Fraunhofer IOSB

# Kurzbericht

| | |
|---|---|
| Vorhabenbezeichnung: | Verbundprojekt SILK: Bewertung von Sicherheitsvorfällen in kritischer Infrastruktur und Generierung von Handlungsempfehlungen für das Personal durch KI-basiertes Text-Mining<br><br>Teilvorhaben: Erforschung angewandter KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und Erkennung von Fehlverhalten |
| Partner: | Fraunhofer IOSB-INA |
| Autoren: | Dr.-Ing. Jens Otto<br>Felix Specht |
| Förderkennzeichen: | 13N15516 |
| Laufzeit des Vorhabens: | 01.08.2021 – 31.12.2024 |

# Kurzbericht: Erforschung angewandter KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und Erkennung von Fehlverhalten

## 1. Aufgabenstellung und technischer Stand

Energie- und Wasserversorgungssysteme sichern das wirtschaftliche und soziale Wohlergehen der Bevölkerung und sind nach der EU-Richtlinie 2008/114/EG kritische Infrastrukturen (kurz KRITIS). Der Schutz von kritischen Infrastrukturen vor Bedrohungen ist daher eine der wesentlichen Grundvoraussetzungen für Lebensqualität und Wertschöpfung in Deutschland. Energie- und Wasserversorgungssysteme bestehen aus verteilten und miteinander vernetzten Teilsystemen, die weitgehend durch ebenfalls verteilte Automatisierungskomponenten, wie z.B. Steuerungen, Sensoren und Aktoren, gesteuert und überwacht werden [1]. Der zuverlässige Betrieb dieser Systeme hängt u.a. direkt von dem Funktionieren der Kommunikationsnetze ab, d.h. diese Netze rücken in den Fokus von Maßnahmen zur Aufrechterhaltung von KRITIS. Mit Netzwerkanalysesystemen kann heute die Kommunikation zwischen Komponenten überwacht und Sicherheitsvorfälle detektiert werden. Solch ein Sicherheitsvorfall, wie der Ausfall einer Kommunikationsverbindung, kann eine Störung der Energie- und Wasserversorgung zur Folge haben.

In diesem Vorhaben wurde die Erforschung und Entwicklung eines KI-basierten Lösungsansatzes zur automatischen Bewertung von Ereignissen und zur Generierung von Handlungsempfehlungen für das Servicepersonal von KRITIS-Betreibern verfolgt. Dieser Ansatz wurde mit Methoden der Künstlichen Intelligenz – insbesondere maschinellen Lernverfahren - realisiert.

Die Aufgabenstellung dieses Teilvorhabens hatte das Ziel, angewandte KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und zur Erkennung von Fehlverhalten zu erforschen. Dieser Ansatz wurde durch Methoden der Künstlichen Intelligenz, insbesondere maschinelle Lernverfahren, realisiert. Die folgende Auflistung fasst die Teilziele des Teilvorhabens zusammen:

- Erforschung von KI-Verfahren zur Verarbeitung von Ereignisberichten mit domänenspezifischen Begriffen durch Natural Language Processing und Transferlernen
- Analyse von Netzwerkkommunikation zur Reduktion von falsch-positiven Sicherheitsereignissen durch KI-Erweiterung von Graphen-basierten Modellen
- Korrelation von Ereignismodell und Netzwerkmodell zur Bewertung von Sicherheitsvorfällen durch Ensemblemethoden und Suchalgorithmen
- Generierung von Handlungsempfehlungen mit Generativen Modellen zur verständlichen Darstellung für KRITIS-Personal

## 2. Ablauf des Vorhabens

Das Forschungsvorhaben startete im August 2021 durch das Projektkonsortium, bestehend aus Industrie- und Forschungspartnern. Die ursprüngliche Projektlaufzeit betrug drei Jahre und wurde im Jahr 2024 kostenneutral um fünf Monate verlängert. Das Projekt war von Abweichungen gegenüber der ursprünglichen Projektplanung betroffen, wodurch sich der Fokus der Arbeiten auf die Analyse von Logdaten änderte. Am Projekt beteiligt waren auf wissenschaftlicher Seite die Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg mit der Professur für Informatik im Maschinenbau sowie das Fraunhofer IOSB-INA in Lemgo mit den Arbeitsgruppen für „Cybersicherheit in der Produktion“. Auf Seiten der Industriepartner waren die Unternehmen Achtwerk GmbH &Co. KG, KASSELWASSER – Eigenbetrieb der Stadt Kassel und Sachenenergie AG (durch die Tochtergesellschaft DREWAG – Stadtwerke Dresden GmbH) am Vorhaben beteiligt.

Das Teilvorhaben gliederte sich inhaltlich in die Arbeitspakete: AP 1) Anforderungsanalyse und Konzeptentwicklung, (AP 2) Erforschung und Entwicklung der Künstlichen Intelligenz, (AP 3) Entwicklung und Implementierung eines Demonstrators und (AP 4) Evaluierung des Lösungsansatzes

im Realbetrieb. Dabei übernahm das Fraunhofer IOSB-INA die Leitung über die Arbeitspakete. Des Weiteren war der Projektpartner mit seiner Arbeitsgruppe für Cybersicherheit in der Produktion in allen Arbeitspaketen maßgeblich beteiligt. Nachfolgend werden die in den jeweiligen Arbeitspaketen durchgeführten Arbeiten kurz dargestellt.

Das AP 1 Anforderungsanalyse und Konzeptentwicklung hatte die Zielsetzung, die vollständigen Anforderungen aller beteiligten Partner, sowie weiterer interessierter Unternehmen zu erfassen und zu dokumentieren. Das AP 1 umfasste speziell die Themen Aufnahme der Netzwerkkommunikation mit standardisierten Dateiformaten, standardisierte Textformate, Formate und Ansätze zur Textgenerierung von Handlungsempfehlungen. Des Weiteren war die Konzeptentwicklung der Softwarekomponenten, die Definition und Analyse von Sicherheitsszenarien, und die Konzeptentwicklung für die Künstliche Intelligenz sowie die Personalintegration Teil des AP 1.

Im Arbeitspaket 2 lag der Schwerpunkt auf der Entwicklung der Kernmodule des selbstlernenden Sicherheitssystems. Aufgrund geänderter Rahmenbedingungen wurde anstelle der ursprünglich geplanten Analyse von Textdokumenten die Verarbeitung und Analyse von Lognachrichten im Syslog-Format in den Mittelpunkt gestellt. Die Datenerfassung und -Aufbereitung erfolgte mittels Python-Software, um strukturierte Templates aus unstrukturierten Syslog-Meldungen zu extrahieren. Für die Anomalieerkennung wurden bestehende Verfahren der IRMA®-Lösung genutzt, die bei Auffälligkeiten entsprechende Lognachrichten generieren. Darauf aufbauend wurde mit CyberLog ein neuartiger Ansatz zur effizienten Erkennung von Angriffsmustern in Logdaten entwickelt, der Log-Meldungen automatisch als normales oder abweichendes Verhalten klassifiziert.

Im Arbeitspaket 3 wurde ein funktionsfähiger Demonstrator des selbstlernenden Sicherheitssystems in der SmartFactoryOWL entwickelt und implementiert. Der Fokus lag auf der Erfassung, Aufbereitung und Analyse von Logdaten im Syslog-Format. Zunächst wurden Trainingsdatensätze auf Basis realer und synthetischer Logdaten erstellt, die verschiedene Sicherheitsszenarien abdecken. Anschließend wurde eine modulare Testumgebung mit realen Automatisierungskomponenten aufgebaut, um typische Betriebs- und Angriffsszenarien realitätsnah nachzustellen. Der finale Demonstrator integrierte die IRMA-Lösung sowie die entwickelten KI-Module zur automatischen Verarbeitung und Analyse von Logdaten, zur Anomalieerkennung und zur Generierung von Handlungsempfehlungen für das Personal. Die Funktionalität und Praxistauglichkeit des Gesamtsystems wurden durch strukturierte Testszenarien überprüft und die erzielten Ergebnisse zur weiteren Optimierung genutzt.

AP 4 Evaluierung des Lösungsansatzes im Realbetrieb hatte die Zielsetzung der Validierung des entwickelten Lösungsansatzes sowohl in der Testumgebung, als auch in der realen KRITIS-Umgebung hinsichtlich spezifizierter Evaluationskriterien. Im Speziellen umfasste dies die Spezifizierung von Evaluationskriterien und Definition eines Testplans und Evaluierung des Lösungsansatzes in der Testumgebung, Evaluierung des Lösungsansatzes in der KRITIS-Umgebung.

## 3. Projektergebnisse

Im Rahmen des Projekts fand ein regelmäßiger Austausch unter den Projektpartnern im Rahmen eines monatlichen Jour Fixe statt. Die im Projekt erarbeiteten Ergebnisse sind ein Demonstrator in der SmartFactoryOWL sowie wissenschaftliche Veröffentlichungen auf internationalen Tagungen [1-2]. Methoden wurden in den Vorlesungen „Advanced Topics in Algorithms“ und „Network Security“ an der Technischen Hochschule Ostwestfalen-Lippe und in den Weiterbildungsangeboten des Fraunhofer Lernlabors Cybersicherheit (Konsortium Produktion) integriert.

1. Specht, Felix; Otto, Jens; Ratz, Daniel: Generation of Synthetic Data to Improve Security Monitoring for Cyber-Physical Production Systems. In: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN), Lemgo, Germany, July 2023
2. Otto, Jens; Specht, Felix: Efficient Cyberattack Detection in Logs of Cyber-Physical Production Systems. In: 2025 IEEE 7th International Conference on System-Integrated Intelligence (SysInt), Bremen, Germany, June 2025

Fraunhofer IOSB

# Sachbericht

| | |
|---|---|
| Vorhabenbezeichnung: | Verbundprojekt SILK: Bewertung von Sicherheitsvorfällen in kritischer Infrastruktur und Generierung von Handlungsempfehlungen für das Personal durch KI-basiertes Text-Mining<br><br>Teilvorhaben: Erforschung angewandter KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und Erkennung von Fehlverhalten |
| Partner: | Fraunhofer IOSB-INA |
| Autoren: | Dr.-Ing. Jens Otto<br>Felix Specht |
| Förderkennzeichen: | 13N15516 |
| Laufzeit des Vorhabens: | 01.08.2021 – 31.12.2024 |

# Inhaltsverzeichnis

1. Einleitung ..... 3
1.1. Motivation und Aufgabenstellung ..... 3
1.2. Voraussetzungen des Vorhabens ..... 5
1.3. Planung und Ablauf des Vorhabens ..... 5
1.4. Wissenschaftlicher und technischer Stand zu Projektbeginn ..... 6
2. Durchgeführte Arbeiten und erzielte Ergebnisse ..... 8
Arbeitspaket 1: Anforderungsanalyse und Konzeptentwicklung ..... 8
Arbeitspaket 2: Erforschung und Entwicklung von Ansätzen der Künstlichen Intelligenz ..... 12
Arbeitspaket 3: Entwicklung und Implementierung eines Demonstrators ..... 16
Arbeitspaket 4: Evaluierung der Sicherheitslösung ..... 25
3. Wichtigsten Positionen des zahlenmäßigen Nachweises ..... 27
4. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten ..... 27
5. Nutzen und Verwertbarkeit ..... 28
5.1 Verwertbarkeit der Ergebnisse ..... 28
5.2 Zusammenarbeit mit anderen Stellen ..... 28
6. Fortschritt auf Gebiet des Vorhabens durch andere Stellen ..... 29
7. Veröffentlichungen ..... 29
Literaturverzeichnis ..... Fehler! Textmarke nicht definiert.

# Sachbericht: Erforschung angewandter KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und Erkennung von Fehlverhalten

## 1. Einleitung

### 1.1. Motivation und Aufgabenstellung

Energie- und Wasserversorgungssysteme sichern das wirtschaftliche und soziale Wohlergehen der Bevölkerung und sind nach der EU-Richtlinie 2008/114/EG kritische Infrastrukturen (kurz KRITIS). Der Schutz von kritischen Infrastrukturen vor Bedrohungen ist daher eine der wesentlichen Grundvoraussetzungen für Lebensqualität und Wertschöpfung in Deutschland. Energie- und Wasserversorgungssysteme bestehen aus verteilten und miteinander vernetzten Teilsystemen, die weitgehend durch ebenfalls verteilte Automatisierungskomponenten, wie z.B. Steuerungen, Sensoren und Aktoren, gesteuert und überwacht werden [1]. Der zuverlässige Betrieb dieser Systeme hängt u.a. direkt von dem Funktionieren der Kommunikationsnetze ab, d.h. diese Netze rücken in den Fokus von Maßnahmen zur Aufrechterhaltung von KRITIS. Mit Netzwerkanalysesystemen kann heute die Kommunikation zwischen Komponenten überwacht und Sicherheitsvorfälle detektiert werden. Solch ein Sicherheitsvorfall, wie der Ausfall einer Kommunikationsverbindung, kann eine Störung der Energie- und Wasserversorgung zur Folge haben.

In diesem Vorhaben wurde die Erforschung und Entwicklung eines KI-basierten Lösungsansatzes zur automatischen Bewertung von Ereignissen und zur Generierung von Handlungsempfehlungen für das Servicepersonal von KRITIS-Betreibern verfolgt (s. Abbildung 1). Dieser Ansatz wurde mit Methoden der Künstlichen Intelligenz – insbesondere maschinellen Lernverfahren - realisiert.

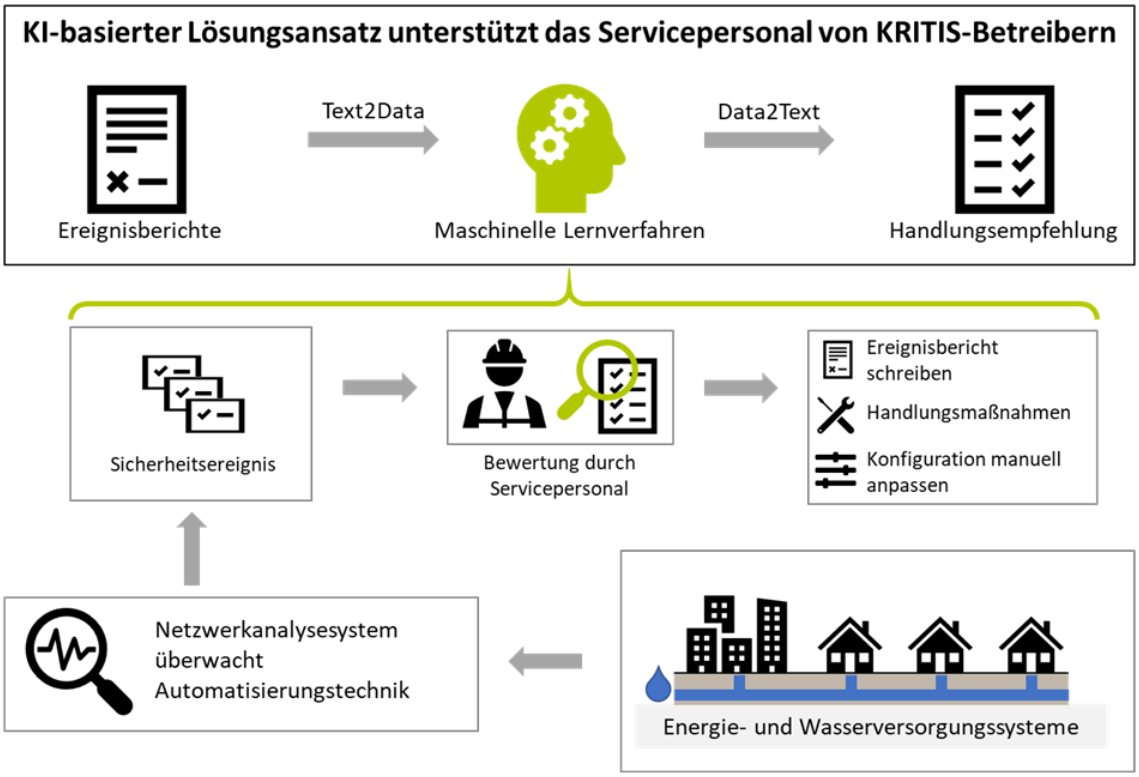



*Abbildung 1: KI-basierte Lösungsansatz unterstützt das Servicepersonal des KRITIS-Betreibers durch automatisch generierte Handlungsempfehlungen bei Sicherheitsvorfällen*

Die Aufgabenstellung dieses Teilvorhabens hatte das Ziel, angewandte KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und zur Erkennung von Fehlverhalten zu erforschen. Dieser Ansatz wurde durch Methoden der Künstlichen Intelligenz, insbesondere maschinelle Lernverfahren, realisiert. Die folgende Auflistung fasst die Teilziele des Teilvorhabens zusammen:

a) Erforschung von KI-Verfahren zur Verarbeitung von Ereignisberichten mit domänenspezifischen Begriffen durch Natural Language Processing und Transferlernen
b) Analyse von Netzwerkkommunikation zur Reduktion von falsch-positiven Sicherheitsereignissen durch KI-Erweiterung von Graphen-basierten Modellen
c) Korrelation von Ereignismodell und Netzwerkmodell zur Bewertung von Sicherheitsvorfällen durch Ensemblemethoden und Suchalgorithmen
d) Generierung von Handlungsempfehlungen mit Generativen Modellen zur verständlichen Darstellung für KRITIS-Personal

Teilziel a): Ein Ziel war die Erforschung von KI-Verfahren zur Verarbeitung von Textdokumenten von KRITIS-Betreibern. Diese Berichte sollten als Textdokumente vorliegen, die historische Sicherheitsvorfälle beschreiben. Da menschenformulierte Textdokumente keine eindeutig definierte Semantik haben, sollten der Einsatz von Natural Language Processing (NLP) und Transferlernen untersucht werden, um ein spezialisiertes Sprachmodell zu entwickeln, das auf die domänenspezifischen Begriffe in den Berichten abgestimmt war.

Teilziel b): In diesem Teilvorhaben wurde die Analyse der Netzwerkkommunikation in KRITIS-Betrieben untersucht, mit dem Ziel, die Anzahl von falsch-positiven Sicherheitsereignissen zu reduzieren. Bestehende Lösungen, wie die von Achtwerk, visualisierten die Netzwerkkommunikation als Graphen, wobei Knoten Geräte und Kanten Kommunikationsverbindungen darstellten. Da es häufig zu falsch-positiven Warnungen kam, insbesondere nach Softwareupdates, wurde erforscht, wie Graphen-basierte Netzwerkmodelle durch KI-Lösungen erweitert werden können. Process-Mining-Ansätze wurden in Betracht gezogen, um Kommunikationsmuster zu erkennen und die Zuordnung als Softwareupdate zu automatisieren.

Teilziel c): In diesem Teilziel wurde die Korrelation des Ereignismodells mit dem Netzwerkmodell angestrebt, um Sicherheitsvorfälle besser zu bewerten. Es wurde untersucht, wie die Modelle miteinander verknüpft werden können, entweder durch Ensemblemethoden, die verschiedene Lernalgorithmen kombinieren, oder durch Suchalgorithmen, die nach eindeutigen Identifikationsmerkmalen suchen. Diese Verfahren sollten helfen, die Modelle so zu verbinden, dass eine umfassende Bewertung der Sicherheitsvorfälle möglich wird, basierend auf den in Teilziel a und b gewonnenen Daten und Erkenntnissen.

Teilziel d): Schließlich wurden Ansätze zur Generierung von Handlungsempfehlungen für das KRITIS-Personal erforscht. Ziel war es, Textdokumente mit verständlich formulierten Arbeitsanweisungen zu erstellen, die für das Personal nachvollziehbar und umsetzbar sind. Dabei mussten technische Informationen, wie die Hardware-Adresse eines Geräts, durch verständliche Darstellungen ersetzt werden, beispielsweise durch Kurzbeschreibungen, Bilder und Positionsangaben. Generative Modelle wurden untersucht, da sie in der Lage sind, basierend auf wenigen Eingabedaten neue und komplexere Ausgabedaten zu erzeugen, die dem Personal klare und hilfreiche Anweisungen bieten.

## 1.2. Voraussetzungen des Vorhabens

Das Teilvorhaben „Erforschung angewandter KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und Erkennung von Fehlverhalten“ gehörte zum Verbundprojekt SILK. Zu den Partnern in SILK gehörten die Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg als Forschungspartner für maschinelles Lernen bzw. künstliche Intelligenz und Achtwerk GmbH & Co. KG, ein technologieorientiertes Unternehmen, das sich auf die Ausfallsicherheit von ICS durch Überwachung von Netzwerkkommunikation spezialisiert. Des Weiteren die Unternehmen KASSELWASSER – Eigenbetrieb der Stadt und SachsenEnergie AG (vertreten durch die Tochtergesellschaft DREWAG - Stadtwerke Dresden GmbH) als Betreiber von kritischer Infrastruktur als Projektpartner beteiligt. Die Koordination des Gesamtprojekts lag beim Fraunhofer IOSB-INA. Anforderungen der Industrie wurden durch die beteiligten Industrievertreter eingebracht. Reale Daten und Anwendungsszenarien wurden sowohl durch die SmartFactoryOWL des IOSB-INA als auch durch die beteiligten KRITIS-Betreiber bereitgestellt.

Das Teilvorhaben war am Fraunhofer IOSB-INA als Schnittstellenthema zwischen künstlicher Intelligenz und Cybersicherheit angesiedelt und wurde von der Arbeitsgruppe „Cybersicherheit in der Produktion“ bearbeitet. Diese Gruppe verfügt über langjährige Erfahrung und Expertise, die sie aus vorhergehenden Verbundprojekten wie dem Spitzencluster BMWI Mittelstand 4.0 Digital in NRW, dem Spitzencluster it's OWL, der BMBF-Initiative des Fraunhofer Lernlabor Cybersicherheit, sowie dem Projekt Hybrid AI Intrusion Prevention for Industrial Control Systems (BMBF HAIP) gewonnen hat. Zusätzlich wurden zahlreiche Projekte mit Industriepartnern durchgeführt, die sich mit typischen industriellen Anwendungsfällen wie Condition Monitoring, Predictive Maintenance oder Security Testing von Industriekomponenten sowie der Absicherung nach IEC 62443 befassen. Darüber hinaus stellte das Fraunhofer IOSB-INA dem Projekt SILK die SmartFactoryOWL (www.smartfactoryowl.de) zur Verfügung. Die SmartFactory OWL ist ein Reallabor mit 2000 qm Produktionsfläche, das diverse reale Produktionsanlagen und Demonstratoren aus den Bereichen Industrial Control Systems und Automatisierungstechnik umfasst und eine realistische Test- und Entwicklungsumgebung bietet.

## 1.3. Planung und Ablauf des Vorhabens

Das Forschungsvorhaben startete im August 2021 durch das Projektkonsortium, bestehend aus Industrie- und Forschungspartnern. Die ursprüngliche Projektlaufzeit betrug drei Jahre und wurde im Jahr 2024 kostenneutral um fünf Monate verlängert. Die folgende Tabelle gibt einen Überblick über den Ablauf des Projekts mit Arbeitspaketen und Meilensteinen.

| | 1. Jahr | | | | 2. Jahr | | | | 3. Jahr | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |
| AP 1 Anforderungsanalyse und Konzeptentwicklung | ■ | ■ | | | | | | | | | | |
| AP 2 Erforschung und Entwicklung von Ansätzen der KI | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | | |
| AP 3 Entwicklung und Implementierung eines Demonstrators | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| AP 4 Evaluierung des Lösungsansatzes im Realbetrieb | | | | | | | | ■ | ■ | ■ | ■ | ■ |
| Meilenstein | | | | | | | ◆ | | | | | |

Das AP 1 Anforderungsanalyse und Konzeptentwicklung hatte die Zielsetzung, die vollständigen Anforderungen aller beteiligten Partnern, sowie weiterer interessierter Unternehmen zu erfassen und zu dokumentieren. Das AP 1 umfasste speziell die Themen Aufnahme der Netzwerkkommunikation mit

standardisierten Dateiformaten, standardisierte Textformate, Formate und Ansätze zur Textgenerierung von Handlungsempfehlungen. Des Weiteren war die Konzeptentwicklung der Softwarekomponenten, die Definition und Analyse von Sicherheitsszenarien, und die Konzeptentwicklung für die Künstliche Intelligenz und die Konzeptentwicklung für die Personalintegration Teil des AP 1.

Im AP 2 Erforschung und Entwicklung einer Künstlichen Intelligenz wurde das Ziel verfolgt, einzelne Module gemäß der Anforderungsspezifikationen aus AP1 zu entwickeln. Das AP umfasste insbesondere die Datenerfassung und Datenaufbereitung, das Modelllernen aus Netzwerkdaten, das Modelllernen aus Textdaten, die Modellkorrelation und die Generierung von Handlungsempfehlungen für das Personal.

Das AP 3 Entwicklung und Implementierung eines Demonstrators verfolgte die Umsetzung eines funktionsfähigen Demonstrators des KI-Lösungsansatzes in der SmartFactoryOWL. Dies umfasste die Teilaspekte Generierung eines Trainingsdatensatz, Entwicklung einer Testumgebung in der SmartFactoryOWL, die Entwicklung des eigentlichen Demonstrators, das Thema Automatische Konfigurationsanpassung des Netzwerkanalysesystems und die Testdurchführung in der SmartFactoryOWL.

AP 4 Evaluierung des Lösungsansatzes im Realbetrieb hatte die Zielsetzung der Validierung des entwickelten Lösungsansatzes sowohl in der Testumgebung, als auch in der realen KRITIS-Umgebung hinsichtlich spezifizierter Evaluationskriterien. Im speziellen umfing dies die Spezifizierung von Evaluationskriterien und Definition eines Testplans und Evaluierung des Lösungsansatzes in der Testumgebung, Evaluierung des Lösungsansatzes in der KRITIS-Umgebung.

Weitere Details zu den Arbeitspaketen und Meilensteinen sind dem Projektantrag zu entnehmen.

## 1.4. Wissenschaftlicher und technischer Stand zu Projektbeginn

Der für dieses Projektvorhaben relevante Stand von Wissenschaft und Technik kann grundsätzlich in zwei Bereiche von maschinellen Lernverfahren gegliedert werden: 1. Modellierung technischer Systeme auf Basis von Netzwerkdaten, sowie 2. Informationsmodellierung aus Textdokumenten und Textgenerierung aus Modellen.

**1. Modellierung technischer Systeme auf Basis von Netzwerkdaten:** In den letzten Jahren wurden vermehrt Verfahren des maschinellen Lernens zur Modellierung von technischen Systemen und zur Bewertung von sicherheitsrelevanten Ereignissen eingesetzt. Die zum Lernen verwendete Datengrundlage bilden üblicherweise die Netzwerkdaten, bzw. die Prozessdaten (z.B. Sensorwerte und Steuerungssignale) von technischen Systemen. Dabei sind maschinellen Lernverfahren wie Deep Learning, aber auch klassische Verfahren im Einsatz [2-5]. Für technische Systeme mit kontinuierlichen Daten werden Differential- oder Differenzengleichungen eingesetzt [6]. Stochastische Modelle ermöglichen die Berücksichtigung von Unsicherheiten – Beispiele sind Kalman Filter, Hidden Markov Modelle und Gaußprozesse [7]. Als Modelle für Systeme kommen oft hybride Automaten [8] zum Einsatz, die kontinuierlich-diskrete Verhaltensmodelle abbilden [9, 10]. Bei der Modellierung technischer Systeme kann auf Vorarbeiten der Antragsteller verwiesen werden [11-15].

Viele Verfahren basieren auf dem Ansatz des überwachten Lernens. Das impliziert eine Klassenzuordnung von Bedrohungen. Unüberwachte Ansätze erlauben hingegen eine Erweiterung auf bisher unbekannte Bedrohungen, indem sie eine Abweichung vom Soll-Verhalten als Grundlage annehmen [16, 17]. Da im KRITIS-Bereich annotierte Daten, die ein überwachtes Lernen ermöglichen, selten vorliegen, sollen in diesem Forschungsprojekt unüberwachte Verfahren des maschinellen Lernens verwendet werden. Unüberwachte Verfahren detektieren Anomalien auf Basis eines Soll-Verhaltens. Hierbei kann auf Vorarbeiten der Antragsteller zurückgegriffen werden [18, 19].

**2. Informationsmodellierung aus Textdokumenten und Textgenerierung aus Modellen:** Um Ereignisberichte von KRITIS-Betreibern zu verarbeiten und Handlungsempfehlungen zu erzeugen, ist die

Informationsmodellierung auf Basis von Textdokumenten und die Textgenerierung aus Modellen relevant.

Dazu müssen u. A. zeitbehaftete kontinuierliche Signale diskretisiert werden, z.B. zum Zweck der Analyse des Zeitverhaltens und der Abfolge von Ereignissen. Dieses relativ neue Forschungsgebiet wurde bereits in unterschiedlicher Weise adressiert. Beispielsweise werden in [20, 21] Continuous-Time Bayesian Networks aus Daten generiert. In [22] wird der Ansatz des hierarchischen Clustering mit anschließender Erkennung verschiedener Modi im kontinuierlichen Systemverhalten verwendet.

Ein anderer Ansatz sind Neuronale Netze: Frühe Beispiele dafür sind Selbstorganisierende Karten, ein Überblick ist in [23] gegeben. Eine wichtige Arbeit zur Diskretisierung neuronaler Netze für sequentielle Daten wurde vor einiger Zeit in [24] publiziert. Hier wurden deterministische Automaten mit endlichen Zuständen aus den rekurrenten neuronalen Netzen zweiter Ordnung extrahiert. Diese Art der Regelextraktion ist Teil der jüngsten Forschung, da sie es ermöglicht, Wissen in symbolischer Form aus verschiedenen stetigen Modellen zu extrahieren [25, 26]. Die Diskretisierung komplexer Daten basierend auf Deep Learning basiert auf mehreren Ebenen der Feature-Extraktion. In [27] werden Sprachspektrogramme mithilfe eines Deep Autoencoder-Netzwerks diskretisiert. Darüber hinaus wird ein ähnliches Deep Belief Network, das aus mehrschichtigen Restricted Boltzmann Machines besteht, zum Extrahieren von Binärcodes aus Textdokumenten angewendet [28].

Frameworks zur Generierung von Texten aus Daten – bestehend aus den Bausteinen Signalanalyse, Interpretation, Ausgabeplanung und Realisierung – wurden bereits 2007 angedacht [29]. Aktuelle neuronale Verfahren, die eine Transformation der Eingabedaten zu Ausgabesätze lernen, verschmelzen jedoch diese Schritte [30]. Solche Sequence-to-Sequence Verfahren erzielen mittlerweile beachtliche Erfolge in der Textgenerierung [31]. Beispielsweise resultierte daraus die Entwicklung von Bidirectional Encoder Representations from Transformers (kurz BERT) und darauf aufbauende Verfahren [32, 33]. BERT ist weitestgehend als der neueste Stand der Technik im Bereich der natürlichen Sprachverarbeitung angesehen. Dies folgt einerseits aus den beachtlichen Ergebnissen in unterschiedlichen Benchmarks der natürlichen Sprachverarbeitung [34]. Andererseits sind vortrainierte BERT-Modelle frei verfügbar. Durch die Kombination mit Ansätzen des Transferlernens können somit in kürzester Zeit neue aufgabenspezifische Modelle mit hoher Leistung erzeugt und der Bedarf an großen Datenmengen reduziert werden.

**Erklärbarkeit von KI:** Maschinelle Lernverfahren agieren typischerweise als Black-Box-Verfahren, die keine Erklärung über eine getätigte Klassifikation eines Sicherheitsvorfalls liefern [35]. Beispielsweise steigern Ansätze auf Basis von Local Interpretable Model-Agnostic Explanations (kurz LIME) die Erklärbarkeit gegenüber dem Anwender enorm [36]. Im Projekt sollen daher vermehrt solche Verfahren des maschinellen Lernens eingesetzt werden, die eine intuitive Inspektion durch den Betreiber und den Einsatz von Expertenwissen erlauben.

**Robustheit von KI:** Insbesondere im KRITIS-Bereich ist die Robustheit der eingesetzten Lernverfahren wichtig und soll daher im Projekt behandelt werden. Dazu kann auf Vorarbeiten der Projektpartner im Bereich Schutz tiefer neuronaler Netze gegen Angriffe mit Adversarial Examples aufgesetzt werden [37], wie auch auf Arbeiten im Bereich Datenbeschreibung und Qualität [38].

Eine Patentrecherche hat gezeigt, dass Verfahren für eine Analyse von Netzwerkkommunikation angemeldet sind, die allerdings die im Projekt anvisierte Vorgehensweise nicht subsumieren (siehe Tabelle 1).

*Tabelle 1: Ergebnisse der Patentrecherche*

| Patentschrift | Anmerkung |
|---|---|
| | |

| | |
|---|---|
| US2016080398A<br>SYSTEM AND METHOD FOR DETECTION OF TARGETED ATTACKS | Behandelt die Auswertung von Indikatoren zur Erstellung einer Wahrscheinlichkeitsabschätzung für einen Angriff. |
| US2016308898A<br>SYSTEMS AND METHODS FOR TRACKING, ANALYZING AND MITIGATING SECURITY THREATS IN NETWORKS VIA A NETWORK TRAFFIC ANALYSIS PLATFORM | Aufzeichnung von Netzwerkverkehr und Analyse mithilfe von Deep-Packet-Inspection. Es sollen Netzwerkbewegungen erfasst werden und Sicherheitsvorfälle gespeichert werden. |
| US2018270082<br>METHOD AND DEVICE FOR MONITORING CONTROL SYSTEMS | Verfahren und Vorrichtung zur Überwachung von Steuerungssystemen der Kommunikation zwischen Endpunkten eines Netzwerks, sowie dem Vergleich aufgezeichneter Daten und Generierung von Alarmen. |
| PCT/EP2018/050028<br>SELF-CONFIGURING MONITORING DEVICE FOR AN AUTOMATION SYSTEM WHICH IS BASED ON AN INDUSTRIAL DATA COMMUNICATION NETWORK | Selbstkonfigurierende Überwachungseinrichtung für ein auf einem industriellen Datenkommunikationsnetzwerk basierendes Automatisierungssystem. |

## 2. Durchgeführte Arbeiten und erzielte Ergebnisse

Das Teilvorhaben wurde im Verbund mit Partnern einer Universität und der Industrie durchgeführt und gliederte sich inhaltlich in die Arbeitspakete: (AP 1) Anforderungsanalyse und Konzeptentwicklung, (AP 2) Erforschung und Entwicklung der Künstlichen Intelligenz, (AP 3) Entwicklung und Implementierung eines Demonstrators und (AP 4) Evaluierung des Lösungsansatzes im Realbetrieb. Dabei übernahm das Fraunhofer IOSB-INA die Leitung über die Arbeitspakete und führte dabei einen Großteil der notwendigen Arbeiten durch. Nachfolgend werden die in den jeweiligen Arbeitspaketen durchgeführten Arbeiten dargestellt.

### Arbeitspaket 1: Anforderungsanalyse und Konzeptentwicklung

Im AP 1 verfolgte das Fraunhofer IOSB-INA das Ziel, die vollständigen Anforderungen aller beteiligten Partnern sowie weiterer interessierter Unternehmen zu erfassen und zu dokumentieren. Im Einzelnen beinhalteten diese Arbeiten mehrere Teile:

#### AP 1.1 Anforderungen an die Datenerfassung:

Die Datenerfassung wird im Rahmen des Projektes durch unterschiedliche Softwarelösungen ermöglicht. Die Partner Achtwerk und Fraunhofer IOSB-INA stellten im Rahmen von gemeinsamen Workshops mögliche Lösungen für die Datenerfassung vor. In der Folge wurden für die Netzwerkdaten folgende Datenquellen festgelegt:

- PCAP-Format (bzw. darauf basierende kompatible Formate wie PCAPNG), aufgezeichnet mittels Wireshark oder ähnlichen Tools
- Netzwerkdaten aus dem IRMA-System
- Kommunikation der IRMA-Events mittels MessageBroker *RabbitMQ*.
- Syslog-Format (Standardisiertes Format für Text-basierte Logmeldungen)

Gemeinsam mit Achtwerk wurde eine Einführung in die Erfassung von Netzwerkdaten gegeben. Sowohl die Netzwerkanalyse mit Wireshark, als auch die eigens von Achtwerk entwickelte Lösung „IRMA (Industrie Risiko Management Automatisierung)“, wurde dabei den anderen Partnern vorgestellt. Die Lösung kann zur Datenerfassung und Vorverarbeitung von Daten im Rahmen des Projekts verwendet werden. Die nachfolgende Abbildung zeigt das zugrundeliegende Vorgehenskonzept für die Datenerfassung und Datenaufbereitung.

Des Weiteren wurde eine auf Open Source basierende Softwarelösung zur Datenerfassung von Netzwerkdaten vom Fraunhofer IOSB-INA vorgestellt. Diese basierte auf der Programmiersprache Python und der offenen Software Wireshark. Eine Anleitung (Tutorial) für die Verwendung der offenen Softwarelösung zur Datenerfassung wurde vom IOSB-INA erstellt und den Partnern bereitgestellt.

In einem mit den Partnern durchgeführten Workshop zur Anforderungsanalyse wurden mit den Partnern Kasselwasser und SachenEnergie folgende generelle Anforderungen ermittelt:

- Zuverlässige Erkennung von Cyberangriffen durch das IRMA-System
- Klassifizierung von Sicherheitsvorfällen
- Generische Alarm-Nachrichten des Systems sollen konkreten Vorfällen zugeordnet werden können
- Vorfilterung der Meldungen des IRMA-Systems
- Handlungsempfehlungen sollten nach Typ klassifiziert werden
- Meldungen aus dem IRMA-System müssen durch die KI aufbereitet und als Klartext an das Personal (IT-, bzw. Nicht-IT-Personal) formuliert werden
- KI sollte Handlungsempfehlungen mit maximal vier Maßnahmen für Nicht-IT-Personal aufbereiten

### AP 1.2 Konzeptionierung der Softwarekomponenten:

Die im Projekt zu entwickelnden Softwarekomponenten für eine KI-Lösung sollen generell einsetzbar sein und der Allgemeinheit zur Verfügung gestellt werden. Die Programmiersprache Python wurde für die gemeinsame Entwicklung ausgewählt. Die Softwarekomponenten sollen sowohl eigenständig nutzbar, als auch in das IRMA-System des Partners Achtwerk integrierbar sein.

Es wurden erste exemplarische Netzwerkdaten mittels Wireshark und mit dem IRMA-System im PCAP-Format aufgezeichnet. Eine erste Datenverarbeitung der aufgezeichneten PCAP-Daten wurde mit den Analysewerkzeugen Pyshark und Pandas (Bibliotheken der Programmiersprache Python) durchgeführt und bestätigt die Auswahl der Analysewerkzeuge. Im Projekt wurde ein Monitoring-Framework und ein Angriffs-Framework konzipiert, siehe nachfolgende Abbildung. Das Angriffs-Framework dient zur Generierung von Testdaten.

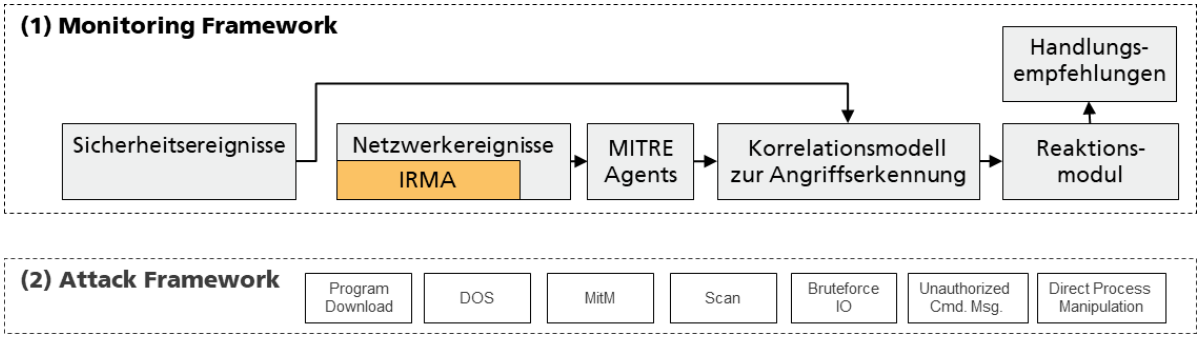



*Abbildung 2: Softwarekomponenten des Monitoring-Frameworks und des Angriffs-Frameworks*

### AP 1.3 Definition und Analyse von Sicherheitsszenarien:

In einem Workshop wurde zunächst ein Überblick über relevante Sicherheitsszenarien gegeben. Daraufhin wurde das *MITRE ATT&CK*-Framework zur Klassifikation von Cyberangriffen vorgestellt. Tabelle 2 zeigt eine exemplarische Liste von konkreten Angriffsszenarien, die in einer Testumgebung des Fraunhofer IOSB-INA nachgestellt werden können.

*Tabelle 2: Liste mit definierten Angriffsszenarien*

| **Angriffsszenario** | **Beschreibung** |
|---|---|
| Normal | Es ist kein Angriff vorhanden. Die normale Kommunikation zwischen einer industriellen Steuerung und einem IO-Gerät wird aufgezeichnet. Dieses Szenario dient zum Vergleich mit anderen Szenarien. |

| | |
|---|---|
| DoS ICMP (IO) | Ein Angriff vom Typ Denial-of-Service (DoS) wird gegen ein IO-Gerät durchgeführt. Dabei werden Netzwerkpakete vom Typ ICMP (Echo) verwendet. Der Angriff führt zu einer Verzögerung der Paketübertragungszeit und/oder zum Verlust von Netzwerkpaketen. |
| Profinet Replay (IO) | Ein Angriff vom Typ Replay wird gegen ein IO-Gerät durchgeführt. Ein zuvor aufgezeichnetes Netzwerkpaket vom Typ Profinet (Aus-Kommando) wird zu einem späteren Zeitpunkt erneut abgespielt. Der Angriff führt zur Beeinträchtigung des Produktionsprozesses (kontinuierliche Ausführung von An- und Aus-Kommandos). |
| Profinet Scan | Ein Angriff vom Typ Reconnaissance, durch einen Profinet-Scan, wird gegen sämtliche Geräte eines Netzwerks durchgeführt. Das Subprotokoll Profinet-DCP wird zur Detektion von Profinet-Geräten und zur Akquise von Informationen im Netzwerk missbraucht. |
| Port Scan | Ein Angriff vom Typ Reconnaissance, durch einen Scan von TCP-Ports, wird gegen eine industrielle Steuerung ausgeführt. Der Angriff erfolgt durch einen TCP-Verbindungsversuch zu einem TCP-Port der industriellen Steuerung, um Informationen über verfügbare Netzwerkdienste zu akquirieren. |
| Manipulierte Steuerungssoftware | Bei diesem Angriff wird das Steuerungsprogramm einer industriellen Steuerung durch ein manipuliertes Programm ausgetauscht. Das manipulierte Steuerungsprogramm verändert das Zeitverhalten bei der Verarbeitung von Sensorsignalen. Dies hat eine Beeinträchtigung des Produktionsprozesses zur Folge (Verlangsamung der Produktionsschritte). |
| ARP Cache Poisoning | Ein Angriff vom Typ *Man-in-the-Middle* wird durch eine Manipulation des ARP-Speichers gegen ein IO-Gerät und eine industrielle Steuerung durchgeführt. Dabei wird die Kommunikation zwischen der industriellen Steuerung und dem IO-Gerät über einen Angreifer umgeleitet. |

Gemeinsam mit den Projektpartnern wurde sich darauf geeinigt, das *MITRE ATT&CK*-Framework im Rahmen des Projekts zu verwenden. Entsprechend wurde festgelegt, dass die im Projekt zu entwickelnde KI-Lösung Cyberangriffe gemäß dem Framework klassifizieren soll.

### AP 1.4 Konzeptentwicklung für die Künstliche Intelligenz:

Literaturrecherche über den Stand der Technik zu selbstlernenden Mustererkennungsalgorithmen in Netzwerkdaten und Textdaten, Identifizierung und Vorauswahl geeigneter Algorithmen zur Erweiterung der Trainingsdatensätze, Konzeptionierung der Modellkorrelation, indem u.a. Ansätze aus der Informationsfusion dafür in Betracht gezogen werden, Konzeptentwicklung zur Anreicherung der KI-Komponenten mit Erklärungskomponenten.

Das AP 1.4 war im Laufe des Jahres 2023 von Abweichungen ggü. der ursprünglichen Projektplanung betroffen. Die nachfolgenden Arbeiten wurden durchgeführt:

Das Konzept für die künstliche Intelligenz wurde um eine Process-Mining-Komponente erweitert. Die Algorithmen basieren auf der frei zugänglichen pm4py[1] Softwarebibliothek des Fraunhofer FIT. Ein Process-Mining-Algorithmus benötigt ein sogenanntes Event log. Dadurch muss zusätzlich eine Transformationskomponente entwickelt werden, um Syslog zu verarbeiten. Durch die Komponente können Modelle für das Normalverhalten und Modelle für verschiedene Angriffsverhalten gelernt werden (Lernmodus). Die Modelle können dann mit dem aktuellen Systemzustand (Überwachungsmodus) abgeglichen werden. Zur Klassifizierung wird das Modell bestimmt, welches am wahrscheinlichsten den aktuellen Systemzustand widerspiegelt.

[1] https://pm4py.fit.fraunhofer.de/

**Beispiel: Sicherheitsszenario Adversary-in-the-Middle**[2]

Für das Sicherheitsszenario Adversary-in-the-Middle beschreibt *MITRE ATT&CK*-Framework[2] notwendige Aktivitäten, die erkannt werden müssen, um ein Modell des Adversary-in-the-Middle mit der Process-Mining-Komponente zu lernen. Dadurch können z.B. folgende Angriffsvorfälle abgebildet werden: (1) Attack device, (2) Remote code execution und (3) Brute-Force SSH.

*Tabelle 2: Zuweisung von Diagnosefähigkeiten und erkannten Aktivitäten*

| ID | Komponente | Aktivitäten |
|---|---|---|
| DS0015 | Application Log Content | ▪ Configuration changed (Webserver)<br>▪ Login failed (SSH)<br>▪ Login (SSH)<br>▪ Application crashed (Webserver)<br>▪ Web request |
| DS0029 | Network Traffic Content | ▪ ARP cache poisoning<br>▪ Rogue DHCP |
| | Network Traffic Flow | Unknown host |
| DS0009 | Process Creation | Process started (iptables) |
| DS0019 | Service Creation | Service started (dhcp) |
| DS0024 | Windows Registry Key Modification | Registry changed (LLMNR disabled) |

Ein Trainingsdatensatz des Angriffs "Adversary-in-the-Middle" ist in Abbildung 3 (links) dargestellt, während ein entsprechender Testdatensatz rechts abgebildet ist.

```
case:concept:name;concept:name;time:timestamp
1;Unknown host;2023-05-23 12:00:00
1;Rogue DHCP;2023-05-23 12:10:00
2;Web request;2023-05-23 13:00:00
2;Process started (sh);2023-05-23 13:10:00
2;Process started (iptables);2023-05-23 13:20:00
2;Configuration changed (iptables);2023-05-23 13:30:00
2;ARP cache poisoning;2023-05-23 13:30:05
…
4;Login failed (SSH);2023-05-23 14:30:05
4;Login (SSH);2023-05-23 14:40:00
4;Process started;2023-05-23 14:40:05
4;Unknown host;2023-05-23 14:40:10
4;Rogue DHCP;2023-05-23 14:40:15
```

```
case:concept:name;concept:name;time:timestamp
1;Unknown host;2023-05-23 12:00:00
1;Rogue DHCP;2023-05-23 12:10:00
2;Rogue DHCP;2023-05-23 16:10:00
3;Web request;2023-05-23 13:00:00
3;Process started (sh);2023-05-23 13:10:00
3;Process started (iptables);2023-05-23 13:20:00
…
3;ARP cache poisoning;2023-05-23 13:30:05
```

*Abbildung 3: Trainingsdatensatz (links) und Testdatensatz (rechts) des Angriffs Adversary-in-the-Middle*

Die Process-Mining-Komponente verwendet diesen Datensatz, um ein Modell des Angriffsverhaltens zu lernen (Lernmodus), siehe nachfolgende Abbildung 4.

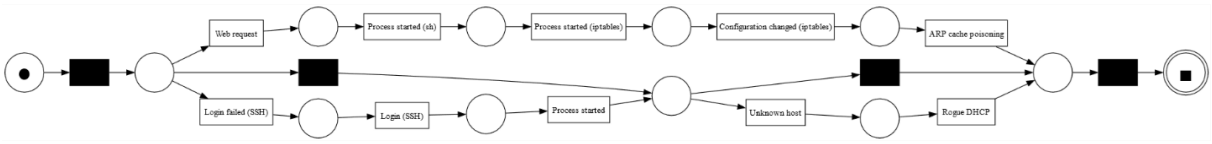



*Abbildung 4: Gelerntes Modell (Petri-Netz) für den Angriff Adversary-in-the-Middle*

Das Modell wird nun mit dem aktuellen Systemzustand (Überwachungsmodus) in Form des Testdatensatzes abgeglichen, siehe Abbildung 5.

[2] https://attack.mitre.org/techniques/T0830/

| Variant | Alignment | | | | | | |
|---|---|---|---|---|---|---|---|
| Variant 1 (1 occurrences) | (MM)None | (MM)None | Unknown host | Rogue DHCP | (MM)None | | |
| Variant 2 (1 occurrences) | (MM)None | (MM)None | (MM)Unknown host | Rogue DHCP | (MM)None | | |
| Variant 3 (1 occurrences) | (MM)None | Web request | Process started (sh) | Process started (iptables) | Configuration changed (iptables) | ARP cache poisoning | (MM)None |

*Abbildung 4: Berechnete Zuordnung des Testdatensatz zum gelernten Modell*

Die Einbindung von Syslog und eine Schnittstelle in die IRMA Lösung von Achtwerk wurde geplant und implementiert. Achtwerk stellt dem IOSB-INA die Schnittstelle bereit, um Syslog an die Transformationskomponenten anzubinden.

### AP 1.5 Konzeptentwicklung für die Personalintegration:

Durchführung eines Workshops mit dem Personal der KRITIS-Betreiber, Anforderungsdefinition aus Sicht der späteren Benutzer, Konzeptdefinition für eine entsprechenden Schnittstelle für den Austausch zwischen KI-Software und dem Personal.

**Anmerkungen zu AP 1.5**

Für das Personal werden folgende Informationen durch das Monitoring-Framework aufbereitet:

- Grad der Bedrohung
- Betroffene Systembereiche
- Mögliche Aktionen
- Einzuhaltende Meldekette

## Arbeitspaket 2: Erforschung und Entwicklung von Ansätzen der Künstlichen Intelligenz

Das Ziel des AP 2 für das Fraunhofer IOSB-INA war die Entwicklung der einzelnen Module der selbstlernenden Sicherheitslösung gemäß der Anforderungsspezifikationen aus AP1.

Das AP war von Abweichungen ggü. der ursprünglichen Projektplanung betroffen. Im Projektverlauf hat sich ergeben, dass die Anzahl der Textdokumente der KRITIS-Betreiber nicht in ausreichender Menge verfügbar ist, um eine sinnvolle Analyse durchzuführen. Der Ansatz des Arbeitspaketes wurde daher angepasst und die Entscheidung zwischen dem Konsortium und dem Projektträger schriftlich festgehalten. Nach Änderungen in der Projektplanung, wurde der Fokus im Projekt auf die Erfassung und Verarbeitung von Log-Nachrichten im Syslog-Format abgeändert.

### AP 2.1 Datenaufbereitung:

Die Datenerfassung und Datenaufbereitung wurde als Software in der Programmiersprache Python implementiert, welche Daten aus verschiedenen Datenquellen und Formaten einlesen kann:

- Log-Meldungen (Syslog) aus der IRMA (Lösung des Unternehmens Achtwerk)
- Log-Meldungen (Syslog) aus lokalen Dateien

Bei der Datenaufbereitung von Syslog-Meldungen ergaben sich mehrere Herausforderungen. Zunächst sind Syslog-Meldungen oft unstrukturiert und können in ihrer Form stark variieren, was die automatisierte Verarbeitung erschwert. Zudem enthalten die Meldungen komplexe Informationen, die nicht immer eindeutig interpretierbar sind, was eine sorgfältige Analyse erfordert, um relevante Daten zu identifizieren und zu extrahieren. Die Vielfalt der Formate stellt eine weitere Herausforderung dar, da Syslog-Meldungen aus unterschiedlichen Systemen und Anwendungen stammen, die jeweils eigene Formate und Protokolle verwenden. Eine einheitliche Verarbeitung dieser verschiedenen Formate ist daher kompliziert. Darüber hinaus fehlt eine Standardisierung für die Struktur und den Inhalt von Syslog-Meldungen, was die Entwicklung von universellen Templates erschwert. Schließlich enthalten Syslog-Meldungen oft eine Vielzahl von irrelevanten Informationen, die herausgefiltert werden müssen, um die Effizienz der Datenverarbeitung zu steigern. Das Ziel im AP 2.1

war es, aus diesen unstrukturierten, teilweise komplexen Syslog-Meldungen einfache Templates zu extrahieren, die eine strukturierte und konsistente Darstellung der relevanten Informationen ermöglichen.

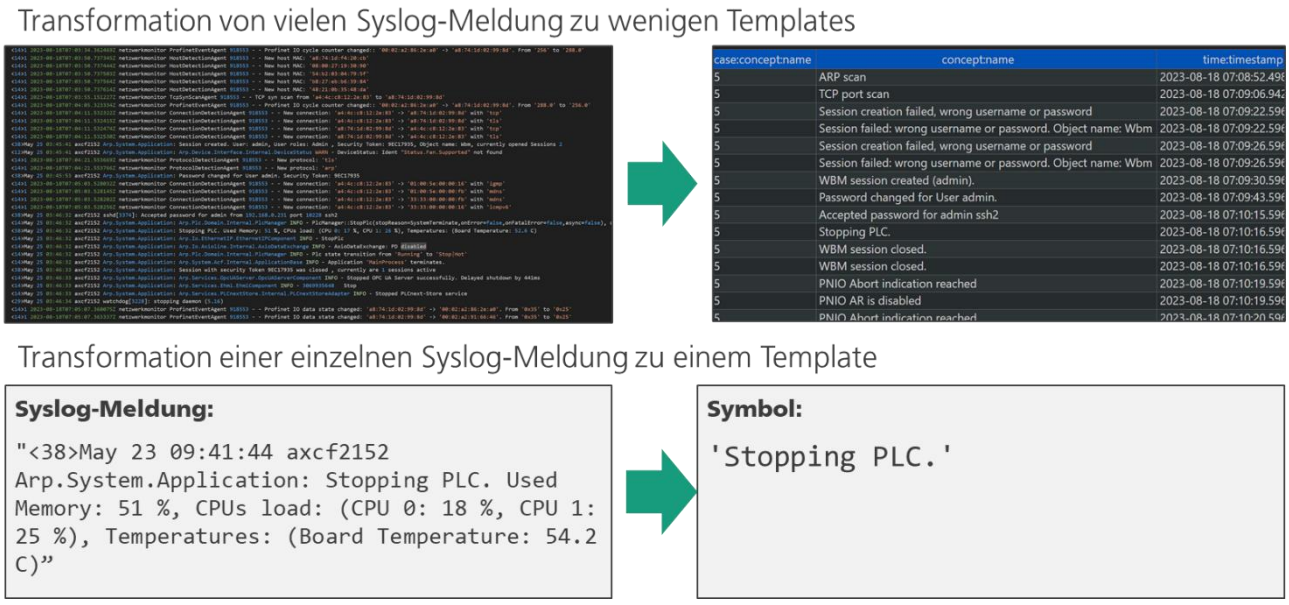



*Abbildung 5: Syslog-Meldung werden in vereinfachte Templates umgewandelt*

### AP 2.2 Modelllernen aus Netzwerkdaten:

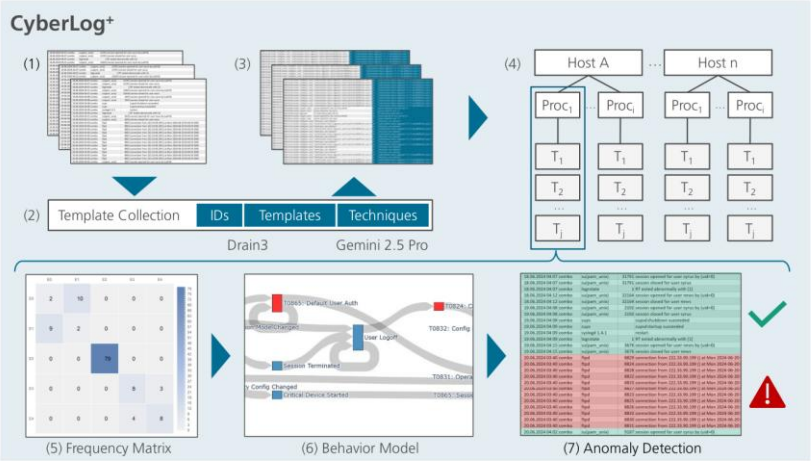
Das Arbeitspaket war von Abweichungen gegenüber der ursprünglichen Projektplanung betroffen. Aufgrund des verschobenen Fokus hin zur Analyse von Log-Nachrichten wurde der ursprünglich geplante wissenschaftliche Ansatz nicht weiterverfolgt. Stattdessen kamen bestehende Verfahren zur Anomalieerkennung der IRMA® zum Einsatz, die Netzwerkdaten analysiert und bei Auftreten von Anomalien entsprechende Log-Nachrichten generiert. Die IRMA® ist in der Lage, Sicherheitsvorfälle zu erkennen und passende Log-Nachrichten zu erstellen, wie in Tabelle 2 dargestellt. Diese Log-Nachrichten können anschließend für das Modellernen aus Textdaten im Arbeitspaket 2.3 und 2.4 genutzt werden. Durch diese Anpassung wurde sichergestellt, dass die Analyse von Sicherheitsvorfällen weiterhin effizient und zielgerichtet durchgeführt werden konnte, indem bewährte Technologien zur Anomalieerkennung integriert wurden. Die Entscheidung, auf bestehende Lösungen zurückzugreifen, ermöglichte eine fokussierte Weiterentwicklung und Anpassung der Projektziele.

### AP 2.3 Modelllernen aus Textdaten:

Das Arbeitspaket war ebenfalls von Abweichungen gegenüber der ursprünglichen Projektplanung betroffen. Anstelle der ursprünglich vorgesehenen Textdokumente mit Beschreibungen von Sicherheitsvorfällen wurden Log-Nachrichten im Syslog-Format verwendet. Log-Nachrichten sind ebenfalls Texte in natürlicher Sprache und können zum Training von Modellen genutzt werden. Die Arbeiten des AP 2.3 wurden in einem wissenschaftlichen Beitrag zum Thema „Erkennung von Cyberangriffen auf cyber-physische Produktionssysteme auf Basis von Logdaten“ auf einer internationalen Konferenz vorgestellt und sind unter der folgenden Referenz zu finden.:

*Jens Otto and Specht, Felix. "Efficient Cyberattack Detection in Logs of Cyber-Physical Production Systems." 7th International Conference on System-Integrated Intelligence (SysInt), 2025.*

Die nachfolgende Abbildung zeigt die Kernidee des veröffentlichten Ansatzes:




| ID | Technique |
|---|---|
| T0806 | Brute Force I/O |
| T0813 | Denial of Control |
| T0815 | Denial of View |
| T0822 | External Remote Services |
| T0824 | UNKNOWN |
| T0831 | Manipulation of Control |
| T0832 | Manipulation of View |
| T0836 | Modify Parameter |
| T0837 | Loss of Protection |
| T0846 | Remote System Discovery |
| T0859 | Valid Accounts |
| T0865 | Spearphishing Attachment |
| T0886 | Remote Services |

MITRE ATT&CKs

*Abbildung 6: CyberLog⁺ ist ein Ansatz zu Erkennung von Angriffsmustern in Log-Nachrichten*

Im Arbeitspaket wurde der Ansatz CyberLog⁺ entwickelt, welcher auf der wissenschaftlichen Konferenz vorgestellt wurde. Dabei handelt es sich um einen neuartigen Ansatz zur effizienten und automatischen Erkennung von Cyberangriffen in großen Mengen von Log-Daten. CyberLog⁺ verarbeitet Log-Daten aus Komponenten eines cyber-physischen Produktionssystems und liefert annotierte Log-Daten, die jedes Log-Eintrag entweder als normales Verhalten oder als Anomalie kennzeichnen. Der Ansatz umfasst sieben Schritte:

1. Die Erfassung von Rohlogdaten, bei denen jede Log-Nachricht einen Zeitstempel, den Komponentennamen, den Anwendungsnamen und eine Log-Nachricht enthält, die ein entsprechendes Ereignis beschreibt.
2. Extraktion von Log-Templates, die gemeinsame Muster für bestimmte Arten von Log-Nachrichten repräsentieren.
3. Zuordnung der Log-Templates zu jeder Log-Nachricht.
4. Verteilung der Log-Nachrichten nach Host und Prozess, gefolgt von der Aufteilung in Zeitfenster.
5. Für jede Kombination von Host und Prozess wird eine Frequenzmatrix durch Zählen der Übergänge von einem Log-Template zu einem anderen gelernt.
6. Transformation jeder Frequenzmatrix in einen gerichteten Flussgraphen, der die normale Abfolge von Log-Nachrichten darstellt, die von einer Anwendung generiert werden.
7. Nach Abschluss der Trainingsphase wird das trainierte Verhalten mit dem beobachteten Verhalten verglichen, um Anomalien zu erkennen.

Eine ausführliche Beschreibung ist dem veröffentlichten Beitrag zu entnehmen.

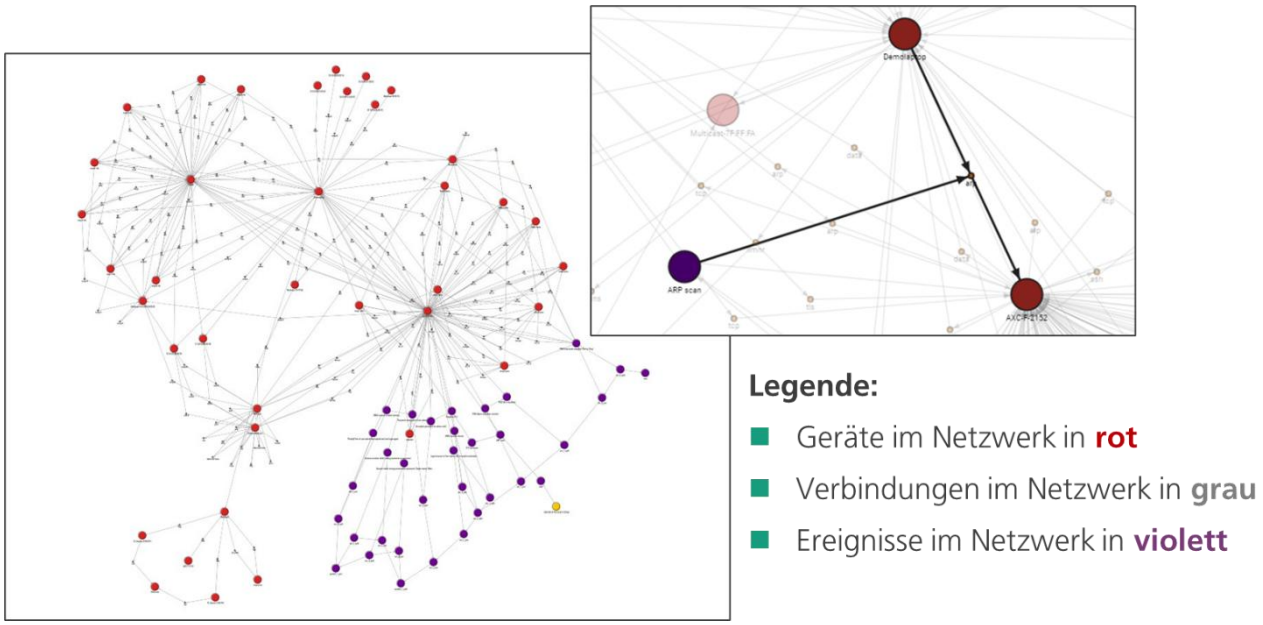
### AP 2.4 Modellkorrelation:

Ziel dieses Arbeitspakets ist die strukturierte Zusammenführung und Verknüpfung unterschiedlicher netzwerkrelevanter Datenquellen (Topologie, Verbindungen, Ereignisse) in einer Graph-Datenbank. Auf Basis dieser Datenbank sollen mittels gezielter Abfragen die Zusammenhänge zwischen physikalischer und logischer Netzwerktopologie, einzelnen Geräten und den generierten Ereignis- bzw. Log-Daten transparent gemacht sowie visualisiert werden. Dadurch wird eine fundierte Grundlage für Analysen, Alarmierung und Berichterstattung geschaffen.

Konkret wurden dabei folgende Schritte sukzessive umgesetzt:

1. **Erfassung und Import der Topologieinformationen** – Alle relevanten Gerätedaten (Hostname, IP-Adressen, Rolle und Standort) wurden aus der IRMA-Datenbank extrahiert und in die Graph-Datenbank überführt. Durch eine saubere Datenaufbereitung ließ sich jeder Netzwerk-Knoten eindeutig als „Gerät" abbilden, sodass spätere Abfragen zuverlässig darauf zugreifen können.
2. **Integration der Verbindungsinformationen** – Anschließend wurden sämtliche Netzwerk-Links in Form von Kanten zwischen den Geräte-Nodes importiert. Diese Verknüpfungen ermöglichen es, Kommunikationspfade und Abhängigkeiten zwischen einzelnen Komponenten lückenlos nachzuvollziehen.
3. **Einspielung der Ereignisdaten (Ergebnisse aus AP 2.3)** – Die aus dem Modelllernen stammenden Log-Meldungen, Anomalie-Alerts und Zeitstempel-Einträge wurden als separate Ereignis-Knoten in die Datenbank gebracht und jeweils mit den betroffenen Geräten verknüpft. So ist sichergestellt, dass jedes Ereignis direkt auf den korrekten Host oder Netzwerkpunkt referenziert und kontextualisiert wird.
4. **Definition und Ausführung gezielter Abfragen** – Mit speziell entwickelten Datenbankabfragen wurden Korrelationen zwischen Topologie, Geräten und Log-Nachrichten realisiert. Beispiele hierfür sind das Nachverfolgen eines bestimmten Alarmtyps entlang des Netzwerkpfades oder das Auffinden von Hotspots, an denen Ereignisse gehäuft auftreten.

Die nachfolgende Abbildung zeigt die Visualisierung nach einer Anfrage an die Graph-Datenbank, welche die Zusammenhänge zwischen der Topologie, einzelnen Geräten und den Log-Nachrichten darstellt. In der Grafik sind die Geräte als rote Punkte zu erkennen, die Verbindungen in dezentem Grau, und die Ereignisse in kontrastreichem Violett hervorgehoben – so wird auf einen Blick deutlich, welche Gerätepfade bei einem bestimmten Log-Eintrag involviert sind und wie sich ein Vorfall räumlich innerhalb des Netzwerks ausbreitet.




### AP 2.5 Handlungsempfehlung für das Personal:

Identifikation eines geeigneten Vokabulars mit Basiskomponenten für Erklärungsmechanismen, Erweiterung um Erklärungsmechanismen, Erforschung und Entwicklung von Algorithmen zur Textgenerierung auf Basis der korrelierten Modelle und des ausgewählten Vokabulars, Aufbereitung der Ergebnisse für den Anlagenbetreiber (Ereigniserklärung, -bewertung und Handlungsempfehlung).

Eine Anforderungsdefinition aus Sicht der späteren Benutzer wurde erstellt. Ein Konzept für eine entsprechende Schnittstelle für den Austausch zwischen KI-Software und dem Personal wurde definiert.

Eine Anbindung an ein Leitsystem zur Übertragung von Alarmmeldungen durch Achtwerk ist geplant. Das IOSB-INA hat eine prototypische Umsetzung einer Weboberfläche zur Interaktion mit den Algorithmen des Monitoring-Frameworks implementiert, siehe nachfolgende Abbildung. Die Inhalte der Weboberfläche sollen zukünftig in der Oberfläche der IRMA® eingebettet werden. In diesem Arbeitspaket kam es zu keinen Abweichungen.

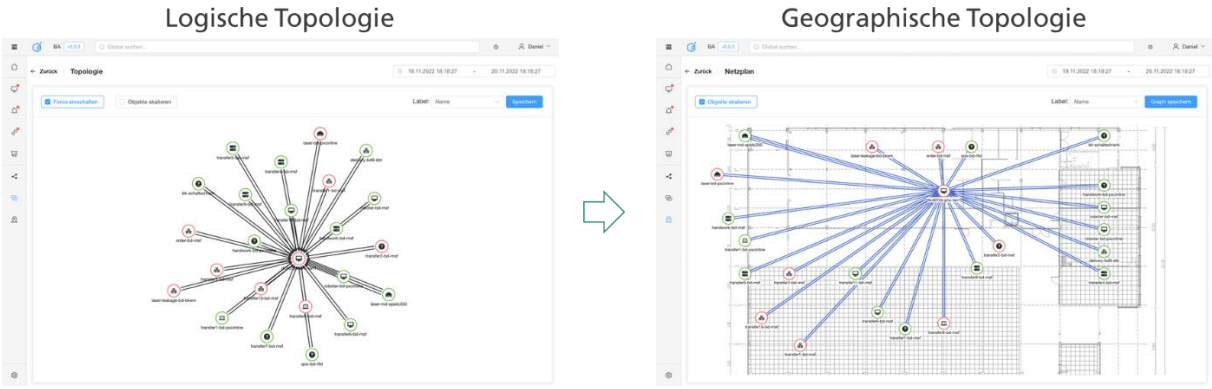



*Abbildung: Prototypische Umsetzung einer Weboberfläche zur Interaktion mit den Algorithmen des Monitoring-Frameworks*

Aufgrund der Abweichungen ggü. der ursprünglichen Projektplanung, wurden weitergehende Tätigkeiten in diesem Arbeitspakete nicht näher verfolgt.

## Arbeitspaket 3: Entwicklung und Implementierung eines Demonstrators

Im AP 3 gab es für das Fraunhofer IOSB-INA die Zielsetzung: Es sollte ein funktionsfähiger Demonstrator des selbstlernenden Sicherheitssystems in der SmartFactoryOWL entwickelt und implementiert werden. Im Einzelnen beinhalteten diese Arbeiten folgende Teile:

### AP 3.1 Generierung von Trainingsdaten:

Erstellung eines Trainingsdatensatz auf Basis von den in AP 1.3 definierten Sicherheitsszenarien, Sammeln von realen Kommunikationsdaten und Textdokumenten mit Ereignisberichten, Kontinuierliche Erweiterung der Trainingsdatensätze mit Realdaten und generierten Daten.

Es sind für die zwei definierten Szenarien: (1) Zulauf-Pumpwerk KW und (2) Alaunplatz DREWAG Datensätze erstellt. Die Datensätze liegen als Syslog und als Packet Capture (PCAP) vor. Die PCAP-Dateien erlauben die Reproduktion der Ergebnisse.

Ein wissenschaftlicher Beitrag zum Thema Generierung von Trainingsdaten wurde 2023 auf einer internationalen Konferenz vorgestellt und ist unter der nachfolgenden Referenz zu finden:

*Specht, Felix, Jens Otto, and Daniel Ratz. "Generation of Synthetic Data to Improve Security Monitoring for Cyber-Physical Production Systems." 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). IEEE, 2023.*

Die nachfolgende Abbildung zeigt die Kernidee des veröffentlichten Ansatzes:

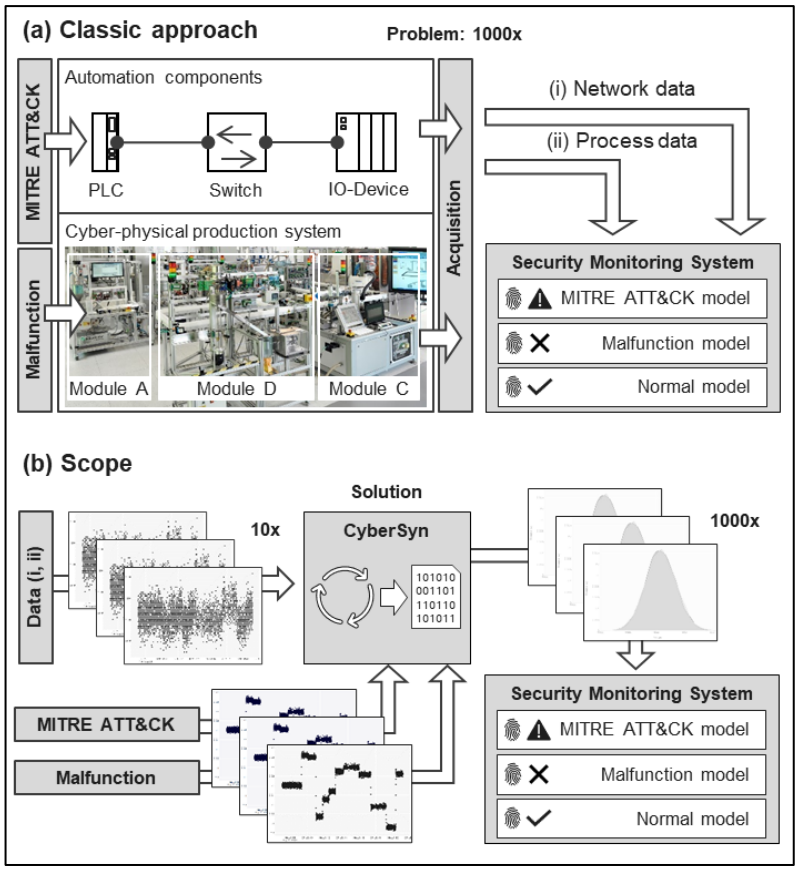



*Abbildung: Die allgemeine Idee des Ansatzes zur Erzeugung von synthetischen Daten zur Verbesserung der Sicherheitsüberwachung für cyber-physische Produktionssysteme.*

Es wurden wissenschaftliche Arbeiten zur Erzeugung synthetischer Datensätze für auf maschinellem Lernen basierende Sicherheitsüberwachungssysteme durchgeführt. Die generierten Datensätze werden anhand von Datenqualitätsmetriken analysiert. Zwei Szenarien aus der Prozessfertigung und aus industriellen Kommunikationsnetzen werden verwendet, um den vorgestellten Ansatz zu bewerten. Der vorgeschlagene Ansatz ist in der Lage, synthetische Datensätze für beide Szenarien zu erzeugen.

Die o.g. Abbildung verdeutlicht die Grundidee unseres Ansatzes. In einem ersten Schritt werden Prozess- und Netzwerkdaten während der Laufzeit eines Produktionssystems aufgezeichnet. In einem zweiten Schritt werden mit einem KI-Algorithmus Regressionsmodelle gelernt. In einem letzten Schritt werden synthetische Datensätze erzeugt. Diese Datensätze können dann von Algorithmen für maschinelles Lernen verwendet werden, um spezifische Modelle für die Erkennung von Cyberangriffen oder Fehlfunktionen zu trainieren. Eine ausführliche Beschreibung ist dem veröffentlichten Beitrag zu entnehmen.

### AP 3.2 Entwicklung einer Testumgebung:

Basierend auf der Festlegung aus AP1.1, die Konzepte *Netzwerkdaten aus IRMA*, sowie *Kommunikation via RabbitMQ* zu verwenden, wurde eine Testumgebung konzipiert (siehe nachfolgende Abbildung), sowie daraufhin aufgebaut.

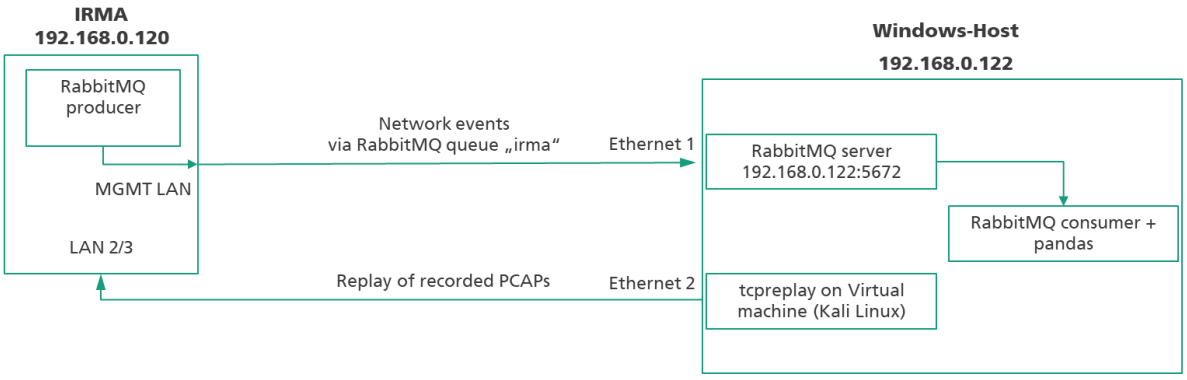



*Abbildung: Konzept der Testumgebung*

Der Aufbau besteht neben dem IRMA-System zunächst aus einem Windows-Rechner, welcher zuvor vorgefertigte Angriffe in das IRMA-System über die Schnittstelle *Ethernet 2 → LAN 2/3* einspielt und die generierten Netzwerkereignisse als RabbitMQ-Verkehr aus dem IRMA-System über die Schnittstelle *MGMT LAN → Ethernet 2* entgegennimmt. Die Netzwerkdaten werden dann in der KI-Software analysiert.

In einem zweiten Schritt wurden reale Automatisierungskomponenten bestehend aus einer SPS (Siemens S7), sowie einem IO-Gerät (Renesas TPS-1) in die Testumgebung integriert und ein tatsächlicher Angriff (*DoS ICMP*) auf diese Systeme durchgeführt. Über den *Mirrorport* wurde der Netzwerkverkehr an das IRMA-System weitergeleitet (siehe nachfolgende Abbildungen).

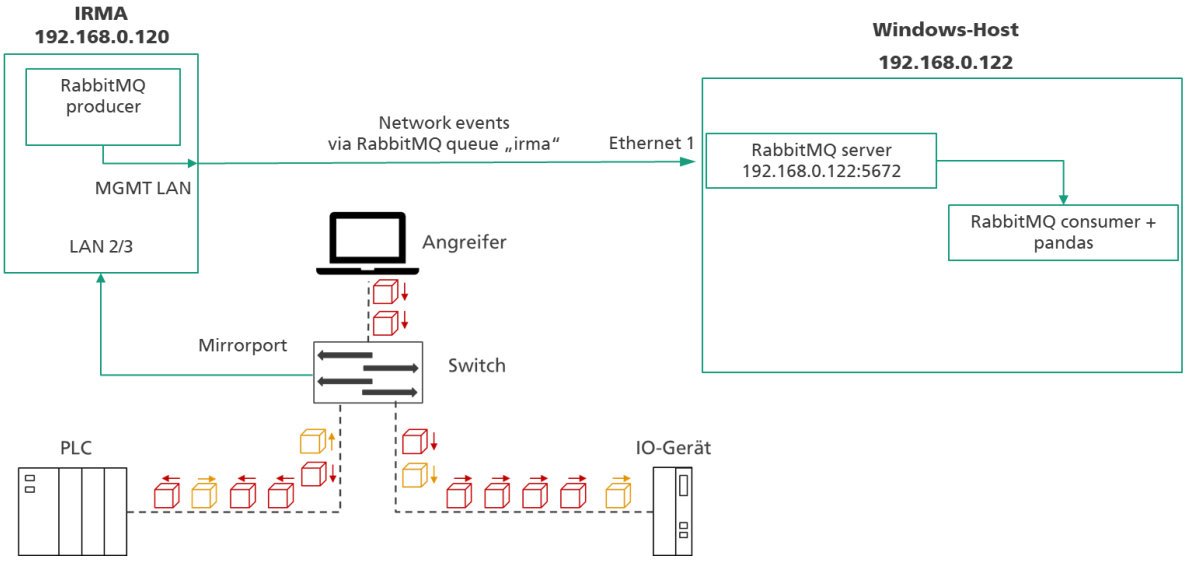



*Abbildung: Erweiterung der Testumgebung um Automatisierungskomponenten sowie Durchführung einer Cyberangriffs*

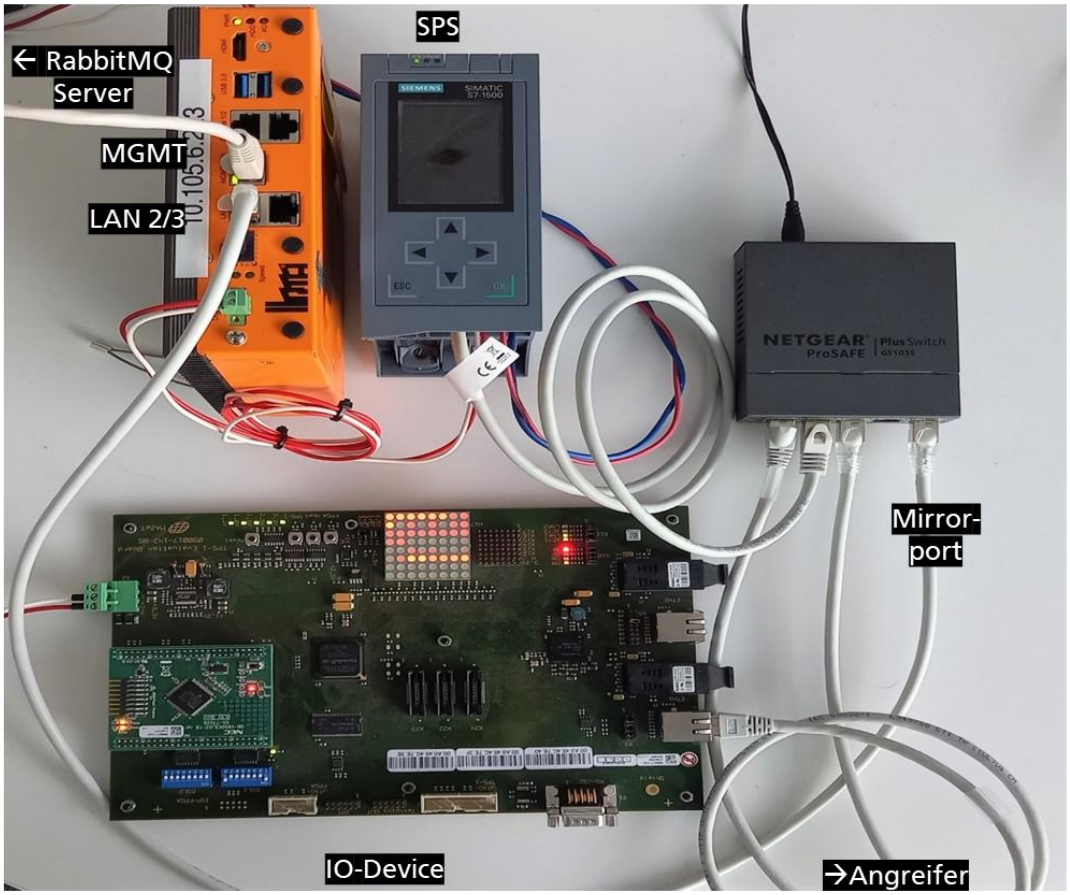
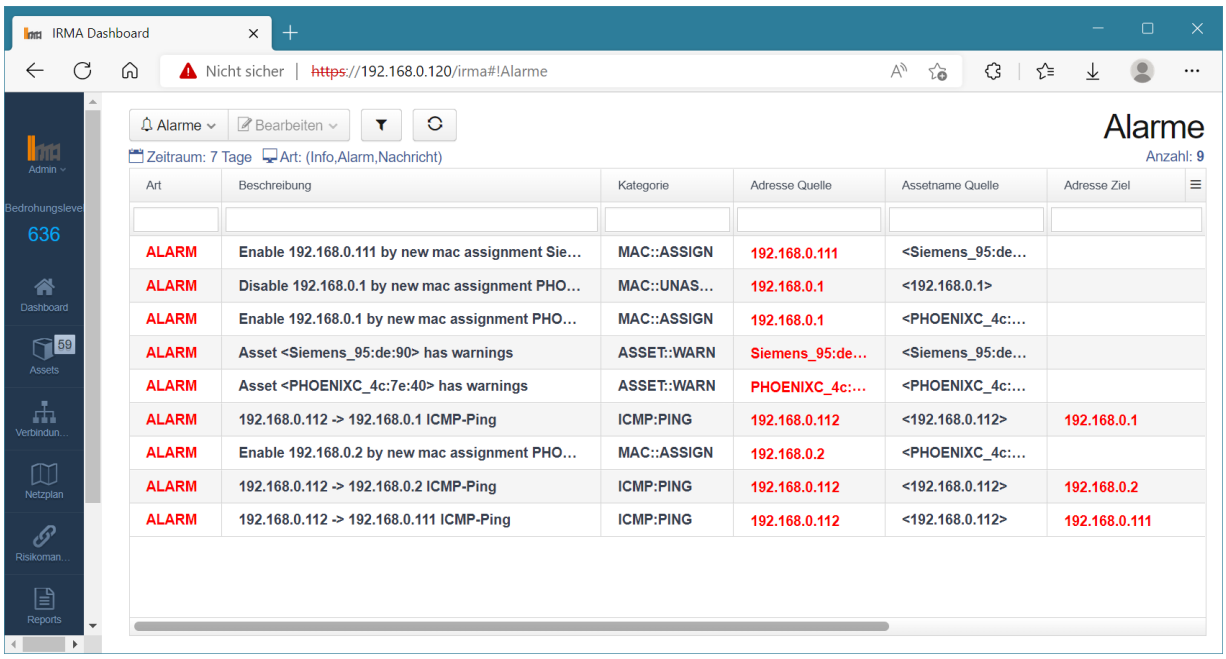



*Abbildung: Erster Prototyp der Testumgebung*

Die nachfolgende Abbildung zeigt, dass das IRMA-System in der Lage ist, diesen Angriff zu erkennen und als Netzwerkereignis an die KI weiterzuleiten.




| Art | Beschreibung | Kategorie | Adresse Quelle | Assetname Quelle | Adresse Ziel |
|---|---|---|---|---|---|
| ALARM | Enable 192.168.0.111 by new mac assignment Sie... | MAC::ASSIGN | 192.168.0.111 | <Siemens_95:de... | |
| ALARM | Disable 192.168.0.1 by new mac assignment PHO... | MAC::UNAS... | 192.168.0.1 | <192.168.0.1> | |
| ALARM | Enable 192.168.0.1 by new mac assignment PHO... | MAC::ASSIGN | 192.168.0.1 | <PHOENIXC_4c:... | |
| ALARM | Asset <Siemens_95:de:90> has warnings | ASSET::WARN | Siemens_95:de... | <Siemens_95:de... | |
| ALARM | Asset <PHOENIXC_4c:7e:40> has warnings | ASSET::WARN | PHOENIXC_4c:... | <PHOENIXC_4c:... | |
| ALARM | 192.168.0.112 -> 192.168.0.1 ICMP-Ping | ICMP:PING | 192.168.0.112 | <192.168.0.112> | 192.168.0.1 |
| ALARM | Enable 192.168.0.2 by new mac assignment PHO... | MAC::ASSIGN | 192.168.0.2 | <PHOENIXC_4c:... | |
| ALARM | 192.168.0.112 -> 192.168.0.2 ICMP-Ping | ICMP:PING | 192.168.0.112 | <192.168.0.112> | 192.168.0.2 |
| ALARM | 192.168.0.112 -> 192.168.0.111 ICMP-Ping | ICMP:PING | 192.168.0.112 | <192.168.0.112> | 192.168.0.111 |

*Abbildung: Vom IRMA-System erkannte Netzwerkereignisse*

Die generierten Netzwerkereignisse wurden mittels vorgefertigter Daten eines Sicherheitsprotokolls, welches von Kasselwasser bereitgestellt wurde, bezüglich des Zeitpunktes korreliert (siehe nachfolgende Abbildung).

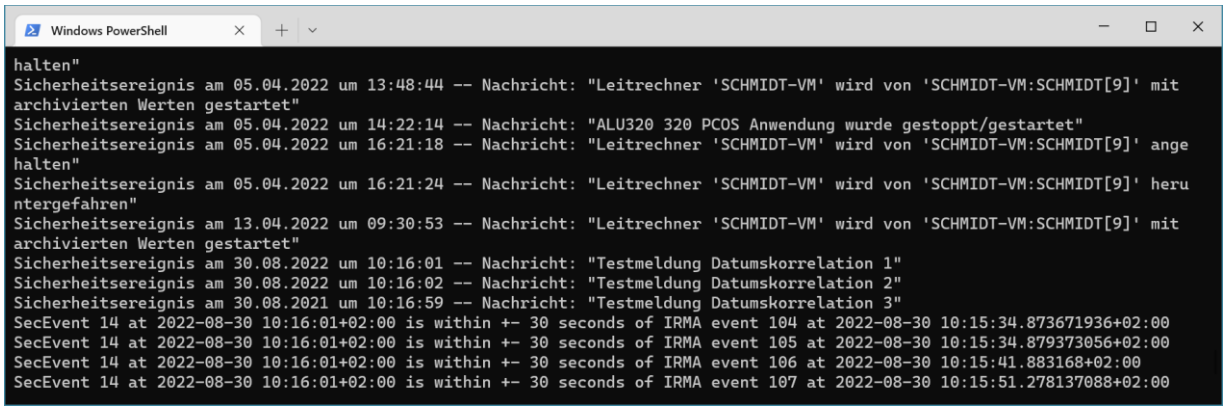
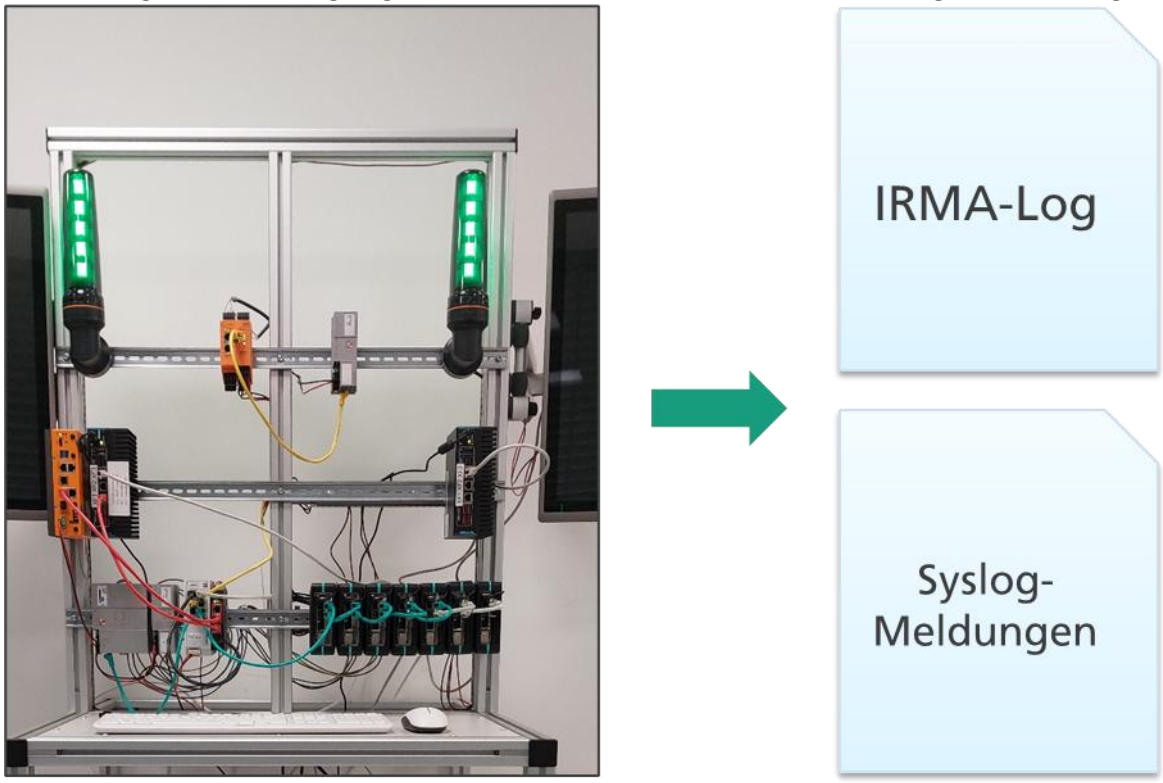



*Abbildung: Analyse der Netzwerkereignisse*

Die nachfolgende Abbildung zeigt den finalen Stand des AP 2.3 zur Entwicklung einer Testumgebung.




Die finale Testumgebung stellt einen Vorläufer des im Arbeitspaket 3.3 entwickelten Demonstrators dar und umfasst eine Reihe von Automatisierungskomponenten, die IRMA der Firma Achtwerk sowie einen Industrie-PC. Dieser Industrie-PC ist dafür verantwortlich, Log-Nachrichten zu aggregieren und mit KI-Ansätzen zu verarbeiten, um die Effektivität und Effizienz der entwickelten Lösungen zu demonstrieren.

### AP 3.3 Entwicklung eines Demonstrators:

Der finale Aufbau des Demonstrators ist nachfolgend dargestellt:

*Abbildung: Finaler SILK-Demonstrator in der SmartFactoryOWL*

Die Vorderseite beinhaltet die nachfolgenden Komponenten:

- 2x Speicherprogrammierbare Steuerung
- 7x Industrielle Feldgeräte
- 1x Industrieller Netzwerkswitch
- 1x Industrie-PC
- 2x HMI-PC

Des Weiteren sind auf der Rückseite folgende Komponenten montiert:

- 1x Industrie PC
- 1x Mini-PC
- 1x Industrieller Netzwerkswitch
- 2x LED-Signage-Leuchte

Auf der Hardware ist das nachfolgende Konzept realisiert:

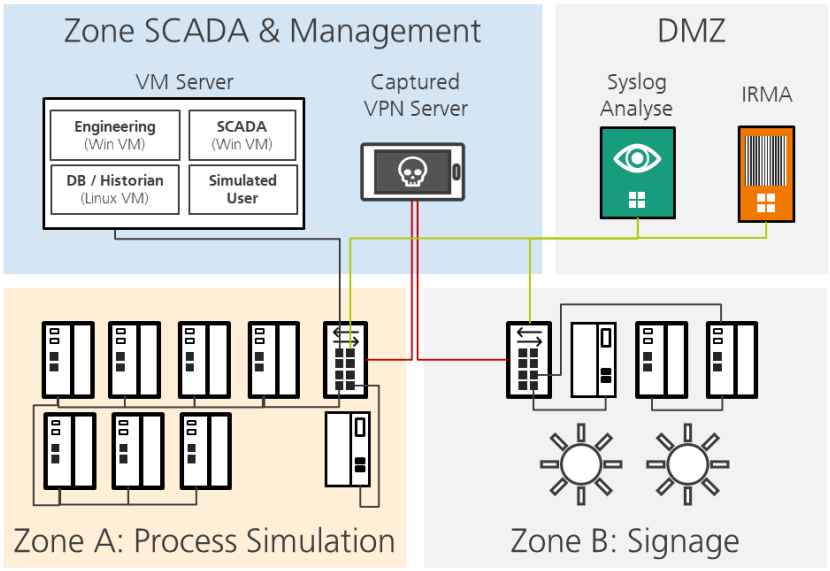



*Abbildung: Demonstrator-Konzept*

Das Demonstrator-Konzept besteht aus den folgenden Zonen:

- **Zone A Process Simulation:** Die erste Zone beinhaltet eine Prozess-Simulation (vgl. Abbildung links unten). Die Komponente beinhaltet einen physikalischen Aufbau und kombiniert Automatisierungskomponenten mit einer Simulation eines Prozesses. Der in der Literatur oftmals verwendete *Tennessee Eastman Prozess* ist eine Simulation einer chemischen Produktion. Im Arbeitspaket wird die Simulation um ein Netzwerk aus IO-Geräten, Netzwerkswitch und Steuerung (PLC) erweitert. Somit wird eine reale Produktion inkl. Netzwerkkommunikation abgebildet. Der physikalische Aufbau des Prototyps ist die Grundlage, um Cyberangriffe in einer kontrollierten Umgebung durchzuführen und die KI-basierte Sicherheitslösung zu evaluieren. Der simulierte Prozess wird mit einer Speicherprogrammierbare Steuerungen und sieben industriellen Feldgeräte abgebildet.
- **Zone B Signage:** Die zweite Zone beinhaltet einen Prozess der Signage. Dabei wird eine industrielle Steuerung verwendet, um Signage-Lampen anzusteuern und so Feedback an das Personal zu geben, z.B. signalisiert eine grüne Lampe einen Gutzustand, während eine rote Lampe einen Ausfall einer Maschine anzeigt. Der Prozess ist mit einer Speicherprogrammierbare Steuerung und zwei LED-Signageleuchten physikalisch abgebildet.
- **Zone SCADA & Management:** Die SCADA & Management Zone beinhaltet typische Komponenten, die in einer Leitstelle und zum Management von Maschinen verwendet werden. Ein SCADA-System ist eine Softwarelösung, die es ermöglicht, industrielle Prozesse zu überwachen, zu visualisieren und zu steuern. Eine Engineering Station wird verwendet, um Speicherprogrammierbare Steuerungen zu programmieren. Ein Historian ist eine Datenbank, die kontinuierlich Prozessdaten erfasst und speichert, um später analysiert und ausgewertet zu werden. Des Weiteren wird ein menschlicher Benutzer simuliert, der mit der Prozess-Simulation interagiert. Ein Cyber-Angriffs-Framework wird verwendet zur Durchführung von Störungen in der Prozess-Simulation (1). Störungen sind z. B. Cyberangriffe basiert auf den Vorarbeiten des AP 1.3 und implementiert eine Auswahl von MITRE Angriffstechniken. Somit können auch komplexere Cyberangriffe, bestehend aus mehreren Angriffstechniken, in der Testumgebung ausgeführt werden.
- **Zone DMZ:** Diese Zone beinhaltet die Komponenten des Monitoring Frameworks, die IRMA von Achtwerk und die vom Fraunhofer IOSB-INA realisierte Software zur Überwachung. Hier werden KI-basierte Modelle auf Basis von Netzwerk- und Logdaten umgesetzt. Mit Hilfe der

Modelle sollen einerseits Netzwerkdaten der Prozess-Simulation (1) analysiert und ausgewertet werden. Andererseits sollen zu (1) passende Ereignisdaten von Sicherheitsereignissen ausgewertet werden. Die Ergebnisse der Modelle werden anschließend korreliert, um eine passende Handlungsempfehlung für das Anlagenpersonal zu generieren. Das Zusammenspiel zwischen IRMA und Fraunhofer-Software ist in der nachfolgenden Abbildung dargestellt:

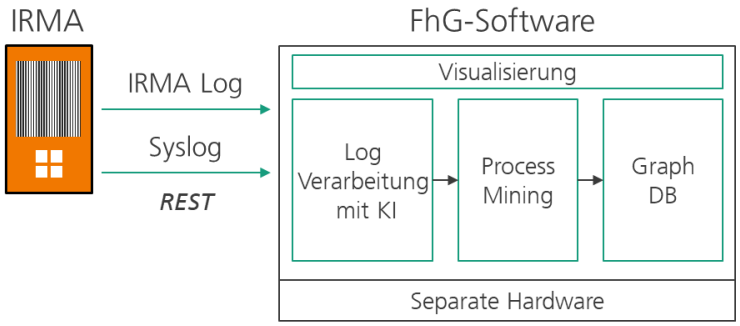



*Abbildung: Architektur und Schnittstelle zwischen IRMA und Fraunhofer-Software*

Der Aufbau des Demonstrators ist abgeschlossen und dieser steht für die Testdurchführung in der SmartFactoryOWL bereit (s. AP 3.5). Der Demonstrator dient im Projekt zur Erprobung der technischen Machbarkeit und zur Überprüfung der parallel entwickelten Software.

### AP 3.4 Automatische Konfigurationsanpassung des Netzwerkanalysesystems:

Erweiterung des Netzwerkanalysesystems um automatisierte Konfigurationsanpassung auf Basis der KI-Entscheidungen und generierten Handlungsempfehlungen. Ein wissenschaftlicher Beitrag, der teilweise auf den Arbeiten aus SILK basiert, und sich mit dem Thema „Automatische Konfigurationsanpassung des Netzwerkanalysesystems“ befasst, wurde 2022 auf einer internationalen Konferenz vorgestellt und ist unter der folgenden Referenz zu finden:

*F. Specht, J. Otto and J. Eickmeyer, "Cyberattack Impact Reduction using Software-Defined Networking for Cyber-Physical Production Systems," 2022 IEEE 20th International Conference on Industrial Informatics (INDIN), Perth, Australia, 2022, pp. 188-194*

Lösungen für automatisierte Maßnahmen bei Störvorfällen auf Netzwerkebene, z.B. bei Cyberangriffen, werden in der Literatur als Intrusion Prevention System klassifiziert. Ein solches System erkennt zunächst einen Cyberangriff und führt anschließend eine vordefinierte technische Maßnahme aus, um den Angriff zu unterbinden. Der entwickelte Ansatz für automatisierte Verteidigungsmaßnahmen ist in der nachfolgenden Abbildung dargestellt. Software-Defined Networking (SDN) wird dabei als Technik eingesetzt. SDN ermöglicht die Kontrolle der Netzwerkkommunikation an zentralen Kommunikationspunkten (Netzwerk-Switch). Der Ansatz besteht aus drei Teilen: (1) einer Angriffsplattform, (2) einem auf SDN basierenden Präventionssystem und (3) Metriken zur Messung der Auswirkungen von Cyberangriffen und Präventionstechniken.

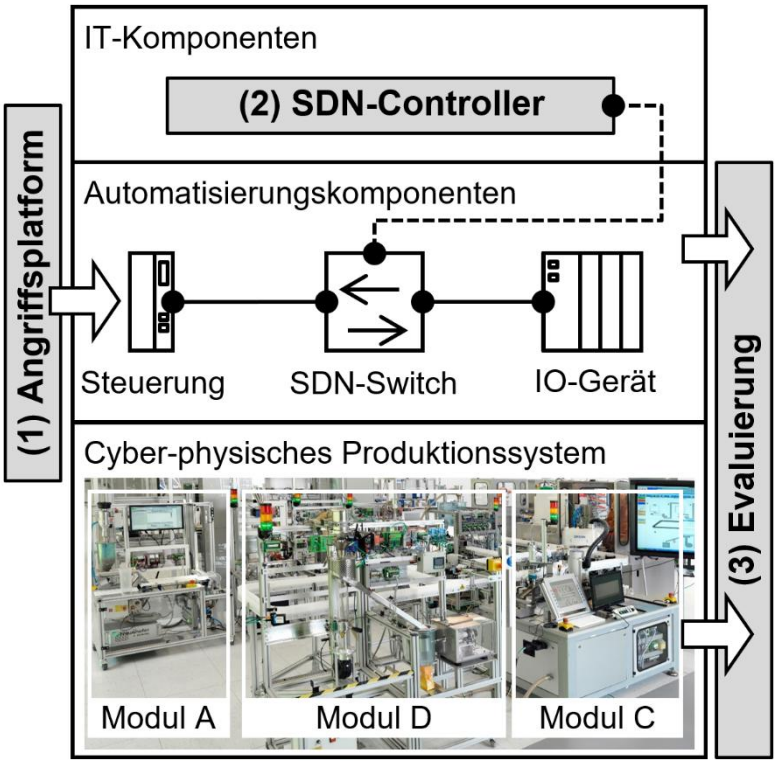



*Abbildung: Ansatz zur automatisierten Verteidigung gegen Cyberangriffe mit Software-Defined-Networking.*

Es wurden erste wissenschaftliche Arbeiten zur Bewertung des Ansatzes durchgeführt. Dabei wurden vier verschiedene Präventionstechniken als Gegenmaßnahme für einen Cyberangriff umgesetzt. Zur Bewertung des Ansatzes wurde ein Real-Szenarien aus der diskreten Fertigung verwendet. Fünf ausgewählte Angriffstechniken (vgl. AP 1.3) wurden in dem Real-Szenario ausgeführt und automatisiert durch die Präventionstechniken abgewehrt. Erste Ergebnisse zeigen, dass der Ansatz die durchschnittlichen Auswirkungen der ausgewählten Cyberangriffe um bis zu 98,1% reduziert. Eine ausführliche Beschreibung ist dem veröffentlichten Beitrag zu entnehmen.

### AP 3.5 Testdurchführung in der SmartFactoryOWL:

Im Rahmen von AP 3.5 wurde ein strukturierter Testplan erstellt, der die in AP 1.3 definierten Sicherheitsszenarien berücksichtigt. Die Testergebnisse wurden iterativ ausgewertet, in die laufende Entwicklung und Forschung rückgespiegelt und der Lösungsansatz schrittweise an die Anforderungen eines realen Industrieeinsatzes angepasst. Abschließend erfolgte eine Prüfung und Bewertung der Resultate durch unsere KRITIS-Partner auf Vollständigkeit und Relevanz.

Die Testdurchführung in der SmartFactoryOWL, basierend auf den Vorarbeiten aus AP 2 und AP 3, lässt sich wie folgt zusammenfassen:

1. **Erstellung des Testdatensatzes und Szenarioaufbau:** Am IOSB-INA Demonstrator wurde zunächst ein umfassender Testdatensatz erzeugt. Dieser umfasst zwei Hauptszenarien:
    - **Normale Produktion (unauffälliger Betrieb):** Hierbei simulieren wir den regulären Produktionsablauf ohne sicherheitsrelevante Vorfälle.
    - **Mehrstufiger Cyberangriff:** In Anlehnung an das im Meilensteintreffen (September 2023) definierte Angriffsszenario wird ein komplexer Angriffsverlauf simuliert.

Für beide Szenarien werden einerseits **IRMA-Logs** (Protokolle des IRMA-Geräts, die Betriebszustände und potenzielle Sicherheitsereignisse dokumentieren) und andererseits **Syslog-Meldungen** (z. B. aus der Industriesteuerung, die Systemaktivitäten abbilden) erzeugt. Insgesamt enthält der Datensatz 44 verschiedene Typen von Lognachrichten, die für die nachfolgende Analyse und Anomalieerkennung zur Verfügung stehen.

2. **Transformation der Lognachrichten für Process Mining:** Um die umfangreichen und heterogenen Logdaten für Process-Mining-Algorithmen nutzbar zu machen, wurden die Roh-Logs zunächst in eine symbolische, standardisierte Form überführt. Dabei steht jeder Symbolwert stellvertretend für eine definierte Teilmenge der ursprünglichen Logdaten (z. B. „Host X meldet ungewöhnlich hohen Netzwerktraffic“). Diese Verdichtung in abstrakte Symbole ist die Voraussetzung dafür, dass die Process-Mining-Algorithmen Muster in den Abläufen zuverlässig erkennen können. Die folgende Abbildung zeigt beispielhaft, wie eine Lognachricht in ein solches Symbol übersetzt wird.
3. **Erlernen eines Verhaltensmodells per Process Mining:** Anschließend wurden die symbolisierten Security-Logs sequenziell analysiert, um das „normale“ Verhalten der Umgebung – beispielsweise einer industriellen Steuerung – zu modellieren. Die Process-Mining-Algorithmen identifizieren typische Ereignisfolgen und Prozesse, die in der unauffälligen Produktion ablaufen. Aus diesem Referenzmodell lassen sich Abweichungen oder Anomalien (beispielsweise aus dem mehrstufigen Cyberangriff) automatisch erkennen, weil sie nicht zur beobachteten Normalverhaltens-Sequenz passen. So können potenzielle Sicherheitsvorfälle frühzeitig detektiert und visuell dargestellt werden.
4. **Verknüpfung von Verhaltensmodell und Netzwerktopologie:** Für eine ganzheitliche Sicherheitsanalyse wurde das so entstandene Verhaltenmodell mit der physischen und logischen Netzwerktopologie der SmartFactoryOWL verknüpft. Dafür werden beide Strukturen – das Prozessmodell (Zustände und Ereignisse) und der Netzwerk-Graph (Geräte und Verbindungen) – in einer Graph-Datenbank abgebildet. Durch gezielte Abfragen lassen sich dann Beziehungen zwischen einem bestimmten Ablauf im Prozessmodell und den involvierten Netzwerkkomponenten herstellen. Auf diese Weise wird ersichtlich, **welche Gerätedatenpfade** zu einer Abweichung geführt haben oder **welche Geräte – beispielsweise eine SPS – im Netzwerk betroffen sind**, wenn das Prozessmodell eine Anomalie meldet. Die nachfolgende Abbildung illustriert beispielhaft diese Verknüpfung aus Verhaltensmodell und Netzwerktopologie.

Durch diese vier Schritte wurde gewährleistet, dass die Sicherheitsszenarien aus AP 1.3 vollständig getestet, die Ergebnisse in das Monitoring- und Erkennungssystem übernommen und das Gesamtkonzept auf seine Praxistauglichkeit geprüft wurde. Die iterative Anpassung und die abschließende Evaluierung durch KRITIS-Partner sicherten dabei eine hohe Aussagekraft und Relevanz für den realen Industrieeinsatz.

## Arbeitspaket 4: Evaluierung der Sicherheitslösung

Im AP 4 verfolgte das Fraunhofer IOSB-INA das Ziel der Evaluation des entwickelten Lösungsansatzes sowohl in der Testumgebung, als auch in der realen KRITIS-Umgebung hinsichtlich spezifizierter Evaluationskriterien. Im Einzelnen beinhalteten diese Arbeiten mehrere Teile:

### AP 4.1 Definition von Evaluierungskriterien:

Die Spezifizierung von Evaluationskriterien und Definition eines Testplans basierte auf Vorarbeiten des AP 3. Die Evaluationskriterien umfassen die nachfolgenden Punkte:

1. **Funktionale Korrektheit und Erkennungsrate**

- True Positive Rate ≥ 95 %
  Mindestens 95 % aller tatsächlich relevanten Sicherheitsvorfälle müssen korrekt erkannt werden (Vergleich mit Ground-Truth-Daten).
- False Positive Rate ≤ 5 %
  Fehlalarme dürfen 5 % der generierten Alarme nicht überschreiten, um Operator-Überlastung zu verhindern.
- False Negative Rate ≤ 5 %
  Maximal 5 % echter Vorfälle dürfen unentdeckt bleiben.

- Time-to-Detect ≤ 60 Sekunden
  Die Zeitspanne vom Eintreten eines sicherheitsrelevanten Ereignisses bis zum Alarm darf zwei Sekunden nicht überschreiten.

**2. Performance und Skalierbarkeit**

- Verarbeitungsdurchsatz ≥ 10 000 Ereignisse/s
  Log- und Netzwerkereignisse müssen in Echtzeit verarbeitet werden, ohne Meldestau.
- Abfragelatenz ≤ 2 Sekunden
  Standard-Graphabfragen (z. B. „Gerätepfade der letzten 5 Minuten“) müssen in maximal zwei Sekunden Resultate liefern.
- Ressourcen-Kennzahlen (RAM/CPU)
  Unter Lasttests wurde dokumentiert, welche Systemkapazitäten bei Verdoppelung der Gerätzahl erforderlich sind, um horizontale Skalierung zu ermöglichen.

**3. Robustheit und Zuverlässigkeit**

- Verfügbarkeit ≥ 99 % (24×7-Betrieb)
  Geplante Ausfallzeiten (z. B. Wartung) pro Quartal ≤ 22 Stunden.
- Fehlertoleranz und Fallback-Modus
  Bei Ausfall einer Log-Quelle oder Datenbank soll das System in einen Pufferungsmodus wechseln, um Datenverlust zu vermeiden.
- Recovery-Zeit ≤ 5 Minuten
  Nach einem ungeplanten Crash müssen alle Dienste (Graph-DB, KI-Module, Web-Frontend) binnen fünf Minuten wiederhergestellt sein.

**4. Genauigkeit der Erklärungen (Explainability)**

- Vollständigkeit der Erklärungstexte ≥ 90 %
  Automatisch generierte Texte müssen in mindestens 90 % aller Fälle folgende Elemente enthalten:
  1. Betroffene Geräte
  2. Zeitstempel des Ereignisses
  3. Art des Vorfalls
  4. Konkrete Handlungsempfehlung
- Kohärenz und Verständlichkeit (Lesbarkeitsindex ≥ 60)
  Die Texte müssen ohne zusätzliche Dokumentation für Operateure sofort verständlich sein.
- Risikopriorisierung (Übereinstimmung ≥ 95 %)
  Algorithmisch errechnete Risikostufen („hoch“, „niedrig“ etc.) müssen in ≥ 95 % der Fälle mit einer Expertenbewertung übereinstimmen.

Diese Kriterien bilden die Grundlage für die anschließende Evaluierung. So wird sichergestellt, dass das Monitoring-, Korrelation- und Explainability-System technisch, operativ und aus Anwendersicht den Anforderungen an KRITIS-Umgebungen entspricht.

### AP 4.2 Integration in den KRITIS-Betrieben:

Das IOSB-INA war nicht aktiv am Teilarbeitspaket 4.2 beteiligt, stand den Partnern jedoch bei Fragen und Problemen mit Hilfestellung zur Verfügung.

### AP 4.3 Evaluierung des Lösungsansatzes in der Testumgebung:

Im Rahmen von AP 4.3 wurde der entwickelte Lösungsansatz in der SmartFactoryOWL iterativ geprüft und weiterentwickelt. Dazu wurde zunächst der entwickelte Demonstrator in die SmartFactory-Umgebung integriert, inkl.: industrielle Steuerungen, Switches, IRMA-Gerät, Graph-Datenbank, KI-Module und Web-Frontend liefen gemeinsam im Echtzeitbetrieb. Anschließend wurden repräsentative Szenarien aus AP 1.3 (“normale Produktion” vs. “mehrstufiger Cyberangriff”) durchgespielt und die Systemreaktionen anhand der in AP 4.1 definierten Kriterien erfasst.

Während der Tests wurde kontinuierlich überwacht, ob die Erkennungsrate von Anomalien ≥ 95 % erreicht wurde, ob Fehlalarme < 5 % ausmachten und ob Alarme binnen ≤ 60 s gemeldet wurden. Parallel dazu liefen Lasttests mit bis zu 10 000 Log-Events pro Sekunde, um Performance, Latenzzeiten (Graph-Abfragen unter 2 s) und Ressourcenauslastung zu messen. Bei simulierten Ausfällen (z. B. Datenbankunterbrechung) stellte sich heraus, dass das System in unter 5 Min. wiederhergestellt war und Puffermechanismen zuverlässig Daten sammelten, bis die Verbindung stand – die angestrebten Kennzahlen wurden damit weitgehend erfüllt.

Schließlich wurde die Bedienbarkeit im Web-Frontend anhand kurzer Practitioner-Tests überprüft: Bediener fanden nach maximal zehnminütiger Einweisung alle relevanten Informationen (Alarm, Erklärung, Handlungsempfehlung) in unter 5 Min. Die automatisch erstellten Texte erreichten einen Lesbarkeitsindex ≥ 60 und stimmten in 95 % der Fälle mit Expertenbewertungen überein. Die Testergebnisse wurden dokumentiert und direkt an die Entwicklung zurückgespielt. Kleinere Anpassungen (etwa Feintuning der Fallback-Logik und Optimierung der Erklärungstexte) wurden umgesetzt, um die Praxistauglichkeit weiter zu erhöhen.

## 3. Wichtigsten Positionen des zahlenmäßigen Nachweises

Im Teilvorhaben wurde die Grundausstattung für einen Mitarbeiter gefördert, sodass diese mobile und für die Arbeiten mit maschinellen Lernverfahren geeignete Arbeitsmittel zur Verfügung hatten. Die im Projekt geförderten Investitionsmittel dienten dem Aufbau eines physikalischen Prototyps (Demonstrators), welcher die Projektergebnisse praxisnah veranschaulicht (vgl. Arbeitspaket 3).

Die geförderten Personalmittel ermöglichten die Projektbearbeitung durch einen wissenschaftlichen Mitarbeiter aus der Arbeitsgruppe „Cybersicherheit in der Produktion“. Die wissenschaftlichen Mitarbeiter Herr Felix Specht und Herr Nils Koch waren anteilig auf einer TVL-E13 Stelle beschäftigt. Zu den Aufgaben der wissenschaftlichen Mitarbeiter zählten dabei die Kooperation und Kommunikation mit den Projektpartnern, Analyse und Erhebung von Anforderungen, Konzeptionierung von Algorithmen und Softwaremodulen, Datenerfassung und -aufbereitung, Implementierung der Softwarelösung, Publikation von wissenschaftlichen Ergebnissen sowie die Planung, Konzeption und Umsetzung des Prototyps als Demonstrator in der SmartFactoryOWL. Des Weiteren war Herr Dr.-Ing. Jens Otto anteilig auf einer TVL-E14 Stelle im Projekt tätig für die übergeordnete Projektkoordination, Mitarbeiterplanung und das Projektmanagement.

## 4. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die angestrebten Ziele des Teilvorhabens wurden vollständig erreicht und die Ergebnisse auf entsprechenden internationalen Konferenzen veröffentlicht. Aus Geheimhaltungsgründen sind Versuche mit Daten der Partner nicht zur Veröffentlichung verfügbar. Während der Projektlaufzeit kam es zu Abweichungen ggü. der ursprünglichen Projektplanung. In Zusammenarbeit konnten die Verbundpartner und der Projektpartner die Ziele und Arbeiten erfolgreich anpassen. Zur sachgemäßen Durchführung der Arbeiten wurde eine kostenneutrale Verlängerung um fünf Monate beantragt. Diese war zum einen durch die Anpassung der Projektziele notwendig, was eine angepasste Zeitplanung erforderte. Zum anderen führten pandemiebedingte Arbeitsweisen (Covid-19) zu Verzögerungen bei der Erprobung der Verfahren für die Praxis, die bei der Beantragung des Projekts nicht vorhersehbar waren. Der Zeitplan wurde innerhalb dieser Rahmenbedingungen eingehalten.

Die beantragten Mittel für Personal wurden für die Durchführung der Forschungs- und Transferaktivitäten benötigt und sachgerecht verwendet. Die beantragten Personalmittel waren dabei angemessen, um die betrachteten Herausforderungen zu bewältigen, da es sich um ein aktuelles und komplexes Thema handelt, dessen Bedeutung aufgrund der Entwicklungen im Bereich der Nutzung von KI zur Generierung heterogener Angriffe voraussichtlich zunehmen wird. Ein klarer Fokus lag dabei

auf der Forschung und Umsetzung der Erkenntnisse in konkreten industrielle und KRITIS-nahe Szenarien, wobei die angesetzten Personalkosten aufgrund der Vielfältigkeit der betrachteten Aufgaben notwendig waren.

Die Förderung des Projekts durch das BMBF war aufgrund der Ausrichtung der Forschungstätigkeit und des Konsortiums notwendig. Aufgrund der klaren Orientierung an der Umsetzbarkeit und der Größe des Projektkonsortiums war eine Förderung durch grundlagenorientierte Forschungsförderungen wie die DFG nicht möglich. Bedingt durch die Konsortialstruktur mit regionalen Partnern war eine europäische Förderung auszuschließen. Zudem konnten die Maßnahmen in dieser Form nicht direkt durch industrielle Partner übernommen werden, da die Maßnahmen noch keine Marktreife besitzen und somit das Risiko nicht von den Partnern getragen werden konnte.

# 5. Nutzen und Verwertbarkeit

## 5.1 Verwertbarkeit der Ergebnisse

Im Laufe des Projekts wurde eine Software als Demonstrationslösung erstellt, die Verhaltensmodelle auf Basis von Log-Nachrichten von Automatisierungskomponenten lernt und diese nutzt, um Cyberangriffe zu erkennen und dem Benutzer Hinweise bereitzustellen. Die Ergebnisse des Projektes SILK gehen über die Fähigkeiten bestehender Lösungen (z. B. Security Information and Event Management-Systeme) hinaus, welche einen hohen manuellen Aufwand erfordern und somit insbesondere kleinen und mittleren Unternehmen die Einführung entsprechender Maßnahmen erschwert.

Alle Projektergebnisse und der entwickelte Software-Demonstrator wurden an die Firma Achtwerk übergeben. Achtwerk plant, die erarbeiteten Komponenten langfristig in ihre IRMA®-Lösung zu integrieren, um dort eine effiziente, leicht bedienbare und kostengünstige Cyberabwehr für industrielle Anwender bereitzustellen.

Das Fraunhofer IOSB-INA beabsichtigt, die Projektergebnisse sowohl kommerziell als auch wissenschaftlich zu nutzen. Der entwickelte Demonstrator verbleibt nach Projektabschluss fest in der SmartFactoryOWL und wird regelmäßig auf Informations- und Weiterbildungsveranstaltungen potenziellen Industrie- und Wirtschaftspartnern präsentiert. Auf diese Weise dient er auch als Instrument zur Akquise von Industriekooperationen. Geplante, weitergehende wirtschaftliche Verwertungen in Zusammenarbeit mit weiteren Industriepartnern befinden sich zum Zeitpunkt dieser Berichterstattung zwar noch in der Konzeptphase, sind jedoch fest vorgesehen. Darüber hinaus fließen die Projektergebnisse in das Weiterbildungsangebot des Fraunhofer IOSB-INA ein und werden Schulungsteilnehmern aus Industrie und Wirtschaft in Form von Unterrichtseinheiten bereitgestellt.

Die wissenschaftliche Nutzung erfolgt über neue Forschungs- und Entwicklungsprojekte, in denen insbesondere der Demonstrator als Basis für weiterführende Untersuchungen dient. Ein Schwerpunkt ist dabei die Verknüpfung mit Sprachmodellen (künstliche Intelligenz), um Mitarbeiter in Industrieunternehmen künftig noch besser zu unterstützen. Zu diesem Zweck wurde bereits ein entsprechender Förderantrag eingereicht.

Da Mitarbeiter des Fraunhofer IOSB-INA parallel als Lehrende an der Technischen Hochschule Ostwestfalen-Lippe tätig sind, wurden die Projektergebnisse auch in die Lehrveranstaltungen „Advanced Topics in Algorithms“ und „Network Security“ integriert und tragen so zum Kompetenzaufbau und zur Wissensvermittlung bei den Studierenden bei.

## 5.2 Zusammenarbeit mit anderen Stellen

Am Projekt beteiligt waren auf wissenschaftlicher Seite die Helmut-Schmidt-Universität / Universität

der Bundeswehr Hamburg mit der Professur für Informatik im Maschinenbau sowie das Fraunhofer IOSB-INA in Lemgo mit den Arbeitsgruppen für „Cybersicherheit in der Produktion“. Auf Seiten der Industriepartner waren die Unternehmen Achtwerk GmbH &Co. KG, KASSELWASSER – Eigenbetrieb der Stadt Kassel und SachenEnergie AG (durch die Tochtergesellschaft DREWAG – Stadtwerke Dresden GmbH) am Vorhaben beteiligt. Die Projektkoordination wurde durch den Partner Fraunhofer IOSB-INA übernommen.

Im Rahmen des Projekts fand ein regelmäßiger Austausch unter den Projektpartnern im Rahmen eines monatlichen Jour Fixe statt. Im Projekt erarbeitete Ergebnisse konnten auf internationalen Tagungen und in Journalen publiziert und somit Dritten kommuniziert werden. Durch die Umsetzung eines Demonstrators in der SmartFactoryOWL stehen Projektergebnisse für zukünftige Forschungs- und Entwicklungsprojekte zur Verfügung. Zudem werden die Ergebnisse durch regelmäßig stattfindende Informationsveranstaltungen Vertretern aus Industrie und Wirtschaft präsentiert.

## 6. Fortschritt auf Gebiet des Vorhabens durch andere Stellen

Während der Projektlaufzeit wurde durch die Beteiligten regelmäßig die aktuelle Literatur recherchiert und hinsichtlich ihrer Relevanz für das Projekt bewertet. Im Teilvorhaben des Fraunhofer IOSB-INA lag der Fokus auf dabei insbesondere bei der Mustererkennung in Logdaten von Automatisierungskomponenten und industriellen Systemen.

Im Rahmen des Projekts wurden signifikante Fortschritte in der automatisierten Analyse von Logdaten erzielt, insbesondere bei der Identifikation von Mustern in Syslog-Meldungen. Die Ergebnisse des Teilprojekts differenzieren sich durch die Integration spezialisierter Algorithmen, die auf maschinellem Lernen basieren, um Anomalien in der Netzwerkkommunikation zu erkennen. Während herkömmliche Systeme oft auf statischen Regeln beruhen, nutzt unser Ansatz dynamische Modelle, die aus realen Logdaten trainiert werden.

Durch die Anwendung fortschrittlicher Techniken, wie der Extraktion von Log-Templates und der Modellierung von Normalverhalten, konnten präzise Vorhersagen über potenzielle Sicherheitsvorfälle getroffen werden. Diese Methodik ermöglicht eine drastische Reduktion von manuellen Aufwänden und verbessert somit die Effizienz der Sicherheitsanalyse.

Die rasante Entwicklung von künstlicher Intelligenz in den letzten Jahren, insbesondere im Bereich der großen Sprachmodelle, wirkt sich auch auf das Gebiet der Cybersicherheit aus und wurde bereits in einigen Anwendungen der klassischen IT-Sicherheit erfolgreich eingesetzt. Das Fraunhofer IOSB-INA sieht großes Potenzial für den Einsatz von Sprachmodellen, insbesondere beim Einbinden menschlicher Benutzer in die Verwendung von Cybersicherheitssoftware. Bereits während der Projektlaufzeit wurde der Einsatz von Sprachmodellen experimentell erprobt, um die Ergebnisse der Log-Analyse auszuwerten und für Nicht-Experten verständlich aufzubereiten. Dadurch könnten auch Firmen ohne Mitarbeitende mit Cybersicherheitsexpertise eine solche Sicherheitslösung effizient nutzen. Die Verwendung und Kombination von Sprachmodellen mit Sicherheitslösungen, wie sie im Projekt entwickelt wurden, sollen zukünftig in weiteren Förderprojekten erforscht und in die Praxis überführt werden.

## 7. Veröffentlichungen

Die Ergebnisse des Teilvorhabens *Erforschung angewandter KI-Ansätze zur automatisierten Analyse der Netzwerkkommunikation und Erkennung von Fehlverhalten* wurden unter Angabe des Förderkennzeichens in folgenden wissenschaftlichen Artikeln veröffentlicht:

- Specht, Felix; Otto, Jens; Ratz, Daniel: Generation of Synthetic Data to Improve Security Monitoring for Cyber-Physical Production Systems. In: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN), Lemgo, Germany, July 2023
- Otto, Jens; Specht, Felix: Efficient Cyberattack Detection in Logs of Cyber-Physical Production Systems. In: 2025 IEEE 7th International Conference on System-Integrated Intelligence (SysInt), Bremen, Germany, June 2025

**Literaturverzeichnis**

[1] IT-Sicherheit – Branchenstandard Wasser/Abwasser (B3S), DVGW-Merkblatt W 1060, DVGW Deutscher Verein des Gas- und Wasserfaches e.V. - Technisch-wissenschaftlicher Verein

[2] E. Hodo, X. J. A. Bellekens, A. Hamilton, C. Tachtatzis und R. C. Atkinson, „Shallow and Deep Networks Intrusion Detection System: A Taxonomy and Survey,“ CoRR, abs/1701.02145, 2017.

[3] M. Almseidin, M. Alzubi, S. Kovacs und M. Alkasassbeh, „Evaluation of Machine Learning Algorithms for Intrusion Detection System,“ CoRR, abs/1801.02330, 2018.

[4] C. Größer et al., "Data Mining-driven Manufacturing Process Optimization," in The World Congress on Engineering (WCE), London, 2012.

[5] C. J. Romanowski, R. Nagi, "Data mining for design and manufacturing," in Analyzing maintenance data using data mining methods, Kluwer Academic Publishers, 2002.

[6] M. Lange-Hegermann, "Algorithmic Linearly Constrained Gaussian Processes", Advances in Neural Information Processing Systems 31 (NeurIPS), 2018.

[7] S. Thewes, M. Krause, C. Reuber, M. Lange-Hegermann, R. Dziadek, M. Rebbert, “Efficient In-Vehicle Calibration by the Usage of Automation and Enhanced Online DoE Approaches,” Simulation and Testing for Vehicle Technology, pp. 299-316 (2016).

[8] A. Vodenčarević, H. Kleine Büning, O. Niggemann and A. Maier, "Using Behavior Models for Anomaly Detection in Hybrid Systems," in XXIII International Symposium on Information, Communication and Automation Technologies (ICAT), 2011.

[9] O. Niggemann, B. Stein, A. Vodencarevic, A. Maier and H. Kleine Büning, "Learning Behavior Models for Hybrid Timed Systems," Conference on Artificial Intelligence, 2012.

[10] A. Maier, A. Vodenčarević, O. Niggemann, R. Just and M. Jäger, "Anomaly Detection in Production Plants using Timed Automata," 8th International Conference on Informatics in Control, Automation and Robotics (ICINCO), 2011.

[11] Hranisavljevic, N.; Niggemann, O.; Maier, A.: A Novel Anomaly Detection Algorithm for Hybrid Production Systems based on Deep Learning and Timed Automata, 2016.

[12] Eiteneuer, B.; Hranisavljevic, N.; Niggemann, O.: Dimensionality Reduction and Anomaly Detection for CPPS Data using Autoencoder. In: 20th IEEE International Conference on Industrial Technology Melbourne, Australien, Feb 2019.

[13] Eiteneuer, Benedikt; Niggemann, Oliver: LSTM for model-based Anomaly Detection in Cyber-Physical Systems. In: Proceedings of the 29th International Workshop on Principles of Diagnosis Warsaw, Poland, Aug 2018.

[14] Maier, A.; Köster, M.; Paiz Gatica, C.; Niggemann, O.: Automated Generation of Timing Models in Distributed Production Plants. In: IEEE International Conference on Industrial Technology (ICIT 2013), Cape Town, South Africa, Feb 2013.

[15] Windmann, S.; Lang, D.; Niggemann, O.: Learning Parallel Automata of PLCs. In: 22nd IEEE International Conference on Emerging Technologies And Factory Automation, September 2017.

[16] N. Thi, V.Cao und N.-A. Le-Khac, „One-class Collective Anomaly Detection based on Long Short-Term Memory Recurrent Neural Networks," CoRR, abs/1802.00324, 2018.

[17] J. M. Beaver, R. C. Borges-Hink und M. A. Buckner, „An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications," in 2013 12th International Conference on Machine Learning and Applications, 2013.

[18] Li, P.; Niggemann, O.; Hammer, B.: On the Identification of Decision Boundaries for Anomaly Detection in CPPS. In: 20th IEEE International Conference on Industrial Technology (ICIT 2019) Melbourne, Australia, 2019.

[19] Li, P.; Niggemann, O.; Hammer, B.: A Geometric Approach to Clustering Based Anomaly Detection for Industrial Applications. In: 44th Annual Conference of the IEEE Industrial Electronics Society (IECON) 2018.

[20] Shuo Yang, et. al; Learning Continuous-Time Bayesian Networks in Relational Domains: A Non-Parametric Approach; In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016.

[21] Christian R. Shelton, Yu Fan, William Lam, Joon Lee, and Jing Xu. "Continuous Time Bayesian Network Reasoning and Learning Engine." Journal of Machine Learning Research, 1137-1140, 2011.

[22] Sürmeli, B. G., Eksen, F., Dinc ¸, B., Schüller, P., Tümer, B.: Unsupervised mode detection in cyber-physical systems using variable order Markov models. In: Industrial Informatics (INDIN), 2017 IEEE 15th International Conference on. IEEE, pp. 841–846

[23] D. Miljković, "Brief review of self-organizing maps," 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1061-1066, 2017.

[24] Omlin, C.W., Giles, C.L., Constructing deterministic finite-state automata in recurrent neural networks. J. ACM 43, 937–972. 1996.

[25] Jacobsson, H. Rule extraction from recurrent neural networks: A taxonomy and review. Neural Computing, 2005.

[26] Wang, Q., Zhang, K., Ororbia II, A.G., Xing, X., Liu, X., Giles, C.L., An Empirical Evaluation of Rule Extraction from Recurrent Neural Networks. Neural Computing. 30, 2568–2591, 2018.

[27] Deng, L., Seltzer, M., Yu, D., Acero, A., Mohamed, A.-R., Hinton, G.E., Binary Coding of Speech Spectrograms Using a Deep Auto-encoder. Interspeech 1692–1695, 2010.

[28] Hinton, G., Salakhutdinov, R. Discovering binary codes for documents by learning deep generative models. Top. Cogn. Sci. 3, 74–91, 2011.

[29] E. Reiter. An architecture for data-to-text systems. In Eleventh European Workshop on Natural Language Generation. URL https://aclanthology.info/papers/W07-2315/w07-2315, 2007.

[30] R. Puduppully, L. Dong, and M. Lapata. Data-to-text generation with content selection and planning. CoRR, abs/1809.00582. URL http://arxiv.org/abs/1809.00582, 2018.

[31] F. Nie, J. Wang, J. Yao, R. Pan, and C. Lin. Operation-guided neural networks for high fidelity data-to-text generation. In 2018 Conference on Empirical Methods in Natural Language Processing, pages 3879–3889, 2018.

[32] Devlin, J., et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[33] Lan, Z., et al. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, 2019.

[34] Storks, Shane; GAO, Qiaozi; CHAI, Joyce Y. Commonsense reasoning for natural language understanding: A survey of benchmarks, resources, and approaches. arXiv preprint arXiv:1904.01172, 2019.

[35] A. Adadi und M. Berrada, „Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),“ IEEE Access, Bd. 6, pp. 52138-52160, 2018.

[36] Ribeiro, M. T., Singh, S., & Guestrin, C., "Why should i trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016.

[37] F. Specht, J. Otto und B. Hammer, „Generation of Adversarial Examples to Prevent Misclassification of Deep Neural Network based Condition Monitoring Systems for Cyber-Physical Production Systems”, IEEE 16th International Conference on Industrial Informatics (INDIN), 2018.

[38] Li, P.; Niggemann, O.: A Data Provenance based Architecture to Enhance the Reliability of Data Analysis for Industry 4.0. In: 23th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA) Sep 2018.

[39] BMBF Projekt AVA – Lernen von zeitbehafteten Automaten für die Anlagendiagnose, 2011-2014

[40] EU Projekt IMPROVE - Innovative Modelling Approaches for Production Systems to Raise Validatable Efficiency, 2015-2017

[41] BMBF Initiative Lernlabor Cybersicherheit, 2016 - 2021

[42] BMBF Projekt GLACIER - Erkennung von IT-Sicherheitsvorfällen durch effiziente Analyse von Netzwerkdaten, URL https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/glacier. 2019-2021

[43] BMBF Projekt AICAS - Sicherheitsvorfälle in Industrieanlagen wirksam erkennen und abwehren. URL https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/aicas. 2019-2021