEdge bezeichnet eine Komponente, welche für das Sammeln von Konfigurations- und Fehlerdaten, sowie das Erstellen eines Asset-Inventars von Schutzrelais verantwortlich ist.

Das Erkennen solcher physischer Attacken ist sehr kompliziert und bei sorgfältiger Ausführung ist es nahezu unmöglich diese zu erkennen. Im Falle dieses Pentests wurden seitens der Angreifer jedoch Spuren hinterlassen. So wurde dem kompromittierten GridEdge nach dem Austausch für eine sehr kurze Zeit eine APIPA-Adresse (Automatic Private IP Addressing) zugewiesen. Diese Adresse wird von Betriebssystemen verwendet, um selbstständig eine IP-Adresse zu konfigurieren, falls kein DHCP-Server erreichbar ist. Die APIPA-Adresse ist immer in dem Adressbereich 169.254.x.x, welche sich somit von den anderen IP-Adressen bzw. IP-Adressbereichen deutlich unterscheidet und als Anomalie detektiert werden kann.

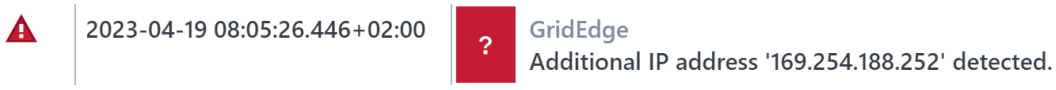


Abbildung 9: Detektion einer APIPA Adresse für GridEdge

Mit dieser Information haben wir anschließend das Verhalten des GridEdge genauer untersucht. Bei der Analyse wurde festgestellt, dass die periodischen Abfragen dieses Gerätes für eine kurze Zeit unterbrochen wurden, was

genau auf diesen Austausch auf das kompromittierte Gerät samt GridEdge-Funktionalität hindeutet.

Nach der Kompromittierung von GridEdge wurden unerlaubte HTTP-Nachrichten an mehrere Switche gesendet. Ein Angriff, der HTTP-Verbindungen von OT-Geräten ausnutzt, ist als äußerst kritisch zu bewerten und deutet stark darauf hin, dass ein Gerät kompromittiert wurde.

Im weiteren Verlauf wurden **Verbindungsversuche zu verschiedenen IEDs** über Port 102 detektiert. Dieser Port wird in der Schutz- und Leittechnik von dem IEC 61850 MMS-Protokoll verwendet und dient zum Austausch von Schaltsignalen, Messwerten, Konfigurationsdaten oder Ereignisberichten. Das IEC 61850 MMS ermöglicht somit die effiziente Kommunikation zwischen den IEDs und den höheren Entitäten wie RTUs und SCADAs.

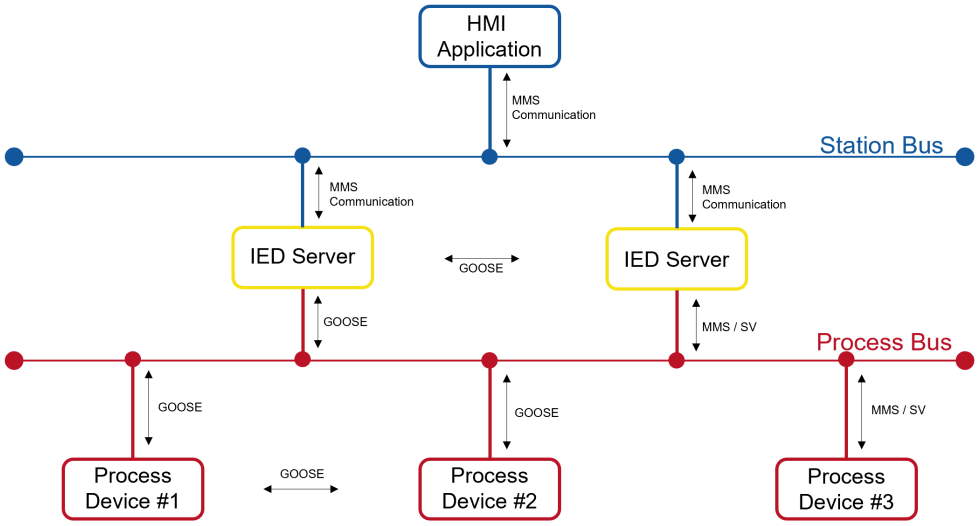


Abbildung 10: Protokollstruktur eines Schutz- und Leittechnik-Netzwerkes

Anschließend wurde mittels eines weiteren IEC 61850-Protokolls angegriffen. In der Analyse war ersichtlich, dass nach dem Verbindungsaufbau auf einigen Switchen GOOSE Nachrichten im Netzwerk nicht mehr sichtbar waren und Statuswerte verändert wurden.

GOOSE wird in Schutz- und Leittechnik Netzwerken für die horizontale Kommunikation zwischen IEDs verwendet. Dazu werden in regelmäßigen Abständen bestimmte Statuswerte sowie Ereignismeldungen (z. B. Auslösesignal, Einschalten des Leistungsschalters) gesendet. Dadurch können andere IED oder SCADA Systeme entsprechende Folgeaktionen ausführen, wie beispielsweise Interlocking.

Bei der ausgeführten Attacke wurden diese GOOSE Nachrichten verändert oder unterbunden. Dadurch sind betriebsrelevante Informationen für die anderen Netzwerkteilnehmer wie beispielsweise aktuelle Statuswerte nicht korrekt oder nicht mehr verfügbar. Dies kann zu Fehlfunktionen, Störungen bis zum Ausfall der Betreiberanlage führen.

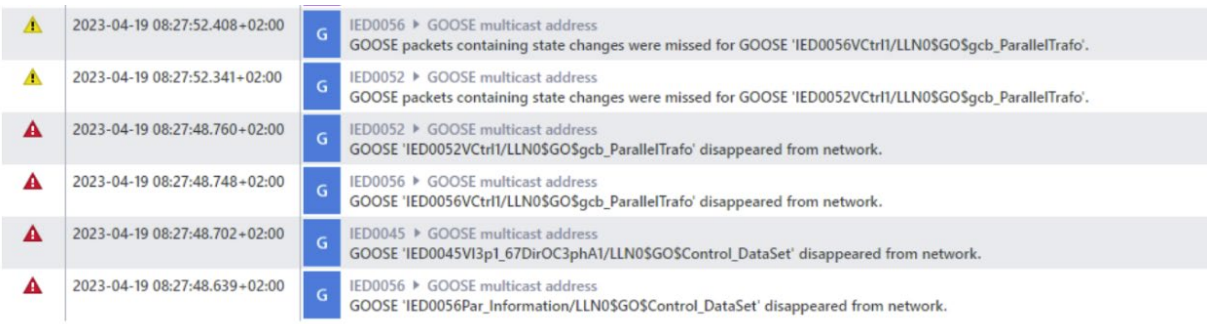


Abbildung 11: Fehlende-GOOSE-Alarme in StationGuard

4 Lessons Learned und Handlungsempfehlungen

Die Analyse des Penetrationstests verdeutlicht erneut die Notwendigkeit und Effektivität eines Systems zur Angriffserkennung (IDS) bei der Identifizierung von Cyberangriffen. Insbesondere wurde anhand der visualisierten Angriffe deutlich, dass viele dieser Bedrohungen nur durch ein auf Operational Technology (OT) spezialisiertes IDS wie StationGuard detektierbar sind. Eine zentrale Erkenntnis aus dem Praxistest ist, dass bereits ein einzelner Alarm oft ein Indikator für einen potenziellen Angriff sein kann.

IDS können Netzwerke in Echtzeit überwachen und dabei helfen, ungewöhnliche Aktivitäten sofort zu erkennen. Zusätzlich zu einem IDS empfehlen wir jedem Energieversorger, sein Netzwerk korrekt zu konfigurieren und zu segmentieren, auch indem nicht benötigte Dienste oder Applikationen deaktiviert werden. Durch diese Maßnahmen wird der Angriffsvektor erheblich reduziert und potenzielle Attacken werden deutlich erschwert. Dies trägt maßgeblich dazu bei, die Sicherheit im Energieversorgungsbereich zu stärken und Kritische Infrastrukturen vor Cyberbedrohungen zu schützen.

Secure IoT Communication in Critical Infrastructure Using Attribute Based Access Control

Moritz Volkmann[1], Sascha Kaven[1], Kai-Hendrik Wöhnert[1], Volker Skwarek[1]

Abstract: In this work, the authors combine the technologies of machine learning based intrusion detection, self sovereign identity, and attribute based access control to create a framework for the secure energy community of the future with active grid management and peer-to-peer flexibility trading.

Keywords: ABAC, IoT, SSI

1 Introduction

The transition of energy grids from traditional centralized systems to decentralized smart grids is a key development toward a sustainable energy future [4]. The traditional centralized energy grid is characterized by a predominantly unidirectional flow of electricity from large power plants to consumers. In contrast, a decentralized smart grid allows the bidirectional flow of electricity, so that excess energy generated by local renewable sources can be stored and shared among consumers. While the electrical grid is still regulated as centralized grid, today it is often operated decentrally through the increasing amount of renewable energy sources and consumers like electric vehicles [26]. This regularly drives the grid to the limits of power stability, with the regulated fallback of centralization.

As a strategy for a more stable decentralized grid, households take on the role of both consumer and producer and act as a prosumer. This enables greater flexibility in energy management, reduces energy waste, and promotes the use of renewable energy sources [19]. Furthermore, the decentralization of the energy market allows prosumers to interact directly in a local peer-to-peer (P2P) market. P2P energy trading is made possible by the use of digital platforms that allow participants to communicate and exchange energy securely and efficiently [7].

To enable safe P2P energy trading and grid stability, an untampered identity of market participants must be secured to ensure authenticity for preventing market manipulation and attacks. In addition, P2P trading requires constant availability of data that serves as the basis for trading. In the future energy community, such data is provided not only by the prosumers themselves via

[1] Forschungs- und Transferzentrum CyberSec, Hochschule für Angewandte Wissenschaften (HAW) Hamburg

their smart meter gateways, but also by IoT sensors in the grid itself, such as the cable distribution stations or grid substations. However, with a grid that is communicating frequently and through various channels with the distribution service operator (DSO), protecting these data sources from attacks is essential to ensure the integrity of the P2P energy market [25]. This also leads to new security challenges since traditional network security mechanisms have weaknesses in providing security assurance in such a complex and dynamic network environment. A main weakness this paper intends to overcome are CWE-284: improper access control and CWE-732: incorrect permission assignment for critical resource. Both weaknesses include unauthorized actors causing harm in the system, which can have a big impact in energy communities.

This paper aims to answer the question: ***How can security measures be integrated into energy communities to ensure a stable energy supply and secure P2P trading?***

The paper will be structured as follows: section 2 describes the current state of the art, section 3 will lay out the scenario and use case, section 4 features the developed architecture, section 5 contains the implementation and evaluation of the architecture, and section 6 will finish with a conclusion and recommendations.

2 State of the art

This section features the state-of-the-art methodologies to be employed in the future energy community, namely intrusion detection and identity management concepts and their usage in current literature works regarding smart grid systems.

2.1 Security in smart grid

The smart grid is susceptible to a myriad of cyber-attacks that exploit its connectivity and reliance on digital technologies. Relevant attacks on smart grid in literature are denial of service (DoS) attacks [14], intrusion attacks such as user to root or remote to local attacks [11], and false data injection [12]. Also occurring are replay attacks [18], port scanning [11], and faking a fault of a resource by manipulating the power measurements being reported [17].

In response to these escalating threats, robust intrusion detection systems have become paramount in safeguarding smart grid environments. Machine learning-based methods exhibit high accuracy in identifying both known and unseen threats. The most common approach is ensemble learning, where

several weak learners are combined. [9] use the XGBoost framework for gradient boosting and [15] uses random forest. [11] takes the resource limitations of participating IoT devices within the smart grid into consideration by creating a lightweight boosting model. However, approaches with higher resource usage, such as deep learning, are used as well [20, 10]. The third common method are support vector machines [17]. Smart grids are decentralized and consist of heterogeneous devices from low-resource devices built into the local energy grids, over more powerful edge devices that function as a gateway between local energy communities and the network control center, finally to a central server at the DSO's network control center. Attacks that may find their entry point at low-resource devices, such as the mentioned injection attacks, can have high impact to the overall electrical grid if not detected at an early stage, resulting in the state estimation outputting faulty data [25]. Therefore, an early detection of attacks is essential. An IDS needs to monitor the overall network traffic, from small devices to the network control center.

Recently, the NIS 2 directive from ENISA and the European union [24] has laid out requirements for cyber security in the domain of critical infrastructure, which holds companies and organizations accountable for violations. These requirements also active intrusion detection and identity management. In the light of the upcoming implementation of NIS2 into national laws, this work provides a suggestion as to how some of those requirements can be satisfied.

2.2 Identity management and access control in smart grid

In recent years, the integration of Self-Sovereign Identity (SSI) principles into smart grid systems has emerged as a promising avenue for addressing various challenges associated with identity management and security. SSI represents a paradigm shift from centralized identity models to user-centric and decentralized approaches. This shift is particularly relevant in the context of smart grid systems, where secure and efficient identity management plays a pivotal role in ensuring the reliability, privacy, and resilience of the grid infrastructure. As a pivotal stakeholder in critical infrastructure, the DSO bears the crucial responsibility of maintaining equilibrium between energy demand and supply. Consequently, it is imperative that only reliable prosumers are allowed to participate in the energy delivery system. The subsequent literature offers insights into strategies for guaranteeing such reliability through the application of SSI principles.

Cali et al. [3] propose an architecture for local energy markets with an SSI identity management system. Similarly, Volkmann et al. [21] propose a framework for local energy markets, in which SSI is used for the trading authorization process in the peer-to-peer market. Dehalwar et al. [6] also endorse the usage of SSI in smart grid systems.

Given the imperative requirement for rigorous supervision of a P2P energy market by the DSO to uphold network stability, the incorporation of secure and self-sovereign identity management can be complemented by robust access control paradigms. Various research efforts have pursued similar objectives, exemplified by the works of Ma et al. [13] and Hong et al. [8]. Notably, Belchior et al.'s proposition of combining SSI with attribute-based access control [2] emerges as an ideal solution tailored to the specific demands of this scenario, which will be implemented in this work.

3 Use case

This work assumes a local energy community in the near future, such as the year 2030, specifically a group of prosumers in the low-voltage grid and the devices and infrastructure of the low-voltage grid itself, such as the cable distribution stations. As shown in figure 1, the grid operator uses cloud-based active grid control paradigms to monitor and control the 5G/6G enabled cable distribution stations and to communicate with the smart meter gateways of the prosumers. The prosumers are interconnected via a secure P2P energy market system to trade excess energy – depicted as a green arrow - as described in [21], where SSI technologies combined with attribute-based access control [2] are used to ensure privacy, trust, and secure trading between the peers. The SSI communication channels are marked with blue connections. Moreover, IoT systems are utilized in various ways inside the network to monitor energy production, consumption, and network activity. While the aforementioned technologies offer advantages in grid management, power efficiency, and general digitization, they also open attack surfaces to potential attackers.

The following section will describe the network architecture and its components, focusing especially on the security features that guarantee the reliable operation of the critical infrastructure.

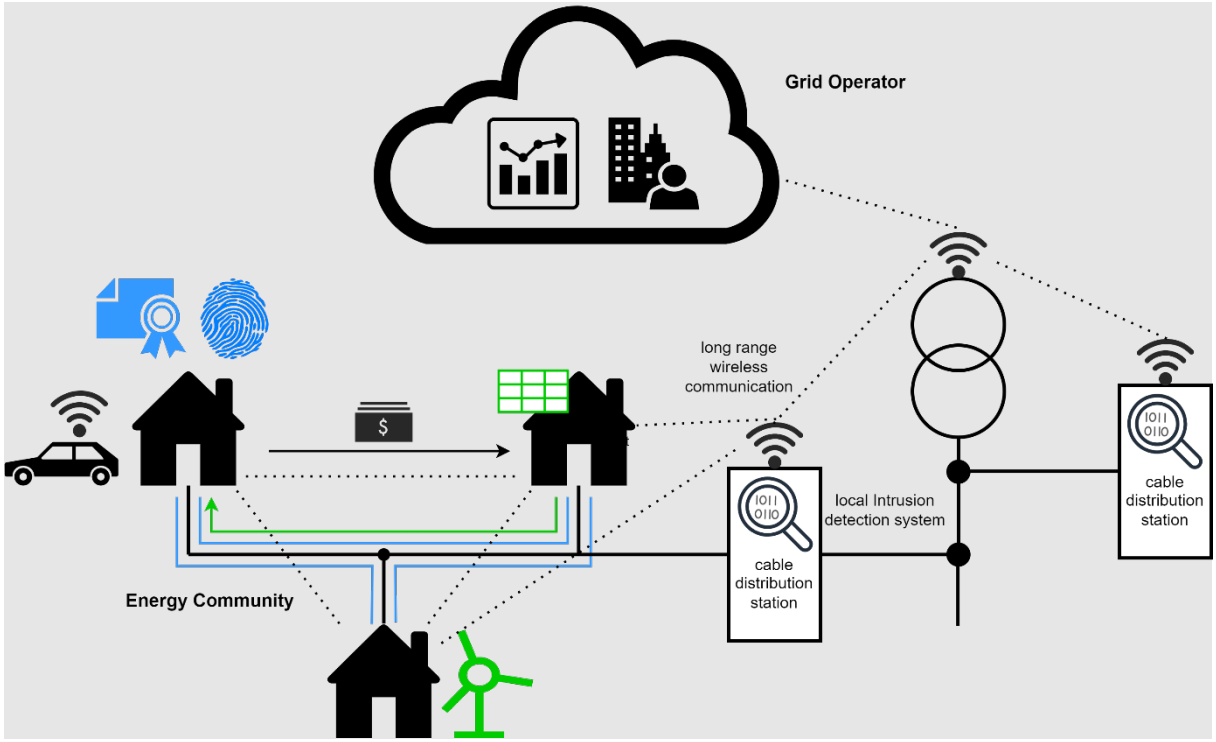


Figure 1: Overview of a communication and power distribution network.

4 Network architecture design for future energy networks

To ensure the security of the grid, it is paramount to envision a network architecture that is not only robust but can preemptively mitigate potential cyber threats.

The inclusion of private households as trading partners makes the task of reliably verifying the identities of each entity crucial. The usage of SSI enables prosumers to control their verifiable credentials which contain all the relevant information for participating in P2P trading activities while preserving prosumer privacy by enabling selective disclosure and verification through zero-knowledge proof [5].

However, these digital credentials are not limited to identity verification alone. The information embedded within these verifiable credentials can be strategically leveraged to enhance network security. Attributes contained within the credentials serve as a basis for instituting an Attribute-Based Access Control (ABAC) model. This model facilitates a granular and precise rights management.

By employing ABAC, the network benefits from increased confidentiality and integrity. It ensures that access permissions are meticulously allocated based on distinct user attributes, significantly minimizing the potential for

unauthorized access and ensuring a more secure, efficient, and reliable network.

In essence, the comprehensive and strategic use of SSI and ABAC not only solidifies the security foundation of the network but also enhances the efficiency and reliability of the energy trading process, ensuring a seamless and robust platform for all participating entities. To achieve this, additional security measures need to be implemented on the distribution grid level.

Given the indispensable role of early detection in curtailing the adverse impacts of cyber-attacks, it is proposed that IDS are integrated even into low-level devices with high resource limitations such as restricted computing power and memory capabilities. This approach is chosen to leverage the benefits of rapid response and build a continuous monitoring system rooted at the foundation of the network's infrastructure. The proposed architecture deploys a lightweight federated machine-learning-based IDS to complement the IDS on higher-hierarchy devices with supplemental information from the lower level devices while using the computational power of the higher-hierarchy device to offset the lower level devices' resource constraints. Central to this IDS is a machine-learning model structured as an ensemble learner composed of weak learners. For efficient classification, feature selection using correlation and principal component analysis is implemented. Additionally, a central component is also needed that can detect an attack on the whole network. This central component is in the hands of the DSO responsible for the distribution grid of the energy community.

Figure 2 shows an architectural overview of the components utilized in the energy community that is envisioned in this work. The different platforms, namely the P2P trading platform with its SSI agent, the P2P trading order book and the ABAC system, which will be described in more detail in section 4.2 and the Hyperledger Indy Blockchain responsible for the hosting of the SSI communications, as well as the storing of the verification registry for the verifiable credentials, the actors such as prosumers with their own SSI and energy trading agents and the DSO, who is responsible for state estimation and congestion management, as well as the grid components like power stations, grid substations and cable distribution stations, are depicted as components of the architecture with their software features depicted as subcomponents.

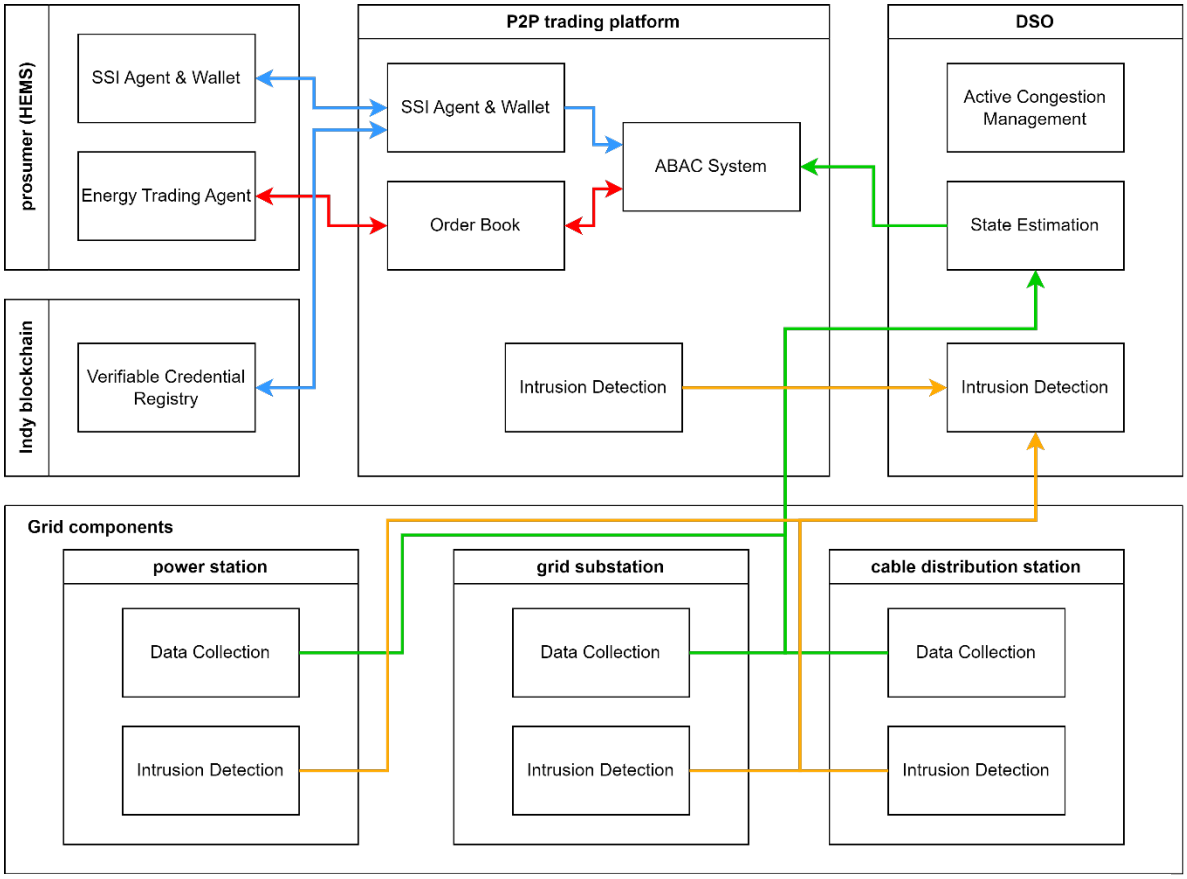


Figure 2: Component diagram of the distribution grid systems

The communication between the components and subcomponents is shown in colored arrows to distuingish the communication purposes and contents from one another. Marked with yellow arrows, the federalized IDS can be seen as a subcomponent of nearly all grid components communicating with the central IDS component in the hands of the DSO. The green arrows depict the grid data that is collected by the IoT devices in the grid components and the P2P trading platform, which is then sent to the DSO, where the data is used to conduct the grid state estimation, which is then returned to the P2P trading platform in the form of the aforementioned traffic light signal. The blue arrows show the SSI communication between the consumers, the P2P trading platform and the Hyperledger Indy blockchain. The SSI communication is described in detail in section 4.1. The red arrows symbolize the trade communication channel between the prosumers and the P2P trading platform. The intersection between state estimation, SSI communication and trade communication can be seen at the ABAC system, which is responsible for deciding whether a trade offer can be published to the P2P market order book. This decision process will be described in more detail in section 4.3.

4.1 Verifiable Credential trading licence

To ensure the security of the P2P market not only in a cyberphysical way but also regarding the security of energy delivery, each participant of the P2P market needs to be trustworthy not only for the DSO, but also for the prosumer who buys their energy. This is why the P2P market utilizes an SSI identity management system with verifiable credentials (VC) that all market participants are issued so they can prove that they are able to supply their energy offers and held accountable if they try to act in a malicious manner. The authentication process of each prosumer to become a trusted market participant is conducted along the description by Volkmann et al. [21] and is out of scope for this work, which will instead focus on the authorization process for offering energy on the P2P market.

Verifiable Credentials are certificates with claims that can be verified with zero-knowledge proof against the verification registry of the SSI blockchain [5]. They are one of the core components of SSI and offer various opportunities for P2P systems, where prosumers need to be able to trust each other without knowledge of their peers. With a VC trading license, a prosumer can prove that they can indeed supply the energy that they want to offer on the P2P market by providing a zero-knowledge proof of the claim "maximum production" that was issued to them by the DSO and therefore has been validated by a trusted source. In the case of malicious activity or expiry, VCs can be revoked by the DSO, and the prosumer cannot provide the zero-knowledge proof anymore, since the VC will be erased from the credential registry on the blockchain immediately, barring the holder from trading and ensuring security in the market [22].

4.2 Attribute representation

In the context of a Self-Sovereign Identity (SSI) based ABAC system the authorization of P2P trades, the attributes from the trading license VC, the state estimation and the order book are the deciding factors for the authorization decision. These attributes are categorized into four main types, subject, object, action and environment attributes. In the following, the relevant attributes will be listed and categorized to lay the foundation of the decision process.

Subject attributes: The subject attributes of the prosumers participating in the P2P market encompass their maximal energy production, their registered power production units, their name and address and the validity of their trading license. Moreover, their trust level derived from past transactions, as well as their geographical location for consideration of local

trading constraints are considered subject attributes. The subject attributes are contained in the trading license VC.

Object attributes: Related to the digital and physical resources within the trading ecosystem, object attributes cover energy data (production, consumption, storage), P2P order book records and real-time or historical meter readings.

Action attributes: These attributes specify the operations proposed by subjects on objects. For the P2P market, the relevant actions are "buy" and "offer".

Environment Attributes: Environment attributes include the current date and time affecting contract validity, grid load in the form of a traffic light signal along the lines of the BDEW grid traffic light concept [23], as well as current weather conditions and predictions. While the grid load can limit P2P trading during congestion phases, the weather conditions can be used to validate the plausibility of energy offers.

The aforementioned attributes lay the foundation for the authorization process, which will be described in more detail in the upcoming section.

4.3 P2P trading authorization

This section describes the authorization process of a prosumer who wants to offer their energy on the P2P energy market. This process is carried out every time a prosumer wants to trade energy on the P2P market and is one of the most fundamental pillars of the security concept described in this work. The implementation and validation of this process will be described in section 5.

As depicted in figure 3, a prosumer who wants to offer energy on the P2P market can send a trade offer to the P2P platform, which will then request a presentation of the prosumers trading licence VC. The prosumer provides the VC presentation including the information needed for the zero-knowledge proof of relevant claims to the P2P platform, which can then verify this presentation with the credential registry stored on the Hyper-ledger Indy blockchain. When the claims are verified, they are combined with the trade offer and sent to the ABAC system, which will gather the environment attributes and decide based on the ABAC policy. This policy contains access rules which ensure that the prosumer is certified to supply the energy they want to offer, the current grid load allows P2P energy trading, and that the current weather and time of day makes the energy offer plausible.

Additionally, it includes rules to make certain that no prosumer can put up more than one energy offer per time period and that they cannot offer energy for more time periods than can be predicted reliably. This prevents problems through faulty predicitions and platform misuse. After the decision has been made, it is returned to the P2P trading platform, which can then accept or deny the trade offer based on the ABAC decision. This process ensures that the prosumer can provide the offered energy, is a trusted trade partner and that the grid state allows P2P trading all in one decision.

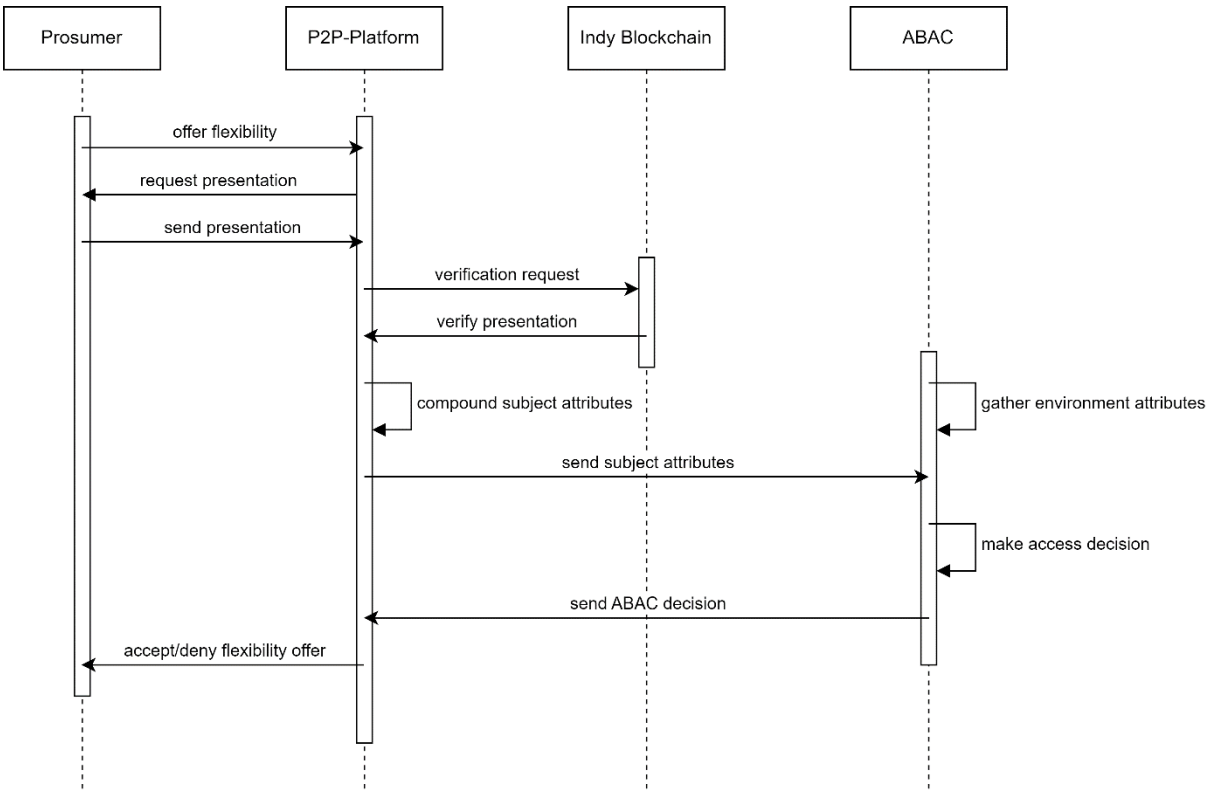


Figure 3: Sequence diagram of the P2P trading authorization process

5 Evaluation

This section introduces the test environment and shows the results of a DoS attack on the environment.

5.1 Test setup

To evaluate the presented concept, a simulation environment was established. This setup includes a Docker network featuring an SSI implementation based on the VON Network, an open-source Hyperledger Indy-based system and Aries-Cloudagent-Python, an open-source implementation of the Hyperledger Aries SSI agent framework. In addition, the ABAC system was developed using the Authzforce open-source PDP (Policy Decision Point) engine. The communication within this system is facilitated through

REST APIs. The ABAC policy set is composed of 10 attributes, which are categorized into the groups described in section 4.2. The SSI system itself is built on four Hyperledger Indy nodes along with nine prosumer agents, and one P2P trading platform agent.

To further assess the robustness and performance of both the SSI and ABAC PDP systems, the setup will undergo testing through a DoS (Denial of Service) attack. This test aims to analyze how well the SSI system, along with the ABAC PDP, can handle and mitigate the effects of such attacks, thereby ensuring the reliability and security of the system under stress conditions.

5.2 Evaluation of System Resilience through DoS Attack Simulation

A DoS attack constitutes a malicious attempt to disrupt the normal operation of a targeted server, service, or network by flooding the target with internet traffic or by sending information that causes the system to crash. Such attacks exploit the limits of network and system resources, aiming to render the system inoperative and deny access to legitimate users. In the context of distributed systems, particularly those involving critical infrastructures like SSI and ABAC systems, resilience against DoS attacks is critical for ensuring uninterrupted service delivery.

In evaluating the resilience of our deployed systems against potential cyber threats, a DoS attack simulation was conducted using the Locust framework against the specified test environment. This environment was configured with 8 GB of RAM and dual AMD EPYC-Rome Processor cores operating at 3 GHz. Conversely, the attacking framework was equipped with 32 GB of RAM and twelve cores of the same processor family. The assault commenced with a single attacker, with an additional attacker joining every second until reaching a total of 128 attackers. The system demonstrated capability to manage up to 20 attackers and sustain 3,000 requests per second. Beyond this threshold, the response time escalated to a level where the system was deemed unavailable, approximately at 5,000 requests per second. The cumulative outcomes are presented in Table 1, highlighting a significant disparity between the minimum and maximum response times, indicating the system's inability to handle the load.

Achieving a maximum throughput of approximately 5,000 requests per second exceeds the anticipated requirement for a local energy community, indicating the system's adequacy under normal operating conditions. Nevertheless, the demonstrated vulnerability to a DoS attack necessitates the implementation of security measures to prevent the SSI and ABAC systems from processing illegitimate requests. One effective countermeasure to mitigate the impact of DoS attacks is the implementation of ABAC in lower

network layers that block malicious requests before reaching the application layer of the system where an investigation of a requests means a higher processing effort.

Number of requests	Median response time	Average response time	Min response time	Max response time	Average requests/s
844806	9 ms	16.99 ms	1.16 ms	784.30 ms	4757.05

Table 1: Results of DoS attack on test environment. The attack included 844,806 sent requests done by up to 128 attackers running on 12 cores. The average and maximum response time in comparison to the minimum response time show that the system was overloaded.

6 Conclusion

In this study, the authors have developed a comprehensive framework aimed at securing IoT communications within critical infrastructure sectors, particularly focusing on energy systems. Answering the research question posed in section 1, "How can security measures be integrated into energy communities to ensure a stable energy supply and secure P2P trading?", the framework integrates machine learning for intrusion detection, SSI for authentication, and ABAC for authorization, targeting the enhancement of security in the local energy community of the future.

The core of the framework is a novel network architecture that combines SSI and ABAC, facilitating not only the verification of participants' identities but also the management of access permissions to ensure a secure P2P market authorization. This is further supported by the deployment of intrusion detection systems across devices on every grid level, enhancing the early detection of cyber threats and improving the grid's overall cybersecurity posture.

Evaluation against DoS attacks demonstrated the framework's effectiveness in maintaining operational resilience, although it also highlighted areas where additional security measures could further strengthen system robustness. This study contributes to the field by offering a practical and adaptable solution for enhancing the security of future energy communities, underscoring the critical role of integrating advanced technological components to address the cybersecurity challenges faced by critical infrastructures. Especially in the light of recent regulation such as the EU's NIS 2, the proposed framework offers a promising solution to cybersecurity requirements.

7 Acknowledgement

This research was performed within the project PEAK and supported by the German Ministry for Economic Affairs and Climate Action (BMWK Grant No. 03EI6035G).
The authors of these publications have committed themselves to the guidelines of good scientific practice of the German Research Foundation.

References

[1] Zakaria Abou El Houda, Bouziane Brik, and Lyes Khoukhi. “Ensemble learning for intrusion detection in sdn-based zero touch smart grid systems”. 2022. IEEE 47th Conference on Local Computer Networks (LCN).

[2] Rafael Belchior et al. “SSIBAC: self-sovereign identity based access control”. IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2020.

[3] Umit Cali et al. “Improved Resilience of Local Energy Markets using Blockchain Technology and Self-Sovereign Identity”. IEEE 1st Global Emerging Technology Blockchain Forum: Blockchain & Beyond (iGETblockchain). 2022.

[4] Francesco Caputo, Barbora Buhnova, and Leonard Walletzk `y. 2018. “Investigating the role of smartness for sustainability: Insights from the Smart Grid domain.” Sustainability Science 13, 2018.

[5] Matthew Davie et al. “The trust over ip stack”. IEEE Communications Standards Magazine 3, 4 , 2019

[6] Vasudev Dehalwar et al. “Blockchain-based trust management and authentication of devices in smart grid”. Cleaner Engineering and Technology 8, 2022

[7] Tassos Dimitriou and Ghassan Karame. “Privacy-friendly tasking and trading of energy in smart grids”. In Proceedings of the 28th Annual ACM Symposium on Applied Computing, 2013.

[8] Seongho Hong and Heeyoul Kim. “Vaultpoint: A blockchain-based ssi model that complies with oauth 2.0”. Electronics 9, 8 , 2020.

[9] Chengming Hu, Jun Yan, and Chun Wang. “Robust feature extraction and ensemble classification against cyber-physical attacks in the smart grid”. In 2019 IEEE Electrical Power and Energy Conference (EPEC). IEEE, 2019.

[10] Vasiliki Kelli et al. “Attacking and Defending DNP3 ICS/SCADA Systems”. In 2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS). IEEE, 2022

[11] Tala Talaei Khoei et al. “Ensemble learning methods for anomaly intrusion detection system in smart grid”. In 2021 IEEE international conference on electro information technology (EIT). IEEE, 2021

[12] Pengyi Liao et al. “Divergence-Based Transferability Analysis for Self-Adaptive Smart Grid Intrusion Detection With Transfer Learning”. IEEE Access 10, 2022

[13] Binhao Ma et al. “A secure and decentralized SSI authentication protocol with privacy protection and fine-grained access control based on federated blockchain”. Plos one 17, 9 , 2022

[14] Mohamed Massaoudi, Shady S Refaat, and Haitham Abu-Rub. “Intrusion detection method based on smote transformation for smart grid cybersecurity”. In 2022 3rd International Conference on Smart Grid and Renewable Energy (SGRE). IEEE, 2022

[15] Vagner E Quincozes et al. “Feature Extraction for Intrusion Detection in IEC-61850 Communication Networks”. In 2022 6th Cyber Security in Networking Conference (CSNet). IEEE, 2022

[16] Keyvan Ramezanpour and Jithin Jagannath. “Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN”. Computer Networks 217, 2022

[17] Puja Sen and Sumit Waghmare. “Machine Learning Based Intrusion Detection System for Real-Time Smart Grid Security”. In 2021 13th IEEE PES Asia Pacific Power & Energy Engineering Conference (APPEEC). IEEE, 2021

[18] R Sriranjani et al. “Machine Learning Based Intrusion Detection Scheme to Detect Replay Attacks in Smart Grid”. In 2023 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2023

[19] Wayes Tushar et al. “Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges”. Applied Energy 282, 2021

[20] Radhakrishnan Vijayanand, D Devaraj, and B Kannapiran. “A novel deep learning based intrusion detection system for smart meter communication network”. In 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). IEEE, 2019

[21] Moritz Volkmann et al. “Privacy in Local Energy Markets: A Framework for a Self-Sovereign Identity based P2P-Trading Authentication System”. In 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). IEEE, 2023

[22] Preukschat, A., & Reed, D. “Self-sovereign identity”. Manning Publications, 2021

[23] BDEW. “Konkretisierung des Ampelkonzepts im Verteilungsnetz” https://www.bdew.de/media/documents/Stn_20170210_Konkretisierung-Ampelkonzept-Verteilungsnetz.pdf, 2017

[24] "Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive)." http://data.europa.eu/eli/dir/2022/2555/2022-12-27 . Official Journal of the European Union, 2022

[25] Inayat, Usman et al. "Cybersecurity Enhancement of Smart Grid: Attacks, Methods, and Prospects" https://doi.org/10.3390/electronics11233854. Electronics 11, 2022

[26] Lifshitz, Y. R. "Private Energy". Stan. Envtl. LJ, 2018

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Sensibilisierung von Mitarbeitern zu KI-gestützten Cyber-Angriffen

Martin Helge Dependahl[1]

Kurzfassung:
KI-Systeme haben sich in der letzten Zeit rasant weiterentwickelt. Neben legalen Anwendungsmöglichkeiten, sowohl im privaten als auch im geschäftlichen Umfeld nutzen auch Cyber-Kriminelle im Rahmen von Angriffen, sowohl gegen Privatpersonen als auch Unternehmen, die Möglichkeiten, die KI-Systeme ihnen bieten. Als Beispiel lassen sich hier die verschiedensten Formen von Social Engineering wie E-Mail-Phishing, Social-Engineering-Angriffe am Telefon oder Deepfakes anführen. Um diesen neusten Entwicklungen zu begegnen, sollten Unternehmen eine Strategie für den Umgang mit solchen Angriffen erarbeiten. Mangels ausreichender technischer Schutzmaßnahmen ist die Sensibilisierung von Mitarbeitern ein wesentlicher Bestandteil einer solchen Strategie. Exemplarisch wird im Rahmen dieses Beitrags ein digitaler Escape Room zur Sensibilisierung von Mitarbeitern zu KI-gestützten Cyber-Angriffen beschrieben, wie er bei der EnBW Energie Baden-Württemberg AG in die Praxis umgesetzt wurde. Neben dem Ablauf des Escape Rooms an sich wird hierbei auch auf die technische Implementierung sowie die Evaluation des Escape Rooms als Sensibilisierungsmaßnahme näher eingegangen.

Stichworte: Awareness, Deepfakes, Escape Room, KI, Mitarbeitersensibilisierung, Phishing, Social Engineering

1 Motivation

Seit der Veröffentlichung des Chatbots ChatGPT im November 2022 haben das Thema künstliche Intelligenz (KI) und ihre Auswirkungen auf Unternehmen und die Gesellschaft eine erhöhte Aufmerksamkeit in der Öffentlichkeit erhalten. Dies lässt sich durch zahlreiche Artikel und Diskussionen in den etablierten Medien belegen. Neben der Ausgabe von Texten, wie bei ChatGPT, gibt es noch eine Vielzahl anderer Anwendungen, bei denen Künstliche Intelligenz sowohl im unternehmerischen als auch im privaten Kontext in der Praxis eingesetzt wird. Hierzu zählen beispielsweise die Bild- und Sprachverarbeitung.

Die Möglichkeiten, die KI-Werkzeuge bieten, werden jedoch nicht nur zu legalen Zwecken von Unternehmen und Privatpersonen genutzt. Auch Cyber-Kriminelle verwenden vermehrt KI-Werkzeuge im Rahmen von Cyber-An-

[1] Hochschule Karlsruhe

griffen. Aufgrund der vielfältigen Szenarien von KI-gestützten Cyber-Angriffen sollten Unternehmen Strategien entwickeln, wie mit diesen Cyber-Bedrohungen umgegangen werden soll.

Bei den meisten dieser Angriffsszenarien ist der Mitarbeiter die jeweilige Schwachstelle des Angriffs. Daher kommt mangels ausreichender technischer Schutzmechanismen den Themen Aufklärung und Sensibilisierung von Mitarbeitern als präventive Maßnahme eine wichtige Rolle zu.

Im Rahmen dieses Beitrags wird ein Sensibilisierungskonzept zu KI-gestützten Cyber-Angriffen vorgestellt, wie es exemplarisch bei der EnBW Energie Baden-Württemberg AG (Karlsruhe) im Zeitraum zwischen Oktober 2023 und Januar 2024 in die Praxis umgesetzt wurde. Der Hauptbestandteil dieses Sensibilisierungskonzeptes ist ein digitaler Escape Room, auf welchem der Fokus dieses Beitrags liegt.

2 Gefahren aus KI-gestützten Cyber-Angriffen

Als Grundlage für eine erfolgreiche Sensibilisierungskampagne zu KI-gestützten Cyber-Angriffen müssen zu Beginn die verschiedenen Formen von KI-gestützten Cyber-Angriffen identifiziert werden. Eine Auswahl von verbreiteten Angriffen, welche auch später im Escape Room aufgegriffen werden, wird im Folgenden näher beschrieben.

2.1 Social Engineering

Social-Engineering-Angriffe zählen zu den weltweit am weitesten verbreiteten Cyber-Angriffen auf Unternehmen und Privatpersonen. Ein Social-Engineering-Angriff kann dabei grundsätzlich sowohl in Form eines Gesprächs (z.B. am Telefon) als auch in textueller Form, beispielsweise per Brief, auf Social-Media-Plattformen oder per E-Mail durchgeführt werden.

2.1.1 Vorbereitung von Social-Engineering-Angriffen

Ein wesentlicher Beitrag für den Erfolg eines Social-Engineering-Angriffs ist die Sammlung von Informationen über das Opfer durch den Angreifer. Hierfür kann ein Angreifer beispielsweise eine OSINT-Analyse (Open Source Intelligence) durchführen. Dabei werden öffentlich verfügbare Informationen als Quelle genutzt, beispielsweise soziale Netzwerke oder Webseiten.

Neben einer manuellen Internetrecherche können Angreifer auch auf zumeist kostenpflichtige OSINT-Tools zurückgreifen. Diese durchsuchen das Internet anhand der eingegebenen Attribute und geben die gefundenen Informationen in Form eines Datengraphen aus. Es zeichnet sich ab, dass diese Art von Tools in Zukunft neben den bereits verwendeten Algorithmen auch vermehrt auf KI-Technologien zurückgreifen werden.

Zusätzlich zu OSINT-Tools kann ein Angreifer auf KI-Systeme zurückgreifen, die über eine Suchmaschinenanbindung verfügen. Die Qualität der erhaltenen Daten hängt dabei von den Privatsphäre-Einstellungen ab, die das Opfer in den jeweiligen sozialen Netzwerken gesetzt hat. Viele dieser KI-Systeme können derzeit nur auf die öffentliche Ansicht von sozialen Netzwerken zugreifen.

Darüber hinaus können KI-Systeme den Angreifer dabei unterstützen, die gefundenen Informationen aufzubereiten. So ist es mithilfe von KI möglich, sich Steckbriefe über das Opfer erstellen zu lassen.

Zur Prävention von Social-Engineering-Angriffen sollten Unternehmen daher ihre Mitarbeiter für die Gefahren bei der Nutzung und für das Verhalten in Social-Media-Netzwerken sensibilisieren. Mitarbeiter sollten dazu angehalten werden, nur unbedingt notwendige Informationen im Internet zu veröffentlichen und die Privatsphäre-Einstellungen der einzelnen Plattformen zu nutzen, um das Sammeln von Informationen zu erschweren.

2.1.2 Phishing-Angriffe

Das Versenden von nicht personalisierten Phishing-E-Mails sowie von Spear-Phishing-E-Mails ist seit Jahren weit verbreitet. Zu den Haupterkennungsmerkmalen von Phishing-E-Mails zählten vor der Verbreitung von KI-Systemen unter anderem die verwendete Wortwahl, ein fehlerhafter Satzbau oder Mängel im Bereich der Rechtschreibung.

Durch die Verwendung von KI-Systemen haben Angreifer nun die Möglichkeit, mit nur wenigen Prompt-Eingaben eine qualitativ hochwertige Phishing-E-Mail zu verschiedensten Themen nach ihren Wünschen zu erstellen. Diese hohe Qualität umfasst sowohl inhaltliche als auch linguistische Aspekte. So muss ein Angreifer durch die Nutzung von KI-Systemen weder die Sprache des Textes noch die Eigenheiten einer Sprache, wie z.B. übliche Redewendungen, beherrschen. Der Angreifer kann seine Phishing-E-Mail darüber hinaus auch stilistisch anpassen, indem er dem Prompt einen Schreibstil, wie die Verwendung von Emojis, mitgibt. Die hohe Qualität von Phishing-E-Mails erschwert die Erkennung solcher E-Mails sowohl für Spamfilter als auch für die Mitarbeiter.

Aufgrund der kurzen Erstellungszeiten ist absehbar, dass in Zukunft neben der Qualität auch die Quantität von Phishing-E-Mails zunehmen wird. Dies erhöht die Wahrscheinlichkeit eines erfolgreichen Angriffs weiter.

Erste Untersuchungen haben ergeben, dass die Klickraten auf in der E-Mail eingebettete Links bei KI-generierten E-Mails im Vergleich zu manuell erstellten E-Mails noch etwas geringer sind. Es ist jedoch davon auszugehen,

dass durch die Weiterentwicklung von KI-Systemen die Klickraten bei KI-generierten Phishing-E-Mails weiter ansteigen und in Zukunft möglicherweise über den Klickraten von manuell erstellten Phishing-E-Mails liegen werden.[2]

Daher ist es als präventive Maßnahme notwendig, dass Unternehmen ihre Mitarbeiter auf die gestiegene Gefahr von Phishing-E-Mails hinweisen und regelmäßig die Erkennungsmerkmale und den Umgang sowie den Meldeweg beim Empfang einer Phishing-E-Mail mit den Mitarbeitern trainieren.

2.1.3 Social-Engineering-Angriffe am Telefon

Eine weitere Form von Social-Engineering-Angriffen, welche durch die Verbreitung von KI-Systemen erst ermöglicht wird, sind Social-Engineering-Angriffe am Telefon (bekannt unter anderem als Enkeltrick). Während früher Angerufene oft von falschen Polizisten oder ähnlich vermeintlich wichtigen Personen unter Druck gesetzt wurden, ist es heute mithilfe von KI-Systemen möglich, Stimmklone, von Bekannten des Opfers zu erstellen und in einem solchen Anruf zu verwenden.

Für einen solchen Angriff reichen bei modernen KI-Systemen bereits wenige Minuten Stimmmaterial aus, um einen hochwertigen Stimmklon der jeweiligen Person zu erstellen. Da der Angerufene in der Regel keine Vergleichsmöglichkeit mit der originalen Stimme des Anrufers besitzt und durch den Anrufer in ein Gespräch verwickelt wird, ist diese Form des Angriffs aus Sicht des Angreifers sehr erfolgversprechend.

Da es derzeit kaum technische Schutzmaßnahmen gegen solche Angriffe gibt, kommt der Sensibilisierung von Mitarbeitern eine besondere Rolle zu. Mitarbeiter sollten dazu angehalten werden, in Telefongesprächen auf auffällige Merkmale zu achten. Dies können zum Beispiel ein falscher Dialekt oder fehlende Füllwörter sein. Außerdem sollten Mitarbeiter darauf achten, keine persönlichen Informationen oder Geschäftsgeheimnisse am Telefon preiszugeben.

Wenn Mitarbeiter Zweifel an der Echtheit eines Anrufers haben, sollten sie den Anrufer zurückrufen, sofern ihnen dessen tatsächliche Rufnummer bekannt ist. Außerdem sollten sie gezielte Nachfragen oder Kontrollfragen stellen und die im Telefonat erhaltenen Informationen unabhängig verifizieren.

[2] Vgl. z.B. SoSafe GmbH, Jeder Fünfte klickt auf KI erstellte Phishing-Mails, https://sosafe-awareness.com/de/ueber-sosafe/presse/jeder-fuenfte-klickt-auf-ki-erstellte-phishing-mails/

2.2 Deepfakes

KI bietet mit der Möglichkeit, Audio-, Bild- und Videoaufnahmen zu manipulieren, eine weitere Methode, Unternehmen anzugreifen. Diese Art der Manipulation wird umgangssprachlich auch als Deepfake bezeichnet.

Deepfakes stellen insbesondere für Unternehmen eine Gefahr dar. Ein Angreifer könnte beispielsweise einen Deepfake als Nachricht im Kontext von Fake News im Internet veröffentlichen und über eine seriöse Aufmachung sowie eine gezielte Platzierung dafür sorgen, dass diese Nachricht in der Öffentlichkeit für echt gehalten wird. Die Folgen eines solchen Angriffs können ein Absinken der Unternehmensbewertung oder ein Verlust von Reputation sein.

Im Vergleich zu anderen Angriffsformen ist die Unterstützung durch KI-Lösungen bei der Erstellung von Deepfakes bereits weit fortgeschritten. Bilder und andere Inhalte im Internet sollten daher nach Möglichkeit verifiziert und auf Fehler und Täuschungsversuche hin untersucht werden. Bei Zweifeln an der Echtheit eines Bildes können beispielsweise fehlende Schattenwürfe oder Spiegelungen in Scheiben ein Hinweis auf ein KI-generiertes Bild sein.

2.3 Manipulation von KI-Systemen

Neben KI-gestützten Cyber-Angriffen bieten auch im Unternehmenskontext eingesetzte KI-Systeme Angreifern eine Möglichkeit zur Manipulation. KI-Systeme an sich können grundsätzlich auf verschiedene Arten manipuliert werden.

Hierzu zählen unter anderem die Manipulation der Trainingsdaten eines KI-Systems durch den Angreifer, außerdem sogenannte Indirect-Prompt-Injection-Angriffe oder die direkte Aufforderung des Nutzers, manipulierte Anfragen an ein KI-System zu stellen.

Für den Mitarbeiter, der ein KI-System als Endbenutzer nutzt, spielen die verschiedenen Formen der Manipulation von KI-Systemen nur eine untergeordnete Rolle, da hierfür tiefergehendes technisches Verständnis und Knowhow erforderlich sind. Jeder Mitarbeiter sollte jedoch dazu angehalten werden, die Ergebnisse eines KI-Systems immer kritisch zu hinterfragen und die Plausibilität der angezeigten Ergebnisse zu überprüfen. Wenn Zweifel an der Plausibilität bestehen, sollte der Mitarbeiter versuchen, das erhaltene Ergebnis auf herkömmliche Weise zu verifizieren, ohne dabei das KI-System zu nutzen.

3 Sensibilisierung von Mitarbeitern zu KI-gestützten Cyber-Angriffen

Wie im vorherigen Kapitel beschrieben, sind die Bedrohungen durch KI-gestützte Cyber-Angriffe auf Unternehmen vielfältig. Sensibilisierungs- und Schulungsmaßnahmen für Mitarbeiter und Führungskräfte sind daher ein elementarer Bestandteil einer Strategie zum sicheren und richtigen Umgang mit KI-gestützten Cyber-Angriffen. Das Ziel dieser Maßnahmen ist es, dass die Mitarbeiter trainieren, entsprechende Angriffe zu erkennen und dass sie wissen, wie sie sich in einer solchen Situation sicher und regelkonform verhalten sollen.

3.1 Lernziele

Unter Berücksichtigung der in Kapitel 2 dargestellten Gefahren lassen sich die folgenden Lernziele für eine Sensibilisierung der Mitarbeiter zu KI-gestützten Cyber-Angriffen aufstellen:

- Das Risiko, Opfer eines erfolgreichen Cyber-Angriffs zu werden, steigt durch die weitere Entwicklung von generativen KI-Systemen an.
- Die Veröffentlichung von Informationen im Internet ist mit Risiken verbunden, da öffentliche Informationen im Rahmen einer OSINT-Analyse für einen Angriff verwendet werden können.
- Der Mitarbeiter soll trainieren, Phishing-E-Mails zu erkennen und zu melden.
- Phishing-E-Mails können mithilfe von KI in einer hohen Qualität erstellt werden. Diese werden dadurch zu einer zunehmenden Gefahr für den Mitarbeiter, der sich dessen bewusst sein sollte.
- Die Gefahr eines manipulierten Anrufes, bei dem eine geklonte Stimme verwendet wird, ist durch die Verbreitung von KI deutlich angestiegen.
- Mithilfe von KI können authentische Bilder erstellt oder manipuliert werden.
- KI-Systeme können durch Angreifer manipuliert werden.

3.2 Escape Room

Ziel der Sensibilisierung von Mitarbeitern und Führungskräften ist das sichere und richtige Handeln in Bezug auf die aufgestellten Lernziele. In diesem Zusammenhang soll durch die durchgeführten Maßnahmen eine Verhaltensänderung sowie eine Reflexion des vorhandenen Wissens angestrebt werden. Hauptbestandteil der Sensibilisierungskampagne bei der EnBW AG zu KI-gestützten Cyber-Angriffen ist ein Escape Room, der im Folgenden näher beschrieben wird.

Der Escape Room sollte jedoch nicht alleinstehend betrachtet werden, sondern immer im Zusammenspiel mit weiteren flankierenden Sensibilisierungsmaßnahmen umgesetzt werden. Hierzu zählen sowohl die Maßnahmen zu den aufgestellten Lernzielen als auch die grundlegenden Sensibilisierungs- und Schulungsmaßnahmen zur Informationssicherheit im Allgemeinen.

Der hier beschriebene Escape Room wird in Teams aus zwei bis fünf Mitarbeitern gespielt. Die jeweiligen Teams melden sich im Vorfeld bei ihrem zuständigen Informationssicherheitsmanager für das Spiel an. Dieser kümmert sich um die weitere Organisation und übernimmt während des Spiels in der Regel auch die Rolle des Spielleiters.

Die Bewerbung des Escape Rooms als Sensibilisierungsformat im Konzern erfolgt über Beiträge im Intranet sowie über die Informationssicherheitsmanager und die Abteilungsleiter. Diese können den Sensibilisierungsbedarf ihrer Mitarbeiter gut einschätzen und bei Bedarf den Escape Room als Sensibilisierungsmaßnahme vorschlagen. Darüber hinaus hat sich in der Praxis gezeigt, dass viele Teams den Escape Room auch auf Empfehlung anderer Mitarbeiter von sich aus anfragen.

Das Format des Escape Rooms eignet sich besonders für Teamtage oder als Maßnahme zum Teambuilding, da er im Vergleich zu anderen Formaten wie Workshops oder Präsentationen durch die Gamifizierung der Inhalte einen auflockernden Charakter besitzt.

3.2.1 Allgemeiner Ablauf

Die Durchführung des Escape Rooms an sich unterteilt sich in drei Phasen:

Zu Beginn werden die Teilnehmer durch den Spielleiter (in der Regel der für das Spielerteam zuständige Informationssicherheitsmanager) begrüßt und in das Format des Escape Rooms eingeführt.

In der darauffolgenden Phase spielen die Teilnehmer den Escape Room und versuchen gemeinsam, die einzelnen Rätsel des Spiels innerhalb der vorgegebenen Zeit und mithilfe der bereitgestellten Materialien zu lösen. Bei Bedarf kann der Spielleiter während des Spiels Tipps geben, um ein positives Spielerlebnis für die Spieler zu gewährleisten.

In der abschließenden Phase bespricht der Spielleiter im Rahmen eines Debriefings noch einmal die einzelnen Rätsel des Escape Rooms mithilfe einer vorbereiteten Präsentation. Im Zuge dieses Debriefings lernen die Teilnehmer, wie sie die im Escape Room thematisierten Angriffe in ihrem Arbeitsalltag erkennen und verhindern können. Außerdem erfahren sie, wie sie sich

im Falle eines Angriffs sicher und regelkonform verhalten sollen. In dieser Phase haben die Teilnehmer zusätzlich die Möglichkeit, sich über die vorgestellten Themen auszutauschen und darüber hinaus eigene Erfahrungen mit der Gruppe zu teilen.

3.2.2 Rahmenbedingungen

Für den Escape Room wurden die folgenden Rahmenbedingungen festgelegt:

- Die Gesamtdauer der Sensibilisierungsmaßnahme beträgt in der Regel 90 Minuten. Der Anteil des Spiels liegt hierbei bei maximal 80 Minuten. Bei Debriefings mit hohem Diskussionsanteil kann die Sensibilisierungsmaßnahme in der Praxis bis zu 120 Minuten dauern.
- Das Spiel wird in Kleingruppen mit 2 bis 5 Personen gespielt.
- Jeder Mitarbeiter ist in der Lage den Escape Room zu lösen. Es ist kein Vorwissen erforderlich, das über die allgemeine Grundsensibilisierung zur Informationssicherheit hinausgeht.
- Es ist möglich den Escape Room in allen Besprechungsräumen zu spielen. Ein Remote-Spiel über Microsoft Teams ist ebenfalls möglich. In diesem Fall nimmt der übertragende Spieler die Rolle des Spielführers ein.
- Für die Durchführung wird folgendes Material benötigt: Notebook, eine Internetverbindung, ein Smartphone sowie bei Spielen in Präsenz ein größerer Monitor (32 Zoll oder größer). Ferner sollte bei Präsenzspielen der Blog des Protagonisten in ausgedruckter Form vorliegen.

Als Zielgruppe für diesen Escape Room wurden alle EnBW-Mitarbeiter mit einem PC-Arbeitsplatz definiert. Hierbei unterscheiden sich die unterschiedlichen Mitarbeiter sowohl von ihren Tätigkeitsschwerpunkten als auch von ihrem Wissensstand in den Bereichen der Informationssicherheit sowie der künstlichen Intelligenz. Daher wurden ausgewählte Teilrätsel des Spiels in verschiedenen Schwierigkeitsstufen umgesetzt. Zu Beginn des Spiels kann der Spielleiter anhand des zu erwartenden Kenntnisstands wählen, welche Schwierigkeitsstufe die einzelnen Teilrätsel haben sollen. Die Handlung des Spiels an sich ist in allen Schwierigkeitsstufen identisch.

3.2.3 Ablauf eines Spiels

Zu Beginn des Spiels erhalten die Spieler eine Einführung in das Szenario des Escape Rooms durch einen Brief des Konzernvorstandes.

Der Brief beschreibt, dass der Konzern von Cyber-Kriminellen angegriffen und viele Computersysteme verschlüsselt wurden. Die Angreifer fordern nun ein Lösegeld zur Entschlüsselung der Systeme.

Außerdem erfahren die Spieler, dass der Account des Protagonisten des Spiels im Zusammenhang mit dem Angriff steht. Die Spieler haben nun die Aufgabe, den Angriff nachzuvollziehen und das Passwort zur Entschlüsselung der Computersysteme zu finden. Als Hilfsmittel erhalten die Spieler an dieser Stelle des Spiels einen Steckbrief über den Protagonisten sowie dessen Blog. Dieser Blog enthält Hinweise zu den einzelnen Teilrätseln und trägt somit zur Lösung des gesamten Spiels bei.

Beim Start des Spiels sehen die Spieler einen typisch eingerichteten Arbeitsplatz der EnBW AG, was für einen hohen Wiedererkennungswert sorgt. Viele der hier dargestellten Elemente dienen als Absprungpunkt zu den einzelnen Teilrätseln des Spiels. So kann beispielsweise durch einen Klick auf das Notebook unter anderem der E-Mail-Client des Protagonisten eingesehen werden.

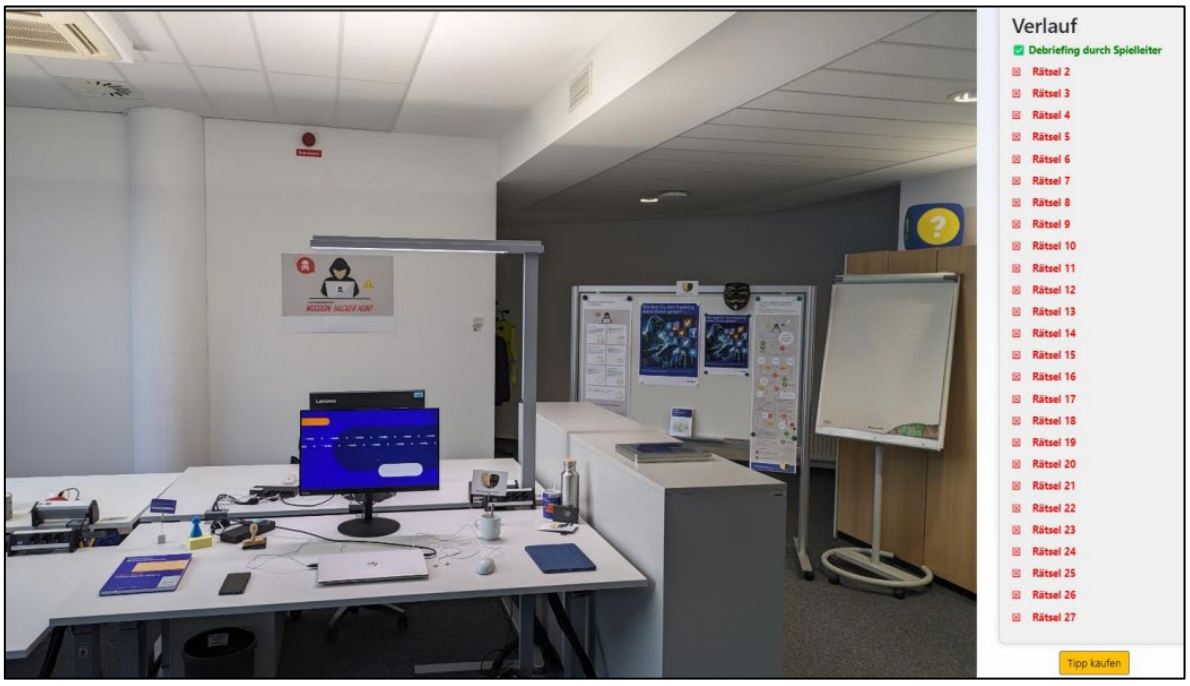


Abbildung1: Hauptoberfläche des Spiels

Im Blog steht zu diesem Teilrätsel beispielsweise folgender Text:

„Phishing Übung

Heute habe ich von der Informationssicherheit eine Übungsmail zum Thema Phishing erhalten. Und natürlich, wie sollte es auch anders sein, bin ich darauf

reingefallen und habe auf den Link in der E-Mail geklickt. Neben den Merkmalen von Phishing-E-Mails wollen Sie auch den Meldeweg mit uns trainieren. Wie meldet man nochmal eine Phishing-Mail?“

Aufgabe des Spielers bei diesem Teilrätsel ist es, den Posteingang nach möglichen Phishing-E-Mails zu durchsuchen und diese über den im E-Mail-Client integrierten Melden-Button zu melden.

Zur Unterstützung der Spieler, falls diese nicht genügend Wissen zu Phishing-E-Mails präsent haben, befinden sich zwei Plakate, welche die typischen Erkennungsmerkmale von Phishing-E-Mails sowie den Meldeprozess enthalten, an der Magnetwand im Hintergrund. Diese werden auch im Rahmen der allgemeinen Sensibilisierung zur Informationssicherheit im Konzern eingesetzt.

Um Zugang zu den weiteren Rätseln zu erhalten, müssen die einzelnen Gegenstände näher betrachtet werden. Zum Beispiel ist in der Flasche ein QR-Code versteckt, der den PIN für das Tablet enthält. Auf diesem sind eine Reihe von Apps installiert, die jeweilige weitere Aufgaben für den Spieler hervorbringen.

Durch das systematische Nachvollziehen und Durchklicken der im Blog beschriebenen Handlungen kann ein Großteil der Rätsel gelöst werden. Darüber hinaus interagiert das Spiel mit den Spielern beispielsweise durch einen Drohanruf auf dem Telefon, welches auf dem Schreibtisch liegt, oder durch den Empfang von neuen E-Mails. Dies wird dem Spieler durch das Abspielen von entsprechenden Klingeltönen akustisch mitgeteilt. Teilweise befinden sich in einigen Gegenständen auch Trigger, sodass einige Rätsel erst nach erfolgreichem Abschluss eines anderen Teilrätsels gelöst werden können.

Am Ende des Spiels, nachdem alle Rätsel gelöst wurden, erhalten die Spieler eine E-Mail von den Erpressern. In dieser E-Mail ist ein Gedicht, welches das Passwort zur Entschlüsselung der Computersysteme enthält.

Nach Eingabe dieses Passworts gelangen die Spieler auf eine Auflösungsseite. Hier haben sie die Möglichkeit, sich mit einem Teamnamen und der verbleibenden Zeit in eine Highscore-Liste einzutragen. Anschließend folgt das Debriefing durch den Spielleiter.

Die Spieler können ihren Fortschritt während des Spiels im rechten Bereich des Bildschirms nachvollziehen. Durch das erfolgreiche Lösen von Rätseln sammeln sie in der Verlaufsanzeige immer mehr Haken. Wenn die Spieler Hilfe benötigen, können sie die Funktion „Tipp kaufen“ nutzen, um einen Tipp gegen Zeitabzug zu erwerben.

Somit ist es auch prinzipiell möglich, das Spiel ohne einen Spielleiter zu spielen. In diesem Fall wird das Debriefing automatisiert als Video gestartet, welches neben den Verhaltenshinweisen für die Mitarbeiter auch die Herleitung der einzelnen Rätsellösungen beinhaltet.

3.2.4 Technische Umsetzung

Die technische Umsetzung des Escape Rooms besteht aus einer Reihe von einzelnen Webseiten, die die gesamte Spielsteuerung sowie die Spielinhalte enthalten. Der Spielleiter erhält die einzelnen Webseiten als Zip-Datei, die er lokal auf seinem Computer entpacken kann. Einzelne Elemente des Spiels wie die im Spiel eingebetteten Videos werden während des Spiels aus Microsoft SharePoint geladen, weshalb trotz der lokalen Implementierung eine Internetverbindung während des Spiels erforderlich ist.

Diese Zip-Datei enthält auch den Blog und den Steckbrief des Protagonisten als PDF-Datei. Bei Präsenzspielen wird empfohlen diese PDF-Datei den Spielern in ausgedruckter Form bereitzustellen. Außerdem ist die Spielleiterdokumentation in dieser Zip-Datei enthalten. Diese enthält als Hilfestellung für den Spielleiter Hinweise für die Konfiguration des Browsers sowie Lösungen für die einzelnen Teilrätsel, damit der Spielleiter bei Bedarf Tipps geben kann.

Die meisten der dargestellten Oberflächen bestehen aus selbst erstellten Fotos, auf denen sich klickbare Bereiche befinden. Die Inhalte der dargestellten Programme, wie dem E-Mail-Client, bestehen größtenteils aus Buttons sowie einfachen Boxen, die Texte enthalten. Für das Thema Social Engineering am Telefon wurde ferner ein Stimmklon mithilfe eines Tools der Firma ElevenLabs erstellt und die für das Spiel benötigten Redeinhalte mithilfe dieses Stimmklons produziert.

Auf eine Einbindung eines echten KI-Systems wurde bei der Umsetzung zu diesem Zeitpunkt verzichtet, da es für die Erreichung der definierten Lernziele ausreichend ist, das Verhalten von KI-Systemen zu imitieren.

Als Programmiersprache wurde dabei eine Kombination aus HTML, CSS und JavaScript gewählt. Die Spielsteuerung und der damit verbundene Ablauf des Spiels erfolgt über JavaScript und Values im Session Storage. Diese Values funktionieren ähnlich wie Cookies. Sie werden aber im Gegensatz zu Cookies beim Schließen des Browsers automatisch gelöscht, sodass auch mehrere Spiele nacheinander auf demselben Gerät gespielt werden können.

Der Escape Room kann in allen auf dem Markt verfügbaren Browsern gespielt werden. Bei Googles Chrome-Browser bestehen jedoch herstellerseitig

Einschränkungen in der Autoplay-Funktion für die automatische Medienwiedergabe. Diese Funktion wird beispielsweise für das Abspielen des Handyklingeltons oder das automatische Starten von Videoinhalten während des Spiels benötigt. Daher wird empfohlen, den Escape Room bevorzugt im Mozilla-Firefox-Browser oder in Microsoft Edge zu spielen. Bei diesen beiden Browsern ist es erforderlich, vor der Durchführung des Spiels die automatische Medienwiedergabe in den Einstellungen zu aktivieren. Bei Spielen auf sehr großen Bildschirmen kann es ferner erforderlich sein, dass im Browser eine andere Zoomstufe gesetzt werden muss. Diese Einstellungen sind beide in der Spielleiterdokumentation entsprechend festgehalten.

Langfristig soll es möglich sein, das Spiel auf eine serverseitige Implementierung umzustellen. Ferner ist es in Zukunft möglich, über eine Schnittstelle, ein echtes KI-System in das Spiel einzubinden. Dabei müssen jedoch datenschutz- und lizenzrechtliche Aspekte beachtet werden. Die Nutzung einer solchen Schnittstelle ist in der Regel kostenpflichtig. Außerdem besteht die Gefahr, dass der Escape Room nicht spielbar ist, falls das angebundenen KI-System ausfällt.

3.2.5 Evaluation

Nach der Fertigstellung konnte der Escape Room im Januar 2024 im Rahmen mehrerer Testspiele evaluiert werden. Insgesamt fanden in diesem Zeitraum fünf Testspiele mit insgesamt 18 Teilnehmern statt. Als Testspieler wurden sowohl Teams mit einem technischen Schwerpunkt als auch Teams mit einem weniger technischen Schwerpunkt ausgewählt.

Nach Abschluss der Testspiele ist der hier beschriebene Escape Room konzernweit für alle Mitarbeiter des EnBW-Konzerns als mögliche Sensibilisierungsmaßnahme zur Verfügung gestellt worden.

Anhand der Auswertung der Testspiele kann festgehalten werden, dass der Escape Room sehr positiv von den Testspielern angenommen wurde. Dies lässt sich durch eine durchschnittliche Bewertung von 4,7 von 5 möglichen Sternen sowie die Nachfrage nach weiteren Escape Rooms zu anderen Themenbereichen der Informationssicherheit belegen.

Auch das Feedback der Teilnehmer in den Gesprächen nach der Durchführung der Spiele war sehr positiv. Besonders positiv wurde hier das gemeinsame Erarbeiten der Rätsellösungen durch die Spieler hervorgehoben, was auch als Teambuildingmaßnahme diente.

Viele Teilnehmer betonten außerdem, dass ihnen erst durch die Teilnahme am Escape Room und das damit verbundene Erleben von Cyber-Angriffen

bewusst wurde, mit welcher hohen Qualität die verschiedenen Angriffsformen durch den Einsatz von KI-Systemen durchgeführt werden können. Auch die Diskussionen im Rahmen des Debriefings trugen maßgeblich dazu bei, die Teilnehmer hinsichtlich KI-gestützter Cyber-Angriffe zu sensibilisieren. So berichteten einige Teilnehmer, bereits selbst Opfer eines Cyber-Angriffs geworden zu sein oder ein solches Opfer in ihrem persönlichen Umfeld zu haben.

Während der Evaluation hat sich jedoch gezeigt, dass das Format des Escape Rooms allein nicht ausreicht, um eine konzernweite Sensibilisierung zum Thema zu erreichen. Pro Spiel können nur wenige Mitarbeiter sensibilisiert werden. Eine konzernweite Sensibilisierung würde somit eine Vielzahl von Spielen erfordern, was mit einem entsprechend hohem Aufwand verbunden ist.

Dieser Escape Room eignet sich daher eher zur Vertiefung von bereits vermitteltem Wissen, als Ergänzung zu anderen Sensibilisierungsmaßnahmen oder für die Erstsensibilisierung von Mitarbeitern, bei denen die Thematik KI-gestützte Cyber-Angriffe von besonderer Relevanz ist.

3.3 Flankierende Maßnahmen

Um die Mitarbeiter für das Thema KI-gestützte Cyber-Angriffe zu sensibilisieren, bedarf es zusätzlich zum Escape Room weiterer flankierender Sensibilisierungsmaßnahmen.

Zu den möglichen Sensibilisierungsmaßnahmen zählen unter anderem ein Themenmonat zu KI-gestützten Cyber-Angriffen im Intranet, Webinare und das Versenden von KI generierten Phishing-E-Mails an alle Unternehmensmitarbeiter. Auch das gezielte Trainieren des Umgangs mit Fake Anrufen ist als flankierende Maßnahme denkbar.

Diese Maßnahmen, genauso wie der Escape Room, sollten im Rahmen eines ganzheitlichen Sensibilisierungskonzeptes zum Themenbereich der Informationssicherheit gesehen werden. Hierbei sollten neben dem Thema der KI-gestützten Cyber-Angriffe auch andere Themenbereiche der Informationssicherheit mit einbezogen werden.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Usable Security gestalten und evaluieren

Dr. Matthias Korn[1], Prof. Dr. Therese Mieth[2], Kristina Unverricht[3]

Kurzfassung:
Effektive Cybersicherheitsmaßnahmen sind ausschlaggebend für eine sichere Digitalisierung. Die praktische Nutzbarkeit dieser Schutzmaßnahmen spielt dabei eine entscheidende Rolle. In diesem Beitrag wird eine Systematik vorgestellt, die die vier wesentlichen Handlungsfelder von Usable Security definiert und beschreibt: Gebrauchstauglichkeit, Transparenz, Zugänglichkeit/Barrierefreiheit und Akzeptanz. Durch die Beschreibung der zentralen Aspekte und Erfolgskriterien kann die Systematik als Hilfestellung bei der Evaluierung von Sicherheitslösungen hinsichtlich des Qualitätsmerkmals Usable Security dienen.

Stichworte: Benutzungsfreundlichkeit, Digitaler Verbraucherschutz, Mensch-zentrierte Cybersicherheit, Usable Security

1 Einführung

In der heutigen Zeit gilt die Digitalisierung als zentrales politisches, gesellschaftliches und wirtschaftliches Ziel, das kontinuierlich vorangetrieben wird. Dies führt dazu, dass Menschen zunehmend auf digitale Dienste angewiesen sind, die nahezu alle Lebensbereiche abdecken. Dieser Trend erfordert verstärkte Anstrengungen für geeignete Cybersicherheitsmaßnahmen digitaler Technologien zum Schutz der Menschen. Die praktische Nutzbarkeit dieser Schutzmechanismen durch IT-Anwenderinnen und Anwender spielt dabei eine entscheidende Rolle, denn ein wirkungsvoller Schutz ist nur gewährleistet, wenn Sicherheitslösungen tatsächlich *genutzt werden* bzw. *genutzt werden können*.

Die Etablierung angemessener Sicherheitsmechanismen stellt eine Herausforderung dar – nicht zuletzt, weil Verbraucherinnen und Verbraucher sehr unterschiedliche Bedürfnisse und Verhaltensmuster in ihrem individuellen digitalen Alltag haben. Insbesondere im Bereich des Staates und der Grundversorgung tragen staatliche Institutionen hierbei eine besondere Verantwortung, die an digitale Dienstleistungen gekoppelten Sicherheitsmaßnahmen für alle Bürgerinnen und Bürger gleichermaßen zugänglich und nutzbar zu machen. Der Schutz vor Cybergefahren darf nicht an bestimmte Bedingun-

[1] Bundesamt für Sicherheit in der Informationstechnik, Bonn/Freital
[2] Hochschule des Bundes für öffentliche Verwaltung, Brühl
[3] Bundesamt für Sicherheit in der Informationstechnik, Bonn/Freital

gen geknüpft sein. Als gesamtstaatliche Aufgabe gilt es hierbei durch Richtlinien und Empfehlungen sowie Beiträge zu Standardisierungs- und Regulierungsprozessen die praktische Nutzbarkeit von Schutzmechanismen voranzutreiben und zu stärken.

Die Gründe dafür, dass die praktische Nutzung von sicherheitsrelevanten Werkzeugen eingeschränkt ist, können vielfältig sein:

- Cybersicherheit ist für Anwenderinnen und Anwender in der Regel kein Selbstzweck, sondern steht meist im Dienst anderer Handlungsziele. Dies kann dazu führen, dass Anwendende Sicherheitsmechanismen als Hürde wahrnehmen und das Nutzungsverhalten entsprechend anpassen (und im schlechtesten Fall solche Hürden ganz umgehen) (siehe auch Kapitel 3.1 Handlungsfeld *Gebrauchstauglichkeit*).
- Der Kreis der Anwenderinnen und Anwender kann sehr inhomogen in Bezug auf Fachkenntnisse und Bewusstsein für IT-Sicherheitsaspekte sein. Beispielsweise kann mangelndes Verständnis für grundlegende Modelle der Public-Key-Verschlüsselung die Nutzung der Ende-zu-Ende-Verschlüsselung von E-Mails behindern. Die Schaffung von geeigneter Transparenz beispielsweise zur Funktion und Wirkungsweise von Cybersicherheit ist eine entscheidende Voraussetzung für wirksame IT-Sicherheitsmaßnahmen im Praxis-Alltag der Nutzerinnen und Nutzer (siehe auch Kapitel 3.2 Handlungsfeld *Transparenz*).
- Individuelle Einschränkungen (z.B. körperlich, emotional oder finanziell) können zu einer mangelnden praktischen Nutzbarkeit von Sicherheitsmechanismen führen. Beispielsweise wären hier Sicherheitsmechanismen betroffen, die ausschließlich auf visueller Wahrnehmung beruhen (wie z.B. einige Captcha-Tests). (siehe auch Kapitel 3.3 Handlungsfeld *Zugänglichkeit und Barrierefreiheit*).
- Häufig werden schlechte Nutzbarkeitseigenschaften als wahrscheinlichste Erklärung für eine geringe oder mangelhafte Nutzung von Sicherheitslösungen herangezogen. Jedoch können die Gründe, die zu einer geringen Akzeptanz der Sicherheitsmaßnahmen führen, vielfältiger sein und sich z.B. auch durch die emotionale Einstellung der Nutzenden ergeben (siehe auch Kapitel 3.4 Handlungsfeld *Akzeptanz*).

Der Begriff **Usable Security** bezieht sich also auf Aspekte der Cybersicherheit, die Akzeptanz und Schutzwirkung erhöhen, indem Sicherheitsmaßnah-

men und -systeme so gestaltet werden, dass sie von den Nutzenden verstanden, angewendet und in den Nutzungsalltag integriert werden können (vgl. [1, 6, 9, 11, 16]).

Usable Security ist ein wesentlicher Beitrag zum Digitalen Verbraucherschutz. Um die Bedienbarkeit von sicherheitsrelevanten Eigenschaften im Alltag der Verbraucherinnen und Verbraucher tatsächlich gewährleisten zu können, sollten die oben genannten Aspekte in eine ganzheitliche Betrachtung der Cybersicherheit von Beginn an miteinfließen und es sollte die gesamte Lebensdauer der Dienste oder Geräte berücksichtigt werden (Usable Security als Bestandteil von Security-by-Design, siehe auch [4, 13]). Zusätzlich trägt das Prinzip der sicheren Standardeinstellungen (Security-by-Default) wesentlich zur einfachen Nutzbarkeit bei. Produkte, die den Prinzipien Usable Security, Security-by-Design und Security-by-Default folgen, könnten dabei gegenüber potenziellen Käuferinnen und Käufer mit einer ganzheitlichen Cybersicherheit werben.

Usable Security wird in Digitalisierungsvorhaben in Verwaltung, Wirtschaft und Gesellschaft vermehrt gewünscht. Dies drückt sich unter anderem in der Cybersicherheitsstrategie für Deutschland 2021 [2] aus, in der die Steigerung der Anwendungsfreundlichkeit sicherheitstechnischer Lösungen ein zentrales Element der anvisierten Handlungsfelder darstellt. Auch der Cyber Resilience Act (CRA) der Europäischen Union [5] soll unter anderem die Cybersicherheit auf Verbrauchermärkten verbessern. Konkretisierende Vorschriften zur Gestaltung von Verbraucherprodukten im Sinne der Usable Security fehlen aus Verbrauchersicht allerdings, wenngleich sie für eine Verbesserung der Cybersicherheit auf Verbrauchermärkten wesentlich wären.

Trotz zunehmender Beachtung finden Aspekte der Usable Security nicht immer vollumfänglich Einzug in Digitalisierungsvorhaben. Dies liegt häufig an der weit verbreiteten Vorstellung, dass sich *Usability* und *Security* nur auf Kosten des einen bzw. des anderen Konzeptes umsetzen lassen. Jedoch gibt es längst Möglichkeiten, diese beiden Konzepte nicht nur theoretisch, sondern auch praktisch miteinander zu verbinden. Eine gesteigerte Usability von Sicherheitsmaßnahmen erhöht nicht nur die Akzeptanz, sondern häufig auch deren sichere Nutzung. Nutzungsfreundlichkeit und Cybersicherheit stehen also nicht im Kontrast zueinander, sondern gehören zusammen und können in die gleiche Richtung wirken.

2 Projekt zwischen BSI und HS Bund

Ein gemeinsames Projekt zwischen dem BSI und der Hochschule des Bundes für öffentliche Verwaltung (HS Bund) widmet sich dem Thema Usable

Security. Ziel des Projektes ist es, Prinzipien und Leitlinien für Usable-Security-Ansätze zu erarbeiten, die einerseits als strukturierte Arbeitshilfe für Mitarbeitende im BSI sowie im behördlichen und wirtschaftlichen Kontext dienen und andererseits ein stärkeres Bewusstsein für das Thema Usable Security schaffen. Langfristig sollen die Ergebnisse und Erkenntnisse auch in Form von Richtlinien etabliert werden sowie in die Standardisierung einfließen und damit als Grundlage für Produktgestaltung dienen.

Es gibt in Wissenschaft und Forschung verschiedene Definitionen für den Begriff Usable Security, die häufig unterschiedliche Dimensionen und Aspekte einbeziehen. Daher haben wir, inspiriert durch die Veröffentlichung [9], in einem ersten Teilprojekt eine Systematik entwickelt, die die wesentlichen Handlungsfelder von Usable Security definiert und dabei zentrale Aspekte beschreibt. Dies soll einerseits den Einstieg in das Thema erleichtern und andererseits zielgerichtet einen systematischen Überblick über die zentralen Faktoren geben, die einen Einfluss auf Usable Security haben. Ziel dieser Systematik ist keine wissenschaftliche Aufarbeitung des Themenfeldes, sondern eine Überführung in die Praxis durch die Schaffung eines Top-Down-Referenzrahmens.

Bei der Entstehung der Systematik haben wir uns darauf konzentriert, existierende Forschungsergebnisse und vorhandene Standarisierungen aus den Bereichen „Usable Security“ und „Human-Centered Security“ zu analysieren und in einer für das Projekt zielführenden Struktur zusammenzufassen (insbesondere [7, 8, 9, 12, 14, 15]). Die entstandene Systematik wird im nachfolgenden Kapitel vorgestellt. Sie richtet sich in erster Linie an sachkundige Anwenderinnen und Anwender, die mit grundlegenden Konzepten der Cybersicherheit vertraut sind und bestehende Sicherheitslösungen oder -systeme dahingehend evaluieren wollen, ob und inwiefern Aspekte der Usable Security Beachtung finden. Die entwickelten Dimensionen der Systematik sind jedoch sehr allgemein formuliert, sodass die einzelnen Handlungsfelder und Kriterien auch für Anwendende ohne tiefer gehende Fachexpertise verständlich sind.

3 Systematik „Usable Security: Vier Handlungsfelder“

Die im Projekt entwickelte Systematik besteht aus vier zentralen Handlungsfeldern mit je drei bis fünf zugeordneten Prinzipien. Diese vier zentralen Dimensionen bzw. Handlungsfelder müssen bei Usable Security von IT-Sicherheitsmaßnahmen beachtet werden, um eine ganzheitliche Alltagstauglichkeit der Sicherheitsmaßnahmen sicher zu stellen. Innerhalb eines Handlungsfeldes sind Einflussfaktoren zu beachten, die auf wesentliche Ziele ein-

zahlen. Die hierfür relevanten Erfolgskriterien werden hier zunächst nur genannt. Eine Ausformulierung dieser Kriterien z.B. in Form von Checklisten, die dem Systemdesign zu Grunde gelegt werden können, gehört zu den nächsten Schritten des Projekts.

Die vier Handlungsfelder und die zugehörigen Prinzipien der Systematik sind in folgender Tabelle aufgezählt:

Gebrauchstauglichkeit (Usability) • Wirksamkeit / Effektivität und Aufgabenangemessenheit • Effizienz • Robustheit gegen Nutzungsfehler • Erlernbarkeit	**Transparenz** • Informationspräsentation • Selbstbeschreibungsfähigkeit • Konsistenz / Erwartungskonformität
Zugänglichkeit und Barrierefreiheit • Einfachheit • Wahrnehmbarkeit • Anpassbarkeit • Nutzungsflexibilität • Vermeidung von Stressoren	**Akzeptanz** • Vertrauen • Joy of Use / Freude • wahrgenommene Nützlichkeit und Minimierung von Störungen

Tabelle 1: Die vier Handlungsfelder von Usable Security

3.1 Das Handlungsfeld „Gebrauchstauglichkeit (Usability)“

Die Usability (auch Gebrauchstauglichkeit oder umgangssprachlich Benutzungsfreundlichkeit) orientiert sich an den Handlungszielen, die eine Anwenderin oder ein Anwender mit der Nutzung eines Systems verfolgt. IT-Sicherheitsmaßnahmen können ein solches Nutzungsziel darstellen – beispielweise bei der Absicherung des Bankkontos vor fremdem Zugriff oder dem Einsatz einer Anti-Virus-Software. Häufig stellen IT-Sicherheitsziele jedoch Rahmenbedingungen oder untergeordnete Nutzungsziele dar und sind den primären Handlungszielen der Nutzenden nachgeordnet – wie bei dem Tätigen von Online-Überweisungen. Bei der Gestaltung von IT-Sicherheitsmaßnahmen ist daher besondere Sorgfalt zu leisten, damit diese gut in die Handlungsabläufe und den Nutzungsalltag integriert werden können und den Anwendenden schlüssig dargeboten werden.

Ohne Zweifel mögen IT-Sicherheitsmaßnahmen das eigentliche Handlungsziel mitunter behindern, können jedoch oft mit Bedacht so gestaltet und kommuniziert werden, dass eine Behinderung gering ausfällt. Usability und

IT-Sicherheit sollten demnach nicht als Balanceakte gegeneinander ausgespielt werden. Vielmehr sollten erhöhte gestalterische Anstrengungen bei der Usability von (notwendigen) IT-Sicherheitsmaßnahmen vorgenommen werden. Werden IT-Sicherheitsmaßnahmen als lästig und unnötig wahrgenommen, kann dies zu deren kreativer Umgehung durch Workarounds der Nutzenden und dadurch zu deutlichen Sicherheitseinbußen in der Nutzung führen.

Die DIN EN ISO 9241-11 [7] definiert Usability als das Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Nutzende in einem bestimmten Anwendungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen. Daraus abgeleitet sind in dieser Systematik vier Prinzipien definiert:

Wirksamkeit/Effektivität bzw. Aufgabenangemessenheit: Anwendende verfolgen mit der Nutzung eines Systems in der Regel ein bestimmtes Ziel – beispielsweise die Erledigung einer Aufgabe oder die Klärung eines Anliegens. Das System sollte Anwendende dabei unterstützen, dieses Nutzungsziel möglichst vollständig und korrekt zu erreichen. Die Nutzerinteraktion sollte dabei nicht auf der zur Erfüllung der Aufgabe gewählten Technologie basieren, sondern auf den charakteristischen Eigenschaften der Aufgabe bzw. des Prozesses beruhen. Die Aufgabenangemessenheit eines Systems besteht, wenn Anwendende nicht vom eigentlichen Aufgabenziel durch die Systembenutzung abgelenkt werden, sondern dieses Ziel durch weniger oder gleichen Aufwand erreichen und es außerdem vollständig oder zumindest genauso erledigen wie zuvor.

Erfolgskriterien für die Wirksamkeit eines Systems sind das Erreichen der Funktionalität, die Identifizierbarkeit der durch das System unterstützten Aufgaben durch die Nutzenden sowie das Setzen von Standardeinstellungen.

Effizienz: Der bei der Nutzung einzubringende Aufwand ist – neben der Zielerreichung an sich – ein entscheidender Einflussfaktor für die Gebrauchstauglichkeit eines Systems.

Erfolgskriterien für eine effiziente Nutzung sind die notwendige Bearbeitungszeit, eine Minimierung der notwendigen Interaktionsschritte, aber auch der Ressourcenverbrauch als ökologisches sowie finanzielles Kriterium. Kognitive und emotionale Ressourcen sind hier ebenso zu berücksichtigen (siehe auch Kapitel 3.3 Handlungsfeld *Zugänglichkeit und Barrierefreiheit*).

Robustheit gegen Nutzungsfehler: Nutzungsfehler können passieren und dürfen weder zu einer Schuldzuweisung oder Kriminalisierung noch zu einer

Einschränkung der Nutzung eines Systems führen. Insofern müssen Systeme so gestaltet werden, dass sie einerseits helfen vorhersagbare Nutzungsfehler zu vermeiden (z.B. durch sinnvoll gestaltete Formulareingabefelder oder durch Plausibilitätschecks der Eingaben) und andererseits auch eine Korrektur von Nutzungsfehlern zulassen oder eine alternative Lösung anbieten.

Erfolgskriterien für das Erreichen einer Robustheit gegen Nutzungsfehler sind eine Minimierung von Fehleranfälligkeit bzw. eine hohe Fehlertoleranz, Maßnahmen zur Fehlervermeidung sowie Maßnahmen zum Fehlermanagement bzw. zur Fehlerbereinigung.

Erlernbarkeit: Durch die Gestaltung der Erlernbarkeit eines Systems kann dies über die Zeit zunehmend effektiver und effizienter genutzt werden. Die Erlernbarkeit schlägt sich dabei bei der initialen Erstnutzung nieder, trägt aber auch fortwährend bei der Nutzung im Alltag positiv zur Nutzungseffizienz bei. Insgesamt fördert eine zunehmende Vertrautheit im Umgang mit einem System auch die Nutzungszufriedenheit.

Erfolgskriterien für eine leichte Erlernbarkeit sind eine assistierte Ersteinrichtung, eine Unterstützung beim Entdecken, Ausprobieren, Erinnern und Wiedererkennen von Bedienfunktionen, ein Lernen durch wiederholte Anwendung bzw. die Schaffung von Routinen, die Aufrechterhaltung (Retention) der Nutzbarkeit des Systems auch nach Phasen der Nichtnutzung sowie eine leicht verständliche Dokumentation.

3.2 Das Handlungsfeld „Transparenz“

Transparenz von Systemen und Sicherheitsmaßnahmen ist für Nutzende wesentlich, um einerseits zu verstehen, warum sie bestimmte Maßnahmen anwenden sollen, aber auch wie sie funktionieren, um ein Vertrauen in Anbieter und System entwickeln zu können. Auch ist Transparenz notwendig, um die vom System erwarteten Handlungen umsetzen zu können. Schaffung von Transparenz ist wesentlich, um den Nutzenden einen souveränen und selbstbestimmten Umgang mit Systemen und Sicherheitsmaßnahmen zu ermöglichen.

Das Handlungsfeld Transparenz bezieht sich auf die folgenden drei Faktoren:

Informationspräsentation: Die Informationen zum System und zur Nutzung des Systems sollten so dargestellt werden, dass Nutzende daraus ein Verständnis entwickeln können, was das System leistet und was als Input von den Nutzenden erwartet wird. Ein System, das sich den Nutzenden als "black box" darstellt, kann schnell zu einer Überforderung und damit zu einer geringen Nutzungsakzeptanz führen.

Erfolgskriterien sind die Verständlichkeit und Vollständigkeit der Informationen, eine Darstellung der Risiken und Vorteile von Sicherheitsmaßnahmen, aber auch die Auffindbarkeit und Inhaltsbezogenheit der Informationen für die Nutzenden.

Selbstbeschreibungsfähigkeit: Der aktuelle Systemzustand sollte für die Nutzenden erkennbar und verständlich sein. Insbesondere dann, wenn eine Interaktion erforderlich ist, müssen mögliche oder erforderliche Aktionen klar und deutlich dargestellt werden.

Erfolgskriterien für die Selbstbeschreibungsfähigkeit des Systems sind, dass Informationen Orientierung zur Verortung im System geben sollten (inkl. dem aktuellen Systemstatus), diese Informationen beherrschbar sein sollten und dass sie mittels Rückmeldungen, Hilfen usw. bei der Nutzung des Systems unterstützen. Auch die Transparenz von Datenerhebung und -verarbeitung sind für Nutzende essenziell.

Konsistenz / Erwartungskonformität: Transparenz kann sich auch auf Nutzungsprozesse beziehen. Ein System sollte in Darstellung und Verhalten vorhersagbar und konsistent im Vergleich zu Erwartungen der Nutzenden und üblichen Konventionen sein. Diese Konsistenz bezieht sich auf das System in sich, d.h. zum Beispiel gleichartige Prozesse bei der Nutzerinteraktion, aber auch im Idealfall auf Systeme untereinander (d.h. die Prozesse anderer Systeme dürfen "kopiert" werden).

Erfolgskriterien für eine solche Konsistenz sind eine einheitliche Darstellung (von Struktur/Aufbau, Begriffen/Metaphern und visuellen Elementen), Vertrautheit und Konformität mit bereits Erlerntem bzw. den Nutzenden bekannten Systemverhalten.

3.3 Das Handlungsfeld „Zugänglichkeit und Barrierefreiheit“

Es gibt zahlreiche Gründe, warum Menschen in Zeiten zunehmender Digitalisierung möglicherweise keinen sicheren Zugang zu digitalen Diensten und Technologien haben können. Die Hürden können vielfältig sein. Nach [10] gibt es folgende drei Arten der Einschränkung in Bezug auf die Zugänglichkeit von Cybersicherheit:

(1) Körperliche und kognitive Einschränkungen: Für Menschen mit Sehbehinderungen kann es schwierig sein, Sicherheitslösungen zu benutzen, die auf visueller Wahrnehmung basieren (wie z.B. Captchas). Sogar die Eingabe eines Passwortes kann zum Problem werden, wenn keine barrierefreien Funktionen wie Vorlese- oder Vergrößerungswerkzeuge zur Verfügung stehen. Zudem gehen viele Sicherheitslösungen von einer gewissen kognitiven Fä-

higkeit (z.B. durch das Erstellen von Passwörtern nach zum Teil sehr komplexen Regeln) oder auch einer gewissen Handgeschicklichkeit (z.B. im Umgang mit Maus/Tastatur) aus.

(2) Finanzielle und externe Einschränkungen: Es gibt externe Einflussfaktoren wie zum Beispiel finanzielle, geografische oder soziale Umstände, die eine eingeschränkte Nutzung von Technologien zur Folge haben können. Nicht alle Menschen verfügen über geeignete Technik, störungsfreies Internet mit ausreichend Datenvolumen oder genügend Zeit. In einkommensschwachen Schichten kann es vorkommen, dass sich mehrere Personen eine Technologie teilen oder nur eingeschränkten Zugang zu digitalen Geräten haben.

(3) Emotionale Einschränkungen: Negative Gefühle wie Sorgen, Stress, Angst oder Hilflosigkeit können den Erwerb von digitalen Kompetenzen und Wissen, um sich vor Cybergefahren zu schützen, beeinträchtigen. Sozio-emotionale Faktoren wie mangelndes Selbstvertrauen in Bezug auf eigene digitale Kompetenzen, eingeschränkte Konzentrationsfähigkeit oder ein geringes abstraktes Denkvermögen können zu einer verminderten Fähigkeit führen, potenzielle Cyberangriffe wie Viren und Phishing-Versuche zu erkennen und erfolgreich abzuwehren.

Solche und ähnliche individuellen Einschränkungen dürfen nicht dazu führen, dass der Zugang zu Werkzeugen und Hilfsmitteln, die für den Schutz vor Cyberbedrohungen erforderlich sind, ebenfalls eingeschränkt ist.

Das Handlungsfeld „Zugänglichkeit und Barrierefreiheit“ hat daher das Ziel, möglichst allen Personen in der Gesellschaft Zugang zu digitalen Angeboten zu verschaffen, aber auch sicherzustellen, dass dies ohne Angst vor digitalen Bedrohungen und Angriffen getan werden kann. In dieser Systematik werden hierfür fünf Prinzipien definiert:

Einfachheit: In der Regel besitzen Anwenderinnen und Anwender keine tiefgreifenden Kenntnisse im Bereich der Cybersicherheit. Häufig steht auch nicht viel Zeit zur Verfügung, um Dokumentationen zu lesen oder sich in komplexe Mechanismen hineinzudenken. Sicherheitsmaßnahmen sollten daher intuitiv sein und keine besonderen Anstrengungen erfordern.

Erfolgskriterien zur Erreichung von Einfachheit sind geringe kognitive und mentale Beanspruchung, geringe motorische Beanspruchung sowie eine Minimierung der zur Systemnutzung geforderten Interaktionen.

Wahrnehmbarkeit: Im Allgemeinen besagt „Wahrnehmbarkeit“ in den Web Content Accessibility Guidelines 2.0 (WCAG) [15], dass die Informationen und Bestandteile der Benutzungsschnittstelle so präsentiert werden müssen,

dass die Nutzenden sie wahrnehmen können. Unterschiedliche Wahrnehmungen, insbesondere im visuellen und auditiven Bereich, müssen daher auch bei der Konzeption von Sicherheitslösungen beachtet werden. Im Gegensatz zum Handlungsfeld Transparenz liegt der Hauptfokus hier auf den Eigenschaften der Anwenderinnen und Anwender.

Erfolgskriterien sind das Angebot von Nutzungsalternativen oder alternativen Technologien, Darstellungsalternativen oder Medien- bzw. Formatalternativen. Eine Unterscheidbarkeit zwischen einzelnen Inhalten sollte beispielsweise durch Farbkontraste und Darstellungsart sichergestellt werden.

Anpassbarkeit: Nutzende haben nicht die gleichen kulturellen und kognitiven Hintergründe, nicht die gleichen Werkzeuge und Ressourcen und auch nicht die gleichen Kontexte, in denen Sicherheitsentscheidungen getroffen werden müssen. Die Bereitstellung adaptiver und personalisierter Einstellungen und Sicherheitslösungen kann helfen, den physischen, technologischen und individuellen Nutzungskontext ganzheitlich zu erfassen. Dabei sollen Struktur, Navigation, Terminologie, Funktionalität und die Präsentation von Inhalten des genutzten Systems an die Wahrnehmung und an den Wissensstand der Nutzenden in Bezug auf sicherheitsrelevante Aufgaben angepasst werden (können).

Erfolgskriterien sind die Anwendungskontrolle, d.h. die Möglichkeit der souveränen Nutzung durch Nutzende, sowie die Berücksichtigung der Leistungsfähigkeit und der Fachexpertise der einzelnen Nutzenden.

Nutzungsflexibilität: Ähnlich wie die Anpassbarkeit hat auch die Nutzungsflexibilität das Ziel, die verschiedenen Nutzungsszenarien und -kontexte ganzheitlich zu erfassen. Während sich die Anpassbarkeit eher auf persönliche Rahmenbedingungen der einzelnen Person bezieht, geht es im Kontext der Nutzungsflexibilität um allgemeine Rahmenbedingungen, die die Umwelt der Nutzenden betreffen. So sollte beispielsweise eine Unabhängigkeit von Sprache oder von kulturellen Gegebenheiten geschaffen werden.

Erfolgskriterien für das Erreichen von Nutzungsflexibilität sind Allgemeingültigkeit, d.h. Anwendung in verschiedenen Nutzungskontexten und Kulturen, Kompatibilität und Interoperabilität mit anderen Systemen (auch mit Assistenzsystemen) und eine generelle Skalierbarkeit des Systems.

Vermeidung von Stressoren: Wenn sich Nutzende unter Druck gesetzt fühlen, kann dies dazu führen, dass alternative Bewältigungsstrategien angewendet und Sicherheitsmaßnahmen umgangen werden. Wenn Logins beispielsweise in stressigen Situationen schnell ausgeführt werden müssen, kann dies dazu führen, dass leichte Passwörter verwendet werden, die sich

schnell eintippen und leicht merken lassen, oder dass mehrstufige Authentifizierungsmethoden umgangen werden, um Zeit zu sparen. Hierunter kann auch sozialer Druck verstanden werden, der beispielweise entstehen kann, wenn Passwörter in Anwesenheit Anderer eingegeben werden müssen.

Erfolgskriterien sind die Vermeidung von Zeit- und Handlungsdruck ebenso wie von Entscheidungsdruck.

3.4 Das Handlungsfeld „Akzeptanz“

Akzeptanz kann ein wesentlicher Erfolgsfaktor für Sicherheitsmaßnahmen sein. In der Praxis mussten Entwicklerinnen und Entwickler teilweise schmerzhaft lernen, dass Sicherheitsmaßnahmen nicht umgesetzt werden, selbst wenn sie noch so gut gestaltet sind und eine hohe Wirksamkeit entfalten. Die vermeintlich weichen Akzeptanzkriterien können dann zu harten Fakten werden, ja sogar zur Ablehnung von Technologien oder Anwendungen führen. Umgekehrt kann man bei manchen wenig wirksamen Sicherheitsmaßnahmen oder gar bei fehlender Sicherheit feststellen, dass Verbraucherinnen und Verbraucher Technologien und Anwendungen nutzen, weil sie diesen (scheinbar blind) vertrauen, Freude bei der Nutzung empfinden oder mit diesen ihre Ziele schneller und besser erreichen.

Die Schaffung von Gebrauchstauglichkeit, Zugänglichkeit und Transparenz unterstützt zwar dabei, Akzeptanz bei den Anwenderinnen und Anwendern zu erreichen, doch ist dies keine Garantie für Akzeptanz. Positive Gründe für die Akzeptanz von Sicherheitsmaßnahmen können sein:

Vertrauen: Akzeptanz kann entstehen, wenn Nutzende dem System vertrauen und sich sicher im Umgang mit dem System fühlen. Auch ein Vertrauen in den Anbieter des Systems ist ein wesentlicher Faktor bei der Schaffung von Akzeptanz.

Erfolgskriterien zur Vertrauensbildung finden sich vor allem in der Art und Weise der Einführung, Heranführung und Begleitung eines (neuen) Systems. So können verschiedene Prozesse und Strukturen wie beispielsweise Methoden des Dialogs und der Einbeziehung von Anwendenden in die Systemgestaltung (User Engagement) und der Unterstützung während der Nutzung (Support) Vertrauen schaffen. Hier bietet es sich an, solche Prozesse und Strukturen gemeinsam mit Partnern bzw. Partnerorganisationen z.B. aus dem Kreis der Nutzenden durchzuführen.

Weiterhin kann die empfundene Freiheit von Risiken bei der Nutzung eines Systems ein Vertrauensfaktor sein, ebenso wie die Verbrauchersouveränität bzw. die Freiwilligkeit der Nutzung, d.h. eine unfreiwillige Nutzung kann negativen Einfluss auf das Vertrauen (und die Akzeptanz) haben.

Joy of Use: Freude bei der Nutzung unterstützt die Akzeptanz eines Systems. Positive Nutzungserlebnisse führen dazu, dass eine ggf. etwas umständliche, aber wirksame Sicherheitsmaßnahme durch die Nutzeneden gerne umgesetzt wird.

Erfolgskriterien für die Schaffung von Freude bei der Nutzung sind Attraktivität und Ästhetik eines Systems, Motivation und Stimulation zu dessen Nutzung, sowie Bequemlichkeit und Komfort der Bedienung. Wenn ein System auf die individuellen Bedarfe und Vorlieben eingestellt werden kann, so wirkt dies akzeptanzsteigernd und führt letztlich auch dazu, dass Nutzende sich mit dem System identifizieren können.

Wahrgenommene Nützlichkeit und Minimierung von Störungen: Nutzende werden Sicherheitsmaßnahmen akzeptieren, wenn sie zufrieden mit der wahrgenommenen Erreichung ihrer Ziele und des hierbei eingebrachten Aufwands sind. Ein Sicherheitsgewinn ist häufig nicht originäres Ziel der Verbraucherinnen und Verbraucher, vielmehr ist es die versprochene Leistung eines Systems, beispielsweise die Pulsmessung durch ein Wearable. Bei der Akzeptanz von Sicherheitsmaßnahme spielt es eine große Rolle, inwieweit diese als Beeinträchtigung der eigentlichen Aufgabenerledigung wahrgenommen werden.

Erfolgskriterien für eine positive Wahrnehmung können sein, inwieweit die IT-Sicherheitsmaßnahmen für die eigene Aufgabenerledigung als relevant und angemessen erscheinen (Aufgabenrelevanz der Maßnahme) und inwieweit der natürliche Workflow zum Erfüllen der eigentlichen Aufgabe durch die Sicherheitslösungen unterbrochen oder verzögert wird (Workflow-Integration). Negativ auswirken können sich z.B. notwendige Re-Authentifikation während der Nutzung eines Systems oder wenn zusätzlich physische Objekte wie etwa Token o.ä. benötigt werden.

4 Erste Erfahrungen aus der praktischen Anwendung der Systematik

Die vorgestellte Systematik kann als Basis, Ausgangspunkt oder zur Orientierung dienen, um die Cybersicherheit von Systemen auf das Qualitätsmerkmal "Usable Security" hin zu analysieren. Dabei können die vier Handlungsfelder und ihre Ausprägungen sowohl in die Gestaltung als auch in die Evaluation einfließen. Die Systematik ist auf einer überblicksmäßigen Abstraktionsebene angesiedelt. Jedoch ist bereits die Formulierung der vier Handlungsfelder und der darunterliegenden Prinzipien und Kriterien zielführend, um für konkrete Anwendungsprojekte systematisch die relevanten Gestaltungs- und Evaluationsbereiche ableiten zu können. Je nach Zielstellung, Ressourcenlage oder Priorisierungsentscheidungen kann dabei entweder die

Abdeckung aller vier Handlungsfelder oder eine Fokussierung auf einzelne Handlungsfelder angestrebt werden. Auch müssen nicht immer alle Prinzipien jedes Handlungsfeldes vollumfänglich betrachtet werden.

Als eine erste praktische Anwendung der Systematik hat das BSI eine gegenüberstellende Betrachtung und Bewertung von verschiedenen Verfahren für die Zwei-Faktor-Authentisierung im Kontext von Online-Diensten vorgenommen [3]. Dabei wurden die drei Aspekte Vertraulichkeit der Daten, IT-Sicherheit und Usable Security von verschiedenen Verfahren der Zwei-Faktor-Authentisierung betrachtet und verglichen. Bei der Betrachtung der Verfahren bezüglich Usable Security stand vor allem das Handlungsfeld Gebrauchstauglichkeit im Fokus. Hier wurde bei den einzelnen Verfahren insbesondere der Nutzungsaufwand und die Handhabbarkeit in Bezug auf unterschiedliche Nutzungsbereiche verglichen, wie beispielsweise die initiale Einrichtung des Verfahrens, dessen (all-)tägliche Nutzung, die Wiederherstellung bei Verlust oder die Migration des zweiten Faktors. Auch die Verwendungsbreite, insbesondere welche Geräte und Dienste vom Verfahren unterstützt werden, wurde betrachtet.

In zwei weiteren Projekten wird die Systematik ebenfalls aktuell zugrunde gelegt und so in der Praxis erprobt und verbessert. In einem ersten Projekt strebt das BSI an, bei der Umsetzungsstrategie der Zulassungs- und Evaluierungsprozesse von IT-Produkten für die Verschlusssachenbearbeitung (VS-IT-Produkte) die Eigenschaft "Usability" stärker zu berücksichtigen. Hierzu wird im ersten Schritt die Erwartungshaltung verschiedener Stakeholder von VS-IT-Produkten hinsichtlich der Usability erhoben. Darauf aufbauend wird ein grundsätzlicher Leitfaden für die Usability von VS-IT-Produkten erarbeitet, der als Grundlage für die künftige Umsetzungsstrategie dienen soll. In einem zweiten Projekt hat das BSI eine Verbraucherbefragung zur passwortlosen Authentisierung (insbesondere Passkeys) in Auftrag gegeben, bei der die Einstellung gegenüber diesem für den Verbraucherbereich neuen Authentisierungsverfahren sowie erste Nutzungserfahrungen, Treiber und Hürden untersucht werden. Zur Erstellung der Befragung wurden die für das Projekt zentralen Aspekte aus den vier Handlungsfeldern identifiziert und priorisiert und daraufhin in Fragen und Items im Fragebogen übersetzt. Beide Projekte sind noch nicht abgeschlossen und nicht veröffentlicht.

5 Zusammenfassung und Ausblick

Die hier vorgestellte Systematik bietet einen überblickshaften, systematischen Top-Down-Referenzrahmen zur Erschließung des Feldes der Usable Security. Die vier Handlungsfelder und ihre zugehörigen Prinzipien sind ein geeigneter Einstieg und Handlungsrahmen, um Aspekte der Usable Security

für spezifische Projekte abzuleiten und anzuwenden. Dies konnte in ersten Projekten bereits getestet werden. Die Erfahrungen daraus fließen in die weitere Arbeit ein. In ihrer aktuellen Ausgestaltung zielt die Systematik auf Cybersicherheits-Expertinnen und Experten im BSI ab und soll ihnen das Themenfeld systematisch erschließen. Später sollen auch weitere Zielgruppen in Staat, Wirtschaft und Gesellschaft adressiert werden.

Das Qualitätsmerkmal "Usable Security" soll perspektivisch auf weiteren Ebenen Berücksichtigung finden. So sollen die im Projekt entwickelten Leitlinien möglicherweise auch bei der Gestaltung von Standards und Spezifikationen (z.B. für Zertifizierungsprozesse) unterstützen oder in Regulierungsprozesse einfließen – z.B. als potenzielle Kriterien für die im Kontext des Cyber Resilience Act angedachten Standardisierungsvorhaben. Ebenso könnten Kriterien für Usable Security als Erfolgskriterien bei Vergabeprozessen der öffentlichen Hand berücksichtigt werden. Für die weiteren Schritte ist auch angedacht, eine überarbeitete Version der Systematik einer öffentlicher Kommentierung zu unterziehen.

Abschließend sei darauf hingewiesen, dass sich Usable Security in Forschung und Praxis kontinuierlich weiterentwickelt und nicht, wie der Name suggeriert, allein auf die Usability von Cybersicherheitslösungen abzielt. Vielmehr wird der Mensch zunehmend mit seinen sozio-psychologischen Fähigkeiten und Kapazitäten und in seinen vielfältigen Handlungskontexten im digitalen Alltag insgesamt wahrgenommen. Auch im Projekt soll sich daher ein Wandel von Usable Security hin zu einer ganzheitlicheren Mensch-zentrierten Cybersicherheit inhaltlich niederschlagen.

Literaturhinweise

[1] Adams, A. & Sasse, M. A. (1999). Users are not the enemy. Communications of the ACM, 42(12):40–46

[2] Bundesministerium des Inneren, für Bau und Heimat (2021), Cybersicherheitsstrategie für Deutschland 2021, abrufbar unter https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2021/09/cybersicherheitsstrategie-2021.pdf?__blob=publicationFile&v=2

[3] Bundesamt für Sicherheit in der Informationstechnik (2023). Technische Betrachtung - Wie sicher sind die verschiedenen Verfahren der 2-Faktor-Authentisierung (2FA)?, abrufbar unter https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Zwei-Faktor-Authentisierung/Bewertung-2FA-Verfahren/bewertung-2fa-verfahren_node.html

[4] Cybersecurity & Infrastructure Security Agency et al. (2023). Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Secure by Design Software, abrufbar unter https://www.cisa.gov/resources-tools/resources/secure-by-design

[5] Europäischen Union (2023). Cyber Resilience Act, Entwurfsfassung vom 20.12.2023, abrufbar unter https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=consil%3AST_17000_2023_INIT

[6] Garfinkel, S. & Lipford, H.R. (2014). Usable Security - History, Themes, and Challenges. Morgan & Claypool.

[7] ISO 9241-11:2018 - Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts, abrufbar unter https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

[8] ISO 9241-110:2020 - Ergonomics of human-system interaction — Part 110: Interaction principles, abrufbar unter https://www.iso.org/obp/ui/#iso:std:iso:9241:-110:ed-2:v1:en

[9] Lennartsson, M., Kävrestad, J., & Nohlberg, M. (2021). Exploring the meaning of usable security – a literature review, Information and Computer Security, Vol. 29 No. 4, pp. 647-663. https://doi.org/10.1108/ICS-10-2020-0167

[10] Renaud, K. & Coles-Kemp, L. (2022) Accessible and Inclusive Cyber Security: A Nuanced and Complex Challenge, SN Computer Science 3, 346, https://doi.org/10.1007/s42979-022-01239-1

[11] Reuter, C., Lo Iacono, L. & Benlian, A. (2022) A quarter century of usable security and privacy research: transparency, tailorability, and the road ahead, Behaviour & Information Technology, 41:10, 2035-2048, doi: 10.1080/0144929X.2022.2080908

[12] Schmitt, H., Gorski, P. & Lo Iacono, L. (2015). USecureD-Qualitätsmodell V.2, abrufbar unter https://www.usecured.de/wp-content/uploads/2022/10/E-1.4-USecureD-Qualitatsmodell-V.2.pdf

[13] TeleTrusT (2023). Handreichung "Security by Design" - Leitfaden für die Entscheidungsebene. https://www.teletrust.de/fileadmin/user_upload/2023-TeleTrusT-Handreichung_Security_by_Design.pdf

[14] Venkatesh, V. & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions, Decision Sciences, Vol. 39 No. 2, pp. 273-315, abrufbar unter https://core.ac.uk/display/144826641

[15] Web Content Accessibility Guidelines (WCAG) 2.0. (2008). abrufbar unter https://www.w3.org/Translations/WCAG20-de/

[16] Whitten, A. & Tygar, J.D. (1999). Why Johnny can't encrypt: A usability evaluation of PGP 5.0., In Proceedings of USENIX 1999. Washington, D.C.: USENIX Press, pp. 169–184

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Work Based Human Factor: Vom Mensch als Störfaktor zum Mensch als Sicherheitsgewährleister

Dr. Dennis Eckhardt[1], Nelli Feist[1], Prof. Dr. Sabine Pfeiffer[1]

Kurzfassung:
Dieses Paper erforscht die Transformation der menschlichen Rolle von einem als störend betrachteten Faktor zu einem zentralen Sicherheitsgewährleister. Human-Factors-Ansätze müssen über individuelles Wissen hinausgehen und Sicherheit als integralen Bestandteil alltäglichen Arbeitshandelns begreifen. Mit ethnografischen Methoden wird gezeigt, wie Bewohnende in Smart Homes eigenständig Sicherheitskonzepte entwickeln. Es wird mit arbeitssoziologischen Fallstudien in Unternehmen verdeutlicht, wie Mitarbeitende durch selbst erworbenes Wissen zu Sicherheitsexperten werden und durch kontinuierlichen Austausch zwischen Abteilungen zur effektiven Umsetzung von Sicherheitsmaßnahmen beitragen. Diese um das menschliche Arbeitsvermögen erweiterte Perspektive auf Human Factors trägt dazu bei, die Gestaltung von Sicherheitsmaßnahmen praxisorientierter und nachhaltig wirksamer zu gestalten.

Stichworte: Arbeitssoziologie, Arbeitsvermögen, Ethnografie, Human Factors, IT-Sicherheit in Smart Homes, Smart Homes

1 Der Mensch als ‚Störfaktor'

Ob bei großtechnischen und/oder High-Risk-/High-Consequences-Systemen oder bei Sicherheitsfragen: Meist wird der Mensch als individueller Störfaktor konzipiert [1] – unberücksichtigt bleiben meist die Arbeitsprozesse und -tätigkeiten und die sich in ihnen entwickelnden Kompetenzen. Dabei liegt gerade hier ein Sicherheitspotenzial, das Awareness nicht nur auf den Aspekt eines ‚Wissens um' Sicherheitsrisiken beschränkt. Human Factors greifen zu kurz, wenn sie den Menschen lediglich als (Nicht-)Wissensträger konzipieren oder nur auf die Gestaltung der Technik blicken. Sicherheit entscheidet sich in einer Welt der vielfältigen und sich dynamisch ändernden Risikofaktoren zunehmend auch auf der Ebene des Handelns – und damit erlangt das alltägliche Arbeitshandeln und nur in konkreten Kontexten entstehende Erfahrungswissen eine bislang in der Sicherheitsforschung und -umsetzung unterschätzte Bedeutung. Zwar versteht sich die Human-Factors-Forschung generell als interdisziplinär, konzentriert sich dabei aber eher auf die klassischen und eher arbeitsteilig vorgehenden „Basisdisziplinen" der Rechtswissenschaften, Ergonomie, Medizin, Pädagogik, Psychologie oder auch Arbeitswissenschaften [2]. Erst in den letzten Jahren werden auch vermehrt

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Soziologie (Technik – Arbeit – Gesellschaft)

sozialwissenschaftliche und ethnologische Ansätze und Methoden eingesetzt, ohne dass diese Disziplinen in ähnlicher Weise als ‚natürliche Partnerinnen' der Human Factors erscheinen (siehe Kapitel 4 und 5).

In unserem Beitrag zeigen wir das Potenzial einer um Arbeit(svermögen) erweiterten Perspektive für die Human-Factors-Forschung und -gestaltung auf, bei der Aspekte des Arbeitshandelns, von Arbeitspraktiken und den Arbeitsbedingungen in Erhebung und Gestaltung einbezogen werden. Arbeitssoziologische Fallstudienforschung in Betrieben verbinden wir mit ethnografischer Feldforschung in Haushalten. Human-Factors-Perspektiven integrieren damit nicht nur den realen Arbeitskontext in allen Dimensionen, sondern auch informelle Arbeitspraktiken und lebendiges Arbeitsvermögen als Ressourcen der Gestaltung im eigenen Zuhause und am Arbeitsplatz.

In diesem Beitrag zeigen wir anhand von Haushalt und Betrieb, wie und warum man den Menschen eher als Sicherheitsgewährleister denn als Störfaktor verstehen sollte. Zunächst gehen wir auf das Konzept des Arbeitsvermögens und die Erhebungsmethoden ein. Im Anschluss zeigen wir Beispiele aus den empirischen Fällen und schließen mit einem Fazit.

2 Das Potenzial von Arbeitsvermögen und Arbeitshandeln

Gemeinhin fragen Human Factors, Awareness-Schulungen oder die IT-Sicherheit nach dem messbaren Wissen von Menschen in Bezug auf Sicherheit. Außer Acht gelassen wird dabei, dass Menschen allein oder auch mit anderen fähig sind, in Situationen, in denen sie mit (auch bislang unbekannten) Herausforderungen konfrontiert sind, souverän zu agieren, sich mit den Anforderungen aktiv auseinanderzusetzen und dabei auch fähig sind, sich selbstständig Kompetenzen anzueignen, die sie für ihre Aufgaben benötigen. Diese situativen, nicht-routinisierten, nicht-formalisierten und nicht von Geschäftsleitung oder Behörden vermittelbaren Kompetenzen fassen wir im Begriff des Arbeitsvermögens. Dieses bezeichnet eine Fülle von nicht immer klar benennbaren Praktiken oder Wissensformen, die eher implizit, im Körper der Menschen oder zwischen Mitarbeitenden gewusst, einfach umgesetzt werden oder auch mal ‚im Gefühl' zu finden sind: „Solange das Subjekt lebt und in irgendeiner Form mit Welt umgeht, verausgabt und bildet sich Arbeitsvermögen." [3]

Das Arbeitsvermögen bezeichnet ein Bündel an Erfahrungen und Erfahrungswissen, das nur situativ und im Konkreten entstehen kann. Und obwohl schon lange belegt ist, dass dieses Arbeitsvermögen auch im Umgang mit IT [4] entsteht und sich sowohl in großtechnischen Hochrisiko-Anlagen [5] als auch bei der Implementierung und dem Monitoring von KI-Systemen im

betrieblichen Kontext [6], [7] von besonderer Bedeutung zeigt, wird diese besondere Kompetenz menschlichen Arbeitshandeln in Bezug auf IT-Sicherheit bislang kaum als Ressource adressiert. Kampagnen, die z.B. über Passwortschutz informieren wollen, adressieren nur kognitives Wissen und adressieren keine Rückbindung an das Handeln in konkreten Arbeitssituationen. Auch Studien, die danach fragen, wann und warum Menschen dubiose Links anklicken suchen eher nach individual-psychologischen Hürden denn nach kontextadäquaten Lösungen. Die Arbeitssoziologie kann dagegen aus der Perspektive des Arbeitsvermögens darstellen, wie Menschen über ihr Erfahrung-Machen, alltägliche Situationen und ständiges Lernen bewältigen und oft arbeits- und kontextbasierte eigenlogische Lösungswege entwickeln, die den Anforderungen in Betrieb, Haushalt, Familie und Kolleg:innenschaft nachhaltig gerecht werden – auch deswegen, weil es das menschliche Arbeitsvermögen braucht, um die oft nicht ausreichend integrierten technischen Lösungen an die komplexen und sich ständig wandelnden organisationalen Bedingungen (immer wieder) rückzubinden [7].

3 Betriebsfallstudie und Ethnografie

In der Arbeits- und Industriesoziologie ist die Betriebsfallstudie fest im methodischen Kanon verankert [8]. In dieser Erhebungsart werden für spezifische Fragestellungen charakteristische und teils kontrastierende Betriebsfälle ausfindig gemacht und dort die Perspektive unterschiedlicher Beschäftigtengruppen und Abteilungen aber auch der Führungsebene und der Interessenvertretung aufgezeigt und verglichen. Sie betont die Kontextbezogenheit des Falles und versucht diesen multiperspektivisch auch durch die methodische Triangulation von qualitativen (beispielsweise Interviews) und quantitativen Methoden (Umfragen) aufzubereiten.

Die diesem Artikel zugrundeliegende arbeitssoziologische Empirie basiert auf Erhebungen in zwei Fällen mit unterschiedlicher Risikoaufschichtung: Krankenhaus (KRITIS) und Lebensmitteldiscounter (KMU). Die hier verwendeten neun Interviews konzentrieren sich auf den Umgang mit IT-Sicherheit aus der Perspektive unterschiedlicher Arbeitsbereiche wie etwa der technischen Wartung (IT), der Verwaltung oder konkreter Bereiche (wie Radiologie oder Lagerhaltung leicht verderblicher Ware).

Die Ethnografie ist dagegen als ein Methodenbündel zu verstehen, das aus ethnografischen Interviews [9], [10], Feldnotizen [11] und Teilnehmender Beobachtung [12] besteht. Sie erhebt explorativ, induktiv und ist dabei radikal gegenstandsbezogen. In der ethnografischen Feldforschung wurden zehn Haushalte besucht. Diese wurden mit einem Schneeballsystem ausfindig gemacht und nach technischer Kontrastion gewählt: Haushalte mit lediglich

einem Saugroboter bis hin zum Selbstbastler-Haushalt (demografische Angaben dienen nicht der Komparatistik oder Typisierung). In Bezug auf IT-Sicherheit und Smart Homes wächst die Anzahl der ethnografischen Studien seit Jahren an [13], [14]. Noch konnten sich diese Ansätze vor allem in Deutschland nicht breit durchsetzen, sodass sie nach wie vor innovativ wirken.

4 Arbeiten am Zuhause: Ethnografische Feldforschung in Smart Homes

Anthropologie forscht immer auch im und mit dem Zuhause der beforschten Feldteilnehmenden. Claude Lévi-Strauss konstatierte, dass wir in einer „société à maisons" – also einer Gesellschaft in Häusern – leben [15]. Allerdings ist der Begriff Zuhause oder ‚Home' wie kaum ein anderer nur schwer zu fassen: Das Zuhause kann der Ort von Gemütlichkeit und Rückzug sein, aber auch von Terror und Gewalt. Es kann gänzlich frei von Technik oder davon überladen und überformt sein. Nichts lässt sich daran klar definieren: „the meaning and study of home ‚all depends'", hielt daher Shelley Mallett fest [16]. Eine Anthropology of Home entwickelt daher einen Fokus auf das Verständnis von Zuhause mehrheitlich auf symbolischer Ebene [17], worin das Verständnis von ‚Home' ethnografisch erfasst wird. Diese Ansätze haben sich auf die Forschung von Zuhause und Migration [18], [19] Obdachlosigkeit in Europa [20], Gewalt im Zuhause [21] oder sogar zum „domicide" [22] – also der Zerstörung von Zuhause – und der Kulturgeografie und der Soziologie erweitert. Alle diese interdisziplinären Studien zu ‚Home' betonen das „home-making" [23], [24] als eine zentrale Praxis von Menschen, die auf das Zuhause Zugriff nehmen (wollen): Sowohl regulativ von staatlicher Seite oder anderen Akteuren – also eher von außen –, als auch einrichtend und aushandelnd von innen: Das „home/house as a node in a net of relations, as well as a structure with an inner logic and action in the world" [24].

Für die IT-Sicherheit ist es daher produktiv, User:innen als Home-Maker und als Bewohnende zu verstehen, die Zugriffe auf ihr Haus regulieren, Funktionen von Techniknutzung im Haushalt aushandeln, Regulationen von außen interpretieren, einbetten oder abwehren, sowie Beziehungen zu anderen Familienmitgliedern, Freunden, möglichen Eindringlingen oder unliebsamen Gästen zwar nicht immer rational, aber eigenlogisch und selbsttätig gestalten. Home-Maker werden noch nicht ähnlich komplex verstanden und untersucht, wie die eigentliche Technik oder IT-Sicherheit. Anders gesagt wird zu wenig untersucht wie viel Arbeit Menschen in die Herstellung ihres Zuhauses stecken – das Home-Making –, worin die IT-Sicherheit eigentlich nur eine Dimension unter vielen ist. Um zu verstehen, was Menschen im Haushalt mit IT-Sicherheit (nicht) machen, ist es daher notwendig die Arbeit am Zuhause

als solche zu beobachten und für Konzepte der IT-Sicherheit produktiver zu nutzen als bisher.

4.1 Basteln, recherchieren, selber machen: Technikintegration im Wohnalltag

Diese Arbeit am Zuhause stellen wir nun exemplarisch an einem der untersuchten Haushalte vor. Dieser Haushalt befindet sich in einer deutschen Großstadt und besteht aus drei Personen: den Eltern und einem Kleinkind. Beide Eltern sind berufstätig, wobei nur der Beruf des Familienvaters erhoben wurde, welcher als Psychologe in der Forschung tätig ist. Der Haushalt hat eine längere Geschichte der Technisierung. Der Gesprächspartner, welcher der Familienvater war, berichtete dem Ethnografen im Gespräch von der Entwicklung des Haushalts hin zu einem Smart Home. Für ihn fing alles mit der Heizung an, da er in der zugigen Altbau-Wohnung oft Probleme mit der Temperatur hatte: „Das Thema Heizen ist irgendwie präsent. Man hat zugige Fenster, es ist immer irgendwie ein bisschen kühl." (Interview vom 07.07.2023 mit Haushalt 6), wie er es im Interview ausdrückte. Er selbst „kam auf die Idee, einfach so einen Schalter zu bauen, dass wenn ich gegangen bin, neben der Tür ich dann draufdrücken konnte und dann ging so alles auf Absenk-Betrieb. Und wenn man zurückkam, musste ich nur auf diesen Schalter drücken und dann war alles quasi wieder so im Präsenzmodus." (Interview vom 07.07.2023 mit Haushalt 6) Der Familienvater sammelte daraufhin erste Erfahrungen mit Smart Home, vor allem mit dem eigenen Automatisieren von Vorgängen, um lästige Prozesse im Alltag zu entfernen oder leichter zu machen.

In der ethnografischen Feldforschung berichteten viele Personen davon, dass sie sich im Laufe der Zeit Fähigkeiten selbst anlernten, um Technik zu installieren, miteinander zu integrieren oder auch Dinge selbst zu programmieren. Völlig unterschätzt wird in der Forschung die reale Rolle von Reddit, Telegram, Discord und YouTube in diesem Feld. Gerade auf der Videoplattform existieren dutzende Kanäle, die erklären, wie man das eigene Zuhause automatisieren kann. Gesprächspartner:innen geben immer wieder an, zumindest phasenweise YouTube-Videos oder Online-Foren mit gesuchten Inhalten zu konsultieren, zu lesen und sich dabei selbst „aufzuschlauen". In einem anderen Haushalt wurde angegeben gerade während der Covid-19-Pandemie die Zeit genutzt zu haben, um Videos zu schauen, die erklären, wie man ein Raspberry Pi installiert und HomeAssistant programmiert. Die Anleitungen müssen dabei von den Nutzenden oft erst passfähig für den eigenen Kontext gemacht werden. Wie stark die Smart-Home-Community auch von diesen Plattformen und Wissensaustauschformaten befördert wird und

wie viel eigenständiger, eigenlogischer sowie pro-aktiver Kompetenzerwerb um diese Angebote herum entsteht, wird in Human Factors nicht wahrgenommen. Dabei bauen Bewohnende darüber ihre technischen Kompetenzen auf, basteln und integrieren selbstständig Technik im Haushalt, selbst dann, wenn ihnen dafür die technische Ausbildung fehlt. Sie integrieren aber dennoch Technik im Haushalt – aus Spaß und Neugier und/oder um konkrete Probleme ihres Umfelds im gewünschten Sinne zu lösen.

Ein Grund dafür kann die Außenwelt sein. So litt der befragte Familienvater nach dem Umzug in eine neue Wohnung unter der Dunkelheit im Inneren. Die Ausrichtung der Fenster, die Lage der Wohnung und Bäume vor dem Haus führten dazu, dass es immer wieder zu dunkel war. Dies führte zu einer Vielzahl von ‚Experimenten' mit Technik im Haushalt. Die so erlernten Kompetenzen kamen dann nach einem weiteren Umzug in eine deutlich größere Wohnung so richtig zum Tragen. Die Möglichkeit, nun alles völlig neu einrichten zu können sah der Familienvater als Chance und hat „dieses Thema Smart-Home, noch mal für [s]ich neu gedacht und überlegt, was brauche ich wirklich? Also was ist eine sinnvolle Automatisierung, die mir hilft?" (Interview vom 07.07.2023 mit Haushalt 6) Von diesem Ausgangspunkt legte er sich einen eigenen Smart-Home-Server mittels eines Raspberry Pis zu, um Geräte im Haushalt herstellerunabhängig miteinander zu verbinden. Der Aufbau wurde im Laufe der Zeit immer größer: Sogar eine Bewässerungsanlage im hausgemeinschaftlich genutzten Garten wird mittlerweile mit dem eigenen Setup gesteuert.

Der Haushalt erscheint damit – wie es die Literatur zu Home-Making eingangs nahelegte – als ein multi-dimensionales Feld, und nicht als einfacher ‚Use Case'. Der Familienvater „schlaut" sich selbst auf, experimentiert und entwickelt dabei technische Kompetenzen, mit denen er auch einen eigenen Server programmieren kann. Im Server sind dabei nicht nur eigene Geräte eingebunden, sondern auch Geräte, die für die ganze Hausgemeinschaft nützlich sind, wie die Bewässerungsanlage im Garten. Dabei zeigt sich nicht nur, dass der hier beschriebene Home-Maker – der Familienvater – auch das Zuhause für die Nachbarschaft mit „herstellt", sondern sich auch über diverse Sicherheitsprobleme bewusst ist. Mit einem „Hack", wie er es nennt, hat er die Geräte mit HomeKit auf einen Apple-TV integriert. Alle Geräte funken daher nicht selbst ins Internet, sondern sind nur lokal miteinander verbunden. Der Familienvater umgeht damit das Problem zu viele Anbieter zu haben, bei denen unklar ist, wie Sicherheitsstandards behandelt werden: „Und das habe ich jetzt zumindest so auf eine kritische Verbindung für mich begrenzt, und so fühlt sich das auch okay an von der Sicherheitskomponente." (Interview vom 07.07.2023 mit Haushalt 6)

Der Familienvater äußerte wiederholt den Wunsch, die Technik möglichst weitgehend selbst in der Hand zu haben, muss dabei allerdings immer wieder Kompromisse eingehen – in diesem Fall mit der Begrenzung auf einen Anbieter (Apple). Er ist sich aber der Sicherheitsprobleme bewusst, die ein Zuviel an Herstellern mit sich bringt, wenn letztere möglicherweise ihre Geräte nicht regelmäßig aktualisieren. Das Ausmaß von individuell investierter Arbeit, Energie und Wissen in die Automatisierung des Haushalts zeigt sich in diesem Fall als eine Arbeit am Zuhause, die mal experimentierend, mal recherchierend und „aufschlauend" geschieht: „[m]aking home is a process of labour" [25].

Der Gesprächspartner zeigt eine breite Palette an Möglichkeiten des Home-Makings, die für eine Work-Based-Human-Factors-Forschung relevant sind: (1) Das Zuhause und die Technik im Zuhause wachsen inkrementell. Sie bauen aufeinander auf, werden neu installiert und auch wieder verworfen. Es entstehen komplexe technische Arrangements. (2) Die Rolle des Zuhauses ist noch unterbewertet. Ein- und Umzug in Wohnung, Lichtverhältnisse, aber auch Lärm und Temperatur sind maßgebliche Auslöser für Automatisierungen. (3) Im Zuhause sind nie nur ein User und ein Gerät gemeint, sondern immer eine komplexe Ansammlung von Technik und anderen Menschen, wie Nachbarn oder Familienmitgliedern. (4) Home-Maker „schlauen sich selbst auf", und holen sich schon längst das Wissen und die Kompetenzen, die sie benötigen, um ihre Haushalte zu technisieren. Offentliche Stellen werden hier nur schwer die Rolle von Peer-to-peer-Ansätzen einnehmen oder ersetzen können. (5) Home-Maker sind bereits jetzt die Sicherheitsgewährleister, nach denen die IT-Sicherheit sucht. Sie täte gut daran das Zuhause als ein „set of social potentials" [26] zu verstehen, und nicht als ein Problem, eine Lücke in Sicherheitsketten, oder gar als DAUs – ‚dümmste anzunehmende User'. Im Gegenteil ist die Realität von Home-Makern schon längst eine andere: Sie integrieren Technik und bedenken – natürlich nicht immer und überall – Sicherheitsprobleme. Will die IT-Sicherheitsforschung tatsächliche Sicherheitsprobleme angehen und verbessern, muss sie erstens verstehen, wie die Arbeit am Zuhause genau aussieht, und zweitens mit diesen Menschen zusammenarbeiten, und sie nicht mono-dimensional als Vermittlungsendpunkte von Sachwissen verstehen. Daraus ließen sich auch tragfähigere Maßnahmen ableiten, um die Bewohnenden zu adressieren, die sich zunächst nicht aus eigenem Antrieb in das Thema vertiefen oder sogar unfreiwillig in ihrem Zuhause mit IT-sicherheitsrelevanter Technik konfrontiert sind.

5 Unsichtbare ‚Sicherheitsgewährleister': Arbeitssoziologische Fallstudien

Die arbeitssoziologischen Fallstudien setzen den Menschen und seine Erfahrungen in das Zentrum der Analyse und des Erkenntnisgewinns, insbesondere im Kontext der in den letzten Jahrzehnten zunehmenden Subjektivierung und Entgrenzung von Arbeit [27], [28]. Auch im Hinblick auf eine nachhaltige Implementierung von Sicherheitsmaßnahmen lohnt sich der Blick in die Betriebe aus einer arbeitssoziologischen Perspektive. Mithilfe der Methoden-Triangulation von Interviews und Umfragen lässt sich nachvollziehen, was im Arbeitsalltag tatsächlich praktiziert wird. Diese Erkenntnisse lassen sich insbesondere für eine bessere Implementierung von IT-Sicherheitskonzepten nutzen, wenn die konkreten Arbeitsanforderungen der Beschäftigten von Anfang an anerkannt und systematisch berücksichtigt werden. Ublicherweise aber werden Sicherheitsmaßnahmen auf Managementebene entschieden und ‚von oben' (top-down) ausgerollt. Teilweise geschieht dies noch nicht mal in Absprache mit dem CISO der IT-Abteilung, eine partizipative Integration der Beschäftigten in den betroffenen Abteilungen bleibt überwiegend aus. Beschäftigte geraten, wenn überhaupt, als Störfaktor in den Blick, die es gilt, durch Awareness-Kampagnen zu IT-Sicherheitsthemen unabhängig von ihrem konkreten Arbeitskontext zu schulen (sprich: sie als Adressat vorgefertigter und abprüfbarer Wissensmodule zu sehen). Das gilt umso mehr, wenn die Management-Ebene sich dem Thema IT-Sicherheit nicht aus genuinem Interesse an der Vermeidung von realen Risiken zuwendet, sondern solche Maßnahmen stärker als Compliance-getriebene Aufgabe sieht, die nachweisbar zu erledigen ist statt sie realitätsnah in den Arbeitsalltag zu integrieren.

Sicherheitsmaßnahmen, die langfristig tragen sollen, dürfen den konkreten Arbeitsalltag nicht einschränken, behindern oder mit zu viel Zeitaufwand belasten. Oftmals aber stehen solche Maßnahmen sozusagen „neben" den eigentlichen und notwendigen Arbeitspraktiken der Beschäftigten. Das zeigt sich deutlich in unseren Interviews im Rahmen von Awareness-Kampagnen. Umgekehrt werden sicherheitsrelevante Hinweise und Nachfragen aus den Arbeitsbereichen, die zu einer substanziellen und real gelebten (statt nur reklamierten) Herstellung von sicheren Arbeitsumgebungen führen könnten, oft nicht gehört bzw. es werden gar keine Räume und Prozesse geschaffen, in denen dies passieren könnte. Gerade in der Kritischen Infrastruktur zeigt sich ein erhöhtes Sicherheitsbedürfnis beim Daten-Management z.B. von sensiblen Patientendaten etc., da Datenschutz neben der Gesundheitsversorgung die oberste Priorität hat, wie ein Informationssicherheitsbeauftragter (ISB) eines Krankenhauses bestätigt. Im Lebensmittelbereich bedingen sich verschiedene Sicherheitsdimensionen zudem wechselseitig: neben reinen

IT-Sicherheitsfragen geht es auch um Lebensmittelsicherheit und Versorgungssicherheit bei teils schnell verderblicher Ware.

Der anhaltende Fachkräftemangel macht sich ebenfalls bemerkbar und wird auch in den Interviews als problematisch thematisiert. IT-Fachkräfte sind gefragt, die Abteilungen häufig unterbesetzt. Laut einer Umfrage im Auftrag des Digitalverbands Bitkom blieben 149.000 Stellen für IT-Fachkräfte im Jahre 2023 unbesetzt.[2] Im Zuge der zunehmenden Digitalisierung von Unternehmen und Verwaltungen wird diese Nachfrage so schnell nicht abnehmen, sondern im Gegenteil mit dem Einzug weiterer Technologien und deren Vernetzung untereinander noch weiter steigen.[3] Zudem wandelt sich das Feld der IT-Sicherheit in Unternehmen erst in jüngster Zeit zu einem attraktiveren Arbeitsplatz innerhalb der IT-Tätigkeiten [29].

Beschäftigte außerhalb der IT können den Mangel an IT-Sicherheitsfachkräften natürlich nicht kompensieren. Wird ihr Arbeitsvermögen aber als Ressource für IT-Sicherheit stärker einbezogen, entlastet das die IT-Sicherheitsfachkräfte und macht zudem deren Tätigkeit attraktiver. In beiden hier empirisch kurz vorgestellten Organisationen zeigt sich, wie das Arbeitsvermögen als zentrale Ressource zur Wartung und Administration von IT-Sicherheit eingesetzt wird. So schildern im Interview etwa Mitarbeitende aus der Radiologie, dass ihr privat angeeignetes Wissen sie dazu motiviert, sich mit IT-Sicherheitsthemen auch innerhalb der eigenen Abteilung weiter auseinander zu setzen und dies mit dem bestehenden Workflow zu verbinden. Sie treten darüber hinaus pro-aktiv mit der krankenhauseigenen IT-Abteilung in Kontakt und stellen so einen anhaltenden Austausch her, der das technische Wissen mit dem betrieblichen Handeln verbindet. Sie gewährleisten damit bereits jetzt Sicherheit im Krankenhaus und erscheinen als weitere Ansprechpartner:innen für spezifische Fragestellungen innerhalb der Abteilung.

In ähnlicher Weise findet so ein Vorgehen auch im Lebensmitteldiscounter statt. Der strukturelle Unterschied besteht darin, dass im Lebensmitteldiscounter eine zentrale IT-Organisation existiert, die Standorte koordiniert und in den Außenstellen Personen in einer Doppelrolle einsetzt. So wird sichergestellt, dass auch in den Niederlassungen eine Ansprechperson für die örtlichen Belange kontaktiert werden kann. Ähnlich zur Radiologie wird zwar aus der selbstgetriebenen Motivation der Mitarbeitenden, sich mit dem Thema zu beschäftigen, geschöpft. Es wird aber vor allem versucht, mit der

[2] https://www.bitkom.org/Presse/Presseinformation/Rekord-Fachkraeftemangel-Deutschland-IT-Jobs-unbesetzt#

[3] https://www.get-in-it.de/magazin/arbeitswelt/it-arbeitsmarkt/so-sieht-der-it-arbeitsmarkt-aus

vorhandenen Arbeitskraft vor Ort möglichst effizient umzugehen. Beide Fälle illustrieren, wie wichtig bereits jetzt die Arbeit ist, mit der KRITIS und KMU sicher gehalten werden. Zur Etablierung einer sicheren Arbeitsumgebung und Instandhaltung dieser gibt es die unsichtbaren ‚Sicherheits-Gewährleister', die an der Schnittstelle der eigenen Abteilung zur Abteilung der IT-Sicherheit sitzen.

Diese ‚unsichtbaren Sicherheitsgewährleister' spielen eine tragende Rolle, wenn es um das Absichern der Netzwerke vor Ort geht. Dabei ist diesem Austausch das Spezialwissen der Mitarbeitenden an der Schnittstelle zur zentralen IT inne. Das bedeutet im Detail, dass in den hochentwickelten und spezifischen Arbeitsumgebungen, wie es beispielsweise in der Radiologie der Fall ist, das Wissen des dort beschäftigten Radiologen in einen Austausch mit der übergeordneten IT-Instanz gerät:

> *Also im Wesentlichen bin ich Medizinphysiker. Das bedeutet andererseits natürlich, ich habe eigentlich jetzt mal formal IT-mäßig in keinster Weise irgendwie eine Ahnung. Nur irgendwer muss es machen. Und da war ich dann halt anscheinend der, der es noch am besten wusste und der Spaß daran hat. Und über die letzten zehn Jahre hab ich es mir dann halt so draufgeschafft, sag ich mal.*
>
> Medizinphysiker in der Radiologie

Aus der Arbeitspraxis heraus ist der Radiologe Experte hinsichtlich der ihm zur Verfügung stehenden Geräte, mit denen er tagtäglich arbeitet. Dieses spezielle Wissen aus der Abteilung bleibt dem CISO verborgen. Im Arbeitsalltag des CISOs ist dieses Spezialwissen für ihn zunächst irrelevant, da er aber in die einzelnen Abteilungen einwirkt, wird die Rückmeldung aus den Abteilungen für ihn wichtig, um Anpassungen vorzunehmen und mögliche Sicherheitslücken zu schließen. Häufig zeigen sich IT-Sicherheitsmaßnahmen als übergestülpt und fernab von der Arbeitspraxis, was dazu führen kann, dass sie vernachlässigt oder nur mit großem Arbeitsaufwand integriert werden. Jede Organisation muss vor ihren kontextuellen Bedingungen und Strukturen betrachtet werden [30], jedoch zeigen die Einblicke unserer Erhebung, dass solche Kooperationen in Betrieben durchaus bestehen. Sie sind kein Einzelfall, auch wenn sie unterschiedlich durch die Beschäftigten gestaltet werden.

Unsere Empirie zeigt: Eine auf IT-Sicherheit bezogene Kooperation und Kommunikation über die Arbeitsbereiche hinweg wirkt selbst in dynamisch sich verändernden Situationen positiv auf die IT-Sicherheit; hat sich ein solcher Austausch erst einmal informell etabliert, profitieren beide Seiten

davon: Die Fachabteilung kann Sicherheit arbeitsintegriert sinnvoll leben, die CISO/IT erfahren mehr und frühzeitig von sich neu ergebenden sicherheitsrelevanten Situationen vor Ort, auf die sie reagieren können. Gleichzeitig mildern solche „Doppelrollen" auch den IT-Fachkräftemangel: Wenn Beschäftigte sich in der Pflicht fühlen, ihre Kompetenzen auch neben der Haupttätigkeit zu erweitern, Schnittstellen aufzubauen und in den Dialog zu gehen, stützen sie als unsichtbare Sicherheitsgewährleister die IT im Hintergrund. Solche Tandems des gemeinsamen Arbeitens an IT-Sicherheit ermöglichen tragfähige weil an die Arbeitserfordernisse konkret angepasste Sicherheitsmaßnahmen.

5.1 Arbeitsvermögen im Krankenhaus und Lebensmit eldiscounter

Der Megatrend der Digitalisierung führt zu einem anhaltenden Wandel der Arbeitswelt und konfrontiert alle zunehmend mit der Frage nach dem Umgang mit erhöhten und sich ständig wandelnden Sicherheitsanforderungen im Unternehmen. Aber nicht nur technisch wandelt sich die Arbeitswelt, auch die Arbeit selbst verändert sich und auch damit gehen neue Sicherheitsanforderungen einher: Der Arbeitsort wird durch entgrenztes Arbeiten im HomeOffice für immer mehr Beschäftigte fluide; neben dieser Entgrenzung kommt es zunehmend zu einer Subjektivierung der Arbeitskraft: Die Regulierung und Begrenzung von Arbeit weicht auf und die Kompetenzanforderungen gehen oft weit über die Stellenaus- und -beschreibung hinaus. Hierarchischen Strukturen brechen zunehmend auf und es ist eine Rücknahme beruflicher Spezialisierungen zu sehen. Dadurch nimmt auch die Selbstverantwortung für IT-Sicherheit zu, die sich auf Basis von Marktbedingungen und personellem Druck auf die einzelnen Beschäftigten auswirkt. Die äußerlichen Markterfordernisse und Rationalisierungslogiken machen sich so auch am Beschäftigtenhandeln im Betrieb bemerkbar.

Beschäftigte sehen sich anhaltend als „Arbeitskraftunternehmer" [31], sie decken Aufgabenbereiche ab, die zuvor von einer mittleren Management-Ebene zumindest aufgefangen wurden. So müssen Beschäftigte immer mehr eigene, private Ressourcen einbinden, eigenständig Arbeitsaufgaben managen und nach Lösungen suchen – was früher nur Hochqualifizierte traf, ist längst auch Anforderung in sogenannter „einfacher" Arbeit. „Selbstorganisation bedeutet im Kern, daß [sic] die Bewältigung von Unbestimmtheit nun mehr und mehr zur Aufgabe der Beschäftigten (oder von Arbeitsgruppen etc.) wird." [32] Aus diesem Konglomerat an äußeren Faktoren, wird das „lebendige Arbeitsvermögen" einst als Störgröße, jetzt als Ressource gehandelt [33]. Diese subjektiven Potenziale könnten zielgerichteter in Kooperationen genutzt werden. Böhle beschreibt dies als kooperativen Moment, der sich am

Gegenstand der Umsetzung von IT-Sicherheit zeigt [34]. Die informelle Praxis, die sich zwischen IT und Fachabteilungen zeigt, war in unseren untersuchten Fällen vom Management nicht vorgesehen, sondern hat sich eigenständig etabliert und konzentriert sich auf den Gegenstand der Umsetzung von IT-Sicherheit im je eigenen Arbeitsbereich (wie etwa im Fall KRITIS in der Radiologie).

Wir haben hier den Fokus sehr auf die Mikroebene des Arbeitshandelns und die Bedeutung des Arbeitsvermögens gelegt. Die letzten Ausführungen zeigen aber auch: Andere Dimensionen des strukturellen Wandels von Arbeit sind als Kontextbedingungen für IT-Sicherheit unterschätzt. So müsste die Human-Factors-Forschung sich stärker mit den Dynamiken des Arbeitsmarkts und beispielsweise auch prekären Arbeitsverhältnissen auseinandersetzen. Dort wo der Fachkräftemangel auf prekäre Arbeitsbedingungen stößt, scheint die IT-Sicherheit besonders in Gefahr. Wie die Einblicke in die Erhebung gezeigt haben, gibt es Sicherheitsgewährleister in den Unternehmen, die ihr Arbeitsvermögen gezielt einsetzen. Dies aber gelingt nur Beschäftigten, die Zeit und Perspektiven haben, sich das dafür nötige Erfahrungswissen auch anzueignen. Mit Prekarität ist IT-Sicherheit daher schlecht, mit Stabilität der Arbeitsverhältnisse dagegen gut beraten. IT-Sicherheit ist viel mehr als nur eine Frage von Awareness, sie steht und fällt mit der Möglichkeit der Beschäftigten, ihr Arbeitsvermögen in die Gestaltung der IT-Sicherheit einzubringen und diese in diesem Prozess auch „on the fly“ weiterzuentwickeln. Das kann auch über Wissensaneignung im privaten Kontext passieren, darf sich darauf aber nicht beschränken. Und dort wo Subjekte den Marktdruck von außen und den Gefahrendruck im Arbeitsalltag auf sich nehmen, sollten diese auch (monetär) dafür honoriert werden.

6 Aus der Perspektive von Arbeit: Für eine Work Based Human Factors Forschung

Dieser Beitrag plädiert anhand zweier empirischer Fälle dafür, in der Forschung von Human Factors die Ebene konkreter Arbeit systematisch zu integrieren – egal ob es um die Arbeit an Sicherheit im Zuhause oder am Arbeitsplatz geht. Wir argumentieren, dass unterkomplexe Vorstellungen des ‚Users‘ – die vom jeweiligen Kontext abstrahieren – problematisch sind, weil sie nicht die tatsächliche Realität von Bewohnenden oder Arbeitenden einbeziehen. Mit der analytischen Perspektive des Arbeitsvermögens zeigen unsere empirischen Fälle, wie sich Menschen im Haushalt mit YouTube-Videos „aufschlauen“, durch Basteln und Experimentieren mit Technik verstehen lernen, was sie an Automatisierungen brauchen oder diese auch wieder verwerfen. Im Betrieb eignen sie sich Wissen privat an, geben es an Kolleg:innen

weiter und sind schon jetzt wichtige Sicherheitsgewährleister in Betrieben – ohne dass sie dafür explizit ausgebildet wurden.

Wir plädieren für die Erweiterung der Human-Factors-Forschung um den Aspekt der Arbeit. Denn nur so lässt sich forschend zeigen (und praktisch gestalten), wie IT-Sicherheit als und im Arbeitsprozess eigentlich und immer wieder aufs Neue hergestellt wird und unter welchen Bedingungen das passiert. Der Ansatz des Arbeitsvermögens ist dabei hilfreich, weil damit die nicht-quantifizierbaren und nicht-formalisierbaren Dimensionen menschlichen Arbeitshandelns in den Blick geraten – sind sie es doch, die Beschäftigte und Bewohnende in besonderem Maße in die Lage versetzen, auch in ungewohnten Situationen souverän agieren zu können und aus diesen Erfahrungen zu lernen: Wo anerkannt wird, dass der Mensch nicht Störfaktor sondern Sicherheitsgewährleister sein kann, kann eine ganz andere Qualität von Sicherheitsmaßnahmen etabliert werden. Diese können nicht nur weit über rein auf Wissen setzende Awareness-Maßnahmen hinausweisen, sie können auch die zunehmend ge- und überforderte und oft unterbesetzte IT-Abteilung entlasten.

Human Factors ist damit aufgerufen Methoden der Sozialforschung auch in arbeitsteiligen interdisziplinären Settings mit zu gestalten, und nicht nur mentale Modelle, rechtliche Bedingungen oder technische Expertisen abzufragen, sondern jenes Arbeitsvermögen zu suchen, das Arbeits- und Wohnverhältnisse alltäglich sichert.

Literaturhinweise

[1] R. A. Stephans, *System safety for the 21st century*. John Wiley & Sons, 2022.

[2] P. Badke-Schaub, G. Hofinger, und K. Lauche, „Human Factors“, in *Human Factors: Psychologie sicheren Handelns in Risikobranchen*, P. Badke-Schaub, G. Ho-finger, und K. Lauche, Hrsg., Berlin, Heidelberg: Springer, 2008, S. 3–18. doi: 10.1007/978-3-540-72321-9_1.

[3] S. Pfeiffer, *Arbeitsvermögen: ein Schlüssel zur Analyse (reflexiver) Informatisierung, 1. Aufl. Wiesbaden: Springer Fachmedien, 2004.*

[4] S. Pfeiffer, *Dem Spürsinn auf der Spur: subjektivierendes Arbeitshandeln an Internet-Arbeitsplätzen am Beispiel Information-Broking*. München: Hampp, 1999.

[5] H. Bauer, F. Böhle, C. Munz, S. Pfeiffer, und P. Woicke, Hightech-Gespür: erfah-rungsgeleitetes Arbeiten und Lernen in hoch technisierten Arbeitsbereichen. Bielefeld: Bertelsmann, 2006.

[6] S. Pfeiffer, „KI im Unternehmen – Herausforderungen an die betriebliche Gestaltung moderner Arbeit“, *DGUV forum*, Nr. 11, 2023, Zugegriffen: 1. März 2024. [Online]. Verfügbar unter: https://forum.dguv.de/ausgabe/11-2023/artikel/ki-im-unternehmen-herausforderungen-an-die-betriebliche-gestaltung-moderner-arbeit

[7] T. Herrmann und S. Pfeiffer, „Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence“, *AI & Soc*, Feb. 2022, doi: 10.1007/s00146-022-01391-5.

[8] J. Pflüger, H. J. Pongratz, und R. Trinczek, „Fallstudien in der Organisationsforschung“, in *Handbuch Empirische Organisationsforschung*, S. Liebig, W. Matiaske, und S. Rosenbohm, Hrsg., in Springer Reference Wirtschaft. , Wiesbaden: Springer Fachmedien, 2017, S. 389–413. doi: 10.1007/978-3-658-08493-6_19.

[9] J. Skinner, Hrsg., *The interview: an ethnographic approach*, English edition. in ASA monographs, no. Volume 49. London ; New York: Berg, 2012.

[10] R. Madden, *Being Ethnographic. A Guide to the Theory and Practice of Ethnography*. London, Thousand Oaks, New Delhi: SAGE, 2017.

[11] D. Eckhardt, „Ethnografisches Feldnotieren in digitalen Feldern: Perspektiven einer Wissens- und Arbeitspraxis“, *Kulturanthropologie Notizen*, Bd. 85, S. 52–77, Sep. 2023, doi: https://doi.org/10.21248/ka-notizen.85.21.

[12] P. Aktinson und M. Hammersley, „Ethnography and participant observation“, *Strategies of Qualitative Inquiry. Thousand Oaks: Sage*, S. 248–261, 1998.

[13] L. K. Aagaard, T. H. Christensen, und K. Gram-Hanssen, „My smart home: an auto-ethnography of learning to live with smart technologies“, *Pers Ubiquit Comput*, Apr. 2023, doi: 10.1007/s00779-023-01725-0.

[14] L. Kocksch und T. Elgaard Jensen, „‚Good‘ Organizational Reasons for ‚Bad‘ Cybersecurity: Ethnographic Study of 30 Danish SMEs“, Aalborg Universitet, Report, Feb. 2023. doi: 10.54337/aau513432435.

[15] J. Carsten und S. Hugh-Jones, Hrsg., *About the House: Lévi-Strauss and Beyond*, 1. Aufl. Cambridge University Press, 1995. doi: 10.1017/CBO9780511607653.

[16] S. Mallett, „Understanding home: a critical review of the literature“, *The sociological review*, Bd. 52, Nr. 1, S. 62–89, 2004.

[17] I. G. Cieraad, „Introduction: Anthropology at home“, in *At home: an anthropology of domestic space*, Syracuse University Press, 2006, S. 1–12.

[18] M. Fathi und C. Ní Laoire, „Conceptualising Home in Migration: An Introduction“, in *Migration and Home: IMISCOE Short Reader*, M. Fathi und C. Ní Laoire, Hrsg., in IMISCOE Research Series. , Cham: Springer International Publishing, 2024, S. 1–20. doi: 10.1007/978-3-031-51315-2_1.

[19] M. Fathi, „Constructing Home through Unhome: Narratives of Resistance by an Iranian Asylum Seeker in Germany“, *Social Sciences*, Bd. 12, Nr. 1, Art. Nr. 1, Jan. 2023, doi: 10.3390/socsci12010016.

[20] M.-T. Haj Ahmad, *Von Ein- und Ausschlüssen in Europa*. Verlag Westfälisches Dampfboot, 2022. doi: 10.56715/398634130.

[21] P. Meth, „Rethinking the 'domus' in domestic violence: homelessness, space and domestic violence in South Africa", *Geoforum*, Bd. 34, Nr. 3, S. 317–327, Aug. 2003, doi: 10.1016/S0016-7185(03)00005-8.

[22] D. Porteous und S. E. Smith, *Domicide: The Global Destruction Of Home*. McGill-Queen's Press - MQUP, 2001.

[23] J. L. Samanani Farhan, „Introduction: Ethnography, dwelling and home-making", in *Home*, Routledge, 2019.

[24] A. Handel, „What's in a home? Toward a critical theory of housing/dwelling", *Environment and Planning C: Politics and Space*, Bd. 37, Nr. 6, S. 1045–1062, Sep. 2019, doi: 10.1177/2399654418819104.

[25] S. Maalsen, „Revising the smart home as assemblage", *Housing Studies*, Bd. 35, Nr. 9, S. 1534–1549, Okt. 2020, doi: 10.1080/02673037.2019.1655531.

[26] P. O'Connor, *Home: The foundations of belonging*. Routledge, 2017.

[27] M. Baethge, „Arbeit, Vergesellschaftung, Identität — Zur zunehmenden normativen Subjektivierung der Arbeit", *Soziale Welt*, Bd. 42, Nr. 1, S. 6–19, 1991.

[28] S. Nies und D. Sauer, „Was wird aus der Betriebsfallstudie? Forschungsstrategische Herausforderungen durch Entgrenzung von Arbeit und Betrieb", *AIS-Studien*, Bd. 3, Nr. 1, S. 14–23, 2010, doi: https://doi.org/10.21241/ssoar.64745.

[29] S. Pfeiffer und Autor*innenkollektiv, „Arbeit und Qualifizierung 2030 – Highlights", FAU Erlangen-Nürnberg, Nürnberg, 2023. [Online]. Verfügbar unter: https://www.labouratory.de/files/downloads/AQ2030-Studie-Highlights-DE.pdf

[30] R. Ø. Skotnes, „Standardization of cybersecurity for critical infrastructures. The role of sensemaking and translation", in *Standardization and Risk Governance A Multi-Disciplinary Approach*, O. E. Olsen, K. Juhl, P. H. Lindøe, und O. A. Engen, Hrsg., London: Routledge, S. 166–180.

[31] G. G. Voß und H. J. Pongratz, *Der Arbeitskraftunternehmer: eine neue Grundform der Ware Arbeitskraft?* in 50, no. 1. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 1998, S. 131.

[32] N. Kratzer und D. Sauer, „Flexibilisierung und Subjektivierung von Arbeit", in *Berichterstattung zur sozioökonomischen Entwicklung in Deutschland*, Soziologisches Forschungsinstitut (SOFI), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Institut für sozialwissenschaftliche Forschung (ISF), und Internationales Institut für empirische Sozialökonomie (INIFES), Hrsg., Wiesbaden: VS Verlag für Sozialwissenschaften, 2005, S. 125–149. doi: 10.1007/978-3-322-80600-0_6.

[33] N. Kratzer, *Arbeitskraft in Entgrenzung: grenzenlose Anforderungen, erweiterte Spielräume, begrenzte Ressourcen*. in Forschung aus der Hans-Böckler-Stiftung, no. 48. Berlin: Ed. Sigma, 2003. [Online]. Verfügbar unter: http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=65929692*

[34] F. Böhle und A. Bolte, *Die Entdeckung des Informellen: der schwierige Umgang mit Kooperation im Arbeitsalltag*. in Veröffentlichungen aus dem Institut für Sozialwissenschaftliche Forschung e.V., ISF München. Frankfurt/Main: Campus-Verl, 2002.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Lieferantenmanagement – ein Blick in die Praxis und in die Zukunft

Lea Calmano[1], Kathrin Schauberger, Michaela Sommer

Kurzfassung:
In einer Zeit, in der Cyber-Crime-as-a-Service angeboten wird, ist es für kleine und mittlere Unternehmen (KMU) zwingend erforderlich, das Thema Informationssicherheit nicht nur im eigenen Unternehmen zu leben, sondern eine Cyber-Resilienz zu entwickeln. Dies bedeutet über die Unternehmensgrenzen hinweg die eigenen Lieferanten in die Informationssicherheit mit einzubinden. Dienste werden zunehmend ausgelagert oder extern verwaltet, wodurch die Bedeutung der Lieferanten für eine sichere Unternehmensinfrastruktur zunehmend wächst. Die eigenen Unternehmenswerte sind nur sicher, wenn sie von allen beteiligten Akteuren, wie z.B. Lieferanten, geschützt werden. Um dies sicherstellen zu können, bedarf es eines qualifizierten Lieferantenmanagements. Wir möchten einen Ansatz für ein Lieferantenmanagement vorstellen, das auch von KMU angewendet werden kann.

Stichworte: Kleine und mittlere Unternehmen, KMU, Lieferantenmanagement, Lieferantenqualifizierung, Supply-Management, Vendor-Management

1 Einleitung in die Lieferantenbewertung

Lieferanten[2] spielen für die Wertschöpfung eines Unternehmens eine wichtige Rolle. Die genaue Bedeutung des jeweiligen Lieferanten für das Unternehmen hängt von verschiedenen Faktoren ab und sollte anhand entsprechender Kriterien bewertet werden, um die Sicherheit des Unternehmens vor z.B. Cyber-Attacken zu gewährleisten.

Durch die komplexen Abhängigkeiten in der Lieferkette kann es durch einen Sicherheitsvorfall beim Lieferanten zu weitreichenden Auswirkungen und negativen Folgen im eigenen Unternehmen kommen. Durch Betriebsunterbrechungen beim Lieferanten kann es zu Engpässen in der eigenen Produktions- und Lieferfähigkeit kommen. Sofern der Lieferant systemseitig an die eigenen Systeme und Prozesse des Unternehmens angeschlossen ist, kann ein Cyberangriff dazu führen, dass das unternehmenseigene Netzwerk ebenfalls von den Auswirkungen des Angriffs betroffen ist oder durch die Schnittstelle selbst angegriffen wird. Das Thema Lieferantenmanagement gewinnt zunehmend an Bedeutung in der Informationssicherheit. Das erkennt man

---

[1] Konica Minolta, Stuttgart

[2] Unter dem Begriff Lieferanten werden in diesem Text auch Dienstleister verstanden. Um die Lesbarkeit zu vereinfachen, werden sie unter dem Sammelbegriff „Lieferanten“ zusammengefasst.

daran, dass es in den meisten Standards und Gesetzen mittlerweile tief verankert ist. Beispiele dafür sind ISO 27001 und branchenspezifischen Standards. Auch in der neuen NIS-2-Richtlinie werden die Anforderungen an das Lieferantenmanagement von Unternehmen verschärft.

2 Best-Practice Methode

Zum Aufbau eines qualifizierten Lieferantenmanagements wird zunächst eine Unterteilung der Lieferanten in unterschiedliche Risikogruppen vorgenommen.

Die Bewertung der Lieferanten in den einzelnen Risikogruppen kann durch unterschiedliche Bewertungsmethoden erfolgen. Diese können individuell kombiniert werden.

Unsere Best-Practice Methode kann genutzt werden, um den neuen Anforderungen der NIS-2-Richtlinie oder den Anforderungen der ISO 27001 zum Lieferantenmanagement zu entsprechen.

Um einen risikobasierten Ansatz der Lieferantenbewertung vorzunehmen, werden die Lieferanten in verschiedene Risikogruppen unterteilt. Voraussetzung für diese Unterteilung ist es, dass im Unternehmen bereits eine Business Impact Analyse durchgeführt wurde bzw. die schützenswerten Informationen bekannt und dokumentiert sind. Basierend auf den schützenswerten Informationen des Unternehmens, kann der Zugriff auf diese durch Lieferanten oder die Verfügbarkeit der Systeme, die vom Dienstleister bereitgestellt werden, als Kriterium für die Einteilung in die verschiedenen Risikogruppen genutzt werden.

In unserem Best Practice Ansatz unterscheiden wir die folgenden fünf Risikogruppen:

2.1 Risikogruppe 1

Lieferanten und Dienstleister der Risikogruppe 1 haben lediglich Zugriff auf öffentliche Daten des Unternehmens und verarbeiten demnach keine besonders schützenswerten Informationen aus Sicht der Informationssicherheit. Da das Risiko eines Angriffs oder einer Schwachstelle bei diesen Lieferanten für die Informationssicherheit des Unternehmens sehr gering ist, wird diese Risikogruppe im weiteren Verfahren zunächst nicht weiter betrachtet. Beispiele für diese Lieferantengruppe sind: Dienstleister für Bürozubehör, Handelswaren, Lebensmittel, etc.

2.2 Risikogruppe 2

Lieferanten und Dienstleister der Risikogruppe 2 haben Zugriff und/oder verarbeiten schützenswerte Informationen der Vertraulichkeitsstufe intern. Beispiele für solche Dienstleister sind Unternehmen, die bei internen Unternehmensabläufen unterstützen.

2.3 Risikogruppe 3

Lieferanten und Dienstleister der Risikogruppe 3 haben Zugriff und/oder verarbeiten schützenswerte Informationen der Vertraulichkeitsstufe vertraulich. Beispiele für solche Dienstleister sind Beratungsunternehmen, die in strategischen Projekten unterstützen.

2.4 Risikogruppe 4

Lieferanten und Dienstleister der Risikogruppe 4 haben Zugriff und/oder verarbeiten schützenswerte Informationen der Vertraulichkeitsstufe streng vertraulich, oder stellen Systeme bereit, die für das Unternehmen eine hohe Verfügbarkeit aufweisen müssen. Beispiele für solche Dienstleister sind Beratungsunternehmen, die in strategischen, streng vertraulichen Projekten, insbesondere in der Geschäftsführung unterstützen. Außerdem sind Lieferanten betroffen, die Systeme bereitstellen, deren Verfügbarkeit für das Unternehmen geschäftskritisch ist.

2.5 Risikogruppe 5

Lieferanten und Dienstleister der Risikogruppe 5 haben administrativen Zugriff auf die IT-Infrastruktur des Unternehmens. Beispiele für solche Dienstleister können IT-Dienstleister sein, die Firewalls betreuen, Netzwerkinfrastrukturen verwalten oder einzelne IT-Anwendungen betreuen und/oder warten und für diese Tätigkeit administrative Zugriffsrechte besitzen.

Bei Bedarf können die Risikogruppen um weitere Parameter erweitert oder in der Anzahl an die Gegebenheiten des eigenen Unternehmens angepasst werden.

3 Bewertungsmethoden zur Beurteilung der Risikogruppen

Die Beurteilung der Risikogruppen für Lieferanten kann anhand von Kriterien aus Informationssicherheit, Datenschutz und der Wichtigkeit des Lieferanten für das Unternehmen erfolgen.

Basierend auf den Risikogruppen können unterschiedliche Bewertungsmethoden für die Lieferanten vorgenommen werden. Diese können individuell kombiniert werden. Als erstes sollten die Lieferanten der höchsten Risikogruppe bewertet werden, da sie das höchste Risiko für das Unternehmen in Bezug auf die Informationssicherheit darstellen.

Für Lieferanten der höchsten Risikogruppe könnten die folgenden Kriterien definiert werden:

3.1 Internes Assessment von bereits vorhandenen Lieferanten-Informationen mittels Risikobewertung

Das interne Assessment erfolgt über die Sammlung und Auswertung von bereits vorliegenden Lieferanten-Informationen mittels einer internen standardisierten Vorgehensweise, die mit einer internen Cyber-Risikobewertung einhergeht. Zu betrachten sind mögliche Bedrohungen oder das Vorhandensein von Schwachstellen, die von dem Lieferanten ausgehen könnten. Grundlage dazu kann eine Checkliste der Lieferantenrisiken sein, die bestimmte Szenarien zugrunde legt, die intern für den Lieferanten bewertet werden.

Bei nicht ausreichenden Informationen über den Lieferanten sollten noch weitere Bewertungsmethoden genutzt werden.

3.2 Lieferantenselbstauskunft und Lieferantenfragebogen mit Risikobewertung

Der Lieferant füllt ein Informationsblatt zur Selbstauskunft aus. Dies beinhaltet die Informationssicherheitsleitlinie zusammen mit den etablierten Rahmenbedingungen u.a. Einhalten von Gesetzen (z.B. EU-NIS-2-Richtlinie oder EU-DSGVO) oder Anwendung von bestehenden allgemeinen oder branchenspezifischen Standards und Richtlinien (z.B. ISO 27001, VdS 10000). Je nach Risikogruppe des Lieferanten werden noch weitere explizite Fragen zur Umsetzung von Informationssicherheitsanforderungen und speziellen Maßnahmen zum Schutz von vertraulichen und sensiblen Informationen mittels Fragebogen an den Lieferanten gestellt. Die Antwortmöglichkeiten sind bereits vorgegeben und müssen von dem Lieferanten nur noch passend ausgewählt werden. Dabei werden den unterschiedlichen Antworten Punkte zugewiesen. Diese Punkte werden für die Auswertung der Lieferantenselbstauskunft genutzt und ergeben ein Gesamtbild der Cyber-Resilienz des Lieferanten. Die Anforderungen an Lieferanten einer hohen Risikogruppe sind entsprechend höher als an Lieferanten einer geringeren Risikogruppe. Je nach interner Risikobewertung der Selbstauskunftsinformationen sind weitere Prüfungen des Lieferanten erforderlich, die in einem persönlichen Gespräch vorgenommen werden können.

Beispiele für mögliche Fragen und Antwortmöglichkeiten sind:

- Haben Sie Zertifizierungen im Bereich der Informationssicherheit?
  1. Keine gültigen Zertifikate vorhanden.

2. Keine gültigen Zertifikate vorhanden, aber notwendige Richtlinien und Prozesse, die eine Informationssicherheit im Unternehmen etablieren.
3. Keine gültigen Zertifikate vorhanden, aber ein etabliertes ISMS nach einem bekannten Standard.
4. Ein Zertifikat ist vorhanden. (Bitte geben Sie ihr Zertifikat an).
5. Mehrere Zertifikate sind vorhanden. (Bitte geben Sie ihre Zertifikate an).

- Schulen Sie Ihre Mitarbeiter regelmäßig zum Thema Informationssicherheit?
  1. Es finden keine Schulungen der Mitarbeiter zur Informationssicherheit statt.
  2. Mitarbeiter werden unregelmäßig über Informationssicherheitsthemen informiert.
  3. Es gibt Informationssicherheitsschulungen, die von allen Mitarbeitern mindestens einmal jährlich absolviert und dokumentiert werden müssen.
  4. Es gibt Informationssicherheitsschulungen einschließlich zielgruppenorientierter Schulungen, die von allen Mitarbeitern mindestens einmal jährlich absolviert und dokumentiert werden müssen.
  5. Es gibt Informationssicherheitsschulungen, einschließlich zielgruppenspezifischer Schulungen, die von allen Mitarbeitern mindestens einmal jährlich absolviert und dokumentiert werden müssen. Es finden zusätzliche Sicherheitstrainings statt (z.B. zum Thema Phishing).

3.3 Lieferantenaudits

Mit einem Lieferantenaudit können Lieferanten individuell geprüft werden. Lieferantenaudits können zu einer besseren Zuverlässigkeit der Informationssicherheit in der Lieferkette beitragen, fördern ein vertrauensvolles Verhältnis zwischen Lieferanten und Unternehmen, verbessern das Verständnis der Prozesse der Lieferanten, und minimieren das Auftreten von Risiken in der Lieferkette.

Audits bilden eine wichtige Maßnahme zur Qualitätssicherung und geben Einblicke in die Umsetzung der Informationssicherheit von Unternehmen. Insbesondere bei risikobehafteten Lieferanten, die auffällige Sicherheitsvorfälle hatten oder nicht allen Mindestanforderungen der Informationssicherheit entsprechen, sind regelmäßige Audits angezeigt. Ein Vor-Ort-Audit dient

der Prüfung von bestehenden Prozessen, Aufzeichnungen sowie der Einhaltung der im Abschnitt 4 genannten vertraglich festgelegten Anforderungen.

Der persönliche Austausch im Rahmen eines Audits kann insbesondere bei kritischen Abweichungen die Zusammenarbeit zwischen Auftraggeber und Lieferant durch einen gezielten kommunikativen Austausch fördern. Anschließend ist die Umsetzung von korrigierenden und präventiven Maßnahmen zu überwachen.

3.4 Digitale Gesamtlösung für externe Lieferantenbewertung

Für die Bewertung der Lieferanten in der höchsten Risikogruppe können auch externe Lieferantenbewertungen durchgeführt werden. Dabei wird von externen Dienstleistern eine Bewertung der angemeldeten Lieferanten vorgenommen.

Die Methoden reichen von einer Analyse und Bewertung des Webauftritts des Lieferanten, über standardisierte Assessments der Lieferanten zu einem Risiko Rating Score, der dem Unternehmen eine Transparenz über die gelebte Informationssicherheit beim Lieferanten verschafft. Dies hat den Vorteil, dass der Lieferant einer standardisierten Kontrolle durch eine unabhängige Instanz unterzogen wird und die Bewertung des Lieferanten in einer transparenten Form erfolgt, die mit anderen Unternehmen verglichen werden kann. Bei festgestellten Risiken erhält der Lieferant eine Auswertung der Schwachstellen und Hinweise zur Umsetzung von Maßnahmen. Damit wird die Widerstandsfähigkeit des auftraggebenden Unternehmens ebenfalls gestärkt.

Ein Beispiel für eine externe Lieferantenbewertung ist das Cyber-Risk Rating[3] durch ein Unternehmen des KSV1870. Der Gläubigerschutzverband hat gemeinsam mit der Initiative „Kompetenzzentrum Sicheres Österreich (KSÖ)“ einen Standard für das Lieferantenmanagement von Unternehmen in Österreich definiert, der den Vorgaben der NIS und der NIS-2-Richtlinie entspricht.

Eine Risikobewertung eines Lieferanten kann dort so aussehen:

[3] https://cyberrisk-rating.at/

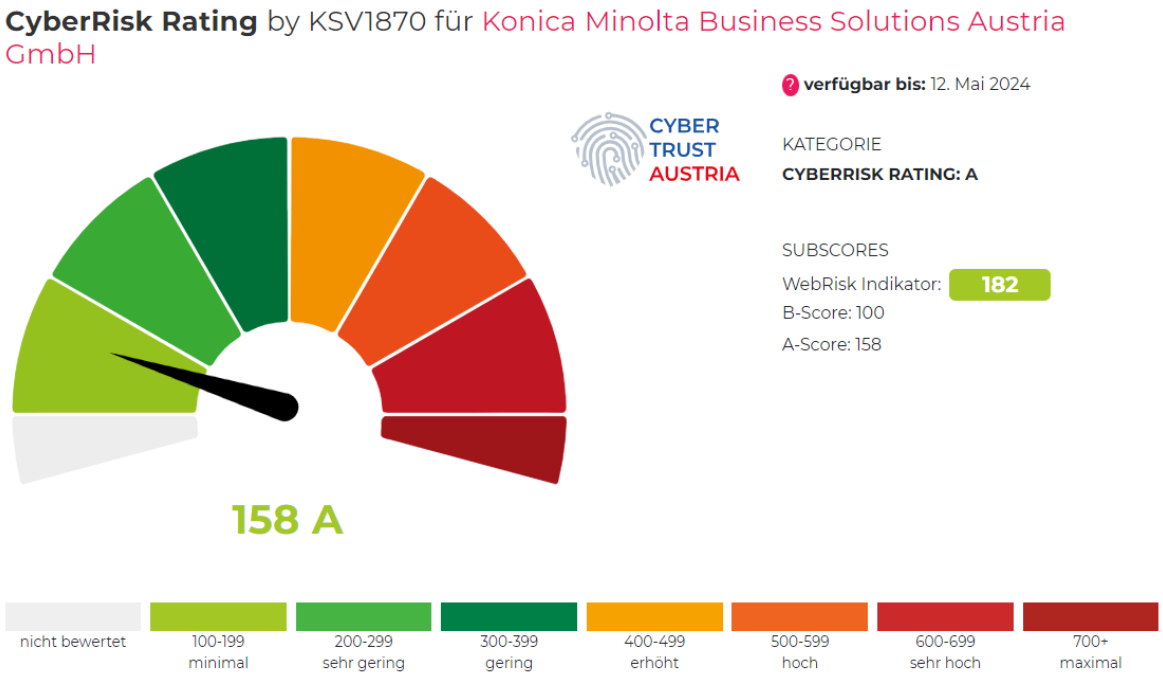


Abbildung 1: Beispiel Cyber-Risk Rating vom KSV1870[4]

In diesem Fall wurde das Cyber-Risk Rating der Konica Minolta Business Solutions Austria erstellt. Dies ist die Ansicht, die andere Unternehmen ebenfalls auf der Website öffentlich einsehen können. Für Unternehmen, die dort bereits gelistet sind, können die Informationen direkt bezogen werden. Ziel des KSÖ und des KSV1870 ist es, eine flächendeckende Abbildung der Unternehmen in Österreich zu erreichen und somit die Cyber-Resilienz der österreichischen Unternehmen zu stärken. Je mehr Dienstleister eine standardisierte und jedem Unternehmen zur Verfügung stehende Risikobewertung vornehmen, desto geringer wird der Aufwand der Lieferantenbewertungen, die sonst jedes Unternehmen neu erstellen müsste.

4 Einbindung des Qualitätsmanagementsystems in die Lieferantenbewertung

Wenn das Unternehmen ein Qualitätsmanagement etabliert hat, sollte dieses in den Prozess der Lieferantenbewertung, Lieferantenauswahl und Lieferantenüberwachung involviert werden, um das Qualifizierungs- und Managementprogramm von Lieferanten, bestehend aus der Evaluierung, Qualifizierung, Genehmigung und der laufenden Überwachung der Lieferantenqualität, als gelenkten und dokumentierten Prozess zu berücksichtigen.

[4] KSV1870, 2023 [1]

In der Evaluierung gilt es, zusammen mit der Fachabteilung, dem Einkauf, der IT und der Qualitätsmanagementeinheit, die richtigen Anforderungen für potenzielle Lieferanten bereitzustellen. In Bezug auf die IT-Sicherheit sind bei der Erstellung der Lieferantenanforderungen Aspekte zu den aktuellen „Best Practices“, wie bereits unter Abschnitt 2 beschrieben, zu prüfen.

Nachstehend sind mögliche Vorgaben zur Informationssicherheit genannt, die in Lieferantenverträgen und Qualitätsmanagementvereinbarungen enthalten sein sollten:

<table>
<tr><th>Lieferantenverträge</th><th>Qualitätsmanagementvereinbarungen</th></tr>
<tr><td>Verfügbarkeit der Systeme</td><td rowspan="3">Verknüpfung von branchenspezifischen Qualitätsstandards mit IT-Sicherheitsanforderungen an Daten und die Einbettung von IT-Sicherheitsprinzipien in das Qualitätsmanagement</td></tr>
<tr><td>Erreichbarkeit der Mitarbeiter und Unterstützung bei Notfällen</td></tr>
<tr><td>Datensicherheit zur Sicherung von vertraulichen Daten</td></tr>
<tr><td>Umgang mit Sicherheitsvorfällen und deren Meldungen</td><td>Festlegung von Leistungskennzahlen, sog. Key Performance Indicators (KPIs), die IT-Sicherheitsaspekte erfassen</td></tr>
<tr><td>Zugriffskontrolle und Management von physischer Sicherheit sowie Schutzmaßnahmen zur Netzwerksicherheit</td><td>Nachweis von Dokumentationen, die die IT-Sicherheitshandhabung belegen</td></tr>
<tr><td>Umgang mit Informationssicherheitsrisiken sowie Business Continuity Management</td><td>Risikobasierte Ansätze, um Auswirkungen von IT- Sicherheitsproblemen über die Fachabteilung hinaus sichtbar zu machen</td></tr>
<tr><td>Management von Sicherheitspatches</td><td>Standardverfahren zum Änderungsmanagement, die Qualitäts- und Sicherheitsänderungen beinhalten und den kontinuierlichen Verbesserungsprozess erfüllen</td></tr>
<tr><td>Compliance mit Sicherheitsstandards und die Möglichkeit zur Überwachung durch Kunden und externe Prüfer</td><td>Einbindung von branchenspezifischen Normen und IT-Sicherheitsnormen in die Anforderungen zur Lieferantenüberwachung</td></tr>
</table>

Eine regelmäßige Lieferantenüberwachung in Form einer wiederholenden Requalifizierung der Lieferanten sollte das Unternehmen in einem Standardprozess vorweisen können, um sicherzustellen, dass die Systeme, Prozesse und Dienstleistungen der Lieferanten für den vorgesehenen Zweck geeignet bleiben. Der Prüfungsrhythmus sollte auf die Risikoeinstufung und Kritikalität des Lieferanten abgestimmt werden.

5 Lieferantenmanagement

Die Anforderungen an die Informationssicherheit sollten bei einer erfolgreichen Erstqualifizierung eines Lieferanten neben der internen Genehmigungsdokumentation, wie bereits im Abschnitt 4 beschrieben, vertraglich mit dem Lieferanten vereinbart sein, damit im laufenden Betrieb die vereinbarten Standards bestehen bleiben. Die getroffenen Vereinbarungen und Anforderungen mit den Lieferanten können vertraglich beispielsweise in den Technischen und Organisatorischen Maßnahmen (TOM), Vertraulichkeitsvereinbarungen, Qualitätssicherungsvereinbarungen, Datentransfervereinbarungen, oder anderen an den Vertrag geknüpften Vereinbarungen dokumentiert werden.

5.1 Erweiterte Maßnahmen zur Sicherstellung der Informationssicherheit

Wenn einer der Lieferanten die vorab geforderten Anforderungen der Informationssicherheit nicht vollumfänglich erfüllt, die Zusammenarbeit jedoch weiterhin bestehen soll, gibt es die Möglichkeit gemeinsam mit dem Lieferanten alternative Maßnahmen zur Behandlung der entstehenden Risiken zu definieren. Dies können technische, organisatorische und das Verhalten der Mitarbeiter betreffende Maßnahmen sein.

5.2 Zusammenfassung des Prozesses der Lieferantenbewertung

Bei dem in den Abschnitten 2 bis 4 beschriebenen Prozess werden die Lieferanten zunächst in eine geeignete Risikogruppe eingeteilt. Die Einteilung basiert auf der Sensibilität der Informationen, auf die der Lieferant Zugriff erhält und der Verfügbarkeit der Systeme, die er bereitstellt. Je nach Risikogruppe kann ein passendes Verfahren zur Lieferantenbewertung genutzt werden.

Für die Risikogruppe 2 bietet sich ein internes Assessment an, bei dem überprüft wird, ob es Risikobereiche bei dem Lieferanten gibt, die für die Informationssicherheit des eigenen Unternehmens wichtig sind. Sollte dies der Fall sein, müssen diese mit dem Lieferanten besprochen und eine Risikobehandlung definiert werden.

Für die Risikogruppen 3 und 4 kann die Lieferantenselbstauskunft oder eine externe Lieferantenbewertung durchgeführt werden. Bei gefundenen Schwachstellen, hohen Risiken oder Abweichungen von den Vorgaben, kann zusätzlich ein Lieferantenaudit durchgeführt werden.

Für Lieferanten der Risikogruppe 5 sollten externe Lieferantenbewertungen oder eine Lieferantenselbstauskunft durchgeführt werden. Dabei sollte besonders auf die Mindestanforderungen Wert gelegt werden. Hier kann ein Lieferantenaudit vor Ort hilfreich sein, um die Umsetzung der Informationssicherheit des Lieferanten an Beispielen zu überprüfen. Wenn der Lieferant bisher keine Zertifizierung in der Informationssicherheit nachweisen kann, sind eventuell zusätzliche Anforderungen, spezifisch angepasst auf den Lieferanten und die jeweilige Dienstleistung, ratsam.

Nach der Bewertung der Lieferanten sollte überprüft werden, ob die Anforderungen der Informationssicherheit weiterhin erfüllt sind. Gegebenenfalls sollten erweiterte Maßnahmen vorgenommen werden.

Anschließend sollten die Verträge mit den Lieferanten überprüft und alle Anforderungen dokumentiert und ergänzt werden, um diese schriftlich zwischen beiden Parteien festzuhalten und die Einhaltung zu gewährleisten.

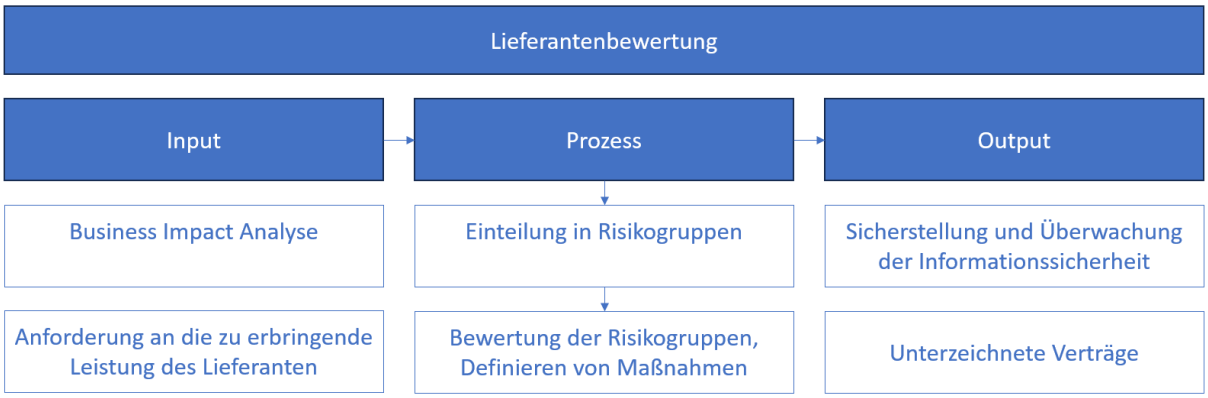


Abbildung 2: Prozess der Lieferantenbewertung[5]

6 Ausblick in die Zukunft des Lieferantenmanagements

Ein Beispiel für ein bereits etabliertes und gelebtes Lieferantenmanagement bietet die Automobilindustrie. Der Verband der Automobilindustrie hat einen eigenen Branchenstandard geschaffen und damit das Thema Informationssicherheit in der Lieferkette selbst in die Hand genommen. Mit dem Trusted Information Security Assessment Exchange (TISAX), einem Prüf- und

[5] Eigene Darstellung

Austauschmechanismus für die Informationssicherheit, beweisen sich Unternehmen untereinander die Einhaltung der von der Automobilindustrie vorgegebenen Mindestanforderungen der Informationssicherheit. Jeder Lieferant der Automobilindustrie muss sich nach dem VDA ISA Katalog zertifizieren lassen. Mit dem TISAX Zertifikat kann ein Lieferant jeden Automobilhersteller und dessen Lieferkette beliefern. Dies fing mit den direkten Zulieferern an und geht in der Lieferkette mittlerweile immer weiter nach außen. Auch IT-Dienstleister und Berater müssen mittlerweile ein TISAX Zertifikat nachweisen.[6]

Das Beispiel der Automobilindustrie zeigt, wo es für Unternehmen in Deutschland und Europa hingehen kann. Mit der NIS-2-Richtlinie, die für die Sicherheit von Netz- und Informationssystemen steht und Unternehmen zu strengeren Informationssicherheitsmaßnahmen verpflichtet, wird das Thema Informationssicherheit gesetzlich verankert. Sie wird europaweit umgesetzt und soll für einheitliche Anforderungen an die Informationssicherheit in Unternehmen sorgen und damit die Cyber-Resilienz von Unternehmen in ganz Europa stärken. Die NIS-2-Richtlinie fordert in ihren Mindestanforderungen eine Sicherheit der Lieferkette. Dem Thema Lieferantenmanagement kommt damit eine noch größere Bedeutung zu und es wird gesetzlich verpflichtend. Selbst Unternehmen, die zunächst nicht unter die Schwellenwerte der NIS-2-Richtlinie fallen, müssen, wenn sie ein Unternehmen beliefern, das zur Einhaltung der NIS-2-Richtlinie gesetzlich verpflichtet ist, die Anforderungen der entsprechenden, von der NIS-2-Richtlinie betroffenen Kunden, an die Informationssicherheit im Unternehmen erfüllen und dies nachweisen.

Diese neuen Anforderungen sind insbesondere für kleine und mittlere Unternehmen (KMU) herausfordernd, da die wenigsten von ihnen derzeit eine Zertifizierung im Bereich der Informationssicherheit nachweisen können und in KMUs teilweise ein Leitfaden für die Anforderungen an die Informationssicherheit fehlt. Um genau diese Zielgruppe besser unterstützen zu können, hat das BSI gemeinsam mit dem Bundesverband mittelständischer Wirtschaft ein Konsortium gegründet und die DIN-SPEC 27076, besser bekannt unter Cyber-Risiko-Check, erstellt. Dieser umfasst 27 Anforderungen aus sechs Themenbereichen der Informationssicherheit und ermöglicht KMU eine Bewertung und Einordnung der eigenen Informationssicherheit.[7]

[6] ENX Association, 2024 [2]
[7] BSI, 2024 [3]

Eine weitere Möglichkeit eines Nachweises der Informationssicherheit für KMU bietet beispielsweise die VdS Schadenverhütung, eine GmbH des Gesamtverbandes der Deutschen Versicherungswirtschaft, mit den VdS-Richtlinien 10000 – Informationssicherheitsmanagementsystem für kleine und mittlere Unternehmen (KMU). Dabei handelt es sich um einen Maßnahmenkatalog, der explizit auf KMU angepasst ist und den Informationssicherheitsstatus eines Unternehmens prüfen und sicherstellen soll.[8]

Der Cyber Trust Europe bietet Unternehmen ein Label für Informationssicherheit basierend auf dem zuvor in Abschnitt 3.4 vorgestellten Cyber Risk Rating, das die Umsetzung der Anforderungen der NIS-2-Richtlinie überprüft.[9]

[8] VdS, 2024 [4]

[9] Cyber Trust Europe, 2024 [5]

Literaturhinweise

[1] KSV1870. Das CyberRisk Rating by KSV1870. Linz: KSV1870. https://cr.nimbusec.com/yourratings (abgerufen am 02.03.2024).

[2] ENX Association. ENX Portal. https://portal.enx.com/de-DE/TISAX/ (abgerufen am 26. 02 2024).

[3] BSI. CyberRisikoCheck Wirkungsvoller Schutz für kleine und Kleinstunternehmen nach DIN SPEC 27076. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/KMU/CyberRisikoCheck/CyberRisikoCheck_node.html (abgerufen am 26. 02 2024).

[4] VdS. Informationssicherheits-Management mit Zertifikat. https://vds.de/kompetenzen/cyber-security/zertifizierung/vds-10000-informationssicherheit-fuer-kmu (abgerufen am 26. 02 2024).

[5] Cyber Trust Europe. Cyber Trust Europe - the European Quality Label for Cybersecurity. https://www.cyber-trust-europe.eu/en/ (abgerufen am 26. 02 2024).

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Bedrohnung von oben: Analyse des Gefahrenpotenzials von frei erhältlichen Drohnen für die Informationssicherheit von Unternehmen

Ben Lutz [1]
Dr. Konrad Meier[2], Prof. Dr. Richard Zahoransky[3]

Kurzfassung:
In diesem Beitrag wird die Bedrohungslage durch den nicht destruktiven Einsatz frei erhältlicher Drohnen zur Beeinträchtigung der Informationssicherheit von Unternehmen untersucht. Unter Verwendung einer Kombination aus Literaturanalyse, Experteninterviews und Fallstudie wird die Vielfalt der Bedrohungen und die Machbarkeit von drohnenunterstützten Angriffen auf die Informationssicherheit detailliert beleuchtet. Dabei zeigt sich, dass Drohnen eine reale und vielschichtige Gefahr für die Informationssicherheit darstellen können. Die Einschätzung der Eintrittswahrscheinlichkeit solcher Angriffe ist aufgrund der dynamischen Entwicklung der Drohnentechnologie und der Vielfalt der Anwendungsmöglichkeiten komplex. Der Beitrag betont die Notwendigkeit einheitlicher Vorgaben für effektive Abwehrstrategien und weist auf die Wichtigkeit weiterer Forschung in diesem Bereich hin, um den Herausforderungen der technologischen Entwicklungen und der sich ständig ändernden Bedrohungslandschaft gerecht zu werden.

Stichworte: Drohne, Drohnenabwehr, Experteninterview

1 Motivation

Wie in vielen Bereichen schreitet der technische Fortschritt auch im Bereich der Drohnen voran. Drohnen werden immer erschwinglicher [1] und die verbaute Kameratechnik entwickelt sich fort [2]. Dieser Fortschritt bringt Vorteile für Unternehmen, die Drohnen zu vielseitigen Zwecken wie der Bewirtschaftung von Agrarflächen [3], der Inspektion technischer Anlagen, dem Verbringen von Waren zu logistischen Zwecken oder künftig auch im Bereich der Telekommunikation [4] einsetzen können. Auch für Privatpersonen bringt der Fortschritt in dieser Branche einen Vorteil. Drohnen können zu vielen freizeitlichen Zwecken eingesetzt werden. So erfreuen sich unter anderem Hobbyfotografen an den Vorzügen beim Erstellen qualitativ hochwertiger Aufnahmen[5]. Auch der Einsatz als Spielzeug oder das Fliegen von Drohnenrennen sind beliebte Einsatzgebiete [6]. Auch in Zukunft wird uns

[1] Hochschule Furtwangen, Furtwangen
[2] BadenIT GmbH, Freiburg
[3] Hochschule Furtwangen, Furtwangen

der vermehrte Einsatz von Drohnen begleiten [7]. Doch nicht zuletzt der Ukraine-Konflikt zeigt [8]: Auch die maliziöse Seite der Gesellschaft profitiert vom technischen Fortschritt im Bereich der Drohnen [9]. Innerhalb von Deutschland lässt sich aktuell kein kriegerischer Einsatz von Drohnen beobachten. Trotzdem können die neuen Möglichkeiten von Drohnen auch in Deutschland eine Gefahr für Unternehmen darstellen. So berichtet der im Rahmen dieser Arbeit interviewte Experte für Werkschutz I1 folgendes.

„Ich würde, Stand jetzt, auch sagen, dass es aktuell in Deutschland sehr leicht ist, mit einer Drohne einen Angriff auf die IT-Sicherheit zu unterstützen." (I1, Pos. 9)

Ob und wie diese These in die Realität zu übertragen ist, wird im Rahmen dieses Beitrags untersucht. Um das aktuelle Bedrohungspotenzial einzuschätzen, wird Literatur analysiert, Experten interviewt und eine praktische Fallstudie durchgeführt. Zusätzlich werden Möglichkeiten zur Abwehr evaluiert, um ein ganzheitliches Bild der Gefahrenlage von Drohnen auf die Informationssicherheit zu erlangen. Abschließend werden die Implikationen der Forschungsergebnisse auf den Fachbereich diskutiert und die Frage beantwortet, ob frei erhältliche Drohnen eine Gefahr für die Informationssicherheit in Unternehmen darstellen.

Weitere Ausführungen zu dieser Forschung sowie zu den Ergebnissen finden sich in der diesem Beitrag zugrunde liegenden, vollständigen Arbeit ([10]).

2 Einleitung

Die Literatur bietet eine Reihe verschiedener Definitionen des Wortes „Drohne", beispielsweise: „*Aerial vehicles that do not carry a human operator, fly remotely or autonomously, and carry lethal or nonlethal payloads are considered as drones*" [11, p. 1] Bei der Betrachtung des Gefahrenpotenzials von Drohnen ist es wichtig zu verstehen, dass Drohnen mobile Plattformen sind. Ähnlich wie bei anderen mobilen Plattformen, wie etwa dem PKW, bietet diese Mobilität eine Vielzahl verschiedener Einsatzmöglichkeiten. Definiert man einen PKW als ein motorisiertes Fahrzeug mit vier Rädern, so zählt sowohl ein Krankenwagen als auch ein militarisierter Pick-up einer Terrororganisation als PKW. Die Zuladung in Form von medizinischem Equipment zur Notfallversorgung kann Menschenleben retten und ist ein gesellschaftlich positiver Einsatzzweck. Bei einem Pick-up mit einer Zuladung in Form von 32 S5 Raketen kann, bei maliziöser Intention, von einem gesellschaftlich negativen Einsatz ausgegangen werden. Das Gefahrenpotenzial ist also weniger abhängig vom Vehikel selbst, sondern von der Zuladung sowie der

Intention des Steuernden. Nach [11] lassen sich Drohnen in sechs Überkategorien mit bis zu 21 Unterkategorien einteilen. Die Vielfältigkeit wird in Abbildung 1 durch einige nicht näher erläuterte Beispiele dargestellt. [12]

Abbildung 1: Beispiele verschiedener Drohnen-Bauformen

Im Rahmen dieses Beitrags sollen ausschließlich Drohnen betrachtet werden, welche für private Akteure erhältlich sind. Außerdem wird im Rahmen des Beitrags davon ausgegangen, dass Angreifer zwar maliziöse Zwecke verfolgen und sich nicht an gesetzliche Richtlinien halten, dabei jedoch moralischen Grundsätzen folgen. Dies schließt die Betrachtung von militärischen oder terroristisch motivierten Angriffen, wie dem Abwerfen von Sprengsätzen oder das absichtliche Steuern zur Induktion von Sach- oder Personenschäden aus.

Einige exemplarische technische Merkmale der untersuchten Drohnengruppe sind zum Verständnis in der folgenden Tabelle 1 aufgeführt. Mithilfe dieses Wissens wird nun im folgenden Abschnitt unter Zuhilfenahme vier verschiedener Methodiken das Gefahrenpotenzial von Drohnen auf die Informationssicherheit von Unternehmen untersucht.

Eigenschaft	Technische Möglichkeit	Literatur
Zuladung	3,5 kg, z.B. (multispektrale) Kameras, Sensoren wie LiDAR, Gas Sensorik, etc.	[12]
Reichweite	15 – 35 km, LTE möglich	[13–15]
Navigation	Automatisiert mittels GPS	[16]
Kameratechnik	Auflösung 5280×3956 Pixel, 48MP, 7x optischer Zoom, mit spezieller Technik bis zu 36x optischer Zoom	[5, 17]

Tabelle 1: Relevante technische Möglichkeiten untersuchter Drohnen

3 Literaturanalyse

In der Literaturanalyse sollen mögliche Angriffe identifiziert werden, welche vom Einsatz von Drohnen profitieren könnten. Statt der vollumfänglichen, systematischen Literaturanalyse sollen in diesem Beitrag besonders prominente Gefahren exemplarisch identifiziert werden, um einen Eindruck der Gefährdungslage zu erhalten. Die dazu betrachtete Literatur wurde anhand verschiedener Kriterien gefiltert. Diese sollen insbesondere die Aktualität, Relevanz und Realitätsnähe der Ergebnisse sicherstellen. Im Rahmen der Analyse lassen sich zum Zeitpunkt der Erstellung ca. 1.750 wissenschaftliche Arbeiten identifizieren. Davon wurden im Rahmen dieser Arbeit ca. 100 Arbeiten grundlegend gesichtet. Durch die Anwendung der Ausschlusskriterien

Kategorie	Exemplarische Angriffe	Literatur
Mitlesen von sensiblen Informationen	Erfassen von Eingaben mittels (spezialisierter) Kameratechnik	[18-22]
Mithören von sensiblen Informationen	Platzieren von Aufnahmegeräten; Lippenlesen; Laser-Doppler-Vibrometer	[23-25]
Abfangen von Funksignalen	Mitschneiden div. Funksignale von z.B. WLAN, DECT, Bluetooth oder ZigBee; „MouseJack“ Angriffe	[26-31]
Senden von Funksignalen	Man-in-the-Middle Angriffe auf Funkprotokolle; Evil-Twin Angriffe auf WLAN; „MouseJack“ Angriffe	[31-33]

Tabelle 2: Gefahrenkategorien aus der Literaturanalyse

flossen schlussendlich ca. 10 % in die Analyse ein. Hierbei wurden vier maßgebliche Kategorien identifiziert. Diese lassen sich grundlegend in die drei Überkategorien visuell, akustisch und Funk einteilen. Die folgende Tabelle 2 stellt prägnant die identifizierten Gefahren dar.

4 Experteninterviews

Mit den Erkenntnissen aus dem vorherigen Kapitel wurden Experten interviewt. Dies soll die Praxisrelevanz einordnen und Eindrücke aus der Realität in die Forschung einbringen. Bei der Auswahl der Experten wurde auf deren Erfahrung in den Fachbereichen Drohnentechnologie und Informationssicherheit geachtet. Die Aussagen von sechs Experten aus den Bereichen Pharmaindustrie, Forschung, Unternehmensberatung, Dienstleistung und einer staatlichen Einrichtung flossen in diesen Beitrag ein.

Jeder der Interviewpartner erhielt im Voraus einen Fragenkatalog als Ausgangspunkt für das semistrukturierte Interview. Dieser bestand aus einführenden Fragen über die aktuelle Lage, mögliche Gefahren und Abwehrmaßnahmen sowie einer Einschätzung der zukünftigen Lage.

Im Rahmen der im Nachgang durchgeführten, qualitativen Inhaltsanalyse konnten insbesondere zwei Kategorien von möglichen Drohnengefahren durch die Experten identifiziert werden.

Als erste Gefahrenkategorie wurden Angriffe mittels Kameratechnik identifiziert. Hierzu zählt das Ausspähen von Informationen mittels Kameratechnik. Laut einem Experten lassen sich auch mit niedrig auflösenden Kameras Informationen auf kurze Distanz ausspähen. Außerdem wurden Gefahren wie das Ausspähen von Unternehmensressourcen zur Vorbereitung auf spätere Angriffe, Industriespionage oder das Anbringen von Kameras an Fenstern als mögliche Szenarien genannt.

Weitere häufig genannte Angriffe fallen in die Kategorie Funktechnik. Experten beschreiben Angriffe auf WLANs, WPANs, Mobilfunk oder LoRaWAN. Neben einer Reihe von Denial of Service Angriffen berichtet ein Experte von einem Angriff auf IoT Geräte im Rahmen eines Proof of Concepts.

Das folgende Schaubild fasst die Menge aller im Rahmen der Literaturanalyse und der Experteninterviews identifizierten Gefahren kategorisiert zusammen und gibt exemplarisch einen groben Überblick über die aktuelle Gefahrenlage.

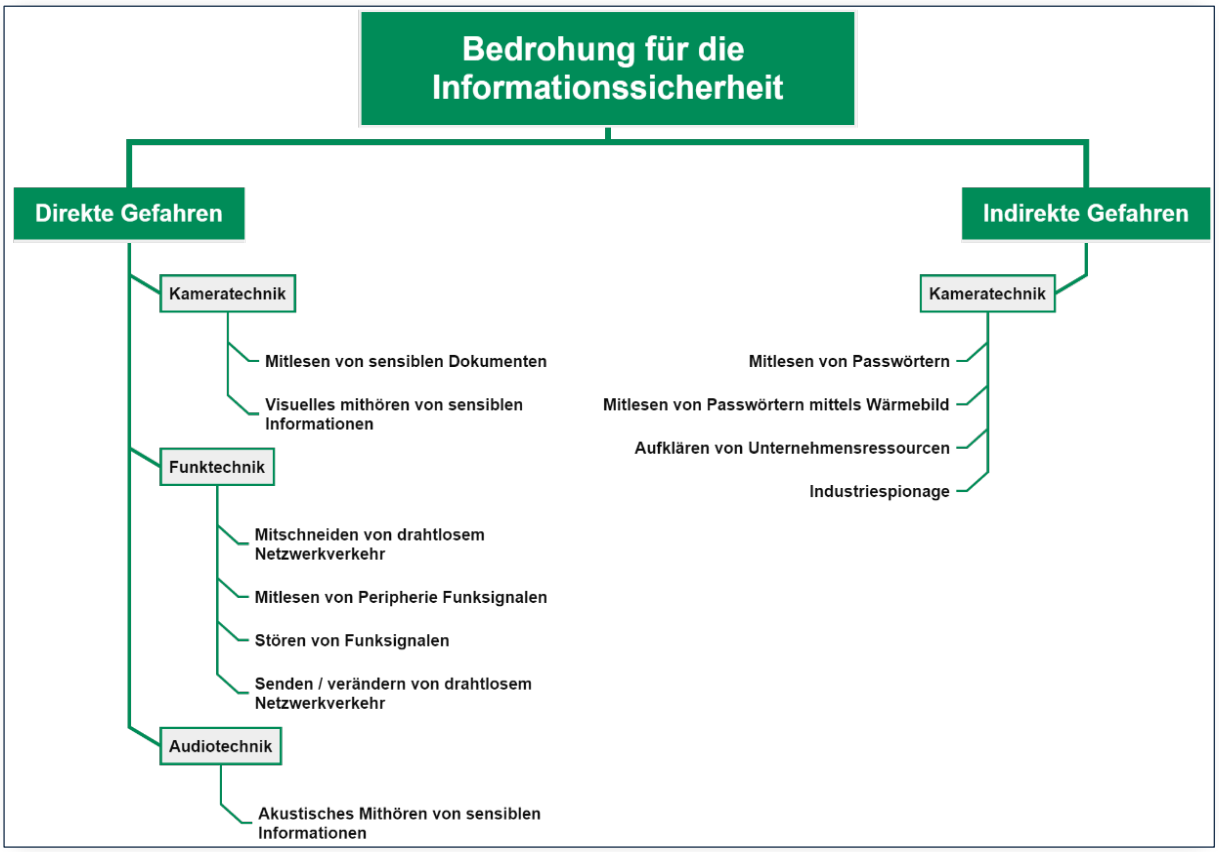


Abbildung 2: Kategorisierung verschiedener Bedrohungen durch Drohnen

Die weiteren Erkenntnisse der Interviews fließen zudem in die Bereiche Abwehrmaßnahmen sowie das abschließende Fazit dieses Beitrags ein.

5 Fallstudie

Um das Gefahrenpotenzial von Angriffen durch Drohnen auf die Informationssicherheit zu untersuchen, wurden in einer praxisorientierten Fallstudie drei relevante Kernfragen untersucht. Als Ausgangsszenario wurde das Ausspähen sensibler Informationen mit Hilfe herkömmlicher Kameratechnik ausgewählt. Hierbei soll im Rahmen der ersten Kernfrage die Machbarkeit bestätigt und das Ausmaß lesbarer Informationen erfasst werden. Die zweite Kernfrage untersucht das Entdeckungsrisiko der Drohne während des Angriffs durch Mitarbeitende. Hierzu wird die Reaktion von Probanden während einer kognitiven Aufgabe beobachtet. Die letzte Kernfrage soll die Effektivität einer, von den Experten genannten, möglichen Abwehrmaßnahme untersuchen. Hierzu wurde nach Anbringen einer Sichtschutzfolie erneut untersucht, inwiefern sensible Daten mithilfe einer Drohne erfassbar sind.

Die unten stehende Abbildung 3 zeigt die Örtlichkeit der Fallstudie so wie die verschiedenen Flugpfade der Drohne (P 1–3). Im Fenster (F) liegen exemplarisch sensible Informationen als Bildschirminhalt (B) und als gedrucktes Dokument (C) aus. Während des Versuchs werden zwei Drohnen der Firma DJI mit einer Auflösung von jeweils 12,35 [34] und 12 Megapixeln [35] eingesetzt.

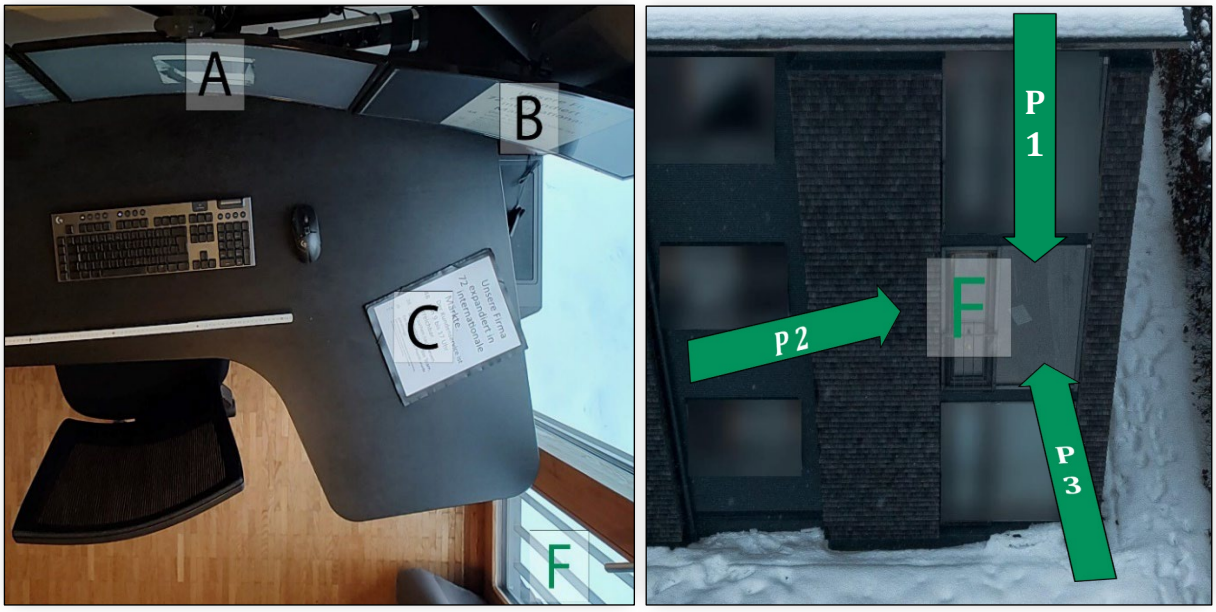


Abbildung 3: Versuchsaufbau während der Fallstudie

Die unten abgebildeten generierten Aufnahmen während des Versuchs belegen die Machbarkeit des untersuchten Angriffsszenarios. So sind Inhalte des gedruckten Dokuments (C) sowie des Bildschirminhalts (B) bis zum Schriftgrad 28 deutlich lesbar. Hierbei gilt es zu beachten, dass mit dem Einsatz aktueller Drohnen ein besseres Ergebnis erzielt werden könnte. So wurde bei einer Referenzaufnahme, mit einer dem aktuellen Stand der Drohnentechnik entsprechenden Kamera, die Lesbarkeit bis Schriftgröße 12 festgestellt. Außerdem ist der Einsatz künstlicher Intelligenz zur Extraktion weiterer Informationen aus schwer leserlichen Texten denkbar [36, 37]. Dies wurde im Rahmen dieser Fallstudie jedoch nicht weiter untersucht.

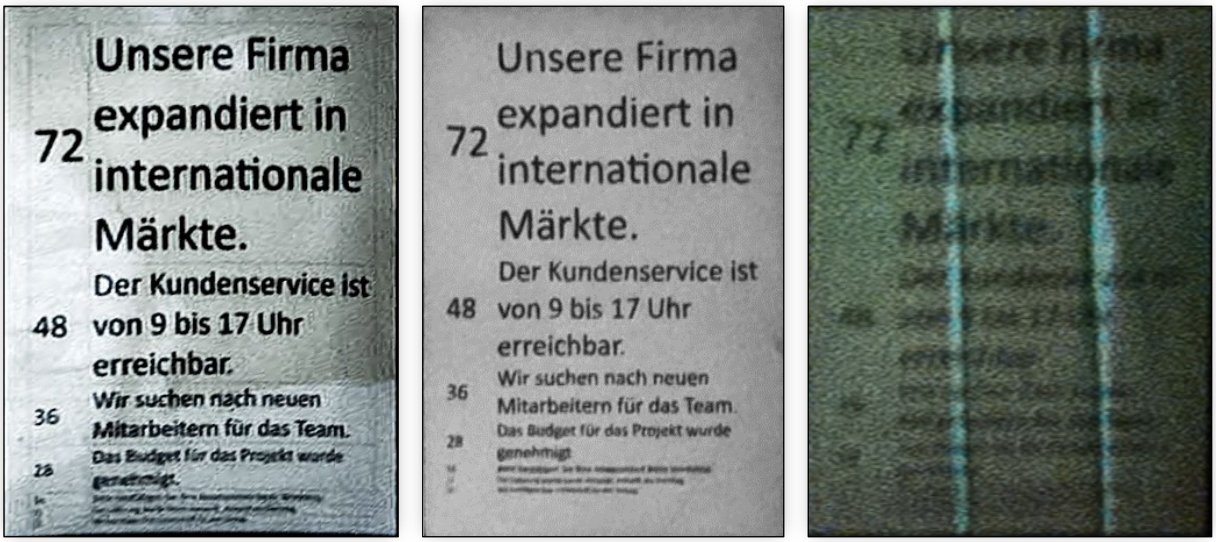


Abbildung 4: Ergebnisse aus Drohnenaufnahmen

Die obenstehende Abbildung zeigt zudem die Effektivität der Sichtschutzfolie als Abwehrmaßnahme. Hierbei konnten bei Nacht, wie dargestellt, beschränkt Informationen ausgelesen werden (SB). Bei Tageslicht ist keinerlei Informationserfassung durch die Drohne möglich. Dies bestätigt bedingt die im Rahmen der dritten Kernfrage untersuchte Effektivität der Abwehrmaßnahme.

Die Untersuchungen der Entdeckungswahrscheinlichkeit der Drohne deuten darauf hin, dass Mitarbeitende die Drohne erst durch Bewegung dieser im peripheren Sichtfeld wahrnehmen. So bemerkten zwei der fünf Probanden die Drohne nicht. Die übrigen drei Probanden entdeckten die Drohne erst auf dem Flugpfad P3, als diese bereits Information auf den vorherigen Flugpfaden erfassen konnte.

6 Abwehrmaßnahmen

Um eine ganzheitliche Erfassung der Gefahrenlage durch Drohnen auf die Informationssicherheit zu ermöglichen, wurde der aktuelle Stand technischer Systeme zur Drohnenabwehr analysiert. Hierbei wurden statt einzelner Systeme, technische Möglichkeiten allgemein analysiert. Dies soll der Schnelllebigkeit der kommerziellen Systeme in diesem Bereich Rechnung tragen und die Übertragbarkeit der Erkenntnisse sicherstellen.

Wie in [38-41] ist die Drohnenabwehr in zwei Bereiche unterteilbar. Den ersten Bereich bildet die Erkennung oder Detektion von Drohnen. Hierbei existieren verschiedene technische Möglichkeiten, welche, abhängig von den äußerlichen Gegebenheiten, Flugobjekte mit variabler Präzision detektieren

können. Den zweiten Bereich bilden Mitigationssysteme. Auch diese bedienen sich verschiedener destruktiver oder nicht destruktiver Maßnahmen, um den Weiterflug unbefugter Drohnen zu unterbinden. Um die Präzision dieser Systeme zu erhöhen, werden sie oft zusammen mit Detektionssystemen eingesetzt.

Auch wenn diese technischen Maßnahmen teils eine hohe Effizienz bei der Drohnenabwehr aufweisen, steht dieser Bereich vor einigen maßgeblichen Herausforderungen, wie in Tabelle 3 zusammengefasst.

Kategorie	**Beschreibung**
„Purpose Detection Problem“	Die Intention und das Vorhaben einer Drohne sind nicht eindeutig bestimmbar [42]
Kosten	Kosten von Abwehrsysteme 300.000-400.000 € (I2) stellen finanzielle Herausforderung dar (I1,2,5,6)
Datenschutz	Beim Einsatz von optischen Detektionssystemen
Umgebungs-abhängigkeit	Die Detektionssysteme sind je nach Interferenzen in der Umgebung unzuverlässig (I3)
Schadensrisiko	Insbesondere bei destruktiven oder störenden Abwehrmaßnahmen besteht Risiko für Sach- und Personenschaden

Tabelle 3: Herausforderungen technischer Drohnenabwehrsysteme

Eine weitere Art der Abwehrmaßnahmen wurde während der Experteninterviews identifiziert. So können Unternehmen neben technischen Maßnahmen auch bauliche und organisatorische Maßnahmen treffen, um sich gegen einige der identifizierten Drohnengefahren zu schützen. Eine Liste der während der Interviews identifizierten Maßnahmen stellt. Diese können, wie auch in der Fallstudie gezeigt, effektiven Schutz gegen Angriffe auf die Informationssicherheit bieten und begegnen vielen der in Tabelle 3 genannten Herausforderungen.

ID	**Art**	**Maßnahme**
I1, I4, I5	Organisatorisch	Schulung der Mitarbeiter zum Erkennen von Drohnen
I1	Organisatorisch	Erstellen eines Notfallkonzepts bei Sichtung einer unbefugten Drohne

I4	Organisatorisch	Regeln zum Verbot von Drohnenflügen auf dem Gelände
I6	Organisatorisch	Clean Desk Policy
I2, I4, I5	Baulich	Abkleben der Fenster mit abschirmender Folie
I4, I5	Baulich	Abschatten der Fenster
I2	Baulich	Besprechungsräume in der Mitte von Gebäuden platzieren
I2	Baulich	Firmengebäude in für Drohnenflug ungünstiges Gelände wie bewaldeten Gebieten bauen

Tabelle 4: Liste möglicher nicht-technischer Maßnahmen zur Drohnenabwehr

Auch wenn diese nicht-technischen Maßnahmen kosteneffizient sind, ist deren Effektivität aktuell noch wenig untersucht. All diese Hürden stellen die Abwehr von Drohnen als herausfordernde Disziplin für Organisationen dar. Die Abwägung einer Reihe wirtschaftlicher und ethischer Faktoren ist notwendig, um über Maßnahmen zur Drohnenabwehr zu entscheiden.

7 Fazit

Dieser Beitrag ging der Frage nach, welche Gefahren frei erhältliche Drohnen auf die Informationssicherheit von Unternehmen darstellen. Dazu wurden vier Ansätze verfolgt: Im ersten Ansatz wurden in einer Literaturanalyse mehrere Gefahrenkategorien identifiziert. Der zweite Ansatz umfasste das Führen und Auswerten von Interviews mit Experten, welche weitreichende Einblicke in die aktuelle und künftige Lage von Drohnengefahren sowie Möglichkeiten zur Drohnenabwehr gaben. Der dritte Ansatz dieses Beitrags umfasste eine praktische Fallstudie, in der mit Hilfe dreier Kernfragen, sowohl die Machbarkeit als auch das Entdeckungsrisiko und die Effektivität einer möglichen Abwehrmaßnahme untersucht wurden. Im letzten Teil wurden Abwehrmaßnahmen und deren Herausforderungen untersucht.

Zu Beginn lässt sich eine Forschungslücke im nicht destruktiven zivilen Einsatz von Drohnen zur Beeinträchtigung der Informationssicherheit von Unternehmen feststellen. Benachbarte Fachgebiete wie die Informationssicherheit in Drohnensystemen selbst oder der destruktive Einsatz von Drohnen so wie die Abwehr im militärischen Bereich sind bereits Bestand aktueller Forschung. Der in diesem Beitrag untersuchte Fachbereich jedoch nicht. Des Weiteren stellt der Beitrag fest, dass Drohnen eine Vielzahl von Gefahren für

die Informationssicherheit darstellen können. Außerdem zeigte die Fallstudie anhand des untersuchten Szenarios, dass Angriffe nicht nur machbar sind, sondern zusätzlich auch ein niedriges Entdeckungsrisiko für Angreifer mit sich bringen. Die Ergebnisse aus den Experteninterviews, der Fallstudie sowie die aktuellen Entwicklungen von Drohnen deuten darauf hin, dass die Machbarkeit und das Ausmaß erlangter Informationen in Zukunft weiter steigen werden. Die Untersuchung verschiedener Methoden zur Drohnenabwehr und Drohnendetektion zeigen deutliche Herausforderungen in Form von Risiken und Unzuverlässigkeit bei technischen Abwehrsystemen. Es wurde jedoch auch festgestellt, dass kosteneffiziente bauliche Nachrüstungsmaßnahmen oder organisatorische Schritte effizienten Schutz gegen bestimmte Drohnengefahren bieten können. Neben der potenziellen Gefahr, die von Drohnen ausgeht, wurden Indikatoren gefunden, welche die Eintrittswahrscheinlichkeit vieler identifizierten Gefahren als eher gering einstufen, da oft ein gewisses Maß von Planung und Ortskenntnis für die Durchführung vonnöten ist.

Die künftige Forschung im untersuchten Bereich ist vielfältig. So könnten zu Beginn genauere Abschätzungen der Eintrittswahrscheinlichkeit oder die Häufigkeit von Angriffen auf die Informationssicherheit durch Drohnen in Form von Umfragen, weiteren Interviews oder Analysen von Zwischenfällen, durchgeführt werden. Sollte künftig eine maßgebliche Gefahrenlage festgestellt werden, könnte die Betrachtung von Angriffsszenarien Einblicke in die Möglichkeiten von Angreifern geben. Somit ließen sich Abwehrmöglichkeiten potenziell verbessern. Diese Untersuchungen können auch von den Ergebnissen der Forschung im Bereich der Angreifer profitieren. Durch Bewertung häufiger Angriffsmuster und aktueller Abwehrmöglichkeiten könnten konkrete Handlungsempfehlungen abgeleitet werden. Diese können dann als Basis in Rahmenwerke zur Absicherung der Informationssicherheit oder in juristische Schriften einfließen. Diese wissenschaftliche Untersuchung des Fachbereichs trägt besonders unter Beachtung aktueller technologischer Trends im Bereich der Drohnen zur allgemeinen Verbesserung der Sicherheit bei.

Literaturhinweise

[1] Statista, Drone average price worldwide 2018 to 2028 (in 1,000 U.S. dollars). [Graph]. [Online]. Verfügbar: https://www.statista.com/forecasts/1399086/drone-average-price-worldwide (abgerufen am: 05.12.2023).

[2] M. G. Moehrle and H. Caferoglu, "Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology," Technological Forecasting and Social Change, vol. 146, pp. 776–784, 2019, doi: 10.1016/j.techfore.2018.07.049.

[3] M. Reger, J. Bauerdick, and H. Bernhardt, "Drohnen in der Landwirtschaft: Aktuelle und zukünftige Rechtslage in Deutschland, der EU, den USA und Japan," (in de), Landtechnik, vol. 73, no. 3, 2018, doi: 10.15150/lt.2018.3183.

[4] BIS Research, Global sky-based communication market revenue 2031. [Online]. Verfügbar: https://www.statista.com/statistics/1289151/sky-based-communication-market-revenue-by-application-worldwide/ (abgerufen am: 05.12.2023).

[5] DJI Official, DJI Mavic 3 Pro – Technische Daten – DJI. [Online]. Verfügbar: https://www.dji.com/de/mavic-3-pro/specs (abgerufen am: 02.12.2023).

[6] Statista, Global: drone market revenue by country 2022. [Online]. Verfügbar: https://www.statista.com/forecasts/1302524/ revenue-of-the-drone-market-worldwide (abgerufen am: 05.12.2023).

[7] Luftfahrt-Bundesamt (LBA), Statistiken Unbemannte Luftfahrtsysteme ab 2021. [Online]. Verfügbar: https://www.lba.de/SharedDocs/Downloads/DE/SBl/SBl3/Statistiken/Betrieb/UAS.html (abgerufen am: 29.09.2023).

[8] D. Kunertova, "The war in Ukraine shows the game-changing effect of drones depends on the game," Bulletin of the Atomic Scientists, vol. 79, no. 2, pp. 95–102, 2023, doi: 10.1080/00963402.2023.2178180.

[9] D. Kunertova, "The Ukraine Drone Effect on European Militaries," CSS Policy Perspectives, doi: 10.3929/ETHZ-B-000584078.

[10] B. Lutz, "Bedrohnung von oben: Analyse des Gefahrenpotenzials von frei erhältlichen Drohnen für die Informationssicherheit von Unternehmen: Bachelorthesis," abgerufen am: 11.02.2023. [Online]. Verfügbar: https://opus.hs-furtwangen.de/frontdoor/index/index/searchtype/latest/docId/10332/start/2/rows/10

[11] M. Hassanalian and A. Abdelkefi, "Classifications, applications, and design challenges of drones: A review," Progress in Aerospace Sciences, vol. 91, pp. 99–131, 2017, doi: 10.1016/j.paerosci.2017.04.003.

[12] S. A. H. Mohsan, N. Q. H. Othman, Y. Li, M. H. Alsharif, and M. A. Khan, "Unmanned aerial vehicles (UAVs): practical aspects, applications, open challenges, security issues, and future trends," Intel Serv Robotics, vol. 16, no. 1, pp. 109–137, 2023, doi: 10.1007/s11370-022-00452-4.

[13] DJI, DJI Inspire 3. [Online]. Verfügbar: https://store.dji.com/de/product/dji-inspire-3 (abgerufen am: 10.12.2023).

[14] ExpressLRS, Long Range Competition - ExpressLRS. [Online]. Verfügbar: https://www.expresslrs.org/info/long-range/ (abgerufen am: 13.09.2023).

[15] I. Bor-Yaliniz, M. Salem, G. Senerath, and H. Yanikomeroglu, "Is 5G Ready for Drones: A Look into Contemporary and Prospective Wireless Networks from a Standardization Perspective," IEEE Wireless Commun., vol. 26, no. 1, pp. 18–27, 2019, doi: 10.1109/MWC.2018.1800229.

[16] J. Glossner, S. Murphy, and D. Iancu, "An Overview of the Drone Open-Source Ecosystem," doi: 10.48550/arXiv.2110.02260.

[17] Yuneec E30Z Optical Zoom Camera Raw Test Footage - 1.2 Mile Line of Sight! - YouTube. [Online]. Verfügbar: https://www.youtube.com/watch?v=l19E7tsZ2_U (abgerufen am: 03.10.2023).

[18] Q. Yue, Z. Li, C. Gao, W. Yu, X. Fu, and W. Zhao, "The Peeping Eye in the Sky," 2018 GLOBECOM, pp. 1–7, doi: 10.1109/GLOCOM.2018.8647787.

[19] Q. Yue, Z. Ling, X. Fu, B. Liu, W. Yu, and W. Zhao, "My Google Glass Sees Your Passwords!," in Black Hat USA 2014. abgerufen am: 22.09.2023. [Online]. Verfügbar: https://www.blackhat.com/docs/us-14/materials/us-14-Fu-My-Google-Glass-Sees-Your-Passwords-WP.pdf

[20] Y. Abdrabou, Y. Abdelrahman, A. Ayman, A. Elmougy, and M. Khamis, "Are Thermal Attacks Ubiquitous?," 2020 AVI, pp. 1–5, doi: 10.1145/3399715.3399819.

[21] N. Alotaibi, J. Williamson, and M. Khamis, "ThermoSecure: Investigating the Effectiveness of AI-Driven Thermal Attacks on Commonly Used Computer Keyboards," ACM Trans. Priv. Secur., vol. 26, no. 2, pp. 1–24, 2023, doi: 10.1145/3563693.

[22] S. A. Macdonald, N. M. T. Alotaibi, M. S. Islam, and M. Khamis, "Conducting and Mitigating Portable Thermal Imaging Attacks on User Authentication using AI-driven Methods," in Proceedings of the Augmented Humans International Conference 2023, Glasgow United Kingdom, 2023, pp. 357–359.

[23] Y. Qu, T. Wang, and Z. Zhu, "Vision-Aided Laser Doppler Vibrometry for Remote Automatic Voice Detection," IEEE/ASME Trans. Mechatron., vol. 16, no. 6, pp. 1110–1119, 2011, doi: 10.1109/TMECH.2010.2077678.

[24] H. Zhang, T. Lv, and C. Yan, "The novel role of arctangent phase algorithm and voice enhancement techniques in laser hearing," Applied Acoustics, vol. 126, pp. 136–142, 2017, doi: 10.1016/j.apacoust.2017.05.024.

[25] H. Chen, W. Li, Z. Cheng, X. Liang, and Q. Zhang, "TCS-LipNet: Temporal & Channel & Spatial Attention-Based Lip Reading Network," in 2023 32nd International Conference on Artificial Neural Networks, pp. 413–424.

[26] Mark Vink, A comprehensive taxonomy of wi-fi attacks. Online, 2020. abgerufen am: 11.02.2023. [Online]. Verfügbar: https://www.ru.nl/publish/pages/769526/mark_vink.pdf

[27] M. Kammerstetter, M. Muellner, D. Burian, C. Kudera, and W. Kastner, "Efficient High-Speed WPA2 Brute Force Attacks Using Scalable Low-Cost FPGA Clustering," in pp. 559–577.

[28] A. Lonzetta, P. Cope, J. Campbell, B. Mohd, and T. Hayajneh, "Security Vulnerabilities in Bluetooth Technology as Used in IoT," JSAN, vol. 7, no. 3, p. 28, 2018, doi: 10.3390/jsan7030028.

[29] S. Khanji, F. Iqbal, and P. Hung, "ZigBee Security Vulnerabilities: Exploration and Evaluating," 2019 ICICS, pp. 52–57, doi: 10.1109/IACS.2019.8809115.

[30] I. Sanchez, G. Baldini, D. Shaw, and R. Giuliani, "Experimental passive eavesdropping of Digital Enhanced Cordless Telecommunication voice communications through low-cost software-defined radios," Security Comm Networks, vol. 8, no. 3, pp. 403–417, 2015, doi: 10.1002/sec.989.

[31] MouseJack, KeySniffer and Beyond: Keystroke Sniffing and Injection Vulnerabilities in 2.4 GHz Wireless Mice and Keyboards, 2016. [Online]. Verfügbar: https://media.defcon.org/def%20con%2024/def%20con%2024%20presentations/def%20con%2024%20-%20marc-newlin-mousejack-injecting-keystrokes-into-wireless-mice-wp-updated.pdf

[32] R. Guo, "Survey on WiFi infrastructure attacks," IJWMC, vol. 16, no. 2, p. 97, 2019, doi: 10.1504/IJWMC.2019.099026.

[33] Matthias Ghering, Evil twin vulnerabilities in wi-fi networks, 2016. abgerufen am: 11.02.2023. [Online]. Verfügbar: https://www.cs.ru.nl/bachelors-theses/2016/Matthias_Ghering__4395727__Evil_Twin_Vulnerabilities_in_Wi-Fi_Networks.pdf

[34] DJI Official, Mavic Pro - Produktinformationen - DJI. [Online]. Verfügbar: https://www.dji.com/de/mavic/info (abgerufen am: 02.12.2023).

[35] DJI Official, DJI Mini 2 - Technische Daten - DJI. [Online]. Verfügbar: https://www.dji.com/de/mini-2/specs (abgerufen am: 02.12.2023).

[36] J. Zhang and H. Qu, "Improvement of super resolution reconstruction method for real text images," 2022 MLISE, pp. 379–382, doi: 10.1109/MLISE57402.2022.00082.

[37] H. Cho, J. Wang, and S. Lee, "Text Image Deblurring Using Text-Specific Properties," 2012 ECCV, pp. 524–537, doi: 10.1007/978-3-642-33715-4_38.

[38] J. Wang, Y. Liu, and H. Song, "Counter-Unmanned Aircraft System(s) (C-UAS): State of the Art, Challenges, and Future Trends," IEEE Aerosp. Electron. Syst. Mag., vol. 36, no. 3, pp. 4–29, 2021, doi: 10.1109/MAES.2020.3015537.

[39] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, and J. Chen, "Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges," IEEE Commun. Mag., vol. 56, no. 4, pp. 68–74, 2018, doi: 10.1109/MCOM.2018.1700430.

[40] H. Kang, J. Joung, J. Kim, J. Kang, and Y. S. Cho, "Protect Your Sky: A Survey of Counter Unmanned Aerial Vehicle Systems," IEEE Access, vol. 8, pp. 168671–168710, 2020, doi: 10.1109/ACCESS.2020.3023473.

[41] F.-L. Chiper, A. Martian, C. Vladeanu, I. Marghescu, R. Craciunescu, and O. Fratu, "Drone Detection and Defense Systems: Survey and a Software-Defined Radio-Based Solution," Sensors (Basel, Switzerland), vol. 22, no. 4, 2022, doi: 10.3390/s22041453.

[42] B. Nassi, R. Bitton, R. Masuoka, A. Shabtai, and Y. Elovici, "SoK: Security and Privacy in the Age of Commercial Drones," 2021 IEEE SP, pp. 1434–1451, doi: 10.1109/SP40001.2021.00005

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Physische Penetrationstests im Kontext des bevorstehenden KRITIS-Dachgesetzes: Ein praxiserprobter Ansatz zur Resilienzerhöhung

Erfan Koza[1], Asiye Öztürk[2], Michael Willer[3]

Kurzfassung:
Das vorliegende wissenschaftliche Papier strebt danach, die bestehende Lücke in der Entwicklung spezialisierter Methoden zur physischen Sicherheitsüberprüfung zu schließen. Es präsentiert einen praxisorientierten Ansatz für den Schutz Kritischer Infrastrukturen, insbesondere vor dem Hintergrund der Anforderungen der Critical Entities Resilience (CER)-Richtlinie. Angesichts dieser Richtlinie ist es für Unternehmen von entscheidender Bedeutung, geeignete Verfahren zur Identifizierung und Stärkung ihrer physischen Widerstandsfähigkeit zu entwickeln. Unsere Studie stellt einen methodischen Ansatz für physische Sicherheitsaudits und Penetrationstests vor, der auf einer empirischen Machbarkeitserprobung in industriellen Umgebungen basiert. Die Wirksamkeit dieser Methodik wurde bereits erfolgreich in Unternehmen mit dem Schwerpunkt für biotechnologische Forschung, Versicherungs- und Finanzdienstleistungen und anderen spezifischen Branchen wie produzierende Betriebe in der Operation Technology demonstriert. Darüber hinaus kann dieser Ansatz als Prüfinstrument nahtlos in bestehende Managementsysteme wie ISMS oder B3S integriert werden. Die Umsetzung dieses Ansatzes ermöglicht es Organisationen, ihre physische Widerstandsfähigkeit zu stärken und partiell die Anforderungen der CER-Richtlinie zu erfüllen, um kritische Einrichtungen vor einer Vielzahl von physischen und menschen-zentrierten Bedrohungen zu schützen.

Stichworte: KRITIS-Dachgesetz, NIS 2.0, Pentesting, Physische Resilienz, Resilienz, Social Engineering

1 Veranlassung

Die zunehmende Einbettung der Informationstechnik (IT) in die vertikale und horizontale Wertschöpfungskette Kritischer Infrastrukturen (KRITIS) sowie die progressive Zunahme distribuierter Automatisierungstechnik haben in den vergangenen Jahren zu einer steigenden Abhängigkeit der KRITIS von störungsfreier und funktionierender Informations- und Automatisierungstechnik geführt. Im Kontext kontinuierlicher organisatorischer und betriebstechnischer Optimierungen durch Digitalisierung und Automatisierung zur Steigerung der betrieblichen Effizienz und Produktivität manifestiert sich jedoch auch eine weitere essenzielle Facette dieser Evolution: die

[1] Clavis Institut für Informationssicherheit an der Hochschule Niederrhein
[2] Clavis Institut für Informationssicherheit an der Hochschule Niederrhein
[3] Human Risk Consulting GmbH

zunehmende Interdependenz, Vulnerabilität und Fragilität digitaler Technologien und Informationssysteme, einschließlich der industriellen Steuerungs- und Automatisierungssysteme wie Industrial Control Systems (ICS), oder speicherprogrammierbare Steuerungen (SPS). Trotz ihrer scheinbaren Virtualität basieren sämtliche digitalen Prozesse und Strukturen letztlich auf physischen Fundamenten. Hardwarekomponenten wie Server, Router, Switche, Patchpanels und andere materielle Entitäten wie Gebäudeleittechnik (GLT), Klimatechnik, Unterbrechungsfreie Stromversorgung (USV) und Netzersatzanlagen bilden das Grundgerüst jeglicher digitalen Infrastruktur. Somit erfordert die Gewährleistung von Sicherheit und Widerstandsfähigkeit nicht allein die Berücksichtigung der digitalen Sphären, sondern gleichermaßen die des physischen Umfelds. Potenzielle Angriffe oder Störungen können sich in beiden Domänen manifestieren, weshalb eine ganzheitliche Schutz-Konzeptualisierung der logischen und physischen Welt unabdingbar erscheint.

Angesichts des Mottos „Die Welt ist im Wandel“ ist nun erkennbar, dass sich nicht nur Prozesse und Verfahren zu unserem Vorteil verändert haben, sondern auch die Angriffsstrategien und -methoden von Sabotagegruppen. Insbesondere besteht gegenwärtig die dringende Notwendigkeit, sich gegen die wachsenden Bedrohungen durch hybride Gefahrenvektoren und Sabotagegruppen zu verteidigen, die möglicherweise im Rahmen des russischen Angriffskrieges und der weltweit veränderten Sicherheitsarchitektur zur Destabilisierung der nationalen Sicherheit eingesetzt werden könnten. Diese Gefahren könnten darauf abzielen, KRITIS gezielt zu stören oder zu lähmen, was schwerwiegende Auswirkungen auf die Daseinsvorsorge der Gesellschaft und infolgedessen auf die nationale Sicherheit hätte.

Ein kohärentes und konkretes Beispiel für die physische Fragilität von KRITIS zeigte sich im Vorfall des Bahnverkehrsausfalls in Norddeutschland. Hierbei wurde die Vulnerabilität von verteilten Systemen und Anlagentechniken deutlich, als Kabel entlang der Bahnanlagen durchtrennt wurden. Dieses Ereignis unterstreicht die unabweisbare Notwendigkeit, auch dezentrale Systeme von KRITIS zu schützen, insbesondere in der Energiewirtschaft, wo verteilte Anlagentechniken in den Netzkoppel- und Übergabepunkten, Umspannwerken, Unterstationen, in der Trinkwasserversorgung und Abwasserbeseitigung wie Wasserwerken, Brunnen und Kläranlagen, sowie in Signalsystemen und Stellwerken im Verkehrssektor entscheidend sind.

Diese Betrachtung erfordert den nachhaltigen umgebungsbezogenen Schutz der physischen Werte der KRITIS, der neben den zentralen Anlagen und

Komponenten der IT und der operativen Technologie (OT) wie Leitzentralen, Rechenzentren und Technikräumen ebenfalls an Bedeutung gewinnt.

In der politischen Arena spielen Initiativen wie die CER-Richtlinie der Europäischen Union und das geplante KRITIS-Dachgesetz in Deutschland eine wichtige Rolle bei der Stärkung des physischen Schutzes der KRITIS. Diese Gesetzesinitiativen zielen darauf ab, die physische Widerstandsfähigkeit kritischer Einrichtungen gegenüber verschiedenen Bedrohungen wie Naturkatastrophen, Terrorismus und Sabotage zu stärken und die Zusammenarbeit der beteiligten Akteure zu verbessern. In diesem Zusammenhang konzentriert sich das Interesse dieser Forschungsarbeit auf die Entwicklung und Anwendung von Methoden zur Resilienzanalyse, insbesondere physischen Penetrationstests. Ziel ist es, einerseits neuralgische und verwundbare Stellen zu identifizieren und andererseits potenzielle Schwachstellen in der physischen Sicherheit von KRITIS aufzudecken, um darauf basierend geeignete Schutzmaßnahmen zu empfehlen. Durch einen methodischen und praxisorientierten Ansatz zielt diese Forschung darauf ab, die Widerstandsfähigkeit von KRITIS Infrastrukturen zu stärken und damit einen wesentlichen Beitrag zur Sicherheit und Stabilität zu leisten.

Die zugrunde liegende Zielsetzung konzentriert sich auf die Prävention und Reaktion gleichermaßen. Analog zur Strukturanalyse und Feststellung des Schutzbedarfs im BSI IT-Grundschutz-Kompendium oder im Kontext der Organisation gemäß Kap. 4 der ISO/IEC 27001:2022 liegt der Fokus darauf, die betroffenen zentralen und dezentralen Einheiten zu identifizieren, sie nach ihrer Ausfall- und Zeitkritikalität zu klassifizieren und somit die wesentlichen neuralgischen Punkte herauszufiltern [1]. Dadurch sollen präventive Maßnahmen zur Vermeidung von Störungen sowie reaktive und korrektive Planungsschritte zum Schutz dieser Einheiten implementiert werden.

1.1 Forschungsdefizit

Das KRITIS-Dachgesetz hat demzufolge das Ziel, die physische Sicherheit und Cyberresilienz von KRITIS zu stärken, um die Gesellschaft vor erheblichen Störungen und Bedrohungen zu schützen. Die physische und menschenzentrierte Sicherheit ist für KRITIS, die bereits sicherheitstechnische Verfahren wie ein Informationssicherheitsmanagementsystem (ISMS) oder einen Branchenspezifischen Sicherheitsstandard (B3S) implementiert haben, von großer Bedeutung.

Ein wesentlicher Bestandteil der Anforderungen betrifft daher die umgebungsbezogene Sicherheit, wie sie in den Normen ISO/IEC 27001:2017 und ISO/IEC 27019 (Abschnitt A.11 „Physische und umgebungsbezogene Sicherheit“) sowie in der ISO/IEC 27001:2022 (Abschnitt C.7 „Physical Controls“)

beschrieben ist [1], [2]. Dies umfasst Themen wie die Definition von Sicherheitsbereichen mittels eines Sicherheitszonenkonzepts, die Einrichtung physischer Sicherheitsperimeter und Zutrittskontrollen, die Sicherung von Büros, Räumen und Einrichtungen, das Arbeiten in Sicherheitsbereichen sowie den Schutz vor externen und umweltbezogenen Bedrohungen durch Aufbau von Redundanzen oder auch den Einsatz des Prinzips n-1-Kriterium. Diese Perspektive beruht darauf, dass KRITIS-Betreiber seit der Einführung der ersten beiden Körbe der BSI KRITIS-Verordnung im Jahr 2017 kontinuierlich mit diesen Themenbereichen konfrontiert wurden und bereits einige Maßnahmen in diesem Bereich implementiert haben.

In Analogie zu einem logischen Penetrationstest stellt sich primär die Frage, inwieweit KRITIS-Betreiber belastbare und objektive Aussagen über den Zustand, die Effizienz und die Wirksamkeit ihrer physischen Sicherheitsperimeter, Zonenkonzepte oder Zutrittssteuerungsmaßnahmen treffen können, wenn diese noch nie zuvor sachkundig und unparteiisch überprüft worden sind. Viele Unternehmen verfügen zwar auf dem Papier über sogenannte Zonenkonzepte mit einer Vielzahl physischer Barrieren und Mechanismen, deren Effizienz jedoch noch nie methodisch überprüft und evaluiert wurde. Zusätzlich stellt sich die Frage, wie die Unternehmen ihre Mitarbeiter in Bezug auf Themen wie Besuchermanagement oder das Arbeiten in Sicherheitsbereichen geschult haben.

Sekundär stellt sich bei KRITIS die Frage nach der Detektion und Reaktion auf physische Sabotageakte. Gemäß den Vorgaben des Prozessbausteins „Detektion and Response (DER)“, sowie der Abschnitte A.16 „Handhabung von Informationssicherheitsvorfällen“ und A.17 „Business Continuity Management“ in der ISO/IEC 27001:2017 sowie dem Abschnitt C.5 „Organizational Control“ der ISO/IEC 27001:2022 sind grundlegende Prozesse, Verfahren, Verantwortlichkeiten und Strukturen für das logische Vulnerability Management, Incident Response Management, Business Continuity Management und Notfallmanagement bereits etabliert [3] [4]. Diese kommen zum Einsatz, wenn Sicherheitslücken, logische Infiltrationen oder Kompromittierungen auftreten.

Vor diesem Hintergrund stellt sich die Frage, ob KRITIS vergleichbare spezifische Organisationsstrukturen sowie Alarmierungs- und Abwehrpläne für physische Sabotageakte vorbereitet und implementiert haben. Wenn ja, wie wirksam und machbar die definierten reaktiven und korrektiven Pläne sind.

1.2 Lösungsansatz

Selbst bei sorgfältiger Implementierung von Sicherheitsmaßnahmen und präventiven Ansätzen bleibt die Frage nach der Wirksamkeit und Effektivität

dieser Maßnahmen bestehen. Eine entscheidende Lücke in diesem Zusammenhang offenbart sich in der Notwendigkeit zur Durchführung von physischen Penetrationstests und Social-Engineering-Penetrationstests (SE-Pentests). Diese Tests dienen nicht nur dazu, bestehende Sicherheitslücken und Schwachstellen in den physischen und menschlich-zentrierten Sicherheitsperimetern sowie bei Zutrittssteuerungsmaßnahmen aufzudecken, sondern auch das Bewusstsein der Mitarbeiter für potenzielle Manipulationsversuche durch den Einsatz von Social-Engineering-Techniken zu schärfen.

Die Bedeutung solcher Verfahren wurde kürzlich vom Bundesamt für Sicherheit in der Informationstechnik (BSI) erneut betont, indem Unternehmen aufgefordert wurden, angesichts der aktuellen hybriden Gefahrenlage wachsam zu bleiben und die Mitarbeiter für Social-Engineering- und physische Angriffe zu sensibilisieren [5]. Der physische Pentest wird ebenfalls zur Überprüfung und Bestimmung des Reifegrads eingesetzt, um in erster Linie die Effektivität der Sicherheitszonenkonzepte und implementierten Resilienzmaßnahmen zu prüfen. Sekundär dient dieser analog zu logischen Pentests dazu, Red-Teaming-Übungen durchzuführen und durch Ausnutzung realer physischer Sicherheitslücken in sensible Bereiche einzudringen, beispielsweise um einen physischen Keylogger zu installieren.

Die Ergebnisse des physischen Pentests werden im Anschluss analysiert. Basierend auf den Erkenntnissen werden Maßnahmen zur Optimierung des erreichten Reifegrads implementiert. Die Durchführung solcher Tests kann dabei helfen, die tatsächliche Wirksamkeit der bestehenden Sicherheitsmaßnahmen zu bewerten und die Resilienz gegenüber physischen Sabotageakten zu erhöhen. Insbesondere in Anbetracht der steigenden Bedrohungen durch hybride Gefahrenvektoren und Sabotagegruppen ist es von entscheidender Bedeutung, diese Tests als integralen Bestandteil des Sicherheitsmanagements zu betrachten.

Durch die systematische Prüfung und Bewertung der physischen Sicherheitsmaßnahmen können potenzielle Schwachstellen identifiziert und behoben werden, bevor sie von Angreifern ausgenutzt werden können. Darüber hinaus tragen solche Tests dazu bei, die zu definierenden Abwehrmaßnahmen besser identifizieren, verorten, planen, umsetzen und fortlaufend anpassen und optimieren zu können.

Innerhalb des physischen Penetrationstests werden nicht nur Resilienzmaßnahmen und Sicherheitskonzepte überprüft sowie Sicherheitslücken identifiziert, sondern es werden auch die betroffenen Organisationseinheiten trai-

niert und in verschiedenen Bereichen sensibilisiert und geschult. Dies umfasst eine breite Palette von Aktivitäten, die darauf abzielen, das Bewusstsein für Sicherheitsrisiken zu schärfen, die Reaktionsfähigkeit auf Bedrohungen zu verbessern und das Sicherheitsverhalten der Mitarbeiter zu stärken. Durch Schulungen und Sensibilisierungsmaßnahmen können Mitarbeiter lernen, verdächtige Aktivitäten zu erkennen, angemessen zu reagieren und Sicherheitsrichtlinien und Verfahren einzuhalten. Auf diese Weise trägt der physische Penetrationstest nicht nur zur Identifizierung von Schwachstellen bei, sondern auch zur Stärkung der Sicherheitskultur und zum Aufbau einer widerstandsfähigeren Organisation.

2 Verfahrensvorstellung

2.1 Bedrohungsmatrix

Zur Operationalisierung der Zielsetzungen bedarf es zunächst einer detaillierten Bedrohungsanalyse zur präzisen Identifikation und Klassifizierung der beteiligten Rollen und physischen Einheiten. Üblicherweise stützt sich dieser Prozess auf Listen von Rollen und Verantwortlichkeiten sowie auf das Asset-Inventar des ISMS, welches insbesondere im Rahmen der Scoping-Phase definiert wurde.

Im Verlauf dieses Prozessschritts werden sämtliche Rollen und physischen Einheiten identifiziert und entsprechend ihrer Ausfallkritikalität für die Versorgungssicherheit klassifiziert, analog zum Prozess der Schutzbedarfsfeststellung. Zu diesem Zweck kommt die sogenannte „Bedrohungsmatrix“ zum Einsatz, welche als Instrumentarium für physische und personenbezogene Cyber Threat Intelligence fungiert. Jede Veränderung in der dynamischen Bedrohungslage führt zu einer neuen Beobachtung und Anpassung. Die visuelle Darstellung dieser Bedrohungslage bietet Chief Information Security Officers (CISOs) und Sicherheitsverantwortlichen eine strukturierte Grundlage für ihre taktischen Wahrnehmungs- und Ausrichtungsprozesse. Zudem ermöglicht es den Entscheidungsträgern, ihre Bedrohungslage entsprechend zeit- und ereignisbasierter Überwachungszyklen zu aktualisieren. Der Leitsatz „Wer schreibt, der bleibt“ betont die Bedeutung einer systematischen Erfassung und Dokumentation von Bedrohungsinformationen. Nur durch eine solche strukturierte Vorgehensweise können Bedrohungen und getroffene Gegenmaßnahmen effektiv ausgewertet, analysiert und kontinuierlich optimiert werden. Im Folgenden werden beispielhafte Anwendungsschritte der Bedrohungsmatrix erläutert.

2.1.1 Erster Schritt: Targeting

Die Phase des „Targeting“ ist nicht bloß als isoliertes Ereignis zu betrachten, sondern vielmehr als ein sich stetig entwickelndes Bewusstsein, das auf fortwährenden, wechselnden Gegebenheiten und unvollständigen Informationen beruht. Innerhalb der Verteidigungsstrategien nimmt diese Phase eine zentrale Stellung ein. Ihr Hauptziel besteht darin, die Herausforderungen der Entropie zu bewältigen, indem neue Informationen über die sich verändernde IT- und OT-Umgebung und die physischen Bedrohungslagen kontinuierlich berücksichtigt werden.

Dabei streben wir danach, ein breites Spektrum an Informationen in einem offenen System abzubilden. Aus diesem Grund wird eine Reihe von Beobachtungszielen und Instrumenten definiert, die sich bereits in der Praxis als effiziente Werkzeuge für das „Beobachten“ und „Orientieren“ bewährt haben. Anschließend werden diese in die Bedrohungsmatrix integriert.

Roles	***human-based Threats (HT)***						
	HT 1	HT 2	HT 3	HT 4	HT 5	HT 6	***Clustering***
R 1	x	x			x	x	Security Staff
R 2	x	x			x	x	
R 3			x				Engineering Maintenance
R 4		x			x	x	OT-Staff
R n+1		x			x	x	
Cluster							
	Department			***security requirement***			
CR 1	Security Staff			Critical (3)			
CR 2	Engineering Maintenance			High (2)			
CR 3	OT-Staff			Critical (3)			
phyiscal Units	***phyiscals Threats (PT)***						
	PT 1	PT 2	PT 3	PT 4	PT 5	PT 6	***Clustering***
P 1	x	x			x	x	Network interconnection point
P 2	x	x			x	x	
P 3	x	x	x		x	x	Network operation center
Cluster							
	Department			***security requirement***			
CP 1	Network interconnection point			Critical (3)			
CP 2	Network operation center			Critical (3)			
Assessment	***Threats Matrix***						
	CP 1		**Value**	***Matrix final value***			
CR 1	x		3 x 3	9			
CR 2	x		2 x 3	6			
CR 3	No allocation		No allocation	No allocation			
	CP 2		**Value**	***Matrix final value***			
CR 1	x		3 x 3	9			
CR 2	No allocation		No allocation	No allocation			
CR 3	x		3 x 3	9			

Abbildung 1: Exemplarische Darstellung einer Bedrohungsmatrix – Teil 1

Im Initialschritt des Verfahrens, dem Targeting, werden sämtliche Rollen und physischen Entitäten identifiziert und den potenziellen Szenarien, Angriffen und Bedrohungen zugeordnet. Dies schließt sowohl interne Rollen als auch externe Einheiten wie Lieferanten oder Dienstleister ein, die eine Funktion im Betriebsablauf oder der Versorgung haben. Nachdem diese Rollen und Einheiten identifiziert wurden, werden auf sie Social-Engineering- und physische Angriffsszenarien angewendet, um die potenziellen Risiken präzise zu erfassen. Mittels kontinuierlicher Beobachtung und Aktualisierung der Beobachtungsziele wird eine konsistente und transparente Darstellung der aktuellen Gefahrenlage und der betroffenen Elemente sichergestellt. Dies ermöglicht eine detaillierte Analyse der konkreten Gefahren und ihrer Auswirkungen auf spezifische Rollen und Systeme. Letztlich führt die Verringerung des Abstraktionsgrades dazu, dass die tatsächliche Gefahrenlage gezielter angegangen werden kann (vgl. Abbildung 1).

2.1.2 Zweiter Schritt: Clustern

Im nächsten Schritt werden Rollen mit vergleichbaren Aufgabenprofilen und ihrem Schutzbedarf zusammengeführt, sofern sie einer gemeinsamen Schnittmenge von Angriffsarten zugeordnet sind. Gleiches gilt für physische Einheiten, die nach ähnlichen Kriterien gruppiert werden können. Auf diese Weise entstehen homogene Cluster von Rollen und physischen Einheiten, die ähnlichen Angriffsvektoren und -arten ausgesetzt sind.

2.1.3 Dritter Schritt: Assessment

Nach dem Zusammenführen homogener Assets wird die Bewertung ihrer kritischen Bedeutung vereinfacht. Im nächsten Schritt erfolgt die Evaluierung der Rollen- und physischen Cluster anhand einer Skala von mäßig bis kritisch. Diese Skala muss individuell für jedes Unternehmen definiert und entsprechend qualitativ oder quantitativ interpretiert werden. Bei der Kritikalitätsbewertung wird die Fragilität, Vulnerabilität und potenzielle Auswirkung von Angriffen auf Personen, Gebäude und die Organisation berücksichtigt, wobei Sicherheits- und Funktionalitätsaspekte sowie die kontinuierliche Betriebsfähigkeit des Systems einfließen.

2.1.4 Vierter Schritt: Mapping

Im vierten Schritt werden die Bewertungen der Rollen- und physischen Cluster zusammengeführt, was zu einer umfassenden Bewertung der einzelnen Clusterbereiche führt. Dabei wird die Gesamtkritikalität ermittelt, was es ermöglicht, die kumulative kritische Bedeutung der Rollen und physischen Einheiten einzeln aber auch in Kombinatorik zu visualisieren.

2.1.5 Fünfter Schritt: Clusterbasierte Entscheidungsfindung

Basierend auf den erzielten Ergebnissen können im Resilienzplan proaktive, reaktive und korrektive Maßnahmen eingeführt werden, die spezifisch auf die identifizierten Rollen, Gebäude und die entsprechenden Gefahren ausgerichtet sind.

Dabei lassen sich grundsätzlich zwei Bereiche definieren. Zum einen die rollenbasierten (cluster role (CR)) und zum anderen die physischen Bereiche (cluster physical (CP)) (vgl. Abbildung 2). Die verschiedenen Arten von Social-Engineering- und physischen Angriffen, die identifiziert wurden, können für jeden Cluster nun individuell behandelt werden. Dies ermöglicht es, Entscheidungen darüber zu treffen, wie Schulungsprogramme für Informationssicherheit gestaltet werden sollen, um den verschiedenen physischen und Social-Engineering-Gefahren präventiv zu begegnen. Je nach Komplexität und Intensität der identifizierten Gefahren können die Bemühungen zur Sensibilisierung und Schulung der Mitarbeiter entsprechend angepasst werden.

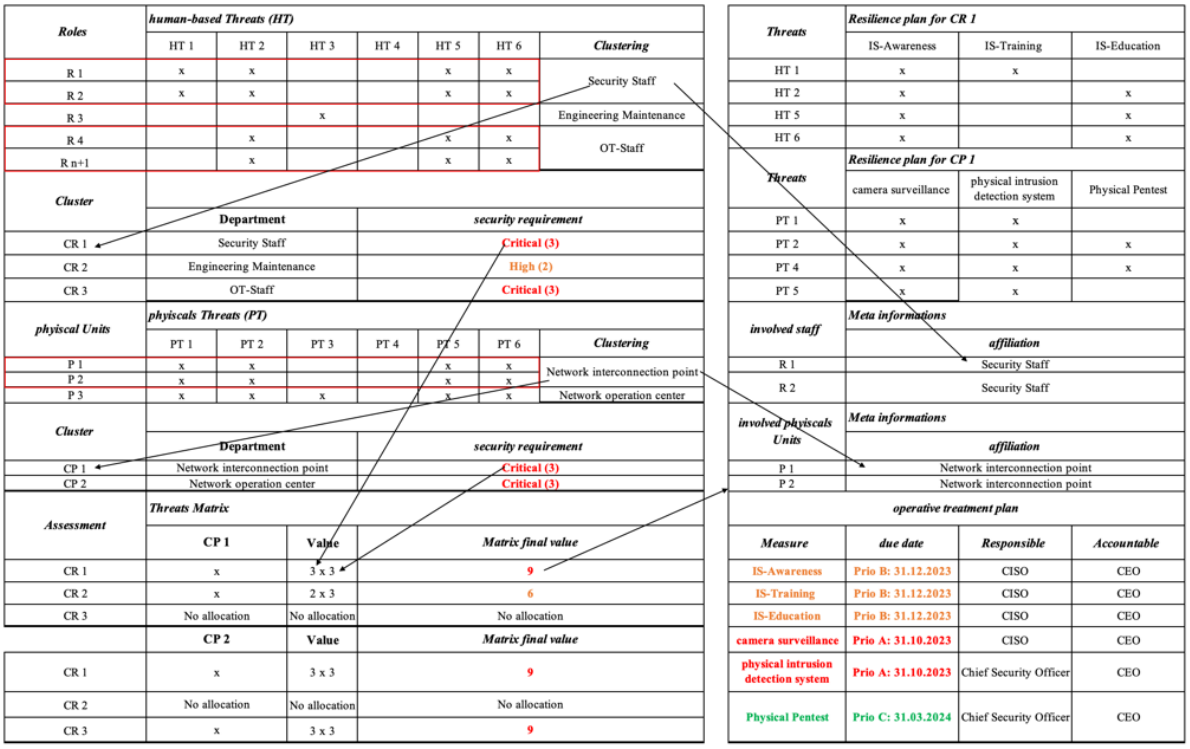


Abbildung 2: Exemplarische Darstellung einer Bedrohungsmatrix – Teil 2

2.1.6 Sechster Schritt: Definition von Verteidigungsmaßnahmen

Die Vielfalt der erkannten und zugeordneten physischen Angriffsarten kann nun spezifisch für jeden Cluster adressiert werden. Hierbei können für jede eingebettete physische Angriffsart Entscheidungen über die Definition von

Verteidigungsmaßnahmen getroffen werden. Es bietet sich die Möglichkeit, festzulegen, welche physischen Bedrohungen mit welcher Genauigkeit und Intensität präventiv, reaktiv, korrektiv oder auch detektierend bekämpft werden müssen. Dieser Prozessschritt kann ähnlich wie der fünfte Prozessschritt durchgeführt werden.

2.1.7 Siebter Schritt: Integration und Priorisierung

Alle festgelegten Maßnahmen aus den vorangegangenen Prozessschritten können in einem Resilienzplan zusammengeführt und zeitlich geplant werden, um sie praktisch umzusetzen und zu überwachen. Dies ermöglicht eine operative Ausführung, die an die spezifischen Bedrohungen und Ressourcenkapazitäten angepasst ist. Der Resilienzplan bietet zahlreiche Vorteile, indem er es ermöglicht, konkrete Bedrohungen aus der Beobachtungsphase gezielt anzugehen und sowohl ihre Ursachen als auch ihre Auswirkungen präzise zu reduzieren. Durch die flexible Struktur des Plans können verschiedene Umsetzungsstrategien entwickelt werden, die den individuellen Anforderungen und Risiken gerecht werden.

Die Planung von Maßnahmen wird durch den Resilienzplan präziser, was eine gezielte Auswahl ermöglicht, um die Wirksamkeit gegenüber spezifischen Bedrohungen zu maximieren. Eine optimale Ressourcennutzung wird durch den Plan erleichtert, was zu einer effizienten Ressourcenallokation führt. Transparente Zielsetzungen verbessern das Verständnis der Auswirkungen jeder Maßnahme und erleichtern die Kommunikation.

Die Implementierung des Plans ermöglicht eine effektivere Nutzung vorhandener Ressourcen und ermöglicht eine präzise Überwachung sowie Auswertung der Wirksamkeit jeder implementierten Maßnahme. Ein ganzheitlicher Ansatz schützt vor verschiedenen Angriffsvektoren und Szenarien. Der Resilienzplan unterstützt die frühzeitige Identifikation von Zielsetzungen und kann kontinuierlich angepasst werden, um auf sich verändernde Bedrohungen und Anforderungen zu reagieren.

2.2 Physisches Pentesting

Der physische Pentest ist eine ganzheitliche Untersuchung der Sicherheitsvorkehrungen innerhalb eines soziotechnischen Umfelds, speziell im Kontext der Interaktion zwischen Mensch und Maschine. Sein Hauptzweck besteht darin, Schwachstellen im Informations- und IT-Sicherheitskonzept einer Organisation sowie in den Resilienzmaßnahmen zu identifizieren und ihre Effektivität und Effizienz zu bewerten. Dieser Test fungiert als messbarer Indikator für das Resilienz-Level einer Organisation gegenüber physischen und Social-Engineering-Angriffen.

Im Rahmen des physischen Pentests werden gezielt Schwachstellen in den menschlichen und physischen Aspekten der Informationssicherheit analysiert. Er dient als Werkzeug zur Überprüfung und Simulation realistischer Social-Engineering-Angriffe oder Sabotageakte und umfasst sowohl logische Angriffsmethoden wie Social Media Intelligence (SOCMINT) und Open Source Intelligence (OSINT) als auch physische Angriffsvektoren wie Tailgating. Der Test folgt einem strukturierten Ansatz, der in vier modularen Prozessschritten durchgeführt wird:

Empirische Beobachtung (Findings)	Erkenntnisse aus der Angriffsphase
Pentester wird als unbekannte und augenscheinlich unternehmensfremde Person zwar von den Mitarbeitenden wahrgenommen, aber nicht angesprochen, obwohl dieser sich z.T. sehr auffällig verhält.	→ {**Mangelhafte oder nicht Vorhandensein von Fehlerkultur**: Unsicherheit und Angst etwas falsch zu machen} → {**Mangelhafte Sicherheitskultur oder mangelhaftes IS-Bewusstsein als kollektive Aufgabe**: Das ist nicht mein Verantwortungsbereich, dafür haben wir einen Empfangsdienst und Security.}
Pentester wird von einem Mitarbeitenden durch einen Nebeneingang einfach mit reingenommen, obwohl sich beide nicht kennen. (Tailgating)	→ {**Fehlende Kommunikation und unzureichende Schulung und Sensibilisierung der Mitarbeitenden**: Mir war nicht bewusst, dass das meine Aufgabe ist, darauf zu achten, wer hier reindarf und wer nicht.} → {**Fehlendes Wissen – Verhaltensabsicht und Salienz:** Warum sollte hier jemand rein wollen, was haben wir schon zu verbergen, wir sind doch nur eine Gemeindeverwaltung.}
Dem Pentester wird durch die Weitergabe eines Mitarbeiterausweises ermöglicht, eine Vereinzelungsanlage zu überwinden.	→ {**Menschliche Eigenschaften – Injunktive wahrgenommene Norm:** Ich wollte nicht unhöflich sein.} → {**Menschliche Eigenschaften – Verhaltensabsicht und Gewohnheit**: Das machen wir mittags immer so, da haben wir immer nur eine Karte dabei, wenn wir zusammen zum Essen gehen.}

Abbildung 3: Vier Phasen des physischen Pentests

2.2.1 Prä-Vorbereitungsphase: Feinplanung

In diesem Stadium werden die grundlegenden Bedingungen für den physischen Pentest festgelegt. Dabei werden klare Ziele definiert, wie beispielsweise das Verlocken von Mitarbeitern zum Klicken auf einen Link in einer Phishing-E-Mail, das Eingeben von Zugangsdaten auf gefälschten Login-Seiten (Pharming), das Preisgeben von Informationen, die Weitergabe von Fehlinformationen innerhalb der Organisation, das unberechtigte Herausgeben oder Versenden von Waren, das Überweisen von Geld, die Änderung von

Stammdaten, der Zugang zu Kundenobjekten sowie das unerkannte Eindringen in spezielle Sicherheitsbereiche, um dort einen simulierten physischen oder logischen Angriff durchzuführen. Neben den Zielen müssen auch die Verfahrensabläufe im Voraus klar definiert werden, wie beispielsweise das Verhalten der Penetrationstester bei Entdeckung durch Mitarbeiter oder Sicherheitspersonal [6].

Bei rein digitalen Social-Engineering-Pentests ist es wichtig, die für die Informationssicherheit verantwortlichen Stellen vorab zu informieren, um sicherzustellen, dass im Ernstfall zwischen dem simulierten Test und einem möglichen echten Angriff unterschieden werden kann. In der detaillierten Planung sollten auch andere relevante Akteure und Interessengruppen aus Querschnittsbereichen wie Datenschutz, Compliance und Betriebsrat einbezogen werden, um Fragen und Erwartungen abzugleichen.

2.2.2 Vorbereitungsphase: Erkundung

Während der Vorbereitungsphase [6] eines physischen Social-Engineering-Pentests erfolgt eine gründliche Vor-Ort-Analyse, um direkt am Zielort des Unternehmens ein umfassendes Verständnis zu entwickeln. Dabei spielen verschiedene Faktoren wie physische Sicherheitsperimeter, Umgebungsfaktoren, Abläufe, Routinen und Verantwortlichkeiten eine entscheidende Rolle bei der Entwicklung der eigentlichen Angriffsszenarien.

Für digitale Social-Engineering-Pentests werden spezifische Erstanalysen mittels OSINT- und SOCMINT-Methoden durchgeführt, um gezielt und präzise Informationen zu sammeln, die als Grundlage für die Entwicklung der Angriffsszenarien dienen. Als Ergebnis dieser Vorbereitungsphase werden die sogenannten Course of Action (COA) definiert, die die Art und Weise der Angriffsszenarien spezifizieren und verfeinern.

2.2.3 Hauptphase: Angriff

Nach den vorangegangenen Phasen folgt der eigentliche Angriff. Im Rahmen eines physischen Social-Engineering-Pentests werden dabei Identitäten, auch als „Legenden“ bezeichnet, entwickelt. Diese Legenden sollen es dem Social-Engineering-Pentester ermöglichen, unbemerkt in das Zielobjekt einzudringen und uneingeschränkte Bewegungsfreiheit zu erlangen. Der Pentester versucht dann beispielsweise durch physische Angriffsvektoren Zugang zum IT-System zu erhalten. Je nach Absprache mit der Zielorganisation können vorbereitete Wechseldatenträger wie USB-Sticks, physikalische Keylogger, WLAN-Sniffer und Abhörmittel eingesetzt werden. Auch analoge Informationen werden vom Pentester vor Ort gesichtet, aufgenommen und gegebenenfalls entwendet. Der physische Social-Engineering-Pentest (Variante I)

ist erfolgreich abgeschlossen, wenn das Zielobjekt unbemerkt und erfolgreich exfiltriert wird. Bei einem digitalen Social-Engineering-Pentest ist der jeweilige COA abgeschlossen, wenn eines der definierten Ziele erreicht wurde, ein weiterer COA eingeleitet bzw. unterstützt wurde oder der Angriff abgewehrt wurde [6].

In der Variante II enden die physischen Social-Engineering-Pentests nach der Durchführung der eigentlichen Angriffssimulation mit der Initiierung von Impulsen für die Sensibilisierung der Informationssicherheit. Beispielsweise kann der Pentester am Ende eines Social-Engineering-Pentests subtiler im Zielunternehmen agieren, bis er von einem Mitarbeiter angesprochen wird (bewusstes taktisches Aufdecken). Dann kann der Pentester beharrlich bei seiner Legende bleiben, sodass der Mitarbeiter einen Vorfall melden muss.

Der Fokus liegt neben der Reaktion der Mitarbeiter auch auf der Funktionstüchtigkeit und Effizienz von Meldewegen, Meldeketten, Alarmierungsplänen und Erreichbarkeiten. Alternativ kann der Pentester die Situation auflösen, den Mitarbeiter über die durchgeführte Simulation informieren und in ein aktives Sensibilisierungstraining übergehen.

Diese Vorgehensweise kann auch für logische Social-Engineering-Pentests angewendet werden, bei denen beispielsweise der Vishing-Angriff zur unautorisierten Erlangung sensibler Informationen oder zur Veranlassung von Fehlverhalten eingesetzt wird. In der Variante II eines logischen Social-Engineering-Pentests gibt sich der Pentester beispielsweise als Mitarbeiter eines bekannten Unternehmens aus und bittet um Hilfe bei einem vermeintlich dringenden betrieblichen Problem. Anschließend löst der Pentester die Situation auf, stellt sich vor, erklärt den Sinn und Zweck der Angriffssimulation und wird zum Sensibilisierungstrainer für Social Engineering [6].

2.2.4 Postphase: Nachbereitung und Ergebnisbesprechung

Die Postphase markiert den Abschluss eines Social-Engineering-Pentests und beinhaltet eine ausführliche Beschreibung der durchgeführten Szenarien sowie des Vorgehens der Social-Engineering-Pentester. Jeder Angriff wird in einer Situationsbeschreibung schriftlich festgehalten, und alle identifizierten Schwachstellen werden in einer gründlichen Analyse bewertet. Um die Mitarbeiter vor möglichen negativen Auswirkungen zu schützen, wird der Bericht anonymisiert und sicher im PDF-Format an die Zielorganisation übermittelt. Nach der Analyse werden die Ergebnisse in Post-Workshops, die als Debriefings bezeichnet werden, mit den Entscheidungsträgern für Informationssicherheit und den betroffenen Mitarbeitern besprochen. Ziel dieser Besprechungen ist es, das gezeigte menschliche Verhalten zu ana-

lysieren, mögliche individuelle Ursachen zu identifizieren und darauf aufbauend geeignete Maßnahmen zur Risikominderung zu entwickeln. Durch den Dialog mit den Mitarbeitern können oft die eigentlichen Probleme aufgedeckt werden, die hinter den Schwachstellen stehen [6].

3 Konklusion

Abschließend lässt sich feststellen, dass physische Pentests einen unverzichtbaren Beitrag zur ganzheitlichen Sicherheitsstrategie von Unternehmen leisten. Die Analyse physischer Sicherheitslücken ist von entscheidender Bedeutung, da diese oft unterschätzt werden und potenzielle Einfallstore für Sicherheitsvorfälle darstellen. Die Integration physischer Pentests ermöglicht es Unternehmen, subtile Schwachstellen in der physischen Sicherheitsinfrastruktur aufzudecken und präventive Maßnahmen zu entwickeln, um sich gegen Sicherheitsbedrohungen zu schützen. Insbesondere ist festzustellen, dass die Ergebnisse physischer Pentests oft auf Aspekte zurückzuführen sind, die indirekt mit Sicherheitsprozessen zusammenhängen. Beispielsweise können falsch konfigurierte Drehkreuze oder Vereinzelungsanlagen mit langen Schleusenzeiten, die Tailgating ermöglichen, zu Sicherheitsverletzungen führen. Ebenso können fehlkonzipierte Prozesse, bei denen Mitarbeiter täglich eine große Anzahl von E-Mails filtern und klassifizieren müssen, die Anfälligkeit für Social Engineering erheblich erhöhen. Darüber hinaus tragen mangelnde Fehler- und Sicherheitskultur sowie ein trügerisches Sicherheitsgefühl, das auf vergangenen Erfahrungen beruht, zu potenziellen Sicherheitsrisiken bei. Diese vielschichtigen Schwachstellen verdeutlichen die Notwendigkeit einer umfassenden Analyse und Stärkung der physischen Sicherheit, die weit über die rein technischen Aspekte hinausgeht. In Zukunft wird die Bedeutung physischer Pentests weiter zunehmen, da sich die Angriffsarten und -methoden ständig weiterentwickeln. Es ist entscheidend, dass Unternehmen nicht nur ihre digitalen Systeme, sondern auch ihre physischen Sicherheitsmaßnahmen regelmäßig überprüfen und verbessern. Dies erfordert klare Vorgaben und Richtlinien seitens des Gesetzgebers, um sicherzustellen, dass Unternehmen angemessene Sicherheitsvorkehrungen treffen und potenzielle Risiken proaktiv identifizieren und adressieren können. Letztendlich wird eine ganzheitliche Sicherheitsstrategie, die sowohl physische als auch digitale Sicherheitskomponenten umfasst, entscheidend sein, um die Informationssicherheit langfristig zu gewährleisten und die Resilienz gegenüber Sicherheitsvorfällen zu stärken.

Literaturverzeichnis

[1] ISO/IEC 27001:2022-10, Informationssicherheit, Cybersicherheit und Datenschutz - Informationssicherheitsmanagementsysteme – Anforderungen (ISO/IEC 27001:2013).

[2] DIN EN ISO/IEC 27019:2020-08, Informationstechnik – Sicherheitsverfahren – Informationssicherheitsmaßnahmen für die Energieversorgung (ISO/IEC 27019:2017, korrigierte Fassung 2019-08).

[3] DIN EN ISO/IEC 27001:2017:06 Informationstechnik – Sicherheitsverfahren – Informationssicherheitsmanagementsysteme – Anforderungen (ISO/IEC 27001:2013 einschließlich Cor 1:2014 und Cor 2:2015).

[4] BSI (2021): Bundesamt für Sicherheit in der Informationstechnik, DER: Detektion und Reaktion, DER.1: Detektion von sicherheitsrelevanten Ereignissen, S. 1-8.

[5] BSI (2023): Bundesamt für Sicherheit in der Informationstechnik: Die Lage der IT-Sicherheit in Deutschland 2023, Appel & Klinger Druck und Medien GmbH, Schneckenlohe, Oktober 2023.

[6] E. Koza, A. Öztürk, M. Willer (2023): Social Engineering Penetration Testing Within the OODCA Cycle–Approaches to Detect and Remediate Human Vulnerabilities and Risks in Information Security, in: International Conference on Applied Human Factors and Ergonomics, S. 72-82.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

ICS/OT-Sicherheit: Evaluation und Validierung der Erkennungsleistung von Stego-Malware in industriellen Steuernetzwerken mittels Synthese und Simulation

Dr.-Ing. Robert Altschaffel[1], Dr.-Ing. Stefan Kiltz[1], Kevin Lamshöft[1], Prof. Dr.-Ing. Jana Dittmann[1]

Kurzfassung:
Kritische Infrastrukturen sind zunehmend Angriffen ausgesetzt. Diese Angriffe setzen teilweise sehr fortgeschrittene Angriffstechniken wie verdeckte Funktionen und Wirkungsweisen ein. Dieser Beitrag beschäftigt sich damit, die Erkennungsleistung gegenüber solchen Angriffen zu evaluieren und zu validieren. Hierfür werden synthetische Netzwerkverkehre plausibel erzeugt und mit sogenannten hidden malicious dummies ohne Schadwirkung versehen um die Grundlage für Evaluierung und Validierung der Erkennungsleistung zu bieten. Dies soll erreicht werden, indem vordefinierte ausschließlich Schadcode-freie Prüfsequenzen (z.B. reine Textsequenzen) als Nachricht codiert und eingebettet werden. Als Voraussetzung für die Synthetisierung solcher Netzwerkverkehre wird ein Netzwerk-Kommunikationsverhaltens-Fingerabdruck (NKV-Fingerabdruck) für die segmentindividuelle Modbus/TCP-Kommunikation innerhalb eines leittechnischen Systems eingeführt und anschließend zur Synthesis und Plausibilitätsprüfung eingesetzt. Zusätzlich greifen wir das Thema der Attribution auf. Wir schlagen dazu die Anwendung des NKV-Fingerabdrucks als Ausgangsbasis vor, einen StegoMalware-Vorfall einem System-of-Origin zuzuordnen und zeigen, wie damit Angreifersignaturen eingegrenzt werden können [1].

Stichworte: Detektion, Industrielle Steuernetzwerke, Kritische Infrastrukturen, OT, Prävention, Steganografie

1 Einleitung

Kritische Infrastrukturen werden heutzutage von Schadcode bedroht. Dabei kommt auch vermehrt Schadcode mit verdeckten Funktionen und Wirkungsweisen, sogenannte StegoMalware, zum Einsatz, um unbeobachtet Schadwirkungen auszuführen (siehe [2]). Dieser Umstand motiviert dazu, frühzeitig Sicherheitsmechanismen, wie Industrial IDS (siehe [3]), auf ihre potenzielle Erkennungsleistung und ihr Reaktionsvermögen gegen diese Bedrohung zu prüfen und in realitätsnahen Umgebungen zu erforschen. Ziel ist die Betrof-

[1] Arbeitsgruppe Multimedia and Security, Institut für technische und betriebliche Informationssysteme, Otto-von-Guericke Universität Magdeburg

fenheit und Resilienz der Leittechnik in industriellen Steueranlagen einschließlich der Sicherheitsleittechnik (SILT) zu bewerten. Das vorliegende Papier aus dem Projekt SYNTHESIS[2] adressiert diese Lücke.

Zielsetzung ist es zu zeigen, wie eine frühzeitige Evaluierung und Validierung im Echtbetrieb von Anlagen bzw. deren Einzelkomponenten sowie Komponentenverbünden mittels nicht aktivem Schadcode erfolgen kann. Systeme sollen frühzeitig prüfbar, validierbar und gegebenenfalls gezielt gehärtet werden, ohne dass ein Angriffsvektor ausgeführt werden kann und darf.

Dieser Beitrag präsentiert Forschungsergebnisse für das weit verbreitete Modus/TCP-Protokoll. Dazu werden zwei zentrale Punkte diskutiert:

- Synthese: ob und wie synthetisch generierte, verdeckte, nicht-funktionell bösartige Nachrichten-Segmente (bezeichnet als hidden mali cious dummies ohne Schadwirkung) für Leittechnik erzeugt werden können und
- Simulation: wie diese zur Simulation in der Leittechnik genutzt werden können, um Netzwerksicherheitsmechanismen auf die Fähigkeit der Erkennung gezielt evaluieren zu können.

Erste Voraussetzung für die Synthese ist, dass die synthetisierten Netzwerkverkehre plausibel für die spezifische Anlage sein müssen, um Funktionsstörungen im Test auszuschließen und eine fälschliche Erkennung als Anomalie auszuschließen. Dafür wird in diesem Beitrag ein Fingerabdruck für die segmentindividuelle Modbus/TCP Kommunikation innerhalb eines leittechnischen Systems vorgeschlagen, welche für Plausibilitätsbetrachtungen genutzt wird. Eine besondere Herausforderung für die Plausibilität für Modbus/TCP und vergleichbare Protokolle in industriellen Einsatzszenarien ist, dass diese typischerweise Freiheitsgrade sowohl a) in der Protokollspezifikation, als auch b) in der Implementierung auf den Protokollstacks der Kommunikationsteilnehmer, als auch c) in den Nutzungsszenarien aufweisen [3], [4]. Die Nutzungsszenarien sind Anlagen-spezifisch. Entsprechend ist eine konkrete Beschreibung von a) – c) notwendige Voraussetzung für die Synthese von plausiblen Netzwerkverkehren, die für Evaluierung und Validierung geeignet sind.

[2] Teile dieser Veröffentlichung entstammen dem Forschungsvorhaben "SYNTHESIS - Synthetisch generierte Datensegmente mit verdeckten Schadcodefunktionen zur Sicherheitsanalyse in der kerntechnischen Leittechnik" mit der Projektnummer FKZ: 1501666A, welches vom Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV) gefördert wird.

2 Hintergrund

Diese Sektion gibt einen Überblick über die in diesem Beitrag verwendeten Begriffe und technischen Grundlagen.

2.1 Industrielle Steueranlagen

Es existieren unterschiedlich verwendete Begrifflichkeiten, die das Feld der industriellen Steueranlagen beschreiben: Leittechnik, Industrielle Steuernetzwerke, Industrial Control Systems (ICS), Operational Technologie (OT). All diesen Begrifflichkeiten ist gemein, dass digitale Rechentechnik dazu genutzt wird, um physikalische Prozesse zu überwachen, respektive zu kontrollieren – man befindet sich daher im Bereich der cyber-physischen Systeme (CPS). Im Rahmen dieses Beitrags werden die verschiedenen Begrifflichkeiten bis auf gekennzeichnete Einzelfälle synonym benutzt.

Generell bestehen Industrielle Steuernetzwerke aus Sensoren, Aktoren sowie Recheneinheiten, welche als Speicherprogrammierbare Steuerungen (SPS), respektive Programmable Logic Controller (PLC) bezeichnet werden. Innerhalb von Verarbeitungszyklen werden die Sensorwerte von den SPS verarbeitet und entsprechende Anweisungen an die Aktoren generiert. Diese Vorgehensweise setzt eine stete Kommunikation zwischen den einzelnen Komponenten voraus.

2.2 Modbus/TCP

Bei Modbus/TCP (eigentlich: MODBUS Messaging on TCP/IP) handelt es sich um eine Verwendung des weitverbreiteten Modbus Protokolls mit TCP/IP und damit verbunden Ethernet als Übertragungstechnologien. Modbus wurde 1979 entwickelt[3] und 2007 an den Einsatz mit TCP/IP angepasst[4]. Modbus verwendet ein Klient/Server-Modell. Der Klient stellt Anfragen an die Server, um Speicherbereiche zu lesen oder zu schreiben.

[3] MODBUS APPLICATION PROTOCOL SPECIFICATION V1.1b3, 2012; https://modbus.org/docs/Modbus_Application_Protocol_V1_1b3.pdf

[4] MODBUS MESSAGING ON TCP/IP IMPLEMENTATION GUIDE V1.0b. 2006; https://www.modbus.org/docs/Modbus_Messaging_Implementation_Guide_V1_0b.pdf

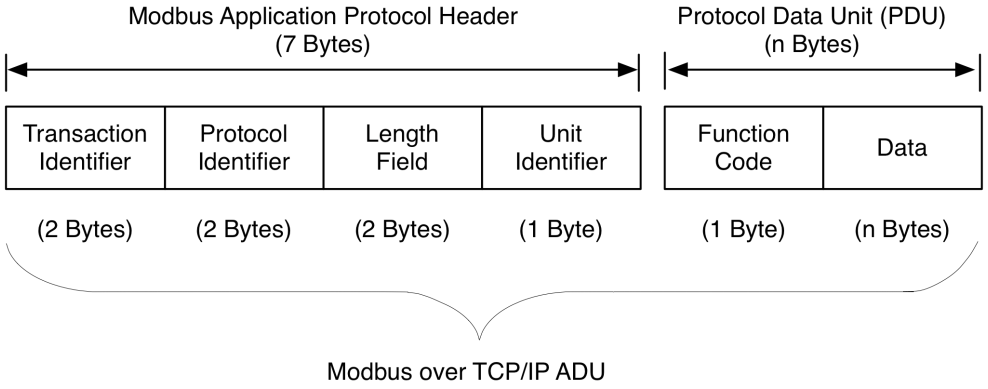


Abbildung 1: Ansicht eines Modbus/TCP Pakets mit Modbus Application Protocol Header (MBAP Header) und Protocol Data Unit (PDU), entnommen aus [5]

Diese Befehle sind in Function Codes (FC) codiert und verfügen je nach Ausprägung über Adressierung und Funktionswerte. Im Fall von Modbus/TCP wird ein MBAP (Modbus Application Protocol) Header verwendet. Dieser besteht aus einer Transaction ID (TID) welche Anfragen und Antworten verbindet, einer Protocol ID die statisch auf 0 gesetzt ist, um Modbus/TCP Pakete zu identifizieren, einem Feld, welches die Länge der Nutzlast (Payload - PL) beschreibt sowie einem Unit Identifier (UID), welcher zur Identifikation nachgeordneter und über serielle Verbindungen angebundener Geräte im Steuerverbund verwendet werden kann. Dem folgt die zuvor beschriebene Nutzlast. Ein solches Modbus/TCP Paket ist in Abbildung 1 dargestellt. Es ist möglich, dass ein System innerhalb eines Netzwerks als Klient und als Server agiert.

2.3 Steganografische Einbettung von Nachrichten in Modbus/TCP

In [5] sind Vorgehensweisen zum steganografischen Einbetten von Nachrichten in Modbus/TCP Kommunikation beschrieben. Dabei ist die Grundidee, dass StegoMalware häufig zum Angreifer zurück kommuniziert. Dies geschieht in Anwesenheit eines Beobachters (Warden). Generell kann die Einbettung in Zeitkanäle (zeitliches Kommunikationsverhalten) oder in Speicherkanäle (Modbus/TCP ADU/PDU Inhaltsveränderungen) unterteilt werden. Die Abbildung 2 aus [5] benennt Zeit- und Speicherkanäle zur Anwendung in Modbus/TCP. Dazu wird die Anwendbarkeit der Zeit- (T1–T8) und Speicherkanäle (S1–S10) auf die Kommunikationsteilnehmer oder das gesamte Netzwerk evaluiert.

Patterns by Mazurczyk et al. (2018)		Our Findings for Modbus/TCP		
ID	**Pattern**	**Master**	**Slave**	**Network**
T1	Inter-Packet Times	✔	✔	✔
T2	Message Timing	✔	✘	✘
T3	Rate/Troughput	✔	✘	✘
T4	Artificial Loss	✔	✔	✔
T5	Message Ordering	✔	✔	✔
T6	Retransmission	✔	✔	✔
T7	Frame Collisions	✘	✘	✘
T8	Temperature	✔	✔	✔
S1	Size Modulation	✔	✔	✔
S2	Sequence Modulation	✔	✘	✔
S3	Add Redundancy	✔	✘	✔
S4	Random Value	✘	✘	✔
S5	Value Modulation	✘	✘	✘
S6	Reserved/Unused	✔	✔	✔
S7	Payload File Size Modulation	✔	✘	✔
S8	User-date Corruption	✔	✔	✔
S9	Modify Redundancy	✘	✘	✘
S10	Value Modulation & Reserved/Unused	✘	✔	✔

Abbildung 2: Zeit- und Speicherkanäle für Modbus/TCP gekürzt aus [5] basierend auf den steganografischen Patterns aus [6] und der Anwendbarkeit für die Kommunikationsteilnehmer

3 Konzept zur Nutzung von Synthese und Simulation zur Evaluierung und Validierung der Erkennungsleistungen von Stego-Malware

Der in diesem Beitrag vorgestellte Ansatz ist ein mehrstufiger Prozess an dessen Ende synthetisierte Netzwerkverkehre mit inaktivem Schadcode stehen. Dieser Prozess ist in Abbildung 3 dargestellt.

Voraussetzung für die Synthese plausibler Netzwerkverkehre ist ein Verständnis des plausiblen Netzwerkverkehrs. Dieses Verständnis wird durch eine Beobachtung von Anlagensegmenten und den darin stattfindenden Modbus/TCP-Netzwerkverkehren erlangt. Zur systematischen Beschreibung dieses plausiblen Netzwerkverkehrs schlagen wir einen Netzwerkkommunikations-Verhaltens-Fingerabdruck (NKV-Fingerabdruck) vor.

Basierend auf diesem NKV-Fingerabdruck werden plausible Datenverkehre synthetisiert und anschließend für eine Qualitätsprüfung gegen den NKV-Fingerabdruck geprüft.

Anschließend erfolgt eine Einbettung von steganografischen Nachrichten ohne Schadwirkung und eine anschließende Prüfung gegen den NKV-Fingerabdruck. Wenn diese erfolgreich ist, liegen synthetisierte Netzwerkverkehre mit hidden malicious dummies ohne Schadwirkung vor.

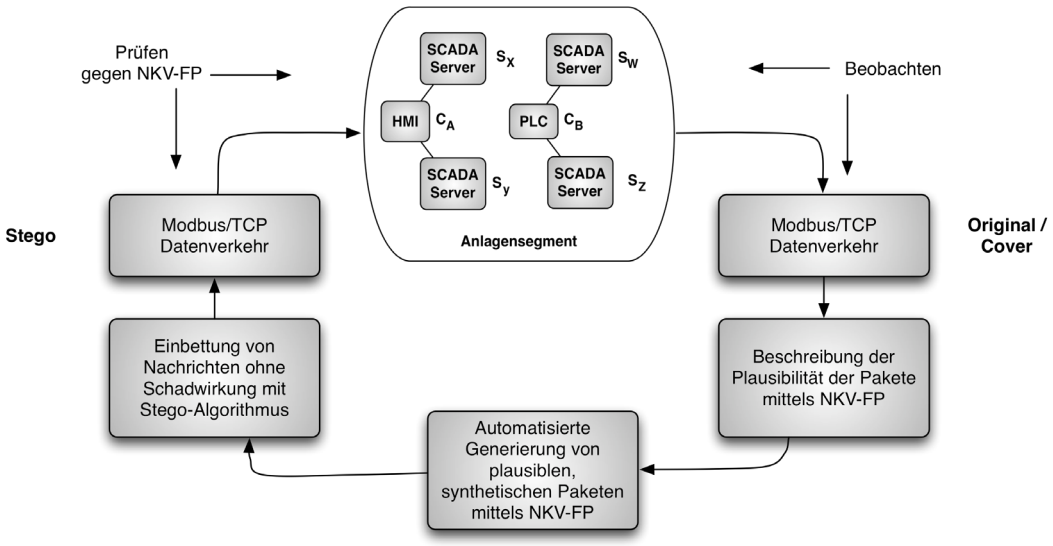


Abbildung 3: Mehrstufiger Prozess zur Synthesierung von Netzwerkverkehren mit hidden malicious dummies ohne Schadwirkung

4 Konzept eines NKV-Fingerabdrucks für Modbus/TCP

Voraussetzung für die Synthese plausibler Netzwerkverkehre ist ein Verständnis des plausiblen Netzwerkverkehrs. Dies unterliegt der Herausforderung verschiedener Freiheitsgrade (siehe a) - c)). Weiterhin besteht der Bedarf nach einer geeigneten Beschreibungsform.

Als grundlegende Vorgehensweise wird die Kommunikation einer leittechnischen Anlage von einer Makrosicht bis hin zu einer Mikrosicht segmentindividuell ganzheitlich als sinnvoll erachtet.

Unter Verwendung der Eigenschaften des Modbus/TCP Protokollaufbaus werden Beobachtungen aus der Sicht der gesamten Kommunikation im Segment gewonnen, andere hingegen konkret aus dem Inhalt einer spezifischen Kommunikationsverbindung oder sogar nur einer einzigen Nachricht und der darauf folgenden Antwort. Die Detailebenen (DE) für eine Modbus/TCP Kommunikation innerhalb eines Netzwerksegments werden wie folgt vorgeschlagen und sind in Abbildung 4 dargestellt:

- DE I: Segmentglobales Kommunikationsverhalten, d.h. Sicht auf alle Kommunikationspakete in einem Netzwerksegment. Eins je Netzwerksegment: DE I (bestehend aus C_A, C_B-S_W, C_B-S_Z)
- DE II: Multilaterales Kommunikationsverhalten eines Klienten mit allen mit diesem in Kommunikation stehenden Servern. Eines je Modbus/TCP Klient: DE II - C_A

- DE III: Bilaterales Kommunikationsverhalten zwischen einem Klienten und einem Server. Eines je Kommunikation zwischen Modbus/TCP Klient und Server: DE III C_B-S_W
- DE IV: Bilaterales und unidirektionales Kommunikationsverhalten (Antwortverhalten) innerhalb der Bearbeitung einer Anfrage eines Klienten an einen Server. Eines je Kommunikation zwischen Modbus/TCP Klient und Server: DE IV C_B-S_Z)

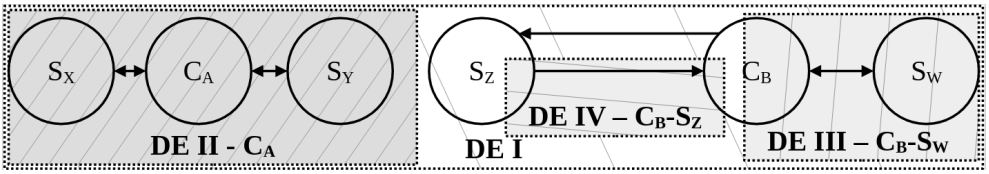


Abbildung 4: Unterschiedliche Detailebenen DE I - DE IV des NKV-Fingerabdrucks einer Modbus/TCP-Kommunikation

Diese Sicht kann auch auf verwandte Klient/Server-Protokolle aus der Leittechnik (z.B. OPC/UA[5]) angewendet werden.

Protokollspezifisch hingegen zugeschnitten auf Modbus/TCP werden die untersuchten Eigenschaften der Kommunikation genutzt:

- Paket: Transaction ID (TID) als Integer [0...65535]
- Paket: Unit ID (UID) als Integer [0...255]
- Paket: Function Codes (FC) als Integer [0...255]
- Paket: Payload (PL) in Bytes
- Kommunikationszyklusverhalten (KZV) in ms

Diese Kommunikationseigenschaften können unterschiedliche Ausprägungen haben: Sie können innerhalb einer Detailebene konstant sein, sich um einen spezifischen Wert inkrementieren, in einem bestimmten Wertebereich liegen oder eine abzählbare Anzahl an Ausprägungen haben. Die Darstellung des Kommunikationszyklusverhaltens ist komplexer und es erfolgt über eine statische Modellierung mittels Mittelwert, Standardabweichung und Mittel/Maximalwert. Diese komplexe Vorgehensweise wird auch für die Nutzlast (PL) notwendig sein – diese wird jedoch in diesem Beitrag noch nicht in der Tiefe betrachtet.

Die Interpretation der Merkmalsausprägungen erfolgt genau auf der jeweiligen Detailebene. Das heißt beispielsweise, dass auf Detailebene II die Zeitstempel aller Kommunikationsbeziehungen, an denen ein bestimmter Klient

[5] OPC Foundation: Unified Architecture - OPC Foundation, https://opcfoundation.org/about/opc-technologies/opc-ua/

CN beteiligt ist, gesamtheitlich (und nicht als Aggregat aus allen einzelnen Kommunikationsbeziehungen) betrachtet wird.

Ebene	TID			UID			FC			KZV – StdDev/Mean in MS		
	C	T1	S6	C	T1	S6	C	T1	S6	C	T1	S6
I	++1	++1	++1	211-214	211-214	100-300	3	3	3	500/ 2,5	500/ 3,5	500/ 2,5
II C_A	++2	++2	++2	211,212	211,212	100-300	3	3	3	1000/ 2,5	1000/ 2,5	1000/ 2,5
II C_B	++2	++2	++2	213,214	213,214	213,214	3	3	3	1000/ 2,5	1000/ 4,5	1000/ 2,5
III C_A-S_Z	++4	++4	++4	211	211	100-300	3	3	3	2000/ 2,5	2000/2,5	2000/ 2,5
IV C_A-S_Z	=	=	=	=	=	=	=	=	=	10/ 1,0	10/ 1,0	10/ 1,0
III C_A-S_Y	++4	++4	++4	212	212	212	3	3	3	2000/ 2,5	2000/ 2,5	2000/ 2,5
IV C_A-S_Y	=	=	=	=	=	=	=	=	=	10/ 1,0	10/ 1,0	10/ 1,0
III C_B-S_X	++4	++4	++4	213	213	213	3	3	3	2000/ 2,5	2000/ 4,5	2000/ 2,5
IV C_B-S_X	=	=	=	=	=	=	=	=	=	10/ 1,0	10/ 1,0	10/ 1,0
III C_B-S_W	++4	++4	++4	214	214	214	3	3	3	2000/ 2,5	2000/ 2,5	2000/ 2,5
IV C_B-S_W	=	=	=	=	=	=	=	=	=	10/ 1,0	10/ 1,0	10/ 1,0

Tabelle 1: Exemplarischer NKV-Fingerabdruck eines Netzwerksegments: Spalte ‚C' mit plausiblem NKV-Fingerabdruck, Spalten ‚T1' und ‚S6' mit NKV-Fingerabdruck für verschiedene hidden malicious dummies ohne Schadwirkung.

Durch die Eigenheiten des Modbus/TCP-Protokolls sollten die Felder TID, UID und FC bei der Antwort eines Servers auf die Anfrage eines Klients (also auf Detailebene IV) denen in der Anfrage entsprechen. Um dies darzustellen, wird hier als Notation „=" verwendet. Um darzustellen, dass eine erwartete Antwort ausbleibt, wird die Notation „-" verwendet. Um die Ausprägung eines Inkrements darzustellen, wird die Notation „++" verwendet. Wenn eine Eigenschaft nicht betrachtet wird (wie Nutzlast in diesem Beitrag), wird diese mit „nb" (nicht betrachtet) notiert.

Eine beispielhafte Darstellung eines plausiblen Fingerabdrucks (Spalte C) sowie beobachteter Fingerabdrücke mit steganografischer Einbettung entsprechend T1 und S6 (siehe Abschnitt 2) in tabellarischer Form findet sich in Tabelle 1. Die Werte für C sind absolut angezeigt - sie können in den Intervallen TID: 0-65535, UID: 211-214, KZV: 500 ms bei Standardabweichung von 2,5 plausibel schwanken. Der FC ist immer 3.

5 Evaluierung des Konzepts zur Erstellung von NKV-Fingerabdrücken

Für die Evaluierung des Konzepts zur Erstellung von NKV-Fingerabdrücken in industriellen Kontrollsystemen wurde ein Datensatz basierend auf der Netzwerkkommunikation zwischen Modicon 241 PLCs erzeugt und genutzt. Dieser umfasst etwa 90 Minuten Netzwerkverkehr. Zur Evaluierung wurde

zusätzlich ein umfangreicher Korpus von Netzwerkmitschnitten aus unterschiedlichen externen Quellen verwendet ([7], [8], [9]).

Aus diesen Samples wurden Fingerabdrücke erzeugt um damit 1) die Machbarkeit der Erzeugung von Fingerabdrücken zu evaluieren, 2) den Prozess der Erzeugung von Fingerabdrücken zu verbessern und 3) weitere Aspekte der NKV-Fingerabdrücke zu diskutieren.

Eine genaue Dokumentation der verwendeten Samples und resultierenden NKV-Fingerabdrücke ist online[6] verfügbar und zeigt die Machbarkeit der Erzeugung von NKV-Fingerabdrücken. Die Erzeugung wurde durch die Schaffung eines Netzwerkdissektors und durch Analyse-Skripte verbessert.

5.1 Ähnlichkeitsmaß zwischen NKV-Fingerabdrücken

Eine zentrale Frage ist ein Ähnlichkeitsmaß für NKV-Fingerabdrücke, um zu entscheiden, ob ein synthetisierter Verkehr plausibel – also dem originären NKV-Fingerabdruck ähnlich genug – ist (Tabelle 1, Spalte C).

Die aus [7] zitierten Samples stammen alle aus der gleichen Anlage mit der gleichen Konfiguration und führen zu einem identischen NKV-Fingerabdruck. Der aus Sample [9] generierte NKV-Fingerabdruck weicht deutlich davon ab. Ein zentrales Merkmal ist die Anzahl an Teilnehmern. Allerdings unterscheiden sich zwei der Samples aus [8] nur dadurch, dass in eines der beiden Samples ein weiterer Kommunikationsteilnehmer eingefügt wurde – es ist naheliegend, dass diese beiden Samples dennoch als ‚grundähnlich' bezeichnet werden können.

Zwei der Samples aus [8] entspringen einer bis auf die Zykluszeit der PLCs identischen Anlage. Hier unterscheiden sich auch die NKV-Fingerabdrücke nur im Merkmal KZV. Ob diese Ähnlichkeit ausreichend ist, hängt vom Anwendungsfall ab – in einem steganografischen Szenario könnte ein Warden jedenfalls eine solche Veränderung des zeitlichen Verhaltens als verdächtig erkennen, womit hier die Ähnlichkeit nicht ausreichend wäre.

Ein mathematisch bewertbares Ähnlichkeitsmaß zwischen NKV-Fingerabdrücken ist eine offene Forschungsfrage. Für den in diesem Beitrag betrachteten Zweck der Synthetisierung von Netzwerkverkehren jedoch ist eine Anwendungsfall-bezogene Betrachtung hinsichtlich der Frage, ob ein Unterschied für einen Warden auffällig wäre, gewählt worden.

6 https://gitti.cs.uni-magdeburg.de/raltschaffel/modbus-fingerprints-bsi-paper

6 Synthese von plausiblen Netzwerkverkehren basierend auf dem NKV-Fingerabdruck mit hidden malicious dummies ohne Schadwirkung

Zunächst wird basierend auf dem NKV-Fingerabdruck plausibler Netzwerkverkehr erzeugt, gegen den NKV-Fingerabdruck geprüft und anschließend mit hidden malicious dummies ohne Schadwirkung versehen.

6.1 Synthese plausibler Netzwerkverkehre

Die Synthese plausibler Netzwerkverkehre geschieht in mehreren Schritten, welche sich entlang der Detailebenen orientieren. Der Gesamtprozess aus Abbildung 3 ist beispielhaft in Abbildung 5 umgesetzt und ergibt eine Baumstruktur bei der sich die DE I aus den einzelnen Ausprägungen der DE II für die jeweiligen Klienten und die DE II aus den einzelnen Ausprägungen der DE III für den jeweiligen Klient zusammenfügt. DE IV nimmt eine Sonderrolle ein.

Dabei wird zunächst die DE III, welche das bilaterale Kommunikationsverhalten zwischen einem Klient und einem Server beschreibt, betrachtet. Hier wird ein Zeitstrahl der Kommunikation zwischen Klient und Server konstruiert, der entlang der im NKV-Fingerabdruck ausgeführten Eigenschaften befüllt wird. Das KVZ gibt dabei die Zeitachse vor. Die DE IV beschreibt das Antwortverhalten der Server, welches direkt in diesem Schritt niederschlägt.

In der Abbildung 5 ist dies beispielhaft basierend auf dem in Tabelle 1 aufgeführten NKV-Fingerabdruck ausgeführt. Für DE III+IV C_A-S_Z schreibt das KZV (in Zeile DE III C_A-S_Z) einen Zyklus mit einem mittleren Abstand von 2000 ms bei einer Standardabweichung von 2,5 vor. Entsprechend werden die Anfragen auf dem Zeitstrahl angetragen. Die Antworten folgen (in Zeile DE IV C_A-S_Z) in einem mittleren Abstand von 10 ms bei einer Standardabweichung von 1,0. Die Belegung der weiteren Felder ergibt sich aus dem NKV-Fingerabdruck: Die UID ist bei den Anfragen 211 und der FC ist 3. Die Antworten entsprechen den Werten der Anfragen. Die TID in den Anfragen inkrementiert jeweils um 4 und wird genauso auf die Antworten übertragen. Hier muss ein Anfangswert gewählt werden – in diesem Beispiel die 4.

Anschließend findet eine Betrachtung auf DE II statt. Diese beschreibt das multilaterale Kommunikationsverhalten eines Klienten mit allen mit diesem in Kommunikation stehenden Servern. Entsprechend wird DE II aus den verschiedenen DE III, an denen ein Server beteiligt ist, zusammengefügt. Dabei ist darauf zu achten, dass die Eigenschaften des NKV-Fingerabdrucks auch auf dieser Ebene erhalten bleiben.

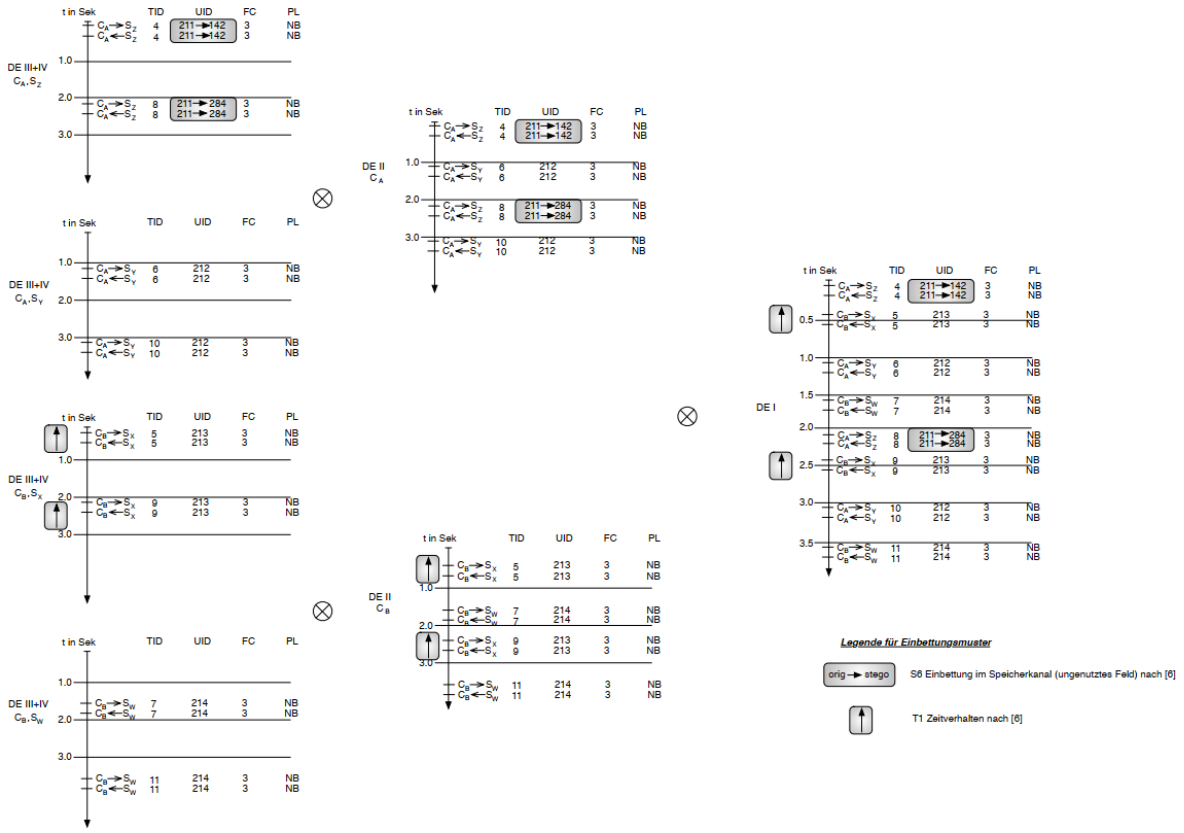


Abbildung 5: Vorgehensweise zur Synthese plausibler Netzwerkverkehre basierend auf NKV-Fingerabdrücken anhand des Beispiels aus Tabelle 1 mit der Einbindung von hidden malicious dummies ohne Schadwirkung; S6 verändert Fingerabdruck DE III+IV C_A-S_Z, DE II C_A und DE I; T1verändert Fingerabdruck DE III+IV C_B-S_X, DE II C_B und DE I

Für das Beispiel in Abbildung 5 gibt die KZV von DE II C_A die KZV die zeitliche Ausrichtung der beiden Kommunikationen, an denen C_A beteiligt ist, vor. Zwischen diesen befindet sich ein Abstand von 1000 ms bei einer Standardabweichung von 2,5 . Das führt zu den abwechselnden Instanzen von DE III+IV C_A-S_Z und DE III+IV C_A-S_Y die hier in einen gemeinsamen Zeitstrahl eingefügt werden.

Anschließend werden alle entstanden Zeitstrahlen für DE III zu einem gesamten Netzwerkverkehr auf DE IV zusammengesetzt, womit das segmentglobale Kommunikationsverhalten erzeugt wird.

In dem Beispiel in Abbildung 5 führt das KVZ von DE I mit dem mittleren zeitlichen Abstand von 500 ms bei der Standardabweichung von 2,5 dabei zu dem dargestellten abwechselnden Verhalten.

6.2 Synthese plausibler Netzwerkverkehre mit steganografischer Nachricht

Die in [5] aufgeführten Vorgehensweisen zum steganografischen Einbetten von Nachrichten in Modbus/TCP Kommunikation dienen als Basis zur Einbettung von hidden malicious dummies ohne Schadwirkung. Damit die daraus resultierenden Daten als plausibler Verkehr gelten, darf die Einbettung den NKV-Fingerabdruck der Kommunikation nicht verändern. Die Tabelle 2 gibt einen Überblick darüber, welche potenzielle Einflüsse auf den NKV-Fingerabdruck sich aus den jeweiligen Vorgehensweisen ergeben.

Vorgehensweisen, die in [5] als für Modbus/TCP nicht zutreffend oder relevant identifiziert wurden, werden nicht dargestellt. Einige weitere Vorgehensweisen beziehen sich auf das zugrundeliegende TCP/IP-Protokoll und sind daher nicht zutreffend.

Für eine Umsetzung wurden die Vorgehensweisen T1 und S6 ausgewählt. Entsprechend ist in Tabelle 1 in der Spalte T1 dargestellt, wie sich das steganografische Einbetten von Informationen durch Veränderung der Zeitabstände zwischen Paketen auf den NKV-Fingerabdruck in Form des KZV niederschlägt. Dies ist auch in Abbildung 5 dargestellt, wo das ZKV von DE III+IV C_B-S_X durch die Einbettung verändert wird. Diese Veränderung propagiert auf die höheren DEs.

Die UID gilt hier als ungenutztes/reserviertes Feld da diese im vorliegenden Netzwerksegment nicht ausgewertet wird. In Tabelle 1. ist in der Spalte ‚S6' dargestellt, welchen Einfluss die steganografische Einbettung in dieses auf den NKV-Fingerabdruck hat. Abbildung 5 umfasst ebenfalls ein Beispiel, bei dem diese Einbettung durchgeführt wurde, wodurch die UID in DE III+IV C_A-S_Z verändert wird. Diese Veränderung propagiert auf die höheren DEs.

Vorgehensweise zum steganografischen Einbetten in Modbus/TCP nach [5]	**Indikator aus dem NKV-FP**
T1 Inter-Packet Times	KZV
T2 Message Timing	Keiner
T3 Rate/Throughput	KZV
T4 Artifical Loss	KZV
T5 Message Ordering	Keiner
T6 Retransmission	KZV
T7 Frame Collisions	n.z.
T8 Temperature	n.z., da keine Seitenkanalinformationen zu den betrachten Samples
S1 Size Modulation	FC+PL
S2 Sequence Modulation	Nicht protokoll-konform umsetzbar ([5])
S3 Add Redundancy	FC+PL
S4 Random Value	Nicht verwendet im Protokoll ([5])
S5 Value Modulation	Nicht zutreffend ([5])
S6 Reserved/Unused	UID
S7 Payload File Size Modulation	FC+PL
S8 User data Corruption	PL
S9 Modify Redudancy	Nicht zutreffend ([5])
S10 Value Modulation & Reserved/Unused	PL

Tabelle 2: Vorgehensweise zum steganografischen Einbetten in Modbus/TCP basierend auf [5]

7 Ausblick - Perspektiven und Möglichkeiten der Attribution

Neben den Möglichkeiten der Evaluation von Sicherheitsmechanismen zur Detektion von StegoMalware ist auch die Attribution des Ursprungs von Interesse. Der NKV-Fingerabdruck liefert hier auch Ausgangspunkte und kann Anwendung im Projekt ATTRIBUT[7] finden, welches auf StegoMalware fokussiert und für diese Schadsoftwareklasse die Attribution von verdeckten (Informations-)Kanälen im Rahmen der Detektion im Bereich Kritischer Infrastrukturen bzw. HSK-Szenarien, sowie Potenziale für Prävention und Reaktion erforscht.

[7] Beauftragung durch die Agentur für Innovation in der Cybersicherheit GmbH: Forschung zu „Existenzbedrohenden Risiken aus dem Cyber- und Informationsraum – Hochsicherheit in sicherheitskritischen und verteidigungsrelevanten Szenarien“ (HSK) https://www.cyberagentur.de/tag/hsk/kritischhttps://attribut.cs.uni-magdeburg.de/

Konzeptionell wird, wie in Abbildung 6 dargestellt, eine erste Anwendung des Fingerabdrucks als Vorverarbeitungsschritt für eine Attribution wie folgt vorgeschlagen: Aufbauend auf dem Steganalyseergebnis im Netzwerkdatenstrom des Netzwerksegments wird im Verdachtsfall oder nach erfolgreicher Detektion von steganografischen Nachrichten der NKV-Fingerabdruck berechnet und auf Anomalien im Netzwerksegment geprüft. Falls im NKV-Fingerprint Anomalien festgestellt werden, erfolgt eine gezielte Prüfung auf und Zuordnung von möglichen Hiding Pattern (Stego-Einbettungsmuster) entsprechend Anomalien nach [5].

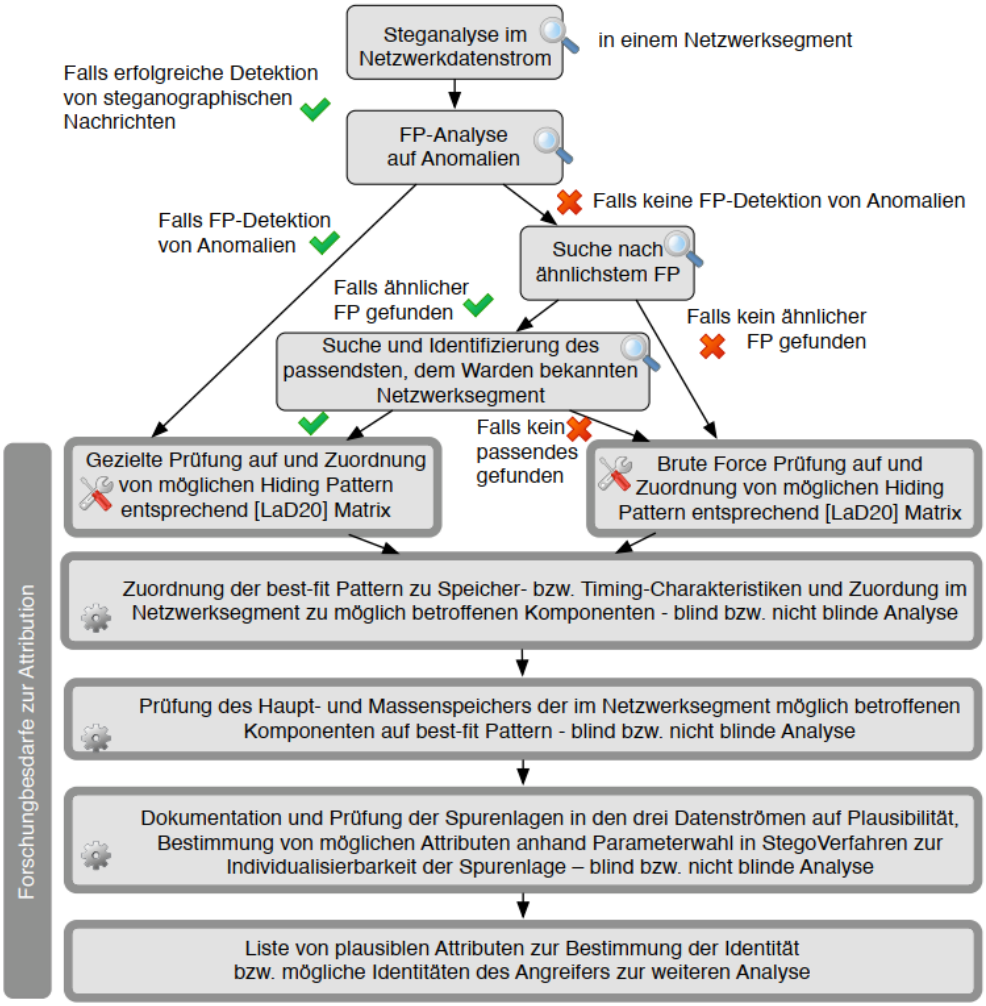


Abbildung 6: Genereller Ansatz zur Attribution von StegoMalware

Werden keine Anomalien beobachtet, erfolgt die Suche nach ähnlichstem NKV-Fingerabdruck in allen Netzwerksegmenten der Gesamtanlage. Werden ähnliche Fingerabdrücke gefunden, erfolgt eine Suche und Identifizierung des zum NKV-Fingerabdruck passendsten bzw. in der Anlage bekannten Netzwerksegment oder allgemein extern bekannte Anlagensegmente, aus

denen der Datenstrom stammen könnte, und es wird gezielt auf steganografische Nachrichten nach [5] geprüft. Falls kein ähnlicher Fingerabdruck gefunden wird: Brute Force Prüfung auf und Zuordnung von möglichen Stego-Einbettungsmustern [5].

Nach der Bestimmung von möglichen Stego-Einbettungsmustern erfolgt die Zuordnung der best-fit Hiding Pattern zu Speicher- bzw. Timing-Charakteristiken und die Zuordnung im Netzwerksegment zu möglichen betroffenen Komponenten. Betroffene Komponenten werden dann im Haupt- und Massenspeicher im Netzwerksegment vertieft auf best-fit Hiding Pattern überprüft. In der Prüfung können Original-bekannte Modbuspakete genutzt werden, falls diese vorliegen (nicht blinde Analyse) oder die Prüfung kann blind, ohne Kenntnis der vorkommenden Modbus-Pakete, arbeiten.

Die Spurenlagen in den drei Datenströmen (nach [10]) werden auf Plausibilität geprüft und mögliche Attribute zur Individualisierbarkeit identifiziert. Ausgangspunkte bilden individuelle Parameter der StegoVerfahren, wodurch eine Liste von plausiblen Attributen erstellt werden kann, die für die Eingrenzung und gegebenenfalls Bestimmung der Identität bzw. möglicher Identitäten des Angreifers zur weiteren Analyse dienen. Weitere Forschungsbedarfe zur Attribution liegen zum Beispiel in der Bestimmung zuverlässiger, plausibler und eindeutiger Attributionsmerkmale in den Spurenlagen.

8 Zusammenfassung und Ausblick

Dieser Beitrag schlägt das Konzept und die Verwendung eines Netzwerkkommunikations-Verhaltens-Fingerabdrucks (NKV-Fingerabdrucks) zur Beschreibung von Netzwerkverkehren in Leittechnik vor. Dies dient als Grundlage für die Generierung von synthetisierten Netzwerkverkehren, die nach Einbringung von hidden malicious dummies ohne Schadwirkung zur Evaluierung von Sicherheitsmechanismen in der Leittechnik verwendet werden können. Diese ersten Überlegungen stehen vor verschiedenen Herausforderungen und Fragen der Generalisierung wie auch der Spezialisierung:

- Verwendung des Payloads ist unvollständig betrachtet und bedarf einer weiterführenden Untersuchung.
- Die Generierung von NKV-Fingerabdrücken muss für weitere Architekturen geprüft und gegebenenfalls erweitert werden.
- Ableitung eines universellen NKV-Fingerabdrucks für verschiedene Protokolle.

Weiterhin wurden in diesem Beitrag erste Anknüpfungspunkte zur Attributierung von StegoMalware vorgestellt.

Literaturhinweise

[1] C. Kraetzer, J. Dittmann; Früherkennung von verdeckten Kanälen in VoIP-Kommunikation; Proceedings of the BSI-Workshop IT-Frühwarnsysteme, Bonn, Germany, July 12th, 2006.

[2] S. Wendzel, L. Caviglione, W. Mazurczyk, A. Mileva, J. Dittmann, C. Krätzer, K. Lamshöft, C. Vielhauer, L. Hartmann, J. Keller, T. Neubert: A Revised Taxonomy of Steganography Embedding Patterns, In Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES '21). Association for Computing Machinery, New York, NY, USA, Article 67, 1–12. DOI: 10.1145/3465481.3470069,2021

[3] W. Gao, T. Morris: On Cyber Attacks and Signature Based Intrusion Detection for Modbus Based Industrial Control Systems, In Journal of Digital Forensics, Security and Law: Vol. 9 : No. 1 , Article 3, pp. 37-56, DOI: 10.15394/jdfsl.2014.1162, 2014

[4] A. Keliris and M. Maniatakos: Remote field device fingerprinting using device-specific modbus information, 2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS), Abu Dhabi, United Arab Emirates, 2016, pp. 1-4, doi: 10.1109/MWSCAS.2016.7870006, 2016

[5] K. Lamshöft, J. Dittmann: Assessment of Hidden Channel Attacks: Targetting Modbus/TCP in IFAC-PapersOnLine, Volume 53, Issue 2, 2020, Pages 11100-11107, ISSN 2405-8963, DOI: 10.1016/j.ifacol.2020.12.258

[6] W. Mazurczyk, S. Wendzel, K. Cabaj: Towards deriving insights into data hiding methods using pattern-based approach. In Proceedings of the 13th International Conference on Availability, Reliability and Security, 10. 2018

[7] I. Frazão, P. H. Abreu, T. Cruz, H. Araújo, H., P. Simões: Denial of Service Attacks: Detecting the frailties of machine learning algorithms in the Classification Process, in 13th International Conference on Critical Information Infrastructures Security (CRITIS 2018), ed. Springer, Kaunas, Lithuania, September 24-26, 2018, Springer series on Security and Cryptology , 2018. DOI: 10.1007/978-3-030-05849-4_19

[8] A. Lemay and J. M. Fernandez: Providing SCADA network data sets for intrusion detection research, in 9th Workshop on Cyber Security Experimentation and Test (CSET 16), 2016;

[9] A collection of ICS/SCADA PCAPs, https://github.com/ITI/ICS-pcap/tree/master

[10] S. Kiltz: Data-Centric Examination Approach (DCEA) for a qualitative determination of error, loss and uncertainty in digital and digitised forensics, https://opendata.uni-halle.de/handle/1981185920/34842, 2020

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Zertifizierungen und agile moderne Softwareentwicklung – ein Widerspruch?

Alexander Küchler[1], Sven Merschjohann[2], Katharina Bogad[1], Konrad Hohentanner[1]

Kurzfassung:
Die derzeitige IT-Sicherheitslandschaft stützt sich stark auf Zertifizierungen, um die Sicherheit eines Produktes zu belegen. Für diesen Zweck muss der Hersteller des Produkts Nachweise in Form von textuellen oder graphischen Beschreibungen erbringen, welche sowohl Entwicklungs- und Qualitätssicherungsprozesse als auch die Sicherheitsleistungen des Produkts beschreiben und somit Rückschlüsse auf das Sicherheitsniveau des Produkts ermöglichen. Dies führt zu signifikanten Aufwänden auf Seiten des Herstellers, aber auch auf Seiten der Evaluatoren, welche die entsprechenden Nachweise verstehen, interpretieren und anschließend bewerten. Insbesondere die regelmäßigen Updates der agilen Softwareentwicklung erzeugen eine stark erhöhte Arbeitslast bei allen Beteiligten, da für jedes Update eine Rezertifizierung erfolgen muss. Zudem kann die manuell durchzuführende Konsistenzprüfung zwischen textuellen Nachweisen und dem Produkt schlimmstenfalls zur Divergenz zwischen dem „zertifizierten“ und „tatsächlichen“ Produkt führen. Um diesen Herausforderungen zu begegnen, muss eine Automatisierbarkeit der Nachweise ermöglicht werden. Dies erfordert allerdings, dass die derzeit verhältnismäßig starren, textbasierten Zertifizierungsprozesse grundlegend neu gedacht werden. Im Folgenden werden die aktuellen Prozesse kritisch beleuchtet und ein Lösungsansatz zur Erhöhung der Automatisierung aufgezeigt.

Stichworte: Agile Softwareentwicklung, Common Criteria, Compliance, DevOps, Rezertifizierung, Zertifizierung

1 Einleitung

Um Nutzern eines Produkts eine Einschätzung zur Sicherheit des Produktes zu geben, werden heutzutage verschiedene Zertifizierungen eingesetzt. Diese sind je nach Einsatzzweck unterschiedlich ausgestaltet und reichen von Konformitätserklärungen des Herstellers des Produktes (beispielsweise die CE-Kennzeichnung) bis hin zu aufwendigen unabhängigen Prüfungen durch externe Parteien (beispielsweise einer Zertifizierung nach Common Criteria). Dieses Vorgehen stammt initial aus der Entwicklung von physischen Gütern, die meist über einen langen Zeitraum nicht oder nur geringfügig modifiziert werden, sodass die getesteten Eigenschaften typischerweise nicht erneut überprüft werden müssen, und wurden in ähnlicher Form für

[1] Fraunhofer AISEC, Garching bei München
[2] Fraunhofer IEM, Paderborn

verschiedenste IT-Produkte adaptiert. Dabei wird sich auf dasselbe Vorgehen gestützt, wobei jedoch nicht unbedingt das Produkt selbst als Grundlage zur Prüfung dient. Stattdessen werden häufig Beschreibungen von Softwareeigenschaften, deren Entwicklungsprozesse, Benutzung und Betrieb als Grundlage der Prüfung herangezogen. Ein international anerkanntes und etabliertes Verfahren ist die Zertifizierung nach Common Criteria. Diese bildet beispielsweise auch die Grundlage zur Einführung angepasster Produktzertifizierungen auf Ebene der EU oder einzelner Nationen, um die Sicherheit oder Standard-Konformität von IT-Systemen zu bewerten.

Im Gegensatz zu physischen Gütern sollten Softwareprodukte immer auf einem aktuellen Stand gehalten werden, was zu vermehrten Updates führt und zukünftig sogar vom Cyber Resilience Act gefordert wird. Dies führt zu hohen Aufwänden, sowohl bei den Herstellern als auch bei den Evaluatoren, um dieser Masse von erforderlichen Rezertifizierungen gerecht zu werden. Kern des Problems ist die in der Praxis langwierige, aufwendige und nicht skalierende manuelle Erstellung und Evaluierung der Nachweise.

Nachfolgend wird das allgemeine Vorgehen bei einer Zertifizierung zusammengefasst. Anschließend werden die aus der aktuellen Umsetzung von Zertifizierungsprozessen entstehenden Herausforderungen mit Blick auf eine moderne Softwareentwicklung erörtert. Zuletzt wird aufgezeigt, wie Prozesse zukünftig gestaltet werden können, um diesen Herausforderungen zu begegnen. Hierfür muss insbesondere ein Fokus auf eine hohe Automatisierbarkeit bei allen Beteiligten gelegt werden.

2 Background

2.1 Aktuelle Zertifizierungsprozesse

Derzeit lassen sich viele Zertifizierungsprozesse vereinfacht durch die folgenden Arbeitsschritte der verschiedenen Stakeholder zusammenfassen:

1. *Standardisierungsgremien*, welche meist aus Unternehmen der Wirtschaft, öffentlichen Einrichtungen und Forschungseinrichtungen bestehen, erarbeiten Anforderungen, die das Produkt oder die Nachweisdokumente einhalten müssen, um standardkonform zu sein. Ein Standard ist meistens in einem natürlichsprachlichen Dokument spezifiziert, das Hinweise zur Umsetzung enthalten kann, aber häufig allgemein bleibt. Erfahrungen mit dem Standard in der Praxis werden typischerweise in eine weitere Iteration des Standards zurückgespielt und eine aktualisierte Version des Standards wird veröffentlicht.
2. Der *Hersteller* beschreibt die Sicherheitseigenschaften des Produkts, dessen Architektur, Abhängigkeiten, die Implementierung und Tests,

dokumentiert Entwicklungsprozesse und schreibt Nutzerhandbücher. Obwohl einige der Dokumente oftmals ohnehin im Rahmen einer professionellen Softwareentwicklung erstellt werden sollten, sind insbesondere die standardkonforme Beschreibung verschiedener Prozesse und die Zusammenstellung der Nachweise nicht unbedingt Teil der Softwareentwicklung. Auch müssen diese – je nach angestrebter Zertifizierung – unterschiedlich aufbereitet und deren Umsetzung auf andere Arten nachgewiesen werden, was erhebliche Aufwände darstellt.

3. Eine *Prüfeinrichtung* liest die zur Verfügung gestellten Dokumente und schätzt daraufhin (ggf. nach Rücksprache mit dem Hersteller) die Produktsicherheit ein. Der Evaluator muss hierfür die Existenz verschiedener Inhalte prüfen, aber auch einschätzen, ob die Beschreibungen ausreichen, um auf eine sichere Produktentwicklung sowie ein sicheres Produkt zu schließen. Diese Einschätzung ist oftmals eine Ermessensfrage, da aufgrund des hochgradig dynamischen Umfelds und sehr unterschiedlicher Produkte in der Regel keine konkreten Handlungsanweisungen existieren. Dies macht die Evaluierung sehr aufwendig. Zudem soll – je nach Zertifizierung – auch die Implementierung oder eine verwandte Darstellung stichprobenhaft betrachtet werden.
4. Nach erfolgreicher Prüfung stellt eine vertrauenswürdige Instanz das Zertifikat aus.

Dieser Prozess erfordert die Einschätzung eines oder mehrerer Experten und ist aufgrund der großen Unterschiede zwischen den verschiedenen Produkten und der Einlieferung vieler natürlichsprachlicher Beschreibungen nicht automatisierbar. Zudem ist ein großer Interpretationsspielraum bei der Bewertung von Maßnahmen zur Sicherstellung der Sicherheitseigenschaften vorhanden, sodass auch hier eine Expertenmeinung nötig wird.

2.2 Requirements Engineering

Die Common Criteria (CC) [1] selbst bietet einen umfassenden Satz von Anforderungen (Requirements) an die Formulierung und Dokumentation von Sicherheitseigenschaften. Besonders relevant sind die Sicherheitsziele, Bedrohungen, organisatorischen Sicherheitsrichtlinien und Annahmen über die Einsatzumgebung des Produkts. Das BSI bietet zusätzliche Ressourcen und Anleitungen, die die internationalen CC-Standards ergänzen.

In der Praxis umfasst das Requirements Engineering zur Nachweiserbringung die Entwicklung von Schutzprofilen und Sicherheitszielen. Schutzprofile definieren dabei Anforderungen an alle Produkte einer Kategorie, wie z.B. Firewalls. Sicherheitsziele hingegen beschreiben die genaue Sicherheitsspezifikation eines konkreten Produkts [2].

Der INCOSE „Guide to Writing Requirements“ [3] bietet detaillierte Anleitungen zur Erstellung von Anforderungen für erfolgreiche Systementwicklungsprojekte. Um die Qualität und Effektivität des Entwicklungsprozesses zu verbessern, sollen Anforderungen definiert werden, die eindeutig, vollständig, konsistent, realisierbar, relevant und testbar sind. So formulierte Anforderungen ermöglichen die Verifikation und Validierung des Systems.

2.3 Automatische Softwareanalysetechniken

In der Entwicklung von Software hat sich der Einsatz verschiedener Analysetools etabliert, um frühzeitig Abweichungen von Qualitätskriterien, Anforderungen oder mögliche Schwachstellen zu identifizieren. Die Verfahren lassen sich grob in statische und dynamische Verfahren einteilen. Statische Verfahren betrachten den Code oder das Softwareartefakt und analysieren diese ohne die Software auszuführen. Im Gegensatz dazu wird bei dynamischen Verfahren das Verhalten eines Produkts während der Ausführung beobachtet. Im Rahmen der Sicherheitsanalyse haben sich dabei die Begriffe *Static Application Security Testing* (SAST), *Dynamic Application Security Testing* (DAST) und *Fuzzing* als Unterkategorie dynamischer Verfahren etabliert.

Neben dieser groben Unterteilung gibt es vor allem bei statischen Analyseverfahren eine feingranulare Unterscheidung in eine Vielzahl verschiedener Vorgehensweisen, die für die Analysen genutzt werden. Für die nachfolgenden Arbeiten ist vor allem eine Unterscheidung in zwei Kategorien relevant:

1. Compliance-Prüftools: Verfahren und Tools, die einen Softwarestand gegen definierte Anforderungen prüfen. Hierbei wird festgelegt, wie das Produkt implementiert werden muss. Diese Verfahren könnten z.B. formale Verifikationstechniken und Model Checking umfassen.
2. Verfahren und Tools, die überprüfen, ob ein Softwarestand potenziell gefährliche Code-Muster aufweist. Dies kann z.B. die einfache Suche nach der Verwendung potenziell gefährlicher Funktionen umfassen.

Derzeit sind vor allem Tools der zweiten Kategorie in der Praxis verbreitet, aber auch für Compliance-Prüftools gibt es erste nutzbare Ansätze.

2.4 Sichere Softwareentwicklungsprozesse

In vielen Unternehmen gibt es für die Softwareentwicklung Prozesse, die die Qualität und Sicherheit der Produkte erhöhen sollen. In den letzten Jahren wurden verschiedene solche Prozesse, auch Secure Development Lifecycle (SDLC) genannt, und Metriken vorgeschlagen. Beispiele für SDLC stammen von Microsoft [4] oder der Cisco [5]. OWASP SAMM2 [6] stellt Metriken zur Einschätzung des Reifegrads solcher Prozesse vor. Die NIST SP 800-218 [7] fasst Modelle und Umsetzungsempfehlungen zusammen.

Alle Empfehlungen beinhalten stets ähnliche Aktivitäten:

1. *Planen, Vorbereiten und Definition organisatorischer Maßnahmen:* Zunächst identifiziert der Hersteller geeignete Maßnahmen, die während der Produktentwicklung umzusetzen sind. Dies umfasst die Verteilung von Verantwortungen innerhalb des Unternehmens oder einzelner Teams, die Definition einzuhaltender Prozesse bei der Softwareentwicklung, die Erstellung von Best Practices zur sicheren Entwicklung, sowie die Erstellung eines Schulungsangebots. Darüber hinaus werden Anforderungen an die Infrastruktur, Zulieferer, sowie Qualitätsanforderungen definiert. Auch die Auswahl geeigneter Tools für die Entwicklung, zur Unterstützung der Entwickler und zur Qualitätskontrolle fallen in diese Aktivitäten. Alle Entscheidungen werden dokumentiert.
2. *Anforderungen an das Projekt definieren:* Sind allgemeine Vorgaben an Entwicklungsprozesse ausgearbeitet, werden die Anforderungen für das jeweilige Projekt definiert. Dies umfasst den benötigten Funktionsumfang und Schnittstellen zu anderer Software oder der Umgebung.
3. *Durchführung einer Risikoanalyse:* Für jedes Produkt wird auf Grundlage der Anforderungen an das Produkt eine Risikoanalyse durchgeführt., deren Ergebnisse dokumentiert werden. Aus der Risikoanalyse werden Maßnahmen zur Minimierung von Risiken abgeleitet.
4. *Design und Entwicklung des Produkts:* Das Produkt wird basierend auf den Anforderungen zunächst architekturell entworfen und anschließend implementiert. Hierbei sind die in Schritt 1 festgehaltenen Maßnahmen einzuhalten und auch die in Schritt 3 festgelegten Risikominimierungstechniken zu implementieren.
5. *Durchführen von Tests:* Das Produkt wird neben Einhaltung funktionaler Anforderungen auch auf Sicherheitsschwachstellen getestet. Dies kann mittels Tools (siehe Abschnitt 2.3), Audits oder Penetrationstests durchgeführt werden. Hierbei sind die in Schritt 1 definierten Anforderungen zu berücksichtigen. Die Art und Tiefe der Tests kann sich je nach Produkt und Sicherheitsniveau unterscheiden. Ergebnisse werden dokumentiert und bei Bedarf werden Schwachstellen korrigiert.
6. *Ausrollen und Incident Response:* Zuletzt wird das Produkt ausgerollt oder veröffentlicht. Hierbei muss eine Integritätsprüfung möglich sein. Zudem müssen Meldewege für Schwachstellen festgelegt und kommuniziert sein, um bei Auftreten von Schwachstellen einen schnellen Patch oder eine Außerbetriebnahme des Produkts zu ermöglichen.

Bei agilen Entwicklungsverfahren werden die Schritte zur Implementierung, Testen und Ausrollen typischerweise mehrfach wiederholt. Potenziell muss auch die Risikoanalyse für neue Produktversionen angepasst werden.

3 Herausforderungen aktueller Zertifizierungsprozesse

Der in Kapitel 2.1 dargestellte Zertifizierungsprozess beinhaltet die Aktivitäten verschiedener Akteure. Diese treten unterschiedlichen Herausforderungen entgegen: Hersteller müssen aufwendig manuell zu erstellende Nachweise erbringen, die anschließend von Evaluatoren manuell geprüft werden. In diesem Kapitel werden die Probleme der aktuellen Nachweisführung beleuchtet. Insbesondere wird diskutiert, wie sich der aktuelle Prozess bei modernen, agilen Softwareentwicklungsmethoden verhält.

3.1 Aufwendige Nachweisführung

Die erste wesentliche Herausforderung ist die sehr aufwendige und zeitintensive Erbringung der geforderten Nachweise. Um diese Aufwände zu reduzieren, verwenden Hersteller einmal erstellte Beschreibungen zu internen Abläufen typischerweise für mehrere Produkte und überprüfen lediglich, ob Änderungen eingetreten sind. Außerdem sind der Umfang und die Erwartungshaltung an die Nachweise oft nicht stark spezifiziert, sondern entstehen erst im Dialog mit den Prüfern. Neue Hersteller müssen sich zunächst an die Gegebenheiten und die Evaluierung anpassen und entsprechende Textbausteine vorbereiten, sodass die Entscheidung eines Herstellers für oder gegen eine Evaluierung oftmals auch eine Frage der Ressourcen und weniger der Sicherheit des Produkts ist.

Zwar gibt es im Rahmen einer Evaluierung durchaus die Möglichkeit, einige Nachweise toolgestützt zu erstellen, allerdings sind auch hierfür signifikante Vorbereitungen notwendig, um die Reports in die geforderten Formate zu überführen. Außerdem umfassen Toolausgaben oftmals nur einen kleinen Teil aller zu erbringenden Nachweise (z.B. die Ausführung von Tests).

Gleichzeitig besteht die Gefahr, dass manuell erstellte Nachweise vom Produkt oder den tatsächlich gelebten Prozessen abweichen. Insbesondere ist eine vollumfängliche Prüfung von Sourcecode nicht unbedingt Bestandteil einer Prüfung, sondern oftmals wird hauptsächlich das zugrundeliegende Konzept geprüft. Dadurch besteht die Möglichkeit, dass trotz der Zertifizierung nicht alle Sicherheitsmaßnahmen korrekt umgesetzt wurden.

3.2 Manuelle Evaluation

Auf der Seite der Prüfeinrichtung hingegen gibt es derzeit so gut wie kein Automatisierungspotenzial, da die Nachweise in jedweder Form erbracht

werden können. Damit muss die Prüfung fast vollständig manuell durchgeführt werden und ist somit zeit- und ressourcenaufwändig.

Neben dem Aufwand für eine manuelle Evaluation sind solche Prozesse auch besonders anfällig für Qualitätsabweichungen, da die Qualität der Bewertung vor allem vom Wissen des Prüfers für das vorliegende Produkt und die darin eingesetzten Technologien, aber auch von der Interpretation der gelieferten Nachweise abhängt. Beispielsweise könnten verschiedene Prüfer einen Textabschnitt anders interpretieren oder sogar bei gleicher Interpretation andere Schlüsse ziehen. Im Extremfall würde ein Produkt durch einen Prüfer als sicher und durch einen anderen Prüfer als unsicher bewertet werden. Um dieses Risiko zu reduzieren, müssen sich in der Regel mehrere Personen abstimmen, was weiteren Aufwand erzeugt und das Risiko dennoch nicht ausschließen kann. Da die Umsetzung von Sicherheitsleistungen und mögliche Schwachstellen oft von den genutzten Programmiersprachen, Frameworks und Bibliotheken abhängt, müssen die Prüfer außerdem detailliertes Expertenwissen der eingesetzten Technologien besitzen. Da sich diese je nach Produkt stark unterscheiden, ist es unwahrscheinlich, dass der Evaluator stets das nötige Wissen besitzt. Deshalb ist davon auszugehen, dass trotz einer gewissenhaften der Prüfung Fehler übersehen werden.

3.3 Schnelle Releasezyklen (DevOps)

Zusätzlich zu den aufwendig zu erbringenden und zu evaluierenden Nachweisen kommen immer schneller werdende Releasezyklen hinzu. Diese sind in der modernen agilen Softwareentwicklung schon lange etabliert. Hierbei wird ein Produkt durch sich ständig wiederholende Phasen stetig (weiter)entwickelt. Ein typischer Prozess ist in Abbildung 1 dargestellt.

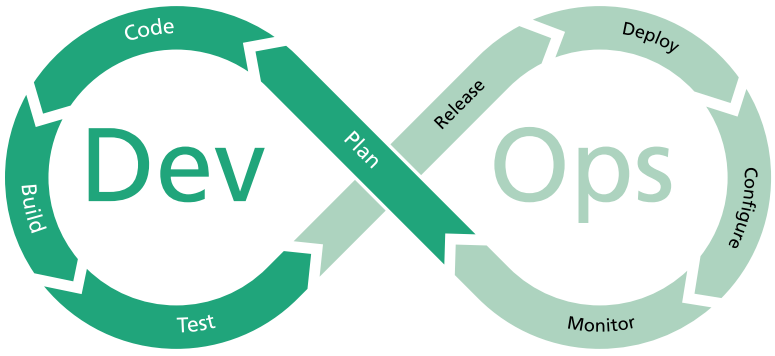


Abbildung 1: Typische Strukturierung eines DevOps-Prozesses und die daraus resultierenden Phasen für die Softwareentwicklung

Für die Entwicklung der Software bedeutet dies, dass eine kontinuierliche Weiterentwicklung des Produkts stattfindet und schneller eine neue Version

veröffentlicht und ausgerollt wird. Für die Zertifizierung bedeutet dies jedoch, dass für jedes Update ein neuer Zertifizierungsprozess angestoßen werden muss, der das Erbringen und Prüfen der Nachweise beinhaltet.

Um trotz umfangreicher manueller Aufwände bei der Zertifizierung zumindest schneller reagieren zu können, ist es möglich, nur die Auswirkungen der Änderungen auf das Sicherheitskonzept zu betrachten. Ob eine Änderung vorliegt und das Konzept angepasst werden muss, bedarf allerdings typischerweise einer menschlichen Entscheidung, da das textuell beschriebene Konzept nicht zwangsläufig eng an den Code gekoppelt ist. Auch hierbei ergibt sich letztlich die Möglichkeit einer Fehleinschätzung.

4 Lösungsansatz

Wir sehen eine gemeinsame Lösung für die zuvor beschriebenen Herausforderungen: mehr Automatisierung. So sieht der Lösungsansatz vor, sowohl zur Erzeugung als auch zur Überprüfung von Nachweisen vermehrt Tools einzusetzen. Als direkte Konsequenz fällt weniger manuelle Arbeit an und der Evaluierungsvorgang wird beschleunigt. Dies ermöglicht damit auch die schnellen Releasezyklen der agilen Softwareentwicklung besser zu unterstützen. So sollten insbesondere bei Updates und Weiterentwicklungen der Großteil an Nachweisen automatisiert erstellt und geprüft werden, um keine Verzögerungen des Releases durch die Evaluation einzuführen.

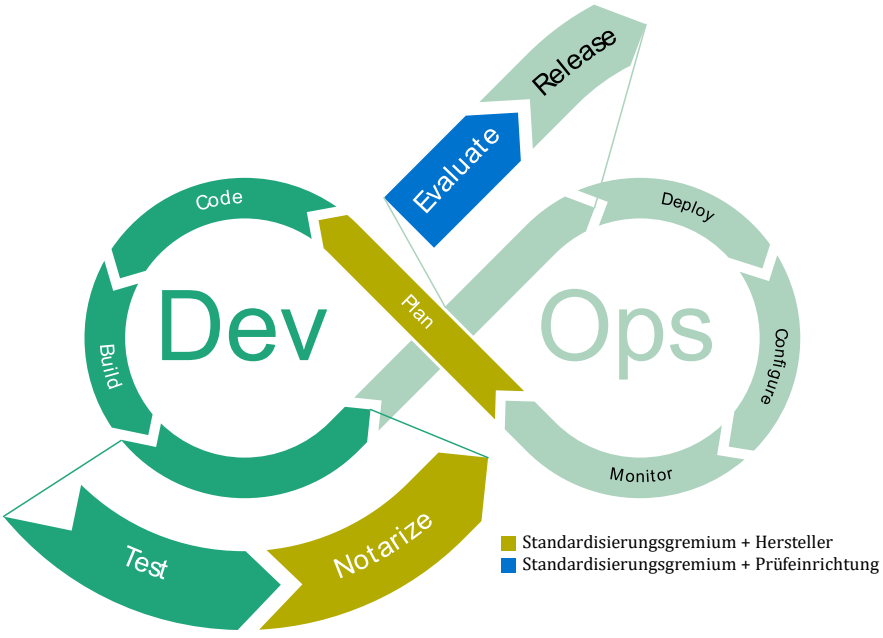


Abbildung 2: Modifizierter DevOps-Prozess zur Notarisierung und Evaluierung

Auf diese Art und Weise können sowohl Hersteller als auch Evaluatoren profitieren: Der Evaluator generiert einen großen Zeitgewinn durch die Automatisierung der bisher manuellen Prüfung der Nachweise bei jedem neuen Release. Der Hersteller andererseits gewinnt zwar ebenfalls diese Vorteile in der Erstellung der Nachweise, hat aber auch unter Umständen einen Mehraufwand durch die Einführung der Werkzeuge, falls diese nicht bereits im Softwareentwicklungsprozess vorgesehen waren. In jedem Fall gewinnt der Hersteller allerdings an Planungssicherheit, da er frühzeitig überprüfen kann, ob die erstellten Nachweise die Prüfkriterien erfüllen. Abbildung 2 stellt den angepassten Entwicklungsprozess mit neuen Zuständigkeiten dar, die in den nachfolgenden Abschnitten erläutert werden.

Um diesen Lösungsansatz in der Praxis umzusetzen, bedarf es allerdings noch einiger Grundlagen, welche in den folgenden Kapiteln erläutert werden. Dies ist einerseits eine genauere Planung der einzusetzenden Werkzeuge, die Bereitstellung von Werkzeugen für Evaluatoren und ein formalisiertes Austauschformat zwischen Herstellern und Evaluatoren. Damit kann die Zertifizierungsgrundlage näher an den erstellten Code und das Produkt rücken, um so vermehrt auf automatisch erbrachte Nachweise zu setzen.

4.1 Planung und Werkzeuge

Der in Abschnitt 3.3 beschriebene DevOps Zyklus und auch die im Rahmen eines SDLC durchgeführten Aktivitäten (vgl. Abschnitt 2.4) liegen derzeit vollständig in der Verantwortung des Herstellers, der diese anschließend dokumentiert und dem Evaluator darlegt, dass diese Aktivitäten ausreichen, um das angestrebte Sicherheitsniveau zu erzielen. Allerdings können einige Aspekte auch vom Standardisierungsgremium übernommen werden.

Soll die Bewertung des Sicherheitsniveaus durch Tools unterstützt oder abgedeckt werden, müssen verbindliche Vorgaben oder Best Practices zum Einsatz von SAST- und DAST-Tools durch Standardisierungsgremien bereitgestellt werden. Dies ist Teil der Phase „Plan“, für welche wir eine Aufteilung der Verantwortung vorschlagen. Produktabhängige Aspekte werden durch den Hersteller, produktunabhängige Aspekte jedoch vom Standardisierungsgremium vorgegeben.

Neben der Auswahl der Tools werden Qualitäts- und Sicherheitsmetriken festgelegt. Auch diese sollten nicht vom Hersteller, sondern vom Standardisierungsgremium vorgegeben werden. Dies umfasst Coverage-Metriken von Tests oder die Interpretation der Ausgaben verschiedener Werkzeuge.

Diese Aufteilung unterstützt einerseits den Hersteller bei der Planung der Rahmenbedingungen für die Produktentwicklung, entlastet bei der Dokumentation der Prozesse und gibt ihm gleichzeitig Planungssicherheit. Zudem kann der Evaluator entlastet werden, da dieser einfacher prüfen kann, ob die Qualitäts- und Sicherheitsmetriken eingehalten wurden.

4.2 Teilautomatisierte interaktive Evaluierung

Im Rahmen einer Zertifizierung muss der Evaluator die Eigenschaften eines (neuen) Produkts untersuchen und bewerten. Hierbei ist in der Regel der Rahmen der Prüfung durch formale Anforderungen oder zu dokumentierende Inhalte vorgegeben, allerdings basieren viele Entscheidungen auf Expertenwissen.

Zur Entlastung der Prüfstellen sollten daher die Automatisierungspotenziale ausgeschöpft werden. Bei allen nicht automatisch entscheidbaren Fragestellungen müssen dem Evaluator alle relevanten Kontextinformationen zugänglich gemacht werden. Dies kann durch ein interaktives System umgesetzt werden, das alle benötigten Informationen bereitstellt.

Eine Aufgabe ist die Prüfung, ob Nachweise konsistent sind und ob alle geforderten Inhalte vorhanden sind. Solche Prüfungen können durch eine Validierung der Nachweise durch ein vorgegebenes Schema automatisiert werden. Nachweise, die diese Prüfung nicht bestehen, müssen nicht weiter betrachtet werden. Der Hersteller kann in diesem Fall schnell nachbessern.

Eine andere Fragestellung ist, ob ein Produkt die Sicherheitseigenschaften (korrekt) implementiert und ob diese im Design berücksichtigt wurden. Die Prüfung auf eine sinnvolle Sicherheitsarchitektur ist weiterhin eine Experteneinschätzung, allerdings ändert sich diese durch Updates selten. Insbesondere mit Blick auf den angepassten DevOps-Zyklus kann eine Betrachtung der Unterschiede zwischen zwei Versionen umgesetzt werden. Gibt es keine Unterschiede an den sicherheitsrelevanten Architekturkomponenten, kann das bereits akzeptierte Konzept weiterhin automatisch akzeptiert werden. Bei den Komponenten handelt es sich um Schnittstellen, voneinander getrennte Systemkomponenten, sowie Funktionen, die Sicherheitseigenschaften umsetzen. Um sicherzustellen, dass die dokumentierten Komponenten mit dem Code übereinstimmen, können Tools eingesetzt werden, die standardisierte Schnittstellenbeschreibungen aus Code, Definitionssprachen oder UML erstellen. Derartige Werkzeuge sind auf dem Markt verfügbar oder leicht zu entwickeln. Wird beispielsweise aus UML automatisch der Code von Schnittstellen erstellt, wie es in der professionellen Softwareentwicklung oft praktiziert wird, kann damit die Übereinstimmung von Code und Dokumentation automatisch sichergestellt werden.

4.3 Standardisierte, strukturierte und formalisierte Datenaustauschformate

Um die Automatisierung auf Seiten der Evaluatoren zu ermöglichen, muss der Nachweis in einem standardisierten und maschinenlesbaren Format vorliegen. Daraus ergeben sich folgende Anforderungen an das Datenaustauschformat zwischen Hersteller und Evaluator:

- Voll *digitalisiert*: Der gesamte Austausch zwischen allen Parteien muss ohne Medienbruch erfolgen.
- *Strukturiert*: Daten müssen auf das Notwendige reduziert in einer definierten Struktur und nicht als Fließtext gespeichert sein.
- *Formalisiert*: Für jedes Feld der Struktur muss genau und abschließend definiert werden, welche Werte erlaubt sind.

In der Praxis dienen Teile der Nachweise zum Verständnis durch Menschen. Hierfür kann ein solches Datenformat nicht erstellt werden, sondern es wird weiterhin Fließtext benötigt. Als Beispiel seien Benutzer- oder Administratorenhandbücher erwähnt; hier ist eine vollständige Formalisierung nicht möglich. Da ohne hinreichend strikte Strukturierung und Formalisierung Automatisierungspotenziale nicht ausschöpfbar sind, muss in der Praxis auf eine Balance hingewirkt werden: grundsätzlich voll digitalisiert, strukturiert und formalisiert, mit Ausnahme von Teilen, die über den Evaluierungsprozess hinaus von Menschen gelesen und verstanden werden müssen.

Außerdem gibt es zusätzliche Anforderungen an das Format, die sich aus dessen Einsatzzweck ergeben. Um die Erstellung und Prüfung der erstellten Nachweise zu vereinfachen und Inkonsistenzen zu vermeiden, dürfen sich Daten innerhalb des Nachweises nicht wiederholen. Duplikate sind durch starke Referenzen zu ersetzen. Zudem müssen Manipulationen an den Daten erkennbar sein, was beispielsweise durch den Einsatz digitaler Signaturen als Integritätsschutz umgesetzt werden kann.

Die Definition eines entsprechenden Standardformats obliegt den Standardisierungsgremien, da diese die Rahmenbedingungen für alle Verfahren schaffen und somit die Automatisierbarkeit maximieren. Die Hersteller setzen das Format für das jeweilige Projekt um und Prüfstellen implementieren die automatischen Prüfungen. Als Beispiel für ein Format sei OSCAL [8] genannt.

Auch die Bereitstellung von Anforderungen an Produktklassen, z.B. in den *Protection Profiles* der Common Criteria, muss überarbeitet werden. Dabei sind die in Abschnitt 2.2 vorgestellten Maßstäbe umzusetzen. Bestenfalls bieten sie Regeln für Compliance-Prüftools wie Codyze [9] [10] oder CogniCrypt [11] an, sodass diese die Umsetzung der Anforderungen prüfen können. Dies

kann Experteneinschätzungen im Prozess reduzieren und Ansätze des Model-Checkings integrieren.

4.4 Wandel der Zertifizierungsgrundlage

Insgesamt soll neben einer Reduzierung von manuellen Aufwänden und Unklarheiten auf Seiten der Hersteller und Prüfer auch ein stärkerer Fokus auf die tatsächliche Implementierung gelegt werden, um Inkonsistenzen schneller zu erkennen. Zudem können dort häufige Fehler relativ gut automatisiert erkannt werden. Dies geschieht dadurch, dass Tool-basierte Nachweise direkt am Source Code, Build-Artefakten oder Dependencies erbracht werden und daher eine starke Kopplung zum tatsächlichen Produkt aufweisen.

In den Fällen, in denen keine Kopplung zum Sourcecode hergestellt werden kann, wie z.B. bei der Bewertung der Softwarearchitektur, sollten standardisierte Formate (wie UML oder Schnittstellenbeschreibungen) genutzt werden, um die Implementierung darzustellen. Daraus können während der Entwicklung auch Teile des Produkts (z.B. Schnittstellen) erstellt werden. All das verfolgt das Ziel, dass der Evaluator auf denselben Daten arbeitet, die der Entwickler zur Erstellung des Produkts genutzt hat.

5 Diskussion des Lösungsansatzes

5.1 Automatisierungspotenziale

Durch den Einsatz von SAST- und DAST-Tools können die Umsetzung von Sicherheitseigenschaften sowie die Existenz von Programmierfehlern im Code automatisch geprüft werden. Eine Prüfung auf Einhaltung formaler Aspekte kann durch ein standardisiertes maschinenlesbares Format automatisch erfolgen.

Um eine weitere Automatisierung zu ermöglichen, kann eine verbindliche Bewertungsmatrix für Tools und Testergebnisse eingeführt werden. Damit kann potenziell automatisch entschieden werden, ob die eingesetzten Tools und Ergebnisse für das vorliegende Produkt ausreichen, um das gewünschte Sicherheitsniveau zu erreichen. Die jeweiligen Prozesse und Testergebnisse müssen dann nicht mehr im Detail manuell nachvollzogen und bewertet werden. Dadurch entstehen auch weniger Interpretationsspielräume, die zu Unklarheiten zwischen Prüfern und Herstellern führen können und zusätzlichen Overhead für eine Evaluierung darstellen.

Stehen maschinell verarbeitbare spezifische Anforderungen an einzelne Produkttypen bereit, kann eine tiefgehende und derzeit zeitaufwendige Analyse automatisiert werden. Somit sind die Wahl der Compliance-Prüftools und die

Gestaltung zugehöriger Regeln ein zentraler Aspekt einer zukünftigen automatischen Zertifizierung und Aufgabe der Standardisierungsgremien.

5.2 Interaktive Evaluation des Zertifizierungsgegenstands

Wie in Abschnitt 4.2 beschrieben, sollte dem Evaluator ein unterstützendes Werkzeug bereitgestellt werden. Analog zu bereits in der Entwicklung etablierten Code-Review-Plattformen kann aufseiten der Evaluatoren eine Plattform zu kollaborativen Evaluierung geschaffen werden. Diese soll die erhaltenen Nachweise in einer bewertbaren, menschenlesbaren Form darstellen. Die Unterstützung kann dabei über das aktuelle Maß hinausgehen, so kann beispielsweise bei Elementen, die Stellen im Quelltext referenzieren, dieser Kontext direkt angezeigt werden.

5.3 Vergleich des Sicherheitsniveaus

Durch die Veränderung der Nachweisführung für eine Zertifizierung sollte das Sicherheitsniveau im Vergleich zu den aktuell gelebten Prozessen mindestens gleich hoch bleiben. Die Auswahl geeigneter Tools und die Interpretation der Ergebnisse sind maßgeblich dafür verantwortlich, ein gleichbleibendes Sicherheitsniveau zu erzielen. Da diese sich je nach angestrebter Zertifizierung stark unterscheiden können, ist an dieser Stelle keine allgemeingültige Aussage möglich. Beispielsweise unterscheidet die Common Criteria selbst zwischen sieben sogenannten Evaluation Assurance Level (EAL), die von funktionalem Testen (EAL1) bis hin zu formal verifiziertem Design und Testen (EAL7) reichen. Dementsprechend müssen hier, je nach angestrebtem EAL, verschiedene Tools zum Einsatz können.

Zwar gibt es Tools zur formalen Verifikation von Software, Protokollen oder Designbeschreibungen, allerdings erfordern diese sehr detailliertes Expertenwissen, um die zu prüfenden Anforderungen und Modelle richtig zu beschreiben, und sie sind in der Praxis noch wenig verbreitet. Zudem weisen sie oftmals gravierende Schwächen in der Skalierbarkeit auf größere Codebasen auf. Dies wirkt sich jedoch nur auf Produkte mit EAL 5 oder höher aus, was im Rahmen typischer Entwicklungsprozesse ohnehin kaum erzielbar ist [12].

Für alle Produkte bis EAL 4 entscheidet hauptsächlich die Qualität der Tool-Ergebnisse, ob das Sicherheitsniveau vergleichbar zu den Einschätzungen eines Evaluators ist. Ein strukturiertes Testen kann beispielsweise durch eine hohe Line- und Branch-Coverage von Unit-Tests nachgewiesen werden, die durch nahezu jedes Testing-Framework ermittelt werden können und auch automatisch abgeprüft werden können. Ist es hierbei gewünscht, dass spezielle sicherheitskritische Bereiche eine besonders hohe Coverage aufweisen,

müssen diese Code-Bereiche ermittelt und anschließend die jeweilige Coverage-Metrik gegen einen Richtwert verglichen werden. Dies hat im Vergleich zum bisherigen Vorgehen den Vorteil, dass die Bewertung direkt am Code gemessen werden kann und nicht an abstrakten Konzepten wie Schnittstellen oder Sicherheitsfunktionen.

Um zu überprüfen, dass keine einfach identifizierbaren Schwachstellen vorhanden sind, können typische SAST-Tools zum Einsatz kommen; je höher das EAL, umso niedriger muss dabei die Anzahl an falsch negativen Ergebnissen sein. Unterstützend zu statischen Verfahren kann mittels Fuzzing aufgezeigt werden, ob es Indikatoren für leicht ausnutzbare Schwachstellen gibt.

Ebenso kann die Qualität der Nachweise besser verglichen werden, da SAST-Tools bei identischer Konfiguration in der Regel weniger Streuung aufweisen als menschliche Einschätzungen. Damit kann eine gleichbleibende Qualität der Prüfung sichergestellt werden.

Insbesondere der Einsatz von Compliance-Tools kann im Gegensatz zum aktuellen Vorgehen das Sicherheitsniveau weiter erhöhen. Hierbei kann, einen geeigneten Regelsatz für den Produkttyp und die darin zu implementierenden Sicherheitsfunktionalitäten vorausgesetzt, automatisch geprüft werden, dass der vorliegende Code diese Funktionalität auch tatsächlich und korrekt umsetzt und somit die Anforderungen erfüllt. Dies verlagert die Komplexität der Prüfung von der Evaluation zu Vorarbeiten, was aber mit Blick auf agile Verfahren und häufig durchgeführte Prüfungen langfristig eine signifikante Zeitersparnis ermöglicht und auch die kontinuierliche Korrektheit der Umsetzung prüfbar macht. Dies ist bei manuellem Vorgehen nicht möglich.

Zusammenfassend kann durch eine geeignete Wahl von Tools das Sicherheitsniveau also langfristig sogar erhöht werden, während die Aufwände reduziert werden.

6 Fazit & Ausblick

Die aktuell gelebten Prozesse in der Zertifizierung von Software sind aufgrund starrer textbasierter Nachweise nur schwer mit den gelebten agilen Entwicklungsmethoden und DevOps vereinbar. Allerdings kann durch den Einsatz geeigneter SAST-, DAST- und Fuzzing-Tools ein gleichbleibendes oder sogar gesteigertes Sicherheitsniveau erzielt und eine schnellere Zertifizierung ermöglicht werden. Hierzu müssen die Zertifizierungsprozesse angepasst werden, um näher an die tatsächlich gelebte Praxis in der Softwareentwicklung zu kommen und dort etablierte Tool-Ausgaben direkt als Nachweise zu etablieren. Gleichzeitig müssen die Anforderungen an Produkte ge-

ändert werden, so dass diese konkret im Code geprüft werden können; idealerweise in Form von Regeln für Compliance-Prüftools wie Codyze oder CogniCrypt. Wenn zudem die einzuliefernden Nachweise strikt spezifiziert und maschinenlesbar sind, entsteht das Potenzial, größtenteils automatisiert Entscheidungen über Produkte mit niedrigem bis mittlerem Sicherheitsniveau zu treffen. In der Kombination dieser drei Maßnahmen (maschinenlesbare Austauschformate, Anpassungen der geforderten Nachweise, Einsatz von Tools) kann der Zertifizierungsprozess erheblich automatisiert werden und somit die Chance erhalten mit dem hohen Entwicklungs- und Updatetempo der agilen Softwareentwicklung Schritt zu halten.

Literaturhinweise

[1] Common Criteria, Common Criteria for Information Technology Security Evaluation, 2017.

[2] D. S. Hermann, Using the Common Criteria for IT security evaluation, CRC Press, 2002.

[3] International Council on Systems Engineering, Guide to Writing Requirements, INCOSE Publications Office, 2023.

[4] Microsoft, „Security Development Lifecycle - SDL Process Guidance Version 5.2," 23.05.2012. https://download.microsoft.com/download/3/4/3/343BAFCB-C685-4A70-9639-FF76BCBB609C/Microsoft%20SDL_Version%205.2.docx. [Zugriff: 23.01.2023].

[5] Cisco, Cisco Secure Development Lifecycle, 2021.

[6] OWASP, „Software Assurance Maturity Model (SAMM) Version 2," [Online]. Available: https://owaspsamm.org/model/. [Zugriff: 11.01.2023].

[7] National Institute of Standards and Technology, „NIST Special Publication 800-218 - Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities," Februar 2022. https://doi.org/10.6028/NIST.SP.800-218. [Zugriff: 11.01.2024].

[8] National Institute of Standards and Technology, „OSCAL: the Open Security Controls Assessment Language," 08.11.2023. https://pages.nist.gov/OSCAL/. [Zugriff: 07.02.2024].

[9] Fraunhofer AISEC, „Codyze: Automated Code Compliance," https://www.codyze.io/. [Zugriff: 20.02.2024].

[10] C. Banse, F. Wendland und K. Weiss, „Automatisierte Compliance-Prüfung von Software-Artefakten," in 17. Deutscher IT-Sicherheitskongress des BSI, 2021.

[11] „CogniCrypt," https://github.com/eclipse-cognicrypt/CogniCrypt [Zugriff: 20.02.2024].

[12] Bundesamt für Sicherheit in der Informationstechnik, „ Evaluation Assurance Level (EAL),“ https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Zertifizierung-und-Anerkennung/Zertifizierung-von-Produkten/Zertifizierung-nach-CC/IT-Sicherheitskriterien/CommonCriteria_v31/eal_stufe.html. [Zugriff: 08.02.2024].

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Who You Gonna Call? Ein taxonomischer Vergleich von Unterstützungsangeboten im Bereich der Incident Response

Dr. Dirk Achenbach[1], Martin Dukek[1], Marc Nemes[1]

Im Verlauf der letzten zwei Jahrzehnte hat sich eine ganze Reihe von öffentlichen Unterstützungsangeboten für Privatpersonen und Firmen entwickelt, die bei einem IT-Sicherheitsvorfall („Cyberangriff") Unterstützung anbieten. Im Gegensatz zur damaligen Situation stellt sich heute die Frage, welches Unterstützungsangebot welche Zielgruppe in welcher Situation anspricht. Wir haben eine Taxonomie entwickelt, die einen systematischen Vergleich der Angebote ermöglicht. Die von uns entwickelte Taxonomie hilft, einen Überblick über die bestehenden Angebote zu erhalten, gegebenenfalls Lücken zu identifizieren und Angebote weiterzuentwickeln.

Stichworte: Hilfe bei Angriffen, Incident Response, IT-Sicherheitsvorfall, KMU, Kontaktstellen, Taxonomie, Unterstützungsangebote

1 Einleitung

Der digitale Wandel hat in der Welt der kleinen und mittleren Unternehmen (KMU) zu bemerkenswerten Fortschritten und Effizienzgewinnen geführt. Durch die zunehmende Digitalisierung ihrer Geschäftsprozesse können KMU ihre Wettbewerbsfähigkeit steigern und neue Märkte erschließen. Diese positive Entwicklung bringt jedoch auch neue Abhängigkeiten mit sich, insbesondere im Hinblick auf Informationstechnologien. Die Studie "Wirtschaftsschutz 2023" des Branchenverbands Bitkom offenbart die Kehrseite dieser Medaille: Mehr als die Hälfte der über 1000 befragten Unternehmen gaben an, dass sie sich durch Cyberattacken in ihrer Existenz bedroht fühlen [1]. Diese Entwicklung unterstreicht nicht nur die zunehmende Abhängigkeit von der Informationstechnologie innerhalb dieser Unternehmensklasse, sondern auch ihre ansteigende Vulnerabilität für Cyberangriffe.

Doch nicht nur KMU stehen vor diesen Herausforderungen; ähnlichen Gefahren müssen sich auch Selbstständige, Privatpersonen, öffentliche Einrichtungen und Behörden stellen. Angesichts dieser Bedrohungslage haben sowohl der privatwirtschaftliche Sektor als auch öffentliche Institutionen eine Vielzahl an Lösungsangeboten entwickelt, um den steigenden Bedarf an IT-Sicherheit zu adressieren. Dabei ist insbesondere im Bereich der Incident Response eine Zunahme an spezialisierten Dienstleistern und Kontaktstellen

[1] FZI Forschungszentrum Informatik, Karlsruhe, Transferstelle Cybersicherheit im Mittelstand

zu verzeichnen. Trotz der zunehmenden Verfügbarkeit von Ressourcen und Diensten bleibt die Frage, inwieweit die bestehenden Angebote das breite Spektrum an Opfern von Cyberangriffen und deren Bedürfnisse abdecken.

In diesem Artikel geben wir einen umfassenden und systematischen Überblick über die gegenwärtigen Angebote im Bereich der Incident Response. Die vorgestellte Taxonomie ermöglicht es, Unterstützungsangebote in diesem Bereich als vergleichbare Acht-Tupel zu kodieren, wodurch eine Vergleichbarkeit der Angebote ermöglicht wird.

1.1 Vergleichbare Übersichten und unser Beitrag

Im Bereich der Computer Emergency Response Teams (CERTs) gibt es eine Tradition der Vernetzung und vertrauensvollen Kooperation. Verbünde wie der CERT-Verbund pflegen eine Liste ihrer Mitglieder [2]. Auch das Netzwerk "Trusted Introducer" pflegt eine Liste europäischer Cybersicherheitsteams [3]. Diese Listen tragen vordergründig dem Gedanken der Vernetzung von CERTs Rechnung, die von ihrer (geschlossenen) Zielgruppe über eigene Wege kontaktiert werden können.

Bei einem Advanced Persistent Threat (APT) handelt es sich um andauernde Bedrohungen, die von Angreifenden mit fortgeschrittenen Fähigkeiten durchgeführt werden. Das BSI hat ein Verfahren entwickelt, um Dienstleister zu identifizieren, die in solchen Situationen qualifiziert unterstützen können, und listet diese online. Der Journalist Jürgen Berke pflegt auf seiner persönlichen Webseite eine Liste mit Notrufnummern bei Cyberangriffen, die der Vermittlung von Soforthilfe dienen soll [4]. Er unterscheidet dabei zwischen den Sicherheitsbehörden des Bundes, den Polizeibehörden der Bundesländer, kommerziellen Anbietern und privaten Cyberwehren.

Unsere Taxonomie liefert eine Systematik, die einen Vergleich verschiedener Hilfsangebote ermöglicht. Darüber hinaus zeigt sie das Feld bestehender Angebote auf und hilft, Lücken zu identifizieren. Anbietern einzelner Hilfsangebote zeigt sie auf, welche Möglichkeiten der Weiterentwicklung sich ihnen bieten – auch im Vergleich mit anderen Angeboten.

2 Taxonomie von Hilfsangeboten

In diesem Abschnitt kategorisieren wir die verschiedenen Arten von Hilfsangeboten systematisch. Zunächst identifizieren wir die Schlüsselmerkmale, die diese Angebote unterscheiden. Anschließend verwenden wir diese Merkmale zur Erstellung einer Taxonomie, um eine klare Übersicht der vorhandenen Angebote zu bieten.

2.1 Rollen

Im Kontext von Hilfsangeboten bei IT-Sicherheitsvorfällen ist die Unterscheidung und Gegenüberstellung von Hilfesuchenden und Hilfeleistenden sowie von Kund*innen und Anbieter*innen notwendig, um die zugrundeliegenden Mechanismen und Beziehungen innerhalb dieses Feldes zu erfassen.

Hilfesuchende und Hilfeleistende

Den Ausgangspunkt bilden die Hilfesuchenden – Individuen, Unternehmen oder öffentliche Einrichtungen, die aufgrund eines IT-Sicherheitsvorfalls externe Unterstützung benötigen. Ihnen gegenüber stehen die Hilfeleistenden, ein breites Spektrum an Fachkräften und Organisationen, die mit ihrem Know-how und ihren Ressourcen auf die Bewältigung der Sicherheitsprobleme abzielen. Diese Gegenüberstellung unterstreicht die initiale Beziehung zwischen dem Bedarf an Unterstützung und dem Angebot von Lösungen.

Kund*innen und Anbieter*innen

Sobald eine formelle Vereinbarung zur Inanspruchnahme einer Dienstleistung getroffen wird, transformieren Hilfesuchende in die Rolle der Kund*innen. Auf der anderen Seite stehen die Anbieter*innen, ehemals Hilfeleistende, die nun im Rahmen dieser Vertragsbeziehung spezifische Dienstleistungen erbringen. Diese Rollen betonen die vertragliche Beziehung und den Austausch von Dienstleistungen gegen Entgelt.

In bestimmten Konstellationen, wie der Interaktion zwischen Privatpersonen und der Polizei zur Strafverfolgung, bleibt das Verhältnis zwischen Hilfesuchenden und Hilfeleistenden erhalten, ohne in eine Kund*innen-Anbieter*innen-Dynamik überzugehen. Hierbei bleibt die Entscheidung über Art und Umfang der Hilfeleistung in den Händen der Hilfeleistenden. Im Kontrast dazu übernehmen in der Kund*innen-Anbieter*innen-Beziehung die Hilfesuchenden die Rolle der Entscheidungsträger*innen.

2.2 Merkmale

Das Ziel dieses Abschnitts ist, die verschiedenen Arten von Hilfsangeboten bei IT-Sicherheitsvorfällen durch die Identifikation und systematische Kategorisierung ihrer Schlüsselmerkmale zu differenzieren. Anhand dieser Merkmale wird eine Taxonomie entwickelt, die eine systematische Klassifikation der verfügbaren Angebote ermöglicht.

I. **Umfang der Unterstützung:** Der Umfang der Unterstützung bezieht sich auf das Ausmaß und die Art der Hilfe, die in Reaktion auf spezifische Bedürfnisse oder Anforderungen bereitgestellt wird. Dieser Umfang kann variieren, abhängig von der Natur des Vorfalls, der Dringlichkeit der Situation, den verfügbaren Ressourcen und der Expertise

der Hilfeleistenden. Wir unterscheiden hierbei zwischen Ersthilfe und Hilfedienstleistungen.

a. **Ersthilfe**: Hierbei handelt es sich um Sofortmaßnahmen zur Bewältigung akuter Sicherheitsvorfälle, wobei die Hilfeleistenden im Regelfall nach einem standardisierten Prozess vorgehen. Diese Phase hat oft eine anamnestische Funktion, da die Hilfesuchenden meist nicht die erforderliche Kompetenz besitzen, um die Situation eigenständig zu analysieren und die nächsten Schritte einzuleiten. Ziel ist es, unmittelbare Schäden zu minimieren, gegebenenfalls Strafverfolgungsmaßnahmen einzuleiten und die Grundlage für weiterführende Maßnahmen zu schaffen.
b. **Hilfedienstleistung**: Hilfedienstleistungen erweitern das Spektrum der Ersthilfe um Angebote wie detaillierte Systemanalysen und weiterführende Maßnahmen, die eine individuell zugeschnittene Unterstützung basierend auf den spezifischen Bedürfnissen und Anforderungen der Hilfesuchenden bieten. Sie sind durch ein klassisches Kund*innenverhältnis gekennzeichnet, welches den Hilfesuchenden mehr Kontrolle über die Ausgestaltung der Dienstleistung gibt. Um diese Kontrolle effektiv ausüben zu können, benötigen die Kund*innen jedoch eine klare Vorstellung von den gewünschten Ergebnissen, die beispielsweise durch eine Ersthilfe präzisiert werden können.

II. **Modell**: Das Modell der Unterstützung legt den Fokus auf das "Wer" hinter der Hilfeleistung und differenziert zwischen Vermittlung, Einbindung Dritter und direkter Unterstützung.

a. **Vermittlung**: Die Unterstützung wird durch Dritte geleistet, wobei die hilfesuchende Partei direkt mit diesen Dritten Verträge abschließt. Die Vermittler agieren als Bindeglied und sind nicht direkt an der Leistungserbringung beteiligt.
b. **Einbindung Dritter**: Beschreibt die Konstellation, bei der die Hilfe leistende Organisation externe Dritte einbindet, um die Unterstützung zu erbringen. Diese Dritten spielen eine wesentliche Rolle im Rahmen der Unterstützung, agieren jedoch unter der Federführung der Hauptorganisation und werden nicht als separate, verantwortliche Parteien wahrgenommen. Obwohl sie essenzielle Dienstleistungen oder Fachkenntnisse beisteuern, bleibt die Gesamtverantwortung für die Qualität und Effektivität der erbrachten Unterstützung bei der initiierenden Organisation.

   c. **Direkte Unterstützung**: Die Hilfedienstleistung wird hier komplett intern von der Organisation, die die Hilfe anbietet, übernommen. Die Unterstützung erfolgt durch eigene Fachkräfte, was eine direkte Kontrolle über die Hilfeleistung ermöglicht.

III. **Perspektive**: Die Perspektive, aus der Unterstützung bei einem IT-Sicherheitsvorfall angeboten wird, ist vielschichtig und abhängig von der Art des Vorfalls sowie den spezifischen Anforderungen, die sich daraus ergeben. Bei einem Vorfall, der personenbezogene Daten betrifft, sind beispielsweise sowohl technische als auch datenschutzrechtliche Aspekte von Bedeutung. Die Unterstützungsangebote können daher aus verschiedenen Blickwinkeln notwendig und hilfreich sein. Um die Vielfalt der Ansätze zu verdeutlichen, kategorisieren wir die Perspektiven eines Unterstützungsangebots in die folgenden Schlüsselbereiche:
   a. **Technisch**: Fokussiert auf die unmittelbaren technischen Herausforderungen, wie die Sicherung von Netzwerken, die Analyse von Malware oder die Wiederherstellung von Systemen.
   b. **Organisatorisch**: Bezieht sich auf die Strukturen, die für das Funktionieren des Unternehmens notwendig sind.
   c. **Juristisch**: Umfasst rechtliche Aspekte, die bei der Bewältigung von Sicherheitsvorfällen eine Rolle spielen, einschließlich Datenschutzgesetzen, Compliance-Anforderungen und der Beratung in Rechtsfragen.
   d. **Kommunikativ**: Beinhaltet die Gestaltung der Kommunikationsstrategien gegenüber internen und externen Stakeholdern, um Transparenz zu wahren und Vertrauen zu sichern.
   e. **Rechtsstaatlich**: Fokussiert sich vorrangig auf Aspekte der Strafverfolgung im Kontext von IT-Sicherheitsvorfällen.

IV. **Kostenstruktur**: Der Kostenumfang ist ein weiteres entscheidendes Merkmal, das bestimmt, ob für die Hilfeleistung Gebühren anfallen.
   a. **Kostenpflichtige Angebote**: Hilfeangebote, bei denen den Hilfesuchenden direkte Kosten entstehen. Dies kann in Form von einmaligen Gebühren, monatlichen Abonnements oder anderen Zahlungsmodellen erfolgen. Die Höhe und Art der Gebühren variieren je nach Umfang und Art der erbrachten Dienstleistung.
   b. **Kostenfreie Angebote**: Werden in der Regel durch staatliche Finanzierung oder durch gemeinnützige Organisationen bereitgestellt, sodass für die Hilfesuchenden keine unmittelbaren Kosten

anfallen. Diese Angebote sind oft darauf ausgelegt, breite Bevölkerungsschichten oder spezifische Zielgruppen wie KMU zu unterstützen.

V. **Verfügbarkeit**: Die Verfügbarkeit von Unterstützungsangeboten bei IT-Sicherheitsvorfällen definiert, wann und wie schnell Hilfe für die Hilfesuchenden verfügbar ist.
   a. **24/7-Unterstützung**: Dieses Angebot gewährleistet, dass Hilfesuchende zu jeder Zeit, also rund um die Uhr und an jedem Tag, direkte Unterstützung erhalten können. Dies beinhaltet nicht nur die Möglichkeit, Kontakt aufzunehmen, sondern auch die Bereitstellung konkreter Hilfeleistungen, wie beispielsweise Sofortmaßnahmen oder Beratung durch Fachpersonal am Telefon.
   b. **24/7-Erreichbarkeit**: In diesem Modell ist zwar eine ständige Kontaktmöglichkeit gegeben, jedoch werden die Anfragen selbst nur innerhalb der regulären Geschäftszeiten bearbeitet. Das bedeutet, dass Hilfesuchende ihre Anliegen jederzeit einreichen können, die eigentliche Unterstützung oder Rückmeldung aber erst später erfolgt.
   c. **Geschäftszeiten**: Die Unterstützung ist hierbei ausschließlich während definierter Geschäftszeiten verfügbar. Außerhalb dieser Geschäftszeiten ist keine Kontaktaufnahme möglich.

VI. **Regionale Reichweite**: Die regionale Reichweite von Unterstützungsangeboten bei IT-Sicherheitsvorfällen wird auf zwei Hauptkategorien begrenzt, um eine klare Vergleichbarkeit zu ermöglichen: regional und bundesweit.
   a. **Regional**: Diese Unterstützungsangebote sind auf spezifische geografische Gebiete beschränkt. Sie zielen darauf ab, Hilfesuchenden innerhalb einer bestimmten Region, wie einem Bundesland, einer Stadt oder einem Landkreis, gezielte Unterstützung zu bieten.
   b. **Bundesweit**: Bundesweite Unterstützungsangebote sind in ganz Deutschland verfügbar und richten sich an Hilfesuchende unabhängig von ihrem Standort.

VII. **Zielgruppe**: Die Angebote können für unterschiedliche Zielgruppen wie Privatpersonen, Unternehmen oder öffentliche Einrichtungen konzipiert werden. Aufgrund der Vielzahl der Möglichkeiten wird an dieser Stelle keine weitere Einschränkung vorgenommen.

VIII. **Kontaktmöglichkeiten**: Effektive Unterstützung erfordert zugängliche Kommunikationswege. Auch hier wird aufgrund der Vielfalt der Angebote keine Kategorisierung vorgenommen.

Ein *Unterstützungsangebot* ist eine Menge von 8-Tupeln (Umfang, Modell, Perspektive, Kostenstruktur, Verfügbarkeit, Regionale Reichweite, Zielgruppe, Kontaktmöglichkeiten). Jedes einzelne dieser Tupel bildet eine konkret abrufbare Unterstützungsleistung ab. Je nach Bedarf können die Unterstützungsleistungen eines/einer Anbieter*in auf mehrere Tupel aufgeschlüsselt werden. Auf diese Weise kann unsere Taxonomie auch komplexe Angebotsstrukturen abbilden.

Unterstützungsangebot := {
(Umfang, Modell, Perspektive, Kostenstruktur, Verfügbarkeit, regionale Reichweite, Zielgruppe, Kontaktmöglichkeiten) |
Umfang ∈{Ersthilfe, Hilfedienstleistung},
Modell ∈{Vermittlung, Einbindung Dritter, direkte Unterstützung},
Perspektive ∈{technisch, organisatorisch, juristisch, kommunikativ, rechtsstaatlich},
Kostenstruktur ∈{kostenpflichtig, kostenfrei},
Verfügbarkeit ∈{24/7-Unterstützung, 24/7-Erreichbarkeit, Geschäftszeiten},
regionale Reichweite ∈{regional, bundesweit},
Zielgruppe (Freitext),
Kontaktmöglichkeiten (Freitext)
}

2.3 Kategorien

Unter Berücksichtigung der vorgestellten Merkmale lassen sich verschiedene Kategorien für Hilfsangebote bilden. Besondere Relevanz für die Kategorisierung haben die Faktoren Anbieterrolle und Kostenstruktur, da sie oft die Schnittstelle zwischen Dienstleister und Klienten darstellen und somit einen wesentlichen Einfluss auf die Inanspruchnahme der Dienstleistung haben. Nach der Bewertung der Angebote auf der Grundlage unserer Recherche bilden sich drei Kategorien heraus: Öffentlich geförderte Vermittlungsstellen, öffentlich geförderte Direkthilfe und kommerzielle Direkthilfe.

2.3.1 Kategorie Öffentlich Geförderte Vermittlungsstelle

Öffentlich geförderte Vermittlungsstellen fungieren als Koordinationszentren, die den Kontakt zwischen den Betroffenen und qualifizierten Hilfeleistern herstellen. Sie sind häufig durch eine ständig erreichbare (24/7) Kommunikationsschnittstelle, beispielsweise eine Hotline, gekennzeichnet. Während Ersthilfe-Maßnahmen für gewöhnlich unentgeltlich sind, werden Hilfedienstleistungen meist kostenpflichtig und über externe, kommerzielle Dienstleister abgewickelt. Der Hauptvorteil dieser Vermittlungsstellen be-

steht in ihrer Rolle als leicht zugängliche Erstkontaktstelle, die eine geordnete Weiterleitung an spezialisierte Dienstleister ermöglicht. Ein möglicher Nachteil könnte die redundante Erfassung von Daten sein, wenn sowohl die Vermittlungsstelle als auch der endgültige Dienstleister ähnliche Informationen benötigen. Durch die Etablierung dieser Vermittlungsstellen strebt der Staat an, eine leicht zugängliche Erstkontaktstelle bereitzustellen, die den Betroffenen einen strukturierten und effizienten Zugang zu spezialisierten Hilfeangeboten ermöglicht.

2.3.2 Kategorie Öffentlich Gefördete Direkthilfe

In dieser Kategorie finden sich öffentlich geförderte Einrichtungen, die direkte Incident-Response-Unterstützung anbieten. Zu den angebotenen Leistungen zählt in der Regel eine kostenlose Ersthilfe. Diese Dienste werden von hausinternen IT-Experten bereitgestellt. Im Hinblick auf die Zielgruppe konzentrieren sich diese Einrichtungen vorrangig auf kleine und mittlere Unternehmen und beschränken ihren Aktionsradius häufig auf bestimmte Bundesländer. Öffentlich geförderte Direkthilfe dient sowohl als wirtschaftspolitische als auch als sicherheitspolitische Maßnahme. Einerseits zielt sie darauf ab, die Resilienz von KMU gegenüber IT-Sicherheitsvorfällen zu stärken und damit die wirtschaftliche Stabilität zu fördern. Andererseits dient sie dem Schutz der öffentlichen Sicherheit, indem durch effektive Incident-Response-Maßnahmen das Risiko einer Eskalation von IT-Sicherheitsvorfällen minimiert wird.

2.3.3 Kategorie Kommerzielle Direkthilfe

Diese Kategorie umfasst IT-Dienstleistungsunternehmen aus dem privatwirtschaftlichen Sektor, die sich entweder vollständig auf Incident Response spezialisiert haben oder diese Dienste als Teil ihres Portfolios anbieten. Die Inanspruchnahme dieser Dienste ist immer kostenpflichtig, wobei die Abrechnung direkt oder über Dritte, beispielsweise Versicherungen, erfolgen kann. Teilweise bieten auch die Versicherungen selbst eine Unterstützung für ihre Kunden an, um den entstandenen Schaden frühzeitig einzugrenzen und Kosten zu senken. Die Dienstleistung umfasst ausführliche Analysen und Maßnahmen, die den Rahmen der vorhergehenden Kategorien oftmals überschreiten. Kostenfreie Ersthilfe ist eher die Ausnahme und wird meist als Marketingstrategie für umfassendere, kostenpflichtige Dienste eingesetzt. Der geografische Wirkungsbereich dieser Dienstleister ist normalerweise regional begrenzt, kann jedoch durch Niederlassungen oder in Ausnahmefällen auf nationaler Ebene erweitert sein. Unternehmen bilden die Hauptzielgruppe, wobei kleinere, kosteneffiziente Dienstleister auch von Privatpersonen in Anspruch genommen werden können.

3 Übersicht existierender Initiativen

Dieser Abschnitt gibt einen Überblick über verschiedene, öffentlich geförderte Cybersicherheitsinitiativen in Deutschland. Wir beschreiben die Initiativen und überführen sie in die Form eines Unterstützungsangebots gemäß unserer Taxonomie. Die gesammelten Unterstützungsangebote bilden eine Momentaufnahme, welche sich in der Zukunft geringfügig verändern kann, beispielsweise wenn ein Hilfsangebot seine Hotline von üblichen Bürozeiten zu einer 24/7-Hotline ausweitet. Im Rahmen unserer Recherche konnten wir einige Werte nicht konkret bestimmen. Wir haben in solchen Fällen Werte ausgelassen und stattdessen einen Stern* gesetzt oder alternativ den Wert angegeben, auf den alle Hinweise deuteten, diesen jedoch zusätzlich mit einem Stern* versehen.

3.1 Chronologische Übersicht

Bevor wir die Initiativen im Detail aufführen, geben wir eine Darstellung in chronologischer Form. Abbildung 1 enthält eine Übersicht über alle Initiativen, für welche wir den Gründungszeitpunkt eindeutig ermitteln konnten.

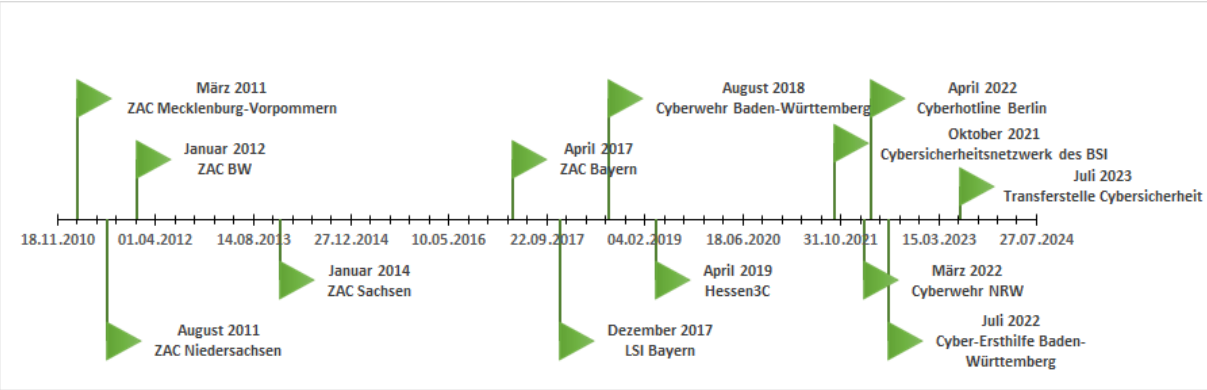


Abbildung 1: Chronologische Übersicht existierender Initiativen

Auffällig ist, dass die Zentralen Ansprechstellen Cybercrime die ersten umfangreichen Anlaufstellen für Betroffene von Cyberangriffen waren, in den letzten Jahren jedoch in verschiedenen Bundesländern vermehrt weitere/spezifische Anlaufstellen etabliert worden sind.

3.2 3.2 Öffentlich geförderte Vermittlungsstellen

Im Folgenden stellen wir die identifizierten öffentlich geförderten Vermittlungsstellen vor.

3.2.1 Cyberwehr Baden-Württemberg [5]

Die Cyberwehr Baden-Württemberg ist ein Projekt eines Konsortiums um das FZI Forschungszentrum Informatik in Karlsruhe [6]. Bis Juni 2022 wurde

die Cyberwehr Baden-Württemberg durch das Ministerium für Inneres, Digitalisierung und Kommunen Baden-Württemberg als Teil der Digitalisierungsstrategie digital@bw gefördert. Sie ist telefonisch rund um die Uhr für kleine und mittlere Unternehmen (KMU) aus Baden-Württemberg verfügbar [7].

{(Ersthilfe, direkte Unterstützung, {technisch,organisatorisch}, kostenlos, 24/7-Unterstützung, regional, {Privatpersonen, Unternehmen, Behörden, öffentliche Einrichtungen}, {Telefon}), (Ersthilfe, Einbindung Dritter, {technisch,organisatorisch}, kostenlos, 24/7-Erreichbarkeit, regional, {Unternehmen}, {Telefon}), (Hilfedienstleistung, Einbindung Dritter, {technisch,organisatorisch,juristisch}, kostenpflichtig, 24/7-Erreichbarkeit, regional, {Unternehmen}, {Telefon})}

3.2.2 Cyber-Sicherheitsnetzwerk (CSN) [8]

Das Cyber-Sicherheitsnetzwerk ist eine Initiative des Bundesamts für Sicherheit in der Informationstechnik (BSI), um bundesweit Unterstützung für KMU und Privatpersonen anzubieten [9]. Wesentlicher Bestandteil ist dabei die Digitale Rettungskette, mit deren Hilfe Vorfälle so lange weiter eskaliert werden können, bis diese behoben werden können. Während die Hotline nur zu den Geschäftszeiten zur Verfügung steht, kann jederzeit über die Webseite auf die Liste der verfügbaren IT-Kräfte zugegriffen werden [10].

{(Ersthilfe, Vermittlung, {technisch,organisatorisch}, kostenlos, 24/7-Erreichbarkeit, bundesweit, {Privatpersonen, Unternehmen}, {E-Mail, Webseite}), (Ersthilfe, Vermittlung, {technisch,organisatorisch}, kostenlos, Geschäftszeiten, bundesweit, {Privatpersonen, Unternehmen}, {Telefon}), (Ersthilfe, Vermittlung, {technisch,organisatorisch}, Kostenpflichtig, 24/7-Erreichbarkeit, bundesweit, {Privatpersonen, Unternehmen}, {E-Mail, Webseite}), (Ersthilfe, Vermittlung, {technisch,organisatorisch}, kostenpflichtig, Geschäftszeiten, bundesweit, {Privatpersonen, Unternehmen}, {Telefon}), (Hilfedienstleistung, Vermittlung, {technisch,organisatorisch,juristisch}, kostenpflichtig, 24/7-Erreichbarkeit, bundesweit, {Privatpersonen, Unternehmen}, {E-Mail, Webseite}), (Hilfedienstleistung, Vermittlung, {technisch,organisatorisch,juristisch}, kostenpflichtig, Geschäftszeiten, bundesweit, {Privatpersonen, Unternehmen}, {Telefon})}

3.2.3 Cyberwehr Nordrhein-Westfalen [11]

Die Cyberwehr in Nordrhein-Westfalen ist ein Projekt des eurobits e.V. für KMU aus Bochum, Essen und Gelsenkirchen. Sie wird vom Ministerium für Wirtschaft, Innovation, Digitalisierung und Energie des Landes NRW

(MWIDE) im Rahmen des RegioCalls EFRE.NRW gefördert und bietet kostenlose vor-Ort-Unterstützung für die Betroffenen [12]. Die Hotline ist werktags zwischen 8 und 18 Uhr erreichbar [11] und vermittelt im Ernstfall IT-Dienstleister aus einem Netzwerk.

{(Ersthilfe, Einbindung Dritter, {technisch, organisatorisch}, kostenlos, Geschäftszeiten, regional, {Unternehmen}, {Telefon}), (Hilfedienstleistung, Einbindung Dritter, kostenpflichtig, Geschäftszeiten, regional, {Unternehmen}, {Telefon})}

3.2.4 Cyberhotline Berlin [13]

Die Cyberhotline Berlin stellt Unternehmen aus Berlin zu den üblichen Bürozeiten eine Hotline sowie ein Kontaktformular zur Verfügung. Das Unterstützungsangebot wird von der DAB Digitalagentur Berlin GmbH angeboten [14]. Eine Ersthilfe ist hier zunächst kostenlos und erfolgt direkt am Telefon. Geht ein Vorfall jedoch in die Tiefe werden kostenpflichtige IT-Dienstleister vermittelt [15].

{ (Ersthilfe, direkte Unterstützung, {technisch, organisatorisch}, kostenlos, Geschäftszeiten, regional, {Unternehmen}, {Telefon}), (Ersthilfe, direkte Unterstützung, {technisch, organisatorisch}, kostenlos, 24/7-Verfügbarkeit, regional, {Unternehmen}, {Kontaktformular}), (Hilfedienstleistung, Einbindung Dritter, {technisch, organisatorisch}, kostenpflichtig, Geschäftszeiten, regional, {Unternehmen}, {Telefon }), (Hilfedienstleistung, Einbindung Dritter, {technisch, organisatorisch}, kostenpflichtig, 24/7-Verfügbarkeit, regional, {Unternehmen}, {Kontaktformular})}

3.2.5 Cyber-Ersthilfe BW [16]

Die Cyber-Ersthilfe BW wird von der Cybersicherheitsagentur Baden-Württemberg (CSBW) betrieben. Sie ist rund um die Uhr telefonisch, per E-Mail oder per Kontaktformular erreichbar und richtet sich im Grunde an alle Opfer von Cyberangriffen in Baden-Württemberg [17]. Bei einem Anruf erhalten die Betroffenen direkte Ersthilfe sowie einen Hinweis auf weitere Anlaufstellen. Teil der CSBW ist auch das CERT BWL, welches als Anlaufstelle für die Landesverwaltung Baden-Württembergs dient [18]. Die Cyber-Ersthilfe BW startete im Juli 2022 [19].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch}, kostenlos, 24/7-Unterstützung, regional, {Privatpersonen, Unternehmen, öffentliche Einrichtungen, Kommunen, Behörden}, {Telefon}), (Ersthilfe, Vermittlung, {technisch, organisatorisch}, kostenlos, 24/7-Unterstützung, regional, {Privatpersonen, Unternehmen, öffentliche Einrichtungen, Kommunen, Behörden}, {Tele-

fon}), (Ersthilfe, Vermittlung, {technisch, organisatorisch}, kostenlos, 24/7-Erreichbarkeit, regional, {Privatpersonen, Unternehmen, öffentliche Einrichtungen, Kommunen, Behörden}, {E-Mail, Kontaktformular})}

3.2.6 Transferstelle Cybersicherheit [20]

Das vom Bundesministerium für Wirtschaft und Klimaschutz geförderte Projekt der Transferstelle Cybersicherheit startete zum Juli 2023 und entwickelt eine Plattform zur bundesweiten Vermittlung von Unterstützungsangeboten an Privatpersonen und Unternehmen [22]. Zum Erstellungszeitpunkt dieses Artikels ist auf der Webseite der Transferstelle eine Ankündigung geschaltet, die auf den Aufbau einer Vermittlungsplattform hinweist [21].

3.3.1 Bundesamt für Verfassungsschutz (BfV)

Das Bundesamt für Verfassungsschutz (BfV) kann deutsche Unternehmen und Behörden unterstützen, falls diese Opfer von Betriebsspionage durch ausländische Geheimdienste werden [4, 23].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch,rechtsstaatlich }, kostenlos, 24/7-Erreichbarkeit, bundesweit, {Unternehmen und Behörden die durch ausländische Geheimdienste angegriffen wurden}, {Telefon})}*

3.3.2 Verfassungsschutz NRW

Speziell Unternehmen und Forschungseinrichtungen aus Nordrhein-Westfalen können sich bei einem Cyberangriff auch telefonisch oder per Mail an den Verfassungsschutz NRW wenden [24, 25].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch,rechtsstaatlich }, kostenlos, 24/7-Erreichbarkeit, regional, {Unternehmen, Forschungseinrichtungen}, {Telefon, E-Mail})}*

3.3.3 Bundeskriminalamt (BKA)

Allgemeine Unterstützung bei verschiedenen Cyberangriffen auf Behörden oder Einrichtungen des Bundes bietet das Bundeskriminalamt [26, 27]. Das BKA ist per Telefon und Mail [28] sowie zusätzlich unter der Behördennummer 115 (werktags 8-18 Uhr [29]) erreichbar [45]. Die Quick Reaction Force (QRF) des BKA ist 24/7 einsatzbereit [27].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch,rechtsstaatlich}, kostenlos, 24/7-Unterstützung, bundesweit, {Behörden, Einrichtungen des Bundes, KRITIS}, {Telefon, E-Mail})}

3.3.4 Zentrale Ansprechstellen Cybercrime (ZAC) [30, 31, 32, 33, 34]

Die Zentralen Ansprechstellen Cybercrime der Polizeien stehen in jedem der 16 deutschen Bundesländern Unternehmen aus dem jeweiligen Bundesland zur Verfügung [35]. Sie sind rund um die Uhr telefonisch, per E-Mail [36] und per Kontaktformular erreichbar, um kostenfrei den Sachverhalt aufzuklären, sowie weitere Angriffe zu verhindern.

{(*Ersthilfe*, d*irekte Unterstützung*, {*technisch*, *organisatorisch,rechtsstaatlich* }, *kostenlos*, *24/7-Erreichbarkeit*, *regional*, {*Unternehmen*, *öffentliche Einrichtungen*, *Behörden*}, {*Telefon*, *E-Mail*, *Kontaktformular*})}

3.3.5 Komm.ONE

Für IT-Sicherheitsvorfälle bei Kommunen in Baden-Württemberg ist der Dienstleister Komm.ONE verfügbar [17].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch}, kostenlos, Geschäftszeiten[2], regional, {Kommunen}, {Telefon})}

3.3.6 Bayern-CERT [37]

Das Bayern-CERT des Landesamtes für Sicherheit in der Informationstechnik (LSI) Bayern dient als Anlaufstelle für bayerische Behörden und ist bei Notfällen zu üblichen Bürozeiten per E-Mail erreichbar [38, 39].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch}, kostenlos, Geschäftszeiten, regional, {Landesverwaltung, Behörden}, {E-Mail})}

3.3.7 Hessen3C [40]

In Hessen ist für mehrere Zielgruppen zusätzlich das Hessen3C rund um die Uhr erreichbar [41].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch}, kostenlos, 24/7-Unterstützung, regional, {Landesverwaltung, Kommunen, KRITIS, Unternehmen}, {Telefon, E-Mail})}*

3.3.8 Onlinewache der Polizei

Privatpersonen können sich bundesweit an die Onlinewachen der Polizeien der Länder oder lokale Polizeidienststellen wenden [42].

{(Ersthilfe, direkte Unterstützung, {technisch, organisatorisch,rechtsstaatlich }, kostenlos, 24/7-Erreichbarkeit, bundesweit, {Privatpersonen}, {Polizeidienstellen, Kontaktformular})}*

---

[2] Nach telefonischer Auskunft

3.4 Kommerzielle Direkthilfe

An dieser Stelle kann unmöglich eine Liste kommerzieller Anbieter gegeben werden, da diese einen wesentlich größeren Umfang als die Angebote der ersten beiden Kategorien aufweisen. Allein im Jahr 2022 betrug der Umsatz für Dienstleistungen im Bereich der IT-Sicherheit in Deutschland 3,6 Milliarden Euro [43]. Neben Selbstständigen und kleinen Unternehmen existieren auch große Systemhäuser mit umfangreichen Hilfsdienstleistungen. Das BSI führt eine Liste von APT-Response-Dienstleistern, welche bei gezielten und besonders schweren IT-Sicherheitsvorfällen kostenpflichtige Direkthilfe anbieten können [44].

4 Fazit

Unternehmen und Privatpersonen, die Unterstützung bei einem IT-Sicherheitsvorfall suchen, stehen vor der Herausforderung, ein verfügbares, angemessenes und wirtschaftliches Angebot auszuwählen. Unabhängige und anbieterneutrale Vermittlungsstellen adressieren diesen Bedarf. Die Herausforderung der Vermittlungsstellen ist, in der Wahrnehmung ihrer Adressaten präsent zu sein.

Wir haben in diesem Artikel eine Taxonomie vorgestellt, die verschiedene Unterstützungsangebote strukturiert vergleichbar macht. Unsere Taxonomie konzentriert sich auf den Raum der Bundesrepublik Deutschland. Eine Betrachtung auf europäischer und internationaler Ebene würde das Bild wesentlich erweitern und auch die bundesdeutschen Angebote in einen größeren Kontext einordnen.

Die Erstellung dieses Artikels wurde im Rahmen der Transferstelle Cybersicherheit im Mittelstand gefördert. Die Transferstelle Cybersicherheit im Mittelstand gehört zu Mittelstand-Digital. Mit dem Mittelstand-Digital Netzwerk unterstützt das Bundesministerium für Wirtschaft und Klimaschutz die Digitalisierung in kleinen und mittleren Unternehmen und dem Handwerk.

Literaturhinweise

[1] Bitkom, Organisierte Kriminalität greift verstärkt die deutsche Wirtschaft an, Presseinformation: https://www.bitkom.org/Presse/Presseinformation/Organisierte-Kriminalitaet-greift-verstaerkt-deutsche-Wirtschaft-an, abgerufen am 26.02.2024

[2] Deutscher CERT-Verbund, Übersichtsseite des CERT-Verbundes: https://www.cert-verbund.de/, abgerufen am 26.02.2024

[3] GÉANT Association, Datenbank europäischer Notfallteams bei Cyberangriffen: https://www.trusted-introducer.org/directory/index.html, abgerufen am 26.02.2024

[4] Jürgen Berke, Übersichtseite verschiedener Anlaufstellen bei Cyberangriffen: https://cyberberke.de/notruf-cyberattacken.html, abgerufen am 21.02.2024

[5] Staatsministerium Baden-Württemberg, Rückblick des ersten CyberSicherheitsForums: https://www.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/erstes-cybersicherheitsforum/, abgerufen am 26.02.2024

[6] Cyberwehr Baden-Württemberg, Projektwebseite: https://cyberwehr-bw.de/, abgerufen am 21.02.2024

[7] Landtag von Baden-Württemberg, Mittelstandsbericht 2021: https://www.landtag-bw.de/files/live/sites/LTBW/files/dokumente/WP17/Drucksachen/1000/17_1550_D.pdf, Seite 104, abgerufen am 21.02.2024

[8] BSI, Informationen über das Cyber-Sicherheitsnetzwerk: https://www.bsi.bund.de/DE/IT-Sicherheitsvorfall/Unternehmen/Ich_suche_Unterstuetzung_durch_das_CSN/Ich_suche_Unterstuetzung_durch_das_CSN_node.html, abgerufen am 26.02.2024

[9] BSI, Informationen zum Cybersicherheitsnetzwerk: https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Cyber-Sicherheitsnetzwerk/cyber-sicherheitsnetzwerk_node.html, abgerufen am 21.02.2024

[10] BSI, Kontaktdaten des CSN: https://www.bsi.bund.de/DE/IT-Sicherheitsvorfall/Unternehmen/Ich-habe-einen-IT-Sicherheitsvorfall-Was-soll-ich-tun/ich-habe-einen-it-sicherheitsvorfall-was-soll-ich-tun_node.html, abgerufen am 23.02.2024

[11] eurobits e.V., Mitteilung zum Start der Cyberwehr NRW: https://www.eurobits.de/cyberwehr-start-ruhrgebiet/, abgerufen am 21.02.2024

[12] Cyberwehr NRW, Projektwebseite: https://cyberwehr.net/, abgerufen am 21.02.2024

[13] DAB Digitalagentur Berlin GmbH, Informationen zum Start der Cyberhotline Berlin: https://digitalagentur.berlin/infothek/pressemitteilung/cyberhotline-fuer-berliner-unternehmen-freigeschaltet/, abgerufen am 26.02.2024

[14] Cyberhotline Berlin, Projektwebseite: https://digitalagentur.berlin/angebote/it-sicherheit/cyberhotline/, abgerufen am 21.02.2024

[15] it's BB, Mitteilung zum Start der Cyberhotline Berlin: https://www.itsbb.net/projekte/projektseite-2-cyberholine/, abgerufen am 21.02.2024

[16] Kommune21, Informationen über die Cyber-Ersthilfe Baden-Württemberg: https://www.kommune21.de/meldung_41753_Direkter+Draht+zur+Cyber-Ersthilfe+BW.html, abgerufen am 26.02.2024

[17] Cybersicherheitsagentur Baden-Württemberg, Projektwebseite der Cyber-Ersthilfe: https://www.cybersicherheit-bw.de/cyber-ersthilfe-bw, abgerufen am 21.02.2024

[18] Cybersicherheitsagentur Baden-Württemberg, Übersichtsseite des CERT BWL: https://www.cybersicherheit-bw.de/das-cert-bwl, abgerufen am 21.02.2024

[19] Land Baden-Württemberg, Mitteilung zur erweiterten Erreichbarkeit der Cyberwehr Baden-Württemberg: https://www.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/erweiterte-erreichbarkeit-der-cyber-ersthilfe-bw-bei-cyberangriffen-1, abgerufen am 21.02.2024

[20] NWB Verlag, Informationen zum Start der Transferstelle Cybersicherheit: https://datenbank.nwb.de/Dokument/1026403/, abgerufen am 26.02.2024

[21] Transferstelle Cybersicherheit, Projektwebseite: https://transferstelle-cybersicherheit.de/akute-hilfe-erhalten/, abgerufen am 26.02.2024

[22] DLR, Förderbekanntmachung der Transferstelle Cybersicherheit: https://projekttraeger.dlr.de/de/foerderung/foerderangebote-und-programme/it-sicherheit-in-der-wirtschaft, abgerufen am21.02.2024

[23] Bundesamt für Verfassungsschutz, Übersichtsseite mit Kontaktdaten des BfV: https://www.verfassungsschutz.de/DE/themen/cyberabwehr/cyberabwehr_node.html, abgerufen am 21.02.2024

[24] Land Nordrhein-Westfalen, Übersichtseite mit Kontaktdaten des LfV NRW: https://www.im.nrw/themen/verfassungsschutz/schutz-von-behoerden-und-unternehmen/wirtschaftsschutz, abgerufen am 26.02.2024

[25] Digital.Sicher.NRW, Soforthilfemaßnahmen bei Cyberangriffen: https://www.digital-sicher.nrw/soforthilfe/erstehilfemassnahmen, abgerufen am 21.02.2024

[26] Bundeskriminalamt, Hinweise zu Cybercrime: https://www.bka.de/DE/UnsereAufgaben/Deliktsbereiche/Cybercrime/cybercrime_node.html, abgerufen am 21.02.2024

[27] Bundeskriminalamt, Abteilung Cybercrime des Bundeskriminalamtes: https://www.bka.de/DE/DasBKA/OrganisationAufbau/Fachabteilungen/Cybercrime/cybercrime_node.html, abgerufen am 26.02.2024

[28] Bundeskriminalamt, Kontaktdaten des Bundeskriminalamtes: https://www.bka.de/DE/KontaktAufnehmen/Kontaktinformationen/Kontaktdaten/kontaktdatenBKA_node.html, abgerufen am 21.02.2024

[29] Bundesministerium des Innern und für Heimat, Erklärung der Behördennummer 115: https://www.bmi.bund.de/DE/service/behoerdennummer-115/behoerdennummer-115-node.html, abgerufen am 21.02.2024

[30] Jörg Bruhn, Vorstellung der ZAC Mecklenburg-Vorpommern: https://www.neubrandenburg.ihk.de/fileadmin/user_upload/News/Innovation_und_Umwelt/ZAC_Joerg_Bruhn.pdf, abgerufen am 26.02.2024

[31] Landtag Baden-Württemberg, Stellungnahme Verhinderung und Aufklärung von Cybercrime Straftaten: https://www.landtag-bw.de/files/live/sites/LTBW/files/dokumente/WP16/Drucksachen/6000/16_6794_D.pdf, abgerufen am 26.02.2024

[32] Sächsisches Staatsministerium des Innern, Webseite der ZAC Sachen: https://www.polizei.sachsen.de/de/47792.htm, abgerufen am 26.02.2024

[33] Gewerkschaft der Polizei, Hinweise zu Cybercrime: https://dpt-statisch.s3.eu-central-1.amazonaws.com/dpt-digital/dpt-25/medien/dateien/284/Leseprobe-Cybercrime.pdf, abgerufen am 26.02.2024

[34] VDI, Cyber Security aus dem Blickwinkel der Bayerischen Polizei: https://www.technik-in-bayern.de/ikt/cyber-sicherheit/cyber-security-aus-dem-blickwinkel-der-bayerischen-polizei, abgerufen am 26.02.2024

[35] Bundeskriminalamt, Übersicht aller ZACs in Deutschland: https://www.polizei.de/Polizei/DE/Einrichtungen/ZAC/zac_node.html, abgerufen am 26.02.2024

[36] Landeshauptstadt Schwerin, Übersicht aller ZACs: https://www.schwerin.de/politik-verwaltung/dienstleistungen/verwaltungsleistungen/Zentrale-Ansprechstelle-Cybercrime-ZAC-Dienststelle-fuer-die-Wirtschaft-kontaktieren/, abgerufen am 26.02.2024

[37] Bayerische Staatskanzlei, Rückblick 5 Jahre LSI: https://www.bayern.de/fueracker-5-jahre-lsi-5-jahre-beschuetzer-der-bayerischen-behoerden-it-und-berater-fuer-kommunen-und-kritis-unternehmen-it-sicherheit-als-grundlage-fuer-digitalisierung/, abgerufen am 26.02.2024

[38] LSI Bayern, Projektwebseite: https://www.lsi.bayern.de/, abgerufen am 26.02.2024

[39] GÉANT Association, Übersichtsseite mit Informationen über das Bayern-CERT: https://www.trusted-introducer.org/directory/teams/bayern-cert-de.html, abgerufen am 26.02.2024

[40] Hessen CyberCompetenceCenter, Projektwebseite: https://hessen3c.de/hessen-cybercompetencecenter/ueber-hessen3c, abgerufen am 26.02.2024

[41] Hessen CyberCompetenceCenter, Projektwebseite: https://hessen3c.de/soforthilfe-bei-cyberangriffen, abgerufen am 21.02.2024

[42] Ministerium für Inneres, Bauen und Sport des Saarlandes, Onlinewache der Polizei zur Anzeigenerstattung: https://portal.onlinewache.polizei.de/de/, abgerufen am 21.02.2024

[43] Bitkom, Presseinformation zum Stand der IT-Sicherheit 2022: https://www.bitkom.org/Presse/Presseinformation/IT-Sicherheit-waechst-2022, abgerufen am 23.02.2024

[44] BSI, Liste qualifizierter APT-Response-Dienstleister: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Dienstleister_APT-Response-Liste.html, abgerufen am 23.02.2024

[45] Bundeskriminalamt, Mitteilung des BKA zur Teilnahme an der Behörden-nummer 115: https://www.bka.de/DE/KontaktAufnehmen/Kontaktinformationen/EinheitlicheBehoerdenrufnummer/einheitlichebehoerdenrufnummer_node.html, abgerufen am 21.02.2024

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Trustpoint: Digitale Identitäten für eine sichere Industrie

Rohit Bohara[1], Florian Handke[2], Alexander Harig[2], Mario Kemper[3], Prof. Dr. Ing. Jan Pelzl[4], Andreas Philipp[5], Dr. Claudia Priesterjahn[3], Christian Schwinne[4]

Kurzfassung:
Die Digitalisierung der Industrie hat zu hochgradig automatisierten und vernetzten Betriebsumgebungen geführt, was die Bedrohung durch Cyberangriffe in industriellen Umgebungen erhöht. Im Rahmen des Forschungsprojekts Trustpoint werden innovative Lösungen entwickelt, um das Vertrauen zwischen den Komponenten und den Netzwerken in industriellen Umgebungen zu stärken und die Interoperabilität zu verbessern. Das Projekt adressiert Herausforderungen wie segmentierte Netzwerke, begrenzte Hardware-Ressourcen und mangelnde Sicherheitskompetenz mit einer benutzerfreundlichen Open-Source-Lösung für das Management digitaler Identitäten, die sowohl benutzergesteuertes als auch automatisiertes Zero-Touch-Onboarding ermöglicht.

Stichworte: Digitale Identitäten, Industrial Security, Onboarding

1 Abstrakt

Die fortschreitende Digitalisierung von Maschinen und Anlagen sowie die zunehmende Vernetzung stellen immer höhere und komplexere Anforderungen an deren IT-Sicherheit. Digitale Identitäten von Maschinen spielen dabei eine entscheidende Rolle, da sie Vertrauen zwischen Herstellern und Betreibern schaffen.

Die Rahmenbedingungen in industriellen Umgebungen stellen allerdings eine Herausforderung für die Einführung und Verwaltung digitaler Identitäten dar. Dies wird häufig durch begrenzte Hardware- und Softwareressourcen sowie mangelnde Interoperabilität zwischen Herstellern erschwert. Ein zentrales Problem besteht darin, dass Hersteller und Betreiber von Maschinen und Anlagen oft nicht über die erforderlichen Kompetenzen und Ressourcen verfügen, um eigenständig komplexe Sicherheits- und Identitätskonzepte zu erstellen und umzusetzen. Dadurch entstehen Schwachstellen und Risiken für den laufenden Betrieb.

[1] asvin GmbH
[2] Campus Schwarzwald gGmbH
[3] achelos GmbH
[4] Hochschule Hamm-Lippstadt
[5] PrimeKey Labs GmbH

Das Ziel des Forschungsprojekts Trustpoint besteht darin, komplexe Prozesse und Mechanismen so zu abstrahieren, dass Hersteller und Betreiber Maschinen und Geräte sicher in ihre Netzwerke integrieren und betreiben können. Trustpoint ermöglicht eine nahtlose und vertrauenswürdige Kommunikation zwischen verschiedenen Akteuren entlang der Wertschöpfungskette durch die Schaffung eines Vertrauensankers (Trustpoint).

Mit diesem Beitrag stellen wir Methodik und erste Ergebnisse von Trustpoint vor, in welchen unter anderem Anforderungen existierender Standards wie OPC UA Part 21[6], BRSKI[7] und FIDO FDO[8] an Hersteller, Integratoren, Betreiber und Anwender sowie den Produktlebenszyklus beschrieben und mögliche Erweiterungen zur praktischen Umsetzung und zur Steigerung des Sicherheitswertes analysiert werden. Der Anwendungsfall 'User-driven Onboarding' wurde bereits exemplarisch in einer Konzeptstudie für bestehende Systeme umgesetzt.

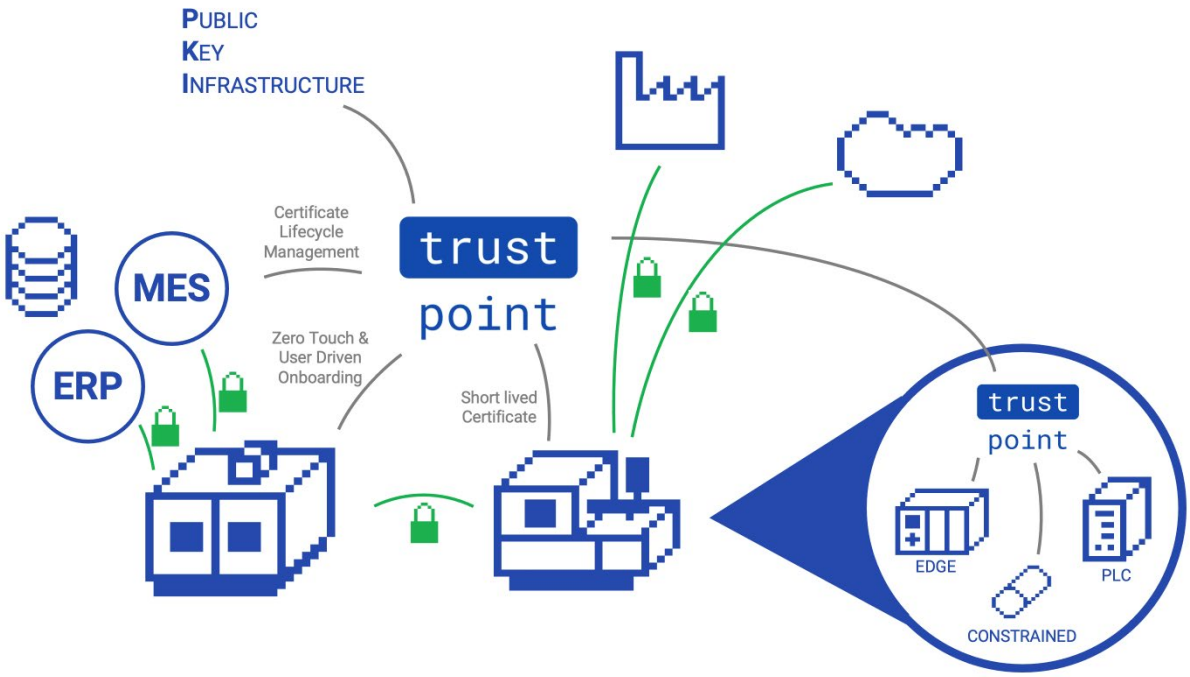


Abbildung 1: Trustpoint Mechanismen

Wie in Abbildung 1 dargestellt, bildet Trustpoint einen Vertrauensanker in Fabriken und Maschinen und verwaltet die im Netzwerk befindlichen Geräte.

6 OPC, „OPC 10000-21: UA Part 21: Device Onboarding," 2022.

7 IETF - Internet Engineering Task Force, „Bootstrapping Remote Secure Key Infrastructure (BRSKI)," 2021. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8995 . [Zugriff am 9 Juni 2023].

8 FIDO Device Onboard Specification, Review Draft, December 02, 2020, https://fidoalliance.org/specs/FDO/FIDO-Device-Onboard-RD-v1.0-20201202.html

Dabei werden verschiedene Schnittstellen zu Vertrauensdiensten wie einer Public Key Infrastructure (PKI) sowie zu den jeweiligen Endgeräten angeboten.

2 Motivation

Die Digitalisierung hat zu einer Transformation der traditionellen Industrie in Richtung hochautomatisierter, vernetzter und intelligenter Betriebsumgebungen geführt. Dies hat zu einem tiefgreifenden Wandel in den operativen Technologien (OT) und den industriellen Prozessen geführt. Betroffen sind nicht nur die Produktion, sondern auch die gesamte Lieferkette. Mit zunehmender Vernetzung und Automatisierung sind industrielle Umgebungen immer stärker von Cyber-Angriffen bedroht. Deshalb gewinnt die Gewährleistung der IT-Sicherheit enorm an Bedeutung.

Dadurch wird die Verwendung von Maschinenidentitäten, wie sie in Standards wie OPC UA [OPC-10000-21] und IEC 62443[9] beschrieben sind, immer wichtiger: Diese spielen eine entscheidende Rolle beim Aufbau von Sicherheit und Integrität für Kommunikation und Daten in industriellen Netzwerken. Maschinenidentitäten ermöglichen die Authentifizierung, die Validierung von Software, den Nachweis der Herkunft von Komponenten und die Schaffung einer sicheren Rückverfolgbarkeit. Damit ermöglichen digitale Identitäten die Herstellung von Vertrauen zwischen Herstellern und Betreibern.

In der Praxis werden Public-Key-Infrastrukturen (PKI) genutzt, um digitale Identitäten in Form von digitalen Zertifikaten bereitzustellen und zu verwalten. Die Installation, Konfiguration und der Betrieb solcher Sicherheitstechnologien erfordern oft spezifisches Fachwissen, das bei Anlagenbetreibern nicht immer vorhanden ist oder teuer hinzugekauft werden muss. Konfigurationsfehler und falsche Bedienung können zu Sicherheitsrisiken führen.

Industrielle Umgebungen sind mit spezifischen Einschränkungen konfrontiert, die es erschweren, digitale Zertifikate und kryptographische Schlüssel zu verwalten. Dazu gehören Netzwerksegmentierung, eingeschränkte Konnektivität, begrenzte Hardware- und Software-Ressourcen, hohe Netzwerkdynamik und organisatorische Einschränkungen. Maschinen- und Anlagenbauer sowie -betreiber stehen vor der Herausforderung, Expertise im Bereich der Informationssicherheit aufzubauen und bestehende Sicherheitslösungen an die spezifischen Anforderungen und Einschränkungen von industriellen Anlagen anzupassen.

[9] IEC 62443, Die internationale Normenreihe für Cybersecurity in der Industrieautomatisierung

So existieren z.B. laut VDMA derzeit keine Lösungen, die mit vertretbarem Aufwand für eine Mehrheit der Betreiber anwendbar sind[10]. Die Autoren des Leitfadens zum Standard IEC 62443 für den Maschinen- und Anlagenbau sehen ein Zertifikat und die dahinterliegende Identität als einen elementaren Bestandteil eines sicheren Konzeptes.

3 Innovation

Trustpoint hat als Forschungsprojekt das Ziel, innovative Lösungen zu entwickeln. Eine der wichtigsten Herausforderungen ist die Verbesserung von Sicherheit und Effizienz. Unternehmen können Maschinen und Geräte sicher in Netzwerke integrieren und betreiben. Dadurch wird das Vertrauen zwischen Komponenten und Netzwerken gestärkt, was zu einer erheblichen Verbesserung der Sicherheit führt. Gleichzeitig wird die Effizienz und Flexibilität industrieller Prozesse erhöht. Dies trägt zur Steigerung der Wettbewerbsfähigkeit bei.

Ein weiterer innovativer Aspekt von Trustpoint ist die Förderung nahtloser Kommunikation. Trustpoint schafft einen sicheren Vertrauensanker, der eine vertrauenswürdige Kommunikation zwischen verschiedenen Akteuren entlang von Wertschöpfungsketten ermöglicht. Dadurch wird die Interoperabilität verbessert und die Zusammenarbeit zwischen Herstellern, Betreibern und anderen Partnern in industriellen Umgebungen erleichtert.

4 Herausforderungen

Diese Herausforderungen sind eng mit den spezifischen Anforderungen und Einschränkungen von Fabriken, Maschinen und industriellen Netzwerken verbunden. Im Folgenden werden einige der wichtigsten Herausforderungen aufgeführt:

Netzwerke sind in industriellen Umgebungen oft stark segmentiert und weisen häufig eine eingeschränkte Konnektivität zu IT-Systemen auf. Diese Segmentierung und eingeschränkte Konnektivität erschweren die nahtlose Integration in erweiterte Wertschöpfungsnetzwerke. Mit der Ausweitung interner Netzwerke auf externe Partner, Kunden und Lieferanten wird es immer dringlicher, Vertrauen zu weiteren Einheiten über die Fabrikgrenzen hinaus aufzubauen.

Hard- und Software in industriellen Umgebungen unterliegen verschiedenen Einschränkungen. Die Aktualisierung von Software und Firmware er-

[10] VDMA, Leitfaden IEC 62443 für den Maschinen- und Anlagenbau, 2021

folgt oft nur unregelmäßig und die Lebensdauer von Komponenten in Maschinen ist deutlich länger als die herkömmlicher IT-Komponenten. Begrenzte Hardwareressourcen in diesen Umgebungen erschweren die Implementierung kryptografischer Funktionen. Zudem fehlen oft sichere Speichermöglichkeiten für kryptografisches Material.

Industrielle Netzwerke sind von Natur aus **dynamisch**. Geräte treten häufig in das Maschinen- oder Fabriknetz ein oder aus ihm aus, was das Identitätsmanagement erschwert.

Die notwendigen Kompetenzen und Ressourcen, um umfassende Sicherheits- oder Identitätskonzepte zu entwickeln und umzusetzen, sind in vielen Unternehmen, seien es Hersteller oder Betreiber, organisatorisch noch nicht vorhanden.

Der Nachweis von Ereignissen im Lebenszyklus einer Maschine erfolgt häufig manuell und ist lückenhaft. Fälschungssicherheit und eindeutige Rückverfolgbarkeit sind jedoch entscheidend für die Vertrauensbildung zwischen Herstellern und Betreibern.

Die **Interoperabilität** zwischen Herstellern und Betreibern sowie zwischen verschiedenen Herstellern ist häufig nicht gegeben. Die Betreiber haben Schwierigkeiten, die von den Herstellern eingebrachten Identitäten effektiv zu verwalten und zu überwachen, um den ordnungsgemäßen Betrieb der Maschinen zu gewährleisten. Zudem sind existierende Sicherheitsmaßnahmen in Komponenten nicht einfach zu bewerten.

Systeme mit langlebigen Komponenten erfordern kryptografische Mechanismen, die langzeitsicher oder zumindest austauschbar sind.

5 Zielsetzung

Trustpoint hat das Ziel, eine benutzerfreundliche Open-Source-Lösung für die sichere Verwaltung digitaler Identitäten von Maschinen und Komponenten im industriellen Umfeld zu entwickeln. Dabei beschäftigt sich Trustpoint mit den komplexen Herausforderungen, die mit IT-Sicherheit und Identitätsmanagement in diesem Umfeld verbunden sind.

Diese Problematik wird durch eine umfassende Betrachtung der Gegebenheiten im Anlagenbau angegangen, wobei Brownfield-Anlagen berücksichtigt werden, um bestehende Anlagen sinnvoll und zukunftsorientiert abzusichern. Das Vertrauensmanagementsystem Trustpoint reduziert die Komplexität für den Betreiber, indem es anwendungsnahe Prozesse für das Device

Onboarding bereitstellt, sinnvolle Sicherheitsfunktionen und Protokolle auswählt und damit das Sicherheitsrisiko durch falsche oder suboptimale Konfiguration verringert.

6 Methodik/Ansatz

Industrielle Umgebungen erfordern innovative Ansätze, um digitale Identitäten von Maschinen und Komponenten sicher zu verwalten. Vor diesem Hintergrund bietet Trustpoint ein skalierbares Konzept, das sowohl ein benutzergesteuertes Onboarding als auch automatisiertes Zero-Touch-Onboarding umfasst. Dieses Konzept soll es Herstellern und Betreibern einfacher machen, den Herausforderungen der dynamischen Industrielandschaft gerecht zu werden und eine flexible Lösung für die Integration von Geräten in komplexe Netzwerke bieten.

7 Anforderungen

Trustpoint strebt eine herstellerunabhängige Lösung an, die eine flexible Architektur gewährleistet und mit verschiedenen Geräten und Herstellern kompatibel ist. Dadurch wird eine reibungslose Integration in heterogene industrielle Umgebungen ermöglicht. Insbesondere im Hinblick auf Legacy-Systeme hat Trustpoint das Ziel, eine Lösung bereitzustellen, die eine nahtlose Integration von Bestandssystemen ermöglicht. Diese Lösung unterstützt eine schrittweise Migration und Integration, ohne den Betrieb der Altsysteme zu beeinträchtigen. Trustpoint berücksichtigt Sicherheitsaspekte und implementiert robuste Sicherheitsmechanismen, die unabhängig von der Verfügbarkeit von Online-Diensten oder der Art der Netzwerkverbindung funktionieren. Trustpoint soll eine anpassungsfähige Lösung bieten, die den individuellen Bedürfnissen und Besonderheiten verschiedener industrieller Umgebungen gerecht wird. Es wird die Anpassung an verschiedene Netzwerkarchitekturen, Gerätetypen und Sicherheitsanforderungen ermöglichen, um den unterschiedlichen Anforderungen gerecht zu werden. Ein wichtiges Ziel von Trustpoint ist die Interoperabilität mit anderen Standards. Trustpoint strebt die Interoperabilität mit anderen Industriestandards an, um eine nahtlose Integration mit bestehenden Protokollen und Systemen zu ermöglichen. Dabei legt Trustpoint großen Wert auf die Unterstützung der Anwender, um eine benutzerfreundliche und intuitiv zu bedienende Umgebung zu schaffen. So können die Anwender fundierte Entscheidungen bei der Auswahl und Implementierung von Onboarding-Mechanismen treffen. Dies betont die Praxisorientierung von Trustpoint und gewährleistet eine effektive Nutzung der Lösung durch die Anwender.

Weiterhin soll Trustpoint die Anforderungen von in der Industrie relevanten Standards und Normen erfüllen. Unter anderem soll das System, in dem Trustpoint zum Einsatz kommt, ein definiertes Sicherheitslevel (SL-C) nach ISO/IEC 62443-3-3 [IEC 62443] erreichen können. Relevante Standards für das Schlüsselmanagement sind u.a. ISO/IEC 11770-1,-3 (Information technology – Security techniques - Key Management)[11] die Special Publication NIST 800-57 (Part 1, 8.1.4 und 8.1.5)[12], und die DIN EN 62351-9[13]. Die bekannten Standards setzen auf eine zentrale, vertrauenswürdige Infrastruktur. Lebenszyklen von kryptografischen Schlüsseln werden u.a. ausführlich in DIN EN 62351-9 für Energiemanagementsysteme beschrieben. Symmetrische und asymmetrische Systeme zum Schlüsselmanagement finden sich detailliert in DIN EN 62351-9 und können auf andere Anwendungen erweitert werden. U.a. könnten symmetrisch basierte Systeme zum Aufbau PQ-resistenter Infrastrukturen dienen. Konzepte für Langzeitsicherheit durch quantencomputerresistente Verfahren befinden sich aktuell in der Standardisierung (NISTIR 8413, Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process)[14]. RFC 8995 (BRSKI) beschreibt ein Zero-Touch Bootstrapping Verfahren von neuen Geräten. IEEE802.1AR[15] spezifiziert sichere Device Identities (DevIDs).

7.1 Zero-Touch Onboarding

Verschiedene Aspekte im Kontext von Maschinenidentitäten wurden bereits in durchgeführten Forschungsprojekten untersucht, wie zum Beispiel im nationalen Referenzprojekt zur IT-Sicherheit in der Industrie 4.0 (IUNO). Trustpoint baut somit auf den gewonnenen Erkenntnissen im Bereich der Risikoanalysen und möglichen Schutzmaßnahmen auf. Das zentrale Anliegen besteht darin, den bisher unterrepräsentierten Aspekt des Lifecycles zu behandeln und die damit verbundenen Hürden bei der Implementierung und Verwaltung zu erläutern. Diese können den Einsatz von IT-Lösungen teilweise unmöglich machen oder erheblich erschweren, wodurch weitere Forschung notwendig wird.

---

11 Standardization, ISO - International Organization for, *ISO/IEC 11770-3:2021 Information security — Key management — Part 3: Mechanisms using asymmetric techniques,* 2021.

12 NIST - National Institute of Standards and Technology, *SP 800-57 Part 1 Rev. 5 Recommendation for Key Management: Part 1 – General,* 2020.

13 DIN, *DIN EN 62351-9 VDE 0112-351-9:2018-05 Energiemanagementsysteme und zugehöriger Datenaustausch – IT-Sicherheit für Daten und Kommunikation,* VDE Verlag, 2018.

14 NIST - National Institute of Standards and Technology, *NISTIR 8413 Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process,* 2022.

15 IEEE, „802.1AR: Secure Device Identity - IEEE Standard for Local and metropolitan area networks–Secure Device Identity," 2017.

Im heutigen IT-Umfeld liegt ein stärkerer Fokus auf Zero-Trust-Architekturen. Traditionell wurde allen Geräten und Nutzern im internen Netzwerk vertraut. Sicherheitsvorkehrungen wurden auf den Schutz der Grenze zwischen internem Industrienetz und externen Netzwerken konzentriert. Dieser Ansatz ist jedoch sowohl aus Sicht der Angriffsmöglichkeiten als auch wegen des Bedarfs nach Fernzugriff auf das Netzwerk nicht mehr zeitgemäß. Im Zero-Trust-Modell müssen sich alle Geräte und Nutzer authentifizieren, da kein implizites Vertrauen besteht. Zudem werden alle vertraulichen Daten verschlüsselt übertragen. Zusätzlich wird das Prinzip der 'Defense in Depth' angewendet, welches mehrere aufeinander aufbauende Sicherheitsmaßnahmen beinhaltet. Des Weiteren wird das 'Least-Privilege'-Prinzip angewendet, bei dem die Berechtigungen von Nutzern, Geräten und Prozessen auf das absolut notwendige Minimum reduziert werden. Die Ziele der Zero-Trust-Architektur sind auch normativ in der ISO/IEC 62443-3-3 festgelegt.

Ein wesentlicher Aspekt des Trustpoint-Projekts ist die Integration von Zero-Touch-Onboarding- und Bootstrapping-Verfahren, die in Standards wie dem untersuchten BRSKI (Bootstrapping Remote Secure Key Infrastructure) [BRSKI], OPC UA Part 21 [OPC-10000-21], FIDO FDO (Fast Identity Online Device Onboarding) [FIDO FDO] und sZTP (secure Zero-Touch Provisioning)[16] beschrieben sind. Zero-Touch bedeutet, dass bei der Inbetriebnahme keine manuellen Schritte zum Vertrauensaufbau notwendig sind und automatisiert ein idealerweise gegenseitiges Vertrauensverhältnis zwischen Gerät und Betreibernetzwerk aufgebaut wird. Diese Verfahren spielen eine zentrale Rolle bei der sicheren Integration von Geräten in industrielle Netzwerke und bei der Bereitstellung digitaler Identitäten.

Trustpoint legt einen besonderen Fokus auf die Untersuchung der Prinzipien von BRSKI, um ein sicheres Onboarding von Geräten in Netzwerke zu gewährleisten. Dieser Prozess ermöglicht es, Maschinen und Komponenten sicher und effizient in ein Netzwerk zu integrieren, indem sichergestellt wird, dass sie über die notwendigen Identitäten und Schlüssel verfügen, um sich zu authentifizieren und sicher zu kommunizieren. Diese Integration erleichtert den Aufbau von Vertrauen zwischen den Geräten und dem Netzwerk. Als weiteren Ansatz für ein sicheres Onboarding wird sich Trustpoint an den Methoden von OPC UA Part 21 orientieren, um eine sichere Integration von

16 PKCS #5: Password-Based Cryptography Specification Version 2.0., Network Working Group, Request for Comments: 2898, B. Kaliski, September 2000

Geräten in OPC UA-basierte Netzwerke zu ermöglichen. Dieser Standard bietet Richtlinien und Empfehlungen für die sichere Konfiguration und Kommunikation von OPC UA Clients und Servern.

7.1.1 BRSKI

Wenn ein neues Gerät, auch Pledge genannt, an das Netzwerk angeschlossen wird, beginnt der Onboarding-Prozess. Der Pledge versucht automatisch, sich mit der Domäne des Betreibers zu verbinden. Der BRSKI-Onboarding-Prozess ist abgeschlossen, wenn eine erfolgreiche gegenseitige Authentifizierung zwischen dem Pledge und einem Registrar in der Betreiberdomäne stattgefunden hat.

Das BRSKI berücksichtigt im Wesentlichen zwei Arten von Parteien: Hersteller und Betreiber. Der Betreiber ist für den Betrieb eines Domain-Registrars verantwortlich, der den Onboarding-Prozess orchestriert und entscheidet, welcher neue Pledge in die Domäne aufgenommen wird. In der Regel betreibt der Betreiber auch eine PKI, um die Pledge-Empfänger und die Registrare des Betreibers mit entsprechenden Zertifikaten zu versorgen. Auf der anderen Seite müssen Hersteller eine Manufacturer Authorized Signing Authority (MASA) betreiben, die Pledge-Zertifikate ausstellt.

Sowohl der Pledge als auch der Registrar müssen mit spezieller Software, Zertifikaten, Truststores, Schlüsseln und Konfigurationen ausgestattet sein, um BRSKI erfolgreich zu unterstützen.

Für das erfolgreiche Onboarding nach BRSKI sind bestimmte Anforderungen an den Pledge zu erfüllen

- **DevID-Zertifikate:** Der Pledge muss mindestens ein DevID-Zertifikat vorweisen, wobei es üblich ist, mehrere mit verschiedenen asymmetrischen Algorithmen zu verwenden.
- **Truststore:** Ein Truststore auf dem Pledge ist erforderlich. Dieser speichert entweder das CA-Zertifikat des Herstellers oder die Vertrauenskette, die zur Validierung der empfangenen Voucher verwendet wird.
- **Online-Onboarding:** Das Standardverfahren für das Onboarding ist das Online-Verfahren. Hierbei ist eine aktive Verbindung zur Registrierstelle sowie eine erreichbare Manufacturer Authorized Signing Authority (MASA) zwingend erforderlich.

Es ist besonders zu betonen, dass ohne eine MASA keine weiteren Voucher ausgestellt werden können, was das Onboarding-Prozessrisiko erheblich beeinflusst. Die BRSKI-Spezifikation umfasst verschiedene Varianten, die das

Sicherheitsniveau beeinflussen, und lässt einige Prozesse, wie die Verteilung von Truststores, offen.

BRSKI ermöglicht eine vollständige gegenseitige Authentifizierung und trägt somit aktiv zur Abschaffung des "Trust On First Use" (TOFU)-Prinzips bei. TOFU bedeutet, dass das Gerät bei Erstinbetriebnahme jedem Netzwerk vertraut. Dies bietet einige Nachteile, zum Beispiel kann u. U. ein Gerät von einem Angreifer vor Inbetriebnahme modifiziert werden und so ein Sicherheitsrisiko darstellen.

7.1.2 OPC UA Part 21

Das Protokoll basiert auf bewährten Mechanismen, die eine zuverlässige Integration in industrielle Netzwerke ermöglichen. Im Mittelpunkt stehen dabei Device-Identity-Zertifikate, wie sie in IEEE 802.1AR [IEEE 802.1AR] definiert sind, sowie entsprechende vom Hersteller bereitgestellte Tickets, die das Gerät identifizieren.

Ein Schlüsselelement von OPC UA Part 21 ist der Registrierungsprozess, der bei jedem Eigentumswechsel eines Gerätes neu gestartet wird. Ein Eigentümer, als zentraler Akteur im Lebenszyklus des Gerätes, muss bei jeder Übertragung das Ticket an den neuen Eigentümer weitergeben. Dabei kann es sich entweder um das originale Hersteller-Ticket oder um ein vom Vorbesitzer signiertes Geräte- oder Verbund-Ticket handeln.

Bei der Implementierung von OPC UA 21 sind jedoch einige Aspekte zu beachten. Das Protokoll erfordert mindestens eine Device Certificate Authority (DCA) auf dem Gerät, was die Anwendung auf leistungsfähigere Geräte beschränkt.

Ein weiterer wichtiger Punkt ist die Identifikation und das Vertrauen in den Registrar. OPC UA 21 bietet keinen Mechanismus, um den Registrar zu identifizieren und ihm zu vertrauen, was zu einem TOFU-Szenario führt. Die Validierung des Gerätes erfolgt durch den Registrar, der das Gerätezertifikat und das zugehörige Ticket überprüft. Hier ist das Vertrauen in die Validierungsschritte jedes Akteurs im Lebenszyklus des Geräts von entscheidender Bedeutung.

Ein zentraler Punkt, der bei der Anwendung von OPC UA 21 berücksichtigt werden muss, ist die Notwendigkeit, dass jeder Akteur im Lebenszyklus eines Gerätes eine Public Key Infrastructure (PKI) betreibt oder zumindest Zugang zu einer solchen hat. Der Lebenszyklus eines Tickets liegt außerhalb des Anwendungsbereichs von OPC UA und erfordert separate, nicht protokollbasierte Verwaltungsprozesse.

Der Aufwand, ein Gerät OPC UA fähig zu machen, ist vor allem dann sinnvoll, wenn es in einer OPC UA Infrastruktur betrieben werden soll, da der Aufwand für das Onboarding allein als zu hoch angesehen wird. Trotz dieser Herausforderungen bietet OPC UA 21 einige positive Aspekte. Es ermöglicht ein vollständig automatisiertes Onboarding, sofern Vertrauensanker und Tickets im Vorfeld bereitgestellt wurden. Das Protokoll erfordert keinen externen Zugriff während des Onboardings und bietet einen Mechanismus zur Aktualisierung der Firmware vor dem Onboarding, um die Sicherheit zu gewährleisten. Es stellt auch sicher, dass keine veraltete und potenziell unsichere Firmware in das Netzwerk integriert wird.

7.2 User-driven Onboarding

Eine ausschließliche Konzentration auf die Unterstützung von Zero-Touch Onboarding Protokollen wird in der Praxis aus verschiedenen Gründen als wenig praktikabel erachtet.

Fabrikbetreiber, die Trustpoint nutzen, sehen sich der Herausforderung gegenüber, dass in ihren Umgebungen sowohl ältere als auch neuere Geräte betrieben werden. Einige dieser Geräte unterstützen teilweise nicht notwendige Komponenten wie beispielsweise BRSKI oder werden als unsicher betrachtet, da sie aktuelle Cipher-Suiten nicht unterstützen.

Ein weiterer Aspekt ist, dass derzeit kein Hersteller Zero-Touch mit BRSKI oder OPC UA Part 21 unterstützt. Selbst wenn einzelne Vorreiter in den kommenden Jahren auftauchen sollten, werden diese sich wahrscheinlich auf geschlossene Systeme konzentrieren, bei denen nahezu ausschließlich auf Komponenten von einem Hersteller zurückgegriffen wird. Dies steht im Widerspruch zu vielen industriellen Umgebungen, in denen die in Maschinen verbauten Komponenten oft von verschiedenen Herstellern stammen. Falls ein Hersteller diese Mechanismen nicht unterstützt, erfordert dies Sonderlösungen, um BRSKI vollumfänglich nutzen zu können.

Obwohl Zero-Touch hochautomatisierte Mechanismen bietet, ist es aus verschiedenen Gründen oft nicht möglich oder gewollt. Eine der Hauptursachen ist die Abhängigkeit von bestimmten Protokollen. Zum Beispiel erfordert BRSKI einen Online-Service des Geräte-Herstellers. In anderen Umgebungen sind nur Offline-Systeme verfügbar, entweder aus technischen Gründen oder aufgrund von Sicherheitsbedenken des Betreibers.

Das Konzept des benutzergesteuerten Onboardings ist ein zentraler Bestandteil von Trustpoint. Es zielt darauf ab, eine flexible und anwenderfreundliche Lösung für die Verwaltung digitaler Identitäten von Maschinen

und Komponenten in industriellen Umgebungen bereitzustellen. Im Folgenden werden die Hauptaspekte dieses Konzepts erläutert:

Das übergeordnete Ziel des benutzergesteuerten Onboardings besteht darin, Geräte mit einem LDevID-Zertifikat auszustatten, welches als Grundlage für die Ausstellung weiterer Anwendungszertifikate dient. Dieser Prozess ermöglicht eine sichere und effiziente Integration von Geräten in industrielle Netzwerke.

Trustpoint nutzt hierzu ein Trustpoint-Registrar-Modul, das Administratoren über die grafische Benutzeroberfläche (GUI) die Konfiguration und Durchführung von benutzergesteuerten Onboarding-Prozessen ermöglicht. Dieses Modul fungiert als zentraler Verwalter für die Durchführung und Überwachung des Onboarding-Prozesses.

8 Konzeptstudie: Demonstrator für User driven Onboarding

Trustpoint bietet eine Client-Software an, die den benutzergesteuerten Onboarding-Prozess erleichtert. Die Software fungiert als Wrapper um PKI- und Krypto-Protokolle und ermöglicht die komfortable Durchführung von Onboarding-Aufgaben sowie die Verwaltung von Zertifikaten auf dem Gerät.

Für Geräte, die die Trustpoint-Client-Software installieren können, wird ein komfortablerer Onboarding-Prozess angeboten. Die Client-Software kann flexibel entweder vom Trustpoint selbst oder aus verschiedenen Repositories bezogen werden.

In Situationen, in denen die Installation der Trustpoint-Client-Software nicht möglich ist, bietet Trustpoint ein manuelles Onboarding über eine Kommandozeilenschnittstelle mittels gängiger Linux-Bordmittel, ohne zusätzliche Software zu installieren. Dabei generiert Trustpoint die notwendigen Befehle für eine höhere Nutzerfreundlichkeit und stellt sie in der Trustpoint-GUI dar. Wenn das Gerät SSH unterstützt, können die Befehle auch automatisiert darüber übertragen werden.

Abbildung 2 zeigt den mehrstufigen Prozess des Onboardings bzw. der Vertrauensbildung. Zu Beginn besteht kein Vertrauen. Im ersten Schritt wird das öffentliche TLS-Zertifikat des Trustpoints über einen vertrauenswürdigen Kanal auf das Gerät geladen. Hierfür können verschiedene Mechanismen verwendet werden. Der Truststore kann direkt zusammen mit dem Trustpoint-Client auf das Gerät übertragen werden. Dies bietet sich an, wenn die Client-Software über einen sicheren Kanal direkt vom Trustpoint bezogen wird. Alternativ kann er auch von einem USB-Speicher eingelesen werden. Ohne zu-

sätzliche Hardware ist es möglich, den Truststore direkt per HTTP vom (angeblichen) Trustpoint herunterzuladen und dessen Integrität bzw. Echtheit mittels einer kryptographischen Prüfsumme wie SHA2 sicherzustellen. Um einen höheren Sicherheitswert zu erreichen und um zu vermeiden, dass der gesamte Hash abgeglichen werden muss, kann das PBKDF2-HMAC-Schema[17] verwendet werden. Dadurch wird ein vom Trustpoint generiertes Einmalpasswort erforderlich, um den Truststore zu signieren und zu überprüfen. Dieses Passwort muss manuell vom Administrator in das Gerät eingegeben werden.

Im zweiten Schritt wird ein asymmetrisches Schlüsselpaar und eine Zertifikatssignierungsanfrage (CSR) erzeugt. Dabei wird das zuvor heruntergeladene Serverzertifikat verifiziert. Diese Anfrage wird im dritten Schritt über eine HTTPS POST-Anfrage an Trustpoint oder die in die Software integrierte Registrierungsstelle (RA) gesendet. Das Gerät authentifiziert sich gegenüber dem Server über HTTP Basic Auth mithilfe eines Einmalpassworts (OTP), welches zuvor über einen sicheren Kanal übertragen wurde. Wenn die Verifikation erfolgreich ist, stellt die in Trustpoint integrierte oder bei Bedarf auch externe ausstellende Zertifizierungsstelle (CA) ein Zertifikat (LDevID - Locally Significant Device Identifier nach IEEE 802.1AR) für das Gerät aus. Zu diesem Zeitpunkt wurde das gegenseitige Vertrauensverhältnis zwischen dem Gerät und Trustpoint hergestellt.

Als letzten Schritt (4) ruft das Gerät die vollständige Vertrauenskette des LDevID-Zertifikats ab. Hierfür wird eine HTTPS GET-Anfrage genutzt, bei der der Server bereits verifiziert ist. Zur Authentifizierung des Geräts gegenüber dem Trustpoint wird das soeben erhaltene LDevID-Zertifikat im Rahmen der TLS-Clientverifizierung verwendet.

Dieses grundlegende Onboarding-Schema soll u.a. noch um Mechanismen zur automatischen Zertifikatserneuerung und zur Prüfung von Zertifikatssperrlisten erweitert werden.

[17] Secure Zero Touch Provisioning (SZTP), Internet Engineering Task Force (IETF), Request for Comments: 8572, K. Watsen, I. Farrer, M. Abrahamsson, April 2019

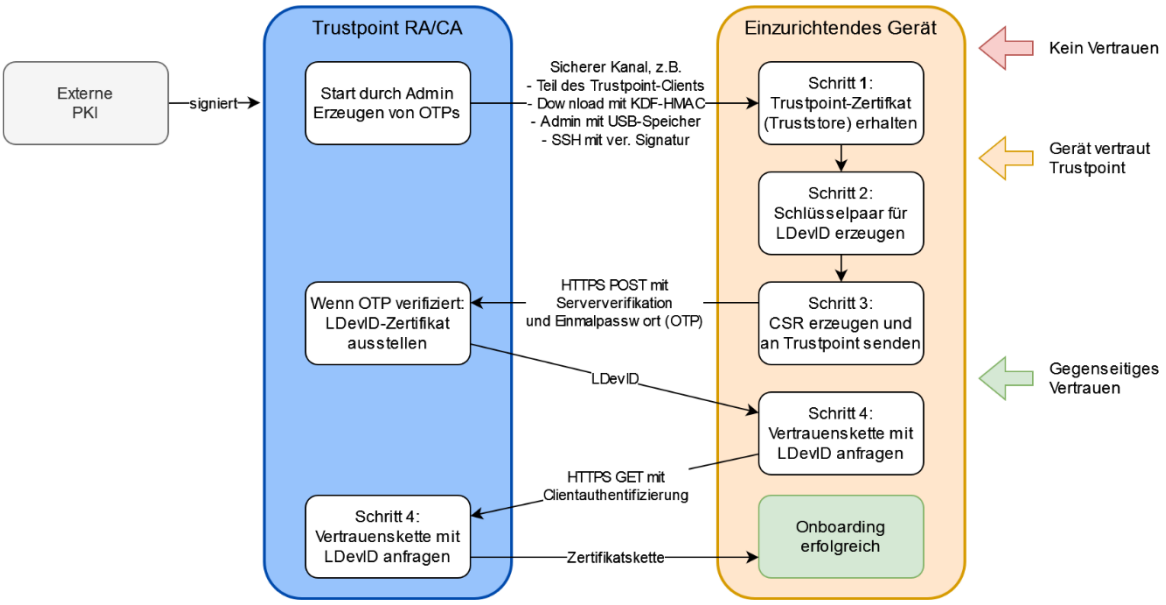


Abbildung 2: User-driven-Onboarding-Schema

9 Zusammenfassung und Ausblick

Ziel dieses Beitrages war die Vorstellung erster Arbeiten für eine benutzerfreundliche Open Source Lösung zur sicheren Verwaltung digitaler Identitäten von Maschinen und Komponenten im industriellen Umfeld. Dabei wurden Anforderungen existierender Standards wie OPC UA Part 21 und BRSKI an Hersteller, Integratoren, Betreiber und Anwender sowie den Produktlebenszyklus beschrieben und mögliche Erweiterungen zur Steigerung des Sicherheitswerts analysiert. Für bestehende Systeme haben wir bereits in einer Konzeptstudie den Anwendungsfall 'User driven Onboarding' beispielhaft umgesetzt.

Darüber hinaus leistet das Projekt einen wichtigen Beitrag zur sicheren Gestaltung des digitalen Wandels im industriellen Umfeld. Damit wird nicht nur die Zukunftsfähigkeit des Industriestandortes Deutschland gestärkt, sondern auch die digitale Souveränität der Unternehmen gewährleistet.

Die Bedeutung digitaler Identitäten als Schlüsselelement für die eindeutige Identifizierung industrieller Geräte, die Authentisierung von Komponenten gegenüber Diensten und die Absicherung der Kommunikation ist anerkannt. In diesem Umfeld einer zunehmend vernetzten und digitalisierten Industrie spielen digitale Identitäten eine entscheidende Rolle.

Zu einem modernen Industriesystem gehört neben der physischen Existenz von Objekten wie Fertigungsmaschinen und ihren Komponenten auch deren virtuelle Abbildung – dies ist auch unter dem Begriff “Digitaler Zwilling“ bekannt. Diese virtuelle Repräsentation von Objekten neben ihrer materiellen

Existenz ist auch als Anforderung des Referenzarchitekturmodells Industrie 4.0 (RAMI 4.0)[18] beschrieben. Bisher fehlt jedoch eine umfassende virtuelle Abbildung für digitale Identitäten. Diese Lücke will Trustpoint schließen und im Rahmen der Asset Administration Shell (AAS) eine generische Definition für digitale Identitäten etablieren. Die Schaffung einer generischen Definition für digitale Identitäten im Rahmen von Trustpoint wird nicht nur den Anforderungen von Industrie 4.0 gerecht, sondern leistet auch einen wichtigen Beitrag zur Standardisierung und Interoperabilität im industriellen Umfeld.

Trustpoint nimmt mit der Unterstützung von Zero-Touch-Onboarding-Verfahren eine Vorreiterrolle im digitalen Identitätsmanagement in industriellen Umgebungen ein. Durch die Vereinfachung der notwendigen Konfigurationshandlungen und die damit einhergehende Reduktion des Risikos für menschliche Fehler leistet Trustpoint einen wichtigen Beitrag zur Verbesserung der Sicherheit in für die Wirtschaft essenziellen Produktionsanlagen.

Trustpoint wurde im September 2023 als Verbundprojekt des Campus Schwarzwald, mittelständischer Unternehmen (PrimeKey Labs GmbH / Keyfactor, asvin GmbH, achelos GmbH) und der Hochschule Hamm-Lippstadt gestartet und wird von namhaften Unternehmen wie ARBURG GmbH + Co KG, HOMAG GmbH, FANUC Deutschland GmbH, PHOENIX CONTACT GmbH & Co. KG und der Siemens AG als assoziierte Partner unterstützt. Ermöglicht wird diese Kooperation durch das Bundesministerium für Bildung und Forschung im Rahmen der Initiative KMU-innovativ mit einer Fördersumme von 1,41 Millionen Euro.

[18] DIN SPEC 91345:2016-04, Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Einfach quantensicher: Integration von Post-Quanten-Kryptografie mit der Kryptobibliothek Botan

Fabian Albert[1], René Meusel[1], Dr.-Ing. Amos Treiber[1]

Kurzfassung:
Um den Angriffsmöglichkeiten eines zukünftigen Quantencomputers entgegenzutreten, ist die reibungslose Integration von Post-Quanten-Kryptografie (PQC; engl. Post-Quantum Cryptography) unabdingbar. In diesem Beitrag zeigen wir auf, wie eine solche Integration schon heute mithilfe der Kryptobibliothek Botan erfolgen kann. Dazu geben wir eine Übersicht über Botan sowie die enthaltenen PQC-Fähigkeiten und zeigen den Einsatz explizit für das Beispiel der Absicherung der Transportschicht mittels des TLS-Protokolls auf, inklusive einer Analyse der resultierenden Performance.

Stichworte: Botan, Kryptobibliothek, Post-Quanten-Kryptografie, Quantencomputer, TLS, Transport Layer Security

1 Einführung

Die potenziellen Gefahren, die Quantencomputer für moderne kryptografische Verfahren darstellen, werden in sicherheitsrelevanten Betrachtungen immer deutlicher. Es sind zwar noch keine kryptografisch relevanten Quantencomputer verfügbar, doch eine adäquate Migration zu Post-Quanten-Kryptografie (PQC; engl. Post-Quantum Cryptography) sollte bereits heute vorbereitet werden [1], sodass zum Zeitpunkt des Einsatzes eines solchen Quantencomputers weiterhin Sicherheit gewährleistet werden kann. Davon abgesehen besteht schon jetzt die Gefahr, dass die heutige verschlüsselte Kommunikation aufgezeichnet und erst später gebrochen wird, sobald der Quantencomputer verfügbar ist (store-now-decrypt-later). So wurde bereits kürzlich PQC in die Messengerdienste von Signal und Apple integriert.

Trotzdem sehen nur 11% der Teilnehmenden einer Marktumfrage deutscher Unternehmen von KPMG und dem BSI eine Chance, rechtzeitig quantensicher zu werden [2]. Denn im Gegensatz zur traditionellen Kryptografie geht PQC mit vielen neuen Aspekten wie längeren Schlüsseln einher, die eine Nutzung und Umstellung erheblich erschweren. Daher ist es unabdingbar, als Fundament für langfristig sichere Technologien eine einfach nutzbare Krypto-Basiskomponente bereitzustellen, die eine möglichst reibungslose Integration

[1] Rohde & Schwarz Cybersecurity GmbH, München

von PQC und damit die Migration zu quantensicheren Technologien ermöglicht. Damit sie weitläufig genutzt und analysiert werden kann, sollte diese grundlegende Softwarebibliothek quelloffen (Open-Source) sein.

Um Basistechnologien zu ermöglichen, haben Standardisierungsgremien neue Standards für quantensichere Verschlüsselungen, Signaturen und Sicherheitsprotokolle wie TLS vorgeschlagen. Zwar sind die dazugehörigen Prozesse größtenteils noch nicht abgeschlossen, sie befinden sich allerdings in den finalen Schritten der Standardisierung. Da nur noch wenige Änderungen der Standards zu erwarten sind und der nötige Vorbereitungsaufwand einer Migration zu PQC erheblich ist, sollten Krypto-Softwarebibliotheken schon jetzt diese neuen Algorithmen und Protokolle integrieren und nutzbar machen. Somit können Organisationen zeitnah deren Einsatz erproben und diesen nach der baldig abgeschlossenen Standardisierung direkt umsetzen.

Aus den oben genannten Gründen hat das BSI mit dem Auftragnehmer Rohde & Schwarz Cybersecurity GmbH Projekte zur Entwicklung einer sicheren Kryptobibliothek für hohen Sicherheitsbedarf durchgeführt [3]. Darin wurde die Open-Source-Kryptobibliothek Botan [4] untersucht und entsprechend weiterentwickelt. Kürzlich wurden in diesem Rahmen neue PQC-Verfahren und -Protokolle integriert, um Quantensicherheit zu gewährleisten.

Unser Beitrag zeigt anhand des weit verbreiteten TLS 1.3 Protokolls auf, wie Anwender schon jetzt mit Botan Kommunikation mit quantensicherer Vertraulichkeit ausstatten können. Dabei stellen wir eine aktuelle und kompakte Übersicht von Botan und der enthaltenen PQC-Fähigkeiten bereit. Eine Schritt-für-Schritt Anleitung im Rahmen einer beispielhaften Case Study zeigt, wie mit Botan eine quantensichere Version des TLS 1.3 Protokolls integriert wird. Zudem analysieren wir das Protokoll bzgl. Performance.

Unsere Hoffnung ist, dass der Beitrag den akuten Handlungsbedarf bezüglich PQC-Integration verdeutlicht und gleichzeitig aufzeigt, wie eine Integration schon jetzt begonnen werden kann.

2 Post-Quanten-Kryptografie und TLS 1.3

Das Faktorisierungsproblem, eine mathematische Annahme des RSA-Verfahrens, oder der diskrete Logarithmus des Diffie-Hellman-Verfahrens können durch Shors Quantenalgorithmus effizient (polynomiell) gelöst werden. PQC umfasst daher asymmetrische[2] Kryptografie, bei der im Gegensatz

[2] Zwar kann auch die symmetrische Kryptografie durch Quantenalgorithmen angegriffen werden, doch die davon ausgehende Gefahr wird als wesentlich geringer eingeschätzt, da durch den hierfür bekannten Algorithmus von Grover keine exponentielle Verbesserung der Schwierigkeit gegeben wird. Eine einfache Verdoppelung der Schlüssellängen kann die Gefahren durch Grovers Algorithmus abdecken.

zur aktuell eingesetzten, „traditionellen“ Kryptografie nicht davon auszugehen ist, dass sie durch einen Quantencomputer gebrochen werden kann. PQC war lange Zeit nur Teil der Forschung, wird jedoch derzeit standardisiert und mittel- bis langfristig zum allgemeinen Standard werden.

2.1 Stand der Standardisierung

Im Standardisierungsprozess der NIST [5] wurde bereits eine Auswahl an PQC-Verfahren getroffen. Davon abgesehen gibt es weitere IETF- und ISO-Standards bzw. -Standardvorschläge. Wir stellen hier nur die relevantesten Standards vor und verweisen auf [6] für eine vollständigere Übersicht.

Der NIST-Prozess [5] steht vor der finalen Veröffentlichung der ersten Standards innerhalb des Jahres 2024. Es werden statt Schlüsselaustauschverfahren (KEXs; engl. Key EXchange Mechanisms) Schlüsselkapselungsverfahren (KEMs; engl. Key Encapsulation Mechanisms) betrachtet. Diese verfügen zusätzlich zur Schlüsselgenerierung über die Operationen `Encapsulate` (Generierung eines Geheimnisses und eines entsprechenden Chiffrats mittels öffentlichen Schlüssels) und `Decapsulate` (Erhalten des Geheimnisses mittels Chiffrats und geheimen Schlüssels). Es werden mehrere Verfahren standardisiert: Kyber/ML-KEM [7] als KEM sowie Dilithium/ML-DSA [8] und SPHINCS+/SLH-DSA [9] als Signaturverfahren. Später wird auch die Falcon Signatur [10] standardisiert werden. In der Auswahl befinden sich noch Classic McEliece [11], BIKE [12] und HQC [13]. Zudem hat die NIST einen neuen Wettbewerb für weitere Signaturverfahren gestartet [5].

Außerhalb des NIST-Prozesses wurden die zustandsbehafteten Signaturalgorithmen LMS/HSS [14] und XMSS/XMSS^MT [15] in der IETF und anschließend auch von der NIST standardisiert. Zudem befindet sich der aus dem NIST-Prozess ausgeschiedene Algorithmus FrodoKEM in einem Standardisierungsprozess der ISO [16]. Für das TLS 1.3 Protokoll gibt es in der IETF einen von der TLS Working Group adoptierten Vorschlag zur Nutzung von KEMs [17], den wir im folgenden Abschnitt 2.2 genauer erläutern.

2.2 Transport Layer Security (TLS) und die Adoption von PQC

Das Transport Layer Security (TLS) Protokoll erlaubt es, beispielsweise über das Internet, geschützt zu kommunizieren. Es stellt die zwei maßgeblichen Schutzziele Vertraulichkeit und Authentizität bereit, indem asymmetrische Kryptografie während des Verbindungsaufbaus (Handshake) und symmetrische Kryptografie während der eigentlichen Kommunikation eingesetzt werden. Mittels Signaturen und einer PKI wird beim Handshake Authentizität

etabliert. Für die Vertraulichkeit erschaffen die Teilnehmer ein geteiltes Geheimnis mittels eines *asymmetrischen Schlüsselaustauschverfahrens*, das einen symmetrischen Schlüssel für die spätere Kommunikation ergibt.

Gefahren durch Quantencomputer. Die heute in TLS eingesetzten traditionellen Schlüsselaustauschverfahren sind bereits jetzt den Gefahren eines Quantencomputers ausgesetzt (store-now-decrypt-later). Im Gegensatz dazu ist „store-now-spoof-later" generell keine relevante Angriffsstrategie gegen Authentizität[3]. Daher fokussieren wir uns in diesem Beitrag nur auf die Vertraulichkeit gegen einen zukünftigen Quantencomputer.

Bisherige Schlüsselinformationen in TLS 1.3. In der aktuellen Revision TLS 1.3 [18] enthält die initiale `ClientHello` Nachricht Algorithmus-IDs („Code Points") und die eigentlichen kryptografischen Bytes. Der Server wählt davon eine ID aus und antwortet mit seinen Schlüsselaustausch-Informationen (`ServerHello`). Dabei ist die konkrete Beschaffenheit der kryptografischen Bytes für die Nachrichten nicht relevant; das Protokoll ist nicht an bestimmte Verfahren gebunden (Kryptoagilität), sodass ein Wechsel von einem traditionellem KEX zu einem Post-Quanten KEM möglich ist.

Die PQ-Integration des Standard-Drafts von [17] stellt genau diesen Wechsel bereit, indem in die kryptografischen Bytes auch Informationen aus PQC-KEMs einfließen können. Allerdings wird empfohlen, die Post-Quanten-KEMs für voraussehbare Zeit nur *zusätzlich* zu einem traditionellen KEX einzusetzen („hybride Verfahren") [19, pp. 42-45]. Die Spezifikation in [17] ermöglicht dies, indem jeder konkreten Algorithmen-Kombination *eine* konkrete Algorithmus-ID zugeordnet und die kryptografischen Bytes sowie geteilten Geheimnisse beider Verfahren schlicht aneinandergefügt werden[4]. Diese Konstruktion stellt PQ-Vertraulichkeit sicher und wird mit ausgewählten Algorithmus-Kombinationen[5] in der Praxis bereits erprobt [20].

---

[3] Unberührt davon gibt es andere Anwendungsszenarien mit „langlebigen" Signaturen (Schutzziel: Nichabstreitbarkeit), die auch heute schon einer vergleichbaren Bedrohung durch Quantencomputer ausgesetzt sind.

[4] Das einfache Aneinanderfügen der Geheimnisse beider Algorithmen ergibt die gewünschten hybriden Eigenschaften erst durch die konkrete Konstruktion der Schlüsselableitung in TLS 1.3 und ist nicht allgemeingültig [28].

[5] Meist eine Kombination aus X25519 (Diffie-Hellman auf elliptischer Kurve) und dem Post-Quanten Verfahren Kyber768.

3 Die Kryptobibliothek Botan und PQC

Botan [4] wurde im Jahr 2001 (zunächst unter dem Namen OpenCL) veröffentlicht. Es ist open-source lizensiert (Simplified BSD), im Gegensatz zu vielen anderen prominenten Kryptobibliotheken in C++ implementiert und hat den Anspruch einer entsprechend einfachen Nutzbarkeit.

3.1 Funktionsumfang

Botan verfügt über alle wichtigen Hash-Funktionen, Strom- und Blockchiffren, Chiffren-Modi und MAC-Algorithmen und unterstützt TLS 1.2 und TLS 1.3 sowie die meisten Public-Key-Standards und traditionellen Public-Key-Algorithmen. Die vollständige Liste ist auf [21] einzusehen.

PQC-Algorithmen werden bereits heute angeboten: Aktuell werden Kyber/ML-KEM, FrodoKEM, Dilithium/ML-DSA, SPHINCS+/SLH-DSA und XMSS unterstützt. HSS/LMS und Classic McEliece sind bereits implementiert und werden voraussichtlich bald in den Hauptzweig integriert. Auch PQC-TLS [17] wird mit vorläufigen „Code Points“ der genannten KEMs angeboten.

3.2 Sicherheit

Botan verwendet verschiedene Mechanismen zur Erhöhung der Sicherheit.

Policies. Um bei der Auswahl der richtigen Algorithmen zu helfen, bietet Botan ein Policy-System [22] an, beispielsweise um nur diejenigen Algorithmen zu aktivieren, die den technischen Richtlinien des BSI [19] entsprechen.

Dokumentation. Botan verfügt über eine zusätzliche Dokumentation der Kryptoalgorithmen und ihrer Implementierungen [23], die in Zusammenarbeit mit dem BSI erstellt wurde. Auf Botans Website werden außerdem gefundene und behobene Sicherheitslücken aufgeführt [24].

Speichersicherheit. Botan verzichtet weitgehend auf die Verwendung von rohen Zeigern und setzt stattdessen Hilfsfunktionen mit zusätzlichen Überprüfungen ein, um potenziell unsichere C++-Funktionen abzusichern.

Tests. Zudem verfügt Botan über eine umfangreiche Testsuite, die eine Testabdeckung von etwa 92% bietet [21].

3.3 PQC-Performance

Um ein Verständnis für die Performanceveränderung durch die Nutzung von PQC zu vermitteln, messen wir hier die PQ-Algorithmen in Botan.

Laufzeit. Wir nutzen Botans `speed` Tool [22], um lokal Laufzeiten der verfügbaren PQ-Algorithmen[6] mit dem prominent genutzten traditionellen Algorithmus ECDH-brainpoolP384r1 zu vergleichen. Der traditionelle Algorithmus ist zwecks Seitenkanalresistenz maskiert und erreicht das NIST-Sicherheitslevel 3 [25], während pro PQ-Algorithmus jeweils die NIST-Sicherheitslevel 1 (AES-128), 3 (AES-192) und 5 (AES-256) repräsentiert sind. Unsere Experimente fanden auf einem Intel Core i7-1165G7 (4 Kerne, Threads, 2.8 GHz) mit 16 GB RAM statt und sind in Abbildung 1 dargestellt. Die evaluierten Algorithmen nutzen hierbei keine Hardwarebeschleunigung.

Es zeigt sich, dass die Laufzeit von Kyber durchaus mit den traditionellen Algorithmen mithalten kann und in Botan sogar schneller ist als beispielsweise ECDH-brainpoolP384r1. FrodoKEM liegt hier ca. eine Größenordnung hinter ECDH. Alle gemessenen Algorithmen sind reine Software-Implementierungen ohne die Verwendung plattform-spezifischer Hardware-Erweiterungen.

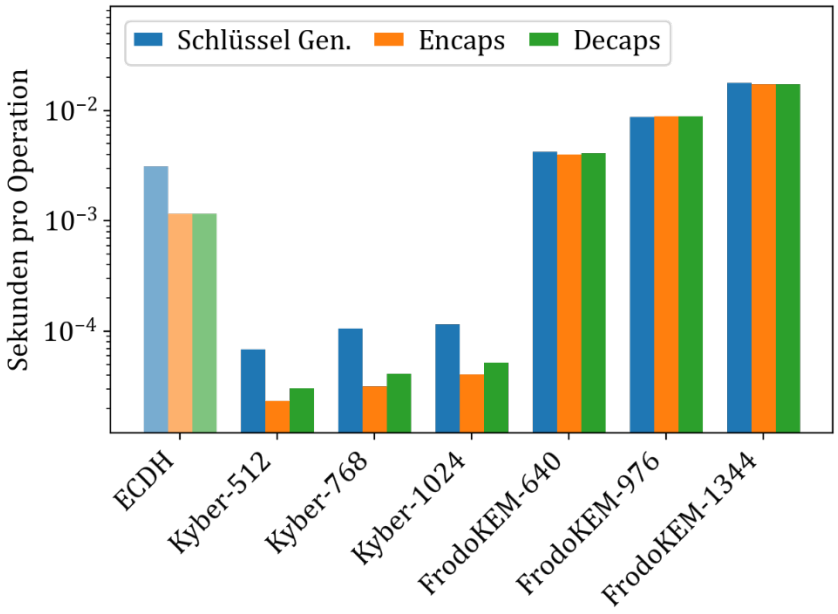


Abbildung 1: Laufzeitmessung von ECDH-brainpoolP384r1 und PQ-Algorithmen, jeweils für Schlüsselgenerierung, Encaps (bzw. Schlüsselvereinbarung für ECDH) und Decaps (bzw. Schlüsselvereinbarung für ECDH). Die Laufzeiten pro Operation sind auf einer logarithmischen Skala dargestellt.

[6] Wir haben hier nur die FrodoKEM-SHAKE Variante berücksichtigt, die AES-Variante ist jedoch auch in Botan verfügbar.

Die Größen des Schlüssels und des Chiffrats sind ebenfalls entscheidende Metriken, da sie bei einem Austausch übermittelt werden müssen. In Abbildung 2 werden die Größen der öffentlichen Schlüssel und Chiffrate der PQ-Algorithmen im Vergleich zu den traditionellen Algorithmen dargestellt. Hier zeigt sich die bisher größte Schwäche der PQ-Algorithmen: Ihre Schlüssel und Chiffrate sind um ein Vielfaches größer als die von ECDH.

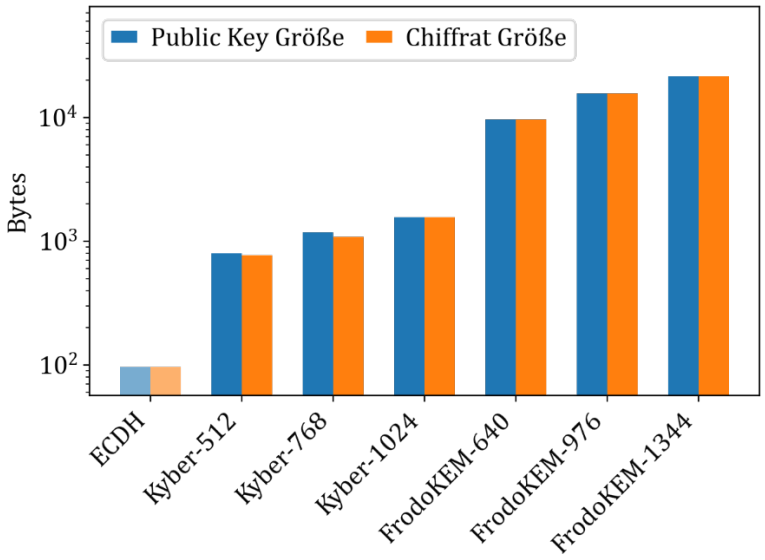


Abbildung 2: Vergleich der Größen der öffentlichen Schlüssel und Chiffrate (Schlüsselmaterial für ECDH) von ECDH-brainpoolP384r1 und PQ-Algorithmen. Die Größenangaben erfolgen in Bytes auf einer logarithmischen Skala.

4 Case Study: Post-Quanten TLS mit Botan

Die in Abschnitt 2 erwähnten PQC-Standards können mit der in Abschnitt 3 vorgestellten Bibliothek Botan eingesetzt werden, inklusive des quantensicheren TLS 1.3 [17]. In diesem Abschnitt zeigen wir anhand einer Case Study detaillierter auf, wie genau das Protokoll mit Botan genutzt werden kann.

Dabei gehen wir von einem typischen Webanwendungsfall aus: Ein Client (Webbrowser) verbindet sich mit einem Webserver, um eine Webseite unter einer Domain abzurufen. Zum Beispiel: pq.botan-crypto.org.

4.1 Mit Botan und Boost: Ein minimaler HTTPS Client

Wir schreiben einen HTTPS-Client, der Botan zum Verbindungsaufbau zu einem Server nutzt, wobei wir bei diesem auch entsprechende Protokoll-Konfigurationen vornehmen müssen, um es mit PQC zu verwenden.

Client-Programm zum Verbindungsaufbau. In Quelltext 1 präsentieren wir einen Ausschnitt aus einem C++20 Beispiel in Botan, das eine TLS-

Verbindung zum Server `host` aufbaut, eine HTTP GET-Anfrage auf die Ressource `http_target` absetzt und die Antwort auf der Konsole ausgibt. Für das vollständige Programm verweisen wir auf Botans Repository[7].

Botan liefert hier die nötigen kryptografischen Algorithmen und TLS 1.3, Boost ASIO die asynchrone Netzwerk-Kommunikation mittels C++20 Co-Routinen und Boost Beast die HTTP-Implementierung. Falls gewünscht kann Botans TLS Implementierung auch in beliebige Kommunikationsarchitekturen mittels einer Callback-Schnittstelle integriert werden.

Die genutzte Funktion `prepare_tls_context_for()` konfiguriert Botans TLS-Implementierung und wird im folgenden Paragraphen genauer erklärt.

Konfiguration. Botans TLS Implementierung kann mit einem Text-Format (der „Policy") konfiguriert werden. Eine Applikation kann zusätzlich jede der Klassen im in Quelltext 2 gezeigten `TLS::Context` Objekt überschreiben und so das Verhalten des Protokolls feingranular beeinflussen [26]. Dies erlaubt beispielsweise die Zertifikatsvalidierung durch proprietäre Logik zu ersetzen, einzelne TLS-Nachrichten zu inspizieren oder um eigene Extensions zu erweitern oder spezielle Kryptohardware zum Schlüsselaustausch oder die Authentisierung einzubinden.

Der gezeigte `TLS::Context` verwendet im Wesentlichen Botans Standard-Implementierungen der Kontext-Klassen. Mit der Text-Konfiguration in der Variable `cfg` schalten wir TLS 1.2 ab[8] und erlauben in diesem Beispiel neben der traditionellen Brainpool-Kurve, die wir zu Migrationszwecken erst einmal weiterhin erlauben, das gewünschte hybride und PQ-sichere Schlüsselaustauschverfahren auf Basis von X25519 und Kyber768. Andere Verfahren sind mit dieser Konfiguration explizit verboten.

[7] Beispiel: github.com/randombit/botan/blob/72ad0a8/src/examples/tls_stream_coroutine_client.cpp

[8] Botan implementiert den hybriden Schlüsselaustausch wie in [17] spezifiziert. Dieser ist aber nicht kompatibel mit TLS 1.2.

```
namespace net = boost::asio;
namespace tls = Botan::TLS;
namespace http = boost::beast::http;


net::awaitable<void>
https_request(std::string host, std::string port, std::string http_target) {
   // Prepare a Botan TLS stream on top of an ASIO TCP stream
   auto stream =
      tls::Stream(prepare_tls_context_for(host, port),
                  create_tcp_stream(co_await net::this_coro::executor));

   // Connect to host and establish a TLS session
   co_await stream.next_layer().async_connect(co_await resolve(host, port));
   co_await stream.async_handshake(tls::Connection_Side::Client);

   // Send HTTP GET request
   co_await http::async_write(stream, HTTP_GET_request(host, http_target));

   // Receive HTTP response and print result
   beast::flat_buffer response_buffer;
   http::response<http::dynamic_body> response;
   co_await http::async_read(stream, response_buffer, response);
   std::cout << response << std::endl;

   // Gracefully terminate TLS session and connection
   co_await stream.async_shutdown();
   stream.next_layer().close();
}
```

Quelltext 1: Aufbau der Verbindung zu einem Server `host`.

```
std::shared_ptr<tls::Context>
prepare_tls_context_for(const std::string& host, const std::string& port) {
   auto cfg = "allow_tls_12 = false\n"
              "allow_tls_13 = true\n"
              "key_exchange_groups = x25519/Kyber-768-r3 brainpool256r1";
   return std::make_shared<tls::Context>(
                  std::make_shared<System_Credentials_Manager>(),
                  std::make_shared<Botan::AutoSeeded_RNG>(),
                  std::make_shared<tls::Session_Manager_Noop>(),
                  std::make_shared<tls::Text_Policy>(cfg),
                  tls::Server_Information(host, "https", std::stoi(port)));
}
```

Quelltext 2: Konfiguration von PQC-TLS 1.3 in Botan.

Ausführung. Die gezeigte Anfrage an `pq.botan-crypto.org`[9] baut damit erfolgreich eine Post-Quanten-sichere TLS 1.3 Verbindung auf und fragt über eine REST-Schnittstelle die Verbindungsdetails ab:

```
$> ./https_client pq.botan-crypto.org 443 /api/connection_details
HTTP/1.1 200 OK
Server: Botan 3.3.0
Content-Type: application/json
Content-Length: 78

{"kex_algo":"x25519/Kyber-768-r3","is_quantum_safe":true,"wire_code":"0x6399"}
```

Dieser X25519/Kyber-768-R3 [17] [27] Schlüsselaustausch ist kompatibel mit anderen experimentellen Implementierungen, wie z.B. von Cloudflare.

4.2 Der quantensichere TLS 1.3 Handshake

Im Folgenden gehen wir auf die technischen Details des in Abschnitt 4.1 genutzten und ausgeführten quantensicheren TLS 1.3 Handshakes ein, um zu veranschaulichen, was die gezeigte Umstellung auf PQC mit sich bringt.

Genereller Ablauf. Nach dem Verbindungsaufbau führen Client und Server den in Abbildung 3 dargestellten *einseitig authentisierten* TLS 1.3 Handshake durch. Im ersten Schritt wird mittels eines geeigneten Schlüsselaustauschverfahrens und durch den Austausch von `ClientHello` und `ServerHello` Nachrichten ein gemeinsamer symmetrischer Schlüssel etabliert. Der Rest des Handshakes findet dann bereits verschlüsselt statt. Nun weist sich der Server mit seinem Zertifikat (`Certificate`) und einer dazugehörigen digitalen Signatur (`CertificateVerify`) als legitimer Eigentümer der angefragten Domain aus. Der Client bleibt in einem typischen Web-Anwendungs-

[9] Ein Test-Server, der auch Botan und Boost einsetzt: https://github.com/reneme/botan-tls-testserver

Abbildung 3: Betrachteter TLS-Handshake

fall unauthentisiert und präsentiert kein Zertifikat. Die verbleibenden Nachrichten in der Skizze sind für unsere Betrachtung nicht relevant. Nach Abschluss des Handshakes können beide Teilnehmer sicher kommunizieren. Die beschriebene Abstraktion trifft sowohl auf das traditionelle als auch auf das quantensichere TLS 1.3 zu.

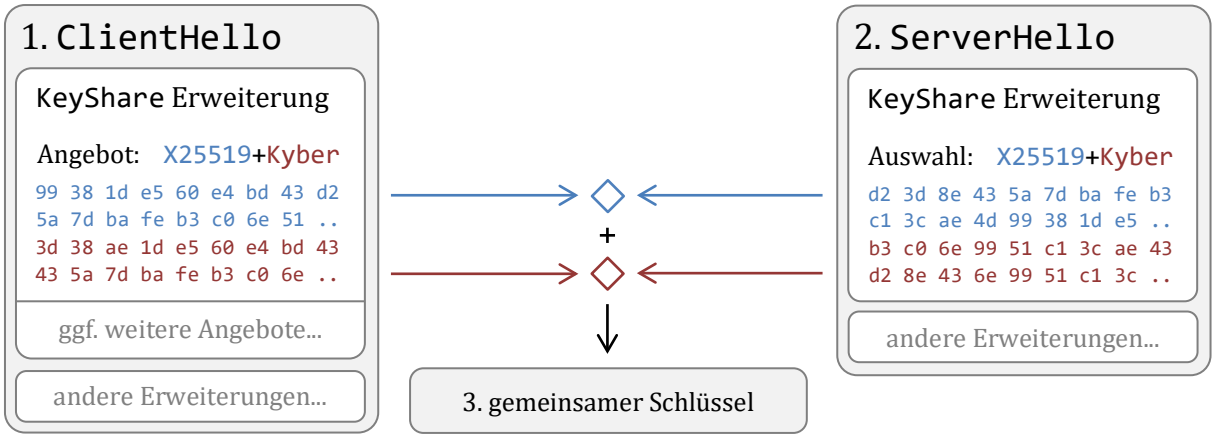


Abbildung 4: Hybrider Schlüsselaustausch in TLS 1.3

PQC-Schlüsselaustausch. Im traditionell eingesetzten Protokoll enthalten die Hello-Nachrichten nur traditionelle Kryptografie. Die „Schlüsselaus-

tausch-Information" in den Hello-Nachrichten stellt aber aus Sicht des Protokolls eine beliebige Byte-Folge dar. Damit ist die Verkettung mehrerer Verfahren wie in [17] möglich, was in Abbildung 4 beispielhaft für X25519 und Kyber768 zu sehen ist und in diesem Fall Quantensicherheit gewährleistet.

Dabei erzeugt der Client im ersten Schritt je ein kurzlebiges Schlüsselpaar für X25519 und Kyber768, verkettet die Byte-Darstellungen der öffentlichen Schlüssel und sendet diese als sein einziges Austausch-Angebot[10] in einer `ClientHello`-Nachricht zum Server.

Im zweiten Schritt erzeugt dieser ebenfalls ein kurzlebiges X25519 Schlüsselpaar, um mit dem öffentlichen X25519-Schlüssel des Clients die erste Hälfte des gemeinsamen symmetrischen Schlüssels zu erhalten. Die zweite Hälfte und das dazugehörige Chiffrat wird mit `Encapsulate` aus dem öffentlichen Kyber-Schlüssel des Clients gewonnen. Dann sendet der Server seinen öffentlichen X25519-Key und das Kyber-Chiffrat als seine Schlüsselaustausch-Information in einer `ServerHello`-Nachricht.

Der Client errechnet im dritten Schritt die erste Hälfte des gemeinsamen Schlüssels mit X25519 aus seinem geheimen Schlüssel und dem empfangenen öffentlichen Schlüssel. Das `Decapsulate` des empfangenen Kyber-Chiffrats mit dem privaten Schlüssel ergibt die zweite Hälfte. Beide Teilnehmer haben nun einen identischen symmetrischen Schlüssel, wobei aufgrund der Schlüsselableitung [18, p. 92] die Vertraulichkeit im Fall einer Schwachstelle in einem der Algorithmen weiterhin gegeben ist [28, p. Abschnitt 3.2]. Für weitere technischen Details, siehe [7] [17] [18] [27] [29].

4.3 Analyse

Die in Abschnitt 4.2 veranschaulichten zusätzlich generierten und übermittelten PQC-Informationen führen natürlich zu Performanceveränderungen. Um ein Verständnis zu vermitteln, welche Performanceveränderungen diese Umstellung auf PQC mit sich bringt, analysieren wir die Kosten des in Abschnitt 4.1 durchgeführten Handshakes. Dafür stellen wir in Abbildung 5 das Datenvolumen des gesamten TLS-Handshakes mit einer traditionellen X25519-Konfiguration der vorgestellten Konfiguration X25519/Kyber768 sowie einer weiteren Kombination X25519/FrodoKEM640 gegenüber.

[10] In TLS 1.3 kann eine ClientHello-Nachricht mehr als ein Austausch-Angebot enthalten, um dem Server eine Algorithmus-Auswahl zu ermöglichen. Clients müssen in der Praxis oft Annahmen über die unterstützten Algorithmen des Servers treffen und senden daher mehrere Schlüsselaustausch-Informationen für unterschiedliche Algorithmen. In diesem vereinfachten Beispiel haben wir uns auf ein Angebot beschränkt, mehrere (hybride) Angebote sind auch in [17] möglich.

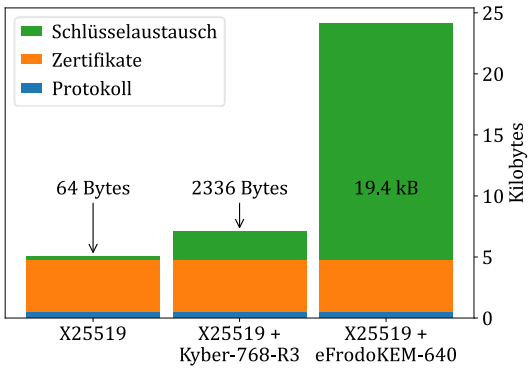


Abbildung 5: Größe des gesamten ausgeführten TLS-Handshakes in Kilobytes für verschiedene KEX/KEM-Kombinationen.

Ergebnisse. Durch das hybride Verfahren X25519/Kyber768 erhöht sich der nötige Datenaustausch in diesem Beispiel um etwa *2,3 KB*. Die nötigen Bytes für die Schlüsselaushandlung (grün) bleiben weit unter den Authentisierungsdaten[11] (orange). Der verbleibende Protokoll-Overhead (blau) enthält die Handshake-Signatur und Message Authentication Tags aus der TLS Record-Verschlüsselung. Die Kombination X25519/FrodoKEM640 hingegen bringt zusätzliche *19,4 KB* mit sich.

Der erwartete Zuwachs in Rechenaufwand ist durch Kyber nicht wesentlich erhöht, weil der Algorithmus nicht mehr Rechenzeit als ECDH-Algorithmen benötigt (siehe Performance-Messungen in Abschnitt 3.3). Bei FrodoKEM ist entsprechend mit einem erhöhtem Rechenaufwand zu rechnen. Wir verzichten bewusst auf eine konkrete Gegenüberstellung der Handshake-Latenzen. Realistische Messungen können beispielsweise in [30] in Erfahrung gebracht werden.

Durch eine technisch bedingte Längenlimitierung von *64 KiB* ist der Einsatz der meisten aktuell betrachteten Post-Quanten KEMs (auch hybrid) möglich, mit Ausnahme des konservativen Code-basierten Ansatzes Classic McEliece, welches vergleichsweise große öffentliche Schlüssel erzeugt. Es kann trotzdem nicht ausgeschlossen werden, dass „Middleboxen" oder niedrigere Netzwerkprotokollschichten Einfluss auf die letztendliche Performance haben.

5 Zusammenfassung

Die aktuell eingesetzte asymmetrische Kryptografie kann von Quantencomputern gebrochen werden. Durch die Gefahr von store-now-decrypt-later

[11] Dieses Beispiel nutzt „Let's Encrypt" mit der RSA-basierten „R3" Zertifikatskette. Es werden drei Zertifikate (End Entity, Intermediate und Cross-Signed Root) sowie eine gepinnte OCSP-Antwort für die End Entity übertragen.

sollten schon jetzt Vorkehrungen getroffen werden, Post-Quanten-Kryptografie zu integrieren. In diesem Beitrag haben wir gezeigt, dass dies schon heute mit der Kryptobibliothek Botan möglich ist, indem wir eine Anleitung für die Nutzung eines quantensicheren TLS-Protokolls aufgezeigt und anschließend analysiert haben.

Das hier betrachtete Szenario sichert allerdings nur die Vertraulichkeit gegen store-now-decrypt-later. Zwar legen neueste Studien [30] nahe, dass sich die Performanceeinbußen auch bei PQ-Authentizität ähnlich in Grenzen halten, doch auch diese Umstellung sollte gerade aufgrund PKI-Umstellungsaufwänden nicht aus den Augen verloren und schon jetzt geplant und angegangen werden – auch in anderen Kontexten als TLS.

Literaturhinweise

[1] BSI, „Kryptografie quantensicher gestalten," 2021. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Broschueren/Kryptografie-quantensicher-gestalten.html. [Zugriff am 26 01 2024].

[2] „Kryptografie und Quantencomputung," BSI und KPMG, 2023.

[3] „BSI-Projekt: Entwicklung einer sicheren Kryptobibliothek," [Online]. Available: https://www.bsi.bund.de/dok/9060550. [Zugriff am 26 01 2024].

[4] J. Lloyd, „Botan GitHub Repository," [Online]. Available: https://github.com/randombit/botan. [Zugriff am 26 01 2024].

[5] „NIST Post-Quantum Cryptography Standardization," [Online]. Available: https://csrc.nist.gov/projects/post-quantum-cryptography. [Zugriff am 26 01 2024].

[6] PQUIP IETF Working Group, „State of Protocols and PQC," 2023. [Online]. Available: https://github.com/ietf-wg-pquip/state-of-protocols-and-pqc. [Zugriff am 05 02 2024].

[7] National Institute for Standards and Technology (NIST), „Module-Lattice-Based Key-Encapsulation Mechanism Standard (FIPS 203 Initial Public Draft)," 2023.

[8] National Institute for Standards and Technology (NIST), „Module-Lattice-Based Digital Signature Standard (FIPS 204 Initial Public Draft)," 2023.

[9] National Institute for Standards and Technology (NIST), „Stateless Hash-Based Digital Signature Standard (FIPS 205 Initial Public Draft)," 2023.

[10] Fouque et al., „Falcon: Fast-Fourier Lattice-based Compact Signatures over NTRU, Specification v1.2," 01 10 2020. [Online]. Available: https://falcon-sign.info/falcon.pdf. [Zugriff am 05 02 2024].

[11] Bernstein et al., „Classic McEliece: conservative code-based cryptography: cryptosystem specification," 2022. [Online]. Available: https://classic.mceliece.org/mceliece-spec-20221023.pdf. [Zugriff am 26 01 2024].

[12] Aragon et al., „BIKE: Bit Flipping Key Encapsulation Round 4 Submission," 2022. [Online]. Available: https://bikesuite.org/files/v5.0/BIKE_Spec.2022.10.10.1.pdf. [Zugriff am 26 01 2024].

[13] Melchor et al., „Hamming Quasi-Cyclic (HQC) Updated Version," 30 03 2023. [Online]. Available: https://pqc-hqc.org/documentation.html. [Zugriff am 26 01 2024].
[14] D. McGrew, M. Curcio und S. Fluhrer, „Leighton-Micali hash-based signatures," RFC8554, 2019.
[15] A. Hülsing, D. Butin, S.-L. Gazdag, J. Rijneveld und A. Mohaisen, „XMSS: eXtended Merkle Signature Scheme," RFC8391, 2018.
[16] Alkim et al., „FrodoKEM: Learning With Errors Key Encapsulation Preliminary Standardization Proposal," 14 03 2023. [Online]. Available: https://frodokem.org/files/FrodoKEM-standard_proposal-20230314.pdf. [Zugriff am 05 02 2024].
[17] D. Stebila, S. Fluhrer und S. Gueron, „Hybrid key exchange in TLS 1.3," IETF Active Internet-Draft, 2023. [Online]. Available: https://datatracker.ietf.org/doc/draft-ietf-tls-hybrid-design.
[18] E. Rescorla, „The Transport Layer Security (TLS) Protocol Version 1.3," RFC8446, 2018.
[19] BSI, „Technische Richtlinie TR-02102-1: Kryptographische Verfahren: Empfehlungen und Schlüssellängen," 2023.
[20] D. O'Brien, „Protecting Chrome Traffic with Hybrid Kyber KEM," Google's Chromium Blog, 2023. [Online]. Available: https://blog.chromium.org/2023/08/protecting-chrome-traffic-with-hybrid.html. [Zugriff am 31 01 2024].
[21] J. Lloyd, „Botan: Crypto and TLS for Modern C++," [Online]. Available: https://botan.randombit.net. [Zugriff am 01 02 2024].
[22] J. Lloyd, D. Neus, R. Korthaus, J. Somorovsky und T. Niemann, „Botan Reference Guide," 2023. [Online]. Available: https://botan.randombit.net/handbook/botan.pdf. [Zugriff am 07 02 2024].
[23] Fischer et al., „Botan: Cryptographic Documentation," Rohde & Schwarz Cybersecurity GmbH, 01 02 2023. [Online]. Available: https://github.com/sehlen-bsi/botan-docs. [Zugriff 2024].
[24] J. Lloyd, „Botan: Security Advisories," [Online]. Available: https://botan.randombit.net/security.html. [Zugriff am 07 02 2024].
[25] NIST, „Post-Quantum Cryptography: Security (Evaluation Criteria)," 01 2024. [Online]. Available: https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization/evaluation-criteria/security-(evaluation-criteria). [Zugriff am 01 02 2024].
[26] J. Lloyd, „Botan Handbook: Transport Layer Security (TLS)," 2023. [Online]. Available: https://botan.randombit.net/handbook/api_ref/tls.html. [Zugriff am 14 02 2024].
[27] B. Westerbaan und D. Stebila, „X25519Kyber768Draft00 hybrid post-quantum key agreement," IETF Active Internet Draft, 2023.
[28] N. Bindel, J. Brendel, M. Fischlin, B. Goncalves und D. Stebila, „Hybrid Key Encapsulation Mechanisms and Authenticated Key Exchange," in Ding, J., Steinwandt, R. (eds) Post-Quantum Cryptography. PQCrypto, 2019.
[29] D. Bernstein, „Curve25519: New Diffie-Hellman Speed Records," in In: Yung, M., Dodis, Y., Kiayias, A., Malkin, T. (eds) Public Key Cryptography - PKC, 2006.
[30] P. Kampanakis und W. Childs-Klein, „The impact of data-heavy, post-quantum TLS 1.3 on the Time-To-Last-Byte of real-world connections," in MADweb, 2024.

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Die Herausforderungen für Finanzdienstleister bei der Migration zu Post-Quanten-Kryptografie (PQK)

Dr. Jan Rosam[1], Dr. Christoph Capellaro[2]

Kurzfassung:
Gerade für Finanzdienstleister stellen Quantencomputer eine ernstzunehmende Bedrohung dar. Das macht die Entwicklung einer Strategie zur Ablösung bestehender kryptografischer Verfahren erforderlich. Von den Regulatoren werden Zeiträume für die Einführung von PQK vorgegeben. Die Einsatzbereiche kryptografischer Verfahren im Finanzsektor sind vielfältig. Finanzinstitute verfügen über eine Reihe typischer Kommunikationsbeziehungen und Datenschnittstellen, deren Schutzmechanismen von unterschiedlichen Stellen festgelegt werden. Abhängig von den derzeit verwendeten kryptografischen Verfahren und ihren Einsatzumgebungen sind unterschiedliche Wege für die Migration zu PQK einzuschlagen. Hieraus ergibt sich die Notwendigkeit für die Entwicklung institutsspezifischer PQK-Migrationsstrategien und der Kooperation mit Partnern, Kunden und Behörden.

Stichworte: Banken, Finanzindustrie, Kreditinstitute, Kryptoagilität, Kryptografie, Kryptostrategie, Migrationsstrategie, Post-Quanten Kryptografie.

1 Die Bedeutung von Quantencomputern für Finanzdienstleister

Für Quantencomputer gibt es hochinteressante Anwendungsfälle, z.B. im Bereich der Lösung von Optimierungsproblemen. Es geht jedoch auch ein besonderes Risiko von dieser Technologie aus, da sie theoretisch dazu geeignet ist, aktuell gängige kryptografische Verfahren zu brechen. Insbesondere die Notwendigkeit, auch das Angriffsszenario eines „harvest now and decrypt later“ berücksichtigen zu müssen, macht eine frühzeitige organisationsspezifische Analyse von Risiken erforderlich.

2 Aktuelle Einschätzung des Fortschritts der Quantentechnologie

In den letzten 15 Jahren hat sich die Leistungsfähigkeit von Quantencomputern verhundertfacht. Einzelne Hersteller kündigen eine weitere Vertausendfachung der Zahl von Qubits eines Quantencomputers in den nächsten Jahren an. Vor diesem Hintergrund haben Standardisierungsgremien – allen voran das US-amerikanische NIST – feste Zeitpläne für den Einsatz von PQK vorgegeben. Demnach wird bereits ab dem Jahr 2025 von Softwareherstellern gefordert, dass PQK als Default in kryptografischen Softwarebibliotheken eingesetzt werden soll. Ab 2030 soll ausschließlich PQK zum Einsatz kommen.

[1] EY Consulting GmbH, Eschborn
[2] EY Consulting GmbH, München

3 Rahmenbedingungen für eine PQK-Strategie

Organisationen aus dem Finanzdienstleistungssektor sehen sich in diesem Zusammenhang vor drei wesentliche Herausforderungen gestellt. Finanzdaten haben i.d.R. einen sehr hohen Schutzbedarf. In manchen Anwendungsbereichen, wie z.B. dem Handel, bestehen an sie zusätzlich sehr hohe Anforderungen an die Verarbeitungs- und Kommunikationsgeschwindigkeit. Aufgrund der stark verteilten Verarbeitung sind sie dabei sehr hohen Risiken ausgesetzt.

Der Finanzsektor ist von einer starken Vernetzung und Digitalisierung geprägt. Ein einzelnes Finanzinstitut kann daher nur sehr beschränkt seine eigene Strategie für die Migration zu PQK umsetzen, ohne sich mit seinem Umfeld abzustimmen. Außerdem sind Sicherheitsstandards einzuhalten, die von unterschiedlichen nationalen oder internationalen Stellen oder auch von Industrievereinigungen verantwortet werden.

Die spezifischen regulatorischen Anforderungen im Finanzsektor insbesondere zum Management der IT als auch zur Steuerung der Dienstleister machen eine strukturierte Planung und Kontrolle des Migrationsprozesses erforderlich.

4 Typische Kommunikationsbeziehungen eines Finanzinstituts

Der Einsatz von kryptografischen Verfahren bedingt die Kooperation zwischen den an der Kommunikation beteiligten Parteien. Gerade Finanzinstitute tauschen sensible Informationen mit einer Vielzahl von Kommunikationspartnern aus. Als typische Kommunikationsbeziehungen eines Finanzinstituts sind beispielsweise zu nennen:

- Für Privat- und Geschäftskunden werden unterschiedlichste Finanzdienstleistungen angeboten. Zu nennen sind in erster Linie neben dem klassischen Online-Banking auch das Verwahren von Wertpapieren oder anderen Finanzprodukten, wie etwa auch Kryptowährungen. Hinzu kommt die Kundenkommunikation, die üblicherweise einer Vertragsschließung vorausgeht und ebenfalls schützenswert ist. Im Fall von Versicherungen ist die Kommunikation im Rahmen der Schadensabwicklung zu nennen.
- Ebenso vertraulich ist die Kommunikation mit Geschäftspartnern, Investoren oder Eigentümern. Typische Kommunikationsinhalte sind u.a. Reports über gemeinsame Projekte oder den aktuellen Geschäftsverlauf, Risikoberichte oder Kommunikation über die gegenwärtige oder zukünftige strategische Ausrichtung des Instituts.

- Mit Zahlungsdienstleistern werden Finanztransaktionen, Kreditkartenzahlungen und anderweitige Geschäfte abgewickelt sowie sonstige Zahlungsinformationen ausgetauscht.
- Nicht nur Transaktionen, sondern auch Reports und andere marktrelevante Kommunikation werden mit den nationalen und übernationalen Zentralbanken ausgetauscht. In diesem Zusammenhang ist auch die Kommunikation mit Aufsichtsorganen, Meldestellen und Ministerien zu nennen. Neben meldepflichtigen Vorfällen ist hier u. a. noch die Beantwortung von Anfragen durch Regulatoren zu nennen.
- Zwischen den Instituten werden Interbankentransaktionen abgewickelt.
- Über Handelsplätze und mit Börsen werden Order, Termingeschäfte, etc. abgewickelt und sonstige Handelsdaten ausgetauscht. Neben der Anforderung an die Sicherheit dieser Daten kommt beim Hochfrequenzhandel noch die an deren schnelle Bearbeitung hinzu.
- Ein weiterer Bestandteil des Finanzsektors sind Informationsdienste, wie Ratingagenturen oder Bonitätsdienste, über die Ratings, Kreditauskünfte, etc. abgerufen werden können.
- Eine vergleichbare Rolle spielen Analysten und Marktdatenanbieter, die Marktdaten, Analysen und andere finanzmarktrelevante Informationen anbieten.
- Schließlich nutzen Finanzinstitute im Rahmen der Erbringung ihrer Dienstleistungen auch externe Dienstleister und sie stehen im Kontakt mit Wirtschaftsprüfern, Kanzleien und anderen Organisationen, mit denen schützenswerte Daten ausgetauscht werden.

Diese Liste der Kommunikationskanäle erhebt keinen Anspruch auf Vollständigkeit. Nicht alle der genannten Kommunikationskanäle sind für ein Finanzinstitut relevant. Welche davon tatsächlich genutzt werden, hängt von dem oder den spezifischen Geschäftsfeldern ab, in denen das Institut tätig ist. Dennoch wird ein Eindruck der komplexen Kommunikationsstrukturen vermittelt. Und deren Absicherung muss unter den Beteiligten organisiert werden.

5 Unterschiedliche Verantwortlichkeiten für die Festlegung von Kommunikationsstandards, Datenformaten und Schutzmechanismen

Viele der im vorherigen Abschnitt genannten Kommunikationsbeziehungen sind standardisiert. Für die Definition dieser Standards sind verschiedene Stellen verantwortlich, deren Umsetzung wird von staatlichen Stellen oder entsprechend akkreditierten Dienstleistern überprüft. Hierzu sind beispielhaft zu erwähnen:

- Der für das Internetbanking eingesetzte HBCI-Standard wird von der Deutschen Kreditwirtschaft (DK), einer Interessenvertretung deutscher Banken verantwortet.
- Aufsichtsorgane, wie z.B. die Europäische Zentralbank oder die Deutsche Bundesbank aber auch Meldestellen, wie das BSI, haben spezifische Anforderungen an den Schutz von Transaktionsdaten und die Absicherung des Meldewesens.
- Die Anbindung an das internationale Zahlungssystem SWIFT unterliegt detaillierten Anforderungen, die im SWIFT Customer Security Program (SWIFT CSP) kodifiziert sind.
- Für die Absicherung von Kreditkartenzahlungen gelten die Anforderungen des PCI DSS, der von einer Vereinigung von Kreditkarteninstituten verantwortet wird.
- Die Policies für Kryptowährungen enthalten Sicherheitsanforderungen und werden oft nach einem Mehrheitsprinzip unter der Gemeinschaft der Währungsinhaber definiert, daraus leiten sich i.d.R. auch Anforderungen an die Verwahrung ab.
- Ratingagenturen, Auskunfteien, etc. definieren die technischen Anforderungen an die Kommunikation, was die kryptografische Absicherung einschließt.
- Mit Dienstleistern sind individuelle Vereinbarungen über den sicheren Austausch von Daten zu treffen.
- Letztlich sind auch die bankinternen Verfahrensweisen zur Authentisierung, Verschlüsselung, zum Integritätsschutz und für digitale Unterschriften vor dem Hintergrund der internen technischen Voraussetzungen und Sicherheitsanforderungen festzulegen.

6 Derzeit verwendete kryptografische Verfahren, deren Einsatzumgebungen und Hersteller

In die Verarbeitung von Finanzdaten sind Software- und Hardwarekomponenten einbezogen. Neben Debit- oder Kreditkarten, Geldautomaten, Smartphones, etc. wird auch Verschlüsselungshardware z.B. zur Beschleunigung der Verschlüsselung von Transaktionsdaten eingesetzt. Eingesetzt werden dabei im Wesentlichen alle gängigen symmetrischen und asymmetrischen kryptografischen Verfahren, wie AES, RSA, Diffie-Hellman, elliptische Kurven, etc. Diese sind entweder in spezifische Anwendungssysteme integriert, werden mittels Bibliotheken bereitgestellt oder es werden Kryptomodule über entsprechende Schnittstellen angesprochen.

Zunächst sind die Hersteller von Anwendungssystemen, Software und Hardware aufgefordert, ihre Produkte mit PQK auszustatten. Vorschläge für geeignete PQK-Algorithmen werden in diesem Zusammenhang von der US-amerikanischen Normierungsbehörde NIST erarbeitet [1]. Konkrete Anforderungen an die Integration dieser Algorithmen in relevante Anwendungen wurden bereits von der NSA definiert [2]. Demnach sollen z.B. ab 2025 in Webbrowsern, Servern und in Cloud-Diensten PQK Algorithmen als Default zum Einsatz kommen. Schon vor diesem Zeitpunkt sollen sie als optionale Komponenten in solchen Produkten enthalten sein. So bietet z.B. Google Chrome seit August 2023 in der Version 116 quantenresistente Verschlüsselung an.

Im kommerziellen Einsatz ist es wichtig, die Systemumgebung an die neuen Algorithmen anzupassen. Der für Schlüssel benötigte Speicherplatz und die für die kryptografischen Funktionen benötigten Verarbeitungszeiten ändern sich. Kryptokonzepte und -policies sind zu überarbeiten, und für jeden Anwendungsfall ist eine Migrationsstrategie zu entwickeln. Hierzu können die Hersteller für ihre jeweiligen Produkte sicherlich Hilfestellung bieten. Eine organisationsweite Strategie kann nur vom Institut selbst erstellt werden. Für die oben genannten organisationsübergreifenden Anwendungsfälle ist die Koordination zwischen den Beteiligten erforderlich.

7 Notwendigkeit der Kooperation bei der Migration in Richtung PQK

Bei der Entwicklung einer Migrationsstrategie in Richtung PQK gilt es für das Finanzinstitut, auch die Verarbeitungsketten zu berücksichtigen. Z.B. beginnt eine typische Internet-Überweisung mit dem Zugriff des Kunden auf die Homebanking-Anwendung der Bank, wird dort weiterverarbeitet, um anschließend über ein Zahlungssystem (SWIFT) transferiert zu werden. Institutsintern geht diese Transaktion u.a. in das Meldewesen und die Rechnungslegung ein.

Für das Kreditinstitut ergibt sich daraus die Notwendigkeit, die Migration zur PQK in drei Richtungen voranzutreiben:

- Die Anforderungen externer Standardgeber, wie der Zentralbanken, Aufseher, Industrievereinigungen, etc. sind aufzunehmen und in die institutsinterne Kryptostrategie zu integrieren.
- Die technischen Möglichkeiten der verwendeten Produkte und Systeme sowie die Möglichkeiten und Einschränkungen der verwendeten Infrastruktur sind zu berücksichtigen.

- Gemäß dem institutsspezifischen IKT-Umfeld und Risikoprofil ist eine geeignete Migrationsstrategie zu entwickeln und mit den jeweiligen Geschäftspartnern und Dienstleistern zu vereinbaren.

Diese drei Hauptstoßrichtungen der PQK-Migrationsstrategie eines Kreditinstituts sind ggf. zu flankieren mit Maßnahmen zur Absicherung von Blockchains oder Kryptowährungen, an denen das Institut beteiligt ist. In jedem Fall können die Migrationsstrategien in diese drei Richtungen – Regulatoren und Standardgeber, Zulieferer und Technologie sowie institutsinterne Vorgaben – nicht unabhängig voneinander entwickelt werden. Eine enge inhaltliche und zeitliche Koordination ist erforderlich.

8 Vorgehensweise zur Entwicklung einer PQK-Migrationsstrategie

Die Herangehensweise zur Entwicklung einer PQK-Migrationsstrategie variiert abhängig von der jeweiligen Organisation und ihrer Zielsetzung. Aufsichtsorgane haben die Funktionsfähigkeit und Resilienz des Markts im Fokus und sind daran interessiert, dass die Dienste, die den Kunden angeboten werden, verlässlich und sicher sind. Verbände möchten gemeinsame Standards vereinbaren, um Entwicklungskosten zu reduzieren und um Kompatibilität zwischen Dienstleistungen zu gewährleisten. Die einzelnen Institute schließlich sind an effizienten Abläufen interessiert, die an ihrem Risikoappetit ausgerichtet sind. Allen gemeinsam ist eine Vorgehensweise zur Definition einer PQK-Migrationsstrategie, die sich an den Schritten Analyse, Design, Implementierung, Laufende Überwachung und Steuerung orientiert.

8.1 Analyse

Die Grundvoraussetzung für eine wirkungsvolle PQK-Migrationsstrategie ist eine vollständige Transparenz hinsichtlich der Nutzung kryptografischer Verfahren. In welchen Einsatzbereichen, Systemen und Infrastrukturen wird Kryptografie eingesetzt und welche Algorithmen werden wo und wofür verwendet? Für die technische Analyse gibt es dazu verschiedene Tools, die z.B.

- im Netzwerk den Handshake zwischen Kommunikationspartnern auswerten und daraus die verwendeten Kryptoparameter ableiten,
- Sourcecode analysieren und auf diese Weise Aufrufe von Kryptobibliotheken identifizieren oder
- solche Aufrufe während der Laufzeit erkennen.

Will man sich der Aufgabe konzeptionell nähern, gilt es zunächst festzustellen, wo Sicherheitsanforderungen den Einsatz von Kryptografie verlangen, um anschließend die gängigen Kommunikationsstandards zu analysieren, die in der jeweiligen Einsatzumgebung verwendet werden. Neben der Frage,

welche Algorithmen mit welchen Parametern für welchen Zweck eingesetzt werden, ist es ebenso wichtig, zu verstehen, wie die Einsatzumgebung technisch gestaltet und betrieblich organisiert ist. Während der Analyse ist also insbesondere aufzunehmen

- ob Hard- oder Software für die Kryptoverfahren eingesetzt werden,
- welche Rechen- und Speicherkapazitäten zur Verfügung stehen,
- wie das Schlüsselmanagement gehandhabt werden kann und
- wie das Softwaremanagement in der Einsatzumgebung gehandhabt wird.

Aus betrieblicher Sicht

- sind Lizenzrechte zu klären,
- Verantwortlichkeiten zu identifizieren und
- die Change-Management-Prozesse zu verstehen.

Schließlich sind die relevanten Bedrohungen und Risiken zu analysieren, zu deren Mitigation Kryptografie eingesetzt wird. Im Zusammenhang mit PQK ist hier vor allem an die Bedrohung durch „harvest-now-decrypt-later“ Angriffe zu denken. Gerade hier entstehen durch die rasante technische Entwicklung der Quantencomputertechnik Unsicherheiten. Für den Anwender sind zwei Szenarien zu unterscheiden. Solche Informationen, die während ihrer Lebenszeit unter vollständiger Kontrolle der Organisation bleiben, können zu dem Zeitpunkt neu geschützt werden, wenn eine tatsächliche Bedrohung vorliegt, also die Quantencomputertechnik entsprechend fortgeschritten ist. Hierzu können die bereits verschlüsselten Daten ein zweites Mal durch einen quantenresistenten Algorithmus verschlüsselt werden. Bei Informationen, die über öffentlich zugängliche Kanäle kommuniziert oder mit externen Kommunikationspartnern austauscht wurden, ist dies nicht möglich. In diesem Fall muss also schon zum Zeitpunkt des Austauschs der Daten abgeschätzt werden, über welchen Zeitraum hinweg der Schutz vor Offenlegung oder Manipulation wirksam sein soll. Es stellt sich also die Frage, wie der weitere Fortschritt der Quantencomputertechnologie abgeschätzt werden kann.

Der Fortschritt in der klassischen Computertechnik ist gut verstanden und wird gern dem Mooreschen Gesetz abgeschätzt. Entsprechend werden von Sicherheitsbehörden, wie z.B. dem BSI-Empfehlungen für geeignete kryptografische Verfahren und Schlüssellängen veröffentlicht [3]. Vergleichbare Erfahrungen liegen im Bereich Quantencomputing nicht vor. Die Empfehlungen der NSA [2] sind eine der wenigen Richtwerte, die in diesem Zusammenhang herangezogen werden können. Dort werden Termine genannt, ab denen

quantencomputerresistente Algorithmen optional bzw. verpflichtend in Produkte zu integrieren sind. Z.B. soll die Kommunikation zwischen Webservern und -browsern per Default schon ab dem Jahr 2025 mit quantencomputerresistenter Verschlüsselung betrieben werden.

Die Analysephase der PQK Migrationsstrategie dient also dazu,

- alle Anwendungsfälle von Kryptografie im Untersuchungsbereich zu identifizieren,
- die aktuell verwendeten Verfahren mit ihren relevanten Parametern aufzunehmen,
- ein Verständnis der technischen wie betrieblichen Einsatzumgebung mit ihren Rahmenbedingungen zu erhalten und
- das mit dem Einsatz der Kryptografie im Zusammenhang stehende Risiko zu verstehen.

8.2 Design

Aus den Ergebnissen der Analysephase kann eine Priorisierung der Anwendungsfälle abgeleitet werden, die sich einerseits am Risiko ausrichtet und andererseits die Komplexität der Umstellungsmaßnahmen berücksichtigt. Sind z.B. die kommunizierten Daten nur für einen kurzen Zeitraum schützenswert, oder sind kompensierende Zugriffskontrollmaßnahmen wirksam, genügt es eventuell, eine Umstellung auf PQK erst dann vorzunehmen, wenn tatsächlich entsprechend fähige Quantencomputer existieren, die die Verschlüsselung brechen können. Ist der „harvest-now-decrypt-later“ Angriff ein reales Szenario, sind die Zeiträume mit zu berücksichtigen, die die Daten diesem Angriff ausgesetzt sind. Das Risiko gibt also vor, wann eine Umstellung auf PQK einzuplanen ist, während die Komplexität ein Hinweis dafür ist, wieviel Zeit, Aufwand und Ressourcen in eine Umstellung zu investieren sind. Ist etwa mit der Umstellung ein Austausch von Hardware verbunden, oder ist eine größere Anzahl von Organisationen als Kommunikationspartner einzubeziehen, so ist auch mehr Zeit für die Umstellung einzuplanen.

Mit allen Beteiligten und unter Berücksichtigung von Empfehlungen seitens Standardisierungsgremien und Expertengruppen sind geeignete PQK-Verfahren auszuwählen. Dabei sind die spezifischen Anforderungen der jeweiligen Einsatzumgebung zu berücksichtigen (zu erfüllende Sicherheitsziele, Performanceanforderungen, Key Management, etc.) und Abhängigkeiten zu vor- und nachgelagerten Prozessen und Verarbeitungsschritten. Zugrunde liegen stets die regulatorischen Anforderungen hinsichtlich Planung und Dokumentation.

8.3 Implementierung

Die Umsetzung des Designs muss zwischen den Beteiligten eng koordiniert werden. Es sind entsprechende Gütekriterien festzulegen und Tests zu definieren, mit denen die effektive und vollständige Umsetzung nachvollzogen werden kann. Fallbackstrategien sind ebenso vorzusehen wie begleitende Maßnahmen, die zu erwartende Schwierigkeiten bei der Produktivsetzung neuer Lösungen adressieren.

Im Zuge der Implementierung quantencomputerresistenter kryptografischer Verfahren ist es sinnvoll, die Ergebnisse aus der Analysephase auch dazu zu nutzen, um einen zukünftigen Austausch der eingesetzten Kryptoverfahren zu vereinfachen. Dieses, als Kryptoagilität bezeichnete Prinzip ist auch deshalb sinnvoll, weil mit PQK auf neue Klassen von Kryptoverfahren gesetzt wird, die nicht im gleichen Maß untersucht sind, wie die etablierten. Es ist also gut möglich, dass sie zwar resistent gegen Quantencomputer sind, dafür jedoch Angriffe auf Basis klassischer Computertechnik bekannt werden.

Daher lohnt es sich, die Ergebnisse aus der Analysephase dafür zu nutzen, ein Inventar der verwendeten kryptografischen Verfahren zu pflegen. So kann man einerseits eine kontinuierliche Transparenz über die verwendeten Verfahren und die damit im Zusammenhang stehenden Risiken aufrechterhalten und andererseits notwendige Anpassungen besser koordinieren.

Dieses Kryptoinventar ist auch dazu zu nutzen, um im Rahmen der Threat Intelligence sich kontinuierlich über das Auftreten neuer Bedrohungen gegen die eingesetzten Verfahren zu informieren, um rechtzeitig auf daraus resultierende Risiken reagieren zu können.

8.4 Laufende Überwachung und Steuerung

Unter Verwendung der im Rahmen der Kryptoagilität etablierten Mittel ist die Sicherheit der eingesetzten kryptografischen Verfahren enger als bisher zu kontrollieren. Zeichnen sich Schwachstellen ab, sind frühzeitig geeignete Gegenmaßnahmen zu ergreifen. Möglich sind sowohl eine Anpassung der für den jeweiligen Algorithmus zu verwendenden Parameter (z.B. Schlüssellänge) als auch ein Wechsel zu anderen Algorithmen. Zwischen den Beteiligten sind entsprechende Prozesse und Kontrollen zu etablieren, mit denen zukünftig der Wechsel zu neuen Kryptoverfahren koordiniert und schnell umgesetzt werden kann.

8.5 Koordination

Zusammenfassend wird die hohe Komplexität der Aufgabe einer Entwicklung und Umsetzung von PQK-Migrationsstrategien im Finanzsektor festgestellt. Es sind nicht nur institutsintern erhebliche Planungsaufwände zu erwarten. Auch dürfte die notwendige Vereinbarung mit Dienstleistern und auch die Bereitstellung neuer Schnittstellen und Token für die Kunden mit erheblichem Aufwand verbunden sein. Erhebliche Energie ist auch in die Abstimmung in Interessensvertretungen und Branchenverbänden zu investieren. Weiter dürfen die auf die Aufsichtsorgane zukommenden Steuerungs- und Überwachungsaufgaben nicht unterschätzt werden.

Fest steht, dass in den einzelnen Instituten diese drei Handlungsstränge zusammenlaufen und bei der Ausgestaltung der PQK-Migrationsstrategie mit berücksichtigt werden müssen. In die Zeitplanung, in die Auswahl geeigneter Technik und in die Gestaltung der Betriebs- und Überwachungsprozesse müssen die Vorgaben aus der Regulierung, wie die Abstimmungsergebnisse aus Branchenverbänden und die Ergebnisse institutsinterner Analysen mit eingehen.

9 Fazit und Empfehlung einer abgestimmten Zeitplanung

Mit dem Aufkommen von Quantencomputern gerät der Bereich der Kryptografie stärker als bisher in den Fokus des taktischen IKT-Risikomanagements (Management von Risiken der Informations- und Kommunikationstechnik) und damit auch in den Aufgabenbereich der digitalen Resilienz im Sinne von DORA [4]. Während in der Vergangenheit sich das Management des Einsatzes von Kryptografie mehr auf den Austausch von Schlüsseln und das Management von Schlüsselzertifikaten konzentrierte, werden jetzt auch die Kryptoalgorithmen und deren Parameter selbst Teil des regelmäßigen Änderungsmanagements.

Mit der Migration zu PQK wird also einerseits die durch Quantencomputer induzierte Gefahr abgewehrt, andererseits aber auch das Prinzip der Krypoagilität in das IKT-Risikomanagement integriert. Für die drei oben genannten Migrationsrichtungen (Standardgeber, Interessensgruppen, Institute) sind geeignete Transparenz- und Steuerungsmechanismen zu etablieren, um möglichen zukünftigen Risiken geeignet entgegentreten zu können. So wird eine kontinuierliche Überwachung und Steuerung des Einsatzes geeigneter kryptografischer Verfahren Teil des Risikomanagements für die digitale Resilienz von Finanzinstituten.

Literaturhinweise

[1] https://csrc.nist.gov/projects/post-quantum-cryptography

[2] CSA_CNSA_2.0_ALGORITHMS_.PDF (defense.gov)

[3] BSI - BSI TR-02102 Kryptographische Verfahren: Empfehlungen und Schlüssellängen (bund.de)

[4] Regulation - 2022/2554 - EN - DORA - EUR-Lex (europa.eu)

Zurück zum Inhaltsverzeichnis
Zurück zum Stichwortverzeichnis

Quantensichere VPN-Infrastrukturen

David Schatz[1], Friedrich Altheide[2],
Prof. Dr.-Ing. Günter Schäfer[1], Dr. Kai Martius[2]

Kurzfassung:

Für die in den kommenden Jahren erforderliche Umstellung zahlreicher „Virtuelle Private Netz (VPN)"-Infrastrukturen auf quantensichere kryptografische Verfahren stehen grundsätzlich drei alternative technologische Ansätze zur Verfügung: Post Quantum Cryptography (PQC), Quantum Key Distribution (QKD) und klassische symmetrische Kryptografie mit ausreichend langen Schlüsseln (≥ 256 Bit). Da es im Gegensatz zur symmetrischen Kryptografie bei den beiden erstgenannten Verfahrenskategorien aktuell noch nicht klar absehbar ist, welche der verfügbaren Kandidaten auch langfristig ein hinreichendes Sicherheitsniveau garantieren können, stellen wir in diesem Beitrag einen integrierten Ansatz zur sicheren Authentisierung und Schlüsselverwaltung in großen VPN-Infrastrukturen vor, der einen komplementären Einsatz aller drei Verfahrenskategorien ermöglicht und dabei Sicherheit in der Tiefe realisiert.

Stichworte: IKE, MKR, PQC, QKD, Quantensicherheit, VPN

1 Einleitung und Motivation

Seit den Veröffentlichungen von Shor [1] in der Mitte der 1990er Jahre sowie Proos und Zalka [2] wenige Jahre später ist bekannt, dass alle klassischen asymmetrischen Authentisierungs- und Schlüsselaustauschverfahren bei Verfügbarkeit praktisch einsetzbarer Quantencomputer unsicher werden. Aufgrund der Bedrohung, dass ein Angreifer bereits heute großflächig Daten abhören und speichern könnte, um sie in einigen Jahren zu entschlüsseln („store now, decrypt later"), besteht allerdings in zahlreichen Anwendungsszenarien bereits heute akuter Handlungsbedarf, die bestehenden asymmetrischen Verfahren zu ersetzen. Bei symmetrischen Verfahren konnte hingegen gezeigt werden [3], dass diese, Stand heute, ab einer Schlüssellänge von 256 Bit auch Angriffen mittels Quantencomputer standhalten, soweit keine anderweitigen Schwächen bekannt werden. Allerdings sind bei ihrem Einsatz für Authentisierung und Schlüsselverwaltung in der Praxis eine Reihe von Herausforderungen zu meistern. Diese umfassen insbesondere die sichere initiale Verteilung paarweiser Schlüssel, sowie ihre regelmäßige sichere Erneuerung.

[1] Technische Universität Ilmenau, Ilmenau
[2] secunet Security Networks AG, Essen

Dementsprechend wird seit Mitte der 1990er Jahre mit großer Intensität an neuartigen asymmetrischen Verfahren geforscht, welche ebenfalls resistent gegenüber Angriffen mittels Quantencomputern sind. Diese Verfahren werden unter dem Begriff Post Quantum Cryptography (PQC) zusammengefasst (siehe [4] für einen Überblick über den aktuellen Stand der Technologie). Allerdings sind die bereits länger bekannten Verfahren (etwa Classic McEliece) aufgrund des erforderlichen Rechenaufwandes und der Länge der Schlüssel sowie der Chiffren/Signaturen für praktische Anwendungen wenig geeignet. Bei den jüngeren und effizienteren Ansätzen gibt es hingegen noch Vorbehalte/Unsicherheit bezüglich ihrer langfristigen Sicherheitseigenschaften. So wurde beispielsweise das Verfahren Rainbow – immerhin Finalist im Standardisierungsverfahren des NIST für PQC-Verfahren – im Jahr 2022 erfolgreich auf einem Laptop gebrochen [5]. Auch wenn der sofortige Einsatz von PQC-Verfahren zwingend nötig ist, wird dementsprechend von nationalen und einigen internationalen Sicherheitsagenturen für den Moment ein „hybrider" Einsatz von PQC in Kombination mit klassischer asymmetrischer Kryptografie empfohlen (vergleiche auch [4]).

Als Alternative zu PQC wird zunehmend die beweisbar sichere Schlüsselverteilung unter Nutzung quantenmechanischer Effekte (QKD) [6] in Betracht gezogen. In der Praxis stellen sich jedoch eine ganze Reihe von Herausforderungen, die mit der aktuell begrenzten Reichweite (ca. 100 km per Glasfaser) sowie offenen Fragen zur Authentisierung und zur Sicherheitsevaluierung zusammenhängen. Dies führt dazu, dass diverse Sicherheitsagenturen QKD aktuell, wenn überhaupt, höchstens als ergänzenden Schutz *zusätzlich* zu PQC empfehlen [7].

Insgesamt ist derzeit allerdings nicht absehbar, unter welchen Voraussetzungen die von der NIST standardisierten PQC-Verfahren sicher eingesetzt werden können. Deshalb kann nicht ausgeschlossen werden, dass zukünftig die jeweils verwendeten PQC-Algorithmen kurzfristig, etwa nach Bekanntwerden von konzeptionellen Schwachstellen, während des Betriebs ausgetauscht werden müssen. Abhängig von der Verwendung der Algorithmen (etwa in Smartcards) oder von den gefundenen konzeptionellen Schwachstellen, wird ein Wechsel möglicherweise nicht immer sofort möglich sein.

Es stellt sich somit insgesamt die Frage, wie „Virtuelle Private Netz (VPN)"-Infrastrukturen während einer solchen Migration beziehungsweise am besten auch über die Migration hinaus, zusätzlich abgesichert werden können. Insbesondere wäre es wünschenswert, wenn durch diese zusätzliche Absicherung der Angreifer gezwungen werden würde, langfristig zu planen, zu agieren, und hierbei große Mengen an Ressourcen aufzuwenden. Diese Frage

wird im weiteren Verlauf am Beispiel der weitverbreiteten VPN-Protokollfamilie IPsec diskutiert. Die dargelegten Ideen können jedoch auch für andere VPN-Technologien angepasst werden. Im Folgenden werden zunächst Anforderungen an quantensichere VPN-Infrastrukturen aus praktischer und implementierungstechnischer Sicht erläutert und auf dieser Grundlage eine Strategie für ihre Realisierung hergeleitet. Diese zielt darauf ab, den erforderlichen Aufwand für einen großflächig agierenden Angreifer erheblich zu steigern und damit die „store now, decrypt later" Problematik im gesamten VPN nachhaltig zu lösen.

2 Anforderungen an quantensichere VPN-Infrastrukturen

Neben den traditionellen Anforderungen an VPN-Infrastrukturen, wie etwa eine Ende-zu-Ende-Sicherung zwischen Sicherheitsendpunkten, Skalierbarkeit mit der Netzgröße und Robustheit gegenüber Denial of Service (DoS)-Angriffen, hat eine Migration zu quantensicheren VPN-Infrastrukturen vor allem folgende Anforderungen:

- **Kryptoagilität**: Die eingesetzten kryptografischen Verfahren müssen während des Betriebs leicht austauschbar sein, um schnell auf Entwicklungen in der Kryptanalyse reagieren zu können.

- **Schlüsselquellenagilität**: Die Implementierung sollte möglichst einfach um weitere Quellen von symmetrischem Schlüsselmaterial (etwa QKD oder manuell ausgetauschte Schlüssel) erweitert werden können. Insbesondere sollte es möglich sein, Quellen zur Laufzeit dynamisch hinzu oder, etwa wegen Wartungsarbeiten, abschalten zu können.

- **Implementierungssicherheit**: Die Integration muss so erfolgen, dass die Implementierungssicherheit des Gesamtsystems gewährleistet ist. Weiterhin sollte eine Implementierung möglichst einfach gehalten werden (kleine Trusted Computing Base, kurz TCB).

3 Sicherheit in der Tiefe für das IKE-Protokoll

Ein konkreter und weit verbreiteter Anwendungsfall für VPN, welche mittels IPsec realisiert sind, ist es, mehrere Standorte über nicht vertrauenswürdige Netze (zum Beispiel das Internet) sicher zu verbinden (siehe Abbildung 1). Dabei tunneln VPN-Gateways sämtlichen Verkehr von Clients zwischen Standorten und sichern diesen kryptografisch (Verschlüsselung, Schutz der

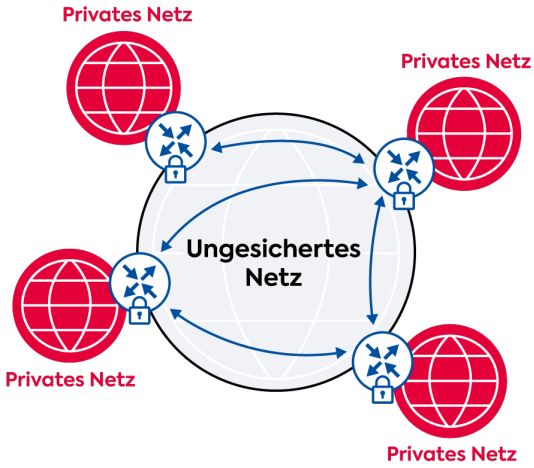


Abbildung 1: Einfache VPN-Infrastruktur mit einem VPN-Gateway und einem privaten Netz pro Standort

Integrität). Die Clients eines Standorts befinden sich dabei etwa in einem gemeinsamen, vertrauenswürdigen lokalen Netzwerk. Der authentisierte Austausch eines symmetrischen Sitzungsschlüssels zwischen zwei Gateways erfolgte bisher durch klassische asymmetrische Kryptografie. Dieser Schlüsselaustausch wird durch das IKE-Protokoll (Internet Key Exchange) durchgeführt. Nach dem erfolgten Austausch kann auf Basis des Sitzungsschlüssels sämtlicher Client-Verkehr zwischen den Gateways abgesichert werden.

Aufgrund ihrer konzeptionellen Ähnlichkeit zu klassischen Verfahren, konnten PQC-Verfahren bereits relativ einfach in die aktuelle Version des IKE-Protokolls (IKEv2) integriert werden, um in Kombination mit etablierten Verfahren verwendet zu werden. Beispielsweise kann durch die Erweiterung IKE_INTERMEDIATE [8] ein hybrider Schlüsselaustausch umgesetzt werden. Hierdurch sind ausgehandelte Sitzungsschlüssel und damit der Client-Verkehr so lange sicher, wie mindestens eines der eingesetzten asymmetrischen Verfahren sicher ist. Allerdings wird das „store now, decrypt later“ Problem nur gelöst, sofern sich die eingesetzten PQC-Verfahren als langfristig sicher herausstellen.

Eine orthogonale Idee ist es deshalb, den von IKE abgeleiteten Sitzungsschlüssel zusätzlich durch „out of band“ ausgetauschtes symmetrisches Schlüsselmaterial zu „verstärken“. Mögliche Quellen für solch zusätzliches Material sind unter anderem:

- **Pre-shared Keys (PSKs)**: VPN-Gateways könnten „per Hand“ mit zusätzlichem symmetrischem Schlüsselmaterial ausgestattet werden. Beispielsweise könnte auf jedem Gateway derselbe Gruppenschlüssel hinterlegt werden, stellenweise auch als „CUG-Schlüssel“ bekannt

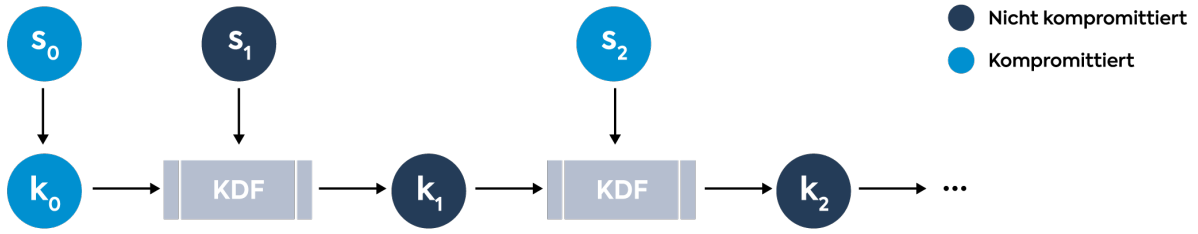


Abbildung 2: Schlüsselkette zur Kombination aller verfügbaren paarweiser symmetrischen Schlüssel s_i zu verschiedenen Master-Schlüsseln k_i über die Zeit. Im gezeigten Beispiel wird angenommen, dass der initial verwendete Schlüssel $k_0 = s_0$ dem Angreifer bekannt ist (kompromittiert). Die Synchronisation des sicheren Schlüssels s_1 sorgt dafür, dass der neue Master-Schlüssel k_1 dem Angreifer nicht länger bekannt ist. Dies ändert sich auch nicht durch die Synchronisation eines kompromittierten Schlüssels s_2.

(Closed User Group). Allerdings bietet ein Gruppenschlüssel keine Ende-zu-Ende-Sicherheit. Ziel sollte es daher sein, den Gateways *paarweise* symmetrische Schlüssel zur Verfügung zu stellen. Dies könnte etwa durch Mitarbeiter auf Dienstreisen oder durch den Versand von Smartcards mit Schlüsselmaterial auf dem Postweg realisiert werden.

- **„Business Trip Key Exchange“**: Anstatt händisch PSKs zu verteilen, könnten (abgesicherte) Laptops von Mitarbeitern verwendet werden, um bei Dienstreisen vollständig automatisiert Schlüssel zwischen den VPN-Gateways der unterschiedlichen Standorte zu verteilen.
- **QKD-Schlüssel**: Sofern einzelne QKD-Strecken sinnvoll/wirtschaftlich betrieben werden können, könnten die Gateways Schlüssel von den entsprechenden QKD-Endpunkten anfragen und einsetzen.

Die hier vorgestellten Schlüsselquellen eignen sich alle, um direkte Punkt-zu-Punkt Verbindungen abzusichern. Vor allem in großen VPNs kann damit jedoch noch nicht gewährleistet werden, dass zwischen allen Kombinationen von VPN-Gateways regelmäßig paarweise Schlüssel ausgetauscht werden. Ein alternatives Konzept, wie dennoch Schlüssel sicher indirekt in großen VPNs ausgetauscht werden können, wird später in Abschnitt 3.4 vorgestellt.

Eine weitere Problematik ist, dass es Angreifern möglicherweise zeitweise gelingen wird, einzelne Schlüssel aus den genannten Schlüsselquellen zu kompromittieren. Deshalb sollten so viele Schlüssel wie möglich und aus unterschiedlichen Quellen kombiniert werden, um den Aufwand für Angreifer dauerhaft zu steigern. Insbesondere sollte neues Schlüsselmaterial nicht einfach das bisherige Schlüsselmaterial ersetzen, welches dem Angreifer mög-

licherweise bereits unbekannt war. Deshalb schlagen wir vor, neues Schlüsselmaterial s_i immer mit dem bestehenden „Master-Schlüssel“ k_i, wie in Abbildung 2 veranschaulicht zu einem neuen Master-Schlüssel k_{i+1} zu kombinieren. Eine solche *Schlüsselkette* wird paarweise für jede Verbindung im VPN gepflegt und sollte mittels einer sicheren Key Derivation Function (KDF) realisiert werden. Stand heute ist die Sicherheit von existierenden, etwa Hash-basierten, KDFs nicht von Quantencomputern bedroht, sofern die einzelnen Ein- und Ausgaben lang genug gewählt sind (≥ 256 Bit pro Einzelschlüssel).

Da neue Schlüssel zu unterschiedlichen Zeitpunkten und in verschiedener Reihenfolge bei den beteiligten VPN-Gateways verfügbar werden können, wird ein *Synchronisationsprotokoll* zwischen den Gateways benötigt. Eine Möglichkeit dies zu realisieren, wäre eine Integration in das IKEv2-Protokoll. Allerdings würde dies die Protokoll- und Implementierungskomplexität weiter erhöhen und wäre aufgrund der Standardisierungsprozesse für IKE nicht kurzfristig umsetzbar. Stattdessen schlagen wir vor, dieses Synchronisationsprotokoll unabhängig durch eine kleine separate Software-Komponente durchführen zu lassen.

Weiterhin stellt sich die Frage, wie der aktuelle Master-Schlüssel am besten eingesetzt wird, um Client-Verkehr zusätzlich abzusichern. Auch hier wäre eine Integration in das IKEv2-Protokoll eine Möglichkeit, wie etwa in [9] und [10] vorgeschlagen. Damit werden Teile der IKE-Pakete durch den Master-Schlüssel geschützt, und der Schlüssel fließt in die Ableitung des finalen Sitzungsschlüssels mit ein. Allerdings können die ersten IKE-Pakete dabei nicht geschützt werden, da die Erweiterungen zunächst eine Einigung auf den richtigen Master-Schlüssel vorsehen. Während dies zumindest im zweiten Vorschlag [10] in der Theorie keinen Nachteil für die Sicherheit darstellt, bietet es dennoch eine gewisse Angriffsoberfläche auf die Implementierung und somit den gesamten IKE-Daemon. Deshalb präferieren wir stattdessen, sämtliche IKE-Pakete über eine separate Komponente –genannt IKE-Proxy – umzuleiten. Dort wird der Master-Schlüssel eingesetzt, um alle Pakete zu überschlüsseln sowie ihre Integrität zu sichern und somit sämtliche mittels IKE etablierten Sitzungsschlüssel implizit abzusichern. Das benötigte Synchronisationsprotokoll kann zudem direkt in den IKE-Proxy integriert werden.

Insgesamt bietet die Überschlüsselung der IKE-Pakete Sicherheit in der Tiefe: Solange der jeweils eingesetzte paarweise Master-Schlüssel k_i dem Angreifer nicht bekannt ist, sind Angriffe auf das IKE-Protokoll (sowohl die

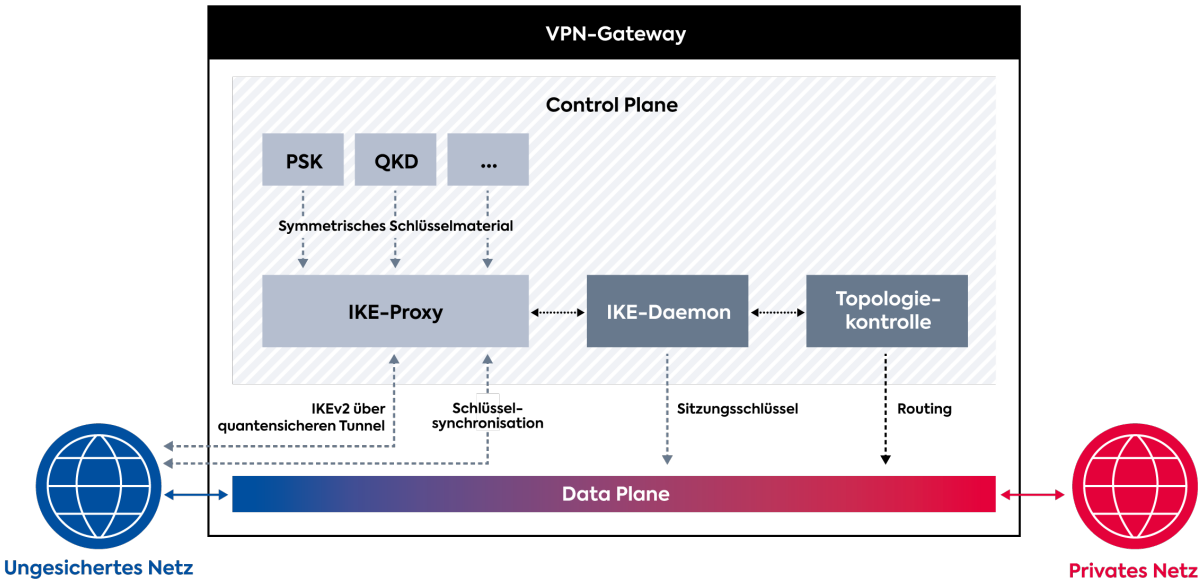


Abbildung 3: Integration des IKE-Proxy in eine VPN-Gateway-Architektur

Implementierung selbst als auch auf die eingesetzten asymmetrischen Kryptografieverfahren) beweisbar ausgeschlossen, selbst wenn zukünftig Schwächen in eingesetzten PQC-Verfahren entdeckt werden. Die Implementierungssicherheit des IKE-Proxys selbst kann zudem aufgrund seiner geringen Komplexität einfacher gewährleistet werden. Des Weiteren ist keinerlei Modifikation des IKE-Protokolls sowie seiner Implementierung nötig, sodass deren Sicherheitseigenschaften beweisbar nicht abgeschwächt werden. Die folgenden Abschnitte geben einen Überblick über die Integration und Funktionsweise des IKE-Proxys und diskutieren die Integration von QKD in das Gesamtkonzept, sowie die Ende-zu-Ende Absicherung von Verbindungen ohne direkte paarweise Schlüsselquellen.

3.1 Integration des IKE-Proxys

Der IKE-Proxy integriert sich in eine typische Architektur eines VPN-Gateways wie in Abbildung 3 gezeigt: Zum einen dient er als Schnittstelle zu allen vorhandenen Quellen für symmetrisches Schlüsselmaterial wie PSKs und QKD. Des Weiteren führt er mit allen VPN-Gateways, zu denen Schlüsselmaterial vorliegt, ein Synchronisationsprotokoll durch, um sich auf einen paarweisen Master-Schlüssel zu einigen, welcher alle bis zu diesem Zeitpunkt jemals beiden Seiten bekannten Schlüssel kombiniert. Dieser Schlüssel wird schließlich benutzt, um alle Pakete, welche vom IKE-Daemon gesendet werden, zusätzlich zu sichern (Verschlüsselung, Integritätsschutz) beziehungsweise eingehende Pakete zu überprüfen und zu entschlüsseln, bevor sie an den IKE-Daemon ausgeliefert werden. Um flexibel auf zukünftige Anforde-

rungen an symmetrische Verfahren reagieren zu können, bietet das verwendete Protokoll eine einfache Möglichkeit der Migration auf neue symmetrische Algorithmen.

3.2 Implementierungsdetails des IKE-Proxys

Dieser Abschnitt fasst einige Implementierungsdetails des IKE-Proxys zusammen, welche für seine Funktionalität und Sicherheitseigenschaften essenziell sind. Weitere Details sind in unserer vorangegangenen Veröffentlichung [11] zu finden.

3.2.1 Synchronisation von Schlüsseln

Die korrekte Synchronisation neuer Schlüssel s_i zu einem neuen Master-Schlüssel ist essenziell für die Verfügbarkeit des VPN: Kann ein IKE-Proxy eingehende Pakete nicht entschlüsseln, weil der Sender einen unbekannten oder falsch synchronisierten Master-Schlüssel verwendet, können keine weiteren IKE-Pakete ausgetauscht werden. Somit würden alle Rekeys sowie der Aufbau neuer Sicherheitsbeziehungen zwischen den betroffenen Gateways fehlschlagen.

Um mögliche Race-Conditions bei der Synchronisation neuer Schlüssel möglichst einfach zu handhaben, schlagen wir vor, zwischen zwei Gateways jeweils zwei Master-Schlüssel zu verwenden, einen in jede Richtung. Zu beachten ist dabei, dass neue Schlüssel s_i stets analog zu Abbildung 2 in beide Master-Schlüssel synchronisiert werden, aber unter Umständen zu unterschiedlichen Zeitpunkten. Ein einfaches Synchronisationsprotokoll sieht vor, dass jeder IKE-Proxy nur die Synchronisation neuer Schlüssel in den Master-Schlüssel für *ausgehende* Pakete initiiert, in einer von ihm festgelegten Reihenfolge. Dies geschieht mit einer einfachen Anfrage, welche eine eindeutige Schlüssel-ID des zu synchronisierenden Schlüssels s_i enthält. Nur wenn die Gegenstelle eine Bestätigung sendet, benutzt der IKE-Proxy den neuen Master-Schlüssel in diese Richtung und löscht den vorherigen. Um sicherzustellen, dass eine fehlerhafte Schlüsselquelle nicht die Verfügbarkeit beeinträchtigen kann, enthält die Bestätigung einen kryptografischen Nachweis über den Besitz von s_i. Dieser kann etwa mittels einer kryptografischen Hash-Funktion h aus s_i berechnet werden: $h(s_i)$. Durch diesen Nachweis können beide Gateways sicherstellen, dass sie denselben Schlüssel von der jeweiligen Quelle erhalten haben, bevor dieser im neuen Master-Schlüssel verwendet wird.

3.2.2 Zuordnung von Master-Schlüsseln

Ein Problem, welches sich aus der strikten Trennung von IKE-Daemon und IKE-Proxy ergibt, ist, dass der IKE-Proxy lediglich die IP-Adresse und Port-

Nummer der jeweiligen Gegenstelle sieht. Im Falle einer dynamischen Änderung, etwa in NAT-Szenarien, muss der Proxy aber dennoch eine statische ID der Gegenstelle ermitteln können, um den richtigen Master-Schlüssel auswählen und verwenden zu können. Dies kann durch ein simples Protokoll gelöst werden, indem der Proxy bei Paketen mit einer unbekannten Kombination aus IP-Adresse und Port-Nummer zunächst eine Anfrage nach der statischen ID der Gegenstelle sendet. Dabei wird sichergestellt, dass ausgetauschte IDs stets mindestens mit einem Gruppenschlüssel abgesichert sind.

3.2.3 Sichere persistente Speicherung von Recovery-Schlüsseln

VPN-Gateways können jederzeit neu gestartet werden, etwa zur Wartung oder nach einem Stromausfall. Damit nach einem solchen Neustart nicht neue paarweise Schlüsselketten verwendet werden müssen, welche beispielsweise nur einen initialen CUG-Schlüssel beinhalten, wird ein Mechanismus benötigt, um die Master-Schlüssel sicher zu persistieren. Eine einfache Methode wäre, alle Master-Schlüssel verschlüsselt auf der Festplatte abzuspeichern. Jedoch widerspricht dieser Ansatz dem Grundsatz der IT-Sicherheit, temporäre symmetrische Sitzungsschlüssel niemals persistent abzuspeichern.

Anstatt also die aktuellen Master-Schlüssel selbst zu persistieren, sehen wir für jeden Master-Schlüssel k_i vor, nur einen verschlüsselten, kryptografischen Hash-Wert $\mathrm{enc}_{\mathrm{kek}}\big(h(k_i)\big)$ zu persistieren („Recovery-Schlüssel“). Der zugehörige „Key Encryption Key“ (KEK) ist nur dem Gateway selbst bekannt und etwa auf einer Smartcard persistent gespeichert. Im Falle eines Neustarts eines der Gateways wird der gespeicherte Hash-Wert verwendet, um innerhalb der Smartcard einen neuen Master-Schlüssel k_{i+1} abzuleiten. In diese Ableitung gehen neben dem Hash-Wert zwei von den Gateways bereitgestellte Nonces n_1, n_2, ihre beiden statischen IDs i_1, i_2, sowie ein Gruppenschlüssel g ein:

$$k_{i+1} = kdf\left(h(k_i), n_1, n_2, i_1, i_2, g\right)$$

Die statische ID des jeweiligen Gateways sowie der Gruppenschlüssel werden hierbei direkt in der Smartcard gespeichert. Letzterer kann so hinterlegt werden, dass sich dieser im Regelfall nicht auslesen lässt.

Diese Art der Persistierung gewährleistet, dass selbst von einem Angreifer mit physikalischem Zugriff auf die Festplatte und die Smartcard, vorherige verwendete Master-Schüssel des angegriffenen Gateways nicht rekonstruiert werden können. Weiterhin kann ein Angreifer mit Zugriff auf eine Smartcard nur gültige (wiederhergestellte) Master-Schlüssel für dieses eine zugehörige Gateway erzeugen, nicht für beliebige andere Gateways.

3.3 Integration von QKD

Getrieben durch aktuelle Standards der ITU und ETSI, etwa [12], sowie durch kommerziell verfügbare Produkte, wird die QKD-Technologie aktuell vor allem mit sogenannten QKD- bzw. „Key Management"-Netzen in Verbindung gebracht. Dort ist die Idee, die eingeschränkte Reichweite von QKD zu überwinden, indem Schlüsselmaterial über mehrere sogenannte „Trusted Nodes" weitergeleitet wird, welche jeweils über eine QKD-Strecke verbunden sind. Dadurch sollen Schlüsselnutzern (etwa VPN-Gateways) transparent paarweise Schlüssel auch ohne direkte QKD-Verbindung bereitgestellt werden. Wesentliche Probleme dieses Vorgehens im Kontext eines VPN sind allerdings mögliche Schwächen im Falle kompromittierter „Trusted Nodes" und der Umstand, dass innerhalb des QKD-Netzwerks viele Funktionalitäten eines VPN erneut implementiert, evaluiert und gegebenenfalls zugelassen werden müssen. Hierzu zählen unter anderem die Authentisierung zwischen den QKD-Geräten zur Realisierung des klassischen Kanals, Routing-Verfahren, sowie Netzwerkmanagement und -monitoring.

Um die Gesamtkomplexität deutlich zu verringern, schlagen wir stattdessen vor, QKD-Strecken jeweils nur Punkt-zu-Punkt zu betreiben und lokal zu konfigurieren. Weiterhin sollte der klassische Kanal zweier QKD-Geräte über die zugehörigen VPN-Gateways mit einer eigenen Sicherheitsbeziehung getunnelt und abgesichert werden. Dadurch kann der klassische Kanal von den bereits etablierten Sicherheitsmechanismen und der gehärteten Implementierung der VPN-Gateways profitieren. Des Weiteren könnte je nach Einsatzbedingungen die Implementierung eigener Authentisierungsprotokolle in den QKD-Geräten entfallen beziehungsweise aus dem Zulassungsprozess genommen werden, um Komplexität einzusparen. Nach einem erfolgreichen QKD-Schlüsselaustausch kann die Sicherheitsbeziehung im VPN durch die Ableitung eines neuen Master-Schlüssels für den IKE-Proxy weiter abgesichert werden. Weitere Synergieeffekte würden sich durch eine Integration von Monitoring-Daten der QKD-Strecke (etwa Schlüsselrate oder Authentisierungsfehler) in bestehende Monitoring-Systeme für VPN-Infrastrukturen ergeben.

3.4 Transitive Ende-zu-Ende Sicherheit

Die bisher vorgestellten Schlüsselquellen sind jeweils darauf ausgelegt, Punkt-zu-Punkt-Verbindungen abzusichern, und skalieren nicht, um in großen VPN-Infrastrukturen paarweise individuelle Schlüssel für alle möglichen Verbindungen bereitzustellen und regelmäßig zu erneuern. Um nun auch

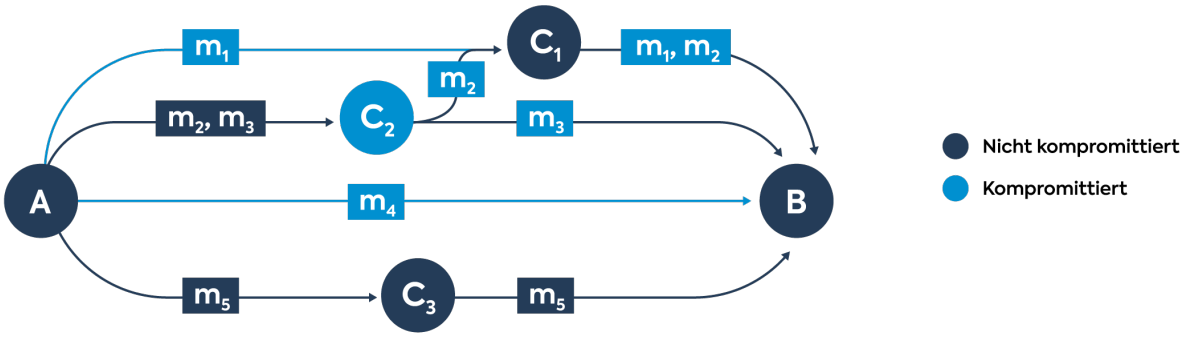


Abbildung 4: MKR-Schlüsselaustausch zwischen Gateways A und B mit insgesamt fünf Pfaden (fünf Teilschlüssel, jeweils Hop-by-Hop über zuvor etablierte Sicherheitsassoziationen übertragen). In diesem Beispiel gelingt es einem Angreifer durch die Kompromittierung einzelner Kanten und Knoten nur vier der fünf Teilschlüssel zu kompromittieren (abzuhören), jedoch nicht den Teilschlüssel m_5. Anzumerken ist noch, dass wir in diesem Beispiel davon ausgehen, dass zwar der Knoten C_2 kompromittiert ist, der Angreifer aber Pakete auf dem Transportweg von und zu C_2 nicht abhört (dunkelblaue Kanten). Dementsprechend sind die Teilschlüssel m_2 und m_3 erst als kompromittiert anzusehen, nachdem sie von C_2 weitergeleitet wurden.

Verbindungen zwischen zwei Gateways absichern zu können, die noch keinen paarweisen Schlüssel besitzen, müsste zwischen diesen ein Pfad existieren auf dem aktuell keine Sicherheitsbeziehung und auch kein VPN-Gateway kompromittiert ist. Existiert ein solcher Pfad, könnte entsprechend über diesen ein sicherer Schlüssel ausgetauscht werden.

In der Realität ist jedoch nicht bekannt, welche VPN-Gateways oder welche Sicherheitsbeziehung durch einen Angreifer zum jeweiligen Zeitpunkt kompromittiert sind. Um mit dieser Problematik umgehen zu können, gibt es den Ansatz, Schlüssel über möglichst viele unterschiedliche Pfade innerhalb des VPN zu verteilen und hierdurch den Aufwand und somit die Kosten für einen Angriff deutlich zu erhöhen. Dieser Ansatz ist auch unter dem Begriff „Multipath Key Reinforcement“ (MKR) bekannt, und wurde unter anderem bereits im Kontext von Wireless-Sensor-Netzen [13] sowie reinen QKD-Netzen [14] untersucht. In Abbildung 4 ist ein solcher Schlüsselaustausch zwischen zwei Gateways A und B dargestellt. Beide Knoten sind über mehrere Pfade miteinander verbunden. Da ihnen nicht bekannt ist, welche Pfade aktuell sicher sind, tauschen sie über jeden der verfügbaren Pfade einen unabhängigen Schlüssel aus. Diese Pfadschlüssel werden anschließend mittels einer KDF zu einem einzelnen MKR-Schlüssel kombiniert. Somit führt bereits ein einzelner sicherer Pfad zu einem sicheren MKR-Schlüssel.

Grundsätzlich lässt sich das beschriebene Verfahren direkt in kleineren und statischen VPNs einsetzen. In größeren VPNs und insbesondere in dynamisch verwalteten VPNs ergeben sich allerdings zwei Grundprobleme:

1. Die Topologie großer VPNs ist nicht statisch (Ausfall von Gateways, dynamischer Auf- und Abbau von Sicherheitsbeziehungen), sodass sich die verfügbaren Pfade jederzeit ändern können.

2. In größeren VPNs existieren gegebenenfalls sehr viele unterschiedliche Pfade zwischen zwei VPN-Gateways. Bei jedem MKR-Austausch alle verfügbaren Pfade zu verwenden, würde die Skalierbarkeit beeinträchtigten. Zur Wahrung der Skalierbarkeit muss deshalb die Anzahl der verwendeten Pfade pro Austausch limitiert werden. Dennoch sollte mittelfristig dieselbe Sicherheit gewährleistet werden, wie wenn immer alle Pfade verwendet würden.

Beide Probleme können leicht gelöst werden. Zwei VPN-Gateways können etwa mittels randomisierten und begrenzten Flutens Pfade zwischen einander suchen, ohne die Skalierbarkeit des VPNs zu beeinflussen. Ebenfalls kann diese Suche vollständig verteilt in den Gateways erfolgen, sodass keine zentrale Instanz benötigt wird. Abhängig von der Anzahl der gefundenen Pfade können die Gateways anschließend entweder alle oder nur eine Teilmenge der Pfade für den eigentlichen MKR-Schlüsselaustausch randomisiert auswählen.

4 Diskussion und Zusammenfassung

Bezüglich der in Kapitel 2 aufgeführten Anforderungen Krypto- und Schlüsselquellenagilität sowie Implementierungssicherheit lässt sich Folgendes festhalten:

Der hier vorgestellte Ansatz des IKE-Proxys bietet den Vorteil, symmetrisches Schlüsselmaterial aus unterschiedlichen Quellen einbeziehen zu können. Diese Eigenschaft ist von Vorteil, wenn die Sicherheit der verwendeten Schlüsselquellen nicht durch einen verständlichen Sicherheitsbeweis nachgewiesen werden kann. So wurde zum Beispiel kürzlich durch das BSI in einem neuen Positionspapier [7] die große Ungewissheit bezüglich der Sicherheit von QKD beziehungsweise aktueller QKD-Implementierungen deutlich. Würde der IKE-Proxy, so wie aktuell vorgeschlagene QKD-Netze, nur auf QKD-Schlüssel setzen, wären alle Vorteile durch diese Ungewissheit gefährdet. Durch die Kombination mit anderen Schlüsselquellen können hingegen selbst großflächige und weitreichende Sicherheitsprobleme, beispielsweise

in einer QKD-Implementierung, automatisch und transparent mitigiert werden.

Durch den konzeptionellen Ansatz, IKE-Pakete vor dem Versand über das Internet immer zu überschlüsseln, wird eine sogenannte „second line of defense“ aufgebaut. Durch diese Maßnahme muss ein Angreifer entweder den aktuellen Master-Schlüssel entwenden können, um direkt mit dem IKE-Daemon interagieren zu können, oder eine gravierende Sicherheitslücke in der Implementierung des IKE-Proxys identifizieren, durch die der gesamte Schutz des IKE-Proxy umgangen werden kann. Aufgrund der geringen Komplexität des Konzepts gehen wir allerdings von einer einfachen und nachvollziehbaren Implementierung aus. Sollte es dem Angreifer dennoch gelingen, Zugriff auf die IKE-Pakete im Klartext zu erlangen, muss er weiterhin die dort implementierten asymmetrischen kryptografischen Mechanismen brechen, um Client-Verkehr entschlüsseln zu können. Der allgemeine Nachteil einer „second line of defense“ ist allerdings, dass diese die Verfügbarkeit und Robustheit des abgesicherten Dienstes negativ beeinflussen kann. Konzeptionell wird diese Problematik durch das Schlüsselsynchronisationsprotokoll sowie die persistente Speicherung von Recovery-Schlüsseln adressiert, welche zusammen die Konsistenz der paarweisen Schlüsselketten sicherstellen. Zudem kann das Risiko eines Fehlerzustandes durch rigoroses Testen minimiert werden. Weiterhin können paarweise Schlüsselketten mit einem Gruppenschlüssel initialisiert werden, sodass auch eine Kommunikation zwischen VPN-Endpunkten stattfinden kann, die bisher keinen einzigen gemeinsamen paarweisen Schlüssel kennen.

Wird ein VPN zusätzlich mit dem IKE-Proxy und den in diesem Artikel vorgestellten Schlüsselquellen abgesichert, erhöhen sich die Kosten für einen Angreifer drastisch. Dies ist damit zu begründen, dass es nicht mehr ausreicht einen einzelnen Sicherheitsmechanismus anzugreifen. Stattdessen müssen alle im Folgenden aufgeführten Sicherheitsmechanismen umgangen werden, um Zugriff auf Client-Verkehr im Klartext zu erhalten:

- **Klassische asymmetrische Kryptografie**: Der Angreifer benötigt einen Quantencomputer, andernfalls kann er die eingesetzte asymmetrische Kryptografie nicht brechen.

- **PQC**: Der Angreifer benötigt Wissen über Schwachstellen in PQC-Algorithmen oder deren Implementierung, andernfalls kann ein sicherer Sitzungsschlüssel mit PQC ausgetauscht werden.

- **QKD**: Der Angreifer muss eine QKD-Verbindung ab der ersten Benutzung durchgängig kompromittieren, oder gravierende Implementierungsfehler identifizieren. Ist ihm dies nicht durchgängig möglich, wird sicheres Schlüsselmaterial mittels QKD ausgetauscht werden, welches anschließend in die paarweise Schlüsselkette im IKE-Proxy einfließt.

- **PSKs**: Der Angreifer muss jeden jemals ausgetauschten PSK, welcher in eine paarweise Schlüsselkette eingeflossen ist, abfangen oder nachträglich in Erfahrung bringen. Sofern ihm ein einzelner PSK fehlt, kann er die damit abgesicherte Verbindung nicht mehr abhören. Wird sichergestellt, dass PSKs nach der Verwendung auf sichere Weise gelöscht werden, kann der Angriff sogar nur noch während des manuellen Schlüsselaustausches erfolgen.

- **„Business Trip Key Exchange“**: Der Angreifer muss jeden mittels eines abgesicherten Laptops automatisch ausgetauschten Schlüssel, welcher in eine paarweise Schlüsselkette eingeflossen ist, in Erfahrung bringen. Hierzu muss jeder jemals verwendete Laptop kompromittiert werden. Wird sichergestellt, dass die Schlüssel nach dem Austausch auf den verwendeten Laptops auf sichere Weise gelöscht werden, muss ein Angriff während den Dienstreisen erfolgen. Demzufolge müssen alle Laptops vom Angreifer durchgängig kompromittiert sein.

- **MKR**: Der Angreifer muss im VPN entweder jede Verbindung durchgängig abhören oder jedes Gateway durchgängig kompromittieren. Ist dem Angreifer dies auch nur kurzzeitig nicht möglich, kann MKR durch seine transitive Eigenschaft das gesamte VPN schrittweise absichern. Demzufolge muss ein Angreifer darauf achten, dass ihm beim Abhören des VPN keine Pakete, etwa durch eine Überlast auf den Verbindungen zu seinem Rechenzentrum, verloren gehen.

Insgesamt sorgt die Kombination dieser Verfahren dafür, dass, sofern ein Angriff überhaupt noch möglich ist, dieser mit erheblichen logistischen und finanziellen Aufwänden für den Angreifer einhergeht.

Danksagung

Diese Veröffentlichung wurde durch dtec.bw – Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr gefördert (Projekt MuQuaNet). dtec.bw wird von der Europäischen Union – NextGenerationEU finanziert.

Literaturhinweise

[1] P. W. Shor, „Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer“, *SIAM J. Comput.*, Bd. 26, Nr. 5, S. 1484–1509, Okt. 1997, DOI: 10.1137/S0097539795293172.

[2] J. Proos und C. Zalka, „Shor’s Discrete Logarithm Quantum Algorithm for Elliptic Curves“, *Quantum Information and Computation*, Bd. 3, Nr. 4, S. 317–344, 2003.

[3] C. H. Bennett, E. Bernstein, G. Brassard, und U. Vazirani, „Strengths and Weaknesses of Quantum Computing“, *SIAM J. Comput.*, Bd. 26, Nr. 5, S. 1510–1523, Okt. 1997, DOI: 10.1137/S0097539796300933.

[4] European Union Agency for Cybersecurity, „Post-Quantum Cryptography: Current State and Quantum Mitigation.“, Publications Office of the European Union, 2021. doi: 10.2824/92307.

[5] W. Beullens, „Breaking Rainbow Takes a Weekend on a Laptop“, in *Advances in Cryptology – CRYPTO 2022*, Bd. 13508, Cham, 2022, S. 464–479. DOI: 10.1007/978-3-031-15979-4_16.

[6] C. H. Bennett und G. Brassard, „Quantum Cryptography: Public Key Distribution and Coin Tossing“, *Theoretical Computer Science*, Bd. 560, S. 7–11, Dez. 2014, DOI: 10.1016/j.tcs.2014.05.025.

[7] French Cybersecurity Agency, Federal Office for Information Security, Netherlands National Communications Security Agency, und Swedish National Communications Security Authority, Swedish Armed Forces, „Position Paper on Quantum Key Distribution“. Zugegriffen: 22. Februar 2024. [Online]. Verfügbar unter: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Crypto/Quantum_Positionspapier.html

[8] V. Smyslov, „Intermediate Exchange in the Internet Key Exchange Protocol Version 2 (IKEv2)“, Mai 2022. Zugegriffen: 22. Februar 2024. [Online]. Verfügbar unter: https://www.rfc-editor.org/rfc/rfc9242.html

[9] S. Fluhrer, P. Kampanakis, D. McGrew, und V. Smyslov, „Mixing Preshared Keys in the Internet Key Exchange Protocol Version 2 (IKEv2) for Post-quantum Security“, Juni 2020. Zugegriffen: 22. Februar 2024. [Online]. Verfügbar unter: https://www.rfc-editor.org/rfc/rfc8784

[10] V. Smyslov, „Alternative Approach for Mixing Preshared Keys in IKEv2 for Post-quantum Security“, Nov. 2022. Zugegriffen: 22. Februar 2024. [Online]. Verfügbar unter: https://datatracker.ietf.org/doc/draft-smyslov-ipsecme-ikev2-qr-alt/

[11] D. Schatz, F. Altheide, H. Koerfgen, M. Rossberg, und G. Schaefer, „Virtual Private Networks in the Quantum Era: A Security in Depth Approach“, in *Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT)*, 2023, S. 486–494. DOI: 10.5220/0012121800003555.

[12] ITU-T, „Overview on Networks Supporting Quantum Key Distribution“, Recommendation ITU-T Y.3800, 2019.

[13] J. Deng und Y. S. Han, „Multipath Key Establishment for Wireless Sensor Networks Using Just-Enough Redundancy Transmission“, *IEEE Trans. Dependable and Secure Comput.*, Bd. 5, Nr. 3, S. 177–190, Juli 2008, DOI: 10.1109/TDSC.2007.70233.

[14] S. Rass und S. König, „Indirect Eavesdropping in Quantum Networks“, in *Proceedings of the 5th ICQNM*, 2011, S. 83–88.

Bundesamt
für Sicherheit in der
Informationstechnik

Deutschland
Digital•Sicher•BSI•

BSI
20. Deutscher
IT-Sicherheitskongress

MITGLIEDER DES PROGRAMMBEIRATS

Dr. Jens Bender – Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit
Dr. Zinaida Benenson – Friedrich-Alexander-Universität Erlangen-Nürnberg
Prof. Dr. Christoph Busch – ATHENE-Darmstadt
Susanne Dehmel – Bitkom e.V.
Dr. Alexandra Gilsbach – TÜV-Verband
Dr. Sven Herpig – Stiftung Neue Verantwortung
Prof. Dr. Hartmut Ihne – Hochschule Bonn-Rhein-Sieg
Michael Jochem – Robert Bosch GmbH
Reinhard Karger – Deutsches Forschungszentrum für Künstliche Intelligenz
Prof. Dr. Stefan Katzenbeisser – Universität Passau
Prof. Dr. Klaus-Peter Kossakowski – Hochschule für Angewandte Wissenschaften Hamburg
Prof. Dr. Michael Meier – Rheinische Friedrich-Wilhelms-Universität Bonn
Dr. Gisela Meister – Eurosmart
Dr. Kim Nguyen – D-Trust GmbH / Bundesdruckerei
Prof. Dr. Wied Pakusa – Hochschule des Bundes für Öffentliche Verwaltung, Brühl
Prof. Dr. Ilia Polian – Universität Stuttgart
Prof. Dr. Reinhard Posch – Technische Universität Graz
Dr. Stefan Pütz –Deutsche Telekom AG
Dr. Jochen Rill – Alter Solutions Deutschland
Markus Schaffrin – eco Verband
Oberst Guido Schulte – Bundeswehr

Autorenverzeichnis

Stichwortverzeichnis