Jürgen Jasperneite
Ulrich Jumar (Hrsg.)

Kommunikation in der Automation

Beiträge des Jahreskolloquiums KommA 2024, Lemgo

OPEN ACCESS

inIT | TH OWL | ifak

Editor:
Benedikt Lücke
Institut für industrielle Informationstechnik - inIT | Technische Hochschule Ostwestfalen-Lippe

Impressum

15. Jahreskolloquium
Kommunikation in der Automation
(KommA 2024)

06. November 2024 • Lemgo

OPEN-Book

ISBN: 978-3-9818463-5-5
DOI: https://doi.org/10.25644/jnac-vp34

Herausgeber:
Jürgen Jasperneite, Lemgo
Ulrich Jumar, Magdeburg

Institut für industrielle Informationstechnik
Technische Hochschule Ostwestfalen-Lippe

Campusallee 6
D-32657 Lemgo
Telefon: +49 5261 702-2400
Internet: www.init-owl.de | www.jk-komma.de

Vorwort

15. Jahreskolloquium Kommunikation in der Automation – KommA 2024

Am 06. November 2024 fand am Institut für industrielle Informationstechnik - inIT der Technischen Hochschule Ostwestfalen-Lippe das 15. Jahreskolloquium der Reihe KommA – Kommunikation in der Automation statt. Das abwechselnd von den Instituten inIT, Lemgo und ifak, Magdeburg veranstaltete Kolloquium ist ein bewährtes Forum für Wissenschaft und Industrie zu allen technisch-wissenschaftlichen Fragen rund um die industrielle Kommunikation. Die Veranstaltung wird durch die ITG und die Gesellschaft für Informatik unterstützt.

Im Fokus des Jahreskolloquiums stehen Kommunikationssysteme – vom Feldbus, über Echtzeit-Ethernet bis zur drahtlosen Kommunikation und der Nutzung von IoT-Technologien. Bei der Systemanalyse und dem Entwurf von Kommunikationssystemen widmen sich die Tagungsbeiträge der formalen Modellierung, der Verifikation und Validierung sowie Interoperabilität, Konformität und Test. Als bedeutsame Aspekte vernetzter eingebetteter Echtzeitsysteme behandelt das KommA-Kolloquium u.a. die Dienstgüte, semantische Interoperabilität, Safety und Security, die Systemintegration und das Engineering. Neben aktuellen Kommunikationstechnologien bilden die vielfältigen Anwendungsbereiche der industriellen Kommunikation einen Schwerpunkt.

Unbenommen vom reichhaltigen Veranstaltungsangebot auf dem Gebiet der automatisierungstechnischen Kommunikation hat das Jahreskolloquium KommA seinen speziellen Platz. Neben den großen internationalen Tagungen wird bewusst der Austausch im kleinen Kreis gepflegt. Bei aller Technologie- und Anwendungsorientierung des Kolloquiums steht die Wissenschaftlichkeit im Fokus. Alle Beitragskurzfassungen werden deshalb von mindestens zwei Vertretern des Programmkomitees begutachtet. Der Verzicht auf parallele Sitzungen befördert den Gedankenaustausch.

Damit trotz der kurzen Tagungsdauer die wichtigsten Themen, Entwicklungen und Trends adressiert werden können, gehören neben den Vorträgen zusätzlich Poster zum Programm. Wie im Fall der Vorträge finden sich auch die Vollmanuskripte der als Poster präsentierten Beiträge im elektronischen Tagungsband.

Prof. Dr. Jürgen Jasperneite

Institut für industrielle Informationstechnik - inIT
Technische Hochschule Ostwestfalen-Lippe,
Fraunhofer IOSB-INA

Prof. Dr. Ulrich Jumar

ifak e.V. - Institut für Automation und Kommunikation e.V.
Otto-von-Guericke-Universität Magdeburg

Komitees

Tagungsleitung:

Prof. Dr. Jürgen Jasperneite

Institut für industrielle Informationstechnik - inIT | Technische Hochschule Ostwestfalen-Lippe, Fraunhofer IOSB-INA

Prof. Dr. Ulrich Jumar

ifak e.V. - Institut für Automation und Kommunikation e.V. | Otto-von-Guericke-Universität Magdeburg

Programmkomitee:

Stefan Bollmeyer

ABB Asea Brown Boveri Ltd

Holger Büttner

Beckhoff Automation GmbH & Co. KG

Prof. Dr. Christian Diedrich

Otto-von-Guericke-University

Prof. Dr. Mathias Fischer

Universität Hamburg

Prof. Dr. Mesut Günes

Otto-von-Guericke-University

Marco Henkel

WAGO GmbH & Co. KG

Gunnar Lessmann

PHOENIX CONTACT Electronics GmbH

Dr. Jan Stefan Michels

Weidmüller GmbH & Co. KG

Markus Rentschler

ARENA2036 e.V.

Detlef Tenhagen

HARTING Deutschland GmbH & Co. KG

Prof. Dr. Thilo Sauter

Technische Universität Wien

Dr. Sebastian Schriegel

Fraunhofer IOSB-INA

Prof. Dr. René Simon

Hochschule Harz

Prof. Dr. Henning Trsek

Institut für industrielle Informationstechnik -inIT

Technische Hochschule Ostwestfalen-Lippe

Dr. Christoph Weiler

Siemens AG

Prof. Dr. Jörg F. Wollert

FH Aachen

Prof. Dr. Martin Wollschlaeger

Techn. Universität Dresden

Organisationskomitee:

Jasmin Zilz

Institut für industrielle Informationstechnik - inIT | Technische Hochschule Ostwestfalen-Lippe

Benedikt Lücke

Institut für industrielle Informationstechnik - inIT | Technische Hochschule Ostwestfalen-Lippe

Tim Esau

Institut für industrielle Informationstechnik - inIT | Technische Hochschule Ostwestfalen-Lippe

Jannik Peters

Institut für industrielle Informationstechnik - inIT | Technische Hochschule Ostwestfalen-Lippe

Integration of TETRA functionalities with 5G MCX: Feasibility Analysis & Research Opportunities

Maxim Friesen[1], Arne Neumann[1], Denis Gustin[2], Björn Kroll[2], Timo Siekmann[2], Pragya Agarwal[3], Philipp Alda[3], Thomas Conrath[4], and Kim Petersen[4]

Abstract: The Terrestrial Trunked Radio (TETRA) system has been essential for mission-critical communications in public safety and operational radio. However, its limitations in bandwidth and scalability are increasingly problematic as demands for higher data rates and real-time multimedia communication grow. This paper explores how 5G Mission Critical Services (MCX) can address these challenges, particularly in Public Protection and Disaster Relief (PPDR) and Autonomous Vehicle Control (AVC) scenarios. By analyzing the communication requirements of these use cases, the study identifies TETRA's shortcomings, proposes a nomadic 5G architecture to enhance the reliability and scalability of future mission-critical communications and discusses future research opportunities.

Keywords: TETRA, 5G, MCX

1 Introduction

The Terrestrial Trunked Radio (TETRA) system has established itself as a major radio standard among public protection authorities and in operational radio [Pu08]. TETRA bundles their necessary communication requirements by creating a uniform, robust, and secure communication system that combines various features, such as voice communication (individual and group calls), data transmission, and emergency call functions into one system. However, given the rapid technological developments and the increasing use of new digital services that require ever larger amounts of data, TETRA systems are increasingly reaching their limits in terms of bandwidth and scalability [Be22]. Given these challenges, the need for an alternative, more scalable communication solution is becoming increasingly clear.

Despite the established position of TETRA, there have been efforts to supplement this technology and its application areas with newer commercial broadband technologies such as 5G mobile communications [Cr21]. To meet the requirements of public protection and disaster relief (PPDR) authorities, the 3GPP has introduced mission-critical (MC) services

1 Technische Hochschule Ostwestfalen-Lippe, Institut für industrielle Informationstechnik, Campusallee 6, 32657 Lemgo, Deutschland, maxim.friesen@th-owl.de; arne.neumann@th-owl.de
2 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Deutschland, denis.gustin@iosb-ina.fraunhofer.de; bjoern.kroll@iosb-ina.fraunhofer.de; timo.siekmann@iosb-ina.fraunhofer.de
3 T-Systems International GmbH, Hahnstraße 43D, 60528 Frankfurt am Main, Deutschland, pragya.agarwal@t-systems.com; philipp.alda@t-systems.com
4 HMF Smart Solutions GmbH, Fritz-Hahne-Straße 7, 31848 Bad Münder, Deutschland, thomas.conrath@hmf-germany.com; kim.petersen@hmf-germany.com

such as Push-to-Talk, (PTT) group calls, as well as authentication and encryption since Release 13. In the 3GPP standard, these services are included under Mission Critical Services (MCX) [3G20b]. Although these standards have laid the groundwork for implementing mission-critical communication through broadband technologies, a practical validation of these concepts is still pending through appropriate application scenarios. To assess the applicability of the 5G MCX services, it is necessary to identify deployment scenarios where TETRA has been used for critical communication tasks, investigate their requirements in terms of the communication infrastructure and subsequently analyze to what extent the 5G MCX standards can meet these requirements.

2 Related Work

The shift from narrowband Land Mobile Radio (LMR) systems to broadband networks like 5G for MC communications has already been investigated in recent research. As per Ericsson's guidelines on migrating to 4G/5G mission-critical services [Er21], the transition from legacy systems to 3GPP-compliant mission-critical broadband services requires a gradual, methodical process to ensure operational needs such as availability, security, and interoperability. In this context, studies such as [Sa24] already analyzed the capacity challenges of 5G networks in supporting MCX, particularly for scenarios involving large-scale group communications. The research highlights key bottlenecks in radio access networks and the difficulties in meeting 3GPP's stringent requirements for MCX services under typical network conditions. The study in [Fr23] further underscores the importance of integrating broadband technologies with smart city infrastructure to enhance emergency response capabilities. It evaluates the potential impacts and benefits of migrating to digital MC systems while addressing the challenges specific to urban environments. [Sa19] analyzes how different network configurations, primarily LTE, LTE with Multi-access Edge Computing (MEC) and 5G, affect latency and other performance indicators for MCPTT application scenarios. They found that while current LTE networks struggle to consistently meet MCPTT requirements, the shift towards 5G, especially with MEC, significantly improves these KPIs, making 5G a promising solution for future MC communications. Similarly, the feasibility of MCPTT over LTE has been investigated in [Ch19], where the study examines the standardization and implementation of MCPTT as an evolution of traditional group communication services for public safety. The research demonstrates that LTE, when combined with the appropriate technologies to minimize end-to-end latency, can effectively meet the stringent performance requirements of MCPTT. The integration of 5G with TETRA has also already been explored in [HL24]. Mainly to enhance safety and information security in smart railway application scenarios. The study analyzes the combined use of 5G and TETRA in railway systems, highlighting how a 5G-TETRA system can support real-time monitoring, safety measures, and operational efficiency, making it a viable solution for modern railway communication needs.

2.1 Contribution

While existing research has already explored the integration of 5G with MC systems like TETRA, it mainly focused on specific benefits like increased capacity and lower latency, or only emphasized a migration to LTE. This study analyzes how 5G can cover TETRA-specific use cases and its requirements. In this context, the feasibility of implementing 5G MCX services in critical communication areas, including public safety and operational radio, is explored to develop a general understanding of the capabilities and limitations of 5G in these specialized application fields.

3 Background

3.1 Terrestrial Trunked Radio (TETRA)

The TETRA standard, developed and standardized by the European Telecommunications Standards Institute (ETSI), is primarily used for mission-critical and business-critical communications within closed user groups. The standard supports narrowband radio technology designed for both voice and data transmission, including a high-quality audio codec optimized for human voice. TETRA operates in two main modes: "network-connected mode"(Trunking Mode Operation, TMO) and "device-to-device mode"(Direct Mode Operations, DMO). In TMO, the routing of communications within the network infrastructure functions similar to how public cellular networks operate. In contrast, DMO allows devices to communicate directly with each other without relying on network infrastructure, which is particularly useful in situations where network coverage is unavailable. Additionally, TETRA supports advanced modes such as DMO repeater mode, which extends the range of direct device-to-device communication, and a network relay mode (TMO-DMO gateway), which bridges communications between devices in direct mode and those connected to the network. These modes ensure reliable communication in enclosed areas with low signal penetration or poor network coverage.

A key feature of TETRA is its PTT functionality, which allows users to initiate group communication (Point-to-Multipoint) with the press of a button without requiring additional call setup procedures. Hereby, TETRA offers typical call setup times of around 300 milliseconds. Although the standard does not specify this as a maximum latency or Key Performance Indicator (KPI), this is crucial for public protection authorities and transport operators to coordinate responses quickly and effectively during emergencies or operational activities.

The most significant feature of TETRA is its group call functionality, which supports a range of standardized call priority parameters, including pre-emptive and emergency priorities. This allows critical communications to be prioritized over routine messages. This is further enhanced by features such as Broadcast Calls, Talkgroup Patching, Late Entry,

Call Forwarding, Presence Indication, and Talking Party Identification (TPI). For individual one to one voice communication, TETRA supports both Full-Duplex calls, similar to commercial telephone networks, allowing for private conversations, and Half-Duplex PTT-operated calls, which are useful for structured communication in operational environments. Additional features like Calling Line Identification (CLI) enhance user convenience, while Ambience Listening (AL) can improve situational awareness by allowing other users to passively monitor an user in an emergency condition. To accommodate dynamic operational needs, TETRA provides Over-The-Air (OTA) assignment of dynamic groups and group memberships. This allows real-time adjustments to communication structures as situations evolve.

For regular data transfers, TETRA includes Short Data Service (SDS) for text messaging to groups and individuals, the transfer of status identifiers (SDS Status) representing defined operational user statuses, and the transfer of geo-locations using standardized formats (Location Information Protocol, LIP). Additionally, TETRA provides basic IP data transfer functionality (TETRA Packet Data, PD), although it lacks the low latency and high bandwidth needed for more demanding applications, such as real-time video streaming or bulk data transfers.

In Trunking Mode, wireless communication resources for both the uplink and downlink are managed by the network infrastructure. Devices communicate over a control channel using the Time Division Multiple Access (TDMA) method, which divides the frequency into four 6,25 kHz bandwidth channels per carrier. Depending on the encoding scheme, this results in a maximum data rate of 7,2 Kbit/s for each slot, totaling 28,8 Kbit/s for all four slots, excluding error correction overhead. Due to the limited bandwidth of the TETRA radio standard, new application scenarios such as real-time video streaming, advanced group communication with simultaneous duplexing, or the control and monitoring of autonomous vehicles are not possible. This leads to an increased reliance on alternative solutions as a means of communication such as public broadband networks by emergency personnel [Sc20]. Public transport operators on the other hand continue to use analog radio for specific applications, such as traffic light prioritization, because the added delays for signing in and out of a traffic controller when using TETRA make a practical implementation infeasible [Be22].

3.2 Mission-Critical Services (MCX)

The MCX framework, developed under the 3rd Generation Partnership Project (3GPP) and supported by the European Telecommunications Standards Institute (ETSI), provides standardized communication services designed for operation over cellular networks, including 5G. These services encompass mission-critical voice (MCPTT), data (MCDATA), and video (MCVIDEO) and are intended to meet the stringent requirements of public safety and other critical communication needs.

MCPTT (Mission-Critical Push-To-Talk) replicates and extends the PTT functionality given in TETRA, enabling similar group communication capabilities. Moreover, MCPTT provides advanced features like dynamic group management, pre-emption priorities, and emergency priority calls. Like TETRA, MCPTT supports group calls with a wide range of priority levels, ensuring critical communications can be prioritized. For individual voice communication, MCPTT supports both Full-Duplex and Half-Duplex calls. This includes features like TETRA's Full-Duplex private calls for confidential communication and Half-Duplex calls for structured communication environments. Moreover, Device-to-Device (D2D) functionality in 5G allows devices to communicate one to one without relying on the network infrastructure, similar to TETRA's Direct Mode Operations.

MCDATA supports SDS text messaging to groups and individuals and the transmission of status updates (SDS Status) and geo-location data, similar to TETRA's Location Information Protocol. The IP data streaming capabilities benefit from the overall low latency and high-bandwidth characteristics of 5G, enabling new use cases for mission-critical Internet-of-Things (IoT) or Internet-of-Life-Saving-Things (IoLST) applications.

MCVIDEO introduces additional capabilities that are not possible with TETRA, such as real-time video streaming for enhanced situational awareness. It includes features like Push-To-Video group calls, allowing users to share live video feeds with options for pre-emptive and emergency priority. Video functionalities extend to Broadcast Calls, private video calls and Video Push/Pull features, enabling video content to be transmitted to and from command centers or mobile units in real-time. These video services are complemented by Ambient Viewing, a video-based variant of TETRA's Ambience Listening, providing visual situational awareness through live video feeds from body cams or similar devices without active user participation.

4 Application Scenarios

4.1 Public Protection and Disaster Relief (PPDR)

TETRA's ability to prioritize communication, manage group calls and maintain a secure communication, make it ideal for PPDR operations. However, its limitations in bandwidth make it unsuitable for serving high-capacity data transmissions such as video streams or handling a large number of simultaneous high-quality voice channels. Consequently, there is a need to enhance the situational awareness and decision-making capabilities of emergency services by utilizing 5G for real-time data transmission. This includes live video streaming via Ambient Viewing to provide a comprehensive situational overview without interfering with first responders' work. Emergency personnel can use video calls and share operational data to receive external assistance, thereby improving response times and overall efficiency. As visualized in Fig. 1, potential PPDR deployment scenarios can involve nomadic 5G networks, used on operational vehicles to enable local communication at the

emergency site independently of the public network. These private networks enable direct and secure communication between the emergency services on site. For communication with the command headquarters (HQ), a gateway to the public mobile network is set up with a custom high priority network slice. The private radio network can offer a flexible, temporary and non-public network environment tailored for emergency situations. A local server within the operational vehicles hosts the radio core and enables digital data processing and communication services through edge applications. Mobile 5G-enabled handhelds are used at the scene to enable necessary voice and video communication with each other and with the control center.

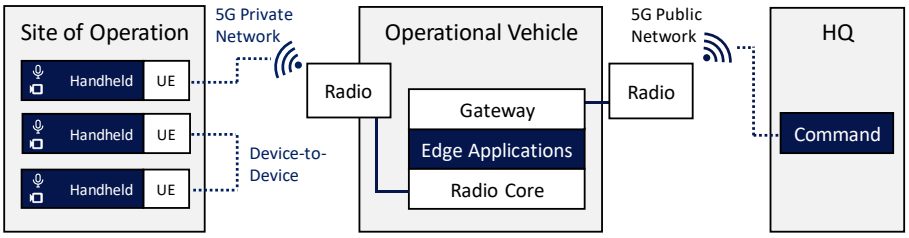


Fig. 1: Nomadic 5G network deployments for mission-critical PPDR communication services.

4.2 Autonomous Vehicle Control (AVC) in Public Transport

TETRA systems have proven to be well-suited for handling direct communication between operators or carrying signal information, even in challenging environments such as underground networks or densely built urban areas. TETRA's ability to integrate with other systems, like Intermodal Transport Control Systems (ITCS) and Remote Bus Location (RBL) systems, and its robust integration of voice and data in a single platform allowed public transport operators to coordinate effectively. However, with the adoption of AVC systems, TETRA's limitations in terms of bandwidth and latency become increasingly critical. Autonomous trains require real-time data exchange to function safely and effectively. This includes the need for low-latency communication for remote control, high-definition video streaming for situational awareness, and real-time sensor data exchange to ensure precise navigation and obstacle detection. In scenarios as visualized in Fig. 2, where multiple vehicles operate in close proximity, such as platooning, maintaining a reliable and low-latency communication link is essential. The vehicles need to reliably communicate with each other, with a central control system and with infrastructure elements like level crossings or traffic management systems. Such Vehicle-to-Everything (V2X) scenarios envision the use of 5G private networks as central platforms that can replace TETRA for signaling and voice communications. They can further integrate comfort services like passenger information and critical applications such as safety system monitoring through edge applications close to the private network.

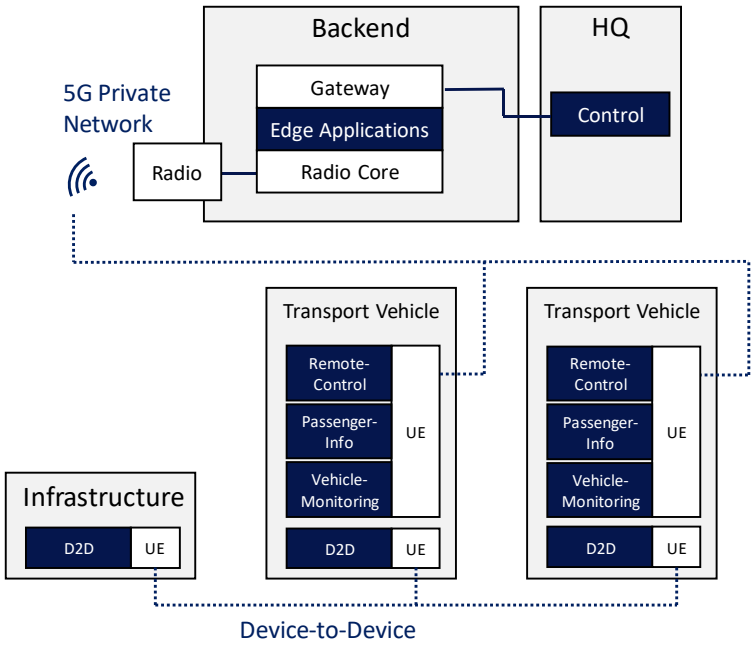


Fig. 2: 5G network deployments for mission-critical V2X communication services.

4.3 Performance Requirements

Based on the selected application scenarios, service requirements towards network performance are detailed in Table 1. These requirements are aligned with the 3GPP defined service requirements for public safety [3G24, 3G20a] and vehicle-to-everything scenarios [3G20c]. The end-to-end latency of less than 30 ms (PPDR-1) and call setup times of less than 300 ms for both group and direct calls (PPDR-2) are necessary for rapid response times during emergencies. MCPTT services must achieve access times of less than 300 ms for 95% of all requests, and 99% of requests for emergency group calls. Additionally, the end-to-end MCPTT access time must be less than 1000 ms when a group call has not been established prior to the MCPTT access request. A minimum bandwidth of 12 Mbit/s per user in both downlink and uplink (PPDR-3) is necessary to support data-intensive applications such as Ambient Viewing. This considers typical traffic scenarios involving a mix of MCPTT, MCVideo, and MCData services, where each operator might utilize multiple devices, including one dedicated to video. The network must handle high uplink traffic efficiently, particularly during high-activity periods, where 30% of devices might transmit simultaneously. The system's ability to support communication for up to 30 UEs within a single building and maintain a communication range exceeding 50 meters indoors (PPDR-4, PPDR-5) ensure reliable communication in complex, multi-level structures and

Tab. 1: Service requirements towards network performance for PPDR and AVC scenarios.

ID	Requirement
PPDR Communication Requirements [3G24, 3G20a]	
PPDR-1	End-to-end latency must be less than 30 ms.
PPDR-2	Access times should not exceed 300 ms for both group and direct calls and less than 1000 ms for calls that were not pre-established .
PPDR-3	A minimum bandwidth of 12 Mbit/s downlink and 12 Mbit/s uplink per user.
PPDR-4	The system must support communication for up to 30 UEs within a single building (100 m x 100 m area, covering up to 3 floors).
PPDR-5	Communication range must exceed 50 meters indoors.
Autonomous Vehicle Control (AVC) Communication Requirements [3G20c]	
AVC-1	Ultra-reliable communication is required with 99.999% availability.
AVC-2	End-to-end latency must be less than 5 ms for device-to-device communication and less than 100 ms for control connections.
AVC-3	Each vehicle must support a minimum uplink data rate of 25 Mbit/s and a downlink rate of 1 Mbit/s.
AVC-4	Communication range must be greater than 350 meters.
AVC-5	Payload sizes of up to 6500 bytes with a transmission rate of 50 messages per second.
AVC-6	The system must support the simultaneous communication of up to 200 UEs within the same network.

are aligned with 3GPP's specifications on network relaying capabilities, similar to TETRA's DMO relay and repeater modes.

The need for ultra-reliable communication with 99.999% availability (AVC-1) reflects the critical nature of AVC operations, where failures are costly and can lead to safety hazards. An end-to-end latency of less than 5 ms for D2D communication and less than 100 ms for control connections (AVC-2) is essential for real-time responsiveness in vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. This is required for maintaining coordination in vehicle platooning scenarios, where vehicles must react to each other as quickly as possible. A minimum uplink data rate of 25 Mbit/s per vehicle, with a downlink rate of 1 Mbit/s (AVC-3), supports the extensive data exchange required between autonomous vehicles and their control centers or edge applications. This includes the transmission of sensor data, video streams, and control commands, needed for AVC operation and situational awareness. The communication range of greater than 350 meters (AVC-4) ensures reliable long-distance communication, assuming vehicle speeds of up to of 130 km/h. However, the system must facilitate message exchanges at absolute speeds of up to 250 km/h. Furthermore, the system must handle payload sizes of up to 6500 bytes with a transmission rate of 50 messages per second (AVC-5) to support the large data volumes typically associated with vehicle control scenarios.

5 Feasibility Analysis & Implications

The transition from TETRA to 5G technologies in mission-critical applications necessitates a comprehensive understanding of how TETRA functionalities can be mapped onto the 5G 3GPP standards. Both TETRA and 5G provide essential services such as voice and data communication, crucial for public institutions like police, fire brigades, and public transport systems. TETRA's established framework includes Direct Mode and Trunked Mode Operation, analogous to the on-network and off-network capabilities of MCX. In 5G, these functionalities are facilitated through mechanisms such as Proximity Services (ProSe) for direct communication and Network Slicing for managed communication within a network.TETRA has been a reliable and indispensable communication system for mission-critical use cases, PPDR scenarios. It meets several key performance requirements, including call setup times of around 300 ms and high-speed operations capable of handling ground speeds up to 300 km/h, making it suitable for high mobility use cases [ET08]. TETRA's range coverage in urban areas typically reaches 4-5 km, extending up to 58 km in open air at 400 MHz, which provides adequate coverage for most communication needs.

However, TETRA's limitations become evident when considering more demanding future scenarios, such as those involving autonomous vehicle control (AVC). The bandwidth provided by TETRA, limited to 28.8 Kbit/s using all four slots, is insufficient for high data rate applications. Additionally, TETRA's SDS exhibits significant delays, with end-to-end transmission times ranging from 0.4 to 1.6 seconds depending on the message size [AX07]. These delays are too long for scenarios where low latency is critical, such as real-time sensor data transmission or V2X communications, which require latencies as low as 5 ms. The limited payload size and transmission rate of SDS further render it unsuitable for high-demand applications requiring rapid transfer of large amounts of data. Moreover, TETRA's capacity for user density is restricted due to its TDMA approach, which requires separate time slots for each user in duplex communication, limiting the number of simultaneous users that can be supported. This limitation is particularly problematic in dense urban environments or large-scale deployments, where a high number of devices need to be connected simultaneously.

One of the critical aspects of TETRA technology that has made it indispensable for mission-critical use cases is its robust framework for preferential access to resources. TETRA allows for comprehensive prioritization mechanisms based on multiple factors, including call priority, type of call, and subscriber-type prioritization, ensuring that critical communications receive priority access. However, while TETRA excels in these areas, its ability to handle more advanced requirements such as those posed by AVC scenarios is limited. In contrast, 5G technologies, particularly under the 3GPP standards, are evolving to incorporate similar, if not more advanced, prioritization mechanisms. Network slicing, a unique feature in 5G, allows for the creation of multiple virtual networks within the same physical infrastructure, each optimized for specific use cases with distinct priority levels. This is particularly beneficial for complex communication environments like those required by AVC and modern PPDR scenarios. Additionally, 5G's application-based prioritization

leverages Quality of Service (QoS) parameters, enabling precise management of network resources based on bitrate reservation, latency requirements, and reliability criteria.

While TETRA has been well-suited for past and current mission-critical applications, its limitations in terms of data throughput, latency, and user density indicate that it may not be fully equipped to handle the advanced requirements of future communication scenarios. The transition to 5G networks, with their enhanced capabilities, is necessary to meet the demands of applications like AVC and modern PPDR scenarios. However, the implementation of 5G MCX core functionality [3G20b] is still in development, and further work is required to fully realize the potential of these new technologies in mission-critical environments.

6 Research Opportunities

Our exploration of integrating TETRA functionalities with 5G systems reveals several research opportunities to further investigate how current communication infrastructures can evolve to meet future demands. A conceptual architecture as shown in Fig. 3 could meet the requirements for both PPDR and AVC use cases. It presents two potential deployment models: a nomadic 5G system with local hosting of the MCX server, and a static 5G system using cloud-hosted MCX services.

In the nomadic setup, the MCX Core is hosted locally. This setup is well-suited for PPDR use cases, where emergency services require autonomous, local network operation in scenarios where access to a broader network might be unavailable. The MCX Core provides critical functionalities (MCPTT, MCVIDEO etc.) for MC communications, ensuring that users can maintain operational capabilities even in isolated environments. The IP Multimedia Subsystem (IMS) manages multimedia services over IP, handling the control layer and enabling seamless transitions between voice, video, and messaging services. For the static setup, the MCX services can be cloud-hosted, connecting to the broader public network via the core infrastructure. This configuration is beneficial for AVC use cases, where reliable and continuous connectivity is required for tasks such as real-time control and monitoring of autonomous vehicles.

The architecture can dynamically scale to adapt to varying communication needs, especially in emergency situations where the number of connected devices and data throughput may fluctuate significantly. While the nomadic system is self-contained and ensures operational usability in the absence of public network connectivity, it can seamlessly integrate with a wider network via the cloud-hosted MCX server when available. This is essential for public safety operations that may shift between urban, rural or disaster-affected areas.

Maintaining high QoS in such a setup requires research into techniques that ensure low-latency, reliable communications even under heavy network load. Both the PPDR and AVC use cases demand stringent QoS management to prioritize voice, video, and data traffic under stressful conditions. Further studies are needed to evaluate how the architecture can

balance these requirements, particularly in resource-limited environments like the nomadic 5G system.

Testing and validation through real-world trials are critical to ensure the architecture meets the performance requirements for MC applications. Collaboration with emergency services and public transport operators will be important to refine and validate the system's reliability, scalability, and operational feasibility.

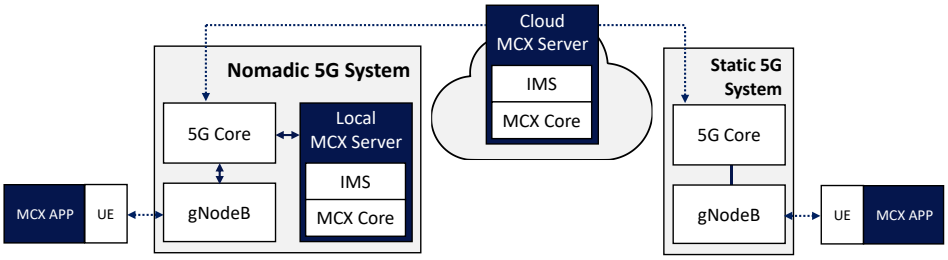


Fig. 3: Flexible 5G MCX System Architecture.

7 Conclusion & Outlook

While TETRA has served as a critical radio standard for various MC applications, its limitations necessitate the adoption of more advanced solutions such as 5G's MCX. These services offer enhanced capabilities for real-time communication, which are essential for modern public safety and operational radio use cases. This paper examines the limitations of TETRA within these application domains and explores the potential for integrating 5G with TETRA functionalities. A conceptual architecture was presented to show how 5G networks can be deployed in a nomadic, localized setup and in more static environments using cloud-based services. The next steps involve the technical implementation and rigorous real-world testing of this architecture to ensure it meets the stringent QoS requirements of the identified PPDR and AVC applications. Collaboration with first responders and transport operators will be crucial to refine and validate the system's performance. By moving from legacy TETRA systems to 5G, this integration aims to enhance the reliability and efficiency of MC communications, addressing the complex demands of modern emergency and operational environments.

Bibliography

[3G20a] 3GPP: Mission Critical Push to Talk (MCPTT); (3GPP TS 22.179 version 16.5.0 Release 16), November 2020.

[3G20b] 3GPP: Mission Critical Services Common Requirements (MCCoRe); (3GPP TS 22.280 version 16.8.0 Release 16), August 2020.

[3G20c] 3GPP: Service requirements for enhanced V2X scenarios (3GPP TS 22.186 version 16.2.0 Release 16), November 2020.

[3G24] 3GPP: Service requirements for the 5G system (3GPP TS 22.261 version 18.13.0 Release 18), May 2024.

[AX07] Axiotis, Dimitrios I; Xenikos, Dimitrios G: On the performance of TETRA short data service-transport layer. Wirel. Pers. Commun., 43(4):1121–1135, November 2007.

[Be22] Berichte der Bundesanstalt für Straßenwesen: Nutzung der C2X-basierten ÖV-Priorisierung an signalisierten Knotenpunkten. Verkehrstechnik, Heft V 353, January 2022.

[Ch19] Choi, Sang Won; Song, Yong-Soo; Shin, Won-Yong; Kim, Juyeop: A feasibility study on mission-critical push-to-talk: Standards and implementation perspectives. IEEE Commun. Mag., 57(2):81–87, February 2019.

[Cr21] Crisis Prevention: "TETRA bleibt noch für viele Jahre entscheidende Technologie". Fachportal für Gefahrenabwehr, Innere Sicherheit und Katastrophenhilfe, February 2021.

[Er21] Ericsson AB: Migration to mission critical 4G and 5G: The technical and operational aspects for adopting 3GPP based broadband services, July 2021.

[ET08] ETSI: ETSI TR 102 753 V1.1.1 (2008-05): TETRA mobiles moving at high velocity. Technical report, May 2008.

[Fr23] Freire, Débora Vanessa Campos: Mission-Critical Communications from LMR to 5G: a Technology Assessment approach for Smart City scenarios. PhD thesis, NOVA University Lisbon, 2023.

[HL24] Hsiao, Liang-Sheng; Lin, I-Long: Integrating 5G and terrestrial trunked radio into railway communication system for railway safety and information security. Sensors and materials, 36(3):1275, March 2024.

[Pu08] Public Manager: Motorola installiert neues TETRA-Digitalfunknetz für die Verkehrs-Aktiengesellschaft Nürnberg. Rolf Soll Verlag GmbH, December 2008.

[Sa19] Sanchoyerto, Aitor; Solozabal, Ruben; Blanco, Bego; Liberal, Fidel: Analysis of the impact of the evolution toward 5G architectures on mission critical push-to-talk services. IEEE Access, 7:115052–115061, 2019.

[Sa24] Santiago, Adrian; De Cerio, Andoni Diaz; Sanchoyerto, Aitor; Liberal, Fidel: Analysis of mission critical services radio access network capacity limitations over 5G. IEEE Access, 12:6191–6203, 2024.

[Sc20] Schmidt, Stephanie: Einsatzmittel Smartphone: Nutzung von Mobiltelefonen im polizeilichen Arbeitsalltag. Institut für Bürgerrechte & öffentliche Sicherheit e.V., May 2020.

Drahtlose Innovation: Wie 5G RedCap die Industrie transformiert

Björn Kroll[1], Denis Gustin[1] und Dimitri Block[2]

Abstract: Automatisierungsanwendungen sind auf eine zuverlässige und echtzeitfähige Kommunikation angewiesen. Gleichzeitig sind neben den funktionalen Anforderungen, den prozesskritischen Parametern auch wirtschaftliche Sichtweisen im Fokus. Damit eine Automatisierungslösung am Markt bestehen kann, müssen die Anschaffungs- und Betriebskosten je nach Anwendung in einem betriebswirtschaftlich darstellbaren Rahmen bleiben. Industrial 5G bietet die Möglichkeit die Kommunikation zwischen den einzelnen Prozessteilnehmern zu vereinfachen, ist bisher aber nur in wenigen Fällen aufgrund der Preisstruktur sinnvoll einsetzbar. Dieses Paper evaluiert das Potential durch 5G-RedCap-Modulen sowohl für etablierten 5G Lösungen als auch für neue Industrial 5G Use Cases und Zielgruppen. Exemplarisch wird eine anlagenunabhängigen Prozessüberwachung mittels Industrial 5G basierend auf den funktionsreduzierten und kostenoptimierten 5G RedCap Modulen untersucht. Die Verbindungsqualität wird durch etablierte Messverfahren wie RFC-2544, Y-1564 sowie RFC6369 evaluiert. In diesem Paper wird der Nachweis erbracht, dass 5G RedCap Module den Return on Investment (ROI) von Industrial 5G Lösungen steigern können und dadurch auch neue Industrial 5G Business Cases realisiert werden können.

Keywords: Industrial 5G, ROI, Business Cases

1 Motivation

Die technologische Reife von Industrial 5G wurde in den letzten Jahren hinlänglich nachgewiesen. Durch die kürzlich erfolgte Standardisierung weiterer Schnittstellen (z.B. O-RAN), ist ein netzseitiger Trend hin zu einem stärkeren Marktwettbewerb wie zum Beispiel O-RAN-Applikationen zur vermehrten Kommunikation der Verkehrsteilnehmenden im Straßenverkehr (engl. vehicle-to-x, kurz V2X) siehe [Li23] oder auch für satellitengestützte 5G-Kommunikation (non-terrestrial networks, kurz NTN) siehe [CAV23]. Das deutlich breiteren 5G-Netz-Angebot steigert folglich die Rentabilität von 5G-Netzen wahrnehmbar.

Die Standardisierungsorganisation (3rd Generation Partnership Project) der fünften Mobilfungeneration (5G) hat in der dritten Hauptversion des 5G-Standards, dem 3GPP Release 17, funktionsreduzierte 5G-Module unter dem Namen "RedCap"(Reduced Capability) spezifiziert [Pr22]. Durch die Funktionsreduktion wurden zusätzlich u.a. der Energiebedarf, die erzeugte Wärme, die mechanischen Abmessungen, die Anzahl der benötigten Antennen und damit folglich die Kosten eines 5G-Moduls reduziert [Ve22]. Durch die neuen 5G-RedCap-Module werden zahlreiche Use Cases für IoT (Internet of Things) Applikationen, feldnahe

1 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Deutschland, bjoern.kroll@iosb-ina.fraunhofer.de; denis.gustin@iosb-ina.fraunhofer.de

2 Weidmüller Interface Interface GmbH & Co. KG, Klingenbergstraße 26, 32758 Detmold, Deutschland, dimitri.block@weidmueller.com

Sensorik, applikationsnahe Steuerungsaufgaben und weitere spezifische Applikationen mittels Industrial 5G wirtschaftlich abbildbar.

In diesem Paper werden dementsprechend unterschiedliche Konfigurationen und Variationen von 5G-RedCap-Modulen betrachtet und evaluiert. Es werden insbesondere der theoretische und praktische Einsatz von 5G-RedCap-Modulen und deren Einordnungen in mögliche Automatisierungsszenarien gegenübergestellt. Dabei werden etablierten Messverfahren wie zum Beispiel basierend auf RFC-2544, Y-1564 sowie RFC-6369 angewendet. Die Verbindungsqualität wird durch CIR (Committed-Information Rate), EIR (Excess-Information Rate) sowie auch die Latenz, Jitter, FLR (Frame-Loss Rate) und den CBS (Committed Burst Size) quantifiziert.

Nach der technischen Leistungsbewertung folgt die Evaluierung der wirtschaftlichen Verwertung. Aktuell finden Mobilfunk-basierte Komponenten vorrangig in der Kommunikationsinfrastruktur Verwendung. Zum Beispiel werden Mobilfunk-Zugangspunkte für die Fernwartung eines drahtgebundenen Automatisierungsnetzes eingesetzt. Durch die kostenreduzierten 5G-RedCap-Modulen ist die 5G-Konnektivität von Komponenten nicht nur für die Kommunikationsinfrastruktur rentabel sondern kann bis zu den Endgeräten wie einem Sensor oder Aktuator wirtschaftlich realisiert werden. Durch weniger drahtgebundener Kommunikationstechnologien in industriellen Applikationen wird letztendlich die Komplexität von Maschinen und Anlagen. Zusätzlich bietet eine Funkkommunikation eine erhöhte Flexibilität und Mobilität für die Applikation.

Wenn 5G-RedCap-basierte Komponenten eine ähnliche Leistungsfähigkeit aufweisen, wie Komponenten mit herkömmlichen 5G-Modulen mit vollem Funktionsumfang, hier 5G-FullCap-Module gennant [BA23], lassen sich allein durch die Kostenreduktion der Komponenten deutlich mehr Business Cases realisieren und damit zwangsläufig die Kosten für weitere Lösungen senken.

2 Problembeschreibung

Ein Einsatz von 5G-RedCap-Modulen in Endgeräten wie einem Sensor oder Aktuator in Automatisierungsapplikationen erfordert eine Kommunikation mittels Feldbus. Um 5G-RedCap-gestützte Komponenten für eine Feldbus-Kommunikation zu qualifizieren, gibt es mehrere Probleme die es zu bewältigen gilt. In [Ca23] ist die Problemstellung die ein Feldbus wie z.B. PROFINET an ein Netzwerk stellt übersichtlich dargestellt. Dabei stellt ein deterministisischer Jitter eine der anspruchsvolleren Anforderungen dar.

Neben den technischen Betrachtungen ist ein weiterer Grund für die Nutzung von 5G-RedCap-Module, die zu erwartenden geringeren Kosten. Diese sind bedingt durch den einfacheren Hardwareaufbau z.B. durch weniger Antennen. Dies kommt insbesondere dem Hardwaredesign bei Automatisierungskomponenten entgegen. Die wichtigste Fragstellung hierbei ist, welche Einschränkungen sind zu erwarten im Bezug auf die Anbindung der

Moduls an vorhandene Systeme (z.b. bedingt durch unterschiedliche Schnittstellen) und welche Änderungen der vorhandenen Hardware insgesamt zu erwarten sind.

3 Stand der Technik: 5G Campusnetze in der Industrie

Seit der Standardisierung in 2019 mit dem 3GPP Release 15 wird 5G in diversen industriellen Applikationen erprobt. Der Fokus liegt dabei auf sogenannte private 5G Campusnetze. Diese Netze können von Unternehmen lokal betrieben werden. Dafür ist in Deutschland das für industrielle Anwendungen relevante Frequenzband zwischen 3700 MHz und 3800 MHz reserviert.

Wie in z.B. [Gu23] oder [Ad20] beschrieben, ermöglichen 5G Campusnetze private Mobilfunknetze für Produktionsumgebungen, Intralogistik und anderen industriellen Anwendungsfeldern. Gegenüber alternaitven WLAN-basierten Campusnetzen ergeben sich hierbei Vorteile vor allem aufgrund der deutlich erhöhten erlaubten Sendeleistung, des koordinierten Mediumszugriffs, die erhöhte IT-Sicherheit und die exklusive Nutzung des Frequenzbandes.

Die Zuverlässigkeit der 5G-Kommunikation kombiniert mit den in 3GPP Release 16 eingeführten Ultra-Reliable-Low-Latency (URLLC) Features ermöglicht geringe Übertragungslatenzen sowie ein relativ geringer Jitter. In [Mo23] wurden zum Beispiel Round-Trip Zeiten von rund 10 ms erreicht bei einem ca. 4 ms Jitter. Diese Optimierungen erfüllen die Echtzeitanforderungen von Feldbus- und Safety-Kommunikation wie PROFINET und PROFISAFE über 5G.

Allein in Deutschland sind Stand August 2024 insgesamt 409 Zuteilungen von Frequenzen für 5G Campusnetze von der Bundesnetzagentur für den für industrielle Anwendung relevanten Sub-6-GHz-Frequenzbereich zwischen 3700 MHz und 3800 MHz erteilt worden[Bu24]. Neben den Forschungs- und Telekomunukationsbranchen wurde jede vierte Frequenzzuteilung für die verarbeitenden Metall- und Elektronik-Branche erteilt. Zusätzliche Evaluierungen können mit hybriden Deployment Varianten unter der Einbindung eines öffentlichen 5G Netz durchgeführt wie in [5G19] dargestellt. Einige der Technologieerprobungen wurden bereits erfolgreich abgeschlossen, sodass die ersten privaten 5G Netze für die produktive Nutzung in industriellen Applikationen geplant und aufgebaut werden wie zum Beispiel [BA23; Jo24].

4 Theoretische Betrachtung erreichbarer Datenraten

5G-RedCap-Module sind bewusst so konzipiert, dass diese einen vermeintlich einfacheren Hardwareaufbau haben. Dies bedeutet aber im Umkehrschluss auch, dass die z.b. nutzbare Bandbreite geringer ist.

Aktuell sind am Markt zwei verschiedene 5G-RedCap-Module verfügbar. In Tabelle 1 sind die Geräten verglichen.

Insbesondere bei der Erreichung der Datenraten sind mehrere Punkte zu beachten.

1. Die Rechnerisch maximale Datenrate
2. Die erreichbare Datenrate der Module
3. Die Empfangsparameter (z.B. RSRQ)
4. Bandbreite des Netzes (max 20 Mhz)

Die Empfangsparameter werden hierbei nicht betrachtet. Für die Bandbreite wird immer der Maximalwert veranschlagt.

Die Rechnerisch mögliche Datenrate errechnet sich laut 3GPP TS 38.306 folgendermaßen:

$$\sum_{j=1}^{J}(v_{Layers}^{j} * Q_{m}^{j} * f^{j} * R_{max} * \frac{N_{PRB}^{BW(j),\mu}}{T_{S}^{\mu}} * (1 - OH^{j})) \quad (1)$$

Hierbei gilt J ist die Anzahl der Frequenzband Komponenten bzw. Kombinationen, v_{Layers}^{j} ist die Anzahl der Mimo PDSCH für den Downlink und PUSCH für den Uplink, Q_{m}^{j} ist die Modulationsreihenfolge (z.B. Qm 2 = QPSK, 4 = 16QAM, 6 = 64QAM, 8 = 256QAM), f^{j} ist der Skalierungsfaktor, R_{max} = 948/1024, $N_{PRB}^{BW(j),\mu}$ ist die Bandkonfiguration nach 3GPP 38.104 (In diesem Fall 20Mhz, $\mu = 30khz$), T_{S}^{μ} ist die Durchschnittliche OFDM Symboldauer und zuletzt $1 - OH^{j}$ definiert den Overhead.

Wenn hierraus ein 5G Netz gebildet wird mit den Werten, die z.B. das Quectel RM255 angibt, ergibt sich hierbei folgende Datenrate (basierend auf dem Slotformat 45 nach 3GPP 38.213 Tabelle 11.1.1-1):

$$\sum_{j=1}^{J}(2^{j} * 6^{j} * 1^{j} * 0.92578125 * \frac{51^{1(j),\mu}}{0.00003571428571428572^{\mu}} * (0.14^{j})) = 93,6\,Mbps \quad (2)$$

Der Upload ist hierbei mit rund 100 Mbps aufgrund des Slot Formates 45 vergleichbar.

In dieser Konfiguration bietet das 5G Netz nicht den maximalen Datendurchsatz der Module. Die Quectel RM255 z.B. sind für Datenraten von maximal 220 Mbps spezifiziert. Dies stellt aber insbesondere für PROFINET kein Hinderniss dar.

Bei PROFINET werden im Durchschnitt 256 Byte pro Paket übetragen. Bei einer Beispielkonfiguration von 1 Endgerät und einer Zykluszeit von 3 ms ergeben sich hier 0,65 Mbps

pro Stream. Diese addieren sich auf Basis der anzusteuernden Endgeräte. Dies sind aber lediglich Durchschnittswerte. Je nach PROFINET-Konfiguration sind maixmal 1440 byte Nutzdaten pro Gerät erlaubt. Diese werden ergänzt durch den Header und zyklische Daten. Typische Konfigurationen sind dann 64 Byte, 256 Byte und 1500 Byte. Dieser werden auch als Service Configurationen für die Tests angenommen.

Die spezifische Datenrate von PROFINET errechnet sich folgendermaßen:

$$Datenrate\ in\ kBit = \frac{(\frac{1000ms}{P/s} * P_{byte}) * 8}{1000} \tag{3}$$

Da die 5G-RedCap-gestützte Geräte tendentiell auf einem PROFINET Busteilnehmer verbaut, ist daher in der Bandbreite keine Einschränkungen zu erwarten. Sollte das Gerät in einer SPS genutzt werden, lassen sich theoretisch 143 Busteilnehmer pro 5G-RedCap-gestütztes Gerät erreichen.

Aktuell muss aber für Layer 2 Traffic zusätzlich ein Tunnel über das 5G System gebildet werden. Dieser hat einen Overhead von ca. 50 byte (VXLAN) wie in [Mu21] dargestellt. Dieser wird in dieser Betrachtung nicht weiter beachtet, da in 3GPP Release 16 Systemen die Möglichkeit besteht eine Ethernet PDU Session aufzubauen, damit ist der VXLAN Tunnel dann überflüssig.

5 Theoretische Betrachtung erreichbaren Latenz

Die Latenzanforderungen der Industrie sind deutlich höher, da hier Protokolle mit festen Zykluszeiten zum Einsatz kommen. Dementsprechend wurde im Kapitel 4 das Slotformat 45 gewählt. Dieses definiert jeweils 6 Download/Upload Symbole und 2 freie Symbole. Dadurch ergeben sowohl in Upload als auch Downloadrichtung ähnliche Datenraten. Aber gleichzeitig auch vergleichbare Latenzen.

Ein Radio Frame ist jeweils 10ms lang in diesem sind 10 Subframes zu je 1ms und bei 30Khz Subcarrier Spacing 2 Slots pro Subframe. Dies ergibt eine theoretische Übertragungsdauer von 500 µs pro Slot.

Zusätzlich können wie in [Mo23] dargestellt Minislots in einem System genutzt werden, um die Latenz und vor allem den Jitter zu optimieren. Minislots benutzen 2 bis 7 OFDM Sysmbolen dauerhaft und erlauben es asynchron zum Slotlayout Daten zu übertragen. Hierdurch lassen sich Daten ausserhalb eines Slot basierten Scheduler übetragen. Bei 2 OFDM Symbolen liegt hierbei dann die Übertragungszeit bei 70 µs.

Die Datenraten werden dadurch aber wieder entsprechend kleiner (siehe Kapitel 4)

Die theoretische Betrachtung betrifft allerdings lediglich die Luftschnittstelle. Da sich die Anzahl der möglichen Netzwerkkomponenten in einem realen System sehr stark unterscheidet, müssen hier individuell Tests durchgeführt werden, um eine Einschätzung der Möglichen Latenz zu bekommen.

6 Testaufbau

6.1 ITU-T Y-1564 Messverfahren

Das ITU-T Y-1564, auch bekannt als Ethernet Service Activation Test Methodology (SAM), ist ein umfassendes Testverfahren zur Validierung von Ethernet-Diensten vor deren Inbetriebnahme. Es wurde entwickelt, um sicherzustellen, dass Netzwerke die Leistungsanforderungen moderner Ethernet-Dienste erfüllen.

Das Verfahren besteht aus zwei Hauptphasen: dem Service Configuration Test (SCT) und dem Service Performance Test (SPT). Der SCT überprüft, ob die Netzwerkdienste gemäß den vereinbarten Service Level Agreements (SLAs) korrekt konfiguriert sind. Hierbei werden mehrere Parameter wie Durchsatz, Frame-Verlustrate und Latenzzeit nacheinander getestet. Der SPT bewertet anschließend die langfristige Leistungsfähigkeit der Dienste gleichzeitig unter Lastbedingungen, indem dieselben Parameter über einen längeren Zeitraum gemessen werden.

Es werden sowohl beim SCT als auch beim Nachfolgenden SPT-Test mehrere Schlüsselparameter überprüft:

1. **Durchsatz (Throughput):** Misst die maximale Datenrate, die der Dienst ohne Frame-Verluste übertragen kann.
2. **Frame-Verlustrate (Frame Loss Rate):** Bestimmt den Prozentsatz der Frames, die während der Übertragung verloren gehen. Ein niedriger Wert ist hierbei entscheidend für die Dienstqualität.
3. **Latenzzeit (Latency):** Erfasst die Verzögerung bei der Übertragung von Datenpaketen über das Netzwerk. Geringe Latenzzeiten sind besonders wichtig für zeitkritische Anwendungen.
4. **Jitter:** Misst die Variation der Latenzzeit zwischen aufeinander folgenden Datenpaketen. Ein niedriger Jitter-Wert ist wichtig für Anwendungen wie VoIP und Videostreaming.
5. **Fehlerrate (Error Rate):** Überprüft die Anzahl der fehlerhaften Frames im Vergleich zur Gesamtzahl der übertragenen Frames.

Zusammengefasst bietet das ITU-T Y-1564 Messverfahren eine strukturierte und effiziente Möglichkeit, die Leistungsfähigkeit und Zuverlässigkeit von Ethernet-Diensten zu validieren, bevor diese in den produktiven Betrieb übergehen.

6.2 Hardwareaufbau

Aktuell sind 5G-RedCap-Module von den Modulherstellern Simcom, Fibocom und Quectel auf dem Markt. In Tabelle 1 finden sich Beispielhaft Datenblätter der Module von Simcom und Quectel. Die Module sind von allen Herstellern ähnlich aufgebaut, daher werden sich diese auch mit der Verfügbarkeit von weiteren Modulen anderer Hersteller nicht substanziell ändern. Alle hier gezeigten Testergebnisse, wurden mit dem Quectel RG255C erstellt.

Feature	SIM8230X 5G (Simcom)	RG255C-GL (Quectel)
Hersteller	Simcom	Quectel
Modellnummer	SIM8230X 5G	RG255C-GL
Technologie	5G NR (RedCap)	5G NR (RedCap)
Frequenzbänder	Unterstützt gängige 5G-Bänder	Unterstützt gängige 5G-Bänder
Formfaktor	M.2 oder LGA	LGA
Max. Download	Bis zu 220 Mbps	Bis zu 220 Mbps
Max. Upload	Bis zu 100 Mbps	Bis zu 100 Mbps
Schnittstellen	USB, PCIe, UART	USB, PCIe, UART
Stromverbrauch	Optimiert für IoT-Anwendungen	
Betriebstemperatur	-40°C bis +85°C	-40°C bis +85°C
Besondere Merkmale	Hohe Integration für IoT-Anwendungen, niedriger Energieverbrauch	

Tab. 1: Vergleich der 5G-RedCap-Modulvarianten

Das 5G Campusnetz wird mittels einer Amarisoft Callbox entsprechend dem 3GPP Release 17 bereitsgestellt. Die gNB ist gemäß 3GPP TS 38.104 mit der Bandbreite 20 MHz und dem Subcarrier Spacing von 30 kHz sowie dem Radio Frame Format 45 (siehe 3GPP 38.213 Table 11.1.1-1) konfiguriert.

7 Testergebnisse

In einer ersten Interation wurden Latenztests durchgeführt. Weiterhin werden zusätzlich Lasttest gemacht, die aber im Sinn der Nutzbarkeit mit PROFINET nur eine begrenzte Aussagekraft haben und dementsprechen hier kein Fokus darauf gelegt wird. Die Datenraten lassen sich auch theoretisch ermitteln, wie in Kapitel 4 erläutert. Fokus ist insbesondere die Zyklusfestigkeit und mögliches Jitterverhalten bei bestimmten Zykluszeiten.

Wie in Abbildung 1 dargestellt ist die durchschnittliche RTT Latenz von 20-40 ms grundsätzlich nicht ideal, insbesondere der Jitter ist mit Durchschnittlich 15 ms sehr hoch. Zusätzlich sind bei einer Payload von 1500 Byte sehr viele Retransmissions zu beobachten. Hierdurch wird die Gesamtlatenz zu hoch und nicht mehr für industrielle Protokolle nutzbar. Mittels eines weiteren Test mit einem 5G-RedCap-Modul Quectel RG255C und mit einem 5G-FullCap-Modul Quectel RM520 im Vergleich wird ersichtlich, wie sich das 5G-RedCap-Modul Quectel RG255C in eine Gesamtperformance einordnet. In Abbildung 2 ist ersichtlich, dass die Modulvarianten durchaus eine ähnliche Performance bieten. Das

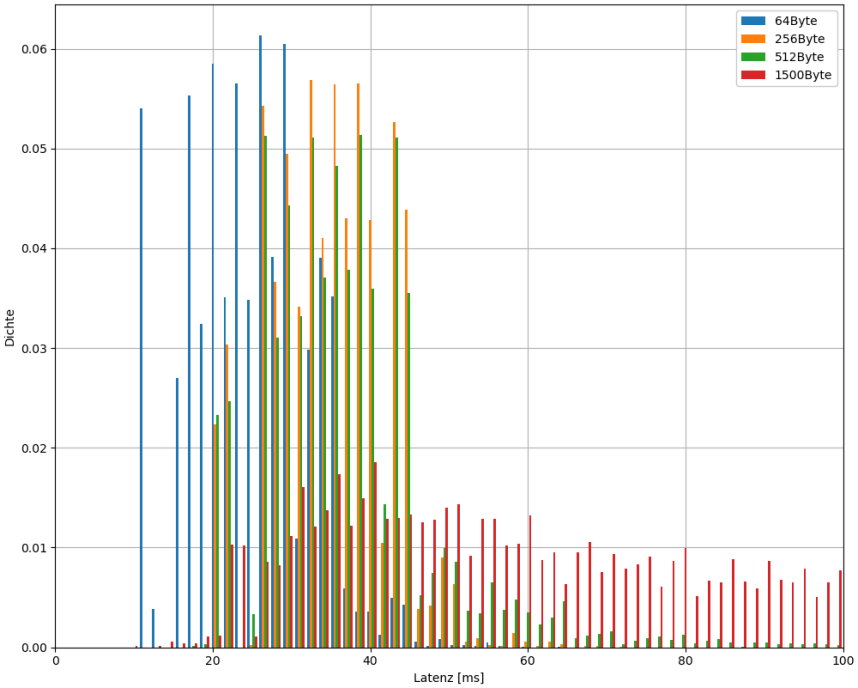


Abb. 1: Histogramm der Round-Trip-Latenzen bei unterschiedlichen Payloads des 5G-RedCap-Moduls Quectel RG255C

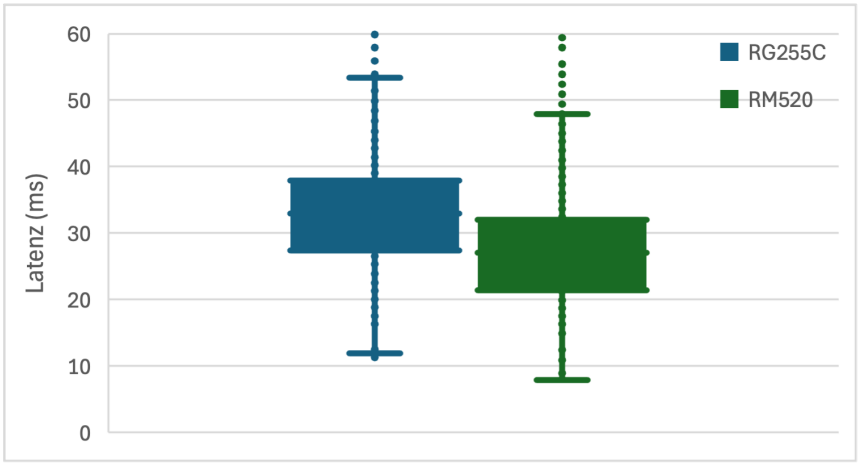


Abb. 2: Round-Trip-Latenz-Messungen in Millisekunden von dem 5G-FullCap-Modul Quectel RM520 und dem 5G-RedCap-Modul Quectel RG255C

5G-RedCap-Modul Quectel RM520 ist für industrielle Protokolle geeignet, dies lässt den Schluss zu, dass beide Varianten geeignet sind. Das Histogramm in Abbildung 1 unterstützt diese Aussage dahingehen, dass der Jitter zwar höher ist, aber sich relativ uniform darstellt. Dies wiederum ist z.B. in PROFINET gut abzubilden. Insbesondere, wenn man die Standardeinstellungen von PROFINET in Betracht zieht, die immer das dreifache der Zykluszeit als Timeout betrachtet. D.H. bei einer Zykluszeit von 24 ms, ist der Timout bei 72 ms. Wenn man die Messung mit 1500 Byte Payload außen vorlässt, wird dieser Wert in keiner Messung erreicht.

8 Zusammenfassung und Ausblick

Die theoretischen Betrachtungen und die Messungen unterstützen das Fazit, dass funktionsreduzierte 5G-RedCap-Module nach dem 3GPP Release 17 für die Kosten eine sehr gute Performance haben und durch die Vorteile, die der einfache Aufbau mit sich bringt, in der Industrie für viele Anwendungen interessant werden. Hier sind insbesondere das Retrofitting prozessunkritischer Sensoren der naheliegende Anwendungsfall. Darüber hinaus ist die Leistung der 5G-RedCap-Module ausreichend, für eine Feldbus-Kommunikation z.B. von PROFINET oder PROFISAFE. Dies eröffnet deutlich mehr Anwendungsfälle insbesondere für dynamischer funktionale Sicherheitsapplikationen in Kombination mit fahrerloser Transportsysteme (FTS).

Dieses Paper stellt einen ersten Eindruck in die mögliche industrielle Nutzung von 5G-RedCap-Module. Sobald die ersten Implementierungen entwickelt sind, werden diese Messungen mit weiteren ergänzt und die Thematik erneut aufgegriffen.

Literaturverzeichnis

[5G19] 5G-ACIA: 5G Non-Public Networks for Industrial Scenarios, Techn. Ber., 5G Alliance for Connected Industries und Automation, 2019.

[Ad20] Adebusola, J. A.; Ariyo, A. A.; Elisha, O. A.; Olubunmi, A. M.; Julius, O. O.: An Overview of 5G Technology. In: 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS). S. 1–4, 2020, DOI: 10.1109/ICMCECS47690.2020.240853.

[BA23] BASF: Vom 5G-Testfeld zum Produktivnetz. 2023.

[Bu24] Bundesnetzagentur: Übersicht der Zuteilungsinhaber für Frequenzzuteilungen für lokale Frequenznutzungen im Frequenzbereich 3.700-3.800 MHz, Techn. Ber., Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2024.

[Ca23] Cainelli, G.; Underberg, L.; Arleving, H.; Sufiye, S.; Master, H.; Velasquez, J.; Sharma, G.; Singh, M.: Performance testing of a 5G network for PROFINET and PROFISafe data transmission from the application perspective. In: 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA). S. 1–8, 2023, DOI: 10.1109/ETFA54631.2023.10275462.

[CAV23] Campana, R.; Amatetti, C.; Vanelli-Coralli, A.: O-RAN based Non-Terrestrial Networks: Trends and Challenges. 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), S. 264–269, 2023.

[Gu23] Gustin, D.; Siekmann, T.; Kroll, B.; Kleen, P.; Schriegel, S.; Jasperneite, J.: Outdoor Field Test of 5G-based V2X Communication for Real-Time Monitoring and Remote Control of a Monorail Vehicle. In: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). 2023.

[Jo24] Jones, D.: Tesla deploys private 5G in its factory in Berlin. Fierce Network, 2024.

[Li23] Linsalata, F.; Moro, E.; Magarini, M.; Spagnolini, U.; Capone, A.: Open RAN-Empowered V2X Architecture: Challenges, Opportunities, and Research Directions. 2024 IEEE Vehicular Networking Conference (VNC), S. 113–116, 2023.

[Mo23] Mondal, N.; Block, D.; Kroll, B.; Klingler, F.: Performance evaluation and application of real-time communication with 5G IIoT, 2023, URL: http://dx.doi.org/10.25673/111641.

[Mu21] Muhammad, T.: Overlay Network Technologies in SDN: Evaluating Performance and Scalability of VXLAN and GENEVE. International Journal Of Computer Science And Technology 5 (1), S. 39–75, 2021.

[Pr22] Project, 3. G. P.: Release 17 Description; Summary of Rel-17 Work Items. 3GPP Technical Report, 2022.

[Ve22] Veedu, S. N. K.; Mozaffari, M.; Hoglund, A.; Yavuz, E. A.; Tirronen, T.; Bergman, J.; Wang, Y.-P. E.: Toward Smaller and Lower-Cost 5G Devices with Longer Battery Life: An Overview of 3GPP Release 17 RedCap. IEEE Communications Standards Magazine 6, S. 84–90, 2022.

Development and Analysis of Deployment Strategies for Smart Services Using Various Hardware Assets for Industrial Applications

Robin Foster[1], Simon Althoff[2], and Henning Trsek[1]

Abstract: This paper analyses several deployment strategies for containerized services in an industrial ecosystem, as there is no definitive standard on deployment yet. To be regulatory compliant for the future, important aspects of the IEC 62443 series of standards are being considered. This includes key aspects from the fields of cybersecurity to procedural aspects. The evaluation of the strategies also includes quantitative arguments such as deployment times and qualitative aspects such as maintainability. It is shown that a solid strategy includes a device management component that ensures proper management of the ecosystem.

Keywords: Smart Service, Digital Twin, DevOps

1 Introduction

On one hand, there are companies that have various machines on the shop floor. These machines have various computing resources such as PLCs or industrial PCs and can collect various data from the machines. On the other hand, smart services are needed to be able to derive added value from this data. Especially, small and medium-sized enterprises often do not have the capacities to demonstrate the necessary competences in the fields of action [MF20]. Manufacturers of machines can, for example, provide services for the remaining useful lifetime, error recognition or for billing purposes. Various strategies are presented in the following to enable a possible modular solution for the deployment of these smart services with different resource requirements and dependencies on different hardware.

One objective of this paper is to enable a smart service to run on various hardware architectures such as ARMv7, ARMv8 or x86 without having to adapt the actual source code of the service for the different architectures. This is needed to be able to use existing and new hardware in the whole cloud-edge spectrum. Various virtualization methods are available to achieve this. Using the Java Virtual Machine (JVM), it is possible to run the same source code on various platforms [Ra01]. Utilizing QEMU, it would be possible to emulate the necessary hardware for a compiled source code [Be05]. Both approaches provide basic ideas for the implementation in this paper. For a modern smart service, the Python interpreter and execution using container-based virtualization are used [Al18]. There

[1] Technische Hochschule OWL, Institut Industrial IT, Campusallee 6, 32657 Lemgo, Deutschland, robin.foster@th-owl.de; henning.trsek@th-owl.de, https://orcid.org/0000-0002-0133-0656

[2] Weidmüller Interface GmbH & Co.KG., Klingenbergstraße 26, 32758 Detmold, Deutschland, simon.althoff@weidmueller.com

are various container runtimes such as containerd or CRI-O which use runC or gVisor at lower levels [Es20]. We rely on the Docker platform, which uses containerd and runC as it is widespread and easy to use. The decision to use Python for all smart services was driven by its growing prominence in machine learning, making it the language of choice for this focus area [PAC16].

The digital twin, in particular, an asset administration shell of a smart service must contain diverse information. This paper touches on this topic only briefly, focusing on the provision of information for an automated orchestration component. For example, these can be interface descriptions to be able to link smart services, information about the hardware resource requirements or other configuration options for a service. Regarding cybersecurity, information such as the software bill of materials can also be stored in CycloneDX or SPDX format. This enhances the ability to conduct vulnerability scans within a smart service ecosystem [Xi23] and should be incorporated into the strategic design of smart services. A possible basis for the digital twin is provided by the "Nameplate for Software in Manufacturing" [ID23].

Software deployments in the IIoT sector are usually carried out in the form of firmware updates. Update clients located in the firmware, such as RAUC or SW-Update, are used for this purpose. These can be orchestrated by frameworks such as hawkBit to roll out updates [SLS21]. These updates solutions usually do not provide additional status information about a rolled-out service. By utilizing a software monitoring solution, it is possible to establish a back-channel for individual deployment strategies, particularly in relation to IEC 62443. Telemetry data must be obtained from the deployment services and infrastructure, which maintain logs and metrics.

2 Related work

Various projects are contributing to the development of smart services ecosystems. These include Siemens Industrial Edge [SIIE] and EdgeX Foundry Platform [EDGE]. Both ecosystems offer data exchange and the management of services on I4.0 IT and OT. Siemens also relies on Docker as a virtualization technology. Both platforms offer an all-in-one solution. A comparison between these two solutions is offered by [Ve23]. Another reference architecture for a smart I4.0 ecosystem is offered by the OpenIndustry 4.0 Alliance [OI4A]. Our paper identifies two key aspects that occur in all architectures, and these are the KPIs for efficient scheduling of smart services. The it's OWL — I4.0 AutoServ project is currently doing research on a solution for small- and medium-sized enterprises in this area [I4AS]. This project also serves as the foundation for the topics we present, including the collection of key performance indicators for smart services and the development of a specialized deployment method. Our approach aims to enable an I4.0 ecosystem to run a service on the smallest possible edge resource and, if this does not provide enough hardware resources, to move the smart service to an on-premise cloud or even a public cloud.

3 Our Proposal for a Smart Service

The perception of a smart service can vary widely. In terms of the consumer market, consumers would describe a self-learning service as smart. Beverungen et al. [Be19] describe the intersection of all descriptions of a smart service as a service capable of being aware of where it is located in an ecosystem and can be addressed by other systems and services in an ecosystem. This also reflects our view of a smart service in the IIoT environment.

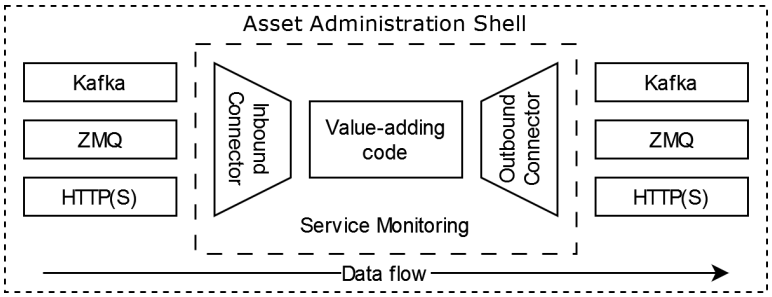


Fig. 1: Smart Service

The smart service we use is illustrated in fig. 1. A value-adding code is embedded in a standardized environment. This value-adding code is a function written in Python, which can fulfill various purposes. It is possible to use several additional Python packages. We have also tested the ONNX runtime for the use of various machine learning models. The value-adding code can communicate with its wrapper code via interfaces and access the configuration of the smart service. Values that are to be used explicitly for the value-adding code can also be stored in this configuration. Interfaces to the inbound and outbound connectors enable to communicate with other systems and services in the ecosystem. The interface is always the same for the value-adding code, regardless of whether Kafka, ZMQ, HTTP(S) or other communication protocols are used externally. This makes it possible to use the service as generically as possible. For example, services can be linked via these data interfaces to create a complex value chain. Data sources and sinks such as databases can also be connected. The wrapping code of the smart service also includes service monitoring. This includes log management of the service and its embedded code. A health check can also be implemented and data on system utilization can be collected. The latter was implemented explicitly for one of our test setups.

In its development phase, the smart service is equipped with an asset administration shell, which is continually enriched in the process of continuous integration. Information about the service itself, test results and hardware requirements are stored in this shell. If an instance of the service is deployed at a later stage, this administration shell is attached to the administration shell of the running instance together with its configuration and additional data.

4 Preparation of a Smart Service

The preparation of a smart service in a continuous integration pipeline is a crucial step before a deployment strategy can be implemented. The service is extensively tested, and the test results are documented. After the completion of the test, this data is subsequently incorporated into the digital twin of the service. The provision of this information serves to satisfy the requirements of IEC 62443.

4.1 Integration Test Setup

The continuous integration pipeline we use for a smart service involves various steps. First, the wrapper code, which serves as a template for all smart services, is tested. This is followed by testing the value-adding code. Static code analysis, code coverage and unit tests are applied. After this step, the system attempts to build the Docker image of the container. To ensure that the intelligent service can run on different end devices with different hardware architectures (ARMv7, ARMv8, x86, . . .), the image is created for the various platforms using the Docker build-x feature. If one of these pipeline steps fails, the developer is informed and has to iterate over his source code again. After the successful build step, the service images are temporarily stored on the host of the ci pipeline system for the next step.

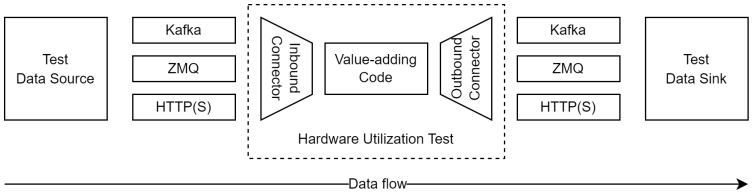


Fig. 2: Integration Test

To determine the hardware resource requirements, the smart service is subjected to an integration test (see fig. 2) in a continuous deployment step. In this test step, test data taken from the ecosystem is passed through the previously built service and the output data is then stored. In addition to the functional test, data on the average inference time and memory requirements can be determined. Technically, the *docker run* command is executed once for each docker image previously created (for each hardware architecture) and the test data is stored on the test server. Subsequently, this test data is stored in the administration shell of the smart service and is available later for the matching between hardware and software.

After the integration test, all the necessary data of the smart service is available for a potential deployment. In the following, the docker images of the service are signed and stored in a container registry for the deployment step.

4.2 Overhead of Testing in an Emulated Environment

One of the objectives of our implementation is to determine the hardware resource requirements of the smart service in the integration test. This is carried out within the ci pipeline, which in our case is executed in an on-premise cloud environment. The hardware platform is x86. The aim of the research was to find out whether the memory requirement of a smart service is identical on the target device when we test it within the integration test. While no deviations are to be expected when running an x86 docker image on an x86 host (was also verified), there are deviations when testing other hardware architectures. More specifically, when an ARMv8 image is tested in the x86 ci pipeline, a different memory requirement is detected on the target device. In the following, a test is carried out to determine whether there is a constant overhead factor. For this purpose, some programming tasks were implemented in python and executed with pseudo random data. The results are visualized in fig. 3 based on the data for the ARMv8 platform.

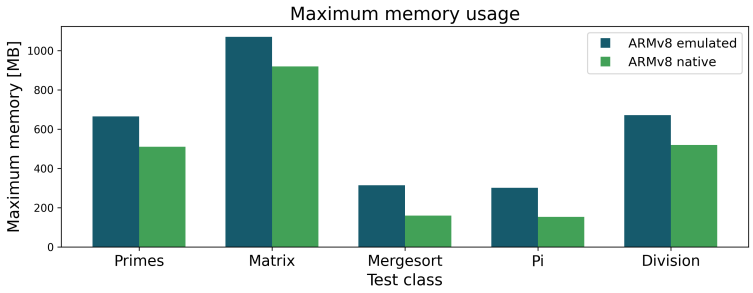


Fig. 3: Memory requirements — ARMv8 emulated vs. ARMv8 native

As the figure illustrates, there is an almost constant overhead for various problems. The problems were chosen so that various data structures have to be used internally. For ARMv8, an overhead of 144 MB±18 MB was measured for the emulated execution. For ARMv7 (not shown) this is almost identical at 145 MB±19 MB. This factor can be included in the matching for the orchestration of service and computing resources.

The technical reason behind this behavior lies in the emulation of the target platform and the variant of measuring the resource requirements. The maximum memory requirement was logged at intervals of 100 ms for the process to be tested. Docker uses QEMU for the emulation. As Cota et al. [Co17] describe, the memory of a QEMU process consists of emulated RAM, memory for I/O and “container” memory regions. This creates an overhead compared to native execution, where the host system’s memory can be used directly by Docker. Richard Jones, Senior Software Engineer at Redhat, describes in his blog [Jo13] a test series in which he examines exactly this overhead and determines an overhead of around 150 MB per QEMU process. This is about the same as the value determined in this test series.

There are two approaches to solving this problem. Firstly, the overhead factors could be considered. Secondly, it would also be possible to use CI Runners that run on the target architecture. In an on-premise cloud, the latter is not an option and extra hardware may have to be maintained purely for this application purpose.

5 Deployment Strategies

An important factor for planning a resource-efficient deployment is the available hardware. In addition to hardware capabilities factors such as connected sensors or machines, available system resources are important for the execution of smart services. The focus is on free memory and disk space, as well as the processor architecture. The ecosystem and the service orchestrator can be informed about these values through administration shells. While the capabilities usually represent a static value, the system utilization is dynamic. If several smart services are installed on a computing resource, these values fluctuate. The administration shells can represent these values dynamically. While the continuous integration of the smart service was concerned with recording the key figures of the service, the key figures of the hardware are now included in the continuous deployments.

Once the continuous integration of a smart service has been completed, it is available in a repository for deployment. A service orchestrator that intends to execute the service subsequently must then perform a matching between hardware and software. To obtain data on the utilization of the ecosystem's existing computing resources, it is necessary to record relevant metrics. For this purpose, it is necessary for the computing resources to report the current metrics via a monitoring pipeline. Metrics acquisition proceeds differently for various deployment strategies.

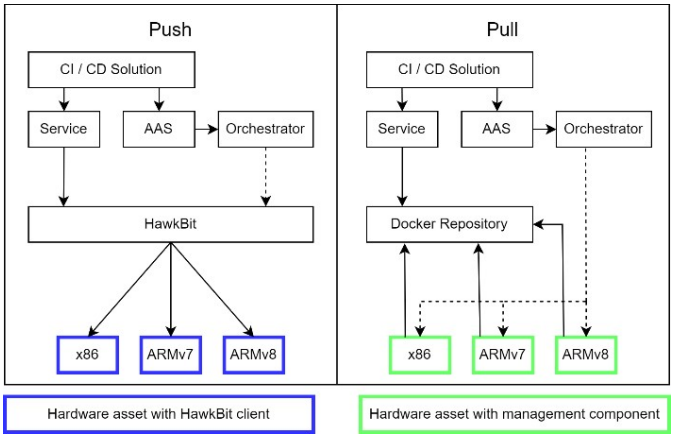


Fig. 4: Push vs. Pull Method

Two deployment strategies are visualized in figure 4. The push method represents the use of the common hawkBit framework. The pull method is our self-developed variant.

5.1 Push Method

A client is installed on each computation resource, which can receive software updates from a hawkBit server (see fig. 5 — SWUpdate Daemon). The docker image of the smart service, an installation script for this image and the configuration of the service are placed in these software update packages. The client periodically requests an update from the hawkBit server. As soon as an update is available, it is downloaded from the hawkBit server and installed. Once the installation is complete, the client reports back to the server whether the installation was successful or not. Logs are not transmitted. The advantage of the push variant is that firmware updates for the computing resource itself can also be transferred via the same channel, as this is hawkBit's actual purpose. The disadvantage of this variant is that the known clients such as RAUC or SW-Update do not send any metric data about the device usage to hawkBit or another service. Another service, including a potential monitoring smart service, would have to be installed on the computing resource for this information. Another disadvantage of this variant is that the polling interval of the client causes delays in the installation of smart services.

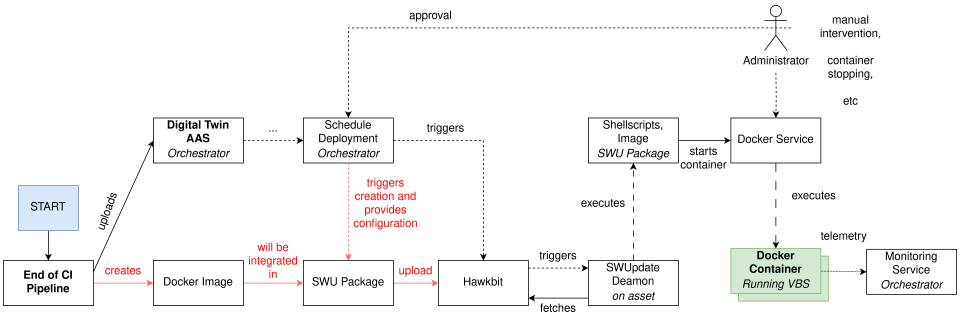


Fig. 5: The Push Method for Smart Service Deployment

5.2 Pull Method

The pull method relies on a smart service, which must be installed on all computing resources in the ecosystem (see fig. 6 — VBS Management Container). This can happen when a new device is onboarded. The management service provides the orchestrator with a REST API that can be used to control the installation, uninstallation, and configuration of a smart service. At the same time, the management service provides the monitoring component, which delivers metric data to the administration shells. Fluent-Bit was used for this in our test setup. Logs from all installed smart services can also be viewed via this container and transmitted to a central location. Data endpoints for the dynamic variables of the administration shells can also be provided. The disadvantage of this method is that an additional channel is required for firmware updates, as the Docker environment does not allow such in-depth changes to the host on an ad hoc basis. Another disadvantage so far is that there is no large developer community behind this component.

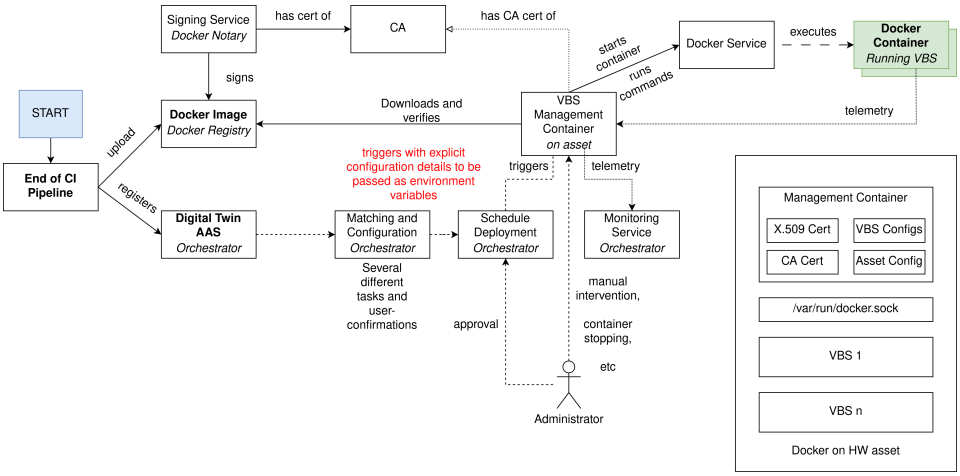


Fig. 6: The Pull Method for Smart Service Deployment

5.3 Evaluation

With a focus on the deployment of smart services, it can be concluded that our pull method offers several advantages over the push method. The resource utilization of the computation asset is reduced during the installation process of a new service, as fewer steps are required for installation. This applies to the deployment time, which could be reduced by around 70 %, as well as an average CPU utilization of around 10 % lower by our method during an active installation process. The nature of Docker images also means that the disk space for the management components can be reduced by using the same basis as the other smart services. RAUC or SW-Update do not offer such an advantage. The disadvantage of the push method is that firmware updates cannot yet be installed via this method. Nonetheless, a significant advantage of our pull approach is that previously installed services can be reconfigured without the necessity of an additional deployment. This is done using an API call. In the push method, the service is required to be rolled out again. From the perspective of IT security within the industrial sector, it is expected that the pull method will cover significant issues in the near future. The primary consideration is the back-channel for service monitoring, which can be implemented in an aggregated manner within a management container. This management container additionally provides an API through which real-time information regarding the hardware asset and installed services can be automatically retrieved, and modifications can be initiated (in addition to the information that is fed into the digital twin). Intervention by an administrator is also possible through the API. All API calls, software containers and communication can be secured centrally in the network using a public key infrastructure (see fig. 6 — CA). Actors (administrators, principals, . . .) can authenticate themselves on devices approved for them.

6 Conclusion and Outlook

This paper presents some of the challenges that can arise in deployment strategies for smart services in the industrial environment. To enable resource-efficient deployment across multiple hardware architectures, various KPIs are required for matching hardware and software. With suitable strategies, these can be collected efficiently. Possible further resource optimizations could be achieved by extending the pull method to also allow firmware updates. In this case, the firmware updater is no longer needed. It would also be conceivable to further adapt the push method by implementing a hawkBit client specialized for smart services. This could probably reduce the overhead of the method. It could also enable the transmission of metric data. This would also close the gap for firmware updates that currently exists in the pull method. It is also feasible to provide signed software through the HawkBit framework; however, this would necessitate the establishment of a Certificate Authority (CA) to serve as a trust anchor, similar to the pull method. It can be concluded that the pull method has clear advantages in terms of resource efficiency and security, primarily for the consideration of smart services.

References

[Al18] Al-Dhuraibi, Y. et al.: Elasticity in Cloud Computing: State of the Art and Research Challenges. IEEE Transactions on Services Computing 11 (2), pp. 430–447, 2018, DOI: 10.1109/TSC.2017.2711009.

[Be05] Bellard, F.: QEMU, a fast and portable dynamic translator. In: Proceedings of the annual conference on USENIX Annual Technical Conference (ATEC '05). USENIX Association, p. 41, 2005.

[Be19] Beverungen, D.; Müller, O.; Matzner, M.; Mendling, J.; vom Brocke, J.: Conceptualizing smart service systems. Electronic Markets 29 (1), pp. 7–18, 2019, DOI: 10.1007/s12525-017-0270-5.

[Co17] Cota, E. G.; Bonzini, P.; Bennée, A.; Carloni, L. P.: Cross-ISA machine emulation for multicores. In: 2017 IEEE/ACM International Symposium on Code Generation and Optimization (CGO). Austin, TX, USA, pp. 210–220, 2017, DOI: 10.1109/CGO.2017.7863741.

[EDGE] EdgeX Foundry: EdgeX Foundry, [Online]. Available: https://www.edgexfoundry.org. [Accessed: 03-Jun-2024].

[Es20] Espe, L. et al.: Performance Evaluation of Container Runtimes. In: Proceedings of the 10th International Conference on Cloud Computing and Services Science - CLOSER. Pp. 273–281, 2020, DOI: 10.5220/0009340402730281.

[I4AS] it's OWL: I4.0 AutoServ, [Online]. Available: https://its-owl.de/projekte/automatisiert-maschinendaten-erheben-und-aufbereiten-i4-0autoserv/. [Accessed: 03-Jun-2024].

[ID23] IDTA: 02007-1-0 Nameplate for Software in Manufacturing, [Online]. Available: https://industrialdigitaltwin.org/en/content-hub/submodels. [Accessed: 04-Jun-2024], 2023.

[Jo13] Jones, R. M.: Overhead of QEMU virtualization, [Online]. Available: https://rwmj.wordpress.com/2013/02/13/what-is-the-overhead-of-qemukvm/. [Accessed: 05-Jun-2024], 2013.

[MF20] msg systems ag; Fraunhofer Institut für Produktionsanlagen und Konstruktionstechnik: Studie Digital Twin Readiness Assessment, [Online]. Available: https://www.msg.group/automotive/studie-digital-twin-readiness-assessment. [Accessed: 02-Jun-2024], 2020.

[OI4A] Open Industry 4.0 Alliance: OI4 Reference Architecture, [Online]. Available: https://openindustry4.com/download-center. [Accessed: 03-Jun-2024].

[PAC16] Portugal, I.; Alencar, P.; Cowan, D.: A Survey on Domain-Specific Languages for Machine Learning in Big Data, 2016.

[Ra01] Radhakrishnan, R. et al.: Java runtime systems: characterization and architectural implications. IEEE Transactions on Computers 50 (2), pp. 131–146, 2001, DOI: 10.1109/12.908989.

[SIIE] Siemens: Industrial Edge, [Online]. Available: https://www.siemens.com/global/en/products/automation/topic-areas/industrial-edge.html. [Accessed: 03-Jun-2024].

[SLS21] Srivastava, A. K.; Lilaramani, D.; Sree, K.: An open-source SWUpdate and Hawkbit framework for OTA Updates of RISC-V based resource constrained devices. In: 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4). Bangalore, India, pp. 1–6, 2021, DOI: 10.1109/C2I454156.2021.9689433.

[Ve23] Venanzi, R.; Solimando, M.; Patrali, M.; Foschini, L.; Chatzimisios, P.: Siemens and EdgeX IIoT Platforms: A Functional and Performance Evaluation. In: ICC 2023 - IEEE International Conference on Communications. Rome, Italy, pp. 834–839, 2023, DOI: 10.1109/ICC45041.2023.10278750.

[Xi23] Xia, B.; Bi, T.; Xing, Z.; Lu, Q.; Zhu, L.: An Empirical Study on Software Bill of Materials: Where We Stand and the Road Ahead. In: 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE). Melbourne, Australia, pp. 2630–2642, 2023, DOI: 10.1109/ICSE48619.2023.00219.

Kommunikationsintegration in verteilten Produktionssystemen auf Basis von Verwaltungsschalen

Melanie Stolze[1], Alfed Barnard[1], Duy Lam Tran[1] und Matthias Riedl[1]

Abstract: Durch steigende volatile Marktanforderungen erhält die Entwicklung flexibler, verteilter und anpassungsfähiger Produktionssysteme immer mehr an Bedeutung. Neben der eigentlichen Konstruktion solcher Produktionssysteme muss auch deren Kommunikationsintegration betrachtet werden, da die Modularität und Flexibilität der Produktionssysteme zu einer häufiger auftretenden Rekonfiguration im laufenden Betrieb führen können. Dieser Beitrag stellt ein Konzept vor, das eine standardisierte, konfigurierbare und damit teilautomatisierte Kommunikationsintegration ermöglichen soll. Neben der statischen Beschreibung der Kommunikationsintegration wird zudem ein Lösungsvorschlag aufgezeigt, wie die Produktionssysteme trotz der Verwendung unterschiedlicher TCP basierter Kommunikationsprotokolle wie OPC UA oder MQTT miteinander Daten austauschen können. Das Hauptaugenmerk der vorgestellten Konzepte liegt auf der Verwendung der Verwaltungsschale.

Keywords: Kommunikationsintegration, Verwaltungsschale, Verteiltes Produktionssystem

1 Motivation und Problemstellung

Die Integration standardisierter und etablierter Internetprotokolle des Internet of Things (IoT) in die industrielle Umgebung wurde unter dem Begriff des Industrial Internet of Things (IIoT) geprägt. Das IIoT bietet im Bereich verteilter Steuerungssysteme das Potenzial, intelligente Fertigungssysteme im Kontext von Industrie 4.0 zu realisieren. Der Beitrag fokussiert sich auf die Protokolle MQTT (Message Queuing Telemetry Transport) und OPC UA (Unified Architecture), die traditionell für die vertikale Integration genutzt werden, perspektivisch auch für die horizontale Kommunikation zwischen IIoT-Geräten. Hiermit wird beabsichtigt, Produktionsprozesse dynamisch anpassbar zu machen und so neuen Anforderungen mit hoher Flexibilität und Effizienz gerecht zu werden. Es wird ein Konzept vorgestellt, das zum einen eine einfache Konfiguration der Kommunikationsinfrastruktur und zum anderen den interoperablen Datenaustausch zwischen heterogenen Kommunikationsprotokollen realisieren soll.

1 Institut für Automation und Kommunikation e.V., IKT & Automation, Werner-Heisenberg-Str. 1, 39106 Magdeburg, Deutschland , Melanie.Stolze@ifak.eu, Alfred.Barnard@ifak.eu, Duylam.Tran@ifak.eu, Matthias.Riedl@ifak.eu

Die Arbeit ist in vier Teile untergliedert. Zuerst werden in Abschnitt 2 die aktuellen Herausforderungen zur Kommunikationsintegration von verteilten Systemen sowie deren industrielle Anforderungen angerissen. Daraufhin folgt in Abschnitt 3 ein Einblick in existierende IIoT-Kommunikationsprotokolle und verschiedene Modellierungssprachen, die für die Kommunikationsintegration von IoT-Geräten in eine Kommunikationsinfrastruktur genutzt werden können. Ein Konzept zur komfortablen Konfiguration der Kommunikationsanbindung sowie der ereignisgesteuerten Prozessbeeinflussung zwischen verteilten Systemen ohne zentrale Orchestrierungsdienste wird in Abschnitt 4 beschrieben. Zum Ende folgt in Abschnitt 5 eine kurze Zusammenfassung der Ergebnisse sowie ein Ausblick auf offene Foschungsaufgaben.

2 Herausforderungen und daraus resultierende Anforderungen

Während eines Produktlebenszyklus werden Daten sowohl auf horizontaler als auch zwischen vertikalen Ebenen ausgetauscht. Zu den horizontalen Ebenen zählen unter anderem der Datenaustausch zwischen Wertschöpfungsnetzwerken in Unternehmen und über Unternehmen hinweg sowie auf der Produktionsebene der horizontale Datenaustausch zwischen und innerhalb von Produktionssystemen. Letzteres ist auch unter dem Begriff Machine-to-Machine (M2M)-Kommunikation bekannt.
Damit die ausgetauschten Daten in beiden Ebenen (horizontal und vertikal) korrekt transferiert und interpretiert werden können, ist die Interoperabilität auf all den von [Di20] beschriebenen Stufen gefordert. Angefangen von der technischen Interoperabilität, in der Systeme technisch in der Lage sind, Daten miteinander auszutauschen, über die syntaktische Interoperabilität, repräsentiert durch einheitliche maschineninterpretierbare Datenformate, bis hin zur semantischen Interoperabilität, in der die Daten dem Kontext entsprechend richtig interpretiert werden. [Di20]

Weitere Herausforderungen ergeben sich durch die volatilen Marktanforderungen, durch die eine flexible Anpassungsfähigkeit von Produktionssystemen notwendig wird. Die Modularisierung und dynamische M2M-Kommunikation bilden die Grundlage für ein effizientes und flexibles Produktionssystem [RZ22]. Durch die Modularisierung wird die Rekonfiguration von Systemen an die sich ändernden Anforderungen erleichtert, was zudem die Ausfallzeiten bei dem Wechsel oder Austausch einer Maschine im Produktionssystem senkt.

In diesem Bericht wird der Fokus auf die horizontale M2M-Kommunikation gelegt. Zur Realisierung dieser Kommunikation ist ein in sich geschlossenes, autonomes Verhalten der modularen Maschinen gefordert. Das heißt, Maschinen tauschen nach [Di20] keine einzelnen Variablenwerte mehr aus, sondern stoßen gegenseitig Funktionen an, die in internen Zustandsmaschinen gekapselt sind. Trotz der technischen Fortschritte gibt es im Bereich der M2M-Kommunikation immer noch erhebliche Herausforderungen zu bewältigen. Die Interoperabilität zwischen verschiedenen Systemen und der Bedarf an standardisierten Kommunikationsprotokollen, erfordern weitere Forschung und Entwicklung. [Lu20]

3 Stand der Forschung und Technik zur Kommunikationsintegration

3.1 Überblick

In den folgenden Abschnitten werden verschiedene gängige Modellierungssprachen aufgezeigt, mit denen die Beschreibung einer Kommunikationsintegration vorgenommen werden kann. Jede der Modellierungssprachen zeigt ihre Stärken in unterschiedlichen Phasen des Produktlebenszyklus auf. Während AutomationML überwiegend in der Engineeringphase von Systemen verwendet wird, fokussiert sich der Einsatz von OPC UA auf die flexible Anbindung von modularen Anlagen bzw. Maschinen (im weiteren Systeme genannt) im operativen Betrieb.

Für die horizontale M2M-Kommunikationsanbindung im operativen Betrieb gibt es neben OPC UA weitere TCP (Transmission Control Protocol) basierte Kommunikationsprotokolle die unter anderem auch als IIoT-Kommunikationsprotokolle zusammengefasst werden. Aus der Verwendung unterschiedlicher Protokolle resultiert eine heterogene Kommunikationslandschaft dezentraler Systeme, die miteinander Daten austauschen wollen. Um die Interoperabilität trotz unterschiedlicher Protokolle zu ermöglichen, wird die Verwaltungsschale als Lösungsansatz eingeführt.

3.2 IIoT-Kommunikationsprotokolle

Das IIoT stellt eine an den Industrieanforderungen ausgerichtete Erweiterung der Konzepte des IoT dar. Als Basis werden die Kommunikationsprotokolle der IoT genutzt, die auf standardisierten Internetprotokollen wie IPv4 (Internet Protocol Version 4) oder IPv6 (Internet Protocol Version 6) basieren. Der Transport der Nachrichten geschieht meist durch TCP basierte Protokolle auf denen sowohl HTTP/REST (Hypertext Transfer Protocol/Representational State Transfer), OPC UA als auch MQTT und Modbus/TCP aufbauen. Zu sehen ist der Aufbau des IoT-Protokoll-Stack in Abbildung 1. [We23]

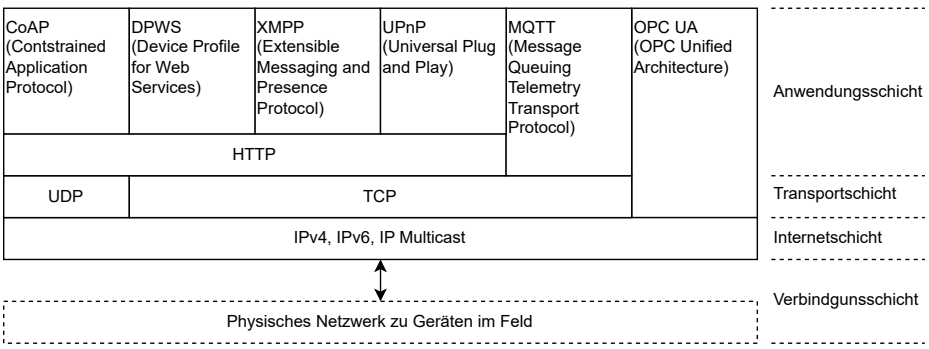


Abb. 1: Protokoll-Stack des IoT (in Anlehnung an [We23])

Im Weiteren wird anstelle des allgemeinen Begriffs IoT der (aufgrund der industriellen Anforderungen an die IoT geprägte) Begriff IIoT verwendet.

3.3 Automation Markup Language

Automation Markup Language (AutomationML) ist kein Kommunikationsprotokoll, sondern dient als eine universelle Beschreibung für einen plattformunabhängigen Datenaustausch. Der Standard zielt darauf ab, ein offenes Datenformat bereitzustellen, das die Schnittstelle zwischen den Planungswerkzeugen der Digitalen Fabrik und der Automatisierungstechnik überbrückt. AutomationML bietet das Potenzial, in allen Planungsphasen eine konsistente Daten- und Informationsverarbeitung sicherzustellen. Speziell zur Beschreibung der Kommunikation zwischen Systemen steht der Teil 5 der IEC 62714 [IE22] zur Verfügung. In diesem sind die Sichten auf die physische Topologie und der logischen Nutzung (gestrichelte Pfeile, siehe Abbildung 2) definiert. Ausgehend von diesen abstrakten Sichten lassen sich konkrete Anwendungen z. B. für die gängigen Echtzeitprotokolle wie PROFINET inklusive der geplanten Kommunikationsbeziehungen mit Links zu den Geräteparametern bzw. Variablenbezeichnern in Steuerungsprogrammen ableiten.

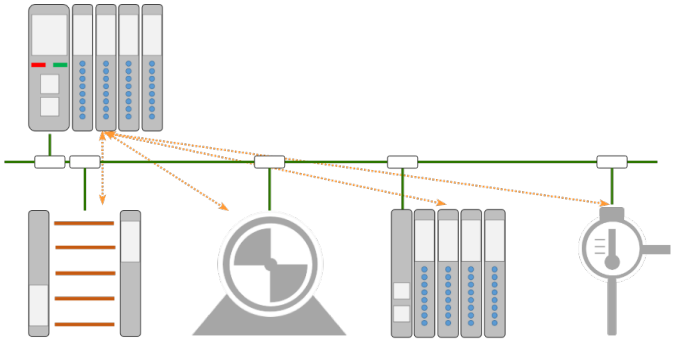


Abb. 2: Beispiel von logischen Kommunikationsbeziehungen [Dr21]

3.4 OPC Unified Architecture

OPC UA wurde mit dem Ziel entwickelt, Systeme aus dem industriellen Bereich miteinander interoperabel zu vernetzen, sodass sie zum einen Informationen miteinander austauschen aber auch Steuerungsmechanismen realisieren können. Damit soll die starre Automatisierungspyramide [We23] zu einem flexiblen Netzwerk von miteinander interagierenden System abgelöst werden. Für die interoperable Beschreibung der Kommunikationsintegration dezentraler Systeme bietet OPC UA das Konzept von Companion Specifications (CS), die im Folgenden beispielhaft aufgeführt sind.

Eine abstrakte CS, auf der viele weitere CS aufbauen, ist die OPC 10000-100 „Devices“ [OP22]. In dieser Spezifikation ist ein Gerät mit seinen Eigenschaften modelliert, wo unter

anderem die vom Gerät unterstützen Protokolltypen dazugehören. Zu diesen Protokolltypen nennt die Spezifikation vor allem Beispiele aus der seriellen Kommunikation wie PROFINET, PROFIBUS (Process Fieldbus) oder HART (Highway Addressable Remote Transducer) sowie weitere Protokolle, die in der Spezifikation nicht näher beschrieben sind. Damit wird eine Stärke von OPC UA ersichtlich, die in der Abstraktion serieller Kommunikationsprotokolle liegt, um bereits in der Feld- und Steuerungsebene einen interoperablen Datenaustausch zu ermöglichen. Bezüglich des TCP basierten Datenaustausches im Fall der M2M-Kommunikation werden in [OP22] keine konkreten Beispiele außer OPC UA selbst als Protokolltyp genannt. Dafür bietet die OPC 10100-1 „WOT Connectivity - API Definition“ [OP24] die Möglichkeit, IIoT-Geräte mit unterschiedlichen Kommunikationsprotokollen wie z. B. MQTT oder HTTP an bestehende OPC UA Server anzubinden. Dies wird mit Hilfe von Mappingregeln zwischen der Web of Things (WoT) Thing Description und dem OPC UA Modell gelöst, die in der Spezifikation [OP24] definiert sind.

3.5 Packaging Machine Language

Neben dem genutzten Kommunikationsprotokoll wie OPC UA oder MQTT für den Datenaustausch spielt zur gegenseitigen flexiblen und interoperablen Prozesssteuerung auch die Standardisierung von Zustandsmaschinen eine wichtige Rolle. Mit der Standardisierung sind die in einem Netzwerk vorhandenen Systeme in der Lage, die von anderen Systemen bereitgestellten Funktionen zu interpretieren, um wiederum ihre Zustandsmaschinen korrekt auszuführen. Ein Standard dafür ist die Packaging Machine Language (PackML), die ursprünglich für Verarbeitungs- und Packetiermaschinen entwickelt wurde. [OM24]

In dem PackML-Standard, der 2008 in die ANSI/ISA TR88.00.02 adaptiert wurde, ist eine generische Zustandsmaschine beschrieben (s. Abbildung 3), die nicht nur für die zuvor genannten Maschinenarten, sondern auch für viele andere Systeme anwendbar ist. [PL]

Mit Hilfe der standardisierten Zustandsmaschine, die je nach Betriebsmodus eine Untermenge der zur Verfügung stehenden Zustände beinhalten kann, ist jedes System in der Lage, die Zustände eines anderen Interaktionspartners zu interpretieren. Zudem ist das Modell auch für OPC UA in der OPC 30050 „PackML - Packaging Control“ [OP20] spezifiziert, sodass Systeme sich gegenseitig über OPC UA flexibel in ihrem Prozessablauf beeinflussen können.

3.6 Verwaltungsschale

Neben der Nutzung der zuvor eingeführten Informationsmodelle AML, OPC UA und PackML werden auch proprietäre Informationsmodelle zur Beschreibung der Kommunikationsintegration eines Systems genutzt. Vor allem bei der Integration mehrere Subsysteme in ein übergeordnetes System führt diese Individualität zu manuellem Mehraufwand. Eine

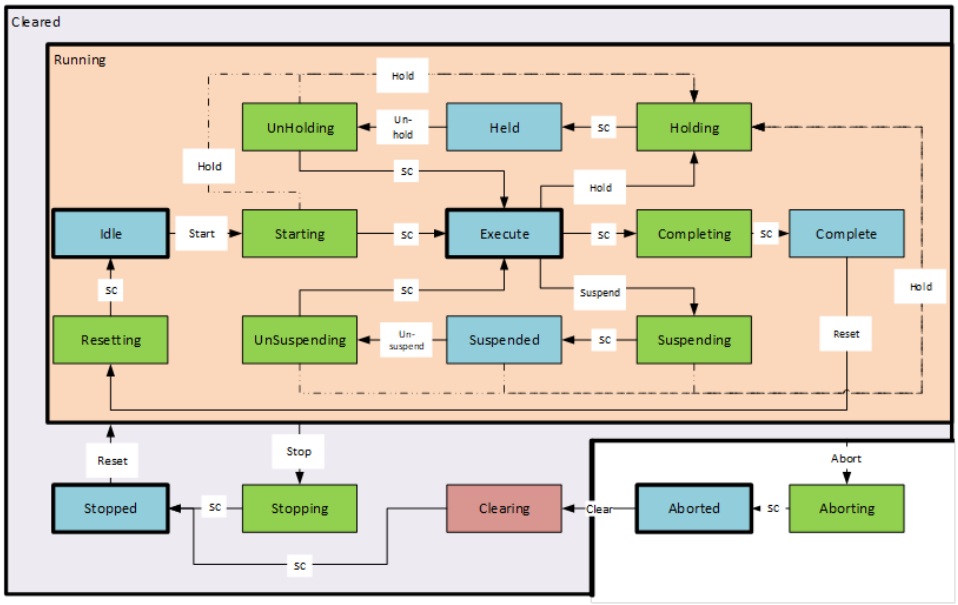


Abb. 3: PackML Zustandsmaschine [OP20]

Lösung zur Auflösung noch existenter proprietärer Formate und Schnittstellen bildet das Konzept der Verwaltungsschale (VWS, engl.: Asset Administration Shell, kurz AAS).

Die VWS sammelt die über den Lebenszyklus eines Systems entstehenden Daten in spezifizierten Teilmodellen und ermöglicht damit einen interoperablen Datenaustausch entlang der Wertschöpfungskette. Sie hat nicht den Anspruch, die proprietären Schnittstellen einzelner Softwareapplikationen obsolet werden zu lassen. Vielmehr bildet die VWS durch die Abstraktion proprietärer Softwareschnittstellen eine Art Middleware zwischen mehreren Softwareapplikationen, die jeweils ihre Informationen in einem spezifizierten Metamodell [In23a] in der VWS speichern und anderen Applikationen über die VWS-Schnittstelle [In23b] bereitstellen.

Bezüglich der Kommunikationsintegration bietet die VWS mehrere Optionen an, von denen zwei im Folgenden näher beschrieben werden. Die erste Option ist die Integration bestehender AML- oder OPC UA-Modelle als Detailmodell in die Teilmodelle der VWS, wie es in [Dr23] beschrieben ist (s. Abbildung 4). Bei dieser Herangehensweise werden bereits durchgeführte Arbeiten genutzt und nicht in einer anderen Modellierungssprache neu designt. Die Detailmodelle in der VWS werden dabei als Beschreibungsmittel mit weitgehend statischen Informationen für die Kommunikationsintegration des von der VWS abstrahierten Systems genutzt. Anhand dieser Informationen können wiederum automatisiert Clients

generiert werden, die sich direkt mit dem von der VWS abstrahierten System verbinden und den Datenaustausch direkt mit diesem ausführen.

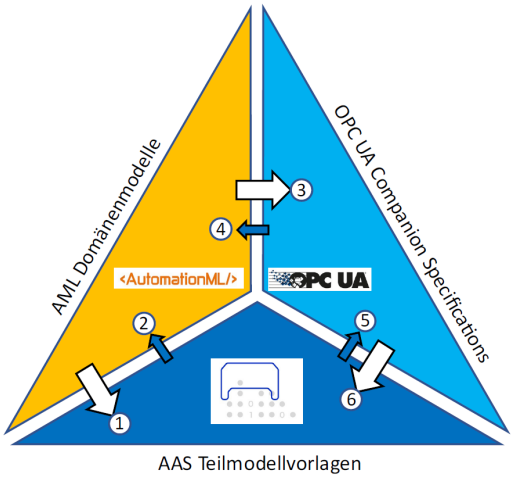


Abb. 4: Harmonisierung von AML, OPC UA und der VWS (engl.: AAS) [Dr23]

Die zweite Option ist die Nutzung der VWS als eigenständiges Kommunikationssystem zwischen den von den VWS abstrahierten Systemen. Dazu werden Aktualdaten aus dem abstrahierten System in der VWS benötigt. Das Teilmodell „Asset Interfaces Description“ (AID) [In24a], das sich an die WoT Thing Description anlehnt, ermöglicht nach dem aktuellen Stand die Beschreibung von Endpunkten und Metainformationen sowie Variablen, die das abstrahierte System bereitstellt, für die Protokolle MQTT, Modbus und HTTP. Sollen die Daten aus dem AID Teilmodell auf mehrere Teilmodelle verteilt werden, braucht es eine Mappingkonfiguration, die eine Zuweisung der Variablen aus dem AID Teilmodell zu den entsprechenden Teilmodellen ermöglicht. Hierfür steht das Teilmodell „Asset Interfaces Mapping Configuration“ (AIMC) [In24b] bereit.

Mit der zweiten Option können, basierend auf den Beschreibungen der beiden Teilmodelle AID und AIMC, automatisiert Clients generiert werden, die sich direkt mit dem abstrahierten System zum Datenaustausch verbinden. Der einzige Unterschied zwischen den beiden beschriebenen Optionen liegt darin, dass die Aktualwerte aus dem abstrahierten System in die VWS transferiert werden. Die Wahl der vorgestellten Optionen hängt stark von den Anforderungen des jeweiligen Anwendungsfalls ab.

4 Konzept zur Kommunikationsintegration verteilter Produktionssysteme

Im Folgenden wird ein Konzept vorgestellt, das vor allem die Nutzung der VWS-Teilmodelle aus Abschnitt 3.6 anstrebt. Das Konzept soll eine möglichst interoperable Kommunikati-

onsintegration dezentraler Systeme ermöglichen, die nicht ein und dasselbe Kommunikationsprotokoll verwenden. Ein einfacher Anwendungsfall soll zur Erläuterung des Konzepts dienen.

4.1 Anwendungsfall

Zur Veranschaulichung des Konzepts wird ein Demonstrator genutzt, der zwei Krane (Kran A und Kran B) enthält. Jeder der Krane kapselt über ein eigenes IIoT-Gerät seine internen Prozessabläufe und stellt über eine Schnittstelle externe Funktionen zur Prozessbeeinflussung zu dem jeweils anderen Kran bereit.

Die Zustandsmaschinen für die Prozessabläufe sind in PackML modelliert. Wie in Abschnitt 3.5 beschrieben, kann die generische Zustandsmaschine für verschiedene Betriebsmodi in der Menge der zur Verfügung stehenden Zustände variieren. Für das Konzept wird vorausgesetzt, dass die Betriebsmodi AUTOMATIC, MANUAL und MAINTENANCE durch eine eigene PackML Zustandsmaschine beschrieben sind.

Für jeden der Krane sind verschiedene Funktionen bzw. Services spezifiziert. Beide Krane bieten den Service „GoTo“ an, um den Kran an eine gewünschte Zielposition zu fahren. Zusätzlich bietet Kran A den Service an, einen „Init_Transport_Tandem“ durchzuführen. Hierzu müssen beide Krane zu einer vorgegebenen Zielposition fahren. Kran B wird dabei von Kran A angestoßen, mittels des „GoTo“ Services und unter Hinzunahme eines Sicherheitsabstands (delta) zur vorgegebenen Zielposition zu fahren. Die gleiche Funktion wird durch den Service „Init_Transport_Tandem“ auch in Kran A angestoßen.

Während Kran A mittels MQTT Daten austauscht und Befehle sowohl sendet als auch entgegennimmt, kommuniziert Kran B nur über OPC UA. Dadurch ist eine direkte Kommunikation zwischen den dezentral gesteuerten Kranen nicht möglich. Zur Lösung dieses Problems soll das im Folgenden vorgestellte Konzept beitragen. Abbildung 5 stellt anschaulich die Problemstellung dar. Die Darstellung der PackML Zustandsmaschinen stellt keinen Anspruch auf Vollständigkeit, da hier nur die unterschiedlichen Zustandsmaschinen je Betriebsmodi hervorgehoben werden sollen.

4.2 Konzept

Dieser Abschnitt geht auf zwei Konzepte ein, mit der die Problemstellung aus Abschnitt 4.1 gelöst werden kann. Das erste Konzept beruht auf der Nutzung eines IIoT-Gateways, dass zwischen den Kommunikationsprotokollen übersetzt (s. Abbildung 6). Mit der Implementierung eines solchen Gateways ist jedoch manueller Aufwand bzgl. Implementierung und Wartung verbunden, der mit steigender Anzahl an Kommunikationsprotokollen ebenso steigt.

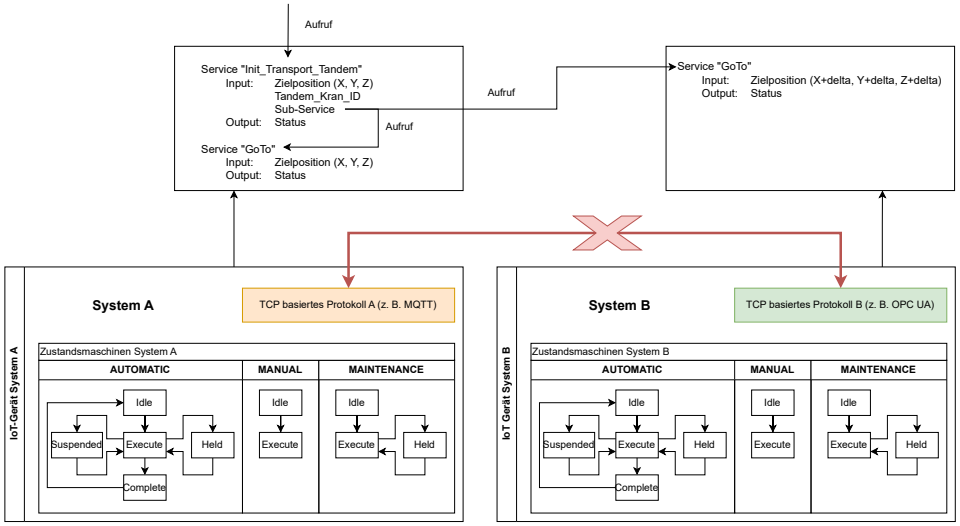


Abb. 5: Anwendungsfall - autonome Kommunikation zwischen Kränen mit unterschiedlichen Kommunikationsprotokollen

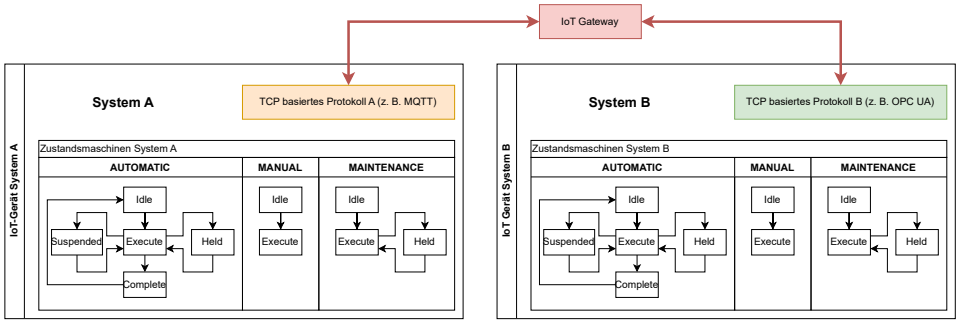


Abb. 6: Konzept 1 - IIoT-Gateway zur Verbindung verteilter Systeme

Ziel sollte es sein, auf bestehenden Standards aufzusetzen und diese zu kombinieren. Deshalb wurde ein zweites Konzept entwickelt, bei dem mit Hilfe der Verwaltungsschale die vorhandenen Kommunikationsprotokolle der verteilten Systeme, im Beispiel Kran A mit MQTT und Kran B mit OPC UA abstrahiert. Die Ansteuerung von Prozessen erfolgt über die von den Systemen bereitgestellten Services, die nach Aufruf intern gekapselte PackML Zustandsmaschinen starten. In diesem Konzept wird die VWS nicht nur als Beschreibungsmittel genutzt. Die VWS um eine eigene Businesslogik erweitert, sodass die von den abstrahierten Systemen bereitgestellten Services über die VWS freigegeben und aufgerufen werden können.

Für das Konzept werden grundsätzlich drei Teilmodelle in der VWS benötigt. Das erste Teilmodell AID dient der Anbindung des Systems an die VWS, siehe Abschnitt 3.6. Hierin werden die für den Anwendungsfall relevanten Signale und Operationen beschrieben. Das zweite Teilmodell dient zur Beschreibung von Services und wird im Forschungsprojekt ESCOM (Förderkennzeichen: 01MD22004G) entwickelt. In den Services werden die externen Funktionen aufgerufen, die das von der VWS abstrahierte System anbietet. Das Mapping der Daten von dem AID Teilmodell zu dem Service Teilmodell erfolgt über das AIMC Teilmodell. Das beschriebene Konzept ist in Abbildung 7 zu sehen.

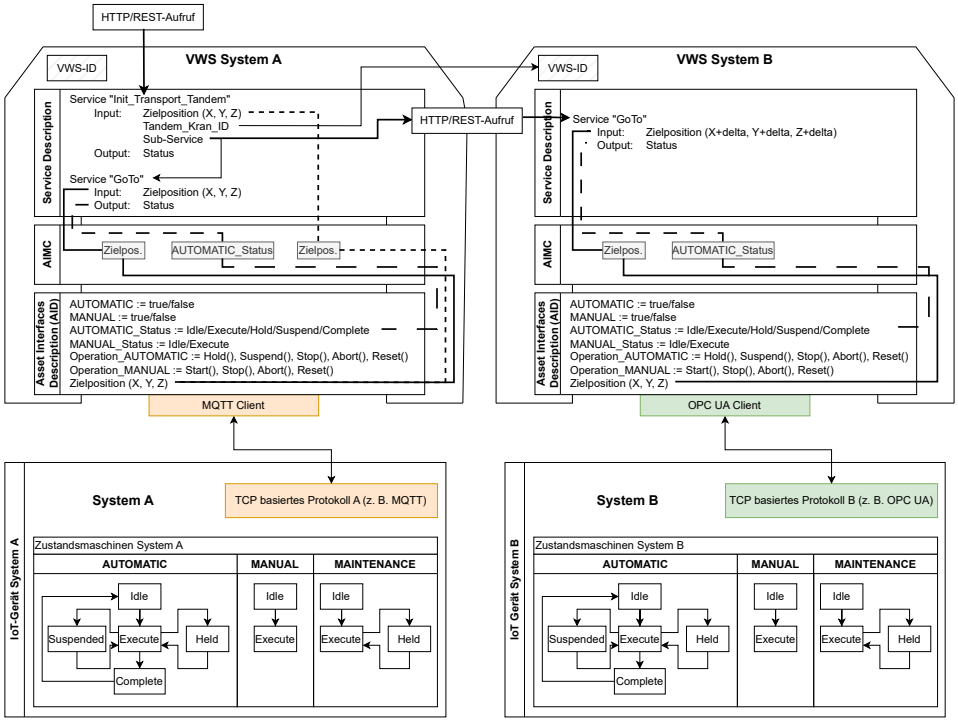


Abb. 7: Konzept 2 - Nutzung der Verwaltungsschale (VWS) zur Verbindung verteilter Systeme

Für die automatisierte Anbindung des Systems an die VWS wird die standardisierte Schnittstellenbeschreibung über das AID Teilmodell genutzt und darauf basierend automatisiert ein Client generiert, der die Kommunikation zwischen der VWS und dem von ihr abstrahierten System übernimmt. Zur Ausführung der Services im Service Teilmodell wird ebenfalls Businesslogik benötigt. Eine Option ist, die Businesslogik als eine Erweiterung des VWS-Servers zu implementieren, auf dem die VWS später gehostet wird. Eine zweite Option ist die Bereitstellung externer Webservices, die von Operationen in der VWS aufgerufen werden. Die dritte Option sieht das Speichern von Skripten in der VWS selbst vor, die dann von einem Deployment-Service gelesen und bei Aufruf ausgeführt werden können. Die

Serviceaufrufe und folglich die Steuerung von Kran B durch Kran A erfolgt über einen interoperablen HTTP/REST-Aufruf der VWS-API.

Der Vorteil dieses Konzepts liegt darin, dass durch die abstrahierte Schnittstelle beider Krane, diese trotz unterschiedlicher IIoT-Kommunikationsprotkolle interoperabel miteinander kommunizieren können. Dies funktioniert jedoch nur für Services, deren Funktionen nicht hochfrequentierte Daten austauschen müssen. Am Beispiel des Anwendungsfalls kann nur die Initialisierung der Tandemfahrt mit dem Anfahren der Startpunkte als Zielposition erfolgen. Das parallele Fahren zweier Krane kann nicht über die VWS realisiert werden, da die Krane sich im Bereich von Millisekunden andauernd synchronisieren müssen. In diesem Fall ist die VWS nicht geeignet.

5 Zusammenfassung und Ausblick

Die Individualität in der Beschreibung von Schnittstellen und Formaten für die Machine-to-Machine-Kommunikation stellt eine Hürde auf dem Weg zur Interoperabilität dar, die zum einen die einfachere Integration von dezentralen Systemen aber auch deren Datenaustausch zum Ziel hat. MQTT, bekannt für seine Einfachheit und Effizienz im Nachrichtentransport, ermöglicht eine nahtlose Kommunikation innerhalb und zwischen verschiedenen Systemen sowie mit externen Anwendungen, die MQTT unterstützen. Andererseits bietet OPC UA ein robustes Framework für einen sicheren, interoperablen und komplexen Datenaustausch, um semantische Interoperabilität auf der Steuerungsebene zu gewährleisten.

Der Beitrag hat die aktuellen Herausforderungen aufgezeigt die bezüglich der horizontalen Kommunikation zwischen Systemen existieren. Es wurden Modellierungssprachen und Informationsmodelle aufgeführt, die zur Lösung der Herausforderungen beitragen. Zu den Herausforderungen zählen unter anderem die steigende Modularität und Flexibilität von Produktionssystemen.

Es wurde ein Konzept entwickelt, das zum einen ein Beschreibungsmittel für eine standardisierte und konfigurierbare Kommunikationsintegration von IIoT-Geräten bietet und zum anderen die Abstraktion unterschiedlicher Kommunikationsprotokolle ermöglicht. Hauptaugenmerk wurde hierbei auf die Verwendung der Verwaltungsschale (VWS) gelegt. Sie kann die vorliegenden Kommunikationsprotokolle der Systeme abstrahieren und somit eine flexible und interoperable Kommunikationsintegration und folglich einen ereignisgesteuerten Prozess von miteinander interagierenden Systemen erreichen. Ausgenommen davon sind jedoch die Übermittlung hochfrequenter Daten zwischen zwei oder mehreren Systemen, da dies mit Hilfe der VWS nicht skalierbar ist.

Für die Kommunikationsintegration eines Systems wurde neben der VWS mit dem „Asset Interfaces Description“ (AID) und „Asset Interfaces Mapping Configuration“ (AIMC) Teilmodell der Standard PackML genutzt. Daneben gibt es jedoch noch den Standard des „Module Type Package“ (MTP) [VD22], für das ebenfalls ein Teilmodell in der

Verwaltungsschale spezifiziert wurde und das in dieser Arbeit nicht betrachtet wurde. Mit dem MTP sollen Prozessmodule nach dem „Plug&Play“ Prinzip flexibel in Anlagen ein- und ausgegliedert werden. Das MTP bietet einen größeren standardisierten Beschreibungsumfang als PackML und sollte in weiteren Forschungsarbeiten evaluiert werden.

Weiterhin beschränkt sich die Kommunikationsanbindung des Systems an die VWS über das AID Teilmodell bisher auf die Protokolle MQTT, HTTP/REST und Modbus. Das Protokoll OPC UA ist noch nicht integriert. Es kann eine proprietäre Erweiterung vorgenommen werden, die der Web of Things Thing Description folgt. Jedoch besteht hier die Gefahr einer proprietären Implementierung, die es zu vermeiden gilt. Neben dem AID Teilmodell gibt es jedoch auch noch das „OPC UA Server Datasheet“ Teilmodell, das sich zur Zeit in Entwicklung befindet. Eine zukünftige Evaluierung wird zeigen, ob die Kombination beider Teilmodelle zukünftig auch die Integration von OPC UA in die VWS ermöglicht.

6 Danksagung

Die Autoren möchten sich beim Bundesministerium für Wirtschaft und Klimaschutz (BMWK) für die Förderung des Projektes ESCOM, FKZ. 01MD22004G, bedanken, durch das wichtige Ideen zur Bereitstellung von Edge-Cloud-Services entwickelt und mit den Projektpartnern umgesetzt wurden.

Literaturverzeichnis

[Di20] Diedrich, C.; Belyaev, A.; Schröder, T.; Werner, T.: Neue Anforderungen an die Interoperabilität aus den Szenarien von Smart Manufacturing. In (VDI Wissensforum GmbH, Hrsg.). 1. Aufl., Bd. 2375, VDI-Berichte, VDI Verlag, Düsseldorf, S. 943–956, 2020, ISBN: 9783181023754, DOI: 10.51202/9783181023754-943.

[Dr21] Drath, R.: AutomationML: The Industrial Cookbook. Walter de Gruyter GmbH, Berlin / Boston, 2021, ISBN: 978-3-11-074597-9.

[Dr23] Drath, R.; Mosch, C.; Hoppe, S.; Faath, A.; Barnstedt, E.; Fiebiger, B.; Schlögl, W.: Diskussionspapier – Interoperabilität mit der Verwaltungsschale, OPC UA und AutomationML - Zielbild und Handlungsempfehlungen für industrielle Interoperabilität, Diskussionspapier, 2023.

[IE22] IEC, Hrsg.: IEC 62714-5 - Engineering data exchange format for use in industrial automation systems engineering - Automation Markup Language, Part 5: Communication, 2022.

[In23a] Industrial Digital Twin Association, Hrsg.: Asset Administration Shell Specification - Part 1: Metamodel Schema, 2023, URL: https://industrialdigitaltwin.org/en/content-hub/aasspecifications/specification-of-the-asset-administration-shell-part-2-application-programming-interfaces-idta-number-01002-3-0-2.

[In23b] Industrial Digital Twin Association, Hrsg.: Asset Administration Shell Specification - Part 2: Application Programming Interfaces, 2023, URL: https://industrialdigitaltwin.org/en/content-hub/aasspecifications/specification-of-the-asset-administration-shell-part-1-metamodel-idta-number-01001-3-0-1.

[In24a] Industrial Digital Twin Association, Hrsg.: IDTA 02017-1-0 - Specification - Asset Interfaces Description, 2024, URL: https://industrialdigitaltwin.org/wp-content/uploads/2024/01/IDTA-02017-1-0_Submodel_Asset-Interfaces-Description.pdf.

[In24b] Industrial Digital Twin Association, Hrsg.: IDTA 02027-1-0 - Specification - Asset Interfaces Mapping Configuration, 2024, URL: https://industrialdigitaltwin.org/wp-content/uploads/2024/06/IDTA-02027-1-0_Submodel_AssetInterfacesMappingConfiguration.pdf.

[Lu20] Lueder, A.: Flexibility in Production Systems by Exploiting Cyberphysical Systems. In. IEEE Press, S. 81–85, 2020, DOI: 10.1109/MC.2019.2949107, URL: https://ieeexplore.ieee.org/document/8960944.

[OM24] OMAC: What is PackML?, Accessed: (13.08.2024), 2024, URL: https://www.omac.org/packml.

[OP20] OPC Foundation, Hrsg.: OPC 30050 PackML - Packaging Control, 2020, URL: https://opcfoundation.org/developer-tools/documents/view/209.

[OP22] OPC Foundation, Hrsg.: OPC 10000-100: Devices, 2022, URL: https://opcfoundation.org/developer-tools/documents/view/197.

[OP24] OPC Foundation, Hrsg.: OPC 10100-1 WOT Connectivity - API Definition, 2024, URL: https://opcfoundation.org/developer-tools/documents/view/345.

[PL] PLCopen: The Mapping of the OMAC PackML State Diagram to IEC 61131-3, URL: https://plcopen.org/sites/default/files/downloads/mapping_of_omac_state_diagram_v8.pdf.

[RZ22] Riedl, M.; Zipper, H.: Selbstorganisation von Softwarekomponenten für produktive Aufgaben. In (VDI Verlag GmbH, Hrsg.). 1. Aufl., Bd. 2399, VDI-Berichte, VDI Verlag, Düsseldorf, S. 319–328, 2022, ISBN: 9783968622309.

[VD22] VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik AND VDI/VDE Society Measurement and Automation AND VDI Verein Deutscher Ingenieure e.V. AND VDI - The Association of German Engineers: VDI/VDE/NAMUR 2658 Blatt 1 Automatisierungstechnisches Engineering modularer Anlagen in der Prozessindustrie - Allgemeines Konzept und Schnittstellen, Düsseldorf, 2022.

[We23] Weyrich, M.: Industrielle Automatisierungs- und Informationstechnik - IT-Architekturen, Kommunikation und Software zur Systemgestaltung. Springer Vieweg Berlin, Heidelberg, 2023, ISBN: 978-3-662-56354-0.

Security-Event-Logging am Beispiel von PROFINET

Karl-Heinz Niemann [1], Julian Göppert [2], Andreas Walz[3] und Dominik Ziegler [4]

Abstract: Dieses Dokument beschreibt die Notwendigkeit des Loggings von Security-relevanten Events am Beispiel des PROFINET Security Konzeptes. Nach einer Einführung in das Thema werden auf Basis verschiedener Normen die Anforderungen an ein Event-Logging erarbeitet. Abschließend wird das für PROFINET spezifizierte Vorgehensweise beschrieben.

Keywords: PROFINET, OT-Security, Event-Logging, Ereignisprotokollierung, Syslog, IEC 62443-4-2

1 Einleitung

Die Ereignisprotokollierung (engl. Event-Logging) erleichtert die Erkennung und Analyse von Anomalien in IT- und OT-Systemen. Sie ermöglicht die kontinuierliche Überwachung und Dokumentation (security-) relevanter Ereignisse, wodurch potenzielle Bedrohungen identifiziert und entsprechende Gegenmaßnahmen eingeleitet werden können. Im Zuge der Erweiterung des PROFINET-Protokolls um Sicherheitsfunktionalitäten, ist es daher sinnvoll das Event-Logging sowohl in die Security-Spezifikation für PROFINET aufzunehmen.

Eine Ereignisprotokollierung war bereits bisher Bestandteil des PROFINET-Protokolls sowie von Automatisierungssystemen auf Basis von PROFINET. In der Vergangenheit war diese Protokollierung dazu gedacht, Informationen über Fehler und Störungen der PROFINET-Kommunikation der Komponenten sowie von prozessbezogenen Ereignissen zu erkennen und zu speichern. In typischen Anwendungsfällen wurden die Protokolle genutzt, um Störungen bei der Inbetriebnahme und während des Betriebs eines Automatisierungssystems zu identifizieren (Event-Logging).

[1] Hochschule Hannover, Institut für Sensorik und Automation (ISA), Ricklinger Stadtweg 120, 30459 Hannover, Karl-Heinz.Niemann@Hs-Hannover.de, https://orcid.org/0000-0001-8931-6789
[2] Hochschule Offenburg, Institut für verlässliche Embedded Systems und Kommunikationselektronik (ivESK), Badstraße 24, 77652 Offenburg, julian.goeppert@hs-offenburg.de, https://orcid.org/0000-0003-0679-3841
[3] Siemens AG, Digital Industries, Factory Automation Architecture, Siemenspromenade 1, 91058 Erlangen, andreas-christoph.walz@siemens.com
[4] Siemens Aktiengesellschaft Österreich, Digital Industries, Factory Automation, Strassganger Str. 315, 8054 Graz, Österreich, dominik.ziegler@siemens.com. https://orcid.org/0000-0002-5930-8216

Mit der Einführung der zusätzlichen Security-Funktionalitäten von PROFINET [NIE2023] hat sich der Anwendungsbereich der Ereignisprotokollierung erweitert und es müssen eine zusätzliche Security-Event-Logging-Funktionalität bereitgestellt werden Diese Funktionen werden über eine zusätzliche SEV-Klasse in der Software realisiert. Mögliche Security-Events sind:

- Sicherheitsrelevante Ereignisse, z. B. fehlgeschlagene Authentifizierung (aufgrund ungültiger oder abgelaufener Zertifikate), unbefugter Zugriff, Aufruf der Wiederherstellungsfunktionen, Zurücksetzen auf Werkseinstellungen) müssen erkannt und dem Ereignisprotokoll hinzugefügt werden.
- Sicherheitsrelevante Ereignisse müssen an übergeordnete Überwachungssysteme übertragen werden und dort müssen möglicherweise Alarme auf der Grundlage vordefinierter Regelsätze erzeugt werden.
- Sicherheitsrelevante Ereignisse müssen für eine forensische Analyse gespeichert werden. Dies kann zum einen direkt im Gerät erfolgen. Hier besteht das Problem, dass der begrenzte Speicherplatz im Gerät ein Überschreiben alter Einträge (Rollover) erfordert. Zum anderen kann die Speicherung in einem übergeordneten Überwachungssystem erfolgen. Hier ist der Speicherplatz in der Regel kein Problem, da hier Massenspeicher zur Verfügung stehen. Eine forensische Analyse wird nach der Erkennung eines Angriffs oder nach der Erkennung eines Systemausfalls, der möglicherweise durch einen Angriff verursacht wurde, durchgeführt. Ziel ist es, die Ursache des Angriffs oder des Ausfalls zu finden. Die forensische Analyse fällt nicht in den Anwendungsbereich dieses Dokuments und wird daher nicht weiter behandelt.

In den folgenden Kapiteln dieses Beitrages wird die Ereignisprotokollierung gemäß der spezifizierten Sicherheitserweiterungen für PROFINET gemäß der aktuellen PROFINET-Spezifikation (Version 2.4. MU5) [PNO2024a], [PNO2024b] behandelt.

2 Grundlagen des Event-Loggings

In diesem Abschnitt werden in Kapitel 2.1 zunächst die allgemeinen Grundlagen des Event-Loggings beschrieben. Danach folgt in Kapitel 2.2 die Funktionsbeschreibung eines solchen Systems. In Kapitel 2.3 folgt dann die Beschreibung eines zugeordneten Security Information and Event Management Systems (SIEM).

2.1 Allgemeines Vorgehen beim Event-Logging

Logmeldungen von einem Computersystem werden typischerweise als Reaktion auf Ereignisse in diesem System erzeugt. Beispiele sind An- und Abmeldungen durch Benutzer, die Änderung der Konfiguration einer Firewall oder Unterbrechungen von Netzwerkverbindungen. Nach [CHU2013] können Logs in folgende Kategorien eingeteilt werden:

    - **Informativ:** Meldungen dieses Typs dienen dazu, Benutzer und Administratoren darüber zu informieren, dass etwas Unbedenkliches passiert ist. Ein System erzeugt beispielsweise Meldungen, wenn es neu gestartet wird. Dabei ist jedoch Vorsicht geboten. Wenn ein Neustart beispielsweise außerhalb der normalen Wartungs- oder Geschäftszeiten erfolgt, könnten auch informative Log-Meldungen Grund zur Sorge bieten.

- **Debug**: Debug-Meldungen werden in der Regel von Softwaresystemen erzeugt, um Softwareentwicklerinnen / Softwareentwicklern bei der Fehlersuche und der Identifizierung von Problemen mit dem laufenden Anwendungscode zu unterstützen.
- **Warnung**: Warnmeldungen beziehen sich auf Situationen, in denen Ressourcen fehlen oder für ein System benötigt werden, deren Fehlen aber keine Auswirkungen auf den Systembetrieb hat. Wenn zum Beispiel einem Programm nicht die richtige Anzahl von Befehlszeilenargumenten übergeben wird, es aber trotzdem ohne sie laufen kann, kann das Programm dies als Warnung für den Benutzer oder Betreiber protokollieren.
- **Fehler**: Fehlerprotokollmeldungen werden verwendet, um Fehler weiterzugeben, die auf verschiedenen Ebenen in einem Computersystem auftreten. Zum Beispiel kann ein Betriebssystem ein Fehlerprotokoll erstellen, wenn es nicht in der Lage ist Puffer auf die Festplatte zu schreiben.
- **Alarm**: Ein Alarm soll darauf hinweisen, dass etwas Besonderes passiert ist. Ein Intrusion Prevention System (IPS) kann z. B. den Datenverkehr an kritischen Stellen des Netzwerks überwachen. Wenn das IPS auf eine Verbindung stößt, die möglicherweise bösartig ist, kann es eine beliebige Anzahl von vorkonfigurierten Maßnahmen ergreifen und auch einen Alarm erzeugen. Auf PROFINET bezogen könnte ein Alarm erzeugt werden, wenn z. B. die Ausgänge einer Ausgabebaugruppe in den sicheren Zustand gegangen sind oder wenn die Baugruppe in den Auslieferungszustand zurückgesetzt wurde.

Die Entscheidung und die gegebenenfalls ergriffenen Maßnahmen werden in einem Protokoll festgehalten, das von den Geräten aufgrund von Ereignissen erstellt wird. Bei den Geräten kann es sich um Computer, Speicherprogrammierbare Steuerungen (SPSen), Remote-IO-Baugruppen, Switches, Firewalls oder Router handeln. Ein solches Gerät kann einen Protokolleintrag in einem lokalen Puffer erstellen, wenn es entscheidet, dass ein Ereignis protokolliert werden sollte. Normalerweise puffert (speichert) das Gerät das Ereignis zumindest für einen bestimmten Zeitraum.

Nach [CHU2013] gibt es verschiedene Gründe für die Erstellung von Protokollen:

- **Security-Event-Logging**: Ziel ist die Dokumentation von Angriffen, Malware-Infektionen, Datendiebstahl oder anderen Sicherheitsproblemen. Ein Beispiel ist die Aufzeichnung von Unterbrechungen von Kommunikationsverbindungen oder das Zurücksetzen von Baugruppen.

- **Operational-Logging** dient der Bereitstellung von Informationen für Systembetreiber. Beispiele sind die Nutzung der Speicherkapazität von Servern oder die erwartete Lebensdauer von Ventilatoren.
- **Compliance-Logging:** Diese Art der Protokollierung überschneidet sich mit der Security-Event-Logging. Sie wird durchgeführt, um die gesetzlichen Vorschriften einzuhalten, z. B. für kritische Infrastrukturen.
- **Debug-Logging:** Wird durchgeführt, um Softwareentwicklern Informationen für die Fehlersuche zur Verfügung zu stellen.

Nachdem verschiedene Logging-Aspekte beleuchtet wurden, konzentriert sich der Rest dieses Beitrages auf das Security Event-Logging.

2.2 Funktionsbeschreibung eines Event-Logging-Systems

Das Ziel eines Event-Logging-Systems ist die Erfassung von Ereignissen, die im System bzw. in den Komponenten eines Systems auftreten. Ein Ereignis ist ein einzelnes Vorkommnis in einer Umgebung, das in der Regel eine versuchte oder erfolgte Zustandsänderung beinhaltet. Ein Ereignis umfasst in der Regel einen Zeitbegriff, das Auftreten und alle Details, die sich explizit darauf oder die Umgebung beziehen, und helfen können, die Ursachen oder Auswirkungen zu erklären oder zu verstehen [MIT2010].

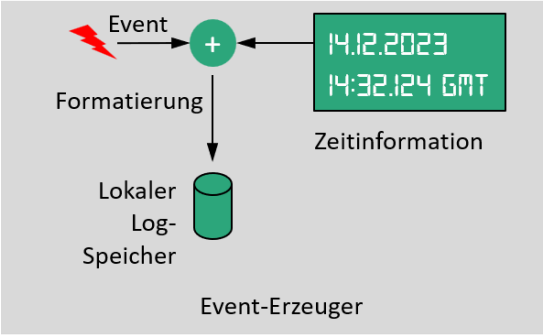


Abbildung 1: Event Erzeugung

Abbildung 1 zeigt die allgemeine Funktion eines Event-Erzeugers. Nehmen wir an, es handelt sich um eine speicherprogrammierbare Steuerung (SPS). Das aufzuzeichnende Ereignis könnte z. B. sein, dass die Spannung der Backup-Batterie unter einen bestimmten Schwellenwert gefallen ist (Betriebsprotokoll) oder dass eine Kommunikationsverbindung unterbrochen wurde (Sicherheitsprotokoll). Wenn das Ereignis eintritt, wird es zusammen mit einer Datums- und Zeitinformation in einem lokalen Speicher abgelegt. Das Speicherformat der Ereignisinformationen kann proprietär

sein. Vorzugsweise wird ein allgemein verwendetes Format verwendet. Ein typischer Eintrag kann z. B. aus den folgenden Informationen bestehen [CHU2013]:

- Zeitstempel, z. B. Datum/Uhrzeit (und idealerweise auch eine Zeitzone) - Art des Protokolleintrags
- System, welches den Eintrag erzeugt hat.
- Anwendung oder Komponente, die den Eintrag erzeugt hat.
- Fehlerbeschreibung / Ereignisbeschreibung.
- Schweregrad, Priorität oder Wichtigkeit einer Protokollmeldung.
- Bei Einträgen, die sich in irgendeiner Weise auf Benutzeraktivitäten beziehen, wird auch der Benutzername aufgezeichnet.

Die Erfassung, die Zeitstempelung, die Formatierung und die Speicherung eines Protokolleintrags ist der erste Schritt, um den Ereignisprotokollierungsprozess zu starten. Wie in Abbildung 1 zu sehen ist, hat die lokale Speicherung der Ereignisse jedoch einige Nachteile:

1. Die Größe des lokalen Speichers ist begrenzt, somit auch die Anzahl der maximal speicherbaren Einträge.
2. Früher oder später ist der lokale Speicher erschöpft. Es sind dann Strategien zum Löschen von Einträgen bzw. zum Überschreiben alter Einträge erforderlich. Dies kann von Angreifern ausgenutzt werden, um kritische Einträge bewusst mit unkritischen Einträgen zu überschreiben.
3. Falls Flash-Speicher verwendet wird, können häufige Schreibzugriffe auf den Flash-Speicher zu einer vorzeitigen Alterung des Speicherbauteils führen.
4. Der Zugriff auf die lokal gespeicherten Log-Informationen ist möglicherweise nur vor Ort möglich.
5. Ein Echtzeit-Monitoring und eine proaktive Reaktion auf Ereignisse sind daher nicht möglich.

Aufgrund dieser Einschränkungen sollte in Erwägung gezogen werden, Log-Meldungen an ein übergeordnetes System zur weiteren Speicherung und Verarbeitung weiterzuleiten. Abbildung 2 zeigt, wie eine zentralisierte Protokollierung und Ereignisverarbeitung aussehen könnten.

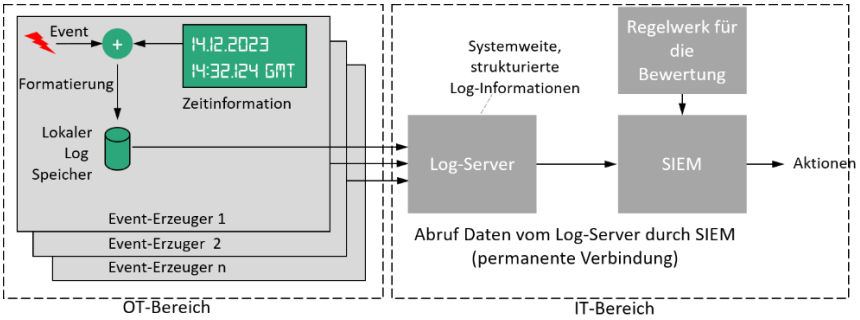


Abbildung 2: Zentralisierte Protokollierung und Ereignisverarbeitung

Es wird angenommen, dass es in der Anlage im OT-Bereich mehrere Event-Erzeuger gibt (1 bis n). Die Event-Erzeuger generieren aus den Protokolleinträgen im Puffer so genannte Ereignismeldungen. Diese Meldungen werden relativ häufig z. B. im so genannten Syslog-Format [RFC_5424] [RFC_6587] [RFC_5848] übertragen und gespeichert. Die Ereignismeldungen aller Event-Erzeuger werden an einen Log-/Syslog-Server übermittelt. Der Übermittlungsmechanismus kann entweder ein Push-Mechanismus sein (Absender sendet die Daten) oder ein Pull-Mechanismus (Log-Server fragt die die Geräte ab und holt die Informationen ab). In dem in Abbildung 2 gezeigten Fall verwenden alle Event-Erzeuger die gleiche Formatierung der Protokolleinträge. So kann der Log-Server die Ereignisse auf der Basis strukturierter Informationen systemweit einheitlich speichern. Da das IT-Management in der Regel solche Log-Server für die IT-Ausrüstung (Server, Workstations, Switches, Firewalls) betreibt, könnte man in Erwägung ziehen, denselben Log-Server auch für die in der OT-Domäne erzeugten Logs zu verwenden. Ein dedizierter Log-Server für die OT-Domäne ist ebenfalls möglich. In einem PROFINET-System können Event-Erzeuger sowohl PROFINET Devices als auch PROFIONET Controller sein.

2.3 Funktionsbeschreibung Security Information and Event Management System (SIEM)

Alle Protokollinformationen an einem einzigen Ort, dem Log-Server, zu speichern stellt zunächst sicher, dass alle Informationen an einem Ort verfügbar sind. Aufgrund der Größe der Produktionsanlagen und der Menge der Protokolleinträge ist eine manuelle Analyse der Daten jedoch nicht möglich. Für diese Aufgabe wird ein **S**ecurity **I**nformation and **E**vent **M**anagement System (SIEM) eingesetzt.

SIEM-Systeme unterstützen das Sammeln und die Analyse von Events. Diese Analyse kann automatisiert, z.B. durch Schwellenwertüberwachung, oder operatorgesteuert, z. B. durch Ereignis-Dashboards mit Navigationsfunktionen, durchgeführt werden. Die Analyse von Syslog-Inhalten erfordert, dass SIEM-Produkte in Bezug auf IETF-definierte [RFC_5424] oder private Namensräume und deren Schlüsselwort/Wert-Paare

konfigurierbar sind (wie man Elemente findet und wie man sie parst). Um das Mining zu unterstützen, erlauben SIEM-Produkte oft die Verwendung von Bewertungsregeln, die von weiteren Parteien bereitgestellt werden, z. B. MITRE ATT&CK [MIT2023] für Enterprise/Mobile/Industrial Control Systems. Für eine SIEM-Produktübersicht siehe den Gartner Magic Quadrant for SIEM [GAR2023]. Das SIEM kann mit einem bestimmten Regelwerk programmiert werden, um die Log-Einträge zu analysieren und aus den Ereignissammlungen Alarme, Statusübersichten, Key-Performance-Indicators, Dashboards und Aktionen, z. B. Arbeitsaufträge, abzuleiten.

3 Anforderungen aus der Security-Normung an das Event-Logging

Verschiedene Normen für den OT- und IT-Bereich definieren Anforderungen an das Event-Logging. Die folgenden Unterkapitel geben einen kurzen Überblick über die wichtigsten Aspekte dieser Normen.

3.1 Anforderungen der ISO 27001 und ISO 27002 an das Event-Logging

Die [DIN_EN_ISO/IEC_27001] umreißt die Anforderungen für die Einrichtung eines Informationssicherheitsmanagementsystems (ISMS), das die Definition eines systematischen Ansatzes für die Verwaltung sensibler Informationen, die Bewertung von Risiken, die Implementierung von Kontrollen sowie die Überwachung und Verbesserung der Sicherheit von Informationswerten umfasst. Der Anhang der Norm definiert in Bezug auf die Protokollierung in Tabelle A1:

- 8.15: Protokollierung: „Protokolle, die Aktivitäten, Ausnahmen, Fehler und andere relevante Ereignisse aufzeichnen, müssen erstellt, gespeichert, geschützt und analysiert werden.“
- 8.16: Netzwerke: Netzwerke, Systeme und Anwendungen müssen auf anormales Verhalten überwacht und geeignete Maßnahmen müssen ergriffen werden, um potentielle Informationssicherheitsvorfälle zu bewerten.
- 8.17: Uhrzeitsynchronisation: Die Uhren der von der Organisation verwendeten Informationsverarbeitungssysteme müssen mit zugelassenen Zeitquellen synchronisiert werden.

In der [DIN_EN_ISO_IEC_27002] finden sich die zugehörigen Umsetzungsempfehlungen.

3.2 Anforderungen des BSI Grundschutzes an das Event-Logging

Das Dokument [BSI2021] bildet die Elemente der [DIN_EN_ISO/IEC_27001] auf die Module des BSI-Grundschutzes ab. Es sind in Bezug auf das Event-Logging die folgenden

BSI-Grundschutzmodule zu betrachten: OPS.1.1.5 Protokollierung, ORP.4.A16 Richtlinien für die Zugriffs- und Zugangskontrolle, OPS.1.1.5.A4 Zeitsynchronisation der IT-Systeme, DER.1.A11 Verwendung einer zentralisierten Protokollierungsinfrastruktur.

3.3 Anforderungen der IEC 62443-4-2 an das Security-Event-Logging

Neben anderen Anforderungen an Automatisierungskomponenten definiert die [DIN_EN_IEC_62443-4-2] einige Anforderungen an das Security-Event-Logging. Die Anforderungen hängen von dem angestrebten Security Level (SL) ab, das von dem Gerät erreicht werden soll. Der Security-Level reicht von SL-1 (geringe Anforderungen) bis SL-4 (sehr hohe Anforderungen). In diesem Beitrag wird von der Sicherheitsstufe SL-2 ausgegangen da dies sowohl vom VDMA [FUH2016] als auch von der ISA [ISA2024] für Standard-Produktionsanlagen empfohlen wird. Anforderungen in Bezug auf das Logging für ein System mit SL-2 sind:

- CR 2.8: Die Komponenten müssen die Fähigkeit haben, Ereignisdatensätze zur IT-Sicherheit in den folgenden Kategorien zu erzeugen:
    a. Zugriffskontrolle;
    b. fehlerhafte Anfragen;
    c. Ereignisse im Automatisierungssystem;
    d. Ereignisse bei der Sicherung und Wiederherstellung;
    e. Konfigurationsänderungen; und
    f. Ereignisse, die aus dem Ereignisprotokoll hervorgehen.

  Die einzelnen Ereignisdatensätze müssen folgende Angaben enthalten:

    a. Zeitstempel;
    b. Quelle (Gerät, Softwareprozess oder Konto des menschlichen Nutzers, on dem das Ereignis herrührt);
    c. Kategorie;
    d. Art;
    e. Ereigniskennung; und
    f. Ergebnis des Ereignisses.CR 2.9: Audit-Speicherplatz. Zuteilung von Speicherkapazität für Auditaufzeichnungen (bis das Ereignis in einem zentralen System verfügbar ist); Schutz vor Ausfällen der Komponente bei Erreichen der Kapazität
- CR 2.9: Speicherkapazität für Ereignisdatensätze
- CR 2.10 - Verhalten bei Verarbeitungsfehlern von Ereignisdaten
- CR 2.11: Zeitstempel. Erstellung von Zeitstempeln für Auditaufzeichnungen.
- CR 3.9: Schutz von Prüfinformation. [EU_3602]
- CR 6.1: Zugriff auf Auditprotokolle. Schreibgeschützter Zugriff auf Auditprotokolle für autorisierte Personen und/oder automatisierte Tools

3.4 Weitere Normen die Anforderungen an das Event-Logging definieren

Neben den bisher genannten Normen existieren noch weitere Normen und Richtlinien, die sich mit dem Security-Event-Logging befassen. Dies sind:

- [NIST_800-92] Guide to Computer Security Log Management
- [NIST_800-137] Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations
- [EU_3602] European Commission Information System Security Policy. Standard on Logging and Monitoring.
- [EUR_2022_272] Regulation of the European Parliament and of the Council on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020. Cyber Resilience act.
- [NAM2023]: Angriffserkennung nach IT-Sicherheitsgesetz 2.0

Aus Platzgründen wird auf diese Normen nicht weiter eingegangen, da eine vollständige Beschreibung den Rahmen dieses Beitrags sprengen würde.

4 Das Konzept für das PROFINET-Security-Event-Logging

Nach einer Analyse der o.g. Anforderungen wurde für das PROFINET-Security-Konzept das Security-Event-Logging spezifiziert. Hierbei wurde insbesondere auf die Anforderungen der [DIN_EN_IEC_62443-4-2] geachtet, da diese für Anwendungen im Bereich der industriellen Automatisierungstechnik die höchste Relevanz aufweist.

Abbildung 3 zeigt das grundlegende Konzept des PROFINET-Security-Event-Loggings. PROFINET-Komponenten (PN Controller und PN Devices) sind in der Lage Events zu erzeugen. Diese Information kann entweder über eine PN-interne Schnittstelle an ein PROFINET-Event-Relay übertragen. Dieses Event-Relay ist in der Lage zeitgestempelte Syslog-Events zu generieren und diese zeitgestempelt an einen Syslog-Server zu übertragen bzw. für diesen bereitzustellen.

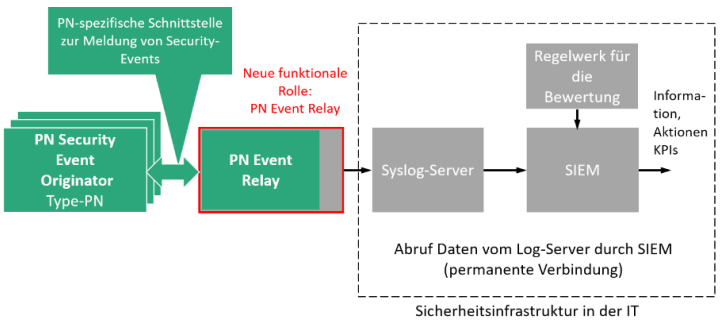


Abbildung 3: Struktur des PROFINET-Logging-Konzeptes

Alternativ kann die PROFINET-Komponente auch über einen eigenen Syslog-Server nutzen, sofern dieser vorhanden ist.

Die optionale Zwischenschaltung des PN-Event-Relays erfolgt in der Regel, weil dadurch nicht jede PN-Komponente über eine Verbindung zu einem Syslog-Server verfügen muss. Die grau dargestellten Bereiche werden als Bestandteil der IT-Sicherheitsinfrastruktur angesehen und nicht weiter erläutert. Sofern eine PN-Komponente schon über einen eigenen Syslog-Client verfügt, kann dieser direkt an den Syslog-Server angebunden werden.

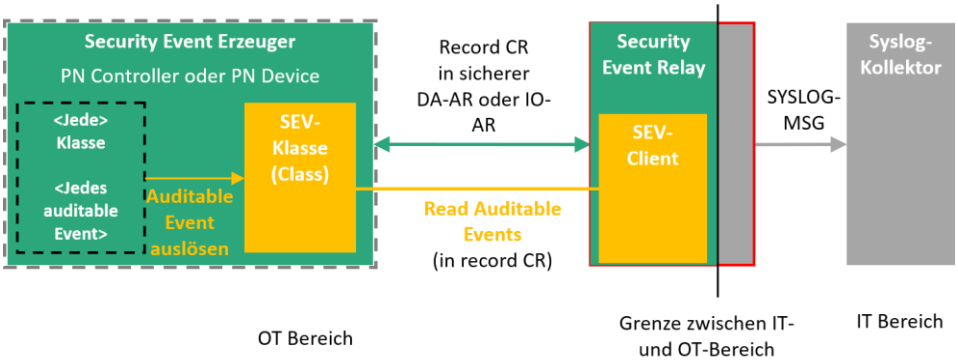


Abbildung 4: Detaillierung des PROFINET-Logging-Konzeptes

Abbildung 4 zeigt eine weitere Detaillierung des PROFINET-Security-Logging-Konzeptes. Hier ist nun zu erkennen, dass die Übertragung der Event- /Log-Information über die Record-Communication-Relation (CR) des PROFINET-Protokolls erfolgt. Da die rechte Seite des Security-Event-Relays an den IT-Bereich angebunden ist, kann von dort – bei Bedarf – auch die synchronisierte (Welt-)Uhrzeit bezogen werden.

5 Zusammenfassung und weiteres Vorgehen

Der Beitrag zeigt, dass für das Security-Event-Handling auf bereits bestehenden Mechanismen aus der IT zurückgegriffen werden kann. Es ist also im Wesentlichen das PROFINET-Protokoll um die Fähigkeit zu ergänzen, Security-Events in bestehende Systeme einzuspeisen. Die Notwendigkeit der Erzeugung von Security-Events wurde durch die Analyse verschiedener Standards nachgewiesen. Das PROFINET-Security-Konzept ist in der aktuellen PROFINET-Spezifikation (Version 2.4. MU5) [PNO2024a], [PNO2024b] dokumentiert. Eine Beschreibung der Security Eventing Class findet sich in [PNO2024b] ab Kapitel 6.3.21.9. Mit dieser Spezifikation ist die Basis für die Implementierung der PROFINET-Protokollstacks gelegt, die Security –einschließlich des Event-Loggings – unterstützen werden.

Danksagung

Die Autoren danken den Mitgliedern der Arbeitsgruppe CB/PG10 der PROFIBUS Nutzerorganisation und insbesondere den Mitgliedern der Arbeitsgruppe TG2 für die Unterstützung und die technische Zuarbeit.

6 Literaturverzeichnis

[BSI2021] Bundesamt für Sicherheit in der Informationstechnik (BSI): Zuordnungstabelle Zuordnung ISO/IEC 27001 sowie ISO/IEC 27002 zum IT-Grundschutz. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Kompendium/Zuordnung_ISO_und_IT_Grundschutz.pdf?__blob=publicationFile&v=5.

[CHU2013] Chuvakin, A. Hrsg.: Logging and log management. The authoritative guide to understanding the concepts surrounding logging and log management. Syngress, Waltham, MA, 2013.

[DIN_EN_IEC_62443-4-2] DKE-Deutsche Kommission Elektrotechnik, Elektronik Informationstechnik DIN und VDE, DIN EN IEC 62443-4-2 (VDE 0802-4-2): IT-Sicherheit für industrielle Automatisierungssysteme - Teil 4-2: Technische Sicherheitsanforderungen an Komponenten industrieller Automatisierungssysteme (IACS) (IEC 62443-4-2:2019); Deutsche Fassung EN IEC 62443-4-2:2019, 2019.

[DIN_EN_ISO/IEC_27001] DIN-Normenausschuss Informationstechnik und Anwendungen (NIA), DIN EN ISO/IEC 27001:2024: Informationssicherheit, Cybersicherheit und Datenschutz – Informationssicherheitsmanagementsysteme – Anforderungen (ISO/IEC 27001:2022); Deutsche Fassung EN ISO/IEC 27001:2023, 2024. URL: https://www.dinmedia.de/de/norm/din-en-iso-iec-27001/370680635.

[DIN_EN_ISO_IEC_27002] DIN-Normenausschuss Informationstechnik und Anwendungen (NIA), DIN ISO/IEC 27002:2024: Informationssicherheit, Cybersicherheit und Schutz der Privatsphäre – Informationssicherheitsmaßnahmen (ISO/IEC 27002:2022); Deutsche und Englische Fassung prEN ISO/IEC 27002:2022 - Entwurf, 2024.

[EU_3602] Eurpean Commission: European Commission Information System Security Policy C(2006) 3602. Standard on Logging and Monitoring. URL: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/1449046/4e9f17de-4589-424c-a670-c0cdc1b5f67b/Annex%205%20Standard%20on%20Logging%20and%20Monitoring.pdf?retry=1.

[EUR_2022_272] European Parliament and of the Council: Regulation of the European Parliament and of the Council on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020. Cyber Resilliance Act. URL: https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act.

[FUH2016] Fuhr, David et al.: Profilierung von IT-Sicherheitsstandards für den Maschinen- und Anlagenbau. URL: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/vdma-security-automation.html.

[GAR2023] Gartner Inc.: Security Information and Event Management (SIEM) Reviews and Ratings. Gartner Peer insights. URL: https://www.gartner.com/reviews/market/security-information-event-management.

[ISA2024] ISA - The International Society of Automation: The Case for ISA/IEC 62443 Security Level 2 as a Minimum for COTS Components. URL: https://www.isasecure.org/hubfs/The-Case-for-ISA-IEC-62443-Security-Level-2-as-a-Minimum-FINAL.pdf.

[MIT2010] Mitre Corporation: Common Event Expression. Architecture Overview. URL: https://cee.mitre.org/docs/CEE_Architecture_Overview-v0.5.pdf.

[MIT2023] Mitre Corporation: MITRE ATT&C ICS Matrix. URL: https://attack.mitre.org/matrices/ics/.

[NAM2023] NAMUR AK 4.18 Automation Security: Angriffserkennung nach IT-Sicherheitsgesetz 2.0. URL: https://www.namur.net/fileadmin/media_www/Dokumente/AK-PRAXIS_AK_4.18_Angriffserkennung_2023-06-14_DE.pdf.

[NIE2023] Niemann, Karl-Heinz; Walz, Andreas; Sikora, Axel: Security Extensions for PROFINET. Concepts, Status, and Prospects. In (Sikora, A. Hrsg.): Embedded World Conference 2023 Proceedings. WEKA Fachmedien GmbH, 2023; S. 99–104.

[NIST_800-137] NIST Computer Security Division, NIST 800-137: Information Security Continuous Monitoring (ISCM) for Federal Information

Systems and Organizations, 2011. URL: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-137.pdf.

[NIST_800-92] Kent; Karen; Souppaya, Murugiah: Guide to Computer Security Log Management. NIST Special Publication 800-92. URL: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-92.pdf.

[PNO2024a] PROFIBUS Nutzerorganisation e.V.: Application Layer protocol for decentralized periphery Technical Specification for PROFINET IO. Version 2.4 MU5. URL: https://www.profibus.com/download/profinet-specification.

[PNO2024b] PROFIBUS Nutzerorganisation e.V.: Application Layer services for decentralized periphery, Technical Specification for PROFINET IO. Version 2.4 MU5 – March 2024. URL: https://de.profibus.com/downloads/profinet-specification/.

[RFC_5424] Network Working Group: The Syslog Protocol. Request for Comments: 5424. URL: https://datatracker.ietf.org/doc/html/rfc5424.

[RFC_5848] Kelsey, J.: Request for Comments: RFC 5848. Signed Syslog Messages. URL: https://datatracker.ietf.org/doc/html/rfc5848.

[RFC_6587] Internet Engineering Task Force (IETF): Transmission of Syslog Messages over TCP. Request for Comments: 6587. URL: https://datatracker.ietf.org/doc/html/rfc6587.

An Industry 4.0 ontology-based framework for interoperability at the field level

Victor Chavez [1], Jörg Wollert [2]

Abstract: Interoperability at the field level is dependent on the specific technology implementation and its semantics. Integrating field devices with different communication protocols is not a simple process, as there is no direct semantic mapping between them. In recent years, standards such as the OPC UA Field eXchange and the Asset Administration Shell have proposed neutral data models to reduce the heterogeneity of field device semantics. However, to integrate different field device standards, a formal mapping between the semantic terms of these standards and the neutral data models must still be defined. A research topic that remains open is how different standards can be automatically mapped to a neutral interface independent of their implementation. In this paper we present a novel approach that generalizes the semantics of field devices at the communication level, enabling the use of inference rules that are independent of specific standards. Our method, based on the Industry 4.0 Field Device ontology, identifies the generic type of any field device and provides an interoperable capability description adaptable to various protocols. The framework includes a semantic broker that automates the creation of device instances, executes inference requests, and generates a generic semantic model for field devices. The objective of this work is to simplify the integration of field device semantics, with a generalization of the application layer and facilitate their mapping to other higher-level data models.

Keywords: Field device, ontology, Industry 4.0, interoperability, capabilities, framework, semantic model

1 Introduction

Industrial organizations and consortia have standardized field device communication protocols and facilitated their integration across multiple vendors. As interoperability has become an important concern for the creation of intelligent industrial systems, generalization and abstraction of communication protocols are required [1], [2]. Organizations working towards the Industry 4.0 (I4.0) strategy [3], have published neutral data formats and interfaces to address this challenge. Two recent examples of this, are the Asset Administration Shell (AAS) [4] and the OPC UA Field eXchange (OPC UA FX) [5]. Although these standards provide a generic representation of field devices, a semantic alignment of the different field device technologies is required. This is typically done via a mapping mechanism that assigns process variables or parameters to the specific semantics of one data model to another [6], [7]. However, since there is no generic mapping model for field device technologies, this semantic alignment results in fragmented data models. Standards such as the Field Device Integration (FDI) and

[1] Institut für angewandte Automation und Mechatronik, Goethestraße 1, 52064 Aachen, chavez-bermudez@fh-aachen.de, https://orcid.org/0000-0001-6419-1641
[2] Institut für angewandte Automation und Mechatronik, Goethestraße 1, 52064 Aachen, wollert@fh-aachen.de, https://orcid.org/0000-0001-7576-1339

companion specifications for OPC UA enable the representation of specific field device technologies over a common interface. Nevertheless, these interfaces only define a data bridge to a specific protocol but do not cover the abstraction of their application data (see Fig. 1).

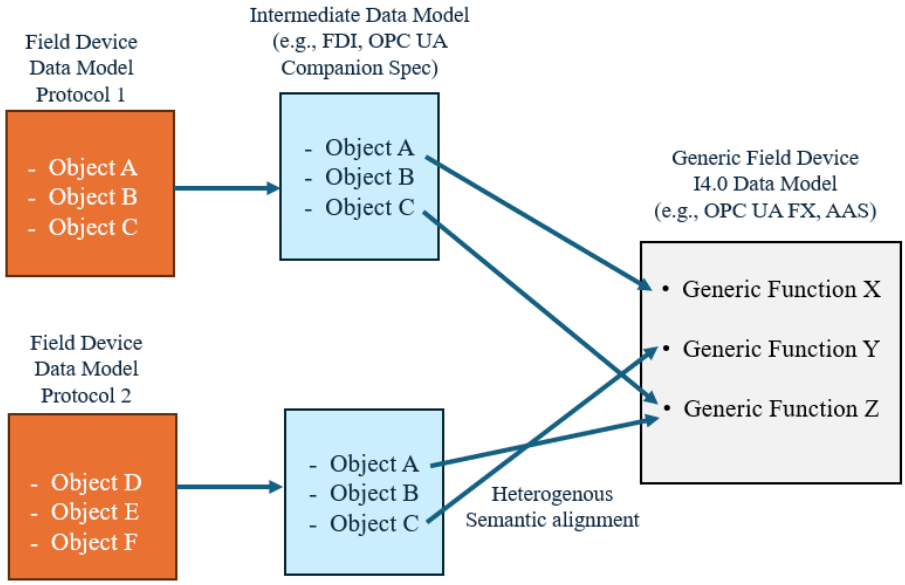


Fig. 1: Semantic alignment heterogeneity for multiple field devices.

This paper presents an ontology-based approach to generalize the semantic representation of application data from field devices. This generalization is employed to develop a software framework that enables the automatic generation of field device capabilities. The motivation for this work is the increasing complexity and diversity of field devices in Industry 4.0 environments, where interoperability remains a significant challenge due to the varying communication protocols and semantic models. This fragmentation leads to inefficient and error-prone mappings between standards, hindering the development of intelligent industrial systems. The objective of this research work is to simplify the integration of field devices, enhance interoperability, and facilitate the seamless exchange of information across diverse industrial protocols.

This paper is structured as follows: Section 2 presents an analysis of various information models utilized in field devices, identifying similarities across standards and gaps. Subsequently, section 3 introduces the ontology the Industry 4.0 Field Device ontology (I40FD) which is the foundational basis for this work. Then in section 4, the software framework that facilitates the automatic representation of generic field devices is presented. In Section 5, we show a small proof-of-concept to validate the correctness of the framework. Lastly, in Section 6, we present the conclusions of this research and future directions for further work.

2 Field Device Information Models

Field devices provide a standardized description of their functionality in the form of an information model. There is not a unique information model among standards but there are similarities between them. An information model for field devices describes identification information, an addressing mechanism, communication properties, and the structure of their application data. An engineering configuration tool is used to load the information model via a device description file. The tool enables the configuration of the device, development of an application over an integrated development environment or maintenance tasks.

Miny et al. [8] conducted a comparative analysis of several information models for representing assets. Among them are the Web of Things, Automation ML, FDI, and Process Automation Device Information Model (PA-DIM). From this analysis, they share the conceptualization of terms such as properties, services, data types, units, and access types. Similarly, in previous work [9], we compared the application layer of two field device standards and proposed the generic conceptualization of application data according to access type, data type, and quantity type.

The similarities presented from different field device standards show that there is a way to generalize and abstract the application layer for field devices. To the best of our knowledge, there is not a semantic model to generalize the application data. Standards such as PA-DIM, FDI, or the OPC UA companion specification provide only a direct mapping to the underlying application data. This only facilitates access to a specific semantic model but does not enable an interoperable interface independent of the protocol as the data model is still tied to a specific implementation.

To fill this gap, we propose to reuse the semantics of the application layer independent of the field device technology. Instead of defining a generic interface with a direct mapping to an auxiliary data model, we derive generic functions that a device can perform based on the application data. In this case, the application data is only used to describe how a field device capability is implemented (see Fig. 2). Compared to current standard field device interfaces (e.g. OPC UA, FDI), we leverage the power of Semantic Web Technologies to infer the context of the application data and define an abstract capability model for field devices. In the next section, this generalization is implemented with an ontology for field devices and a comparison to current approaches is shown.

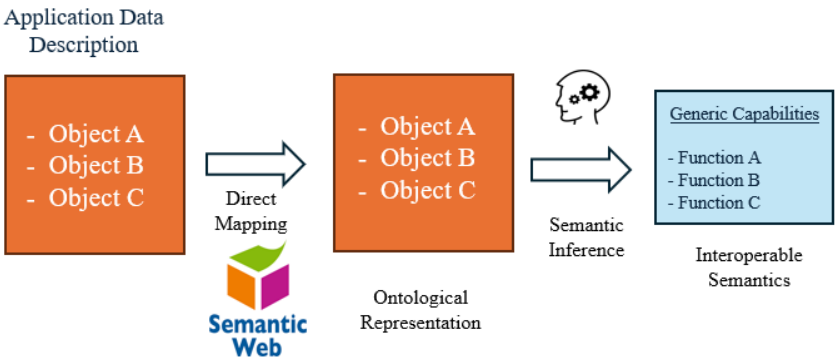


Fig. 2: Interoperable semantic description independent of implementation

3 An Industry 4.0 Field Device Ontology

The use of SWT standards such as the Web Ontology Language and the Resource Description Framework (RDF) is a recurrent topic for interoperability at the field level. Dibowski and Kabitzsch [10] proposed the description of automation devices via an ontology-based description and a triplestore database. The ontology architecture consists of four layers that define the domain's specific vocabulary, platform-specific data, manufacturer-specific types, and manufacturer-specific device descriptions. A test dataset for the building automation domain was used to analyze the performance of queries with the SPARQL Protocol and RDF Query Language (SPARQL).

Hammerstingle and Reinhart [11] present a Plug&Produce architecture for the automatic integration of field devices. The concept entails the loading of the device description, with information access facilitated by dedicated drivers that standardize the access of information. This content is then made available to higher-level systems, allowing them to access the services and capabilities of the model. An automatic generation process is proposed to convert the access mechanisms to a Program Organization Unit within a Programmable Logic Controller (PLC). The concept is tested with an IO-Link controller, two sensors, and a frequency converter that communicates via EtherCAT with the PLC.

Kannoth et al. introduce the Virtual Automation Bus (VAB) [12] to enable communication protocol interoperability. In this work, a mapping between basic data types of the OPC UA data model and VAB is presented. The approach translates the source protocol into an intermediate syntax that other gateways can communicate with. Another similar approach to mapping information models into existing OPC UA Servers is discussed by Müller et al [13]. Through a mapping rule based on DisplayNames, the semantic information of a custom model is transferred to an existing server based on three accuracy levels that match its value.

Despite similar approaches and concepts for defining interoperable models, the reuse and abstraction of existing information from device description files is missing. The conceptual application of Hammerstingle and Rheinhart [11], is the closest but does not have to the ontology or semantic descriptions. With the motivation of enabling an abstract representation of field device semantics as described in the previous sections, we propose the Industry 4.0 Field Device Ontology [14].

The I40FD ontology introduces an abstract definition of the application layer for field devices. The concept of Application Data Object (AO) encapsulates any type of application object that a communication protocol defines. The AO is constituted of one or more Application Data Elements (AE), which are defined by the basic data type of a specific protocol. Fig. 3 presents the conceptual model of the I40FD ontology.

If the AO is used to exchange process data (i.e. synchronous communication) and contains AEs with unit information, generic field device capabilities can be inferred based on the access type (i.e., read or write). This abstraction level allows us to describe any field device protocol that has an application layer and has Process Data Objects. The capability consists of a realization through a communication process that is prescribed by an AE. These relationships are presented in Fig. 4. The I40FD ontology reuses the Quantities, Units, Dimensions, and Types (QUDT) ontology [15], the Basic Formal Ontology [16], and the Industrial Ontologies Foundry Core ontology [17].

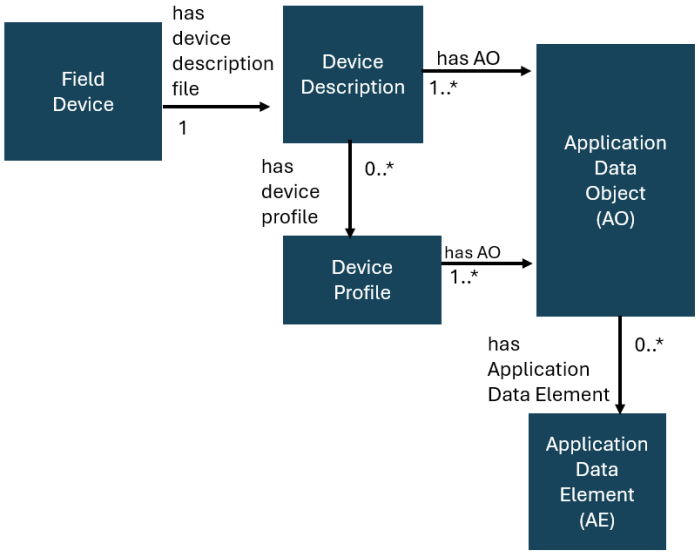


Fig. 3: I40FD ontology conceptual model.

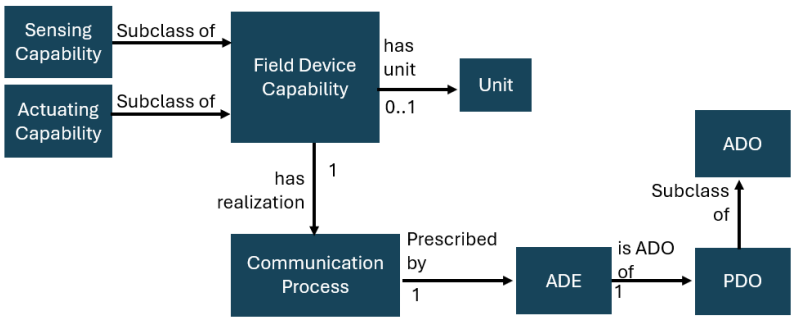


Fig. 4: I40FD ontology conceptual model of a field device capability

The inference of the capabilities is done through the Semantic Web Rule Language (SWRL). The core rule to infer capabilities is based on quantity units linked to any AEs used as process data. The unit description is associated automatically with a protocol unit and a specific encoded value proper of the protocol implementation (see Fig. 6). The units are associated with object property *hasunit* from the QUDT ontology. If the access is of type read it can be inferred that the field device has a sensing capability (see Fig. 7). In contrast, if the access is of type write, then the field device has an actuating capability.

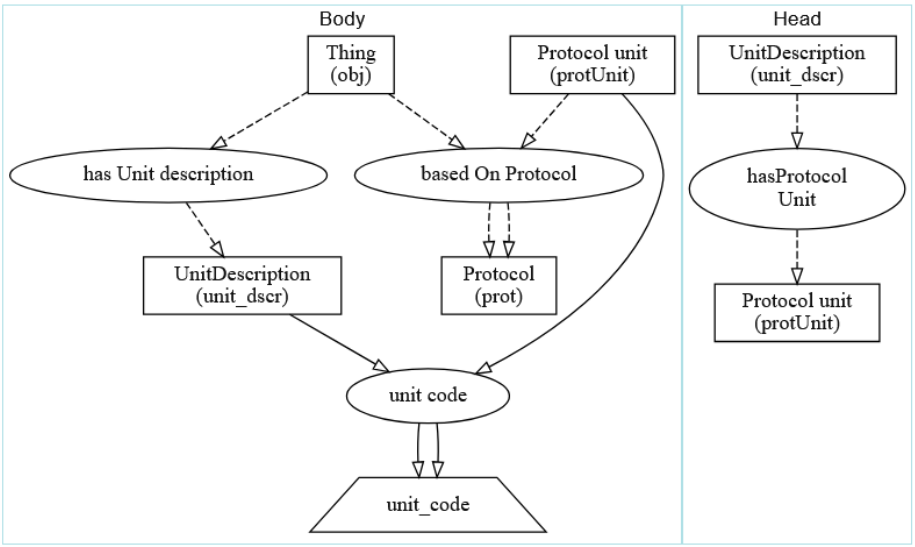


Fig. 6: SWRL inference rule of a protocol unit code.

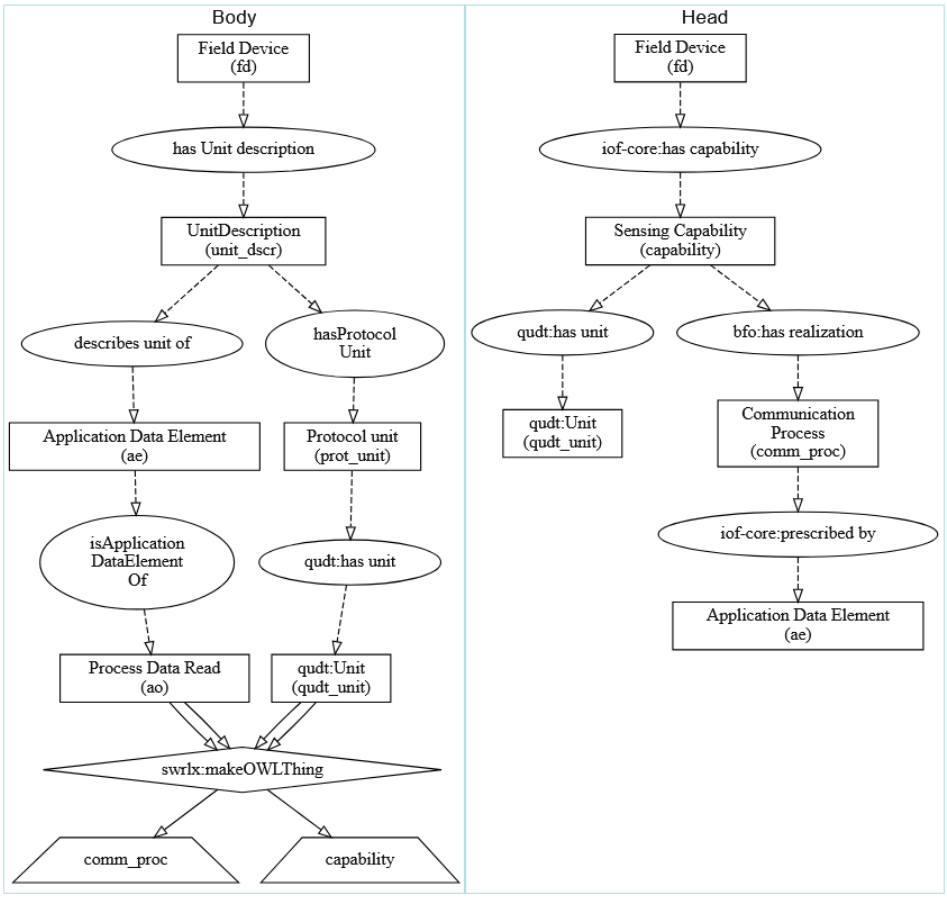


Fig. 7: SWRL inference rule of a sensing capability.

4 I40FD Framework

The I40FD ontology aims to reduce the semantic alignment of field device descriptions with an abstract representation of field device capabilities. To validate the correctness of the ontology and automate the processing of the proposed generic field device capabilities a software framework has been developed. The objective of the I40FD framework is to allow a software client to obtain the generic field device representations that can be used for example for the automatic creation of I4.0 data models (e.g., OPC UA FX or AAS), automatic interpretation of process data or capability matching.

The I40FD framework consists of three main components (see Fig. 8). Firstly, the I40FD ontology includes the abstraction of field device description models, their application data, and specific field device standards. Secondly, the I40FD inference engine that infers field device capabilities. Lastly, the I40FD Semantic broker acts as a bridge between the semantic layer and an end-user application.

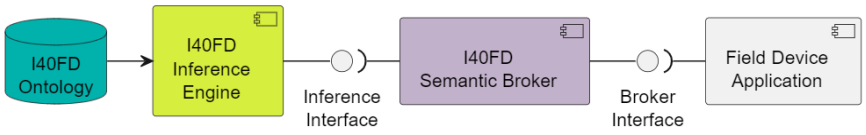


Fig. 8: I40FD Framework component overview.

The I40FD inference engine loads the I40FD ontology, and the SWRL rules and has an interface that sends RDF axioms that represent a field device description according to the I40FD ontology. In general, the axioms that are required for inference are the AOs that the field device has, any device profiles, and unit information. The inference engine loads the axioms and runs a SWRL inference engine. If any field device capabilities are found, then they are reported back to the client (see Fig. 9).

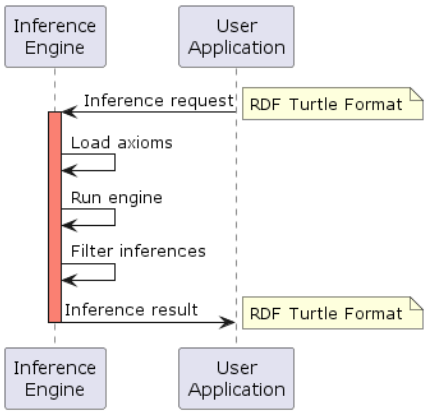


Fig. 9: Sequence diagram of an I40FD inference request.

The semantic broker receives requests from a client that wants to find out any generic capabilities a field device has. The input of the request consists of the specific field device protocol and the device identification. Based on this information, the broker will look for the device description of the device from a local or online catalog. The device description is then mapped to the I40FD ontology, and the axioms are sent to the inference engine. If any capability has been found, the broker will send back the result to the client (see Fig. 10).

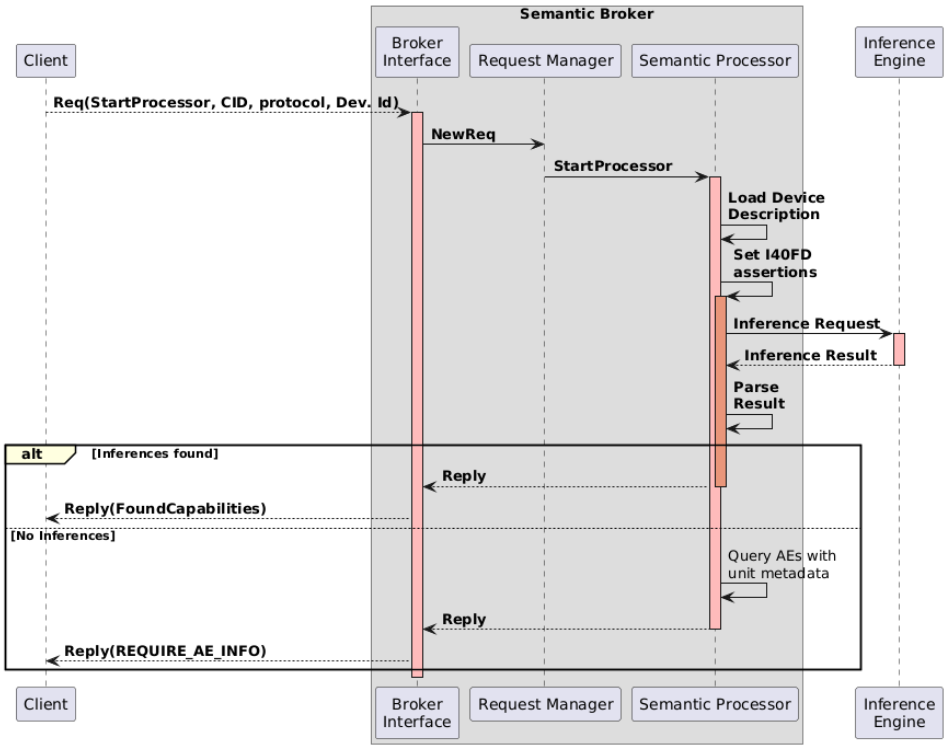


Fig. 10: Field device semantic request sequence diagram.

In the case that the inference result did not result in any capabilities, the broker will query the field device model to check if there is any AE that has unit information associated with process data (see Fig. 11). Then this information will be sent to the client to indicate that more information is required to obtain an inference result. This occurs when a field device has values in its AOs that can only be obtained at run-time.

```
PREFIX i40fd: <http://w3id.org/iaam/i40fd#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?ae ?ae_unit ?ao_unit
WHERE {
  ?device_ao i40fd:hasApplicationDataElement ?ae .
  ?device_ao rdf:type/rdfs:subClassOf* i40fd:ProcessDataObject .
  ?ae i40fd:unitCodeObtainedFrom ?ae_unit .
  ?ao_unit i40fd:hasApplicationDataElement ?ae_unit .
}
```

Fig. 11: SPARQL query for any AE that has unit information associated with process data.

5 Validation

A proof-of-concept has been developed to test the I40FD framework. The application consists of a PLC that runs an I40FD semantic client implementation over a TCP socket. The PLC client has an IO-Link and a CANOpen interface. The client will constantly scan for a new connection from an IO-Link or CAN Open device. When it finds a new device, it will create a request with metadata from the device (i.e., vendor ID, device ID) and await the semantic inference result. The purpose of this application is to validate the software and test its correctness with different protocols.

Fig. 12: I40FD Proof-of-concept with a PLC as a semantic client.

Two example requests are presented in Fig. 13. The requests contain the device identification metadata of an ifm IO-Link pressure sensor PI2899 and an ifm RM9000 encoder with a CANOpen interface. The semantic broker loads from a local database the device description of both devices and sends a semantic request to the broker with the I40FD axioms that represent the field device description file.

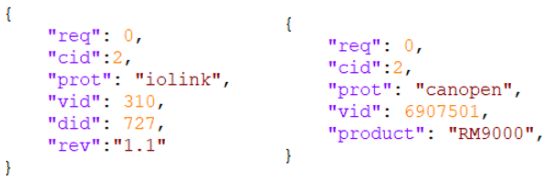


Fig. 13: JSON Request for an IO-Link and CANOpen device.

Afterwards, the semantic broker gets the capabilities found and sends the information back to the PLC client (see Fig. 14). For this proof-of-concept the PLC client simply interprets the capability AE origin to display the process data over the respective communication interface (see Fig. 15).

Fig. 14: Inference result for an IO-Link pressure sensor (left) and CANOpen encoder sensor (right).

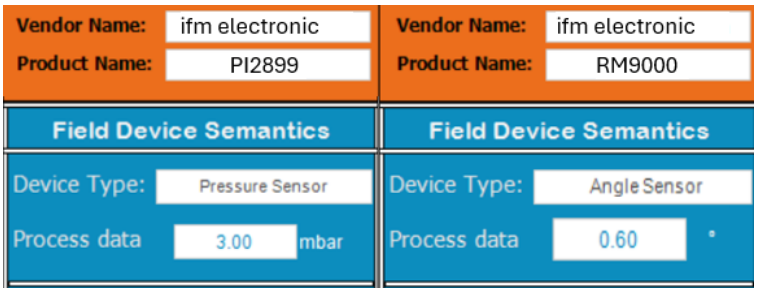


Fig. 15: PLC graphic representation of semantic information.

6 Conclusions

In this work, we have introduced an ontology-based framework to enable the interoperability of field device semantic models. The abstraction model is comprised of a generalization of the application layer and its relationship to device profiles, units, and access types. A software framework was developed based on the I40FD ontology and SWRL inference rules to facilitate the semantic mapping of field device description files into a generic capability-based information model. The specific implementation of the capabilities is based on the AOs that a field device contains.

In contrast with existing approaches, the I40FD semantic model does not suggest the introduction of an additional transport interface for the application layer. Instead, it offers a generic description of field device functionalities and their realization with a specific application layer. The semantic framework was validated using two field devices with disparate semantic models. It is demonstrated that irrespective of the specific semantics, an abstract representation is feasible.

Although a simple proof of concept was demonstrated, the primary objective of this framework is to illustrate that higher-level applications could leverage the use of an abstract model based on capabilities to enable seamless integration with other systems or data models.

In future work, we intend to expand the I40FD vocabulary and device profiles with other field device protocols. Additionally, we will investigate the automatic generation of I4.0 field device interfaces with AAS and OPC UA FX.

Acknowledgement

The authors of this paper would like to thank ifm electronic for the financial support of this research project.

References

[1] Y. Liao, L. F. P. Ramos, M. Saturno, F. Deschamps, E. de Freitas Rocha Loures, and A. L. Szejka, The Role of Interoperability in The Fourth Industrial Revolution Era, IFAC-PapersOnLine, vol. 50, no. 1, pp. 12434–12439, Jul. 2017

[2] R. M. Pereira, A. L. Szejka, and O. Canciglieri Junior, Towards an information semantic interoperability in smart manufacturing systems: contributions, limitations and applications, International Journal of Computer Integrated Manufacturing, vol. 34, no. 4, pp. 422–439, Apr. 2021

[3] Federal Ministry for Economic Affairs and Climate Action - BMWK, Platform Industrie 4.0, https://www.plattform-i40.de/IP/Navigation/EN/Home/home.html, Accessed: 01 September 2024.

[4] Platform Industrie 4.0, Details of the Asset Administration Shell Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC02), 2022.

[5] OPC UA Foundation, OPC 10000-80 UAFX Part 80: Overview and Concepts, 2022.

[6] A. Barth, B. Balakrishna, and A. Willner, Configurable Mapping of EtherCAT field-level devices to OPC UA, in 2022 International Young Engineers Forum (YEF-ECE), IEEE, Jul. 2022, pp. 57–62.

[7] S. Cavalieri and M. G. Salafia, Insights into Mapping Solutions Based on OPC UA Information Model Applied to the Industry 4.0 Asset Administration Shell, Computers, vol. 9, no. 2, p. 28, Apr. 2020.

[8] T. Miny et al., Overview and Comparison of Asset Information Model Standards, IEEE Access, vol. 11, pp. 99189–99221, 2023.

[9] V. Chavez and J. Wollert, An Industry 4.0 Ontology-based Architecture for Interoperability at the Field Level, in 4th IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 / 5.0 (ARCI' 2024), IFSA, 2024, pp. 319–321.

[10] H. Dibowski and K. Kabitzsch, Ontology-based Device Descriptions and triple store based device repository for automation devices, in 2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETFA 2010), IEEE, Sep. 2010, pp. 1–9.

[11] V. Hammerstingl and G. Reinhart, Unified Plug&Produce architecture for automatic integration of field devices in industrial environments, in 2015 IEEE International Conference on Industrial Technology (ICIT), IEEE, Mar. 2015, pp. 1956–1963.

[12] S. Kannoth, F. Schnicke, and P. O. Antonino, Enabling Industry 4.0 Communication Protocol Interoperability: An OPC UA Case Study, in 7th Conference on the Engineering of Computer Based Systems, New York, NY, USA: ACM, May 2021, pp. 1–9.

[13] A. Muller, T. Schnieders, S. Storms, and W. Herfs, Integration method of custom information models into existing OPC UA Servers, in 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, Sep. 2022, pp. 1–7.

[14] V. Chavez and J. Wollert, Sensors & Transducers Development of an Industry 4.0 Ontology to Enable Semantic Interoperability at the Field Level, Sensors & Transducers, vol. 265, pp. 139–147, 2024.

[15] QUDT.org, Quantities, Units, Dimensions and Types, https://qudt.org., Accessed: 01 September 2024.

[16] Robert Arp, B. Smith, and A. D. Spear, Building Ontologies with Basic Formal Ontology. The MIT Press, 2015.

[17] M. Drobnjakovic, B. Kulvatunyou, F. Ameri, C. Will, B. Smith, and A. Jones, "The Industrial Ontologies Foundry (IOF) Core Ontology, in FOMI 2022: 12th International Workshop on Formal Ontologies meet Industry, 2022.

Vorgehensmodell zur Bewertung und Planung omlox-basierter Indoor-Lokalisierungssysteme in Industrieumgebungen

Harry Fast[1], Florian Hufen[1], Holger Flatt[1], Salar Bybordi[2] und Franz Lehmann[2]

Abstract: In diesem Beitrag wird ein Vorgehensmodell zur Bewertung und Planung von omlox-basierten Indoor-Lokalisierungssystemen in Industrieumgebungen vorgestellt. Ziel ist es, die zu erwartende Lokalisierungsgenauigkeit unter Berücksichtigung von Störeinflüssen abzuschätzen und eine fundierte Auslegung der Satelliteninfrastruktur zu ermöglichen. Das vorgeschlagene Modell basiert auf einer zweistufigen Methodik. In der ersten Phase erfolgt eine qualitative Bewertung der Produktionsumgebung anhand spezifischer Eigenschaften wie Sichtverbindung, Funkreflexionen, Funkabsorption, Anwesenheit anderer Funkumgebungen, EMV-Störungen sowie Dynamik im Produktionsumfeld. Diese Einordnung wird auf der Grundlage einer Bewertungsmatrix vorgenommen. Anschließend wird das Ergebnis dieser Bewertung zusammen mit der gewünschten Genauigkeitsklasse in ein Modell integriert, welches durch Genauigkeitsmessungen in drei verschiedenen Umgebungen erstellt wurde. Die Ergebnisse zeigen, dass ein Modell basierend auf der durchschnittlichen Distanz der Satelliten zu einem Referenzraster in Verbindung mit einer kubischen Hermite-Interpolation eine zuverlässige Vorhersage der erreichbaren Lokalisierungsgenauigkeit für das 95% Fehlerperzentil ermöglicht. Die Anwendung des Vorgehensmodells kann Unternehmen eine präzisere Planung und Implementierung von Indoor-Lokalisierungssystemen ermöglichen und dazu beitragen, die Effizienz und Zuverlässigkeit von Produktionsprozessen zu erhöhen und die Integration neuer Technologien in komplexen Industrieumgebungen zu erleichtern.

Keywords: Indoor-Lokalisierung, omlox, Vorgehensmodell, Genauigkeitsmessungen

1 Einleitung

Indoor-Lokalisierungssysteme spielen seit vielen Jahren eine zunehmend wichtigere Rolle in Industrieumgebungen, da sie eine präzise und zuverlässige Erkennung und Verfolgung von Objekten und Personen ermöglichen [FH15][HS22][HF23]. Der omlox-Standard hat sich hierbei als ein vielversprechender Ansatz für die interoperable Implementierung solcher Systeme ohne Vendor-Lock-In etabliert, da dieser durch die Herstellerunabhängigkeit, im Vergleich zu proprietären Systemen, neue und skalierbarere Indoor-Lokalisierungs-Lösungen (z. B. unternehmensübergreifend) ermöglicht [HI20]. Im Bereich der funkbasierten Lokalisierung setzt der omlox-Standard auf

[1] Fraunhofer IOSB, Institutsteil für industrielle Automation (IOSB-INA), Campusallee 1, 32657 Lemgo, {harry.fast, florian.hufen, holger.flatt}@iosb-ina.fraunhofer.de
[2] TRUMPF Tracking Technologies GmbH, Dornierstraße 12, 71254 Ditzingen, {salar.bybordi, franz.lehmann}@trumpf.com

Ultrabreitbandfunk (UWB). Dieser verspricht im Vergleich zu anderen Lokalisierungstechnologien (z.B. WLAN oder RFID) eine Robustheit und Genauigkeit für viele Anwendungen und besteht im Wesentlichen aus der Infrastruktur, den sogenannten ortsfesten Satelliten sowie den mobilen, zu lokalisierenden Tags. Bisherige Untersuchungen von UWB-Indoor-Lokalisierungstechnologien zeigen, dass die erreichbaren Genauigkeiten signifikant von der Umgebung abhängig sind, in der sie betrieben werden. Während im ungestörten Fabrikumfeld Lokalisierungs-Genauigkeiten im Bereich 30 cm erreichbar sind [SB21], erschweren eine Vielzahl an Faktoren wie Sichtbehinderungen, Funksignalreflexionen durch metallische Gegenstände, Funksignalabsorption (z.B. menschliche Körper) oder EMV-Einflüsse eine derartige Präzision in anspruchsvollen Industrieumgebungen [SA22, ZW22]. Derzeit greifen die Infrastrukturhersteller für omlox bei der Planung von Systemen auf Erfahrungswerte von Experten zurück, da formale Ansätze bisher nicht verfügbar oder auf andere Lokalisierungstechnologien beschränkt sind [ZZ16]. Speziell für komplexe Umgebungen mit verschiedenen Störeinflüssen fehlt bisher eine validierte Methodik, die beispielsweise eine Abschätzung der erzielbaren Lokalisierungs-Genauigkeit und der erforderlichen Anzahl installierter Satelliten ermöglicht.

Im Rahmen dieses Beitrages wird ein Vorgehensmodell zur Bewertung und Planung omlox-basierter Indoor-Lokalisierungssysteme in Industrieumgebungen vorgestellt, welches auf der Grundlage von Genauigkeitsmessungen in drei verschiedenen Umgebungen erstellt wurde. Ziel ist es, hiermit Industrieunternehmen dabei zu unterstützen, für spezifische Umgebungen unter Berücksichtigung von Störeinflüssen zu erwartende Lokalisierungsgenauigkeiten abschätzen zu können und eine Hilfestellung zur Auslegung der Satelliteninfrastruktur zu geben.

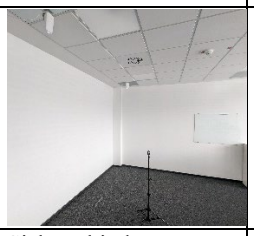
Für die Modellierung wird ein zweistufiges Konzept vorgeschlagen, welches in einer ersten Phase zunächst eine qualitative Bewertung der Produktionsumgebung erfordert. Das Ergebnis dieser ersten Einstufung wird als Eingabe zusammen mit der gewünschten Genauigkeitsklasse in ein Modell gegeben, welches auf der Grundlage von Genauigkeitsmessungen mit einem mobilen Prüfsystem in drei verschiedenen Umgebungen gebildet und evaluiert wurde.

Der Beitrag gliedert sich wie folgt: Abschnitt 2 bildet die Grundlagen zur Klassifizierung von Produktionsumgebungen. Abschnitt 3 stellt den Prüfaufbau zur Durchführung von Genauigkeitsmessungen sowie Ergebnisse einer Messkampagne vor. Auf dieser Grundlage erfolgt anschließend die Modellbildung mit einem Anwendungsvorschlag zur Auslegung von omlox-Infrastrukturen. Abschnitt 4 fasst den Beitrag zusammen und gibt einen Ausblick auf zukünftige Arbeiten.

2 Klassifizierung von Produktionsumgebungen

Die Untersuchungen in dieser Veröffentlichung gehen von der Hypothese aus, dass unterschiedliche Produktionsumgebungen die Lokalisierungsgenauigkeit unterschiedlich

beeinträchtigen. Mit dem Ziel einer möglichst einfachen Anwendbarkeit erfolgt im Rahmen dieses Beitrages eine Differenzierung von Produktionsumgebungen in drei verschiedene Klassen, welche in Tab. 1 vergleichend gegenübergestellt sind.

Eigenschaft	**Ungestörtes Umfeld**	**Produktions-umgebung**	**Produktionsumgebung mit komplexen Störgrößen**
Beispielbild			
Line-of-Sight/ Verschattung	Sichtverbindung zu allen Satelliten vorhanden	Sichtverbindung zu einigen Satelliten vorhanden	Kaum Sichtverbindung zu Satelliten vorhanden
Funk-reflexionen	Kaum bis wenig metallische und reflektierende Flächen im Lokalisierungs-Umfeld	Teilweise Metall im Bereich des Lokalisierungs-Umfelds	Viel Metall im Bereich des Lokalisierungs-Umfelds
Funk-absorption	Kaum bis wenig absorbierende Flächen und Körper (z.B. menschliche Körper)	Teilweise absorbierende Flächen	Viele absorbierende Flächen
Andere Funk-umgebungen (z. B. WLAN, BT, 5G)	Keine / nicht signifikant	Wenige	Viele
EMV-Störungen	Keine / nicht signifikant	z. B. Roboter, Förderbänder, kleine und mittlere Motoren	z. B.: Große Elektromotoren, IGBTs, Schweißzellen
Dynamik	Keine Veränderungen, statisches Umfeld	Wenig (z. B. Roboter-arme, einzelne Personen oder FTS)	Hoch (z. B. viele Flurförderfahrzeuge oder Regalbediengeräte)

Tab. 1: Gegenüberstellung verschiedener Typen von Produktionsumgebungen

In einem **ungestörten Umfeld** besteht eine klare Sichtverbindung zu allen Satelliten, was eine präzise Lokalisierung ermöglicht. Unter derartigen Bedingungen finden beispielsweise Zertifizierungstests für omlox-Core Zone-Komponenten statt [PI23]. Das Vorhandensein von metallischen und reflektierenden Flächen ist minimal, wodurch Funkreflexionen kaum auftreten. Auch absorbierende Flächen und Körper, wie menschliche Körper, sind kaum vorhanden, was die Funkabsorption geringhält. Andere Funkumgebungen wie z.B. WLAN, Bluetooth (BT) oder 5G sind nicht vorhanden oder werden nur mit nicht störender Sendeleistung betrieben. Elektromagnetische

Verträglichkeit (EMV)-Störungen sind ebenfalls nicht vorhanden oder von geringer Bedeutung. Zudem gibt es keine dynamischen Veränderungen in dieser Umgebung.

In einer **Produktionsumgebung** ist die Sichtverbindung zu einigen Satelliten gegeben, was die Lokalisierung bereits etwas erschwert. Metallische Flächen sind teilweise im Lokalisierungs-Umfeld vorhanden, was zu Funkreflexionen führen kann. Ebenso sind absorbierende Flächen teilweise vorhanden, was die Funkabsorption erhöht. Andere Funkumgebungen wie 2,4 GHz oder 5G sind in geringen Mengen vorhanden. EMV-Störungen können durch Roboter, Förderbänder und kleine bis mittlere Motoren verursacht werden. Die Dynamik in dieser Umgebung ist gering, mit gelegentlichen Bewegungen von Personen oder fahrerlosen Transportsystemen (FTS) ist zu rechnen.

In einer **Produktionsumgebung mit komplexen Störgrößen** ist die Sichtverbindung zu Satelliten stark eingeschränkt, was die Lokalisierung erheblich beeinträchtigt. Es gibt viele metallische Flächen, die zu zahlreichen Funkreflexionen führen. Die Anzahl der absorbierenden Flächen ist hoch, was die Funkabsorption weiter verstärkt. Andere Funkumgebungen wie 2,4 GHz oder 5G sind zahlreich vorhanden, was zusätzliche Störungen verursacht. EMV-Störungen sind durch große Elektromotoren, IGBTs und Schweißzellen signifikant vorhanden. Die Dynamik ist in dieser Umgebung hoch, mit vielen Flurförderfahrzeugen und anderen beweglichen Elementen.

3 Prüfaufbau zur Durchführung von Genauigkeitsmessungen

3.1 Auswahl repräsentativer Umgebungen

Anhand der oben identifizierten Klassifizierungen werden für jedes Umfeld umfangreiche Messreihen durchgeführt, um ausgehend von den Ergebnissen ein geeignetes Modell zur Bestimmung der Positionierungsgenauigkeit in Abhängigkeit zur Satelliten-Dichte und -Positionierung erstellen zu können. Folgende Messumgebungen wurden aufgrund ihrer Beschaffenheit als Repräsentanten für die Messung ausgewählt:

Ungestörtes Umfeld (Laborbedingungen)

Für die Messungen in einem ungestörten Umfeld wurde das Institutsgelände des Fraunhofer IOSB-INA in Lemgo (siehe Abbildung 1) als Messfeld ausgewählt. Der leere Parkplatz bietet eine uneingeschränkte Sichtverbindung zwischen allen Satelliten und zum Tag. Zudem gibt es hier kaum metallische und reflektierende Flächen, was die Signalreflexionen und Mehrwegeffekte minimiert, welche die Genauigkeit der Lokalisierung beeinträchtigen könnten. Weiterhin sind auf dem leeren Parkplatz keine absorbierenden Flächen und Objekte wie menschliche Körper vorhanden, was die Dämpfung der Signale reduziert und eine maximal mögliche stabile Verbindung sicherstellt. Darüber hinaus gibt es keine signifikanten Störquellen, welche die Messungen beeinträchtigen könnten. Die Umgebung bleibt zudem über die Zeit der Messungen

unverändert, was ebenfalls einem idealen Umfeld entspricht. Die Sendeleistung für die Messung wurden dabei mit den Einstellungen für den Indoor-Betrieb konfiguriert, um eine Vergleichbarkeit mit den beiden anderen Umgebungen zu erzielen.

Abbildung 1: Messfeld auf dem Institutsgelände des Fraunhofer IOSB-INA in Lemgo: Links Foto, rechts Schemazeichnung Satellitenanordnung

Produktionsumgebung

Für die Messungen in einer Produktionsumgebung (PU) wurde die Produktionsfläche der SmartFactoryOWL in Lemgo (siehe Abbildung 2) als Messfeld ausgewählt. Die Sichtverbindung der Tags ist an den meisten Positionen zu mehreren Satelliten gegeben, jedoch gibt es teilweise metallische Strukturen im Bereich des Lokalisierungs-Umfelds, welche Signalreflexionen und Mehrwegeffekte verursachen könnten. Zudem sind teilweise absorbierende Flächen vorhanden, welche die Dämpfung der Signale beeinflussen könnten. Es gibt einige wenige Störquellen, wie beispielsweise Roboter, Förderbänder, kleine bis mittlere Motoren sowie 5G-Campus- und WLAN-Kommunikationsnetze. Darüber hinaus findet wenig Bewegung durch einzelne Personen oder fahrerlose Transportsysteme (FTS) statt.

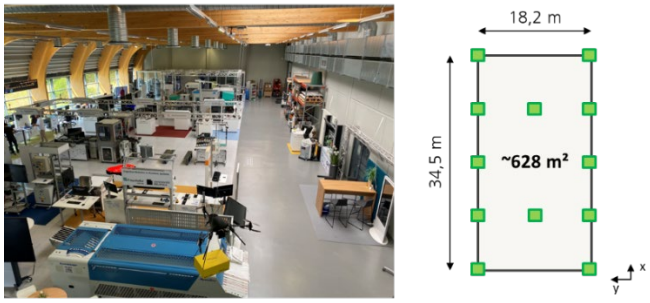


Abbildung 2: Messfeld in der SmartFactoryOWL: Links Foto, rechts Schemazeichnung Satellitenanordnung

Produktionsumgebung mit komplexen Störgrößen

Für die Messungen in einer Produktionsumgebung mit komplexen Störgrößen wurde die Lagerfläche (Stehendlager für Stahl-Coils) des Bilstein Group Werkes in Hagen ausgewählt (sieh Abbildung 3). In dieser Umgebung kann die Sichtverbindung der Tags zu den Satelliten stark eingeschränkt sein, weil diese z.B. an einem Stahl-Coil im Lager an einer Seite befestigt sind, welche von weiteren Coils verdeckt werden. Zudem gibt es im Bereich des Lokalisierungs-Umfelds neben den Stahl-Coils viele metallische Strukturen und Maschinen, die Signalreflexionen und Mehrwegeffekte verursachen können. Um tiefere Einblicke darüber zu erhalten, welche Auswirkungen die Abschirmung der Tags durch die Art der Lagerung der Coils hat, wurden in dem Werk noch ergänzende Messungen in einem weiteren Messfeld mit liegend gelagerten Coils durchgeführt (siehe Abbildung 4 (A)), da sich hier durch die Anbringung der Tags zwischen aufeinander gelagerten Coils eine noch komplexere Lokalisierungsumgebung vorliegt. Ferner enthält diese Umgebung auch zahlreiche potenzielle Störquellen wie große Elektromotoren, welche das Potenzial haben, die Messungen zusätzlich beeinflussen zu können. Um den Einfluss der Motoren gesondert zu erfassen, wurde ebenfalls eine Messung in der Nähe leistungsstarker Elektromotoren durchgeführt (siehe Abbildung 4 (B)).

Des Weiteren stellt der Verkehr von Flurförderfahrzeugen im Lagerbereich eine weitere Störquelle im regulären Betrieb dar.

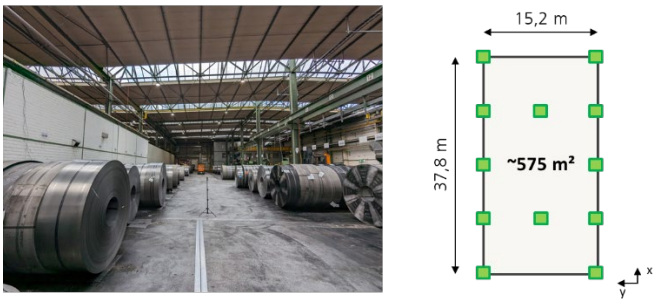


Abbildung 3: Stahl-Coil-Stehendlager des Bilstein Group Werkes in Hagen: Links Foto, rechts Schemazeichnung Satellitenanordnung

(A) (B)

Abbildung 4: Ergänzende Messungen im Werk der Bilstein Group in Hagen: (A) Stahl-Coil-Liegendlager, (B) Leistungsstarke Elektromotoren in der Nähe

3.2 Vorgehen zur Datenaufnahme

Für die Durchführung der Messungen wurde das Coriva-System von der Firma ZIGPOS GmbH genutzt, welches den omlox V2 spezifischen Anforderungen entspricht. Das Testsystem besteht aus 12 Satelliten einer Location-Engine und einem Tag. Die Location Engine verwaltet die Netzwerktopologie der Satelliten und stellt darüber hinaus Möglichkeiten zur Aufzeichnung von Messwerten bereit.

Der jeweils temporäre Aufbau der Satelliteninfrastruktur verfolgt das Ziel, die Abstandsmaße in ähnlichen Größenordnungen und Anordnung zu halten, um vergleichbare Bedingungen für die Genauigkeitsbetrachtung für die Gegenüberstellung der verschiedenen Umgebungen zu erhalten. Ein exakt gleicher Aufbau ist jedoch aufgrund äußerer Gegebenheiten vor Ort wie z.B. Stahlstützen in der Halle oder Laternenmasten, welche zur Montage genutzt wurden, nicht möglich. Je nach Aufbau wurden weitere Satelliten auf Stativen verwendet.

Der Ablauf bei der Durchführung der Messungen an den einzelnen Messfeldern (mit Ausnahme der Messreihe nahe den Elektromotoren) ist wie folgt:

1. Aufbau und Inbetriebnahme der Infrastruktur, sowie Einmessen der Satelliten mit einem 3D-Laser-Abstandsmesser[3] (Einmessen in 3D-Koordinaten möglich).
2. Festlegung und Einmessen von 20 Referenzpositionen. Bei der Festlegung der Positionen wurde berücksichtigt, dass diese eine möglichst große Vielzahl an möglichen Positionen (die im realen Betriebsalltag auftreten können) abdecken.
3. Positionieren des Tags an den Referenzpositionen und Aufzeichnung von 100 Messungen (Berechnung der Position des Tags durch das System) an jeder Referenzposition

[3] Leica DISTO S910 mit P2P-Package

4. Abschalten einzelner Satelliten, um die Satellitendichte zu reduzieren und die Topologie zu verändern. Die einzelnen Topologien für die jeweilige Messreihen sind in der Abbildung 5 aufgeführt.
5. Wiederholen der Messungen (100) für alle 20 Referenzpositionen für alle Satellitentopologien.

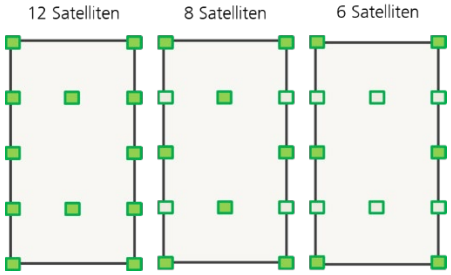


Abbildung 5: Reihenfolge der Abschaltung der Satelliten für die Messfelder: Parkplatz, SmartFactoryOWL, Bilstein - Messfeld 1 (Stehendlager)

Die Messreihe für das Messfeld nahe den Elektromotoren unterscheidet sich vom Messablauf von den vorher beschriebenen Messreihen. Hier wurde eine Messreihe (6 Referenzpositionen mit je 100 Messpunkten) mit laufenden Motoren und eine ohne laufende Motoren durchgeführt.

3.3 Ergebnisse der Genauigkeitsmessung

Die Auswertung der Messreihen betrachtet im Wesentlichen den Messfehler der Positionsbestimmung in Abhängigkeit von der Beschaffenheit des Messfeldes sowie der verschiedenen Satellitentopologien.

Als Messfehler (Positionierungsfehler) wird die Euklidische Distanz zwischen der eingemessenen Referenzposition und der berechneten Position durch das omlox-Sytem in der xy-Achse herangezogen. Für jedes Messfeld in jeder Topologie werden dann über alle Einzelfehler der Mittelwert, der Median, der maximale Fehler und der Fehler für das 95%-Perzentil ermittelt. Die Messwerte sind in der Tabelle 2 aufgetragen.

Für die Charakterisierung der Messfelder werden verschiedene Größen zur Beschreibung der Topologie vorgeschlagen. Die inverse Satellitendichte gibt an, wie viel Quadratmeter des Messfeldes auf einen Satelliten entfallen. Da diese Größe jedoch nicht berücksichtigt, wie gut sich die Satelliten über das gesamte Messfeld verteilen, werden zwei weitere charakterisierende Faktoren vorgeschlagen: durchschnittliche Distanz zu einem Raster (Schachbrettmuster) von Referenzpunkten, welches über das gesamte Messfeld aufgespannt wird sowie durchschnittliche Satellitendistanz, welche das Mittel aller Satellitenabstände zueinander beschreibt. Bei der durchschnittlichen Distanz zu einem Raster wird der durchschnittliche Abstand von jedem Referenzpunkt zu den nächsten drei

Satelliten in der XY-Ebene berechnet. Anschließend wird über alle Referenzpunkte gemittelt, um die durchschnittliche Distanz zu erhalten. Bei der Berechnung wird die Auflösung des Rasters so lange erhöht, bis die durchschnittliche Distanz sich auf 10 cm nicht mehr verändert.

Die Eignung dieser Größen hinsichtlich Zusammenhangs mit dem Positionierungsfehler ergibt sich anhand von Tabelle 2. Durch den Vergleich des Positionierungsfehlers in verschiedenen Umgebungen und Topologien lassen sich eindeutige Ergebnisse ableiten. Mit zunehmender Komplexität der Umgebung verschlechtert sich erwartungsgemäß die Lokalisierungsgenauigkeit. Messungen in der Nähe leistungsstarker Elektromotoren zeigten keine Beeinflussung der Genauigkeit und sind daher in der Tabelle nicht aufgeführt.

Messfeld	**Institutsgelände (ungestörtes Umfeld)**			**SmartFactory OWL (Produktions-umgebung)**			**Bilstein – 1 (PU mit komplexen Störgrößen)**			**Bilstein – 2 (PU mit komplexen Störgrößen)**		
Satelliten	12	8	6	12	8	6	12	8	6	10	8	6
Dichte (m^2/Sat.)	50	76	101	52	79	105	48	71	95	42	52	69
Ø Distanz Ref. (m)	15,4	16,0	17,4	14,2	15,0	16,1	16,5	17,2	18,6	13,4	14,4	15,2
Ø Distanz Sat. (m)	17,9	19,9	23,1	16,5	18,6	21,2	19,7	21,8	24,8	16,3	18,3	20,2
Fehler in m												
Mittelwert	0,09	0,12	0,14	0,16	0,20	0,69	0,30	0,37	1,39	0,54	0,88	1,02
Median	0,08	0,11	0,13	0,12	0,14	0,45	0,20	0,23	0,44	0,34	0,62	0,51
Max	0,27	0,35	0,38	0,80	0,99	6,42	1,25	6,21	7,58	2,13	3,17	3,53
95%-Perzentil	0,22	0,22	0,24	0,50	0,71	2,09	0,83	0,94	6,30	1,55	2,62	2,74

Tab. 2: Messfehler für verschiedene Messfelder und Satelliten-Topologien

Die Verringerung der Satellitendichte führt in den Messreihen innerhalb einer bestimmten Umgebung zu unterschiedlichen Auswirkungen. Während unter idealen Bedingungen (Parkplatz), nahezu keine Verschlechterung feststellbar ist, verstärkt sich der negative Einfluss der Reduzierung mit zunehmender Komplexität der Umgebung, was zu einem erhöhten Positionierungsfehler führt. Dabei zeigt sich, dass mit abnehmenden Satellitendichten signifikant höhere Maximalwerte im Vergleich zum Mittelwert auftreten, was auf eine Zunahme von Ausreißern innerhalb der Messreihe hinweist. Diese Beobachtung wird durch die Divergenz zwischen Mittelwert und Median (welcher robust gegenüber Ausreißern ist) weiter untermauert. Das 95% Perzentil wird einbezogen, da es einen guten Kompromiss zwischen Genauigkeit und Robustheit darstellt. Es gibt an, dass

95% der Positionierungsmessungen innerhalb dieses Fehlerbereichs liegen, was eine verlässliche Einschätzung der Systemleistung ermöglicht. Dadurch können extreme Ausreißer, welche die durchschnittliche Genauigkeit erheblich beeinträchtigen könnten, besser berücksichtigt werden.

Für die Modellbildung wird zunächst ermittelt, welcher der Messfeld beschreibenden Größen (Dichte, durchschnittliche Distanz zum Referenzraster, durchschnittliche Distanz zu Satelliten) am besten mit dem zunehmenden Fehler (95%-Perzentil) in jeder Umgebung korreliert. Aus der Untersuchung geht hervor, dass für alle Messumgebungen eine hohe Korrelation zwischen dem 95%-Perzentil und der durchschnittlichen Distanz zum Referenzraster besteht (Korrelationskoeffizienten 0.93-0.96). Eine Approximation des Messfehlers (95%-Perzentil) in Abhängigkeit von der durchschnittlichen Distanz (Ref) durch eine lineare Funktion ist möglich, liefert insbesondere jedoch Fehler an den gemessenen Datenpunkten (siehe Abbildung 6). Eine kubische Funktion beschreibt den Zusammenhang besser, suggeriert jedoch für die Messumgebung Bilstein-2 in dem Bereich 16,5 – 17,25 m (siehe Abbildung 6 (C)) eine Reduktion des Fehlers, was physikalisch nicht plausibel ist. Eine kubische Hermite-Interpolation ermöglicht sowohl eine Fehlerfreiheit an den Datenpunkten als auch eine realistische Abbildung des physikalischen Verhaltens (ansteigender Fehler bei Erhöhung der durchschnittlichen Distanzen). Sie verwendet hierzu stückweise kubische Polynome, um eine glatte Kurve durch eine Reihe von Punkten zu legen [DV80].

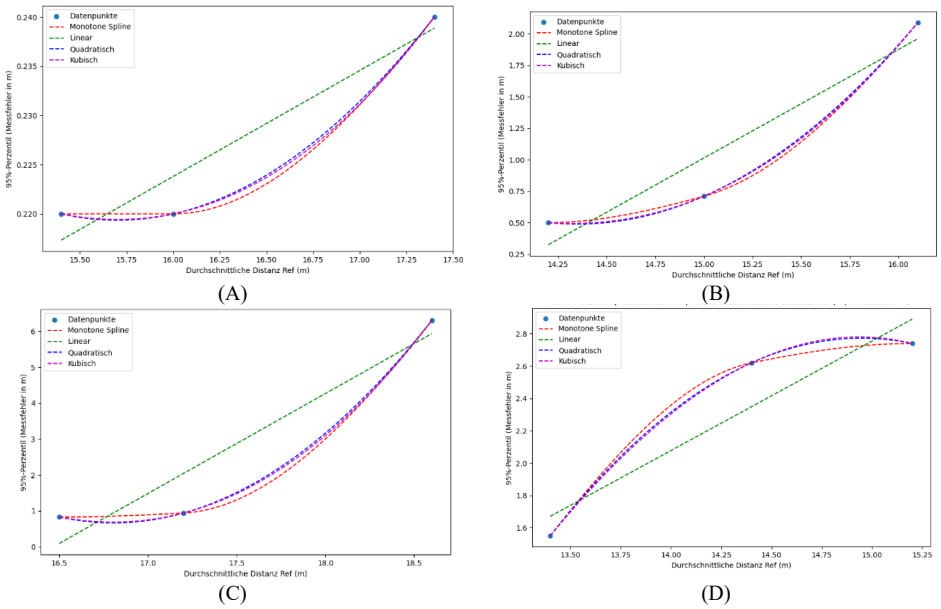


Abbildung 6: Modellierung verschiedener Approximationsfunktionen für die verschiedenen Umgebungen: (A) Institutsgelände, (B) SmartFactoryOWL, (C) Bilstein - 1, (D) Bilstein - 2

3.4 Vorgeschlagenes Vorgehen

Ausgehend von den Ergebnissen dieser Arbeit wird folgendes Vorgehen für die Auslegung einer omlox-Infrastruktur vorgeschlagen: Zunächst sollte das spezifische Umfeld der zu installierenden Umgebung identifiziert werden, wozu Tabelle 1 dieser Veröffentlichung herangezogen werden kann. Anschließend kann unter Verwendung der in Abbildung 6 dargestellten Approximationsfunktionen für das vorliegende Umfeld und die angestrebte Genauigkeit die erforderliche durchschnittliche Distanz der Satelliten zum Referenzraster bestimmt werden. Die Satelliten sollten dann derart positioniert werden, dass diese durchschnittliche Distanz eingehalten wird. Als Anhaltspunkt kann hierzu die in Tabelle 2 dargestellte Satellitendichte in Abhängigkeit von der jeweiligen durchschnittlichen Distanz (Ref) herangezogen werden, welche eine Rückrechnung auf die Abstände zwischen den Satelliten und somit auch dem Gesamtbedarf an Satelliten für die Lokalisierungsumgebung ermöglicht.

4 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Vorgehensmodell zur Bewertung und Planung omlox-basierter Indoor-Lokalisierungssysteme in Industrieumgebungen vorgestellt und validiert. Ziel war es, die zu erwartende Lokalisierungsgenauigkeit unter verschiedenen Störeinflüssen abzuschätzen und eine fundierte Auslegung der Satelliteninfrastruktur zu ermöglichen. Hierzu wurde ein zweistufiges Konzept entwickelt, welches zum einen eine qualitative Bewertung der Produktionsumgebung umfasst und zum anderen für das identifizierte Umfeld einen Prozess bereitstellt, welcher eine Auslegung der Satelliteninfrastruktur auf Grundlage der geforderten Genauigkeit ermöglicht. Es wird hierfür die durchschnittliche Distanz der Satelliten als Berechnungsmaß zugrunde gelegt auf Basis einer kubischen Hermite-Interpolation eine Genauigkeitsabschätzung durchgeführt. Ergebnis ist eine zuverlässige Prognose der erreichbaren Lokalisierungsgenauigkeit für das 95%-Fehlerperzentil. Die Anwendung des Vorgehensmodells kann somit Unternehmen eine präzisere Planung und Implementierung omlox-basierter Indoor-Lokalisierungssysteme ermöglichen, mit dem Ziel die Effizienz und Zuverlässigkeit von Produktions- und Logistikprozessen zu erhöhen.

Für zukünftige Arbeiten wird empfohlen, das Modell weiter zu verfeinern, als Planungs-Software umzusetzen und auf zusätzliche Umgebungen und Szenarien anzuwenden, um die Robustheit und Allgemeingültigkeit zu erhöhen. Es wäre auch sinnvoll, die Integration weiterer Technologien wie Machine Learning zu untersuchen, um die Vorhersagegenauigkeit weiter zu verbessern.

Danksagung

Die Autoren dieses Beitrages möchten sich bei der Firma Bilstein Group aus Hagen bedanken, welche Ihre Produktionsumgebung für Messungen im Rahmen dieses Beitrages zur Verfügung gestellt und die Messungen vor Ort unterstützt hat.

Literaturverzeichnis

[DV80] Dougalis, Vassilios A.; Serbin, Steven M.: On the Superconvergence of Galerkin Approximations to Second-Order Hyperbolic Equations. In: SIAM Journal on Numerical Analysis, 1980.

[FH15] Flatt, Holger; Koch, Nils; Guenter, Andrei; Röcker, Carsten; Jasperneite, Jürgen: A Context-Aware Assistance System for Maintenance Applications in Smart Factories based on Augmented Reality and Indoor Localization. In: IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2015), Luxembourg, Sep. 2015.

[HF23] Hufen, Florian; Siekmann, Timo; Fast, Harry; Flatt, Holger; Schriegel, Sebastian: Positionierung und Vermessung von Komponenten für Indoor-Lokalisierungs- und drahtlose Kommunikationssysteme in Industrieumgebungen. In: Kommunikation in der Automation (KommA 2023), Magdeburg, Nov 2023.

[HS22] Hayward, S.J.; van Lopik, K.; Hinde, C.; West, A.A.: A Survey of Indoor Location Technologies, Techniques and Applications in Industry. In: Internet of Things, Volume 20, 100608, Elsevier, Nov. 2022.

[HI20] Hausladen, Iris: Grundlagen der IT-gestützten Logistik. In: Iris Hausladen (Hg.): IT-gestützte Logistik. Wiesbaden: Springer Fachmedien Wiesbaden, S. 1–28. 2020.

[PI23] PROFIBUS Nutzerorganisation: omlox V2 Core Zone Compliance Test Specification. In: https://www.profibus.com, Version 2.0.0, Order No.: 20.132, Jun. 2023.

[SA22] Schjørring, Allan; Cretu-Sircu, Amalia Lelia; Rodriguez, Ignacio; Cederholm, Peter; Berardinelli, Gilberto; Mogensen, Preben: Performance Evaluation of a UWB Positioning System Applied to Static and Mobile Use Cases in Industrial Scenarios. In: Electronics 11 (20), S. 3294. DOI: 10.3390/electronics11203294. 2022.

[SB21] Schulte, Bastian; Fast, Harry; Flatt, Holger: Lokalisierung von mobilen intelligenten Werkstückträgern innerhalb einer vollautomatisierten industriellen Fertigungsanlage mittels Bluetooth 5.1. In: Automation 2021, digital, Jun. 2021.

[ZW22] Zhao, Wenda; Goudar, Abhishek; Schoellig, Angela P.: Finding the Right Place: Sensor Placement for UWB Time Difference of Arrival Localization in Cluttered Indoor Environments. In: IEEE Robot. Autom. Lett. 7 (3), S. 6075–6082.

[ZZ16] Zhang, Zhaoyu; Di, Xin; Tian, Jun; Chen; Pei: A WLAN planning method for indoor positioning system. In: 2016 International Conference on Information Networking (ICOIN), pp. 303-307, Kota Kinabalu, Malaysia, 2016.

Evaluierung von öffentlichen Mobilfunknetzen als Rückfallebene von privaten 5G Netzwerken für autonomen Bahnverkehr

Timo Siekmann[1], Pragya Agarwal[2], Philipp Alda[3]

Abstrakt: In diesem Beitrag wird die Evaluierung von öffentlichen Mobilfunknetzen als Rückfallebene für 5G-Campusnetzte in autonomen Schienenfahrzeugen betrachtet. Dies geschieht anhand der Forschungsfrage: Unter welchen Bedingungen öffentliche Netzwerke die Hauptkommunikation von dedizierten Netzwerken im Ausfallfall übernehmen können? Es wird dabei im ersten Schritt beschrieben, wie die Anforderungen an das Kommunikationssystem festgelegt werden, indem für die beiden Betriebsarten ATO und RTO jeweils Werte für relevante Dienste (benötigter Uplink/Downlink, Latenzen) festgelegt werden können. Diese Methodik wird durch eine Performance Messung mit einem speziellen Testzug der DB-Systemtechnik GmbH in Minden im 5G Campusnetzwerk und auf der Strecke zwischen Minden und Hannover evaluiert.

Keywords: Öffentliches Mobilfunknetz, private 5G Netzwerke, Automatic Train Operation, Remote Train Operation

1 Einleitung

Die Automatisierung und Digitalisierung des ÖPNV steht aufgrund ihrer Vorteile zunehmend im Fokus der Transportindustrie [1]. Eine wichtige Komponente dabei sind fahrerlose Züge. Diese können das Risiko menschlicher Fehler verringern und gleichzeitig die Effizienz und Sicherheit des Schienenverkehrs verbessern [2]. Für fahrerloses Züge werden vor allem zwei verschiedene Betriebsarten unterschieden, einmal RTO (Remote Train Operation) sowie die ATO (Automatic Train Operation). Beide Betriebsarten stellen hohe Anforderungen an die Kommunikationstechnik zwischen den Fahrzeugen, der Infrastruktur und Kontrollstellen. [3]

Während im ATO Betrieb das Fahrzeug selbst für die Herstellung und Aufrechterhaltung eines sicheren Zustandes verantwortlich ist und dementsprechend die Kommunikation weniger sicherheitskritisch ist, sind insbesondere im RTO Betrieb funktional sichere Verbindungen zwischen Schienenfahrzeug, Fahrdienstleiter sowie Remote Operator

[1] Fraunhofer IOSB-INA, Drahtlose Kommunikationstechnik, Campusallee 1, 32657 Lemgo, timo.siekmann@iosb-ina.fraunhofer.de
[2] *T-Systems* International GmbH, BU Rail&Public Transport, Am Gut Wolf 9a, 52070 Aachen, pragya.agarwal@t-systems.com
[3] *T-Systems* International GmbH, BU Rail&Public Transport, Am Seestern 3, 40547 Düsseldorf, philipp.alda@t-systems.com

essentiell. Private 5G Netzwerke, wie 5G Campusnetzwerke oder zukünftig das FRMCS, versprechen die Erfüllung von enormen QoS Anforderungen. [4]

Um jedoch eine robuste Betriebsfähigkeit der Verkehrsmittel zu gewährleisten und im Falle von Kommunikationsstörungen reagieren zu können, muss untersucht werden, ob öffentliche Mobilfunknetzwerke in einem solchen Störfall als Rückfallebene genutzt werden können.

2 Theoretische Evaluierung von öffentlichen Mobilfunknetzen als Rückfallebene

Im Zuge der theoretischen Evaluierung wurden die Definitionen von zwei Elementen der Untersuchung benötigt. Einerseits wurden die fachlichen Dienste definiert, die mittels des Kommunikationssystems übertragen werden sollen. Andererseits wurden die Charakteristika der Mobilfunktechnologien in diesem Zusammenhang recherchiert und für den Rahmen dieser Untersuchung festgelegt. Die Evaluierung basiert auf der Annahme, dass Mobilfunkfunktionalitäten wie die Netzwahl, die Zellenauswahl usw. nicht betrachtet werden müssen. Die Fragestellung wird aus der Sicht eines (Entwicklers/Produzenten eines) Bordrechners/Zugrechners/On-Bord-Unit eines autonomen Fahrzeugs (MONOCAB) betrachtet, dessen On-Bord-Unit erhält von einer anderen Einheit im Fahrzeug, bspw. dem Router, lediglich die Information welches Netz gerade ausgewählt wurde.

2.1 Kommunikationsanforderungen von ATO und RTO

Die Anforderungen an die Kommunikationstechnik, welche sich aus dem Betriebsmodi GoA (Grade of Automation) 3 und GoA4 ableiten lassen, erfordern eine sichere und zuverlässige Kommunikationstechnologie. Insbesondere das zukünftige FRMCS benötigt eine flächendeckende Kommunikation im gesamten ETCS fähigen Schienennetz. In dieser Veröffentlichung beziehen sich die Kommunikationsanforderungen zwar auf das MONOCAB, dessen Anforderungen wurden aber von Vollbahn abgeleitet. Das MONOCAB ist ein autonomes, kreiselstabilisiertes Einschienenfahrzeugs, das auf wenig frequentierten oder stillgelegten Eisenbahnstrecken eingesetzt wird, um ländliche Gebiete besser zu bedienen. Das Fahrzeug operiert im autonomen Modus (ATO/GoA-4), führt Fahraufträge eigenständig aus und verarbeitet sicherheitsrelevante Informationen selbstständig. Externe Umgebungsinformationen wie Position anderer Fahrzeuge und Wetterdaten können zur Verbesserung des Fahrkomforts und der Systemkapazität genutzt werden, sind jedoch nicht sicherheitskritisch. Wesentliche Dienste im autonomen Betrieb umfassen Netzwerkmonitoring und Logistik/Disposition. Für den Fall, dass das autonome System versagt, ist ein Fernsteuerbetrieb (RTO) vorgesehen, bei dem eine Leitstelle das Fahrzeug aus der Ferne steuert. Dieser Modus ist für Notfälle und niedrigere Geschwindigkeiten (bis 6 km/h) vorgesehen. Er erfordert eine Echtzeitübertragung

kritischer Informationen mit hoher Zuverlässigkeit. Im Fernsteuerbetrieb entfallen Logistik und Disposition zugunsten direkter Fahrbefehle und visueller Informationen für den Bediener. Die wichtigsten Dienste sind hier Video für die Fernsteuerung und Fernsteuerungsbefehle (PLC/Steuerung). Basierend auf Expertenschätzungen und Literaturrecherche werden an dieser Stelle Werte für die Dienste definiert, die die Grundlage der Untersuchung bilden. Die in Tabelle 1 abgebildeten Dienstdefiniotionen bilden nur einen Teil der gesamten Dienste ab. An dieser Stelle werden nur die Dienste, die für den RTO- und ATO-Betrieb unerlässlich bzw. zwingend erforderlich sind aufgeführt. Auf (Komfort-)Dienste wie einen Hotspot für Passagiere kann verzichtet oder dessen Datenrate stark reduziert werden.

Tabelle 1: Dienstdefinitionen [5]

Datenprofil	Datenrate/ DL	Datenrate/ UL	Latenz RTT
Netzwerk Monitoring	100 kBit/s	100 kBit/s	< 50 ms
Logistik u. Disposition	100 kBit/s	100Kbit/s	<50 ms
PLC/ Steuerung	50 kBit/s	50 kBit/s	< 30 ms
Video Fernsteuerung	2 x 100 kBit/s	2 x 2765 kBit/s	< 100 ms

2.2 Charakterisierung von Mobilfunknetzwerken

Im Projekt 5G SIMONE wurden innerhalb des Konsortiums die wichtigsten Mobilfunknetz-Technologien festgelegt. Aus diesem Grund wurde bspw. „3G“ nicht mehr betrachtet, da es in Deutschland zum Zeitpunkt der Erstellung dieser Ausarbeitung keine Rolle mehr im öffentlichen Mobilfunk spielt. Zur Untersuchung wurden folgende Netze herangezogen: 5G, LTE/4G+, 4G, 2G

Dafür wurden für die vier festlegten Netztypen die für diese Untersuchung relevanten Charakteristika festgelegt und Werte hierfür ermittelt. In diesem Fall Uplink (UL), Downlink (DL) und Latenz (LAT). Deren Werte wurden in Maximal-, Minimal- und typischer Wert aufgefächert.

- Der Maximalwert (Max), ist der Wert, der mit dieser Technologie derzeit möglich ist. Im Zusammenhang mit der Latenz wird dieser Fall, aufgrund der umgekehrten Interpretation der Zahlenwerte, als Best Case (BC) bezeichnet.

- Der Minimalwert (Min) ist die Untergrenze, von der bei dieser Technik realisitisch ausgegangen werden kann. Im Zusammenhang mit der Latenz wird dieser Fall, aufgrund der umgekehrten Interpretation der Zahlenwerte, als Worst Case (WC) bezeichnet. Unterhalb dieses Wertes kann davon ausgegangen werden, dass die verantwortliche, hier nicht untersuchte, Einheit dieses Netz nicht auswählen würde.
- Mit dem typischen oder durchschnittlichen Wert (Avg) wurde versucht, einen Wert festzulegen, welcher Wert typischerweise erwartet werden kann in einem öffentlichen Mobilfunknetz. Dieses Mobilfunknetz ist in dem Fall normal ausgelastet, normal viele eingewählte Endgeräte und das eigene Gerät besitzt keine Bevorrechtigung. Der typische beobachtete Wert berücksichtigt verschiedene Faktoren, die nicht gemessen werden können, z.B. Netzwerkzuverlässigkeit und -stabilität.

Als Ergebnis der Recherche ist Tabelle: 2 entstanden.

	UL Min	UL Avg	UL Max	DL Min	DL Avg.	DL Max	LAT WC	LAT Avg.	LAT BC
5G	12 Mbps	35 Mbps	10 Gbps	100 Mbps	260 Mbps	20 Gpbs	10 ms	4 ms	1 ms [6]
LTE/ 4G+	2 Mbps	8 Mbps	1 Gbps	30 Mbps	150 Mbps	1 Gbps	100 ms [7]	25 ms	10 ms
4G	2 Mbps	3 Mbps	50 Mbps	0,2 Mbps	10 Mbps	150 Mbps	100 ms	25 ms	10 ms
2G	9,6 Kbps [8]	10 Kbps	40 Kbps	9,6 Kbps	10 Kbps	100 Kbps	1000 ms	310 ms	300 ms

Tabelle: 2 Netzwerkeigenschaften

2.3 Vergleichsergebnisse

Im Anschluss an die Definition der Dienste und Charakterisierung der Mobilfunknetze, ergänzt durch die Festlegung der Betriebsfälle wurden die Daten zusammengeführt. Hierbei wurden im ersten Schritt die aggregierten Datenraten mit den verfügbaren Bandbreiten der Netzwerke verglichen. Im Best Case/Beim Maximalwert eines Netzwerksstandards zeigten sich alle, bis auf das 2G-Netzwerk, in der Lage die geforderten Dienste bearbeiten zu können. Im Worst Case/Beim Minimalwert der Netzwerkstandards war weiterhin das 5G-Netzwerk vollumfänglich in der Lage, die geforderten Dienste abzubilden. Während das 2G-Mobilfunknetz erwartungsgemäß hier ebenfalls keine Umsetzung der Dienste zuließ, erfüllten die 4G- und LTE/4G+-Netzwerke Teile der Anforderungen. Durch die Ergänzung im Average Case/durchschnittlicher Wert,

bei dem die 4G- und LTE/4G+-Netzwerke deutliche Verbesserungen aufwiesen, wurde abgeleitet, dass durch Verändern der Dienstdefinition an diesen möglicherweise die genannten Netzwerke ebenfalls als Rückfallebene dienen können.

2.4 Mögliche Business Rules zur Nutzung von öffentlichen Mobilfunknetzen als Rückfallebene

Business Rules stellen dabei Regeln dar, die während der Entwicklung definiert werden können. Diese Regeln sollen es einem Gerät, wie beispielsweise einem Bordcomputer, ermöglichen, Entscheidungen zu treffen, welche Maßnahmen in bestimmten Szenarien ergriffen werden sollen. Der erste untersuchte Fall befasst sich mit der Frage, ob die Latenzen verringert werden können. Hierzu wurden zunächst Orientierungswerte für Latenzen herangezogen, wobei eine Verzögerung von bis zu 200 Millisekunden (ms) als akzeptabel für Echtzeitwahrnehmung angenommen wurde. Diese Annahme basiert auf akustischen Signalübertragungen und wird auch für visuelle Wahrnehmungen sowie die SPS-Fernsteuerung übernommen. Im weiteren Verlauf wurde die Relevanz dieser Latenzen für die Verkehrssicherheit des Monocab-Fahrzeugs untersucht. Dabei fokussierte man sich auf die Bremswege, die in Reaktionszeit/-weg und Anhaltezeit/-weg unterteilt sind. Referenzwerte aus dem Straßenverkehr und Straßenbahnbetrieb wurden verwendet, um eine Reaktionszeit von 1 Sekunde und eine Bremsverzögerung von 2,75m/s² anzunehmen. Beim Fernsteuerbetrieb des Monocab, das mit einer Höchstgeschwindigkeit von 6 km/h operiert, wurde ein Regelbremsweg von 2,18m errechnet. Unter Berücksichtigung einer zusätzlichen Latenzstrecke von 0,6m erhöht sich der Bremsweg um ca. 27,5%. Um diese Verlängerung zu kompensieren, wurde eine Reduzierung der Höchstgeschwindigkeit auf 5 km/h analysiert, was den Bremsweg auf 2,25m reduzieren würde. Weitere Reduzierungen der Geschwindigkeit, etwa auf 3 km/h, würden die Auswirkungen der Latenz vollständig ausgleichen und den Einsatz von 4G-Netzen im Notfallbetrieb ermöglichen. Der Netzwerkmonitoring-Dienst ist bezüglich der Latenz entscheidend. Die festgelegten 50 ms Latenz entsprechen typischen 4G-Netzwerten. Selbst bei höheren Latenzen von bis zu 100 ms im Notfallbetrieb würde dies keine sicherheitskritischen Auswirkungen haben, da der Informationsgehalt dieser Dienste nicht sicherheitsrelevant ist. Im autonomen Betrieb wären Latenzen ebenfalls tolerierbar, da dispositive Informationen wie Logistik und Disposition nicht sicherheitskritisch sind. Zusammengefasst können 4G-Netze als Rückfalllösung dienen, wenn die Latenzanforderungen in bestimmten Rahmen reduziert werden, um den sicheren Betrieb sowohl im autonomen als auch im Fernsteuerbetrieb zu gewährleisten. Der zweite Fall befasst sich mit der Frage, ob der Uplink im RTO verringert werden kann. Einsparungen können beispielsweise durch Reduzierung der Videoqualität beim „Video für Fernsteuerung“ erreicht werden, etwa durch Verringerung der Auflösung und Framerate der Kameras. Dadurch kann die Uplink-Bandbreite auf 1,2 Mbps reduziert werden, was den Minimalanforderungen der 4G-Netze entspricht.

3 Beschreibung der Testumgebung und der Testkonfiguration

In diesem Kapitel werden die Testumgebung und die Testkonfiguration beschrieben. Ziel ist die Prüfung und Bewertung der Einhaltung der Kommunikationsanforderungen und der Grad der Abweichung von diesen Anforderungen. Umschaltzeiten oder Redundanzprotokolle wurden in dieser Arbeit nicht näher betrachtet, sollen aber in zukünftigen Arbeiten untersucht werden.

Die Messungen wurde auf dem RailCampus OWL (www.railcampusowl.de) in Minden sowie im öffentlichen Schienennetz der DB Netz zwischen Minden und Hannover durchgeführt. Hierzu lag eine Sondererlaubnis für eine Messfahrt zwischen dem Dienstgelände der DB-Systemtechnik GmbH sowie dem Bahnhof in Minden vor. Auf der Strecke zwischen dem der DB Systemtechnik und dem Bahnhofe Minden war nur das öffentliche Mobilfunknetzwerk verfügbar. Hinzu kommt, dass nur das öffentliche Mobilfunknetz der deutschen Telekom untersucht wurde. Weitere MNOs (Mobile Network Operator) können in Folgenden Arbeiten untersucht werden. Die Distanz zwischen der Strecke beläuft sich auf 53 km.

Abbildung 1: Bahnstrecke zwischen Minden und Hannover

Der Testaufbau in dem Testzug umfasste eine Fraunhofer Mobilfunkmessbox auf dem Zug Dach (siehe Abbildung 2), welche Antennentechnik, sowie 5G Modems beinhaltete, der über ein Netzwerk und Stromverbindung mit dem inneren des Wagenastens verbunden war. Im Wagenkasten waren verschiedene Hardwarekomponenten wie RGB und IR-Kameras, als auch Traffic Generatoren mit dem Mobilfunknetzwerk verbunden.

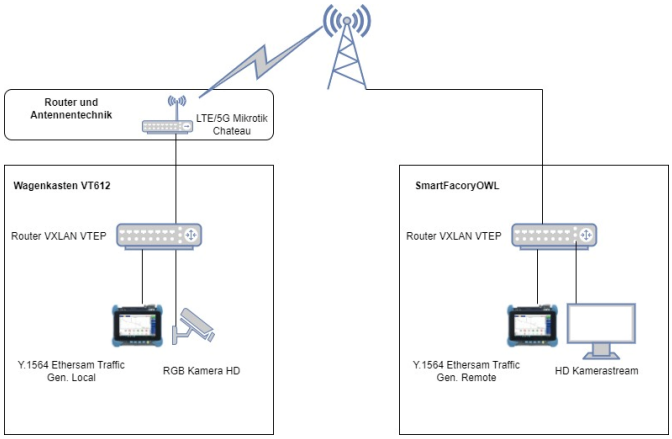


Abbildung 2: Messaufbau und Netzwerktopologie

4 Messergebnisse

Die Messergebnisse wurden an vereinzelten Positionen im parkenden Fahrzeug v=0 km/h aufgenommen. So können diese Messwerte einer genauen GPS Position sowie einer Mobilfunkzelle zugewiesen werden. Zudem basiert der V.1564 Ethersam Test, darauf das die einzelnen Messungen mit den gleichen Netzwerkwerk Bedingungen durchgeführt werden. Die Messwerte wurden im Campusnetzwerk als auch im öffentlichen Netzwerk unter den gleichen Bedingungen aufgenommen.

Dabei zeigen die Messergebnisse, dass das öffentliche 4G Mobilfunknetz dem 5G Campusnetz im besten Fall um den Faktor 5 unterlegen ist. Somit können auch die zuvor

eingeführten Anforderungen aus Tabelle 1 mit dem öffentlichen Netzwerk nicht eingehalten werden.

Die Folgenden Grafiken zeigen die Latenzwerte der verschiedenen Applikationen zuerst im 5G Campusnetz und in der nächsten Abbildung im öffentlichen Mobilfunknetzwerk.

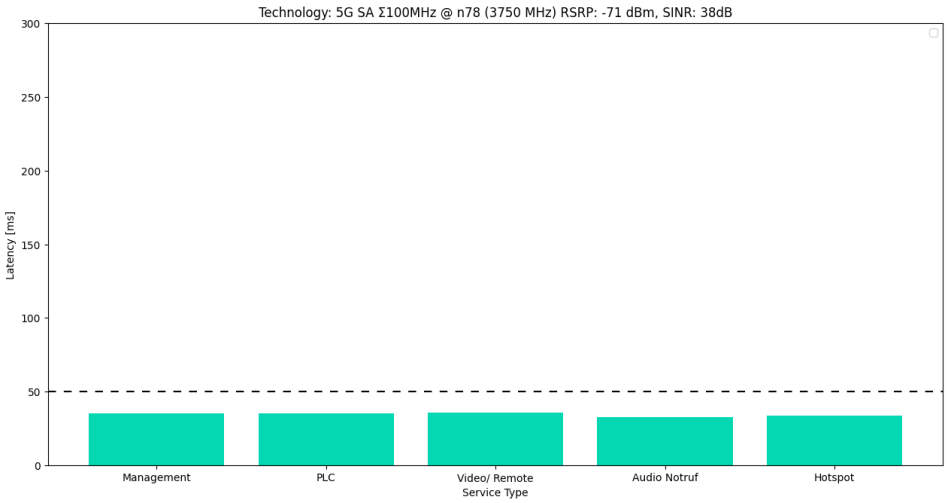


Abbildung 3: Latenzwerte der Datenprofile 5G Campusnetzwerk

Die Messung wurde im Frequenzband n78 bei einer Center Frequency von 3750 MHz ausgeführt. Für die Messung wurde die volle Bandbreite von 100 Mhz der Kurzzeit Frequenzzuteilung der BNetzA genutzt. Sowohl der Referenzwert für das Empfangssignal an der Bahn als auch der SINR (Signal Interference Noise Ratio) waren bei der Messung im sehr guten Bereich. Das UE im öffentlichen Mobilfunknetzwerk hatte einen deutlich schlechteren Funkkanal zur Verfügung. So war der RSRP Wert um 30dB schlechter und der SINR Wert mit 0 dB um 38 dB schlechter. Trotzdem wurden an diesem Messpunkt im öffentlichen Mobilfunk die besten Latenzwerte für die einzelnen Anwendungen erzielt. Die Latenzwerte (Abbildung 5) für das Videoprofil lag mit über 150 ms, bei dem dreifachen der oben eingeführten Anforderung. Auch das Profil für die Steuerbefehle sowie das Netzwerk Management wurde nicht den Anforderungen gerecht.

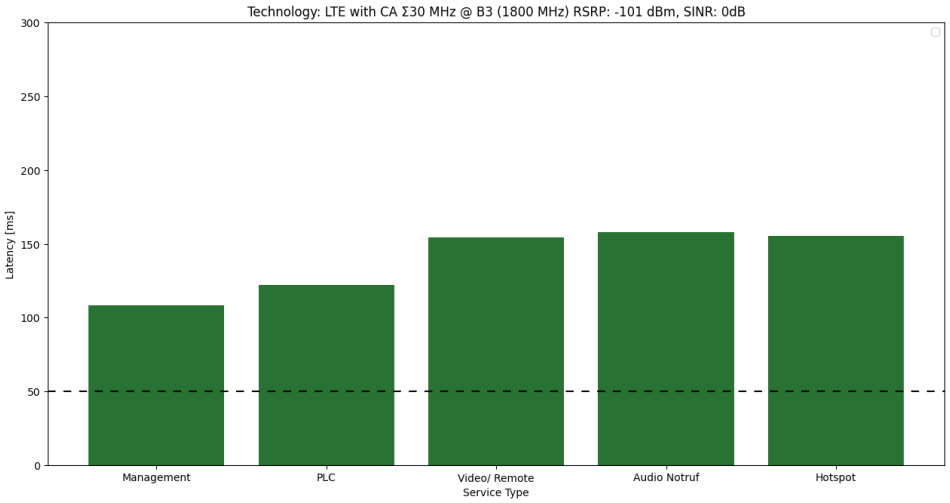


Abbildung 4: Latenzwerte Datenprofile öffentlicher Mobilfunk

Die Leistungsunterschiede zwischen dem Campusnetz und dem öffentlichen Mobilfunknetzwerken sind auf verschiedene Gründe zurückzuführen.

Frequenzspektrum:

Im öffentlichen Mobilfunknetzwerk hingegen bekommt das UE weniger Frequenbandweite zugewiesen, da den Mobilfunkoperatoren weniger Frequenzband zur Verfügung steht. Besonders zusammenhängendes Frequenzband ist sehr rar, daher benutzen eine Vielzahl der öffentlichen Mobilfunknetze die FDD (Frequency Divsiosn Duplexing). Das bedeutet, dass der Downlink und der Uplink zur gleichen Zeit auf unterschiedlichen Frequenzen senden kann. Die BNetzA schreibt für die Campusnetze in Deutschland hingen eine Nutzung des TDD (Time Divison Duplexing) vor. In dieser Technologie senden und empfangen Funkteilnehmer auf der gleichen Frequenz, aber in unterschiedlichen Zeit Slots.

Beide Verfahren haben verschiedene Vor- und Nachteile und werden weltweit von Mobilfunkoperatoren genutzt. Ein Betriebsart bedingte Besonderheit ist, dass man zum Betrieb von FDD immer ein gepaartes Spektrum benötigt, da Up und Downlink jeweils eine eigene Frequenz benötigten. Carrier Aggregation ist eine weitere Technologie die im LTE-Advanced Standard (> 3GPP Rel. 11) eingeführt wurde. Carrier Aggregation löst das Problem der nur kleinen Frequenzblöcke, welche durch Mobilfunkoperatoren bei der BNetzA ersteigert wurde. Mittels der Technologie können mehre BWP (Bandwidth Parts) zu einem großen nutzbaren Spektrum zusammengeführt werden. Der 3GPP Standard sieht eine maximale 5 x 20 MHz, also in Summe 100 MHz zusammenhängendes Spektrum vor. Theoretisch ist mit dieser Technologie eine Datenrate von 1Gbit/s möglich.

Während der 56 km langen Messfahrt wurden Maximal Geschwindigkeiten von 160 km/h gefahren und das UE war mit insgesamt 126 Mobilfunkzellen verbunden. Einige dieser Verbindungen wurden exemplarisch in diesem Paper untersucht. So ist im Datensatz zu erkennen, dass Carrier Aggregation in verschiedenen Bandkombinationen angewandt wurde. Die LTE Technologie arbeitet in verschiedenen Frequenzbändern, die Band Nomenklatur umfasst Bänder mit der Bezeichnung startend mit B1.Während der Messfahrt war das UE mit fünf verschiedenen Bändern verbunden. Den größten Anteil der Verbindungen sind über das Band B1 (2100 MHz) gefolgt von Band B8 (900MHz) und B3 (1800 MHz), abgewickelt worden.

Distanz zwischen dem eNodeB und dem UE:

So zeigt sich in der Abbildung 6, das der Abstand zwischen dem eNodeB und dem UE über 1,27 km beträgt. In diesem Bereich wird das Funksignal aufgrund von Mehrwegeausbreitung sowie Signaldämpfung verschlechtert. Das 5G Campusnetz wurde mit einem kleineren Zellradius betrieben. Somit sind der Empfangspegel und die Qualität deutlich gesteigert.

Abbildung 5: Distanz zwischen eNodeB und UE

Downlink Optimierung im öffentlichen Mobilfunknetzwerken:

In öffentlichen Mobilfunknetzwerken ist der Datenverkehr Downlink orientiert. Daher sind die Netzwerke dementsprechend ausgelegt. Da die im MONOCAB erzeugten Videobilder jedoch in der Uplink Richtung übertragen werden müssen, wird ein Bottleneck erreicht, welches die Datenrate und Latenz Uplinkseitig limitiert.

Unbekannte Anzahl der Netzwerkteilnehmer öffentlichen Mobilfunknetzwerken:

Eine weitere Unbekannte ist die Zellauslastung und die damit pro Teilnehmer verfügbare Kapazität. Diese geht damit einher, dass dem Nutzer keine Information darüber gegeben wird, wie viele Nutzer gerade parallel die Mobilfunkzelle nutzen.

5 Fazit und Ausblick

Zusammenfassend ergibt sich aus den vorliegenden Analysen, dass öffentliche 5G-Netzwerke im Kontext des Notfallbetriebs nahezu vollumfänglich als zuverlässige Rückfallfunktion dienen können. Die hohe Leistungsfähigkeit und niedrigen Latenzen machen sie besonders geeignet, auch kritische Anforderungen zu erfüllen. In Ergänzung dazu zeigen 4G und LTE/4G+ ihre Unterstützungsfähigkeiten, allerdings mit gewissen Einschränkungen und der Notwendigkeit von Dienstverringerungen, um als effektive Rückfallfunktionen zu agieren.

6 Literaturverzeichnis

[1] A. Denisenkov, N. Denisenkova und Y. Polyakova, „Digital transformation of ransport infrastructure: experience of European and Moscow metro systems,“ in *EU and its Neighbourhood: enhancing EU actorness in the Eastern Borderlands*, EURINT 7, 2020, pp. 303-325.

[2] M. E. Lopez-Lambas und A. Alonso, „"The driverless bus: An analysis of public perceptions and acceptability.",“ *Sustainability 11.18,* 2019.

[3] J. Yin und e. al, „Research and development of automatic train operation for railway transportation systems: A survey,“ *Transportation Research Part C VOl 85,* pp. 548-572, 2017.

[4] B. Ai, A. Molisch, M. Rupp und Z.-D. Zhong, „5G Key Technologies for Smart Railways,“ *Proceedings of the IEEE (Vol: 108, Issue 6),* 6 2020.

[5] D. Gustin, *Outdoor Field Test of 5G-based V2X Communication for Real-Time Monitoring and Remote Control of a Monorail Vehicle*, Lemgo: IEEE 21st International Conference on Industrial Informatics (INDIN), 2023.

[6] Y. a. L. H. a. L. Z. a. L. Y. a. Q. F. a. G. L. a. X. X. a. X. T. Li, „A nationwide study on cellular reliability: measurement, analysis, and enhancements,“ 2021.

[7] G. e. al, „"Feasibility study of teleoperated vehicles using multi-operator LTE connection.",“ in *International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, IEEE, 2020.

[8] V. G. a. W. B. Z. Hunaiti, „An Assessment of a Mobile Communication Link for a System to Navigate Visually Impaired People,“ *IEEE Transactions on Instrumentation and Measurement*, 2009.

Performance-Evaluation von Device-to-Device Kommunikation via C-V2X

Denis Gustin[1], Mario Schoppmeier[1], Björn Kroll[1] und Timo Siekmann[1]

Abstract: Zukunftsweisende Mobilität geht mit direkter Kommunikation von Fahrzeugen einher. Dieser Beitrag untersucht die Performance der C-V2X-Kommunikation (Cellular Vehicle-to-Everything) durch reale Messungen und analysiert die Zuverlässigkeit bei steigendem Datenverkehr und mehreren aktiven Teilnehmern. Unsere Ergebnisse zeigen, dass selbst bei hoher Kanalbelastung kurze Latenzzeiten von rund 15ms erreicht werden können. Dies ermöglicht eine effiziente Datenverteilung an alle Netzteilnehmer und bietet Potenzial für zukünftige Anwendungen im Straßenverkehr und in der Automatisierungstechnik.

Keywords: 5G NR, C-V2X, C-V2V, C-V2I, Device-to-Device Communication

1 Motivation

Für die Zukunft der Mobilität ist die direkte Kommunikation von Fahrzeugen untereinander unerlässlich. Durch erhöhten Austausch von Informationen unter allen Verkehrsteilnehmern werden die Sicherheit und die Effizienz des Straßenverkehrs erhöht. Fahrzeuge tauschen zusätzliche Informationen aus und können somit dem Fahrer z. B. Zusatzinformationen bereitstellen oder autonome Fahrzeuge unterstützen. Die hierfür genutzte Technologie wird C-V2X (Cellular Vehicle to Everything) als Sammelbegriff bzw. C-V2V (Vehicle to Vehicle) und C-V2I (Vehicle to Infrastructure) Kommunikation genannt. Ermöglicht werden Anwendungen, wie bspw. die Anzeige der Restdauer von Ampelphasen im Fahrzeug, Kommunikation von Notbrems- oder Unfallmeldungen an umgebende Fahrzeuge, Grünschaltung für Einsatzfahrzeuge oder Platooning (virtuelle Vernetzung von Fahrzeugen zu einer Einheit, wodurch das erste Fahrzeug die Richtung / Geschwindigkeit vorgibt und die restlichen automatisiert folgen) [AA24].

Dieser Beitrag wird die Performance der C-V2X-Kommunikation durch reale Messungen evaluieren und analysieren, wie sich die Zuverlässigkeit mit mehreren aktiven Teilnehmern und somit steigendem Datenverkehr entwickelt. Zusätzlich wird dieser Beitrag einen Ausblick geben, wie die Technik weiterhin unter realen Bedingungen skaliert und ob es ggf. noch weitere Anwendungsfelder gibt.

1 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, denis.gustin@iosb-ina.fraunhofer.de; mario.schoppmeier@iosb-ina.fraunhofer.de; bjoern.kroll@iosb-ina.fraunhofer.de; timo.siekmann@iosb-ina.fraunhofer.de

2 Problembeschreibung

Obwohl C-V2X eine Kommunikationstechnologie für Fahrzeuge ist, wird sie von dem 3GPP in der Standardisierung von 4G/LTE und 5G NR mitberücksichtigt, was die Signifikanz unterstreicht. Aber dadurch, dass Sie im Rahmen der Spezifikation direkte Device-to-Device Kommunikation verwendet, anstatt des Uu Interface zwischen Netz und Endgerät, ist kein Koordinator für die Funkressourcen vorhanden. Dies hat zur Folge, dass das Kanalzugriffverfahren entscheidend für die Performance ist und eine Koordination von vielen Geräten im Spektrum erschwert wird. Genutzt wird das ITS 5,9 GHz-Spektrum, welches auf 5G Sidelink basiert [3G16].

Durch die sicherheitsrelevanten Anwendungen ist die Performance der C-V2X- Kommunikation ein entscheidender Aspekt. Besonders bei Notbrems-, Unfallmeldungen und Regelungsaufgaben, wie z. B. Platooning, sind die Latenz und Paketverlustrate bzw. das Auftreten von Retransmissions ausschlaggebend. Dies ist von besonderem Interesse, da, durch fehlende zentrale Koordination des Kanalzugriffes, eine stark sinkende Performance bei steigenden Teilnehmerzahlen und Datenverkehr zu erwarten ist.

Es stellt sich die Frage, ob sicherheits- und regelungsrelevante Nachrichten noch rechtzeitig kommuniziert werden können, um die entsprechenden Anwendungen korrekt auszuführen. Eine zu stark sinkende Performanz würde im Umkehrschluss bedeuten, dass C-V2X nur bei wenigen Verkehrsteilnehmern gleichzeitig einsetzbar und somit nicht unter realen Voraussetzungen einsetzbar wäre.

3 Stand der Technik

C-V2X bezeichnet eine Sammlung von Kommunikationsprotokollen, die speziell für die Kommunikation zwischen Fahrzeugen (V2V), Fahrzeugen und Infrastruktur (V2I), Fahrzeugen und Fußgängern (V2P) sowie Fahrzeugen und Netzwerken (V2N) entwickelt wurden. Diese Technologie zielt darauf ab, die Sicherheit und Effizienz im Straßenverkehr zu erhöhen, indem sie den Informationsaustausch zwischen Verkehrsteilnehmern verbessert. Die Standardisierung von C-V2X erfolgt durch das 3rd Generation Partnership Project (3GPP), das die Protokolle in den Releases 14 bzw. 15 (für LTE-V2X) [3G20] und 16 bzw. 17 (für 5G NR) [3G19] spezifiziert hat.

3.1 Technologische Grundlagen

5G NR V2X ist eine Weiterentwicklung von LTE-V2X, welches bereits eine direkte Kommunikation zwischen Endgeräten ohne die Notwendigkeit einer zellularen Basisstation ermöglicht. Diese Kommunikation erfolgt über dedizierte Frequenzbänder, insbesondere das ITS 5.9 GHz Spektrum, das weltweit für intelligente Verkehrssysteme reserviert ist.

5G NR Sidelink bietet mehrere Vorteile gegenüber älteren Technologien wie DSRC (IEEE 802.11p), darunter eine größere Reichweite, höhere Zuverlässigkeit und bessere Leistung bei Nicht-Sichtverbindungen (NLOS) [AA18].

3.2 Kommunikationsmodell

Ein zentrales Merkmal von 5G NR Sidelink ist die Ressourcenzuweisung, die entweder durch die Endgeräte selbst (dezentrales Management) oder durch die Netzwerk-Infrastruktur (zentrales Management) erfolgen kann. Es gibt mehrere Modi für die Ressourcenverwaltung [3G20]:

Mode 3: Die Ressourcen der Endgeräte werden zentral über ein Mobilfunknetz koordiniert.

Mode 4: Dezentrale Ressourcenzuweisung ohne zentrale Koordination, basierend auf einer Listen-Before-Talk (LBT) Strategie.

Die Herausforderung beim Kanalzugriff ohne zentrale Koordination besteht darin, Interferenzen zu minimieren und eine effiziente Nutzung des Spektrums zu gewährleisten. 5G NR Sidelink nutzt Mechanismen wie Listen-Before-Talk (LBT) und adaptive Modulation und Codierung (AMC), um diese Herausforderungen zu adressieren. Abbildung 1 illustriert die beiden Kommunikationsmodi.

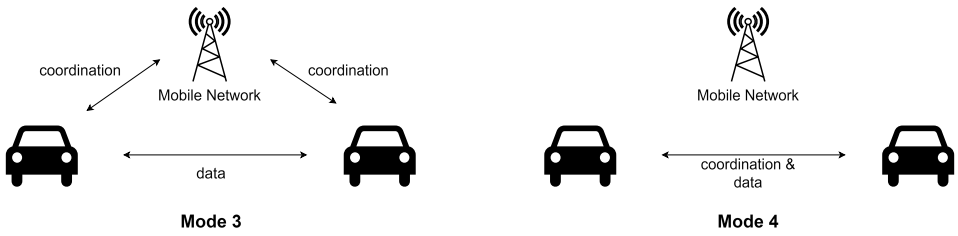


Abb. 1: C-V2X Communication Modes 3 and 4

3.3 Leistungsmerkmale

Die Leistungsfähigkeit von 5G NR Sidelink ist ein wichtiger Aspekt für die Akzeptanz und den Erfolg von C-V2X. Typische Latenzzeiten liegen im Millisekundenbereich und ermöglichen Echtzeitkommunikation, die für sicherheitskritische Anwendungen wie Notbrems- und Unfallmeldungen unerlässlich ist. Die Zuverlässigkeit der Kommunikation wird durch robuste Fehlerschutzmechanismen und Wiederholungsstrategien (HARQ - Hybrid Automatic Repeat reQuest) gewährleistet.

3.4 Vergleich mit anderen Technologien

Im Vergleich zu anderen ITS-Kommunikationstechnologien wie DSRC (IEEE 802.11p) bietet 5G NR Sidelink mehrere Vorteile. DSRC, eine auf WLAN basierende Technologie, wurde ursprünglich für die Fahrzeugkommunikation entwickelt und bietet schnelle Verbindungszeiten und niedrige Latenzen. Allerdings ist die Kommunikationsreichweite von DSRC begrenzt, und die Technologie zeigt Schwächen bei NLOS-Bedingungen.

5G NR Sidelink übertrifft DSRC in mehreren Aspekten [AA18]:

Reichweite: Größere Kommunikationsreichweite, was eine effizientere Informationsverteilung ermöglicht.

Zuverlässigkeit: Höhere Zuverlässigkeit, insbesondere bei NLOS-Bedingungen durch fortschrittliche Fehlerschutzmechanismen.

Flexibilität: Unterstützung einer größeren Bandbreite an Anwendungen durch höhere Datenraten und flexiblere Ressourcenzuweisung.

Die Vorteile von 5G NR Sidelink machen es zu einer vielversprechenden Technologie für die zukünftige Entwicklung von C-V2X-Anwendungen, insbesondere in komplexen und stark frequentierten Verkehrsnetzen.

3.4.1 Nutzdaten

Wie in [Ro18] beschrieben, ist der Zweck der C-V2X Technologie die Übertragung von Nachrichten, die den folgenden Typen entsprechen können:

Cooperative Awareness Messages (CAM): ermöglichen es Fahrzeugen, regelmäßig Informationen über ihre Position, Geschwindigkeit und Richtung auszutauschen, um die Situationswahrnehmung im Straßenverkehr zu verbessern.

Decentralized Environmental Notification Messages (DENM): werden verwendet, um Warnungen über plötzliche oder gefährliche Ereignisse, wie Unfälle oder Hindernisse auf der Straße, in Echtzeit an umliegende Fahrzeuge zu übermitteln.

Map Message / Signal Phase and Timing Messages (MAPEM/SPATEM): liefern Informationen über die Straßeninfrastruktur, wie Kreuzungen und Ampelphasen und ermöglichen es Fahrzeuge, sich sicher und effizient im Verkehrsnetz zu bewegen.

In-Vehicle Information Messages (IVI): übermitteln wichtige Informationen und Warnungen direkt an die Fahrer, beispielsweise Tempolimits, Baustellenhinweise oder spezifische Verkehrsregeln, die unmittelbar im Fahrzeug angezeigt werden.

3.5 Stand der Forschung

Die 5G Atomotive Association (5GAA) hat Vergleichsmessungen zwischen DSRC und C-V2X durchgeführt [AA18]. Hierbei wurde die Überlegenheit von C-V2X gegenüber DSRC festgestellt. Zudem wurden Messungen zur Packet Error Rate (PER) durchgeführt, die verschiedene Szenarien mit und ohne direkte Sichtverbindung abdecken. Devika et al. [DRS23] haben aufgezeigt, dass C-V2X auch für die regelungstechnische Aufgaben Platoonbildung von Elektro-LKW im Vergleich zu DSRC überlegen ist.

Betrachtungen, die die Messungen unter realen Bedingungen durchführen [AA18] sind bisher selten und betrachten dabei eher die Paketfehlerrate.

4 Methodik

4.1 Testaufbau und Messverfahren

Im Rahmen des Testaufbaus wird nur die Übertragung von Datenpaketen mit unterschiedlicher Nutzlast (Payload) in Byte unter variierenden Sendeintervallen und Kanalbelastungen untersucht. Der Messaufbau umfasst dabei drei Szenarien, die jeweils verschiedene Payload-Größen berücksichtigen (Abbildung 1 veranschaulicht den Testaufbau).

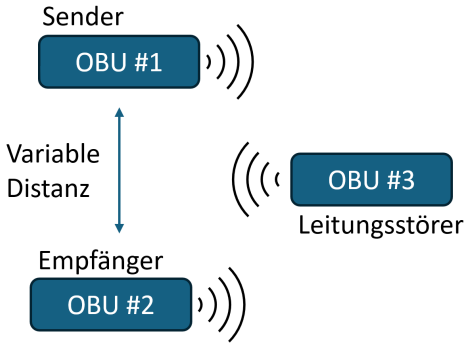


Abb. 2: Testaufbau

Alle Test wurden mit Payloads von 64, 256 und 1500 byte durchgeführt, um einen möglichst großen Bereich möglicher Paketgrößen abzubilden. Zusätzlich wurde in den Tests 2 und 3 jeweils die Kanalbelastung mit weiteren sendenden Geräten erhöht. Die Tabelle 1 zeigt die Tests in der Übersicht. Weitere Einheiten senden mit einem anderen Intervall, um den

Abb. 3: Teststrecke auf dem Innovation Campus in Lemgo

Kanal effektiv gegenüber der Messung zu blockieren. In Messung 3 stört das empfangende Gerät den Kanal ebenfalls.

Messung	ID1 TX	ID2 TX	ID3 TX
1.1	64Byte (10ms)	-	-
1.2	256Byte (10ms)	-	-
1.3	1500Byte (10ms)	-	-
2.1	64Byte (10ms)	-	256Byte (12ms)
2.2	256Byte (10ms)	-	256Byte (12ms)
2.3	1500Byte (10ms)	-	256Byte (12ms)
3.1	64Byte (10ms)	256Byte (12ms)	256Byte (12ms)
3.2	256Byte (10ms)	256Byte (12ms)	256Byte (12ms)
3.3	1500Byte (10ms)	256Byte (12ms)	256Byte (12ms)

Tab. 1: Übersicht der Testkonfigurationen

Die neun unterschiedlichen Messungen wurden unter realen Bedingungen auf dem Innovation Campus in Lemgo durchgeführt. Für jede Messung wurde dieselbe, vorher festgelegte Strecke befahren (siehe Abbildung 3). Während der Tests blieben die Stationen mit den IDs 1 und 3 stationär, während die Station mit der ID 2 mobil war. Cohda Wireless MK6 wird dabei ausschließlich als On-Board Unit eingesetetzt.

5 Auswertung

In diesem Kapitel werden zunächst die aufgezeichneten Messwerte dargestellt und anschließend in einer Diskussion interpretiert.

5.1 Beschreibung der Messergebnisse

Zunächst werden die Latenzwerte der drei Messreihen statistisch ausgewertet. Abbildung 4 zeigt die Histogramme der einzelnen Messreihen und der jeweiligen Paketgrößen auf. Zusätzlich werden die Messwerte in Tabelle 2 zusammengefasst.

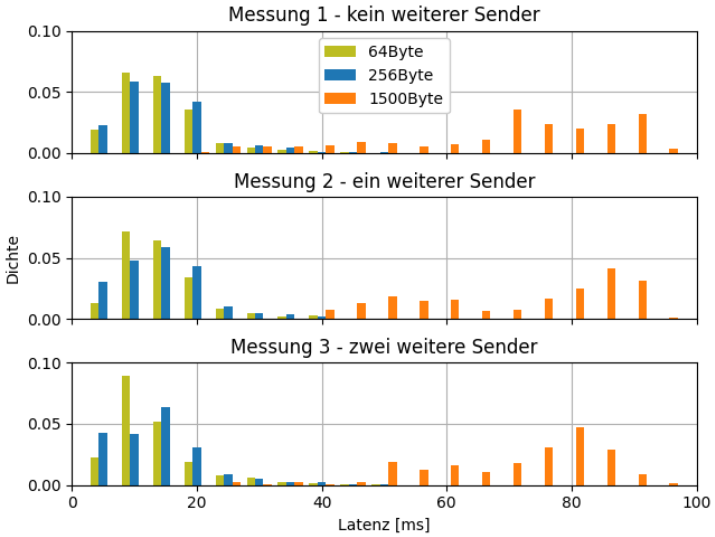


Abb. 4: Histogramm der Latenzen bei verschiedenen Framegrößen und weiteren Sendern

In der ersten Messung, mit nur einem sendenden Gerät, zeigt sich eine Häufung der Latenzwerte für die Paketgrößen 64 und 256Byte bei unter 20ms. Hier sind die Mittelwerte knapp bei ca. 15ms. Die Dichten für beide Paketgrößen unterscheiden sich wenig voneinander. Die meisten Messwerte liegen für beide Paketgrößen in der Klasse zwischen 7,5ms und 12,5ms. Die Paketgröße 1500Byte weicht hiervon stark ab und hat ihren Mittelwert bei 69,7ms. Häufigste Klasse ist hier 67,5ms bis 72,5ms. Latenzwerte unter 17ms sind bei der Paketgröße 1500Byte gar nicht vorhanden, während niedrigsten Latenzen ist bei 64Byte 6,2ms und bei 256Byte 6,0ms sind. Die Maximalwerte hingegen sind bei ca. 51ms für 64 und 256Byte und bei 95,8ms für 1500Byte.

In der zweiten Messung, mit einem zusätzlichen Sender (256Byte alle 12ms), bleibt das Bild ähnlich. Die Mittelwerte der Latenzen bei allen Paketgrößen liegen nach wie vor im gleichen Bereich. Die häufigste Klasse verschiebt sich bei 64Byte auf 7,5ms bis 12,5ms und bei 256Byte auf 12,5ms bis 17,5ms. Für 1500Byte ist die häufigste Klasse 82,5ms bis 87,5ms.

Auch in der dritten Messung, mit insgesamt zwei zusätzlichen Sendern (256Byte alle 12ms), zeigt sich ein größtenteils unverändertes Bild. Hier verschiebt sich nur die häufigste Klasse für 1500Byte Paketgröße auf 77,5ms bis 82,5ms.

Messung	Paketgröße	Mittelwert Latenz	Jitter	Min. Latenz	Max. Latenz
1.1	64Byte	14,4ms	6,6ms	6,2ms	50,8ms
1.2	256Byte	15,1ms	6,8ms	6,0ms	52,1ms
1.3	1500Byte	69,7ms	15,2ms	17,8ms	95,8ms
2.1	64Byte	14,9ms	6,6ms	5,9ms	51,5ms
2.2	256Byte	15,4ms	7,0ms	6,2ms	49,9ms
2.3	1500Byte	70,9ms	13,2ms	17,9ms	95,0ms
3.1	64Byte	13,6ms	7,2ms	6,0ms	48,2ms
3.2	256Byte	14,3ms	7,4ms	5,7ms	48,9ms
3.3	1500Byte	71,1ms	11,9ms	22,9ms	94,9ms

Tab. 2: Übersicht der gemessenen Latenzen

Abbildung 5 zeigt die Latenz der Messwerte in Relation zu der Distanz zwischen Sender und Empfänger für Messung 2. Bei dieser Messung ist leider nur die Auswertung für die Paketgrößen 256Byte und 1500Byte möglich, da bei 64Byte in den Paketen für die Latenzmessung kein Platz mehr für die Positionsdaten des Senders ist. Gut zu erkennen ist, dass sich die Bereiche, in denen die Latenzen für die unterschiedlichen Paketgrößen liegen, fast gar nicht überlappen. Während 99,6% aller Latenzwerte für die 1500Byte Pakete bei über 40ms Latenz liegen, liegen die Pakete für 256Byte zu 99,0% unter 40ms. Die Streuung der Latenzwerte weist augenscheinlich über die Distanz keine Änderung auf. Lediglich bei den 256Byte Paketen treten ab einer Distanz von ca. 180m auch höhere Latenzwerte von über 40ms auf. Tabelle 3 gibt eine Übersicht der Korrelationskoeffizienten zwischen Latenz und Distanz für alle Messungen.

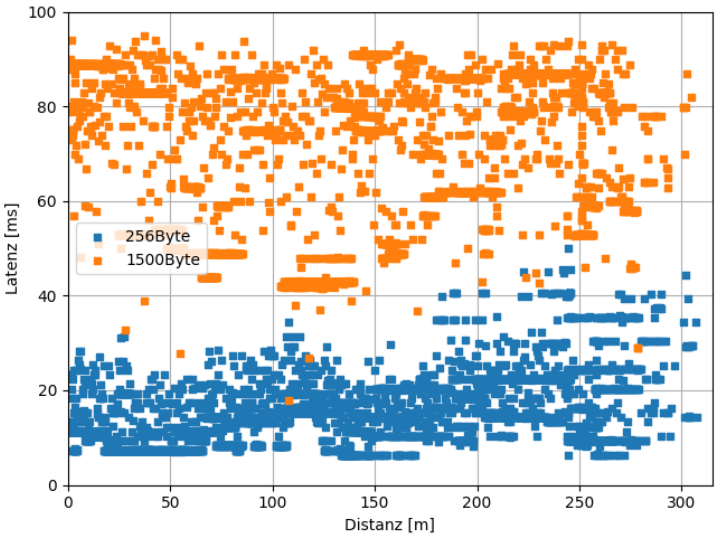


Abb. 5: Latenzwerte bei zwei Störern (Messung 2) in Relation zur Distanz

Messung	Paketgröße	Korrelationskoeffizient
1.2	256Byte	0.481283
1.3	1500Byte	0.144376
2.2	256Byte	0.480737
2.3	1500Byte	0.007573
3.2	256Byte	0.205336
3.3	1500Byte	-0.174214

Tab. 3: Korrelationskoeffizienten für Latenz und Distanz

Des Weiteren wurde der Paketverlust betrachtet, indem die fortlaufenden Paket Sequenznummern aufgezeichnet wurden. Abbildung 6 zeigt die Sequenznummern in Relation zur Distanz. Da hier die Mitschnitte des empfangenden Gerätes gezeigt werden, entsprechen ausgebliebene Sequenznummern verlorenen Paketen. So ist zu erkennen, dass für die Paketgröße 256Byte bis zum entferntesten Punkt auf der Messstrecke noch Pakete empfangen werden konnten. Allerdings ist ab ca. 280m Paketverlust zu verzeichnen. Für die Paketgröße 1500Byte ist bereits ab ca. 240m der erste Paketverlust aufgetreten. Dieser verstärkt sich bis ca. 285m, ab wo (je nach Laufrichtung – dem Messaufbau verschuldet) keine Pakete mehr empfangen werden. Der Knick in den Graphen bei ca. 245m wird durch eine Straßenüberquerung über eine Verkehrsinsel erzeugt.

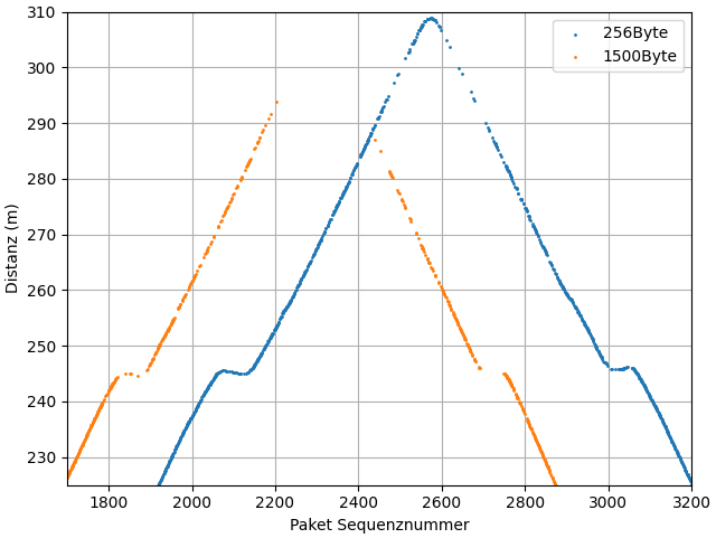


Abb. 6: Grafik der empfangenen Sequenznummern in Relation zu der Distanz

5.2 Diskussion

Die Messergebnisse zeigen deutliche Unterschiede der Latenzzeiten zwischen den verschiedenen Paketgrößen. Besonders auffällig ist der Anstieg der Latenz bei einer Paketgröße von 1500 Byte. Dieser Anstieg ist darauf zurückzuführen, dass die maximale Übertragungseinheit (MTU) bei unter 1500 Byte liegt. Dies führt dazu, dass Pakete in mehrere Telegramme aufgeteilt werden müssen, was zu einer ungefähr doppelten Latenz und somit erhöhtem Jitter führt.

Die zusätzlichen Sender in den Messungen 2 und 3 beeinflussen das Verhalten der Latenz kaum bis gar nicht. Dies ist zu erwarten, da es genügend Kanäle gibt und wir uns im einstelligen Bereich gleichzeitiger Sender befinden. Die C-V2X-Technologie ist darauf ausgelegt, mehrere Hundert Teilnehmer in einem Ballungsraum zu unterstützen. Ein hoher Jitter ist ebenfalls zu beobachten, was auf das „Listen before Talk"-Verfahren mit anschließender eigenständiger Ressourcenbelegung zurückzuführen ist.

5.2.1 Korrelation

256 Bytes: Bei einer Paketgröße von 256 Bytes zeigen die Ergebnisse eine moderate positive Korrelation (ca. 0.48) zwischen Distanz und Latenz in den ersten beiden Messungen. Dies könnte durch externe Faktoren wie Interferenzen, Umgebungsbedingungen oder die Implementierung und Verarbeitung der Signale in den Geräten bedingt sein. Theoretisch sollte die Latenz bei kleineren Paketgrößen weniger stark von der Distanz beeinflusst werden, da die Signale mit der gleichen Geschwindigkeit übertragen werden und die Technologie auf große Reichweiten ausgelegt ist. In der dritten Messung ist die Korrelation geringer (0.205), aber immer noch positiv.

1500 Bytes: Bei einer Paketgröße von 1500 Bytes sind die Korrelationen schwächer und weniger konsistent. In der ersten Messung gibt es eine schwache positive Korrelation (0.144), in der zweiten Messung ist die Korrelation nahezu null (0.007573), und in der dritten Messung gibt es eine schwache negative Korrelation (-0.174). Dies deutet darauf hin, dass bei größeren Paketen die Latenz nicht eindeutig mit der Distanz korreliert und andere Faktoren möglicherweise eine größere Rolle spielen.

Kleinere Paketgrößen (256 Bytes) zeigen eine konsistentere positive Korrelation zwischen Distanz und Latenz, was jedoch theoretisch durch die Technologie und Signalübertragungsgeschwindigkeit nicht zu erwarten wäre. Dies könnte auf externe Faktoren oder spezifische Implementierungsdetails zurückzuführen sein. Größere Paketgrößen (1500 Bytes) weisen keine klare Korrelation auf, was darauf hindeutet, dass bei größeren Paketen andere Einflüsse als die Distanz eine größere Rolle bei der Bestimmung der Latenzzeiten spielen.

5.2.2 Paketverlust

Die Analyse des Paketverlustes zeigt, dass die Technologie grundsätzlich für größere Distanzen ausgelegt ist, jedoch durch verschiedene Nebeneffekte beeinflusst wird. Bei der Paketgröße von 256 Byte konnten bis zum entferntesten Punkt auf der Messstrecke noch Pakete empfangen werden, jedoch trat ab ca. 280 m ein Paketverlust auf. Bei der Paketgröße von 1500 Byte wurden erste Paketverluste bereits ab ca. 240 m beobachtet, die deutlich stärker waren und bis ca. 285 m zunahmen, ab wo keine Pakete mehr empfangen wurden. Als beeinflussende Faktoren sind größtenteils die Umgebungsbedingungen und einen suboptimaler Messaufbau zu erwarten. Während der Messung herrschte keine durchgängige Sichtverbindung, da nach der Überquerung der Straße Bäume, Hecken und Gebäudeteile mit hohem Metallanteil im Weg waren. Zudem ist die Positionierung der Antennen nicht wie in Fahrzeugen üblich, sondern in geringerer Höhe und an einem Handwagen befestigt gewesen. Diese Faktoren führen zu einer deutlichen Reduktion der Leistungsfähigkeit, da die Signalausbreitung massiv eingeschränkt wird.

6 Zusammenfassung

CV2X wird mittelfristig in immer mehr Fahrzeugen oder Roadside-Units verbaut werden. Hier sind vor allem Warnstellen oder Baustellen im Fokus, um Fahrzeuge im Umkreis die entsprechenden Daten zu geben. Dies führt aber auch zu einem erhöhten Datenverkehr und dementsprechend schlechteren Kanalzugriffsmöglichkeiten. Wir haben mit diesem Beitrag gezeigt, wie viele Geräte theoretisch mit welchen Kanalzugriffen Daten austauschen können. Gleichzeitig haben die Messungen gezeigt, dass die Entfernung keinen bzw. nur einen kleinen Einfluss auf die Latenz der Kommunikation hat. Dadurch lassen sich auch weit entfernte Teilnehmer mit kurzen Latenzzeiten von rund 70ms erreichen. Dies erlaubt neben dem effizienten und schellen Verteilen der Daten an alle Netzteilnehmer, auch z.B. einen effiziente Updatemöglichkeit für Daten. Z.B. jede Sekunde ein Update einer Ampel die bald grün wird.

7 Ausblick

Cellular Vehicle-to-Everything (C-V2X) hat das Potenzial, auf die Art und Weise, wie Fahrzeuge miteinander und mit ihrer Umgebung kommunizieren, grundlegend zu verändern. Die infrastrukturlosen Kommunikationssysteme, wie sie für den Automobilsektor entwickelt wurden, bieten auch in der Automatisierungstechnik Potential, um Maschinen und Roboter dezentral zu vernetzen und so Flexibilität und Effizienz zu steigern. Zudem könnte die V2X-Technologie für den Einsatz in bahnähnlichen Verkehrssystemen, wie z. B. MONOCABs [MC], geeignet sein, um durch verbesserte Kommunikation die Sicherheit und Betriebseffizienz zu erhöhen.

Literaturverzeichnis

[3G16] 3GPP: Study on LTE-based V2X Services, Technical Report 36.885, Version 14.0.0, 3rd Generation Partnership Project (3GPP), 2016, URL: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2939.

[3G19] 3GPP: Study on NR V2X, Technical Report 38.885, Version 15.0.0, 3rd Generation Partnership Project (3GPP), 2019, URL: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3521.

[3G20] 3GPP: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description, Technical Specification 36.300, Version 16.0.0, 3rd Generation Partnership Project (3GPP), 2020, URL: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430.

[AA18] 5G Atomotive Association (5GAA), https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/fcc_usdot_cv2x_-_v2.14_w_video1.pdf, Accessed: 2024-09-02.

[AA24] 5G Atomotive Association (5GAA), https://5gaa.org/content/uploads/2020/10/5GAA_White-Paper_C-V2X-Use-Cases-Volume-II.pdf, Accessed: 2024-09-02.

[DRS23] Devika, K. B.; Rohith, G.; Subramanian, S. C.: Impact of V2V Communication on Energy Consumption of Connected Electric Trucks in Stable Platoon Formation. In: 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS). S. 42–47, 2023, DOI: 10.1109/COMSNETS56262.2023.10041288.

[MC] OWL University of Applied Sciences and Arts, https://www.monocab-owl.de/, Accessed: 2024-09-02.

[Ro18] Rondinone, M.: Definition of V2X message sets, https://www.transaid.eu/wp-content/uploads/2017/Deliverables/WP5/TransAID_D5.1_V2X-message-sets.pdf, Accessed: 2024-09-02, 2018.

Automated adaptation of Industrial Communication Networks to meet application requirements through Asset Administration Shell.

Gustavo Cainelli[1], Hasal Kulasekara Pallewaththe Kankanamge[2], Parva Yazdani[1], Lisa Underberg[1], Santiago Soler Perez Olaya[2], Sven Müller[3], Nils Kranefeld[3], Hagen Borstell[4], and Sebastian Rupprecht[4]

Abstract: This paper presents an automated adaptation approach for industrial communication networks using Asset Administration Shells (AAS) to meet specific application requirements in Industry 4.0. The network capabilities and the application requirements are modelled so that virtual entities can negotiate and configure communication settings dynamically. A use case involving a 5G-enabled intralogistics environment demonstrates the approach, in which 5G parameters such as transmission power are adjusted according to the on-demand requirements of cranes and AGVs. To support this, the environment is also modeled, allowing the integration of channel model data into the AAS, which facilitates adaptive, application-specific communication by optimizing energy consumption, enhancing coverage, and reducing interference. The proposed approach demonstrates significant improvements in network performance and adaptability by dynamically aligning communication settings with the operational needs of industrial systems. Future research will concentrate on conducting practical tests to validate the effectiveness of this approach in real-world scenarios.

Keywords: Industry 4.0, Wireless Communication Systems, Digital twin, Asset administration shell, Communication resource management

1 Introduction

Industry 4.0 (I4.0) enables production systems to autonomously reconfigure, reducing the need for rigid planning typical of traditional mass production [As21; De23]. Unlike traditional manufacturing, which prioritizes mass production, I4.0 focuses on flexibility and customization to meet dynamic market demands, minimizing downtime from manual reconfiguration. A crucial aspect of I4.0 is the adaptable communication system that integrates production components, requiring real-time adjustments to meet varying operational needs, such as different uplink or downlink capacities, depending on machine functions [Ca23a; CUR22].

1 Institut für Automation und Kommunikation e.V. , ICT & Automation, Werner-Heisenberg-Straße 1, 39106, Magdeburg, Germany, gustavo.cainelli@ifak.eu; parva.yazdani@ifak.eu; lisa.underberg@ifak.eu
2 Technische Universität Dresden, Dresden, Germany, hasal_kulasekara.pallewaththe@tu-dresden.de; santiago.soler_perez_olaya@tu-dresden.de
3 Demag Cranes and Components Gmbh, Wetter, Germany, sven.mueller3@demagcranes.com; nils.kranefeld@demagcranes.com
4 Thorsis Technologies GmbH, Magdeburg, Germany, hbo@thorsis.com; sru@thorsis.com

This work utilizes digital twins and Asset Administration Shells (AAS) for the holistic management of heterogeneous industrial networks [As21]. This approach models application and communication elements to enable virtual entities to negotiate and decide on optimal configurations of the communication system, aligned with application-specific requirements. A use case demonstrates this concept with 5G technology in intralogistics, where 5G parameters are adjusted to meet specific needs of cranes and AGVs. The orchestrator identifies application requirements from AAS and fine-tunes network settings for optimal performance, such as prioritizing low-latency uplink transmissions.

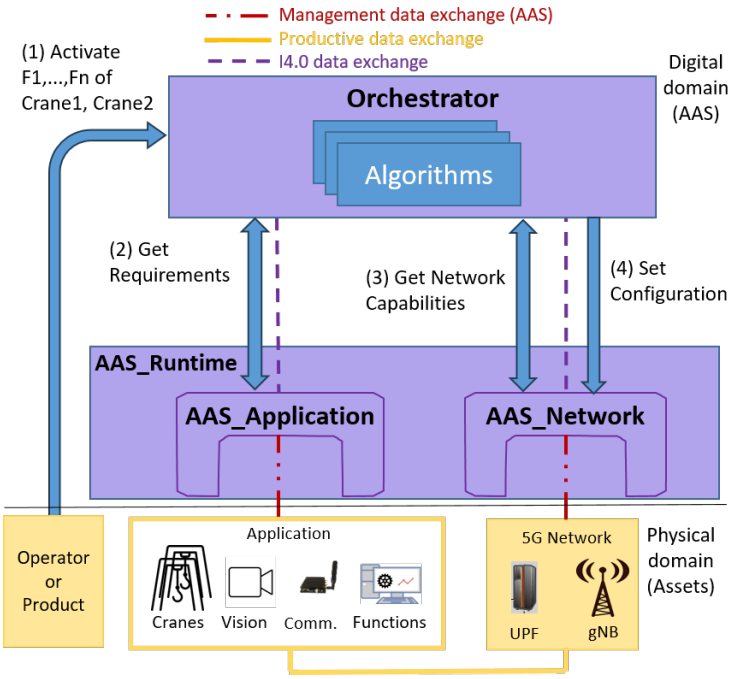


Fig. 1: Use case diagram illustrating the combination of digital and physical domain of the crane system with 5G system.

Building on previous work using a 5G simulator to adapt Time Division Duplexing (TDD) ratio [NSS20; Ol24], this study propose dynamic transmission (TX) power adjustments in industrial settings to enhance energy savings, expand network coverage, and minimize interference, which are crucial for reliable industrial communication. Our AAS-based approach integrates a novel Industry 4.0 submodel that incorporates channel model data, enabling precise TX power adjustments based on specific application needs. This approach aligns with green communication principles by optimizing energy efficiency and communication reliability while minimizing the environmental impact of industrial networks. By proactively responding to transmission condition variations, the system enhances operational

performance and reduces the environmental footprint and costs associated with industrial communications.

The remainder of this paper is organized as follows. Section 2 elaborates on related works including the definition and differentiation of concepts like Industry 4.0, AAS, digital twin, and 5G-NPN and adaptation of the 5G network. Section 2 also elaborates on the channel model and how it can be integrated with AAS. Section 3 elaborates on the use case and the information models. Finally, Section 4 wraps up the results and outlines future work.

2 Related works

2.1 Asset Administration Shell for 5G

5G non-public networks (5G-NPN) are designed for industrial applications, emphasizing quality of service, security, and isolation from public networks, which enhances performance and operational responsibilities [5G21a; Mu23]. Among 5G's critical services, Ultra-Reliable Low Latency Communication (URLLC) is crucial for applications requiring extremely low latency and high reliability [3G18]. The growing importance of the AAS in managing these networks, particularly in URLLC scenarios, highlights its role as a robust management framework [CR21].

In Industry 4.0, AAS plays a vital role in the digital twin architecture by providing a standardized framework for digitally representing industrial assets [Ba]. AAS submodels offer detailed, modular information on assets, including specifications and operational status, enhancing 5G-NPN management while addressing standardization gaps and automating asset management processes [CUR22; Sc21]. Early studies proposed 5G-UE-AAS and 5G-NW-AAS, foundational models that highlighted the need for further communication and application management development [5G21b].

Subsequent research has expanded AAS applications in 5G, including the Wireless System Manager (WSM), which links 5G technology with AAS, enabling adaptive network operations to meet application demands [Ca23b; YC22]. This integration supports sophisticated communication media management and incorporates regulatory considerations for campus networks, enhancing compliance and licensing [5G21b; Fi23; ID24; SSB21]. Recent advancements leverage AAS to dynamically configure networks, improve transmission efficiencies, and support a systematic approach to network customization, emphasizing AAS's evolving role in 5G network management [Ol24].

2.2 Channel modelling

Channel modeling is essential for managing communication systems, especially when using a digital twin approach. It involves capturing the passive influences of the physical

environment, such as obstacles, transmission distances, reflective surfaces, terrain features, and weather conditions, which impact signal propagation. These influences cause effects like fading, path loss, and shadowing, affecting how signals travel from transmitter to receiver. Various methods of channel modeling are available, depending on the required level of detail, and these methods are crucial in understanding and predicting the behavior of communication channels under different environmental conditions [UWR20].

In recent studies, the development of channel models for high-frequency communications has gained significant attention. In [3G22] it is presented the channel models for frequencies up to 100 GHz and outlines their importance in enabling reliable wireless communication. It addresses challenges like diffraction, scattering, and penetration losses at higher frequencies, which are critical for 5G and beyond systems. Additionally, tools like QuaDRiGa provide a detailed framework for simulating quasi-deterministic radio environments by employing and modifying the stochastic channel models provided in [3G22]. This tool helps researchers in replicating real-world conditions for system evaluation, and allows for a comprehensive analysis of various propagation phenomena, including small-scale fading and multi-path effects, for studying the performance of modern wireless systems [Fr21].

Among the various channel models, the one introduced in a recent study by Shiba [Sh24] is particularly noteworthy and is considered in this work. This model is developed from precise measurements that reflect localized environmental conditions typical of industrial settings, such as factory halls. It emphasizes the need for tailored models that consider unique environmental and operational factors specific to different industrial environments. By tuning parameters like maximum receivable power at the receiver, this model provides a more accurate reflection of real-world conditions compared to conventional, generalized simulation-based models, enhancing the predictive capabilities and design of 5G communication systems.

Shiba's channel model specifically accounts for the relationship between transmission distance and maximum received power relative to the transmitted power, and it includes the effects of signal interference and noise specific to certain environments. The model is not intended to generalize across different systems or interference levels but focuses on integrating channel model data into Industry 4.0 information models, particularly the AAS submodels. Although tailored to specific conditions, the template of the AAS submodel can be adapted to include different channel model parameters, making it applicable to other environments with varying communication system conditions.

2.3 Transmit power

The optimization of power consumption in advanced wireless networks is a well-researched area, with many studies focusing on strategies that align closely with the objectives of adapting TX power in 5G systems to enhance energy efficiency and coverage.

The studies by [Oi23] and [ILL23] address the challenge of growing energy consumption in 5G networks, proposing power-saving strategies for base stations. In [Oi23] it is introduced a power consumption model for 5G base stations, using dynamic TX power control and energy-saving techniques like deep sleep, light sleep, and micro sleep modes to minimize energy use based on traffic demands and channel conditions. Similarly, [ILL23] explores methods like dynamic power scaling and deep sleep modes, adjusting downlink TX power to reduce energy consumption while maintaining network performance, using simulations and 3GPP technical reports to demonstrate effectiveness.

While the focus of [Oi23] and [ILL23] is on optimizing 5G networks, [Ma21] and [Wa20] extend these concepts to 6G and advanced link adaptation strategies. [Ma21] introduces POLITE, a framework for 6G that balances power consumption and throughput using delay-aware resource allocation in high-density scenarios. This approach addresses extreme reliability and low-latency requirements of 6G. In contrast, [Wa20] focuses on 5G link adaptation, dynamically adjusting transmission parameters like TX power and modulation order based on real-time channel quality information, enhancing signal quality and optimizing data rates in response to varying channel conditions.

These studies address the optimization of power consumption and network performance in advanced wireless networks through dynamic power control, bandwidth management, and link adaptation strategies, highlighting adaptive mechanisms that balance performance and energy efficiency in 5G and future 6G networks. However, these approaches often overlook the integration of the channel model with the AAS, which is essential for representing physical communication characteristics like signal propagation, interference, and path loss to enhance network reliability. The Industrial Digital Twin Association (IDTA) offers a structured framework for digital twins, including a Wireless Communication Submodel that standardizes communication parameters such as TX power and frequency bands. However, it lacks channel model integration to capture dynamic environmental impacts on signal performance. This study bridges this gap by incorporating the channel model into the AAS, enhancing predictive accuracy and optimizing network performance in I4.0 scenarios.

3 Use case: Adapting the TX power of a 5G system using AAS

This section presents a detailed analysis of a use case within the context of an industrial intralogistics scenario in order to highlight the advantages of employing AAS for channel model representation and radio transmission power adaptation. This use case demonstrates how AAS can enhance the adaptability and efficiency of a 5G system. By examining this specific application, the broader benefits of AAS in enhancing communication efficiency and system performance in such environments can be more clearly understood.

3.1 Overview of the use case

This use case involves a Demag crane system that operates in a 5G-enabled environment (Fig. 1), using a control mechanism and wireless communication system for remote and autonomous operations. The primary actor is an operator who utilizes the "Go To" function to move the crane or AGV to a specific target position.

To optimize network performance and reduce power consumption, the Tx power is not set to the maximum by default. High Tx power could unnecessarily increase energy usage and potentially interfere with other networks. Since positions with weak signal coverage are rarely used by the cranes or AGVs, setting the Tx power to maximum is not justified. Instead, the network dynamically adapts when the AAS identifies that the crane or AGV is moving to a position with poor signal quality.

The AAS utilizes a channel model (related studies are presented in the Section 2) to estimate signal strength at the target position. The channel model, based on the work in [Sh24] and developed from factory hall measurements, helps predict whether the current position will have adequate or poor signal coverage. Other models could also be integrated. Additionally, simulations can be run in parallel using the channel model data stored in the AAS to estimate or predict when an increase in Tx power is necessary. This predictive approach allows the network to adapt dynamically, ensuring efficient communication and minimizing operational costs.

Workflow:

- The operator activates the "Go To" function for the largest distance target position.
- The control system executes the corresponding algorithm.
- The AAS checks the channel model data to estimate signal quality at the target position.
- If the signal is estimated to be weak, the AAS adjusts the Tx power accordingly.
- The crane autonomously moves to the target position.
- The operator waits several minutes at the current location.
- The operator activates the "Go To" function for the nearest distance target position.
- The algorithm is executed again.
- The AAS evaluates the signal conditions and adjusts the Tx power as needed.
- The crane autonomously moves to the new target position.

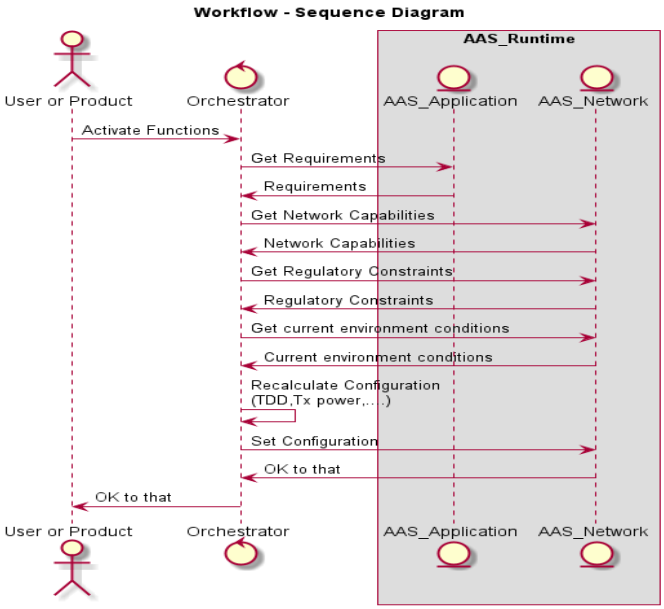


Fig. 2: Workflow sequence diagram

3.2 Tx Power variations and control

An adaptive signal transmission subsystem, governed by information from the AAS submodel with integrated channel model data, allows the Tx power to be adjusted according to the specific needs of the application. There can be many Tx power variation mechanisms and the proposed approach can be implemented in systems supporting software-based continuous or discrete Tx power adjustments.

Tx Power Variation Mechanisms:

- **Hardware Adjustments:** Some devices may include hardware switches or potentiometers for power output adjustments through physical control panels.
- **Software Interfaces:** Some devices may provide software interfaces allowing users to adjust power levels through software applications.
- **Hardware and Software-based Adjustments:** Some devices may support both hardware and software-based power adjustments.
- **Continuous Tx Power Variations:** Some devices may allow Tx power to be varied continuously within a specified range.

- **Discrete Tx Power Variations:** In cases where continuous power adjustment is not possible, a step-wise variation with limited discrete values can be considered. Different device arrangements may support varying levels of Tx power variation, ranging from large variations to only a few discrete values.

3.3 Proposed Information Models: AAS Submodels

The UML diagram of Figure 3 represents the "Environment Conditions" submodel within the "AAS_Network" in Figure 1, specifically focusing on the critical environmental conditions for optimizing network management. This submodel contains information about the maximum allowable distance between user equipment (UEs) and access points, maximum signal attenuation, humidity level, the environment density (determination of congestion level), channel model data, as well as factors related to intervisibility. The structure of the "AAS_Network" and "AAS_Application" together with the other submodels are presented in the previous studies; [So24] and [Ol24]. The proposed approach in this study modifies the "Environment Conditions" submodel with the integration of emphasized channel model data.

The "Environment Conditions" submodel integrates a simplified channel model based on measurements described by [Sh24]. This channel model is encapsulated in a Submodel Element Collection (SMC) that includes key parameters such as location, line-of-sight (LOS), non-line-of-sight (NLOS), and measurement device data. The LOS and NLOS elements contain parameters that represent the mathematical curve that defines the channel model. The "parameter0" and "parameter1" represent a linear equation in accordance with the model approximations in [Sh24]. The complexity of the channel model is abstracted away, as it relies on real-world measurements and calculation of essential parameters. The "Environment Conditions" submodel is crucial for calculating the Reference Signal Received Power (RSRP) at specific locations, based on the base station's transmit (Tx) power and channel model parameters. If the calculated RSRP falls below a certain threshold, which would inhibit proper application functionality, the submodel suggests increasing the Tx power as a potential solution. The orchestrator relies on the information provided by this submodel to assess network conditions and make decisions about adjustments, such as increasing the Tx power, to maintain adequate signal strength and ensure smooth operation.

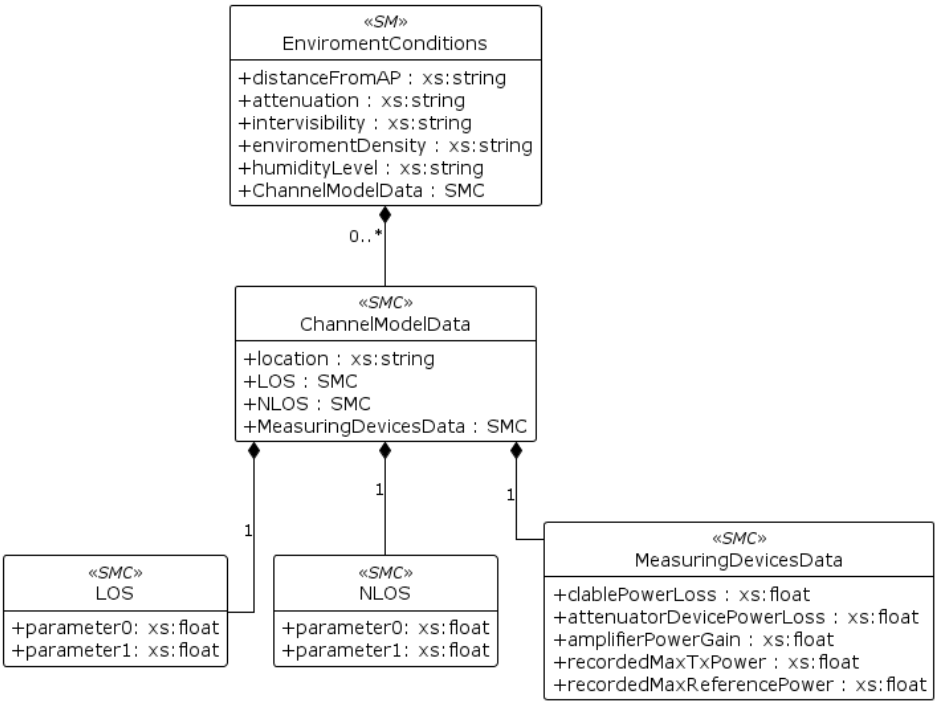


Fig. 3: UML diagram of the environment conditions submodel within the AAS_Network, showing key parameters for channel modelling and network management adjustments by the orchestrator.

Figure 4 illustrates an instance of the AAS submodels on the TU Dresden server, showing the practical application of the previously described environmental conditions submodel. In this context, the channel model, now instantiated on the server, contains specific values obtained from measurements performed at the "Galileo Testfeld" at the "Port of Science" in Magdeburg, using the ifak 5G network.

For example, the value of the location parameter for channelModelData0 is set to “ifak”, indicating that the data were collected at the mentioned location. These values, including channel model parameters such as LOS and NLOS, were obtained using the methods described by [Sh24] and reflect the complexity of a real environment, simplified and encapsulated through the measurements. These data are accessible through standardized interfaces, allowing the orchestrator developed in this work to query the parameters in real-time to make decisions, such as the need to increase the Tx power to ensure efficient and reliable communication.

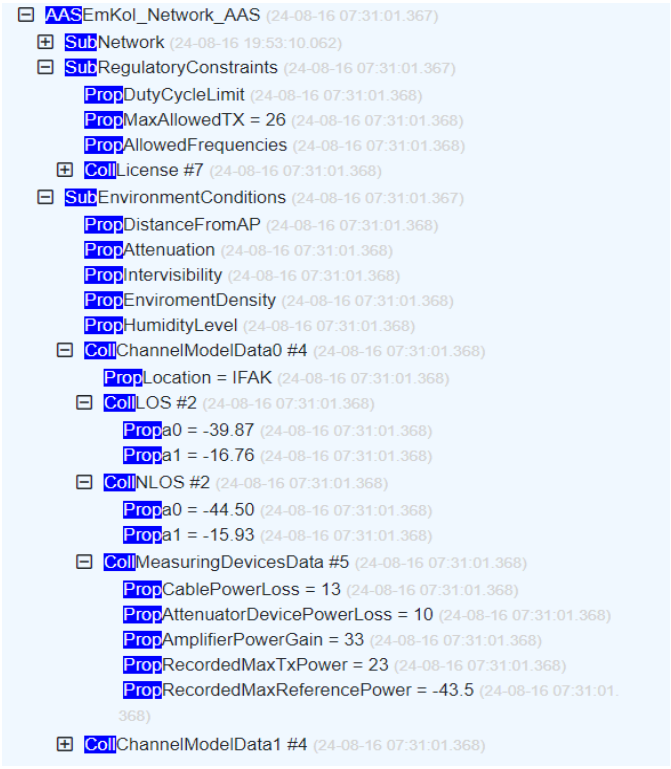


Fig. 4: Instancing of AAS Submodels on the TU Dresden Server.

3.4 The advantages of the proposed approach over the conventional approaches

The conventional network configuration sets Tx power based on worst-case requirements, relying on operators and wasting energy. This study proposes an adaptive reconfiguration method using the AAS ecosystem to dynamically adjust Tx power based on current needs. The AAS stores channel model data, enabling predictions of received signal power to optimize Tx power for energy efficiency and reduced interference. The following list summarizes the key advantages of the proposed approach.

- **Energy Efficiency:** The proposed approach reduces overall energy consumption. This reduction translates to energy savings and lower operating costs.
- **Improved Matching Functions:** More accurate matching between application requirements and communication technology capabilities reduces downtime, enhancing system reliability.

- **Interference Management:** The integration allows for interference level estimation through received signal power calculations, which further supports the decision-making process.
- **Interoperability:** The proposed integration maintains interoperability across the production ecosystem by creating an environment conditions submodel template. This template, containing channel model data, can accommodate different types of channel models for various communication systems under different conditions.

4 Conclusion

In adapting WCS to meet the diverse requirements of modern applications, numerous parameters can be managed to optimize performance such as TDD ratio, bandwidth, and 5QI policies. Among these, Tx power adaptation has been a focal point of this study. By modelling environmental conditions, we demonstrate that managing Tx power through the AAS can enhance energy efficiency across diverse application requirements.

A comprehensive model that captures both functional and environmental parameters is essential for the effective adaptation of the WCS. Our methodology utilizes a specific channel model derived from empirical measurements, providing a realistic depiction of environmental impacts on signal propagation. We advocate for the integration of this model into the AAS framework to enable more adaptive and robust communication systems. Integrating the channel model's parameters into the AAS enables precise estimation of received signal strength, enhancing the system's adaptability.

Future research will be focused on validating this approach over practical experiments. Moreover, further channels models can be tested to verify the necessity of extending the AAS model. This ongoing work aims to refine the adaptability of WCS by exploring advanced modelling techniques and more complex environmental scenarios by leveraging the enhanced capabilities of the AAS.

Acknowledgement

The authors would like to express their sincere gratitude for the funding provided for this work by the project EmKoI4.0 under contracts 01MJ22014A, 01MJ22014B, 01MJ22014C and 01MJ22014D funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK), Germany.

References

[3G18] 3GPP: 5G NR Physical layer procedures for control, tech. rep. TS 38.213, 3rd Generation Partnership Project (3GPP), 2018, URL: https://www.3gpp.org/DynaReport/38213.htm.

[3G22] 3GPP: Technical Specification Group Radio Access Network; Study on channel model for frequencies from 0.5 to 100 GHz, tech. rep. TR 38.901 v17.0.0, 3rd Generation Partnership Project, 2022.

[5G21a] 5G Alliance for Connected Industries and Automation (5G-ACIA): 5G Non-Public Networks for Industrial Scenarios (White Paper), Available:https://5g-acia.org/wp-content/uploads/2021/04/WP_5G_NPN_2019_01.pdf [Accessed: 19.01.2024], 2021.

[5G21b] 5G Alliance for Connected Industries and Automation (5G-ACIA): Using Digital Twins to Integrate 5G into Production Networks (White Paper), Available:https://www.5g-acia.org/publications/ [Accessed: 19.01.2024], 2021.

[As21] Association, I. D. T.: Was ist die Verwaltungsschale aus technischer Sicht?, 2021, URL: https://industrialdigitaltwin.org/wp-content/uploads/2021/10/2021_Was-ist-die-AAS-3.pdf.

[Ba] Bader, S. et al.: Plattform Industrie 4.0 - Details of the Asset Administration Shell-Part 1, Accessed: 2024-04-25, URL: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/Details_of_the_Asset_Administration_Shell_Part1_V3.html.

[Ca23a] Cainelli, G.; Yazdani, P.; Underberg, L.; Jumar, U.; Pereira, C.: Managing the 5G system based on production requirements using an Asset Administration Shell approach. In: 21st IFAC World Congress. Yokohama, Japan, 2023.

[Ca23b] Cainelli, G. P.; Yazdani, P.; Underberg, L.; Jumar, U.; Pereira, C. E.: Managing the 5G system based on production requirements using an Asset Administration Shell approach. In: IFAC-PapersOnLine. Vol. 56 2. 2, 22nd IFAC World Congress, pp. 2108–2114, 2023, DOI: https://doi.org/10.1016/j.ifacol.2023.10.1113, URL: https://www.sciencedirect.com/science/article/pii/S2405896323015161.

[CR21] Cainelli, G.; Rauchhaupt, L.: Introducing resilience in industrial 5G systems using a digital twin approach. In: 2021 17th IEEE International Conference on Factory Communication Systems (WFCS). Pp. 33–36, 2021, DOI: 10.1109/WFCS46889.2021.9483618.

[CUR22] Cainelli, G.; Underberg, L.; Rauchhaupt, L.: Influences of logical link design in 5G campus systems. In: 2022 IEEE Future Networks World Forum (FNWF). IEEE, 2022, DOI: 10.1109/fnwf55208.2022.00072.

[De23] Deutsches Institut für Normung e.V. (DIN): Normungsroadmap Industrie 4.0 Version 5, 2023, URL: https://www.din.de/resource/blob/907744/61cf955a9830c84aefee02747b3d9fa0/nrm-industrie-4-0-version-5-2023-engl-web-data.pdf.

[Fi23] Fischer, C.; Schneider, J.; Krummacker, D.; Salzmann, D.; Holoyad, T.; Schotten, H. D.: A Regulators Perspective on Digital Twinning for Mobile Communications. In: Mobile Communication - Technologies and Applications; 27th ITG-Symposium. Pp. 37–42, 2023.

[Fr21] Fraunhofer, H. H. I.: QuaDRiGa-Quasi Deterministic Radio Channel Generator, user Manual and Documentation, tech. rep. "v2.6.1", Fraunhofer Heinrich Hertz Institute, 2021.

[ID24] IDTA: AAS Submodel Templates - Wireless Communication, Accessed 2024-06-19, 2024, URL: https://github.com/admin-shell-io/submodel-templates/tree/main/published/Wireless%20Communication/1/0.

[ILL23] Islam, T.; Lee, D.; Lim, S. S.: Enabling Network Power Savings in 5G-Advanced and Beyond. IEEE Journal on Selected Areas in Communications 41 (6), pp. 1888–1899, 2023, DOI: 10.1109/JSAC.2023.3273706.

[Ma21] Mandelli, S.; Lieto, A.; Weber, A.; Wild, T.: Power Optimization and Throughput Enhancement in 6G Networks by Delay-Aware Resource Leverage. In: 2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit). Pp. 176–181, 2021, DOI: 10.1109/EuCNC/6GSummit51104.2021.9482470.

[Mu23] Muzaffar, R.; Ahmed, M.; Sisinni, E.; Sauter, T.; Bernhard, H.-P.: 5G Deployment Models and Configuration Choices for Industrial Cyber-Physical Systems – A State of Art Overview. IEEE Transactions on Industrial Cyber-Physical Systems 1, pp. 236–256, 2023, DOI: 10.1109/TICPS.2023.3311394.

[NSS20] Nardini, G.; Stea, G.; Sabella, D.: Simu5g: A system-level simulator for 5g networks. In: Proceedings of the 10th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2020). Pp. 68–80, 2020.

[Oi23] Oikonomakou, M.; Khlass, A.; Laselva, D.; Lauridsen, M.; Deghel, M.; Bhatti, G.: A Power Consumption Model and Energy Saving Techniques for 5G-Advanced Base Stations. In: 2023 IEEE International Conference on Communications Workshops (ICC Workshops). Pp. 605–610, 2023, DOI: 10.1109/ICCWorkshops57953.2023.10283643.

[Ol24] Olaya, S. S. P.; Kankanamge, H. K. P.; Cainelli, G. P.; Gambal, B.: Flexible Reconfiguration of Industrial 5G Networks over Asset Administration Shell. In: IEEE ETFA 2024 - 29th IEEE International Conference on Emerging Technologies and Factory Automation. 2024.

[Sc21] Schweizer, H.; Braunisch, N.; Alt, R.; Schmitz, K.; Wollschlaeger, M.: Orchestrierung und Choreografie von Inbetriebnahme-Prozessen in verteilten Automatisierungssystemen mittels proaktiver Verwaltungsschalen und B2MML. at - Automatisierungstechnik 69 (3), pp. 242–255, 2021, DOI: doi:10.1515/auto-2020-0118, URL: https://doi.org/10.1515/auto-2020-0118.

[Sh24] Shiba, T.; Furuichi, T.; Cainelli, G.; Müller, S.; Gambal, B.; Kranefeld, N.; Ohori, F.; Osuga, T.; Persico, G.; Underberg, L.; Itaya, S.; Suematsu, N.: Proposal of Channel Model Measurement Method under Local 5G System Operation. In: Proceedings of the 20th IEEE Asia Pacific Wireless Communications Symposium (APWCS 2024). 2024.

[So24] Soler Perez Olaya, S.; Pallewaththe Kankanamge, H. K.; Pedroso Cainelli, G.; Gambal, B.; Wollschlaeger, M.: Management of Industrial 5G Networks over Asset Administration Shell. In: 2024 20th IEEE International Conference on Factory Communication Systems (WFCS). 2024.

[SSB21] Stock, D.; Schneider, M.; Bauernhansl, T.: Towards Asset Administration Shell-Based Resource Virtualization in 5G Architecture-Enabled Cyber-Physical Production Systems. Procedia CIRP 104, http://dx.doi.org/10.1016/j.procir.2021.11.159, pp. 945–950, 2021, ISSN: 2212-8271.

[UWR20] Underberg, L.; Willmann, S.; Rauchhaupt, L.: Traffic model integration of a 5G systeminto industrial communication. 2020.

[Wa20] Wang, J.; Han, Y.; Li, X.; Jin, S.: Design and Implementation of a 5G NR-based Link-adaptive System. In: 2020 IEEE/CIC International Conference on Communications in China (ICCC). Pp. 196–201, 2020, DOI: 10.1109/ICCC49849.2020.9238946.

[YC22] Yazdani, P.; Cainelli, G.: Managing communication resources based on production requirements in an Industry 4.0 scenario using a digital twin approach. In: EKA 2022 - Entwurf komplexer Automatisierungssysteme, 17. Fachtagung. Vol. 1, Available:https://www.researchgate.net/publication/381551869_Managing_communication_resources_based_on_production_requirements_in_an_Industry_40_scenario_using_a_digital_twin_approach [Accessed: 2024-07-03], 2022.

Potentials of Asset Administration Shell in Secure System Architecture Design

Ali M.Hosseini [1], Thilo Sauter [1] und Wolfgang Kastner [2]

Abstract: Integrating security decisions early in the system development process can enhance the cost-effectiveness and security of the final product, thus making the principle of "*security by design*" increasingly relevant. In industrial control systems (ICS), developing the system architecture during the design phase plays a crucial role in bolstering security. This involves identifying potential threats, implementing controls, ensuring compliance, and monitoring through a structured and layered approach. However, effective security decision-making depends on the collection and use of diverse data related to each asset, often sourced from various documents, such as asset data sheets and experiments. This paper examines how incorporating a Digital Twin, implemented via the Asset Administration Shell (AAS), can enhance ICS security. The AAS serves as a comprehensive digital representation of an asset, containing detailed asset-related information. We evaluate the current capabilities of AAS and its role in integrating security into the ICS architecture. A use case demonstrates how AAS can be utilized in system architecture to address attack simulations and conduct consequence analysis, thereby supporting informed security decision-making.

Keywords: Asset Administration Shell, System Architecture, Industrial Cybersecurity

1 Introduction

Incorporating innovative technologies, frequently sourced from external vendors and originally developed in the Information Technology (IT) sector, has become a prevalent trend among various vendors in Industrial Control Systems (ICS). This trend is crucial for staying competitive and meeting the fast-changing demands of customers. However, it has also raised considerable security concerns, as demonstrated by the increasing number of successful cyberattacks in recent years. [ST23].

To reduce the higher costs associated with addressing security issues in later stages, especially during operation, it is crucial to consider system security early in the design process [Hu22]. Since system architecture design is one of the initial activities in the design phase, it provides an opportunity to ensure that the security-by-design principle is effectively applied [HSK23]. System architecture development adopts a proactive approach by setting clear requirements, specifications, and system designs. It also remains responsive to feedback during the detailed design, implementation, and testing phases [HSK22].

1 TU Wien, Institute of Computer Technology, Karlsplatz 13, 1040 Vienna, Austria, ali.hosseini@tuwien.ac.at, https://orcid.org/ 0000-0003-3323-1924 ; thilo.sauter@tuwien.ac.at, https://orcid.org/0000-0003-1559-8394

2 TU Wien, Institute of Computer Engineering, Karlsplatz 13, 1040 Vienna, Austria, wolfgang.kastner@tuwien.ac.at, https://orcid.org/0000-0001-5420-404X

Fig. 1 depicts a representative ICS network architecture comprising several common components. Leveraging this architecture for tasks such as risk analysis, simulation, security standards compliance, and security documentation requires a wide range of information. Given the involvement of different vendors, there was a need for a standardized, and machine-readable information catalog.

A significant trend in the industry is the standardized digitalization driven by Plattform Industrie 4.0, which has introduced the concept of the Asset Administration Shell (AAS) as the digital representation of assets. The AAS organizes information about asset characteristics and functionalities into a structured and standardized format, known as a submodel. This structure facilitates interoperability among stakeholders throughout the entire system lifecycle [Wa17]. Given that AAS can provide requisite information in a standardized format, it can be leveraged during the architectural design phase to facilitates the retrieval of necessary information crucial for modeling, analyzing, and verifying a system architecture in accordance with requirements and constraints.

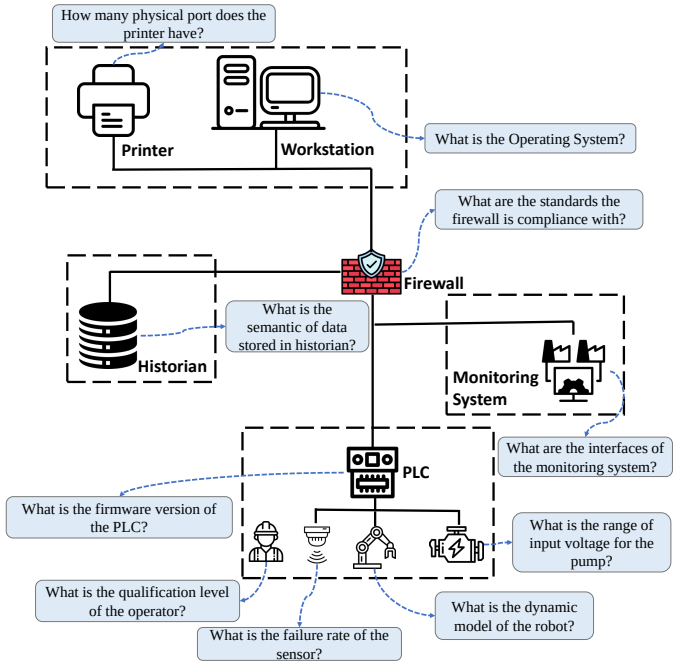


Fig. 1: A generic ICS network architecture emphasizes the importance of having relevant information about assets from a security perspective.

The main research questions (RQ) are:

- **RQ1**: What are the essential components of the AAS that have the potential to contribute to security by design?

- **RQ2**: What is the current state of AAS development in relation to its potential use in architecture design?
- **RQ3**: How can the AAS be integrated into the architectural design process?

2 Background

2.1 Security by Design During Architecture Design

Security by design refers to the practice of shifting cybersecurity responsibilities from the system's end users to its engineers. To alleviate the burden on end users during the operational phase, critical security decisions should be made during the engineering process [Fl23].

Security decisions are often made late in the engineering process, resulting in the implementation of costly and complex security measures, such as adding extra components and extensive authentication and authorization systems. However, more practical and straightforward security measures, like designing secure network architecture or selecting components with built-in authentication mechanisms, should be decided early in the engineering workflow, such as during system architecture development [FDF22; HSK23].

The system architecture contributes significantly to security by establishing a robust framework that integrates security considerations throughout the development lifecycle. Early integration of security requirements facilitates adherence to the security-by-design principle. Moreover, it is mandated by ICS security standards, such as IEC62443. This standard provides architecture-level guidelines and security requirements, including network segmentation and the assignment of security levels to each segment [TSK20]. A well-defined architecture serves as a clear blueprint for security testing, enabling focused unit, integration, system, and acceptance testing by clarifying which security aspects need to be validated at each stage of the V-Model [AR13]. Additionally, system architecture ensures consistent security implementation through detailed documentation and enables traceability of security decisions and their impacts, aiding in compliance with security standards and regulations [Il23].

2.2 Asset Administration Shell

The Asset Administration Shell (AAS), introduced by Plattform Industrie 4.0, represents the digital counterpart of physical assets. From the Industry 4.0 perspective, the AAS models asset-related information and services in a structured format, facilitating access and exchange among stakeholders. An asset can be either a physical entity—such as a sensor, actuator, or field worker—or a non-physical entity, like software, documents, or models [Sc16].

The AAS comprises three main components: (1) Passive, (2) Active, and (3) Application Programming Interface (API). The passive part organizes asset-relevant information into groups called submodels, following a standardized meta-model. The active part encompasses the asset's functionalities and services, while the API allows external access to the AAS content [HSK22]. Fig. 2 illustrates the core elements of the AAS meta-model. Each submodel groups various data properties into submodel elements. As shown in Fig. 2, submodels effectively capture a wide range of information due to the detailed semantics of the AAS. This information can be accessed and utilized by stakeholders such as asset vendors, system integrators, and operators.

Given that the AAS holds asset-relevant information that can be efficiently exchanged, it has been applied in various areas, including predictive maintenance [Ra23], fault detection [Ma23], security risk management [BKS21], and ICS cybersecurity [HSK21].

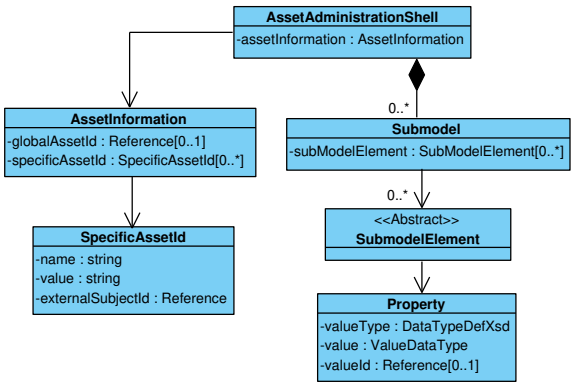


Fig. 2: A simplified data model of an AAS organizes asset properties into submodels, with each submodel grouping relevant properties for a specific domain, such as a security submodel [EB22].

The Asset Administration Shell (AAS) has the potential to provide essential security information throughout all phases of the system lifecycle, including design, integration and implementation, and operation. However, the necessary information must be identified and categorized into submodels. Various submodels have been proposed to address the needs of different fields. The adoption of AAS has been promoted by various governmental and industrial sectors. A key player in this effort is the Industrial Digital Twin Association (IDTA)[3], which fosters a common understanding among industry stakeholders, associations, and researchers by providing the necessary specifications for AAS and creating and harmonizing submodels. Moreover, the research community actively contributes to enriching AAS submodels.

3 https://industrialdigitaltwin.org/en/

3 AAS Submodels Relevant to Security

We examine the present ability of AAS to supply information for system architecture development. To this end, we identify and categorize submodels mainly from IDTA's website[4]. Our focus is specifically on three critical architectural-level tasks: system design and analysis, simulation, and compliance with security standards.

Tab. 1 illustrates submodels proposed by IDTA with potentials to be used in security engineering in design phase, especially in system architecture design.

3.1 Design and Analysis

During the design phase, various analytical methods can be employed to assess and design systems. These methods utilize asset-related information, including interfaces, physical ports, communication protocols, configuration, capabilities, and security attributes. For example, the *Asset Interface Description* submodel provides essential information about asset interfaces, which can be used to model and analyze the system network. Similarly, the *Wireless Communication* submodel supplies information relevant to wireless technologies and their security properties, such as encryption algorithms.

3.2 Simulation

Conducting system simulations is crucial for analyzing security-related design aspects. In the field of security, the resilience of systems can be assessed by modeling and simulating scenarios involving security attacks, such as Denial-of-Service (DoS) attacks [SZS19] or man-in-the-middle attacks [Wl21]. The *Provision of Simulation Models* submodel facilitates the seamless provision of simulation model files for assets using an AAS. This submodel aids asset manufacturers by providing simulation models that can be automatically integrated into specific simulation environments during system engineering.

3.3 Compliance with Security Standards

Security standards generally require risk assessment, mitigation, validation, and documentation. The AAS supports interoperability among stakeholders, making it a valuable tool for ensuring compliance. Compliance necessitates cooperation and detailed information exchange, including asset inventories, communication protocols, user accounts, configuration settings, and logs. The IDTA's *Handover Documentation* submodel standardizes information exchange, while the *Time Series Data* submodel enhances interoperability for time series data throughout the asset lifecycle.

4 https://industrialdigitaltwin.org/en/content-hub/submodels

Submodel	Description	Architectural Application
Asset Interface Description	This submodel defines an information model and a standardized representation for describing the interface(s) of an asset service or asset-related service	Design and Analysis, Simulation, Standard Compliance
Hierarchical Structures enabling Bills of Material	This submodel is designed to offer hierarchical structures for industrial equipment in an interoperable way, primarily utilizing Entities and Relationship Elements from the AAS Metamodel	Design and Analysis, Standard Compliance
Data Model for Asset Location	The location of static or mobile objects (such as assets, goods, or trackables), as well as the origin and destination of transport processes, are crucial pieces of information	Design and Analysis, Standard Compliance
Provision of Simulation Models	The presented submodel facilitates the interoperable delivery of simulation model files for an asset through the asset administration shell	Simulation, Design and Analysis
Handover Documentation	The submodel Handover Specification establishes a standardized format for exchanging information or documentation related to a particular asset	Standard Compliance
Wireless Communication	The primary goal of this submodel is to integrate wireless industrial communication within the I4.0 framework	Simulation, Design and Analysis
Nameplate for Software in Manufacturing	This submodel ensures effective and efficient use and management of this software, it is essential to collect relevant information in a standardized format	Standard Compliance

Tab. 1: Relevant submodels selected from IDTA's website.

4 AAS Integration into Secure Design Processes

We examine how to integrate AAS information into secure design process. To this end, we show that AAS can be used for designing and describing a system architecture and simulating security attacks. Fig. 3 shows the designed use case that is based on the Purdue Model, widely recognized as a foundational model in this field and underpinning most ICS security standards [ST23]. The Purdue Model organizes industrial control systems into different hierarchical levels, each representing a different function in an enterprise.

The process being controlled involves regulating the liquid level in a tank. The tank receives liquid input through a proportional valve, while the output supplies liquid at a constant rate to other processes. A Programmable Logic Controller (PLC) manages the valve to maintain the desired tank level, and a level sensor provides real-time measurements of the liquid level. The system architecture is implemented in MATLAB/SIMULINK.

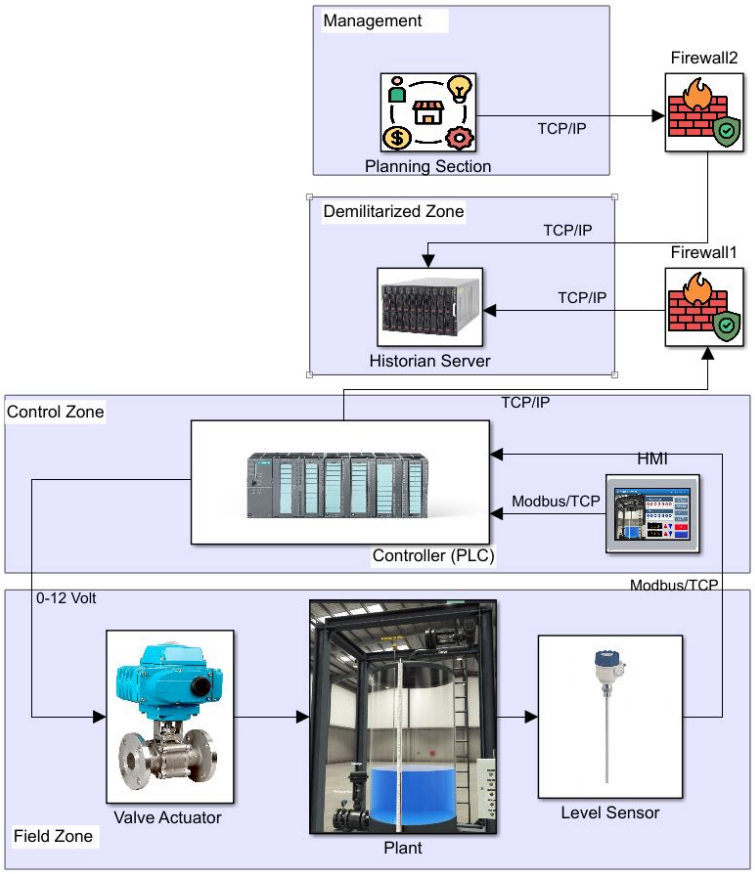


Fig. 3: An illustrative system architecture based on Purdue Model.

4.1 Asset Interface Description

It is possible to describe an asset interface using ”*Asset Interface Description*” submodel of AAS prvoided by IDTA. This submodel defines an information model and a standardized representation for describing the interfaces of an asset service or a related service. With this information, connections to such services can be initiated, allowing requests or subscriptions to data points, and/or the execution of operations. For example, data points provided by a system service might include various sensor readings or status values, while operations could involve actions like turning a motor ”on” or ”off.”

Fig. 4 shows the description of a Modbus interface for level sensor which is connected to a PLC. Consider a level sensor configured as a Modbus server. The server holds the current level reading in a holding register. When the Modbus client (PLC) sends a request to read

this register, the server responds with the level value. As it is shown in Fig. 4, the security features of the interface is described based on a particular security schema. The Modbus interface has no built-in security function.

When the interface between the PLC and the level sensor is accurately described using this AAS submodel, a connection between these two assets can be established for simulation purposes, taking into account real-world properties like timeouts. The submodel was created using AASX Package Explorer[5] and is available in JavaScript Object Notation (JSON) format. The necessary information for the simulation can be extracted from the JSON file.

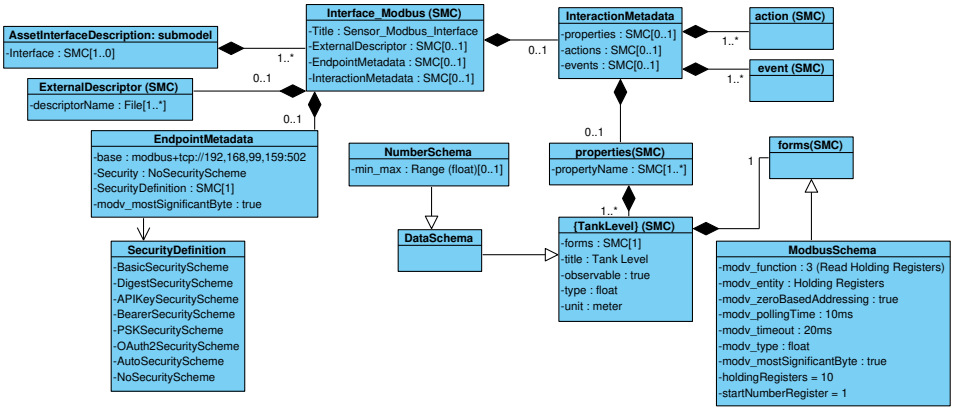


Fig. 4: Asset interface description submodel modelling a the level sensor Modbus interface.

4.2 Provision of Simulation Models

The objective is to offer a variety of simulation models for different types of simulations on the architectural level using AAS. The "*Provision of Simulation Models*" submodel enables the provision of information that simplifies the search for suitable models and their integration into a simulation environment. For simulating the dynamic behaviour of the system, we need to have the dynamic model of each asset shown in Fig. 3. For illustration purpose, we show how this submodel can deliver required information of the valve for simulation purposes.

Fig. 5 presents the data model associated with the valve's *Provision of Simulation Models* submodel. This model includes a nonlinear representation of the valve with defined inputs and outputs, each assigned to different ports with specific properties. These ports allow the simulation models to interact and exchange signals, facilitating the construction of a more complex system. The model is designed specifically for the SIMULINK environment, and the numerical simulation has been validated using the Runge-Kutta solver [BW96].

5 https://github.com/eclipse-aaspe/package-explorer

When integrating individual models into a system model, compatibility is essential. This includes ensuring consistency in the simulation environment, solver, data types, and other relevant factors.

The interfaces described and illustrated in Sec. 4.1 should be assigned to the respective ports, resulting in a complete system model for simulation. The submodels were developed using AASX Package Explorer.

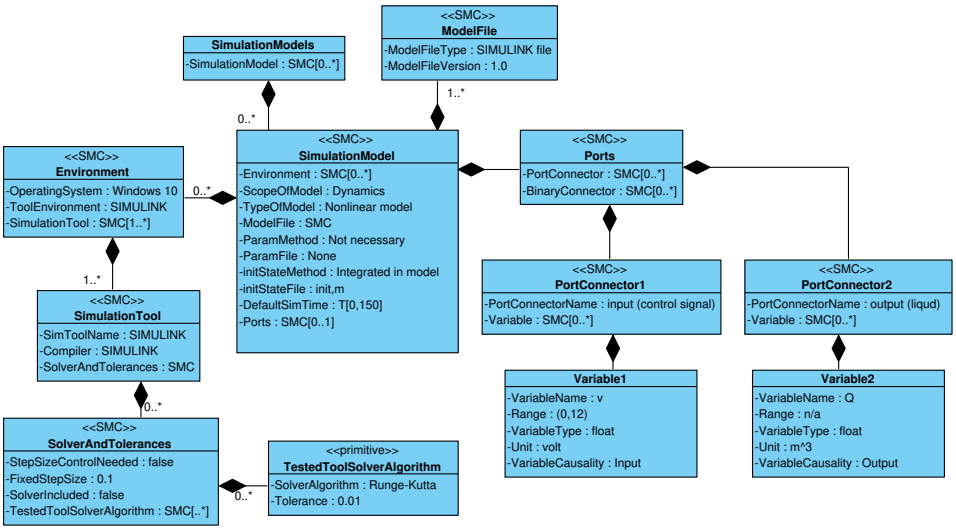


Fig. 5: The ”provision of simulation models” submodel including the dynamic model of valve used in the use case to be used for simulation purposes in SIMULINK.

5 Simulation Results

By utilizing the AAS and its submodels, one can obtain the critical information required for a specific simulation environment. Fig. 6 illustrates a configuration window in SIMULINK used for simulating the use case presented in Fig. 3. All the highlighted information in this window can be sourced from the AAS submodel.

In the proposed use case, the level sensor measures the level of a tank and transmits this data to PLC via Modbus TCP, the security vulnerabilities inherent in the Modbus protocol are a significant concern. The Modbus protocol, which operates in plain-text and lacks any form of authentication, poses a critical security risk. This vulnerability allows an attacker with direct network access to the PLC (e.g. insider attacker) to potentially control the entire Modbus-based ICS.

In the Modbus TCP mode, each device, including the level sensor and the PLC, is assigned a unique IP address, facilitating their identification and communication over an Ethernet

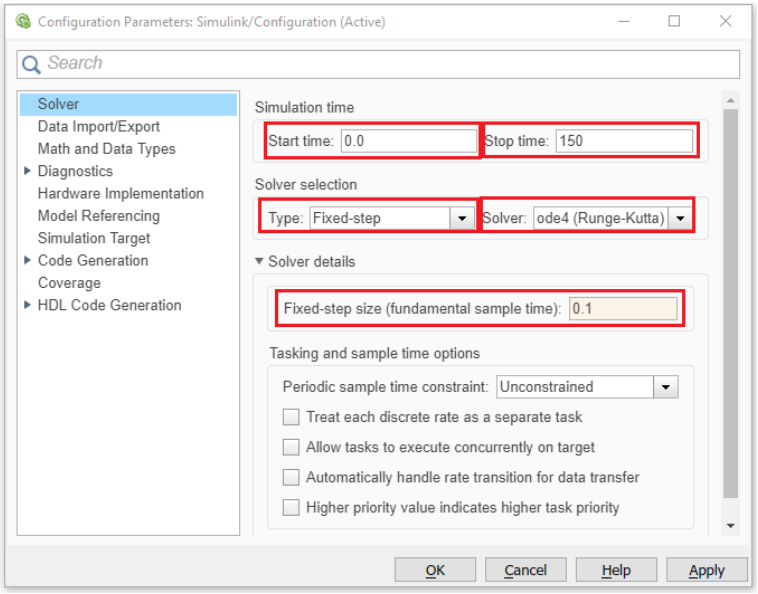


Fig. 6: Configuring SIMULINK environment according to the AAS submodel from Fig. 5.

network. However, the plain-text communication between the Modbus Master (in this case, the PLC) and the Modbus Slave (the level sensor) makes the system particularly susceptible to cyberattacks. Since the protocol does not include any authentication mechanisms, an attacker who gains access to the network can easily fabricate query packets in the correct Modbus format and send them to the Modbus Slave.

To execute such an attack, the attacker's machine needs only two conditions to be met: (1) the machine must be able to send packets to the Modbus Slave (the level sensor), and (2) the packets sent must conform to the Modbus protocol format.

Once these conditions are satisfied, the attacker can manipulate the registers or coils of the Modbus Slave, potentially altering the data sent to the PLC or even disrupting the entire system's operation. This scenario highlights the importance of implementing additional security measures, such as network segmentation, encryption, and intrusion detection systems, to protect Modbus-based communication in industrial environments.

In a potential attack scenario, it is assumed that the attacker has exploited a vulnerability in the HMI, allowing the attacker to send traffic to the system. Hence, the attacker manipulates the Modbus communication by injecting false responses to the PLC's queries (spoofing attack). Instead of relaying the actual tank level from the level sensor, the attacker adjusts the holding registers to send a fabricated value, leading the PLC to take incorrect actions.

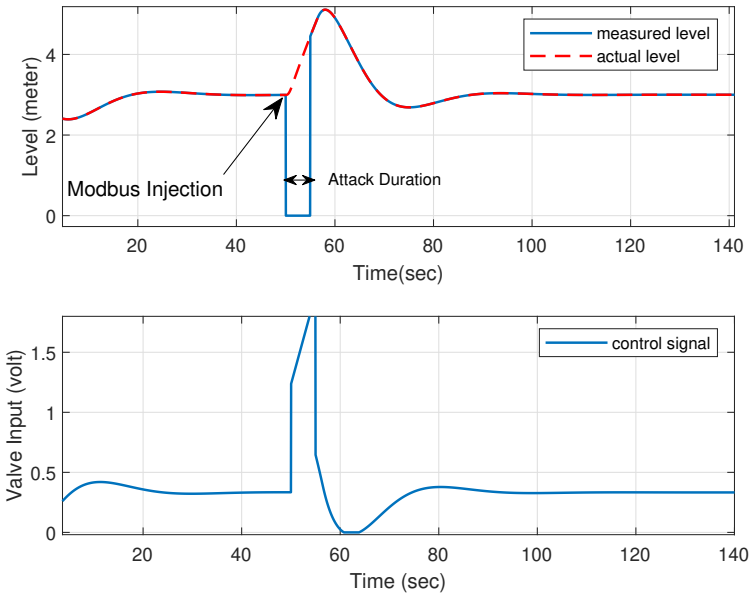


Fig. 7: Simulation of a Modbus injection attack where the level measurement values are manipulated for five seconds to falsely indicate an empty tank.

Fig. 7 demonstrates a Modbus injection attack in which the level measurement values from a tank's level sensor are deliberately altered for a period of five seconds.

During this time, the system is manipulated to falsely report that the tank is empty, regardless of its actual level. This type of attack can lead the PLC to make incorrect decisions based on the fabricated data, potentially causing operational issues or safety hazards within the system. As it can be seen in Fig. 7, the level increased during the attack. The simulation highlights the vulnerability of Modbus communication to such attacks, where unauthorized data injection can disrupt normal system functions.

6 Conclusion

In this paper, we explored the integration of AAS into ICS architecture to enhance security during the design phase. Our investigation centered on how AAS can support critical architectural tasks such as system design and analysis, simulation, and compliance with security standards. We identified and categorized relevant AAS submodels that contribute to these tasks, emphasizing their potential to strengthen security by providing detailed asset

information, facilitating simulations of security attacks, and ensuring compliance with standards (RQ2 is answered in Sec. 3 and RQ1 in Sec .2).

Our use case, built on the Purdue Model, demonstrated how AAS can be applied to model and analyze ICS systems. Specifically, we used the AAS submodels to describe asset interfaces and simulation models to validate system response in case of cyberattacks. The findings revealed that AAS can significantly enhance the security architecture of ICSs. However, many aspects of security engineering still are not supported by AAS, such as vulnerability management and risk analysis (RQ3 is answered in Sec. 4 and Sec. 5).

Overall, incorporating AAS into ICS design processes provides a robust framework for integrating security considerations early in system development. By leveraging AAS for detailed asset descriptions, simulation, and compliance, we can better anticipate and mitigate security risks, ultimately leading to more secure and resilient industrial systems.

References

[AR13] Apvrille, L.; Roudier, Y.: SysML-Sec: A SysML environment for the design and development of secure embedded systems. APCOSEC, Asia-Pacific Council on Systems Engineering, S. 8–11, 2013.

[BKS21] Bhosale, P.; Kastner, W.; Sauter, T.: A centralised or distributed risk assessment using asset administration shell. In: 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, S. 1–4, 2021.

[BW96] Butcher, J. C.; Wanner, G.: Runge-Kutta methods: some historical notes. Applied Numerical Mathematics 22 (1-3), S. 113–151, 1996.

[EB22] for Economic Affairs, F. M.; (BMWK), C. A., 2022, URL: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/VWSiD%20V2.0.html, Stand: 21. 08. 2024.

[FDF22] Fluchs, S.; Drath, R.; Fay, A.: Security Decision Base: How to Prepare Security by Design Decisions for Industrial Control Systems–Analysis of concepts for organizing security-relevant information from software engineering, requirements engineering, and systems engineering. 2022.

[Fl23] Fluchs, S.; Taştan, E.; Trumpf, T.; Horch, A.; Drath, R.; Fay, A.: Traceable Security-by-Design Decisions for Cyber-Physical Systems (CPSs) by Means of Function-Based Diagrams and Security Libraries. Sensors 23 (12), S. 5547, 2023.

[HSK21] Hosseini, A. M.; Sauter, T.; Kastner, W.: Towards adding safety and security properties to the Industry 4.0 Asset Administration Shell. In: 2021 17th IEEE International Conference on Factory Communication Systems (WFCS). IEEE, S. 41–44, 2021.

[HSK22] Hosseini, A. M.; Sauter, T.; Kastner, W.: A safety and security reference architecture for asset administration shell design. In: 2022 IEEE 18th International Conference on Factory Communication Systems (WFCS). IEEE, S. 1–8, 2022.

[HSK23] Hosseini, A. M.; Sauter, T.; Kastner, W.: Formal Verification of Safety and Security Properties in Industry 4.0 Applications. In: 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, S. 1–8, 2023.

[Hu22] Humayun, M.; Jhanjhi, N.; Almufareh, M. F.; Khalil, M. I.: Security threat and vulnerability assessment and measurement in secure software development. Comput. Mater. Contin 71, S. 5039–5059, 2022.

[Il23] Illescas, J.; Ehrlinger, L.; Denk, G.; Buchgeher, G.: Towards an Ontology for Technical Security Standards. In: 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, S. 1–8, 2023.

[Ma23] Mailer, D.: Dynamic Deployment of Fault Detection Models-A Use Case of the Asset Administration Shell, Diss., Wien, 2023.

[Ra23] Rahal, J. R.; Schwarz, A.; Sahelices, B.; Weis, R.; Antón, S. D.: The asset administration shell as enabler for predictive maintenance: a review. Journal of Intelligent Manufacturing, S. 1–15, 2023.

[Sc16] Schweichhart, K.: Reference architectural model industrie 4.0 (rami 4.0). An Introduction 40, 2016.

[ST23] Sauter, T.; Treytl, A.: IoT-enabled sensors in automation systems and their security challenges. IEEE Sensors Letters 7 (12), S. 1–4, 2023.

[SZS19] Sun, Q.; Zhang, K.; Shi, Y.: Resilient model predictive control of cyber–physical systems under DoS attacks. IEEE Transactions on Industrial Informatics 16 (7), S. 4920–4927, 2019.

[TSK20] Tanveer, A.; Sinha, R.; Kuo, M. M.: Secure links: secure-by-design communications in IEC 61499 industrial control applications. IEEE Transactions on Industrial Informatics 17 (6), S. 3992–4002, 2020.

[Wa17] Wagner, C.; Grothoff, J.; Epple, U.; Drath, R.; Malakuti, S.; Grüner, S.; Hoffmeister, M.; Zimermann, P.: The role of the Industry 4.0 asset administration shell and the digital twin during the life cycle of a plant. In: 2017 22nd IEEE international conference on emerging technologies and factory automation (ETFA). IEEE, S. 1–8, 2017.

[Wl21] Wlazlo, P.; Sahu, A.; Mao, Z.; Huang, H.; Goulart, A.; Davis, K.; Zonouz, S.: Man-in-the-middle attacks and defence in a power system cyber-physical testbed. IET Cyber-Physical Systems: Theory & Applications 6 (3), S. 164–177, 2021.

Generische Schnittstellengenerierung für digitale Zwillinge der Industrie 4.0

Nico Braunisch [1], Uwe Schmidt [1], Tom Gneuß [1] und Martin Wollschlaeger [1]

Abstract: Die Asset Administration Shell (AAS) zur Darstellung digitaler Zwillinge, sowie die Kommunikation mit AAS-Systemen, sind wesentliche Bestandteile der Industrie 4.0 (I4.0). Es wird ein Ansatz zur automatischen API-Generierung für AAS-basierte Systeme vorgestellt. Anstelle einer manuellen Übersetzung der abstrakten I4.0 API-Spezifikation, wird diese formalisiert und anschließend automatisch in eine für ein Zielsystem lesbare API-Definition übersetzt. Dies erleichtert die SDK-Generierung, was die Entwicklung und die Wartung von AAS-Systemen optimiert, sowie Fehler minimiert. Als Serialisierungsformat wird YAML/ OpenAPI genutzt. Aufbauend werden die Schnittstellenbeschreibungen Protobuf, SOAP, GraphQL und WSDL untersucht. Perspektivisch soll der Ansatz die Integration von AAS-Systemen in die dynamische I4.0-Landschaft unterstützen.

Keywords: Industrie 4.0, Asset Administration Shell, API-Generierung, modellgetriebene Entwicklung

1 Einleitung

In der sich stetig ändernden Landschaft der I4.0 ist die AAS zur Darstellung digitaler Zwillinge maßgebend. Hierzu zählt neuerdings auch ihre interoperable Zugriff auf AAS-Systeme über ihre Anwendungsprogrammierschnittstelle (API). Bisher wird die API für AAS manuell – durch die Übersetzung der in Form eines Buches veröffentlichten abstrakten Spezifikation [PIP21] – durchgeführt. Dieser Prozess ist langwierig, fehleranfällig und nicht allgemein umsetzbar. Dies hat negative Auswirkungen auf die Wartung bestehender und die Entwicklung neuer AAS-Systeme bei Aktualisierungen der AAS-Spezifikation [PIP22]. Der Beitrag präsentiert einen Ansatz zur automatischen Generierung einer API-Spezifikation anhand der formalisierten I4.0-Spezifikation. Zur Umsetzung der Formalisierung wird vorgeschlagen, YAML/ OpenAPI aufgrund seiner Verbreitung, Übersichtlichkeit und Einfachheit als Serialisierungsformat zu verwenden. Aufbauend werden weitere Schnittstellenbeschreibungen wie Protobuf, SOAP, GraphQL und WSDL untersucht, was die Notwendigkeit klar definierter Schnittstellen für eine effiziente Nutzung von AAS-APIs unterstreicht. Dieser Ansatz garantiert Skalierbarkeit sowie Flexibilität zu entwickelnder und wartender Lösungen, insbesondere bei der SDK-Generierung verschiedener AAS-Schnittstellen unabhängig spezieller Serialisierungsformate. Schlussendlich wird so die reibungslose Integration der Lebenszyklen digitaler Zwillinge in der dynamischen I4.0-Landschaft unterstützt.

[1] Technische Universität Dresden, Institut für Angewandte Informatik, Lehrstuhle für Prozesskommunikation, 01062 Dresden, Deutschland, nico.braunisch@tu-dresden.de, https://orcid.org/0009-0000-2432-5529;
uwe.schmidt1@tu-dresden.de, https://orcid.org/0000-0000-0000-0000;
tom.gneuss@tu-dresden.de, https://orcid.org/0009-0003-5587-3503;
martin.wollschlaeger@tu-dresden.de, https://orcid.org/0000-0002-4646-4455

2 Problemstellung

Das Konzept des digitalen Zwillings ist ein grundlegendes Element im Rahmen der I4.0. Es ermöglicht die Darstellung eines physischen oder realen Vermögenswerts im virtuellen Raum (d.h. im Cyberspace) — eine wesentliche Voraussetzung für die Entwicklung komplexer cyber-physischer Systeme. Ein Vermögenswert (auch Asset) beschreibt in diesem Zusammenhang jede Einheit von Wert, die sich im Besitz oder unter der Verwaltung einer Organisation befindet. Die AAS, auch als Verwaltungsschale bezeichnet, steht für eine auf I4.0 zugeschnittene standardisierte Darstellung des digitalen Zwillings eines Assets[PIP22].

Die Interaktion und Nutzung der Schnittstellen eines digitalen Zwillings variiert je nach Kontext und Lebenszyklus. Eine konkrete AAS nimmt dabei die Form einer explizit strukturierten Dateneinheit an, die sich an das wohldefinierte Metamodell der AAS [PIP22] hält. Das AAS-Metamodell gliedert sich in verschiedene Elemente, die das Asset ganzheitlich beschreiben. Sie reichen von einfachen Datenelementen wie Eigenschaften über Ereignis- und Betriebsdefinitionen bis hin zu Beziehungen (z.B. Asset-Fähigkeiten), was die Grundlage der Interoperabilität in der I4.0 schafft. Der Beitrag beschränkt sich auf die Perspektive, die AAS als interoperables Werkzeug zum Informationsaustausch innerhalb der I4.0-Wertschöpfungskette zu verwenden.

Auch wenn I4.0-Systeme in verschiedenen Programmiersprachen für unterschiedliche Laufzeitumgebungen implementiert werden, der Informationsaustausch zwischen ihren Komponenten erfolgt hauptsächlich durch Datenübertragungen oder Remote Procedure Calls (RCP) [PIP21] unter Verwendung spezifischer APIs. Die allgemeine API für den Datenaustausch wird in [PIP21] abstrakt definiert und erweitert die Asset-Beschreibungen um Begriffe von Operationen, sowie entsprechende Datenstrukturen. Eine Realisierung der AAS-APIs erfordert die Verfügbarkeit in verschiedenen Austauschformaten, um eine Vielzahl von Dateiübertragungsparadigmen und Programmiersprachen zu unterstützen.

Derzeit werden serielle Informationsaustauschschemata von AAS-Systemen für jedes Format manuell implementiert – oft von verschiedenen Autoren. Dieser arbeitsintensive Prozess umfasst die Interpretierung und Kodifizierung der offiziellen Spezifikation [PIP21] in verschiedene Schnittstellen unter Erhalt der Integrität zum AAS-Metamodell. Durch die Verwendung verschiedener Laufzeitumgebungen, Unterschieden in Programmiersprachen, verschiedenen Übertragungsformaten und -technologien, sowie unterschiedlichen Interpretationen der Spezifikation durch die Entwickler, stellt dieser Prozess eine nicht zu unterschätzende Herausforderung dar. Diese bedarf einer langfristigen, automatisierbaren und allgemein nutzbaren Lösung, auch bezogen auf die stetige Weiterentwicklung des AAS-Metamodells, für die Entwicklung und Wartung von AAS-Systemen.

3 Motivation

Das Ziel der Interoperabilität innerhalb des I4.0-Ökosystems, wie es in den „Referenzarchitekturen, -standards und -normen“ der Arbeitsgruppe „Plattform Industrie 4.0“ skizziert wird, ist von grundlegender Bedeutung für die nahtlose Integration und Kommunikation zwischen verschiedenen Systemen und Geräten. Die Kommunikationsschnittstellen werden repräsentiert durch weit verbreitete Datenaustauschformate und spezifische APIs, wobei die APIs den ausgetauschten Daten bestimmte Beschränkungen (Constraints) auferlegen.

Wie in Abb. 1 dargestellt, werden die verschiedenen Ebenen konzeptionell in „technologieneutral“, „technologiespezifisch“, „Implementierung“ und „Laufzeit“ unterteilt. Der Schwerpunkt der API-Spezifikation [PIP21] liegt jedoch auf der Bereitstellung einer konkreten HTTP/ REST-API und der Definition der Nutzdaten gemäß der Datenspezifikation [PIP22]. Dies bedeutet, dass derzeit nur eine technologiespezifische API in Form von HTTP/ REST bereitgestellt wird und keine technologieunabhängige API-Definition existiert [PIP21].

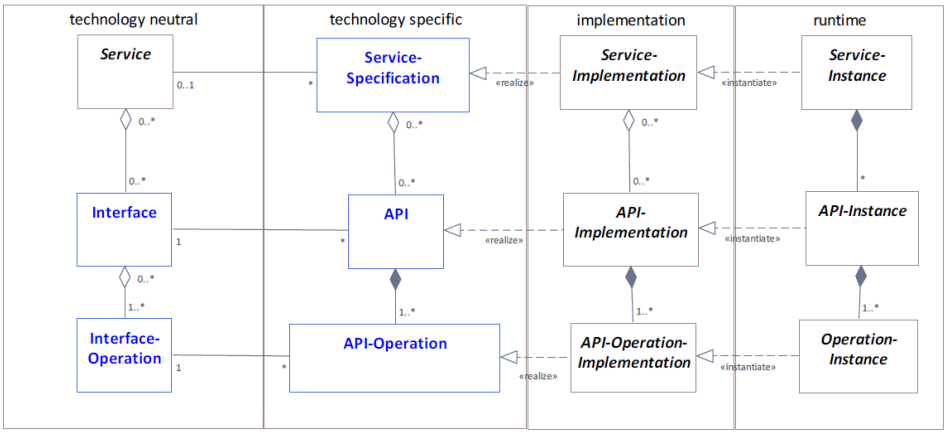


Abb. 1: Dienste, Schnittstellen, APIs and Operationen [PIP21]

Die bereitgestellte HTTP-API auf GitHub oder SwaggerHub [AA] sind jedoch nicht umfassend oder standardisiert, insbesondere im Hinblick auf die OpenAPI [PIP21] Definitionen für HTTP-APIs. Wie in Kap. 2 beschrieben, führen die daraus manuell erzeugten Schnittstellen und Informationsmodelle, aufgrund unterschiedlicher Interpretationen der Spezifikation durch Entwickler, zu Inkonsistenzen und Divergenzen in der anschließenden Kommunikation. Entsprechend sind die erstellten Systeme nicht interoperabel.

Zur Förderung eines kohärenten und interoperablen I4.0-Ökosystems sollte der Schwerpunkt auf Konsistenz in Design, Entwicklung und Wartung gelegt werden. Die Generierung von APIs kann einen bedeutenden Schritt in diese Richtung darstellen, was eine standardisierte Darstellung des Modells über verschiedene Systeme und Geräte hinweg ermöglicht.

Dieser Beitrag motiviert den Entwurf einer modellgesteuerten automatischen Generierungs-Toolchain zur Erzeugung von API-Schemata für verschiedene Technologien auf der Grundlage des Grundprinzips einer „einzigen Quelle der Wahrheit“. Wodurch semantische Gleichheit zwischen den Schemata garantiert und eine nahtlose Interoperabilität zwischen allen Komponenten innerhalb des I4.0-Systems gewährleistet werden kann.

4 Stand der Technik

I4.0 stellt ein neues Paradigma in der Fertigung dar, das sich durch die Integration digitaler Technologien in den Produktionsprozess auszeichnet. In diesem Zusammenhang spielen der Aufbau von Schnittstellen und der Informationsaustausch über APIs eine entscheidende Rolle, um die Kommunikation und Zusammenarbeit zwischen verschiedenen Systemen und Geräten zu ermöglichen. In diesem Abschnitt werden die bestehenden Datenaustauschformate und Informationsmodelle, die in I4.0 verwendet werden, sowie die Methoden zur Erstellung ihrer Schemata und Transformationen aufgeführt.

4.1 Manuelles Kodieren von Schnittstellen

Die manuelle Kodierung von Schnittstellen ist der traditionelle Ansatz, bei dem die Entwickler die Schnittstellendeklarationen direkt in der zu verwendenden Programmiersprache schreiben [LZL22]. Dieser Ansatz bietet Flexibilität und Kontrolle über die Schnittstellendefinition, kann aber zeitaufwändig und fehleranfällig sein, insbesondere bei komplexen Schnittstellen [PP22].

4.2 Verwendung von Codegenerierungswerkzeugen

Code-Generierungswerkzeuge automatisieren den Prozess der Erzeugung von Schnittstellen aus bestehenden Code-Basen, Dokumentationen oder anderen Quellen [JZZ22]. Generierungswerkzeuge können die Entwicklungszeit und den Aufwand erheblich reduzieren, insbesondere bei sich wiederholenden Aufgaben und Änderungen. Die Genauigkeit und Nachhaltigkeit der generierten Schnittstellen hängt jedoch von der Qualität der Eingabedaten und den Fähigkeiten des jeweiligen Werkzeugs ab [BK21].

4.3 Nutzung von IDE-Funktionen

Integrierte Entwicklungsumgebungen (IDEs) bieten häufig integrierte Funktionen für die Erstellung von Schnittstellen. Mit diesen Funktionen können Entwickler in der Regel schnell Schnittstellen aus vorhandenem Code oder Vorlagen erstellen [SLW20]. IDE-basierte Schnittstellengenerierung kann den Entwicklungsprozess rationalisieren und den Bedarf an manueller Kodierung reduzieren [WZZ21].

4.4 Schablonen basierte Ansätze

Schablonen basierte Ansätze verwenden vordefinierte Schablonen zur Generierung von Schnittstellen aus verschiedenen Eingabequellen wie UML-Modellen (Unified Modeling Language), beispielsweise Klassendiagrammen, oder domänenspezifischen Beschreibungen [YZL19]. Diese Ansätze können die Integrität der Schnittstellendefinition zu dem zugrundeliegenden Design bzw. den Anforderungen sicherstellen [LL21].

4.5 Dynamische Schnittstellengenerierung

Dynamische Schnittstellengenerierungstechniken erstellen Schnittstellen zur Laufzeit auf der Grundlage bestimmter Bedingungen oder Benutzereingaben [WHJ20]. Dieser Ansatz bietet Flexibilität bei der Anpassung an sich ändernde Anforderungen oder bei der Bereitstellung kontextabhängiger Schnittstellen. Er erfordert jedoch eine sorgfältige Abwägung der Performance-Kosten und der fehlenden Typsicherheit [LZL22].

4.6 Web Service Description Language

Die Web Services Description Language (WSDL) ist ein Schlüsselelement bei der API-Entwicklung und dient als standardisierte Schnittstellenbeschreibung für Webdienste. Sie definiert die Operationen, Datentypen und Protokolle, die in dem Dienst verwendet werden. Ein WSDL-Dokument, das in der Regel in XML geschrieben ist, dient als Vertrag zwischen Dienstanbietern und -nutzern und ermöglicht eine nahtlose Kommunikation, sowie Interoperabilität [AAE06].

4.7 Simple Object Access Protocol

SOAP ist ein Standard-Nachrichtenprotokoll, welches Anwendungen ermöglicht, Daten und Dienste auszutauschen, unabhängig von den zugrunde liegenden Plattformen oder Programmiersprachen. Es kann verwendet werden, um Schnittstellen für verschiedene Anwendungen zu erstellen, die eine nahtlose Interaktion zwischen verschiedenen Systemen ermöglichen [LL21].

5 Entwurf

Im Rahmen des Beitrags wird ein generativer und modellgetriebener Ansatz zur Generierung von konkreten APIs für AAS-Systeme vorgeschlagen, welcher auf der Arbeit

zur Schema-Generierung [Br23a] und SDK-Entwicklung [Br23b] aufbaut und durch die Schnittstellendefinition von [PIP21] erweitert wird.

Zunächst wird das domänenunabhängige AAS-Metamodell mit all seinen Elementen, Attributen, Konzepten und Constraints aus „Details of the Asset Administration Shell – Part 1 und 2“ [PIP22, PIP21] vollständig in eine Zwischenrepräsentation (IR) übersetzt. Die IR wird dann durch Schemageneratoren in verschiedene gewünschte Schnittstellenbeschreibungen übersetzt, basierend auf den Designzielen und Implementierungsregeln von [PIP22]. Abb. 2 veranschaulicht den Prozess.

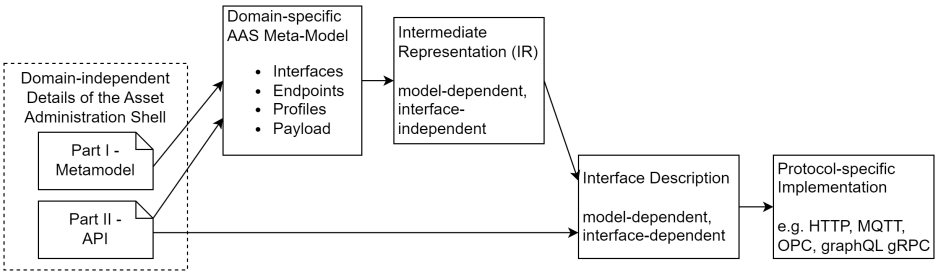


Abb. 2: Modellgetriebener Prozess zur API-Generierung für AAS

5.1 Domaiänenunabhängiges Modell

Das domänenunabhängige Modell von [PIP22] stellt das AAS-Metamodell mit strukturellen und funktionalen Elementen, sowie den Anforderungen und Constraints dieser Elemente dar. Darüber hinaus werden in [PIP21] die Schnittstellen mit allen Endpunkten, Profilen, Dienstspezifikationen, Zugriffsmethoden, erforderlichen und unterstützten Protokollen und die Nutzdaten beschrieben. Die genannte Spezifikation in UML-Diagrammen, Tabellen und Texten ist dabei teils informell, was eine formalere Modellierungssprache verlangt. Seit der Entwicklung von aas-core-meta [So] gibt es zwar eine maschinenlesbare Version von „Part 1“ [PIP22], aber noch keine für „Part 2“ [PIP21].

5.2 Domaiänenspezifisches Modell

Um die spezifischen AAS-Austauschschnittstellen zu generieren, muss eine formale Darstellung erzeugt werden, die alle Datenstrukturen und Constraints der offiziellen abstrakten Spezifikation in „Teil 2“ [PIP21] erfassen kann. Dafür wird auf der früheren Arbeit von aas-core-works zur Modellierung des AAS-Metamodells [Br23a, Br23b] aufgebaut. Darüber hinaus muss diese Darstellung auch die Zugriffs- und Änderungsoperationen auf Modellelemente und deren Constraints sowie Konsistenzprüfungen enthalten. In Anlehnung an die gängige Terminologie im Compilerentwurf wird diese Darstellung als Intermediate

Representation (IR) [Ny21] bezeichnet. In diesem Stadium ist keine schnittstellenspezifische Funktionalität enthalten.

5.3 Schnittstellenspezifisches Modell

Die API-spezifischen Elemente fehlen in dem domänenspezifischen Modell und müssen für die beteiligten Schnittstellen definiert werden. Entsprechend sind die Klassen, ihre Attribute, Klassen-Invarianten sowie Vor- und Nachbedingungen der Operationen auf Darstellungen einer konkreten Schnittstelle abzubilden. So werden z.B. die Klassenattribute des Metamodells auf die Eigenschaften aus OpenAPI abgebildet, siehe Abschnitt 6. Dazu gehören auch zusätzliche implementierungsspezifische Details, die in der IR nicht erfasst werden können, z.B. technologie- oder plattformspezifische Funktionen, die in den angegebenen Constraints Verwendung finden. Aus diesem Grund werden folgende API-spezifischen Elemente definiert:

Schnittstellen-Definition In einer ausgewählten Technologie bildet eine definierte Schnittstelle einer Komponente, die mit anderen funktionalen Komponenten verbunden werden kann und aus einer Menge mit Operationen und Ereignissen als Elementen besteht, die API [PIP21].

Endpunkt-Definition Die Endpunkte/ Zugangspunkte der Schnittstellen beschreiben, wo und wie auf bestimmte Teile des instanziierten ASS-Metamodells zugegriffen werden kann [PIP21].

Zugriffsoperationen Zugriffsoperationen (Access Operations) beschreiben die Operationen (Prozeduren), die über eine API-Schnittstelle aufgerufen werden können und definieren Interaktionsmuster unter Verwendung der angegebenen Schnittstelle [PIP21].

Profil-Definition Profile spezifizieren die anwendbaren Lade- oder Serialisierungsformate, sowie die verfügbaren Funktionen, z.B. Seitennummerierung (pagination) oder asynchrone Operationen. Jedes Profil wird dafür eindeutig identifiziert und durch eine API-Definition dargestellt [PIP21].

Dienstspezifikation Dienstspezifikationen werden in Profilen verfeinert, die unterstützte API-Operationen, Modifizierungen und Pfadkombinationen regeln. Die verschiedenen Profile einer Dienstspezifikation haben gleiche Attribute, aber unterschiedliche Versionen. Das Attribut „Version“ umfasst die Haupt- und Nebenversion sowie das Profil selbst [PIP21].

5.4 Beschreibung der Schnittstelle

Im letzten Schritt wird das schemaspezifische Modell für die Austauschformate auf der Grundlage des domänenspezifischen Modells generiert. Dies ermöglicht eine Skalierung

sowohl bei Änderungen am Metamodell als auch bei der Anzahl der unterstützten Schemata, sowie Austauschformate. Das domänenspezifische Modell ist nur einmal zu aktualisieren und die Änderungen werden anschließend automatisch auf die schemaspezifischen Modelle übertragen.

5.5 Protokollspezifische Implementierung

Für jedes der gewünschten Protokolle wird das protokollspezifische Modell definiert, wie in Abb. 3 dargestellt, indem zusätzliche Elemente aufgenommen werden, welche anderen Protokollen fremd sind. So fehlen beispielsweise bestimmte Details der AAS REST API in anderen Protokollen wie der OPC UA API.

6 Implementierung

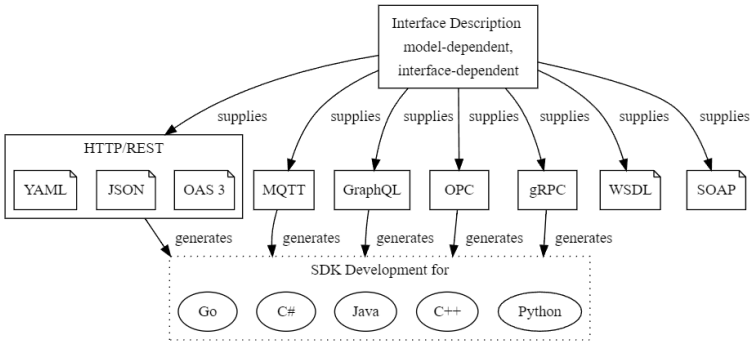


Abb. 3: Generierung von Schnittstellenbeschreibungen und SDKs in Programmiersprachen

Für die Formalisierung des AAS-Metamodells wird eine Untermenge der Programmiersprache Python [Pyb], weiterhin als Simplified Python bezeichnet (SP), verwendet. Dabei werden Datentypen als Python-Klassen und implementierte Metamodell-Constraints als Klasseninvarianten, d.h. Bedingungen, die immer gelten sollten, spezifiziert. Im Folgenden werden ...

6.1 Vereinfachtes Python

Aufgrund seiner spezifischen und eingeschränkten Domäne unterstützt SP nur eine bestimmte Teilmenge von Sprachkonstrukten und eingebauten Funktionen, wie bereits in [Br23b] und [Br23a] vorgestellt. Berücksichtigt wird somit eine sehr begrenzte Teilmenge der eingebauten Funktionen und Typen von Python. Beispielsweise sind verschachtelte

Funktionen, Kontextmanager oder verschachtelte Klassen nicht zulässig. Die Definitionen des AAS-Metamodells, also Aufzählungen (Enumerations), primitive Datentypen (zzgl. eventueller Constraints) und Klassen, werden als Python-Klassen modelliert. Die entsprechenden Constraints werden als Klasseninvarianten mit der icontract-Bibliothek [ic] erfasst. Darüber hinaus werden für alle Elemente des AAS-Metamodells die entsprechenden APIs und ihre jeweiligen Endpunkte, wie in [PIP21] beschrieben, definiert. Die Vor- und Nachbedingungen einer Operation werden ebenfalls als Verträge mit der icontract-Bibliothek definiert. Meist sind die Verträge beliebig komplex und können nicht durch begrenzte Schemasprachen wie OpenAPI erfasst werden.

6.2 Zwischenrepräsentation (IR)

Der abstrakte Syntaxbaum (AST) von Python [Pya] hält nun die definierten Klassen und Datentypen des Metamodells. Er ist jedoch für die weitere Verarbeitung zur Codegenerierung ungeeignet, es fehlen beispielsweise Verweise zwischen den Klassen. Daher übersetzen wir den AST in eine gröbere Form, die IR, mit Hilfe derer wir die Strukturen des Metamodells leicht iterieren und (de-)referenzieren können. Wir stellen in der IR Klassen für alle Datenbausteine wie Aufzählungen, Primitive, abstrakte und konkrete Klassen, Konstanten usw. bereit. Außerdem übersetzen wir Invarianten, Vorbedingungen und Nachbedingungen aus dem Python-AST in unseren eigenen domänenspezifischen Untermengen-AST. So sind Analyse und Typinferenz viel prägnanter zu implementieren, wie bereits in [Br23b] beobachtet.

6.3 API-erzwingbare Constraints

Python ist zwar Turing-vollständig und kann beliebige Constraints modellieren, aber viele Datenformate für den Informationsaustausch über APIs können nur eine begrenzte Anzahl von Constraints durchsetzen. So kann zum Beispiel eine Constraint, die zwei oder mehr Eigenschaften in OAS3 gemeinsam betrifft, nicht in einem Schema modelliert werden. Für solche Formate analysieren wir die Invarianten, indem wir einen Musterabrgleich mit AST der Invarianten durchführen, und leiten die folgenden Constraints ab, die in einem Schema für eine Klasse durchgesetzt werden können.

6.4 Zwischendarstellung zur Schnittstellenbeschreibung

Die IR von aas-core-sdk-gen wird zur Generierung der AAS-Schnittstellen verwendet [Br23b]. Zu diesem Zweck wird das aas-core-meta um eine Schnittstellenbeschreibung erweitert, die Elemente zur Beschreibung verschiedener Schnittstellen, APIs, Zugriffsmethoden, Endpunkte usw. enthält. Es ist wichtig zu beachten, dass die Schnittstellenbeschreibung unabhängig von der konkreten Implementierung ist, z.B. von der verwendeten Protokoll- oder API-Technologie.

6.4.1 Intermediate Representation zu Schnittstellenbeschreibungssprachen

Verschiedene Schnittstellenbeschreibungssprachen (IDLs), darunter OAS, WSDL, Protocol Buffers (ProtoBuf), Apache Thrift, CORBA und GraphQL Schema Definition Language (SDL), spielen eine zentrale Rolle bei der Softwareentwicklung, da sie standardisierte Mittel zur Beschreibung und Definition von Schnittstellen bereitstellen. Diese IDLs dienen als Brücke zwischen verschiedenen Programmiersprachen und Systemen und ermöglichen eine nahtlose Kommunikation und Interoperabilität. Durch die Erfassung der Struktur- und Verhaltensaspekte von Softwarekomponenten dient die Zwischendarstellung in diesen IDLs als Blaupause für die Generierung von APIs in einer standardisierten und sprachunabhängigen Weise. Dieser Ansatz rationalisiert den Entwicklungsprozess, gewährleistet Konsistenz und reduziert die Möglichkeit von Fehlern, während er gleichzeitig die Erstellung von APIs ermöglicht, die den Industriestandards und Best Practices entsprechen.

6.4.2 Zwischendarstellung zu Anwendungsschnittstellen

Die Automatisierung der Generierung von APIs aus einer Zwischendarstellung in Kombination mit der Schnittstellenbeschreibung beinhaltet einen systematischen und regelbasierten Prozess. Die erste Phase umfasst eine umfassende Analyse der Zwischendarstellung, bei der Endpunkte, Datenmodelle und zugehörige Verhaltensweisen identifiziert werden. Unter Ausnutzung der inhärenten Struktur und der in dieser Repräsentation kodierten Regeln übernehmen dann automatisierte Werkzeuge die Generierung von OpenAPI-3-konformer Dokumentation, Code-Skeletten und Konfigurationen. Unser Ansatz stellt ein ausgeklügeltes Mittel dar, um die Feinheiten der Zwischendarstellung in voll funktionsfähige APIs zu übersetzen, und zeigt die Leistungsfähigkeit der regelbasierten Automatisierung in der API-Entwicklung.

7 Zusammenfassung

Wir haben einen transformativen und modellgesteuerten Ansatz vorgestellt, der darauf abzielt, Schnittstellenschemata für offiziell anerkannte APIs zu generieren, die für die Entwicklung von I4.0-Komponenten entscheidend sind. In unserem Bestreben, einen interoperablen Austausch für die Asset Administration Shell (AAS) zu schaffen, haben wir ein Generierungskonzept für drei verschiedene Protokolle (Interface Description, OpenAPI und SOAP) auf der Grundlage des formalisierten AAS-Metamodells innerhalb unserer domänenspezifischen Sprache erstellt. Derzeit liegt unser Schwerpunkt auf der Bereitstellung der neuesten Version des AAS-Metamodells, wie in "Details of Asset Administration Shell - Part 2"[PIP21] veröffentlicht. Die daraus resultierenden Schnittstellenschemata für OpenAPI haben die Zustimmung der Joint Working Group ÄAS"der IDTA, der ÄG1"Plattform Industrie 4.0, des IDTA Workstreams ÄASünd der "Working Group 24 for Digital Twins"der International Electrotechnic Commission (IEC) erhalten. Diese

API-Definitionen können in der offiziellen Veröffentlichung der AAS durch das IDTA verwendet werden, die in einem GitHub-Repository zugänglich ist [PIP21]. Mit Blick auf die Zukunft erstreckt sich unsere Vision auf die Bereitstellung einer Toolchain, die interoperable Implementierungen von I4.0-Komponenten ermöglicht und dabei das in [Br23b] beschriebene domänenspezifische Modell von AAS nutzt. Als Teil dieser Vision veröffentlichen wir eine Reihe von SDKs, Frameworks und Bibliotheken, die auf die Implementierung von sprachspezifischen Lösungen zugeschnitten sind. Diese Werkzeuge werden ausgestattet sein, um:

1. lesend und schreibend auf die AAS-Schnittstellen über die verschiedenen Austauschformate zuzugreifen,
2. Definitionen in verschiedenen IDLs wie OpenAPI, WSDL, SOAP, GraphQL Thrift, usw. bereitzustellen,
3. Constraints zu überprüfen sowie den AAS-API-Endpunkt und die Zugriffsmethoden zu validieren,
4. Constraints zu überprüfen und die Nutzdaten zu validieren, und
5. andere domänenspezifische Schnittstellen wie HTTP OPC-UA, MQTT, AMQP, usw. zu verwenden.

Unsere Roadmap umfasst die Erstellung von APIs für verschiedene Schnittstellen wie HTTP/REST, MQTT, OPC, AMQP und gRPC. Außerdem planen wir, unsere Unterstützung auf verschiedene Schnittstellenbeschreibungssprachen wie Protobuf, SOAP, WSDL, GraphQL und andere zu erweitern. Dies entspricht unserem Engagement, die Interoperabilität zu fördern und umfassende Lösungen für verschiedene I4.0-Ökosysteme anzubieten.

Literaturverzeichnis

[AA] AAS API Definitions. https://github.com/admin-shell-io/aas-specs-api. [Accessed 2024-09-02].

[AAE06] Awad, Ahmed M.; Aboulnaga, Hamed A.; Elmagarmid, Ahmed K.: A Survey of Web Services Description Languages. ACM SIGMOD Record, 35(2):106–115, 2006.

[BK21] Bogdanas, C.; Kuncak, V.: A Framework for Code Generation from Formal Specifications. IEEE Transactions on Software Engineering, 47(12):2980–3000, 2021.

[Br23a] Braunisch, N.; Ristin, M.; Lehmann, R.; Wollschlaeger, M.; van de Venn, H. W.: Generation of Digital Twins for Information Exchange Between Partners in the Industrie 4.0 Value Chain. In: Industrial Informatics (INDIN). 2023.

[Br23b] Braunisch, Nico; Ristin-Kaufmann, Marko; Lehmann, Robert; Wollschlaeger, Martin; van de Venn, Hans Wernher: Empowering Industry 4.0 with Generative and Model-Driven SDK Development. In: Industrial Electronics Society (IECON). S. 1–6, 2023.

[ic] icontract. https://pypi.org/project/icontract. [Accessed 2023-4-9].

[JZZ22] Jiang, C.; Zhang, Y.; Zhang, C.: An Automatic Code Generation System for Adaptive Systems. IEEE Transactions on Software Engineering, 48(12):3182–3198, 2022.

[LL21] Liu, X.; Liu, Y.: A Model-Driven Approach for Generating Dynamic Web Services from Service-Oriented Architectures. IEEE Software, 38(3):44–52, 2021.

[LZL22] Li, Y.; Zhang, D.; Luo, J.: A Systematic Approach to Interface Design and Implementation for Enterprise Applications. IEEE Transactions on Software Engineering, 49(7):1341–1359, 2022.

[Ny21] Nystrom, R.: Crafting Interpreters. Genever Benning, 2021.

[PIP21] Plattform I4.0 (Publisher): Details of the Asset Administration Shell. Part 2 - Interoperability at Runtime - Exchanging Information via Application Programming Interfaces (Version 1.0RC02). Bericht, 11 2021.

[PIP22] Plattform I4.0 (Publisher): Details of the Asset Administration Shell. Part 1 The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC02). Bericht, 05 2022.

[PP22] Patel, N. S.; Patel, H.: A Survey on Code Generation Tools for Java Applications. IEEE Transactions on Software Engineering, 47(9):1386–1412, 2022.

[Pya] Python Abstract Syntax Tree. https://docs.python.org/3/library/ast.html. [Accessed 2023-4-9].

[Pyb] Python Language. https://www.python.org. [Accessed 2023-4-9].

[SLW20] Sadi, A.; Li, N.; Wang, L.: A Comparative Analysis of Interface Generation Features in Integrated Development Environments. IEEE Transactions on Software Engineering, 45(10):1978–2000, 2020.

[So] Source Code of aas-core-meta. https://zenodo.org/record/7807680. [Accessed 2023-4-9].

[WHJ20] Wang, H.; Hu, X.; Jiang, X.: A Framework for Dynamic Interface Generation in Java Applications. IEEE Transactions on Software Engineering, 45(6):739–765, 2020.

[WZZ21] Wang, X.; Zhang, Z.; Zhang, Y.: An Empirical Study on Interface Generation in Java IDEs: Usage Patterns and Preferences. IEEE Transactions on Software Engineering, 46(10):2124–2142, 2021.

[YZL19] Yang, H.; Zhang, A.; Luo, J.: A Template-Driven Approach for Generating Java Interfaces from Class Diagrams. IEEE Transactions on Software Engineering, 44(5):635–653, 2019.

Cloud-native und -agnostische Digital Twin-Plattformen zur Umsetzung des digitalen Produktpasses: Eine Motivation

Magnus Redeker[1], Sven Simikin[2] and Frank Marek[3]

Abstract: Interoperable digital Twins (dt. digitale Zwillinge, DT) vereinfachen und erleichtern die Umsetzung von Industrie 4.0 (I4.0)-Anwendungsfällen wie (kollaborative) Zustandsüberwachung (engl. collaborative condition monitoring, CCM), CO2-Fußabdrücke von Produkten und Unternehmen (engl. product and corporate carbon footprints, PCF, CCF) und des digitalen Produktpasses (engl. digital product passport, DPP) – mit den Zielen der Steigerung der Resilienz und Ressourceneffizienz, der Einhaltung gesetzlicher Anforderungen und der nachhaltigen Stärkung der Wertschöpfung. Plattformen beschleunigen die Entwicklung und den Einsatz von DTs und Services und sorgen für deren stabilen Betrieb, um insbesondere alle relevanten Informationen aus dem Kerngeschäft zu erfassen, beispielsweise um PCFs, CCFs und DPPs zu aggregieren oder Nachhaltigkeitskennzahlen zu dokumentieren – und zwar vollständig automatisiert.

Plattformbetreiber sollten vor allem Vendor-Lock-ins vermeiden, um jederzeit flexibel auf die Anforderungen der Anwendungsfälle und Veränderungen der Vendor-Rahmenbedingungen reagieren zu können. Die Technologie muss überschaubar und die Kosten beherrschbar sein. Ziel ist es, ein Gleichgewicht zwischen Sicherheit, Kosten, Zukunftsfähigkeit und Beherrschbarkeit, kurzer Implementierungszeit und motivierter Nutzung zu erreichen. Solche Anforderungen können derzeit nur mit Cloud-Infrastrukturen erfüllt werden. Ein natives Design ist ebenso Voraussetzung für die Erreichung der Ziele wie ein agnostischer Entwurf, der darüber hinaus sicherstellt, dass ein Wechsel der Infrastruktur möglich ist und sowohl sicher als auch schnell erfolgen kann.

Zur Umsetzung des DPP wird in diesem Paper die Verwendung zustandsloser Cloud-nativer und -agnostischer DT-Plattform motiviert, die spezialisierte, zustandslose, lose gekoppelte DTs und Services bereitstellt, die innerhalb eines Service-Mesh interagieren. Diese Plattformen können problemlos in jeder öffentlichen, privaten und hybriden Cloud ausgeführt werden und lassen sich ohne weiteres horizontal skalieren.

Keywords: Digital Twin-Plattformen, Digitaler Produktpass, Industrie 4.0

1 Einleitung

Die Ecodesign for Sustainable Products Regulation (ESPR) ist integraler Bestandteil des Green Deal der Europäischen Union (EU) [Thd; The]. Der Green Deal ist eine umfassende EU-Initiative mit dem Ziel, Europa bis 2050 klimaneutral aufzustellen und eine nachhaltige und ressourceneffiziente Wirtschaft aufzubauen. Die Netto-Treibhausgasemissionen müssen im Vergleich zu 1990 bereits bis 2030 um mindestens 55% reduziert werden. Die Erstellung

1 Fraunhofer IOSB, IOSB-INA Lemgo, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Germany, magnus.redeker@iosb-ina.fraunhofer.de
2 New York University Shanghai, China,
3 Novatio-Solutions GmbH, Germany,

von Berichten über CO2-Fußabdrücke beginnt für die größeren Unternehmen bereits im Jahr 2025 [Thb].

ESPR unterstützt jene Ziele, indem sie Maßnahmen einfordert zur Förderung nachhaltiger Produkte und zur Verringerung der Umweltauswirkungen während der gesamten Produktlebenszyklen. Unter dem Strich geht es darum, die Nutzungsdauer von Produkten zu verlängern und den Verbrauch wertvoller Ressourcen zu minimieren.

Ein wichtiger und obligatorischer Teil des ESPR ist der digitale Produktpass (DPP) – ein digitales Dokument, das Informationen über den gesamten Lebenszyklus eines Produkts aggregiert und bereitstellt, einschließlich der Umweltauswirkungen, Materialzusammensetzung, Reparierbarkeit, Recyclingfähigkeit und Verwertbarkeit eines Produktes. Der DPP soll Verbrauchern, Herstellern und Behörden helfen, fundierte Entscheidungen zu treffen und Transparenz entlang der Lieferketten zu erzeugen.

Das DPP4.0-Konzept des Verbands der Elektro- und Digitalindustrie (ZVEI) und der Industrial Digital Twin Association (IDTA) bietet einen geeigneten technologischen Rahmen zur Umsetzung der ESPR-/DPP-Anforderungen [Di23; Ine]. Er basiert auf dem Asset Identification Link (IL) der IEC 61406 [IEa] und der Asset Administration Shell (AAS) der IEC 63278 [IEb; Inc; Ind]. Der DPP4.0 ist so konzipiert, dass Produzenten und Hersteller nicht nur in die Lage versetzt werden, die regulatorischen Anforderungen zu erfüllen, sondern darüber hinaus allen Marktteilnehmern neue digitale Services anzubieten – die notwendigen technologischen Implementierungen zur Erfüllung der regulatorischen Anforderungen können zur Schaffung zusätzlicher Mehrwerte genutzt werden.

Die wesentliche Grundlage zur Umsetzung des DPP4.0 ist eine zuverlässige, skalierbare und weitestgehend automatisierte DT-Plattform.

Die Unternehmen setzen Technologien nach strengen Kriterien ein. Eines der wichtigsten ist Vertrauen durch Kontrolle. Entscheider befürworten den Einsatz von Technologien, wenn hinreichend belegt ist, dass die Technologie für das Unternehmen sinnvoll und handhabbar ist und die Kosten des Einsatzes beherrschbar sind. Ziel ist es, ein Gleichgewicht zwischen Sicherheit, Kosten, Zukunftsfähigkeit und Kontrollierbarkeit, kurzer Implementierungszeit und motivierter Nutzung zu erreichen. Entscheidend für die Akzeptanz ist, dass die Methoden, Inhalte und Prozesse auch für den technisch versierten Fachmann einfach zu bedienen sind. Diese Anforderungen können derzeit nur mit Cloud-Lösungen erfüllt werden.

In diesem Papier wird die Nutzung einer zustandslosen Cloud-native und -agnostische Plattform zur Umsetzung des DPP4.0 motiviert, die Zuverlässigkeit, Skalierbarkeit und Portabilität gewährleistet. Spezialisierte, zustandslose, lose gekoppelten Cloud-native DT-Services können innerhalb eines Service-Mesh interagieren. Gemeinsam mit den integrierten Assets bilden die DT-Services interoperable Ökosysteme, in denen Industrie 4.0 (I4.0)-Anwendungsfälle kosteneffizient implementiert werden können.

Cloud-native und -agnostische Plattformen sind in jeder Cloud-Umgebung lauffähig: öffentliche, private und hybride Clouds. Cloud-Provider-Lock-ins werden per Design vermieden. Darüber hinaus ist die Ausführung in hybriden Clouds besonders attraktiv, da sensible Daten in privaten Clouds gut geschützt sind und gleichzeitig Services in öffentlichen Clouds für die Partner angeboten werden können.

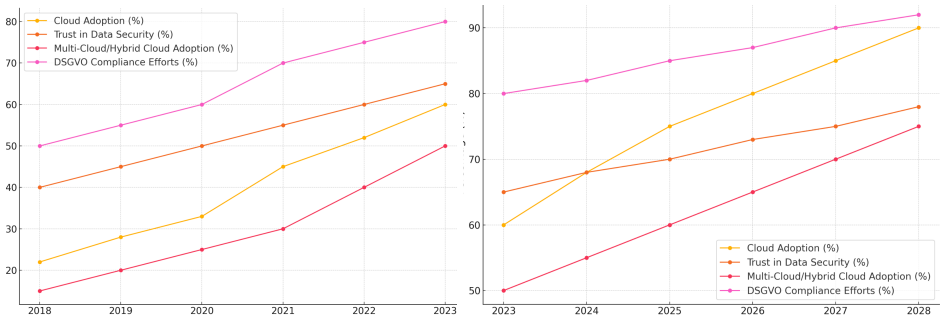
Die Gliederung dieses Papers ist wie folgt. In Abschnitt 2 wird der Bedarf der Unternehmen an Cloud-Lösungen und DT diskutiert und in Abschnitt 3 wird der aktuelle Stand der Technik bzgl. Cloud-Plattformen und AAS-basierten DTs dargelegt. Abschnitt 4 beschreibt die Anwendung zustandsloser Cloud-nativer und -agnostischer DT-Plattformen zur Umsetzung des DPP4.0. Schließlich fasst Abschnitt 5 dieses Paper zusammen.

2 Der Bedarf der Unternehmen an Cloud-Lösungen und digital Twins

Für die Nutzung und Verbreitung der hier beschriebenen Technologie ist es notwendig, zwischen verschiedenen Nutzergruppen zu unterscheiden. Im Allgemeinen betrachten wir die Gruppe der Technologieanwender, die alles von einfachen Geräten bis hin zu hochkomplexen Systemen nutzen oder bedienen. Die DT-Technologie eignet sich für die Beschreibung und den Betrieb beliebig komplexer Szenarien. Wir gehen davon aus, dass nur Institutionen und größere Unternehmen hochkomplexe Systeme verwenden. Jedoch ist die Relevanz und der Nutzen des DT bereits heute als kritischer Pfad erkennbar, unabhängig von der Größe des Unternehmens.

Aufgrund der Alterung der Gesellschaft und des zunehmenden Fachkräftemangels besteht ein Bedarf an Automatisierung und Vereinfachung bei der Nutzung von Technologien, in diesem Fall des DT. Der kritische Pfad wird immer dynamischer, sodass ausscheidende Mitarbeitende kaum die Möglichkeit und vor allem die Zeit haben werden, ihr Wissen weiterzugeben. Die Anzahl der Erwerbspersonen in Deutschland wird bis 2030 voraussichtlich um etwa 3,5 Millionen Menschen (8 Prozent) zurückgehen [Be18]. Dies kann nur teilweise durch zunehmende Automatisierung kompensiert werden. Die Bewahrung von Wissen muss kontinuierlich in Systemen erfolgen. Diese Systeme müssen den Lehrplänen von Schulen, Universitäten und Berufsausbildungsprogrammen entsprechen und allgemein verfügbar sein. Sie müssen auch schnell installiert werden können. Unternehmen stehen vor einer Mammutaufgabe: Die rasche Digitalisierung aller Prozesse bei gleichzeitigem Personalverringerung und Umstrukturierungen.

Hinzu kommt die Öffnung des Marktes – der Wettbewerber wird zum Partner in einer interoperablen Wertschöpfungskette. Dies erfordert erhebliche Investitionen. Bestehende Systemlandschaften müssen homogenisiert werden, um Prozesse zu vereinfachen und Kosten zu senken. Um die beschriebenen Szenarien umzusetzen, benötigen Unternehmen Administratoren und DevOPS-Ingenieure, die in der Regel Engpassressourcen sind. Der Fachkräftemangel kann in absehbarer Zeit nicht ausgeglichen werden. Bis 2060 wird es in Deutschland einen Mangel an einem Drittel aller benötigten Fachkräfte geben.



(a) Trends 2018 to 2023 (b) Prognosen 2024 to 2028

Abb. 1: Darstellung der Trends bei der Einführung von Cloud- und Multi-/Hybrid-Cloud-Services: Vertrauen in die Datensicherheit und Bemühungen zur Einhaltung der DSGVO in den Jahren 2018–2023 sowie Prognosen für 2024–2028. Darstellung jeweils als prozentualer Anteil an Unternehmen. Die Daten wurden auf der Grundlage mehrerer Studien des Statistischen Bundesamtes und des Europäischen Statistikamtes generiert [Eu; Fe].

Unternehmen implementieren Technologien, wenn sie die Technologie und ihre Kosten effektiv verwalten können. Das Ziel besteht darin, ein Gleichgewicht zwischen Sicherheit, Kosten, Nachhaltigkeit, Kontrolle, schneller Implementierung und motivierter Nutzung herzustellen. Derzeit werden erhebliche Anstrengungen unternommen, um sichere Datenräume zu schaffen, die den Widerstand gegen Veränderungen verringern.

Der Umgang mit Cloud-Anbietern bezieht sich in der Regel auf einen Anbieter, wie z. B. Microsoft, der bereits mit anderen Produkten im Unternehmen verankert ist und eine Service-Landschaft aufgebaut hat. Die Motivation, in einer Multi-Cloud oder einer erweiterten agnostischen Cloud zu arbeiten, ist oft gering und wird von den Cloud-Anbietern nicht unterstützt. Verträge mit Cloud-Anbietern werden nicht von IT-Abteilungen, sondern von Rechtsabteilungen geregelt, die wiederum lieber mit direkten Vertragspartnern zusammenarbeiten und fixe Verträge bevorzugen, statt variable Verträge, die für eine temporäre Nutzung über das Internet abgeschlossen werden. Darüber hinaus führen Aspekte wie die erforderliche Bereitstellung einer Kreditkarte dazu, dass Unternehmen lieber mit Vertragspartnern zusammenarbeiten als direkt in der Cloud zu bestellen.

Darüber hinaus haben Industriekunden den Fokus und die Verpflichtung, vorhersehbare Auswirkungen auf die IT-Infrastruktur zu identifizieren. Sie müssen sicherstellen, dass ihre Organisation jederzeit einsatzbereit ist: Produktions- und Lieferbereitschaft. Sie legen besonderen Wert auf die Verfügbarkeit ihrer Systeme. Diese Nutzer betreiben in der Regel eine On-Premise-Cloud als Backup, die somit Teil einer Multi-Cloud wird.

Insgesamt steigt das Vertrauen in Cloud Computing im Vergleich zu herkömmlichen On-Premise-Lösungen weltweit. Die bestehenden Nachteile in Bezug auf Datensicherheit, Anbieterabhängigkeit und die Notwendigkeit einer zuverlässigen Internetverbindung für den

Zugriff auf Cloud-Services sowie das standardisierte Leistungsspektrum der Cloud-Anbieter schränken den Anwendungsbereich ein und werden On-Premise-Lösungen in absehbarer Zukunft nicht ersetzen.

Hybridlösungen bieten die größten Vorteile und die größte Flexibilität [KP]. Die EU-Datenschutz-Grundverordnung (DSGVO) ist ein Treiber für objektives Vertrauen [Thc], ebenso wie EU's Intelligence (AI) Act [Tha] und die europäische Initiative Gaia-X [Ga]. Die Vorteile des DT werden nur durch reale industrielle Lösungen wie Catena-X und Factory-X Akzeptanz erzeugen. Für jene Akzeptanz ist es entscheidend, dass die Methoden, Inhalte und Prozesse für den technisch versierten Experten einfach zu handhaben sind. Diese Anforderungen können derzeit nur mit Cloud-Infrastrukturen erfüllt werden. Der anbieterunabhängige Ansatz stellt sicher, dass Wechsel der Infrastruktur möglich sind und sicher und schnell durchgeführt werden können.

Die Nutzung von Cloud-Technologien hat sich zwischen 2018 und 2023 rasant entwickelt [Jo]: Die Wachstumsraten lagen bei bis zu 34% pro Jahr. Abbildung Abb. 1 veranschaulicht das Wachstum der Cloud-Nutzung, das Vertrauen in die Datensicherheit und die Bemühungen zur Einhaltung der DSGVO in den letzten fünf Jahren. Dieses Wachstum wird sich entsprechend der Prognosen fortsetzen.

3 Stand der Technik

3.1 Cloud-native und -agnostische Plattformen

Das Aufkommen des Cloud Computing hat die Art und Weise, wie Unternehmen arbeiten, revolutioniert, da es einen bedarfsgerechten Zugang zu Computeressourcen über das Internet bietet. Obwohl die öffentliche Cloud erhebliche Vorteile in Bezug auf Skalierbarkeit, Benutzerfreundlichkeit und Kosten bietet, ist sie wahrscheinlich nicht für alle Unternehmen geeignet, insbesondere nicht für solche mit sensiblen Daten und strengen gesetzlichen Anforderungen [Ma11].

Private-Cloud-Anwendungen können unter Verwendung offener Standards und in einer Cloud-agnostischen Architektur erstellt werden, so dass sie auf jeder Cloud-Infrastruktur laufen können, einschließlich öffentlicher, privater und hybrider Clouds. Diese Flexibilität ist einer der wesentlichen Vorteile von Private-Cloud-Anwendungen, da sie es Unternehmen ermöglicht, jederzeit den am besten geeigneten Cloud-Anbieter zu wählen und entsprechend die Gefahr von Vendor-Locks-ins bei Cloud-Anbietern zu mindern. [OST16].

Cloud-native Technologien ermöglichen es Unternehmen, skalierbare Anwendungen in modernen, dynamischen Umgebungen wie öffentlichen, privaten und hybriden Clouds zu erstellen und auszuführen; Container, Service Meshes, Microservices, immutable Infrastruktur und deklarative APIs sind Beispiele für diesen Ansatz [Cla]. Zustandslose Microservices bieten zahlreiche Vorteile wie erhöhte Flexibilität, Skalierbarkeit und Zuverlässigkeit. Einer

der Hauptvorteile ist ihre lose Kopplung, die es ermöglicht, einzelne Services unabhängig von anderen Services zu entwickeln, zu testen und bereitzustellen. Dies ermöglicht eine schnelle Entwicklung neuer Funktionen ohne den Rest der Anwendung zu beeinträchtigen.

Microservice-Architekturen werden in der Regel in Container-Laufzeitumgebungen umgesetzt, die automatisches Lebenszyklusmanagement, ausgefeilte Vernetzung und Skalierung bieten. Die Services innerhalb einer Microservice-Architektur können im Allgemeinen in Plattform- und Kern-Services unterteilt werden. Plattform-Services bieten grundlegende Funktionen wie Messaging, Service Discovery und Lastausgleich, die für die Verwaltung von Microservice-basierten Anwendungen im großen Maßstab unerlässlich sind. Kern-Services bilden die spezifische Geschäftslogik einer Plattform ab.

Mit zunehmender Komplexität von Microservice-Architekturen wird deren Verwaltung immer schwieriger. Hier kann ein Service Mesh erhebliche Vorteile bieten: Eine zentralisierte Verwaltung und Konfiguration wesentlicher Funktionen, einschließlich Sicherheit, Service Discovery, Lastausgleich, Trafficmanagement und Überwachung [W 19]. Ein Service Mesh ermöglicht es Entwicklern, sich auf die Entwicklung und Bereitstellung von Services zu konzentrieren, ohne sich um die zugrunde liegende Infrastruktur kümmern zu müssen.

Die Cloud Native Computing Foundation (CNCF) Service-Landschaft [Clb] umfasst Cloud-native Open-Source-Projekte, die für i) die Bereitstellung von Plattformen, ii) die Bereitstellung von Kern- und Plattform-Services, iii) die Konfiguration von Service Meshes und iv) als Plattform-Services wie Gateways, Monitoring, Tracing, Logging, Messaging und Dashboards verwendet werden können.

3.2 AAS-basierte digital Twins, Server und Plattformen

Das Referenzarchitekturmodell Industrie 4.0 (RAMI4.0) beinhaltet das Konzept einer I4.0-Komponente, bestehend aus Asset und Asset Administration Shell (AAS) [DI16; IEb]. Ein Asset definiert ein physisches oder logisches Objekt, das sich im Besitz oder unter der Obhut einer Organisation befindet und das einen angenommenen oder tatsächlichen Wert für die Organisation hat [Inc]. Das AAS-Konzept unterstützt bei der Implementierung von DTs für I4.0 und der Schaffung von Interoperabilität zwischen den Lösungen verschiedener Anbieter [Inb], wobei ein digital Twin eine digitale Darstellung eines Assets definiert, die den Anforderungen der umzusetzenden Anwendungsfälle genügt [Inf].

AAS- und Submodel-Registries und -Server ermöglichen eine standardisierte Interaktion mit AAS-basierten DTs und deren verlinkten Teilmodellen [Inc; Ind]. Standardisierte (und ggf. proprietäre) Teilmodelle integrieren diejenigen Elemente, die zur Umsetzung der Anwendungsfällen benötigt werden [Ina].

Die populären Open-Source *Eclipse AASXServer, BaSyx*, und *FA^{3}ST* Server-Komponenten [Ec] sind konform zu den AAS-Spezifikation [Inc; Ind] und folglich miteinander kompatibel. Jedoch sind sie im aktuellen Entwicklungsstand nicht Cloud-nativ.

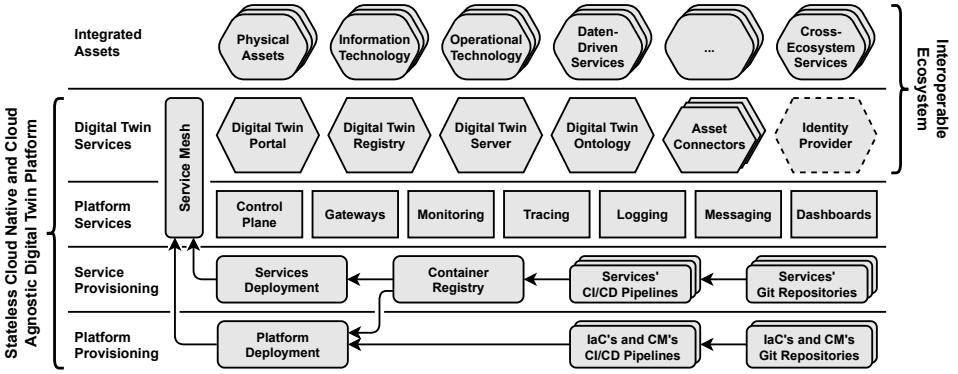


Abb. 2: Eine zustandslose Cloud-native und -agnostische Digital Twin-Plattform [M 24]: Assets und ihre digital Twins bilden ein interoperables Ökosystem, welches eine einfache und zielgerichtete Umsetzung von Anwendungsfällen wie (kollaborative) Zustandsüberwachung, CO2-Fußabdrücke von Produkten und Unternehmen und digitale Produktpässe ermöglicht.

4 Cloud-native und -agnostische Digital Twin-Plattformen zur Umsetzung des digitalen Produktpasses

Die in [M 24] entwickelte zustandslose Cloud-native und -agnostische DT-Plattform besteht, wie in Abb. 2 visualisiert, aus drei Schichten für die Bereitstellung der Basisplattform und Services sowie einer darauf aufbauenden Schicht, die die Geschäftslogik bereitstellt.

Die durch die DT-Services integrierten Assets bilden ein interoperables Ökosystem, welches die einfache Umsetzung von Anwendungsfällen ermöglicht. Darüber hinaus können ökosystemübergreifende Konnektoren verschiedene Ökosysteme miteinander verbinden und somit Anwendungsfälle über Systemgrenzen hinweg ermöglichen. Die Plattform basiert auf modernsten Cloud-nativen und -agnostischen Architekturmustern und passt sich den spezifischen Anwendungsfällen an.

Jeder DT-Service der Plattform erfüllt die in Abschnitt 3 beschriebenen Cloud-Native-Anforderungen, d.h. sie sind solide aufgebaut, werden kontinuierlich weiterentwickelt und integriert, sind API-gesteuert und bieten Monitoring-, Tracing- und Protokollierungsfähigkeiten. Diese Services sind zustandslos, was bedeutet, dass jeder Zustand (persistierte Daten jeglicher Art) außerhalb eines Containers gespeichert wird, in einer beliebigen Anzahl von Orten – wie Ereignis- oder Systemprotokollen, relationalen Datenbanken und Dokumenten- oder Objektspeichern [IB].

Ein Container kann jederzeit sauber heruntergefahren und entfernt werden, ohne dass ein Datenverlust zu befürchten ist. Wenn ein neuer Container erstellt wird, um den alten zu ersetzen oder von einer Überlast zu befreien, verbindet sich der neue Container einfach mit demselben Datenspeicher [Go], was die horizontale Skalierung des Services erheblich

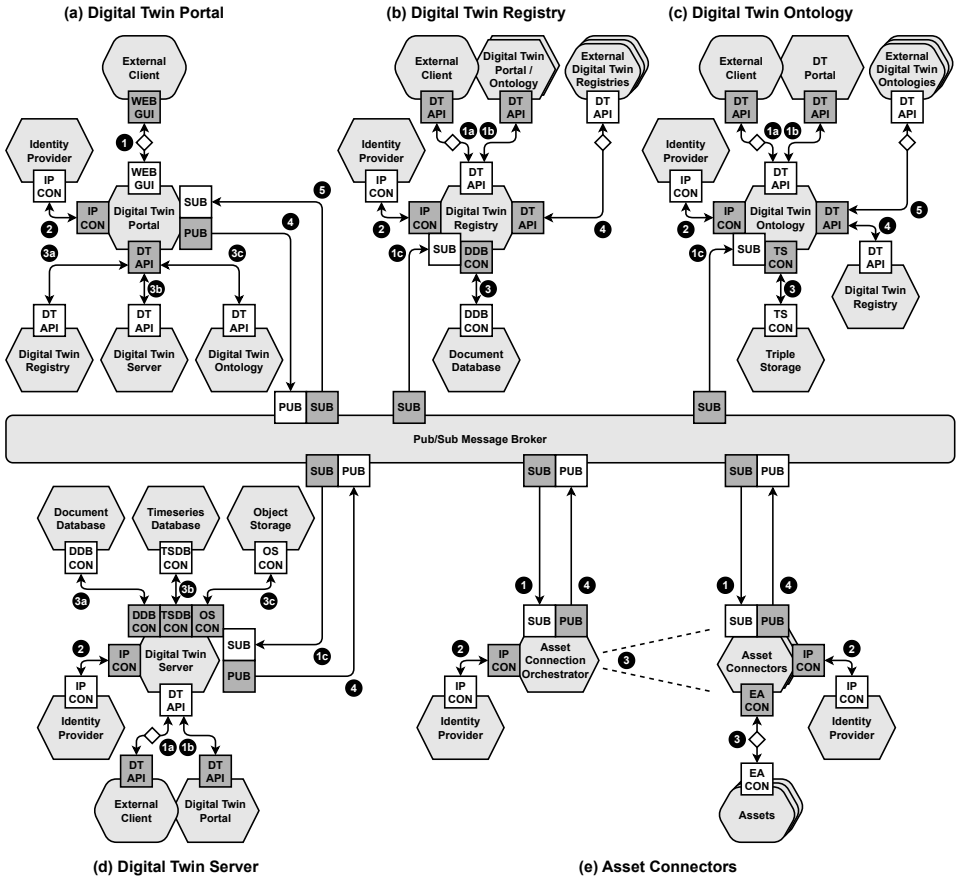


Abb. 3: Interaktionen innerhalb der Digital Twin-Plattform von Digital Twin Portal, Registry, Server, Ontology und ihren individuellen Datenbanken sowie Asset Connection Orchestrator und Asset-Connectors [M 24]. Pfeile visualisieren den Datentransfer, graue Rechtecke die auslösenden Kommunikationspartner und weiße Rechtecke deren Gegenstücke. Die Zahlen geben Prozessschritte an. Anfragen und Antworten, die in die Digital Twin-Plattform eingehen oder sie verlassen, passieren die Gateway-Services der Plattform, die durch eine weiße Raute gekennzeichnet sind. Der Asset Connection Orchestrator konfiguriert, implementiert und entfernt einzelne Asset-Connectors.

vereinfacht. Die Verwaltung von zustandsbehafteten Anwendungen und verteilten Systemen auf Kubernetes hingegen ist ein umfangreiches und komplexes Thema [Ku].

Jeder DT-Service verfügt über eine AAS, die in der AAS-Registry der Plattform registriert ist. Folglich ist jeder DT-Service für autorisierte Clients auffindbar und auslösbar. *DT Portal*, *DT Register*, *DT Server* und *DT Ontology* bieten Zugang für plattformexterne Clients, wobei

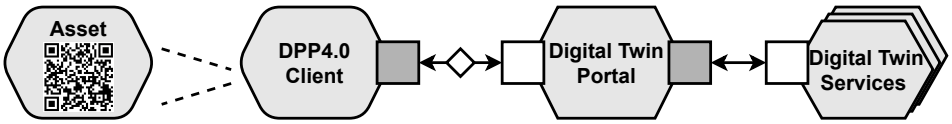


Abb. 4: Abstrahierter Abruf eines digitalen Produktpasses: Ein Client verbindet sich über einen Asset-Identification-Link, der vom Asset abgescannt wird, mit dem Digital Twin Portal, welches asset- und kundenspezifische Daten aus den weiteren DT-Services der Plattform in den DPP4.0 integriert und diesen an den Client übergibt.

das DT-Portal die erste und einzige direkte Anlaufstelle für Clients ist, die einen DPP4.0 anfordern. Intern sind die DT-Services entkoppelt und interagieren durch asynchrones Messaging, wobei die Topics eindeutig durch die Strukturen der AAS der DT-Services und der integrierten Assets spezifiziert sind. Detaillierte Beschreibungen der Plattform- und DT-Services finden sich in [M 24].

Der DPP4.0 versetzt Unternehmen in die Lage, die ESPR-/DPP-Anforderungen zu erfüllen und gleichzeitig digitale Services auf dem Markt anzubieten [Di23; Ine; Thd]. Wie in Abschnitt 1 beschrieben, basiert das Konzept auf einem Link zur Identifizierung eines Assets, der physisch an dem Asset angebracht ist, z.B. QR-Code oder RFID-Chip, und der vom Hersteller des Assets bereitgestellten AAS.

Für einen externen DPP4.0-Client kann, im Falle einer zustandslosen DT-Plattform, die erste und einzige direkte Anlaufstelle das DT-Portal sein, siehe Abb. 4. Ein Client verbindet sich mit dem DT-Portal bspw. über einen Link zur Identifizierung eines Assets, der die herstellerspezifische ID des Assets enthält [Schritt 1 in Abb. 3]. Unter der Voraussetzung, dass die Berechtigung des Kunden bestätigt wird [Schritt 2 in Abb. 3], generiert das DT-Portal in Zusammenarbeit mit den weiteren DT-Services eine asset- und kundenspezifische Sicht auf die AAS und Submodelle des Assets [Schritt 3 in Abb. 3] und sendet den angeforderten DPP4.0 an den Client zurück.

Die DPP-relevanten Assetdaten werden durch die DT-Services beim Entstehen aus den Quellsystemen in die AAS und Submodelle des Assets integriert. Quellsysteme können beispielsweise Produktentwicklungs-, Warenwirtschafts- und Fertigungssysteme sowie Zeitreihendatenbanken sein.

Unter dem Strich ist die Umsetzung von Anwendungsfällen wie dem DPP4.0 mit der entwickelten zustandslosen Cloud-nativen und -agnostischen DT-Platform aufgrund des hohen Automatisierungsgrades der Plattform einfach und kostengünstig.

5 Conclusion

Die alternde Gesellschaft und der Rückgang der Anzahl der Erwerbspersonen im Allgemeinen, spezifische regulatorische Anforderungen, wie die Berichterstattung über den CO2-Fußabdruck von Unternehmen und Produkten eingebettet in digitale Produktpässe (DPP), sowie die Möglichkeit, traditionelle Märkte durch das Angebot zusätzlicher digitaler Services zu erweitern, bedingen bei Wirtschaftsunternehmen einen Bedarf an technologischer Automatisierung und Vereinfachung. Darüber hinaus werden Wettbewerber zu Partnern in interoperablen Wertschöpfungsketten, so dass bestehende Systemlandschaften homogenisiert werden, um Prozesse zu straffen und Kosten zu senken und so einen reibungslosen Betrieb und ein nachhaltiges Unternehmenswachstum zu unterstützen. Ziel der Unternehmen ist es, ein Gleichgewicht zwischen Sicherheit, Kosten, Realisierbarkeit und Kontrolle, schneller Implementierung und motiviertem Technologieeinsatz herzustellen.

Diese Paper motivierte entsprechend den Einsatz Cloud-nativer und -agnostischer digital Twin (DT)-Plattformen zur Umsetzung des DPP, der gemäß des ZVEI-Konzeptes DPP4.0 auf Asset Identification Links (IL, IEC 61406) und Asset Administration Shell-basierten DTs (AAS, IEC 63278) basieren kann. Aufgrund des Cloud-nativen und -agnostischen Plattformdesigns und der Zustandslosigkeit der Services ist die notwendige horizontale Skalierbarkeit garantiert und Cloud-Provider Lock-ins werden vermieden. Insbesondere wird das angestrebte Gleichgewicht zwischen Sicherheit, Kosten, Zukunftsfähigkeit und Beherrschbarkeit, kurzer Implementierungszeit und motivierter Nutzung erreicht.

DPP-relevante Assetdaten können durch die DT-Services einer solchen DT-Plattform beim Entstehen der Daten aus den Quellsystemen in AASs der Assets automatisiert integriert werden. Externe Clients können sich mit einem DT-Portal zum DPP-Abruf verbinden. Auf Abruf und in Zusammenarbeit mit den weiteren DT-Services kann ein solches Portal eine Asset- und Client-spezifische Sicht auf den DPP aggregieren.

Acknowledgment

Das Forschungs- und Entwicklungsprojekt „Industrie 4.0 Ökosystem für den automatisierten Einsatz von datengetriebenen Services“ (I4.0AutoServ) wird mit Mitteln des Ministeriums für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen (MWIKE) im Rahmen des Spitzenclusters Intelligente Technische Systeme OstWestfalenLippe (it‘s OWL) gefördert und vom Projektträger Jülich (PtJ) betreut.

Das Forschungs- und Entwicklungsprojekt „Energieeffiziente Analyse und Steuerungsprozesse im dynamischen Edge-Cloud-Kontinuum für die industrielle Fertigung“ (EASY) wird mit Mitteln des Bundesministeriums für Wirtschaft und Klimaschutz (BMWK) gefördert und durch DLR Projektträger (DLR-PT) betreut.

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literaturverzeichnis

[Be18] Bertelsmann Stiftung: Arbeitskräfte und Arbeitsmarkt im demographischen Wandel, 2018, URL: https://doi.org/10.11586/2018009.

[Cla] Cloud Native Computing Foundation: CNCF Cloud Native Definition v1.0, URL: https://github.com/cncf/toc/blob/main/DEFINITION.md.

[Clb] Cloud Native Computing Foundation: CNCF Cloud Native Interactive Landscape, URL: https://landscape.cncf.io/.

[DI16] DIN: DIN SPEC 91345: Reference Architecture Model Industrie 4.0 (RAMI4.0), 2016, URL: https://dx.doi.org/10.31030/2436156.

[Di23] Dieter Wegener, VP Siemens AG & ZVEI-Speaker Industrie 4.0: DPP4.0 – Big Picture. In: IDTA AAS Tech Days. 2023, URL: https://www.zvei.org/fileadmin/user_upload/Themen/Industrie/Fachverband_Automation/2023-09-15_IDTA_AAS_Tech_Days_DPP4.0_Wegener.pdf.

[Ec] Eclipse Foundation: Eclipse Digital Twin, URL: https://projects.eclipse.org/projects/dt.

[Eu] European Union: European Statistical Office, URL: https://ec.europa.eu/eurostat.

[Fe] Federal Statistical Office of Germany: Federal Statistical Office of Germany, URL: https://www.destatis.de/EN/Home/_node.html.

[Ga] Gaia-X European Association for Data and Cloud AISBL: Gaia-X, URL: https://gaia-x.eu/.

[Go] Google Cloud Arch Center: Best practices for operating containers, URL: https://www.ibm.com/topics/cloud-native.

[IB] IBM: What is cloud native?, URL: https://cloud.google.com/architecture/best-practices-for-operating-containers.

[IEa] IEC: IEC 61406-1:2022: Identification Link - Part 1: General requirements, URL: https://webstore.iec.ch/publication/67673.

[IEb] IEC: IEC 63278-1:2023 Asset Administration Shell for industrial applications - Part 1: Asset Administration Shell structure, URL: https://webstore.iec.ch/publication/65628.

[Ina] Industrial Digital Twin Association: AAS Submodel Templates, URL: https://industrialdigitaltwin.org/en/content-hub/submodels.

[Inb] Industrial Digital Twin Association: Asset Administration Shell Reading Guide, URL: https://industrialdigitaltwin.org/wp-content/uploads/2022/12/2022-12-07_IDTA_AAS-Reading-Guide.pdf.

[Inc] Industrial Digital Twin Association: Specification of the Asset Administration Shell Part 1: Metamodel – IDTA Number: 01001-3-0-1, URL: https://industrialdigitaltwin.org/wp-content/uploads/2024/06/IDTA-01001-3-0-1_SpecificationAssetAdministrationShell_Part1_Metamodel.pdf.

[Ind] Industrial Digital Twin Association: Specification of the Asset Administration Shell Part 2: Application Programming Interfaces – IDTA Number: 01002-3-0-2, URL: https://industrialdigitaltwin.org/wp-content/uploads/2024/06/IDTA-01002-3-0-2_SpecificationAssetAdministrationShell_Part2_API.pdf.

[Ine] Industrial Digital Twin Association e. V. and ZVEI e. V.: DPP4.0 – The Digital Product Passport for Industry 4.0, URL: https://dpp40.eu/.

[Inf] Industry IoT Consortium: IIC: The Industrial Internet of Things Vocabulary, URL: https://www.iiconsortium.org/pdf/Vocabulary-Report-2.3.pdf.

[Jo] Jones, E.: Cloud Market Share: A Look at the Cloud Ecosystem, URL: https://kinsta.com/blog/cloud-market-share/.

[KP] KPMG: Cloud-Computing bietet mehr als nur Kosteneffizienz, URL: https://kpmg.com/de/de/home/media/press-releases/2022/06/cloud-computing-bietet-mehr-als-nur-kosteneffizienz.html.

[Ku] Kubernetes: Documentation - Stateful Applications - StatefulSet Basics, URL: https://kubernetes.io/docs/tutorials/stateful-application/basic-stateful-set/.

[M 24] M. Redeker, S. Simikin and F. Marek: Towards a Stateless Cloud Native and Cloud Agnostic Digital Twin Platform for the Digital Product Passport: Out-of-the-box, Reliable, and Scaling. In: 29th IEEE ETFA. 2024.

[Ma11] Marston, S.; Li, Z.; Bandyopadhyay, S.; Zhang, J.; Ghalsasi, A.: Cloud computing — The business perspective. Decision Support Systems, 2011, URL: https://www.sciencedirect.com/science/article/pii/S0167923610002393.

[OST16] Opara-Martins, J.; Sahandi, R.; Tian, F.: Critical Analysis of Vendor Lock-in and Its Impact on Cloud Computing Migration: A Business Perspective. J. Cloud Comput. 2016, URL: https://doi.org/10.1186/s13677-016-0054-z.

[Tha] The European Commission: AI Act, URL: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[Thb] The European Commission: Corporate Sustainability Reporting, URL: https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en.

[Thc] The European Commission: Data protection in the EU, URL: https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en.

[Thd] The European Commission: Ecodesign for Sustainable Products Regulation, URL: https://commission.europa.eu/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/sustainable-products/ecodesign-sustainable-products-regulation_en.

[The] The European Commission: The European Green Deal, URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.

[W 19] W. Li et al: Service Mesh: Challenges, State of the Art, and Future Research Opportunities. In: IEEE SOSE. 2019, URL: https://doi.org/10.1109/SOSE.2019.00026.

Investigation of Complex Cybersecurity Attacks on Synchronisation and Time-Aware Shaping in a TSN-Network

Mike Neelen[1], Tobias Ferfers[1], Alexander Biendarra[1], Philippe Saey[2], Sebastian Schriegel[1], Henning Trsek[3], and Jürgen Jasperneite[1]

Abstract: Industry 4.0 requires a convergence of the *Information Technology (IT)* and *Operational Technology (OT)* world. Therefore, a common communication medium such as Ethernet is required to fulfill the demands of both worlds. *Time-Sensitive Networking* (TSN) was implemented by the *Institute of Electrical and Electronics Engineers (IEEE)* to enable the real-time capabilities for Ethernet, which are required in the OT world. However, TSN neglect security aspects, visible in the high amount of vulnerabilities according to prior research. This paper investigates the influence of given synchronisation attacks onto the associated Time-Aware Shaper (TAS) of the attacked device. For this a test bench was created, in which the effects of the attacks on the synchronisation are emulated. Hereby the influence of common attack patterns, such as a *Denial-of-Service (DoS), an Adversary-in-the-Middle (AitM)* and a replay attacks are emulated onto the synchronisation of the network. The impact of the attack emulation on the *TAS* was then investigated. The results show that the replay attack as well as the manipulation of the timestamps as an AitM, shift the time of frame transmission in their respective ways. The DoS attack on the synchronisation also lead to an asynchronous behaviour of the device's TAS. This however happened, because the PTP client stops its synchronisation with the connected device, after a given threshold of missed frames was exceeded. Furthermore, this impact on the *TAS* can be used as an indication for an intrusion detection system, that a potential attack on the synchronisation has occurred. This can be achieved by observing the behaviour of the gate events from the observed device.

Keywords: TSN, gPTP, Cyber-attacks, Network-diagnostics, Time-Aware-Shaping

1 Introduction

The combination of the *Information Technology (IT)* and *Operational Technology (OT)* requires a common communication medium for both worlds. *Ethernet* is a proper solution for this, since both worlds utilised it in their respective field already. However, for this to work, real-time requirements e.g. due to bounded low latency motion control applications in the *OT* world have to be implemented. *Time-Sensitive Networking (TSN)* is the solution for this issue, integrating real-time communication on the *MAC* layer per *IEEE* Standard, without relying on proprietary solutions on top. [WSJ17]

1 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Germany, mike.neelen@iosb-ina.fraunhofer.de; tobias.ferfers@iosb-ina.fraunhofer.de; alexander.biendarra@iosb-ina.fraunhofer.de; sebastian.schriegel@iosb-ina.fraunhofer.de; juergen.jasperneite@iosb-ina.fraunhofer.de
2 KU Leuven, ESAT-ELECTA, Gebroeders De Smetstraat 1, 9000 Gent, Belgium, philippe.saey@kuleuven.be
3 Technische Hochschule OWL, Institut Industrial IT, Campusallee 6, 32657 Lemgo, Germany, henning.trsek@th-owl.de

Two key technologies in a *TSN* network are the network synchronisation via the *generalised Precision Time Protocol (gPTP)* according to *IEEE 802.1AS-2020* as well as the *Time-Aware Shaper (TAS)* according to *IEEE 802.1Q-2018*. These two technologies are intertwined, since the *TAS* relies on a synchronisation of the device with the network. Nevertheless, the development of *TSN* standards neglects the aspect of security, visible in its high amount of vulnerabilities [Er21]. Furthermore, these two standards can be attacked by commonly used cyber-attacks, such as *replay attacks*, *Denial-of-Service (DoS)* attacks, or *Adversary in the Middle (AitM)* attacks, provided that the attacker has infiltrated the network [WLH22]. The relation between the two technologies, as well as the ease of attacking the synchronisation, leads to the following research questions (RQ's):

- RQ1: How do cyber-attacks influence the synchronisation?
- RQ2: How does the influence on the synchronisation from RQ1 impact the *Time-Aware Shaper*-functionality of the devices?

This paper aims to answer these two questions starting with a brief introduction of the topics themselves as State of the Art in section 2. Afterwards section 3 describes the methodology used in the paper. The test results are described in section 4 followed by a conclusion and future work in section 5.

2 State of the art

The State of the Art describes a brief description of *TSN* Standards and cyber-security attacks. Furthermore, an overview of related work in cyber attacks on *TSN* is given. Please refer to the sources provided for further information.

2.1 IEEE 802.1 Time-Sensitive Networking

Time-Sensitive Networking is standardised in the *Institute of Electrical and Electronics Engineers (IEEE)* 802.1 Working Group. Other parts of *IEEE* 802.1 include the maintenance and security of the the *MAC* layer. For the network to be time-sensitive, it first requires a common understanding of time via the synchronisation of all participants.

The *generalised Precision Time Protocol (gPTP)* is a profile of the *IEEE 1588 Precision Time Protocol (PTP)*. Hereby the devices utilise a hierarchical structure of *Time Transmitter (TT)* and *Time Receiver (TR)* with the main *Time Transmitter* called *Grandmaster* [IE24]. The *TT* transmits the time information towards all of the *gPTP*-capable participants as a timestamp embedded in the Frame. The *TR* will then compensate the timestamp provided by the *TT* with the path delay time, previously calculated via path_delay messages (See left-hand side of Figure 1). The times t1-t4 are defined as the times at which the devices receive/ transmit the frames respectively. This time compensation is especially important,

when these participants have other *gPTP*-capable participants connected afterwards to relay the time including a delay compensation, in which case the path delay time is embedded into the frame via a correction field. The right hand side of Figure 1 shows a *two-step* method, where the timestamp is located in a follow-up frame. [IE20]

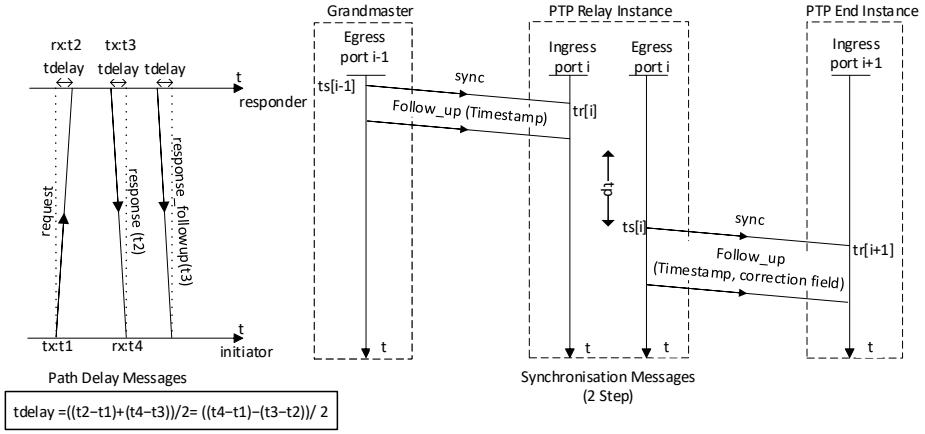


Fig. 1: gPTP message exchange (2-step)

The *Time-Aware Shaper* is defined in the *IEEE* 802.1Q standard. The main concept is a time division of the device's transmission. For this, the duration of time slots as well as a transmission cycle and traffic classes are defined. Within the time slots, the device then allows certain traffic classes, which are stored in transmission queues, to be transmitted. This works with a *gate control list (GCL)*, which enables the different queues to transmit the queued frames within the time they are active (see left-hand side of Figure 2). This GCL

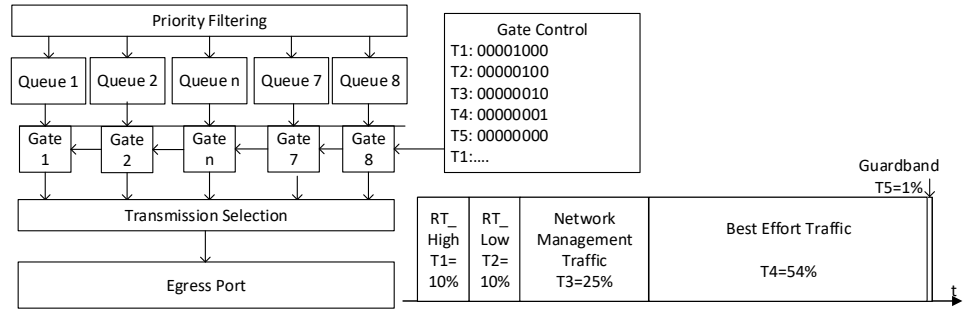


Fig. 2: Time-Aware Shaper

is then periodically repeated. In case multiple classes are transmitted within the same time slot, a transmission selection prioritises the frames further. [IE22]

2.2 Cybersecurity attacks in real-time Ethernet networks and their detection

Common attack patterns used in this paper include the replay attack, which can be combined with the *AitM*. Hereby the device sniffs packages, which are of importance e.g. for accessing confidential information or performing certain actions and replays given frames, thus accessing the information or redoing the requested action. [Ka]
Another attack pattern used is the *Adversary-in-the-middle (AitM)* attack, in which the attacker masquerades as the opposite sides of the communication. Therefore, the attacker has access to the communication between the two devices, enabling the attacker to modify or drop the frames. [Mia]
The third attack pattern is the *Denial-of-Service (DoS)*, in which the attacked device is e.g. directly flooded with a high amount of traffic, rendering the device incapable of responding to other messages, since the device is required to process all the messages. [Mib]
These attacks are used in related work to investigate the devices behaviour when under attack. Furthermore, the detection of such attacks can be done via multiple approaches. Due to the page limitation only the approaches relevant for industrial automation are described. Hereby the system utilises different technologies, such as anomaly detection or another pattern recognition, to compare the current state of the network with its idle but also potential attack states. When an oddity is detected, the *intrusion detection system (IDS)* device inserts a response, where the potential breach is fixed. [BS22a]

2.3 Related work on cybersecurity attacks on *TSN*-mechanisms

The external threat influence on *TSN*-devices has been a topic of many papers. Ergenç et al. present a selection of commonn attacks, which impact *TSN* mechanisms. Hereby they found more than 30 vulnerabilities within the entire structure of *TSN* standards mostly related to the tampering/modification of frames within the network or the impersonation of a key asset within the network [Er21]. Wang et al. utilised a model-based vulnerability mining method for the impact of cyber-attacks on gPTP as well as the TAS. Their results showed, that the mechanisms are easily manipulated via classic attacks such as *AitM*, replay attacks or *DoS*. [WLH22]
Furthermore, Fotouhi et al. investigated the impact of securing *gPTP* with security additions of the *PTP* protocol against spoofing. Their results show that the devices are protected against external spoofing, however, when the attacker compromised the network, the protection is non-existing. [Fo23]
Pena et al. investigated the impact on the end-to-end characteristics of implementing *MACsec* on the *TSN* traffic, to secure the devices against attacks. Their result shows, that *MACsec* implemented in hardware is a feasible solution to protect the devices against tempering.

However, implemented in software, the device's output decreases by 44%. [Pe22]
Whilst all the prior mentioned papers describe the results of attacks on the *TAS* or *gPTP*, none of the above investigate the relation, occurring between the attacks on *gPTP* and their impact on the *TAS*. The goal of this paper is to answer this research gap.

3 Methodology for the investigation

3.1 Measuring equipment and testbed

To investigate the impact of the described attacks on the synchronisation, a testbed was created (see figure 3).

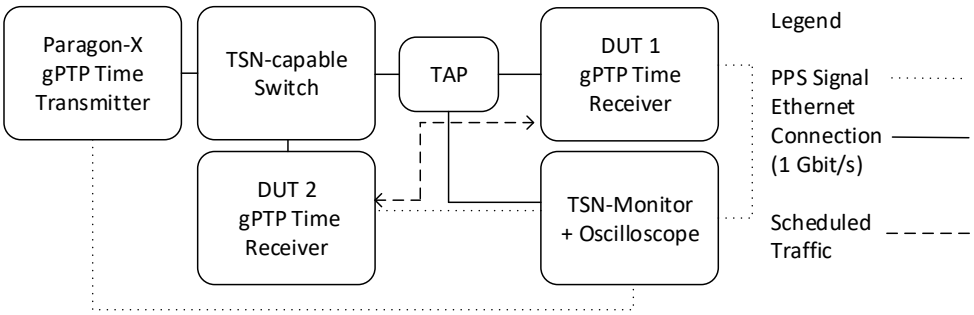


Fig. 3: Testbed for the impact investigation

Linux-based DUTs were implemented as Time Receiver devices, which also include a *Time-Aware Shaper* implementation. A specialised Synchronisation device called *Paragon-X* is used as the *Time Transmitter* [Ca18]. The last component, a *TSN*-capable switch, is not only used to connect the devices, but also as an insertion point of the attack pattern. To visualise the quality of the synchronisation, the devices *Pulse-per-Second (PPS)* Signals are connected to an oscilloscope and their deviation from another measured. On top of this, the network exchange between the devices is duplicated with a *TAP* and visualised via a combination of the special FPGA-based device, called *TSN-Monitor*, developed by Fraunhofer IOSB-INA, and the digital IOs of the oscilloscope. With this visualisation of the Ethernet traffic it is possible to measure each of the time slots. Hereby the *Paragon-X*'s feature to manipulate the transmission/ frame structure of the synchronisation is used. Therefore, the attacks are emulated as following:

- Replay attack: Previously sniffed sync and follow-up frames are inserted again into the network via a PC's network card and mirrored onto the *DUT1*'s Port.
- *AitM*: Since the *AitM* can manipulate the frame and its transmission, the attack is emulated via a constant offset on the timestamp.

- *DoS*: Force frame drops into the transmission of the sync and follow-up frames from the *Paragon-X*.

Furthermore, to create a schedule and stimuli frames, the devices within the network utilise the same *GCL*, influenced by the *PROFINET* scheduled communication cycle. Therefore, the cycletime of 1ms is separated into five phases, of which 10% are dedicated for traffic with a high real-time requirement, 10% are dedicated for traffic with a low real-time requirement, 25% are dedicated for network management, 54% for best effort and 1% is utilised as a guardband to prevent best effort frames from entering the time critical time slot of the next cycle (compare with right hand side of fig. 2). A software based network traffic generator is implemented in C, creating network traffic exchange to visualise the traffic between *DUT1* and *DUT2*. This network traffic consists of minimum sized *Ethernet* frames with a *VLAN tag* (68 byte).
Note that the resulting time slots visualised as the grayed out rectangle in the following figures are an overlay of multiple time slots, thus a deviation in the slots mean that the measured time slot may be bigger than the expected value due to jitter. Furthermore, the *PPS*-Signal of the *TT* is used as a relative starting point of the *TAS*.
Since the *TSN-Monitor* is using the same frequency as the *PHY* of the PCB it is implemented on, its measurement accuracy is around 8ns i.e. the cycle time for a 125MHz clock. The used oscilloscope has a maximum sample frequency of 20GSa/s. Furthermore, the *TAP* utilised in this setup has a delay of 425ns. These inaccuracies have to be considered when interpreting the results. To visualise the traffic of multiple *TAS* cycles the persistence feature is used, triggering on the *PPS*-Signal of the *Time Transmitter*. This is combined with a non synchronised frame transmission, leading to a standing picture of the *TAS*'s time slots (e.g. visible in figure 4).

4 Test results

4.1 Idle Network behavior

To compare the test results, first the devices idle situation is recorded, visible in figure 4. Hereby it is visible that the device separates the network traffic accordingly to the required time slots. The figure is a result of multiple *TAS* cycles being super positioned in a manner that the whole time slot is visualised. Furthermore, the time slots can be measured with the cursor on the oscilloscope. The results of that measurement are stated in table 1.

Tab. 1: Measured idle time slots of the TAS

Time slot	T1 [μs]	T2 [μs]	T3 [μs]	T4 [μs]	T5 [μs]
Expected time slot width	100,0	100,0	250,0	540,0	10,0
Measured time slot width	101,6	101,4	251,0	544,0	8,4

The differences between the values can be explained by transmitting frames exceeding the

time slots and manual measurement of the time slots as well as the delay inserted via the *TAP*. Furthermore, the post decimal accuracy is the result of the oscilloscopes resolution.

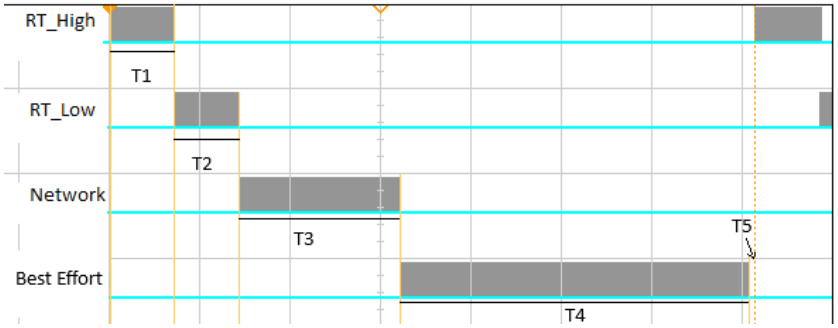


Fig. 4: Behaviour of the network without attacks, i.e. idle state

4.2 Influencing the integrity of the Protocol

The integrity of the frame has a high priority in the automation process, since a manipulated frame may lead to a faulty execution, potentially harming humans or damaging machines. Furthermore, it is possible to violate it via a multiple of attacks on the network. By influencing the integrity of the synchronisation, the *TR*-devices may not work as intended.

4.2.1 Emulating a replay attack on the gPTP Sync and follow-up frames

Prior to the attack, network traffic between the *DUT* and switch is captured via the *TAP* on a PC with Wireshark. This captured traffic includes the Synchronisation and follow-up frames. It is inserted into the switch, which mirrors the traffic onto the Switches Ethernet port connected with the *DUT*. Figure 5 shows, that the *DUT*'s transmission of the traffic, represented by the PCP7/RT_High time slot is not aligned with the *TT PPS*-Signal (Used as the trigger, visible with the orange line) and therefore, it drifts apart, visible in the greyed-out parts signal and the yellow signal. Here the yellow signals represent the time slot at the time of the screenshot and grey the overlay of multiple drifting time slots. This implies that, a

Fig. 5: TAS Time Slot RT_High after a replay attack

measuring of the time slots is not feasible over a longer period. Furthermore, this effect is the result of the device constantly adjusting between the replayed frames and the actual frames sent in the setup. Depending on the gPTP instances setting, the devices may halt the synchronisation process due to the time jumps occurring all the time. Also depending on the transmission, certain follow-up frames are omitted, due to the transmission alignment from the replay and original frames.

4.2.2 Emulating an Adversary-in-the-Middle attack on the Synchronisation

This attack requires the attacker to be set up between the *TT* and *TR*. Afterward the attacker manipulates the timestamp of the frames ”on the fly”, by inserting an additional offset of $100\mu s$ on the timestamp from the *TT*. This offset then leads to an offset in the *TR*'s time understanding, thus the *TAS* is shifted $100\mu s$ ahead (compare DUT time slots with ”Expected time slots” which represent the prior measured idle state in figure 6).

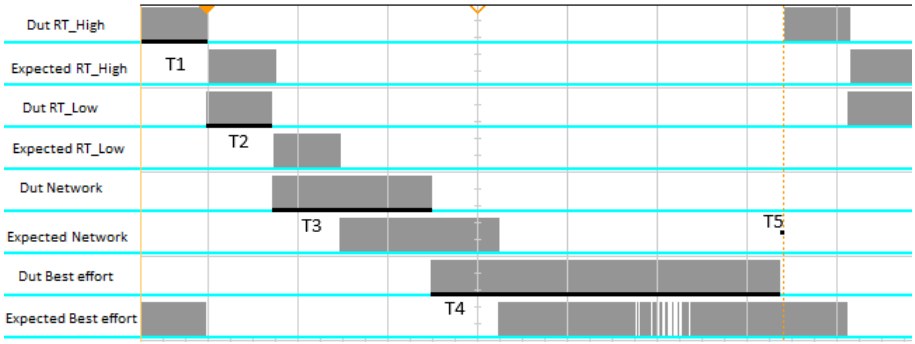


Fig. 6: TAS Time Slots with an asymmetrical delay (TS)

This offset detection has been the topic of other papers, including Ferfers et al. [FSJ23a]

4.3 Influencing the availability of the Synchronisation

The availability of the network participants has a high priority in the industrial automation [NI23]. Therefore, the detection of availability loss has to have a high priority. For this the influence on the availability via a *DoS*-attack has to be discussed.

4.3.1 Emulating a *DoS* on the synchronisation frames

The *DoS*-Attack is performed by dropping the synchronisation frames from the *Time Transmitter*, the *Paragon-X*. Hereby, the drop rate is increased almost to the point where the *gPTP* instance stops the synchronisation process, i.e. two out of three Frames are dropped.

The measured time slots can be seen in table 2.

Tab. 2: Measured *TAS* time slots after a partial *DoS*-attack

Time slot	T1 [μs]	T2 [μs]	T3 [μs]	T4 [μs]	T5 [μs]
Expected time slot width	102,9	101,6	250,6	541,4	8,4
Measured time slot width	101,6	101,4	251,0	544,0	8,4
Measured start of the time slot	1,5	101,6	202,0	450,0	993.2

This shows, that a (limited) *DoS* has no significantly measurable effect on the time slots. However, a full *DoS-attack* leads to an asynchronous frame-transmission of the attacked device, thus the time slots wander off in a similar manner to the replay attack making it impossible to measure.

4.4 Interpreting the results

It is visible, that the synchronisation errors directly translate to the devices *TAS* behaviour, since this disturb the time awareness of the *TAS*. Therefore, it is possible to derive potential errors on the synchronisation via the *TAS* behaviour. One example would be the shift of the devices time slots. This shift in the time slot means that the devices application may send out its data in a wrong time slot. e.g. when it is behind on the schedule, its data does not arrive in time on the receiving end, thus the real-time capability of the system is broken.

4.5 Potential detection of cyber-attacks by monitoring of scheduled traffic

Whilst the attacks on the synchronisation have an impact on the *TAS*, it is also important to draw a conclusion of the synchronisation quality via the observation of the *TAS*. One example could be the delay (positive or negative) of a transmission, which may be the result of a manipulation of the device's synchronisation. For this to work, a system has to be implemented, monitoring the behaviour of the devices within the network. However, the physical location of this is currently uncertain, since additional devices within the network may result in additional complexity, whilst integrating such monitoring solutions into the device itself is problematic, due to them being synchronised with the same *Time Transmitter* via the same *gPTP* protocol. Another solution would be a redundant synchronisation e.g. used in certain *NTP* clients to increase the reliability of the system [BS22b]. Furthermore, it is important to implement the *gPTP* instance in a manner, where threads have a low likelihood of occurring. Therefore, the following steps have to be considered:

- Prevention: preventing the attacking party to enter. e.g. via secure conduit architecture of the network(s). Another example would be the increasing resilience against manipulation, e.g. via hardware implemented *MAC* Security or authentication mechanisms.
- Detection: implementing an *IDS*, to detect potential manipulation of the devices synchronisation. e.g. via *honeypots* or anomaly detection.

- Elimination: elimination of the thread via dedicated methods. This can be done by (logical) isolation of the infected device or maintenance of the device, either automated or via trained personal. E.g. blocking of protocols, devices interfaces, addresses or modifying the devices address as described by Specht et. al. [SOE22].

5 Conclusion and future work

This paper utilised prior work, in which the synchronisation via *gPTP* was stated to be highly affected by cyber-attacks. Furthermore, the impact of these cyber-attacks on the *TAS* of the affected device was discussed. An example can be the manipulation of the timestamp, which directly shifts the device time according to the offset. But also denial of service attacks or the replay of a previously sniffed frame, where all of the prior stated attacks disturb the *TAS* functionality by desynchronising the device. Therefore to answer the research questions:

- RQ1: How do cyber-attacks influence the synchronisation?

As prior work established, synchronisation is influenced in many ways. This includes the complete stop of the synchronisation with a *DoS* and the direct manipulation of the frames e.g. from an adversary in the middle.

- RQ2: How does the influence on the synchronisation from RQ1 impact the *Time-Aware Shaper*-functionality of the devices?

The influence on the function of the *TAS* highly depends on the effect on the synchronisation. This is because the synchronisation errors produced by such attacks directly translate to the timing of the gate events, thus impacting the *TAS* general function by shifting with an inserted offset/timestamp manipulation or disabling the time awareness via a *DoS* of the *TAS*. Table 3 shows a summary of the attacks and their influences in the synchronisation as well as the *TAS*.

Tab. 3: Summary of the attacks impact on Synchronisation and *TAS*

Attack	**Influence Synchronisation**	**Influence *TAS***
DoS	Synchronisation halts	Loss of time awareness
AitM	Manipulation of Timestamps	Offset in the scheduled traffic
Replay Attack	Time jumps	Loss of time awareness

Future work focuses on the detection of common synchronisation attacks via the behaviour of the *TAS*. This should work as an intrusion detection system, focusing on the *TSN*-mechanisms. A way to implement this, could be a honeypot application, in which the device synchronises to the network via two independent *TTs*, comparing the times of the *TTs* and assess their offsets from each other. Therefore, the transmission timestamps can be monitored and a potential offset in them reported. The feasibility of such systems however

and potentially other concepts have to be evaluated in the future. Furthermore, such systems can be used complimentary to other systems, which detect root cause for *TSN* as proposed by [FSJ23b].

References

[BS22a] BSI: BSI-Leitfaden zur Einführung von Intrusion-Detection-Systemen, Guidline, Berlin, Germany: Bundesamt für die Sicherheit in der Informationstechnik (BSI), 2022, URL: https://www.bsi.bund.de/dok/6624202.

[BS22b] BSI: OPS.1.2.6 NTP-Zeitsynchronisation, OPS, Berlin, Germany: Bundesamt für die Sicherheit in der Informationstechnik (BSI), 2022, URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/IT-GS-Kompendium_Einzel_PDFs_2022/04_OPS_Betrieb/OPS_1_2_6_NTP_Zeitsynchronisation_Edition_2022.pdf?__blob=publicationFile&v=1.

[Ca18] Calnex: Paragon-X Getting Started Guide, Accessed: 17.08.2024, Calnex Solutions Ltd, 2018.

[Er21] Ergenç, D.; Brülhart, C.; Neumann, J.; Krüger, L.; Fischer, M.: On the Security of IEEE 802.1 Time-Sensitive Networking. In: 2021 IEEE International Conference on Communications Workshops (ICC Workshops). Pp. 1–6, 2021, DOI: 10.1109/ICCWorkshops50388.2021.9473542.

[Fo23] Fotouhi, M.; Buscemi, A.; Jomrich, F.; Koebel, C.; Engel, T.: Evaluation of PTP Security Controls on gPTP. In: 2023 IEEE Symposium on Computers and Communications (ISCC). Pp. 783–789, 2023, DOI: 10.1109/ISCC58397.2023.10218285.

[FSJ23a] Ferfers, T.; Schriegel, S.; Jasperneite, J.: Investigation in Automatic Fault Detection for Scheduled Traffic and Frame Preemption in Time-Sensitive Networks. In: Kommunikation in der Automation Beiträge des Jahreskolloquiums KommA 2023, Magdeburg. Pp. 145–154, 2023, DOI: 10.25673/111633.

[FSJ23b] Ferfers, T.; Schriegel, S.; Jasperneite, J.: Automated Root Cause Analysis in Time-Sensitive Networks based on Fault Models. In: International IEEE Symposium on Precision Clock Synchronization for Measurement, Control and Communication ISPCS 2023. London, United Kingdom, 2023.

[IE20] IEEE: IEEE Standard for Local and Metropolitan Area Networks–Timing and Synchronization for Time-Sensitive Applications. IEEE Std 802.1AS-2020 (Revision of IEEE Std 802.1AS-2011), pp. 1–421, 2020, DOI: 10.1109/IEEESTD.2020.9121845.

[IE22] IEEE: IEEE Standard for Local and Metropolitan Area Networks–Bridges and Bridged Networks. IEEE Std 802.1Q-2022 (Revision of IEEE Std 802.1Q-2018), pp. 1–2163, 2022, DOI: 10.1109/IEEESTD.2022.10004498.

[IE24] IEEE: IEEE Standard for Local and Metropolitan Area Networks–Timing and Synchronization for Time-Sensitive Applications Amendment 1: Inclusive Terminology. IEEE Std 802.1ASdr-2024 (Amendment to IEEE Std 802.1AS-2020 as amended by IEEE Std 802.1AS-2020/Cor 1-2021), pp. 1–222, 2024, DOI: 10.1109/IEEESTD.2024.10460486.

[Ka] Kaspersky: What is a Replay Attack?, https://www.kaspersky.com/resource-center/definitions/replay-attack, Accessed: 03.06.2024.

[Mia] Mitre: Adversary in the Middle, https://attack.mitre.org/techniques/T0830/, Accessed: 10.06.2024.

[Mib] Mitre: Network Denial of Service: Direct Network Flood, https://attack.mitre.org/techniques/T1498/001/, Accessed: 20.05.2024.

[NI23] NIST: Guide to Operational Technology (OT) Security, tech. rep., Washington, D.C.: U.S. Department of Commerce National Institute of Standards and Technology, 2023, DOI: https://doi.org/10.6028/NIST.SP.800-82r3.

[Pe22] Peña, R. A.; Pascual, M.; Astarloa, A.; Uribe, D.; Inchausti, J.: Impact of MACsec security on TSN traffic. In: 2022 37th Conference on Design of Circuits and Integrated Circuits (DCIS). Pp. 01–06, 2022, DOI: 10.1109/DCIS55711.2022.9970155.

[SOE22] Specht, F.; Otto, J.; Eickmeyer, J.: Cyberattack Impact Reduction using Software-Defined Networking for Cyber-Physical Production Systems. In: 2022 IEEE 20th International Conference on Industrial Informatics (INDIN). Pp. 188–194, 2022, DOI: 10.1109/INDIN51773.2022.9976140.

[WLH22] Wang, H.; Li, X.; Hu, X.: A Vulnerability Mining Method for IEEE bv in TSN System Based on Timed Automata. In: 2022 China Automation Congress (CAC). Pp. 6644–6649, 2022, DOI: 10.1109/CAC57257.2022.10056002.

[WSJ17] Wollschlaeger, M.; Sauter, T.; Jasperneite, J.: The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0. IEEE Industrial Electronics Magazine 11 (1), pp. 17–27, 2017, DOI: 10.1109/MIE.2017.2649104.

VirtuBench: Leistungsbewertung für virtuelle Industriesteuerungen

Felix Specht[1], Jens Otto[1], Henning Heutger[2] und Jens Frieben[2]

Abstract: Die Virtualisierung speicherprogrammierbarer Steuerungen ermöglicht eine flexiblere und effizientere Nutzung von Hardware-Ressourcen, eine schnellere Anpassung an veränderte Produktionsprozesse und eine vereinfachte Wartung und Durchführung von Softwareupdates. Gleichzeitig müssen virtuelle Steuerungen die hohen Anforderungen an Automatisierungskomponenten erfüllen, beispielsweise ein deterministisches Zeitverhalten. Durch gezielte Leistungsmessungen kann die Einhaltung dieser Anforderungen überprüft und sichergestellt werden. In diesem Zusammenhang wird die Softwarelösung VirtuBench zur Leistungsbewertung von virtuellen Industriesteuerungen vorgestellt. Der Schwerpunkt liegt auf der Bewertung relevanter Metriken wie Echtzeitverhalten, Latenzzeiten und Auslastung. Die innovative Lösung umfasst eine Auswahl an physikalischen sowie softwarebasierten Benchmark- und Monitoring-Werkzeugen, die eine umfassende Analyse ermöglichen.

1 Einleitung

Im Zuge der IT-OT-Konvergenz steigt der Bedarf an Flexibilität und Anpassungsfähigkeit von industriellen Produktionssystemen [ON15; OVN18]. Diesem Bedarf wird zunehmend durch den Einsatz von Virtualisierungstechnologien in der Automatisierung begegnet [GJF13; Qu23]. Die Virtualisierung speicherprogrammierbarer Steuerungen ermöglicht eine flexiblere und effizientere Nutzung von Hardwareressourcen, eine schnellere Anpassung an veränderte Produktionsprozesse sowie eine vereinfachte Wartung und Durchführung von Softwareupdates [Sc20]. In diesem Beitrag stellen wir die Softwarelösung VirtuBench (siehe Abbildung 1) zur Leistungsbewertung virtueller Industriesteuerungen vor. VirtuBench erfasst und analysiert Metriken in den Kategorien Host, Netzwerk und Input-Output an verteilten Messpunkten in unterschiedlichen Testszenarien. Dazu wird ein Industrie-PC als Host für die virtuellen Steuerungen, ein industriellen Netzwerkswitch, drei IO-Geräte, ein Oszilloskop sowie ein Test-PC verwendet. Die Kommunikation zwischen den virtuellen Steuerungen und den IO-Geräten erfolgt über Profinet. Die virtuellen Steuerungen können sowohl als virtuelle Maschinen als auch als Container bereitgestellt werden. Der Test-PC ermöglicht die Steuerung der Testumgebung und die Erfassung diverser Leistungsmetriken während der Ausführung von Testszenarien. Die nachfolgenden Kapitel sind wie folgt strukturiert: Kapitel 2 gibt einen Überblick über verwandte Arbeiten. Kapitel 3 beschreibt die Softwarelösung VirtuBench, Metriken zur Leistungsbewertung, den Messpunkten im Versuchsaufbau, den Testszenarien und dem Ablauf eines Messvorgangs. Kapitel 4 zeigt die

1 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Germany, felix.specht@iosb-ina.fraunhofer.de; jens.otto@iosb-ina.fraunhofer.de

2 Phoenix Contact GmbH, Dringenauer Straße 30, Bad Pyrmont, Germany, hheutger@phoenixcontact.com; jfrieben@phoenixcontact.com

Ergebnisse der Leistungsbewertung anhand eines Beispiels. Kapitel 5 fasst die Ergebnisse zusammen und gibt einen Ausblick auf zukünftige Arbeiten.

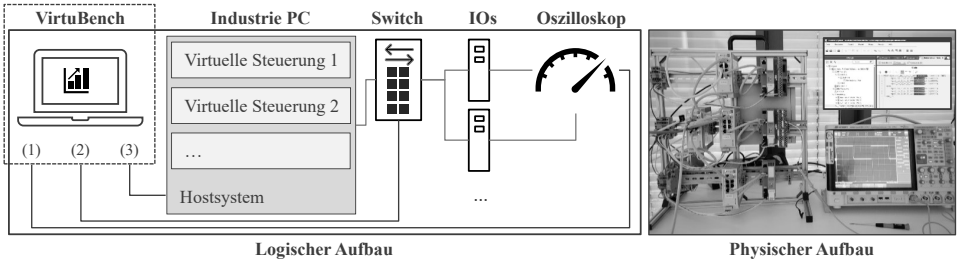


Abb. 1: VirtuBench: Leistungsbewertung für virtuelle Industriesteuerungen.

2 Stand der Technik

Die Virtualisierung industrieller Steuerungen ist dabei ein zentraler Aspekt und kann sowohl in Form einer virtuellen Maschine (VM) als auch in Form von virtuellen Containern umgesetzt werden [Se14]. VM-Virtualisierung beinhaltet das Ausführen eines Hypervisors, der die Ressourcen der Host-Hardware über mehrere VMs aufteilt und verwaltet. Jede VM führt ihr eigenes Betriebssystem und Anwendungen aus [Sh16]. Im Gegensatz dazu verwendet die Container-Virtualisierung Ressourcen des Betriebssystems, beispielsweise den Kernel. Dabei werden Betriebssystemprozesse und ihre Abhängigkeiten als Container gekapselt, die vom zugrunde liegenden Host-Betriebssystemkern gemeinsam verwaltet werden. In einem Überblick untersuchen die Autoren verschiedene Echtzeit-Container-Plattformen. Hauptaspekte der Untersuchung waren die Linux-Echtzeitunterstützung sowie die Herausforderungen in der Echtzeitvirtualisierung in Kombination mit Sicherheitsproblemen und fehlenden Latenztests [St20]. Der Ansatz von Goldschmidt und Hauck-Stattelmann basiert auf Software-Containern für industrielle Steuerungen und bewertet die Echtzeitfähigkeit von Docker mit Cyclictest [GH16]. In einem Beitrag wird eine Testumgebung zur Qualitätsbewertung verschiedener Industrial-Ethernet-Protokolle vorgestellt [STV18] und zeitkritische, industrielle Anwendungen in Docker-Containern evaluiert [SLV19]. Die Qualitätsbewertung basiert auf Latenzzeit, Jitter, Durchsatz sowie dem Auftreten von Fehlern oder dem Verlust von Paketen im Netzwerk. Tasci et al. stellen eine containerbasierte Architektur für Echtzeit-Steuerungsanwendungen vor, welche Interprozesskommunikation zwischen Containern per UDP mit der ZeroMQ-Bibliothek realisiert [TMV18]. Die Bewertung erfolgt anhand von Latenzmessungen der Zykluszeit sowie Paketumlaufzeit. Sisinni et al. untersuchen die Paketumlaufzeit von industriellen Edge-Anwendungen mit MQTT-Kommunikation, die als Container-Virtualisierung auf einem Siemens Industrial Edge Gerät ausgeführt werden [Si23]. Die Autoren messen eine Paketumlaufzeit von 10 ms und argumentieren, dass dies für die Realisierung von Prozessleitsystemen und Überwachungsanwendungen ausreichend ist. In einem Artikel werden die Auswirkungen der Virtualisierung auf die Zeitverzögerungen

einer Softwarekomponente basierend auf Docker-Containern untersucht, indem Messungen auf einem Hardware-Testbed in einem realistischen Anwendungsfall bereitgestellt werden. Die Experimente zeigen, dass Docker-Virtualisierung die Soft-Real-Time-Anforderungen erfüllen und in der industriellen Automatisierung eingesetzt werden kann [Qu23]. Ein weiterer Artikel bewertet die Auswirkungen der Docker-Containerisierung auf weiche Echtzeitanwendungen anhand von Netzwerkmetriken, wie die Paketumlaufzeit [So20]. Die Ergebnisse der Autoren zeigen, dass die Echtzeitanforderungen an Automatisierungsanwendungen erfüllt werden, jedoch ist mit einer zusätzlichen Verarbeitungsverzögerung in einer Größenordnung von 150 µs zu rechnen. In einem Beitrag stellen die Autoren eine Methodik zur Leistungsevaluierung von containerbasierten Infrastrukturen vor. Die Autoren messen die Paketumlaufzeit bei unterschiedlichen Netzwerkpaketgrößen sowie CPU- und Speicherauslastung bei Anwendungen des maschinellen Lernens. Die Ergebnisse zeigen, dass die Containerisierung keine spürbare Leistungsverschlechterung in Bezug auf Kommunikations- und Rechenfähigkeiten verursacht [Li21]. Gaffurini et al. vergleichen die Fähigkeit zur Echtzeitkommunikation zwischen einer virtuellen Steuerung (Siemens vPLC CPU1582v) und einer klassischen Steuerung (Siemens S7-1500) anhand der Paketumlaufzeit und den Übertragungslatenzen [Ga24].

3 Lösungsansatz

In diesem Kapitel wird das Lösungskonzept von VirtuBench vorgestellt, welches aus den folgenden Punkten besteht: (1) Metriken zur Leistungsbewertung von virtuellen Steuerungen, (2) das Konzept des Versuchsaufbaus mit Messpunkten zur Erfassung der Metriken und (3) das Konzept zur Durchführung von Testszenarien und zur Auswertung der Metriken zur Leistungsbewertung von virtuellen Steuerungen.

3.1 Metriken zur Leistungsbewertung von virtuellen Steuerungen

Metriken sind ein zentraler Bestandteil des Lösungsansatzes. Sie sind Indikatoren, um die Leistungsfähigkeit und Effizienz von virtuellen Steuerungen zu bewerten. Tabelle 1 zeigt eine Übersicht der unterschiedlichen Metriken, die in drei Hauptkategorien unterteilt sind: (1) Host, (2) Netzwerk und (3) Input-Output.

In der Kategorie Host werden vier Metriken betrachtet: Die CPU-Auslastung m_1 misst den Prozentsatz der CPU-Auslastung und gibt an, wie stark der Prozessor ausgelastet ist. Die Auslastung des Arbeitsspeichers m_2 misst die Menge des verwendeten Arbeitsspeichers in Gigabyte. Die Datenträgerauslastung m_3 bestimmt die kombinierten Lese- und Schreibzugriffe auf den genutzten Datenträger in Byte pro Sekunde. Die Datenträgerlatenz m_4 gibt die Verzögerung beim Zugriff auf den Datenträger in Mikrosekunden an. In der Kategorie Netzwerk werden vier Metriken betrachtet: Die Paketumlaufzeit m_5 misst die Zeit für eine Hin- und Rückübertragung eines Signals in Millisekunden. Die Metrik Netzwerkpakete

pro Sekunde m_6 misst die Anzahl der übertragenen Netzwerkpakete pro Sekunde. Die Zwischenankunftszeit m_7 zeigt die durchschnittliche Zeit zwischen aufeinanderfolgenden Netzwerkpaketen in Millisekunden an. Die Metrik Jitter der Zwischenankunftszeit m_8 misst die Variabilität der Zeit zwischen aufeinanderfolgenden Netzwerkpaketen in Mikrosekunden. In der Kategorie Input-Output werden vier Metriken betrachtet: Die IEC 61131-Zykluszeit m_9 misst die Zykluszeit eines Takts der IEC 61131 Laufzeitumgebung in Millisekunden. Die IEC 61131-Verzögerung m_{10} Diese Metrik gibt die Verzögerung zum letzten Takt der IEC 61131 Laufzeitumgebung in Millisekunden an. Die IO-Zykluszeit m_{11} gibt die Zeit für einen kompletten Input/Output-Zyklus in Millisekunden an. Der Jitter der IO-Zykluszeit m_{12} misst die Variabilität der IO-Zykluszeit in Millisekunden.

Tab. 1: Metriken zur Leistungsbewertung von virtuellen Steuerungen

Kategorie	Metrik	Beschreibung	Einheit
Host	m_1	CPU-Auslastung	%
	m_2	Arbeitsspeicherauslastung	GB
	m_3	Datenträgerauslastung	B/s
	m_4	Datenträgerlatenz	µs
Netzwerk	m_5	Paketumlaufzeit	ms
	m_6	Netzwerkpakete pro Sekunde	-
	m_7	Zwischenankunftszeit	ms
	m_8	Jitter der Zwischenankunftszeit	µs
Input-Output	m_9	IEC 61131-Zykluszeit	ms
	m_{10}	IEC 61131-Verzögerung	ms
	m_{11}	IO-Zykluszeit	ms
	m_{12}	Jitter der IO-Zykluszeit	ms

3.2 Messpunkte im Aufbau

Diese Messpunkte sind essenziell für die Erfassung und Analyse der unterschiedlichen Metriken. Jeder Messpunkt ist mit spezifischen Metriken verknüpft und wird nachfolgend beschrieben: Der Host ist ein IPC mit einem Betriebssystem bzw. Hypervisor, welcher die Virtualisierung in Form von virtuellen Maschinen oder Containern ermöglicht. An diesem Messpunkt werden die Metriken m_1 CPU-Auslastung, m_2 Arbeitsspeicherauslastung, m_3 Datenträgerauslastung und m_4 Datenträgerlatenz erfasst. Der Test-PC ist ein Computer, der zur Durchführung spezifischer Tests und zur Validierung der Ergebnisse eingesetzt wird. Dieser Messpunkt ermöglicht es, die Testbedingungen zu kontrollieren und die Auswirkungen von Änderungen in der Systemumgebung zu beobachten. Des Weiteren wird die Leistungsmetrik m_5 Paketumlaufzeit erfasst. Der Switch ist das zentrale Netzwerkgerät, welches den Datenverkehr zwischen virtueller Steuerung und IO-Gerät weiterleitet. Die Messungen der Metriken an diesem Punkt sind kritisch für die Analyse der Netzwerkperformance und zur Identifikation von Engpässen in der Datenübertragung. Die Metriken m_6 Netzwerkpakete pro Sekunde, m_7 Zwischenankunftszeit und m_8 Jitter der Zwischenankunftszeit werden an

diesem Messpunkt erfasst. Virtuelle speicherprogrammierbare Steuerungen sind in Form von VMs bzw. Containern auf dem Host realisiert und führen ein IEC 61131-3 Steuerungsprogramm aus. An diesem Messpunkt werden die Metriken m_9 IEC 61131-Zykluszeit und m_{10} IEC 61131-Verzögerung erfasst. Sie ermöglichen die Untersuchung der Effizienz und Reaktionszeit von industriellen Steuerungs- und Regelungsaufgaben. Das Oszilloskop ist mit den Ausgängen eines IO-Geräts in der Testumgebung verbunden und erfasst die elektrischen Signale. Dieser Messpunkt ist entscheidend für die genaue Analyse der Signalqualität und der zeitlichen Charakteristik von Signalen, die zwischen einer virtuellen Steuerung und einem IO-Gerät ausgetauscht werden. Die Metriken m_{11} IO-Zykluszeit und m_{12} Jitter der IO-Zykluszeit werden an diesem Messpunkt erfasst.

3.3 Durchführung und Auswertung von Testszenarien

VirtuBench unterstützt die Ausführung von verschiedenen Testszenarien, jeweils unter Verwendung von VM-basierter oder containerbasierter Virtualisierung. Die Testszenarien unterscheiden sich hinsichtlich der Anzahl von virtuellen Steuerungen, der Anzahl von IO-Geräten und den Zykluszeiten in den IEC 61131-Steuerungsprogrammen. Jedes Testszenario wird durch eine eindeutige ID identifiziert, welche sich aus den Eigenschaften des Testszenarios zusammensetzt. Beispielsweise steht die ID 1104 für ein Testszenario mit einer virtuellen Steuerung, einem IO-Gerät und einer Zykluszeit von 4 ms. Die ID 2216 steht für ein Testszenario mit zwei virtuellen Steuerungen, zwei IO-Geräten und einer Zykluszeit von 16 ms.

Zu Beginn eines Messvorgangs wird ein definierter Ausgangszustand hergestellt. Anschließend erfolgt der Start der VirtuBench-Software zur zentralen Handhabung der Testumgebung, der Konfiguration, dem Start und Stopp der Tests sowie dem Sammeln und Speichern der Messdaten. Der Ablauf des Messvorgangs umfasst folgende Schritte: (1) Konfiguration und Start der IEC 61131-Projekte in der virtuellen Steuerung, (2) Start der Software zur Erfassung der Metriken, (2.1) Aufbau einer SSH-Verbindung zum Host, (2.2) Ausführung der Paketumlaufzeit-Messung am Test-PC, (2.3) Start einer Netzwerkaufzeichnung am Spiegel-Port des Switches, (2.4) Aufbau einer Engineering-Verbindung zur virtuellen Steuerung, (2.5) Start der Messungen am Oszilloskop über eine Netzwerkschnittstelle, (3) Aufzeichnung der Messdaten für ein konfiguriertes Zeitfenster, (4) Abschluss der Aufzeichnung und zentrale Speicherung der Messdaten, (5) Auswertung der Messdaten und Analyse der Metriken. Durch diesen strukturierten Ablauf wird sichergestellt, dass alle relevanten Messdaten erfasst und ausgewertet werden können. Ein Messvorgang wird mehrfach ausgeführt, um die Stabilität und Reproduzierbarkeit der Ergebnisse zu gewährleisten. Während eines Messvorgangs werden die Daten über einen festgelegten Zeitraum erfasst. Anschließend wird für jedes Testszenario der arithmetische Mittelwert der einzelnen Metriken über alle Messvorgänge berechnet. Die Ergebnisse werden schließlich in Tabellen und Diagrammen dargestellt und dem Benutzer der Testumgebung zur Verfügung gestellt.

4 Ergebnisse

Die Ergebnisse wurden mit folgender Konfiguration erstellt: Die VM-Virtualisierung auf dem IPC erfolgt mit der Hypervisor-Lösung ESXi in Version 7.0 U3. Die Container-Virtualisierung wurde durch Docker in Kombination mit Debian Linux 12 Stable RT-Kernel realisiert. Als IPC wird ein Produkt von Phoenix Contact eingesetzt, welches einen Intel Core i5 Prozessor mit 4 Kernen, 64 GB Arbeitsspeicher und zwei SSD-Festplatten. Die virtuelle Steuerung ist durch die Software Virtual PLCnext Control von Phoenix Contact realisiert, welche als VM bzw. als Container zur Verfügung steht. Die Virtual PLCnext Control Software sieht vor, dass ein Prozessorkern für den Hypervisor bzw. das Betriebssystem reserviert wird. Ebenfalls wird jeweils ein Prozessorkern für eine Steuerungsinstanz reserviert, wodurch die Anzahl der Steuerungsinstanzen auf maximal 3 begrenzt ist. Abhängig vom jeweiligen Testszenario wird der IPC mit der einen oder anderen Festplatte gestartet, welche jeweils die VM- bzw. Container-Virtualisierung beherbergen. Die VirtuBench-Software ist in der Programmiersprache Python implementiert. Für die nachfolgenden Ergebnisse wurde jeder Messvorgang 10-mal wiederholt, wobei die Dauer jedes Messvorgangs 60 Sekunden betrug. Die Messergebnisse sind in Tabelle 2 zusammengefasst. Die Tabelle verwendet die in Kapitel 3 beschriebenen Metriken und Testszenarien. Für jedes Testszenario repräsentiert als ID sind Metriken jeweils für Container (CT) und virtuelle Maschinen (VM) erfasst.

Tab. 2: Ergebnisse: m_1 : CPU-Auslastung, m_2 : Arbeitsspeicherauslastung, m_3 : Datenträgerauslastung, m_4 : Datenträgerlatenz, m_5 : Paketumlaufzeit, m_6 : Netzwerkpakete pro Sekunde, m_7 : Zwischenankunftszeit, m_8 : Jitter der Zwischenankunftszeit, m_9 : IEC 61131-Zykluszeit, m_{10} : IEC 61131-Verzögerung, m_{11} : IO-Zykluszeit, m_{12} : Jitter der IO-Zykluszeit.

ID	**Typ**	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}
1104	CT	0.08	0.67	14.26	76.61	2.07	1008.82	2.0	111.23	3.53	4.78	4.04	0.15
	VM	2.52	2.09	7.17	82.43	5.36	1005.33	2.0	119.87	7.43	18.10	4.04	0.15
1108	CT	0.07	0.67	12.26	67.36	0.97	1007.84	2.0	84.54	4.00	5.31	8.04	0.18
	VM	2.45	2.10	7.80	80.90	5.81	1054.06	2.0	127.64	7.87	18.38	8.04	0.18
1116	CT	0.09	0.67	10.10	60.31	0.91	1007.97	2.0	81.57	3.94	5.21	16.04	0.07
	VM	2.44	2.11	7.30	81.77	6.43	1053.07	2.0	138.53	8.12	18.90	16.04	0.08
2204	CT	0.52	0.86	11.00	59.76	0.92	2071.51	2.0	88.66	5.23	6.96	4.05	0.16
	VM	5.12	2.66	11.13	87.03	6.61	2056.04	2.0	92.01	7.40	18.68	4.04	0.15
2208	CT	0.13	0.86	10.15	59.40	0.99	2007.43	2.0	96.86	5.64	6.36	8.04	0.18
	VM	4.74	2.67	11.90	89.77	7.03	2009.02	2.0	86.63	7.90	18.84	8.04	0.18
2216	CT	0.14	0.86	10.00	58.68	5.80	2013.10	2.0	155.92	5.70	6.23	16.04	0.07
	VM	4.80	2.68	10.73	87.93	6.05	2017.12	2.0	103.54	8.23	19.19	16.04	0.06
3304	CT	0.41	1.05	20.00	57.39	5.97	3065.52	2.0	107.06	7.70	9.04	4.02	0.14
	VM	8.28	3.24	14.93	89.93	7.51	3065.36	2.0	89.78	7.52	17.81	4.05	0.16
3308	CT	0.60	1.05	20.00	56.58	6.62	3069.62	2.0	108.56	7.76	8.04	8.04	0.18
	VM	7.70	3.27	13.90	91.20	6.81	3015.39	2.0	86.52	8.23	17.95	8.04	0.18
3316	CT	0.60	1.05	20.00	55.49	5.99	3068.76	2.0	95.61	7.52	17.56	16.04	0.08
	VM	7.69	3.27	16.57	88.30	6.11	3012.65	2.0	92.38	8.19	17.83	16.04	0.06
Einheit		%	GB	B/s	µs	ms	-	ms	µs	ms	ms	ms	ms

Bei der CPU-Auslastung m_1 ist ein deutlicher Unterschied zwischen den Testszenarien mit jeweils einer, zwei und drei virtuellen Steuerungen zu beobachten. Die CPU-Auslastung steigt mit der Anzahl der Instanzen von virtuellen Steuerungen. Erwartungsgemäß ist die CPU-Auslastung bei der containerbasierten Virtualisierung im Vergleich zur VM-basierten Virtualisierung geringer. Während VM-basierte Testszenarien bis zu 8% CPU-Auslastung erreichen, liegt die CPU-Auslastung bei containerbasierten Testszenarien bei unter 1%. Ein ähnliches Bild zeigt sich bei der Leistungsmetrik m_2 Arbeitsspeicherauslastung, bei der die Auslastung mit der zunehmenden Anzahl von Instanzen steigt. Ebenfalls ist die Auslastung bei der containerbasierten Virtualisierung im Vergleich zur VM-basierten Virtualisierung geringer. VM-basierte Testszenarien belegen zwischen 2,1 und 3,3 GB Arbeitsspeicher, während containerbasierte Testszenarien zwischen 600 MB und 1,1 GB belegen. Die Datenträgerauslastung m_3 zeigt eine leicht höhere Auslastung bei den containerbasierten Testszenarien, jedoch ist die Auslastung bei beiden Virtualisierungsarten gering. Die Datenträgerlatenz m_4 ist bei der containerbasierten Virtualisierung im Vergleich zur VM-basierten Virtualisierung geringer. Jedoch sind hier keine nennenswerten Unterschiede ersichtlich.

Die Leistungsmetrik m_5 Paketumlaufzeit zeigt für alle VM-basierten Testszenarien ähnliche Messergebnisse zwischen 5 ms und 7,5 ms. Die containerbasierten Testszenarien zeigen dagegen z.T. deutliche Abweichungen im Vergleich untereinander. Während die Testszenarien mit drei Instanzen ähnliche Latenzen zeigen wie die VM-Szenarien, sind bei einer und zwei Instanzen zum Teil deutlich geringere Latenzen von 1-2 ms zu beobachten. Bei der Anzahl der Netzwerkpakete pro Sekunde m_6 ist zwischen VM-basierter und containerbasierter Virtualisierung kein Unterschied festzustellen. Erwartungsgemäß steigt die Anzahl der Netzwerkpakete gleichmäßig mit der Anzahl der Instanzen. Die Zwischenankunftszeit m_7 und der Jitter der Zwischenankunftszeit m_8 sind bei beiden Virtualisierungsarten nahezu identisch. Die Leistungsmetrik m_7 zeigt für alle Testszenarien eine identische Inter Arrival Time von 2 ms, entsprechend der Konfiguration mit 500 Paketen pro Sekunde. Bei m_8 sind zwar leichte Unterschiede zu erkennen, jedoch in einer sehr geringen Größenordnung von 84 bis 156 µs.

Die Metriken m_9 IEC 61131 Zykluszeit und m_{10} Verzögerung im letzten Takt zeigen für alle VM-basierten Testszenarien ähnliche Ergebnisse mit m_9 im Bereich von 8 µs und m_{10} im Bereich von 17,5 µs. Dagegen ist bei den containerbasierten Testszenarien eine Abhängigkeit von der Anzahl der Instanzen zu erkennen. Mit steigender Anzahl steigt ebenfalls die Zykluszeit und die Verzögerung. Jedoch sind die Werte in einer sehr geringen Größenordnung von 5 bis 17,5 µs. Die beiden letzten Metriken m_{11} IO-Zykluszeit und m_{12} Jitter der IO-Zykluszeit wurden mit einem Oszilloskop am Feldbus-IO-Gerät gemessen. Die Ergebnisse der IO-Zykluszeit zeigen eine sehr gute Übereinstimmung mit den durch die Testszenarien definierten Zykluszeiten von 4 ms, 8 ms und 16 ms. Der Jitter der IO-Zykluszeit zeigt geringe Schwankungen von 0,075 bis 0,175 ms.

5 Zusammenfassung und Ausblick

Dieser Beitrag stellt VirtuBench vor, eine Softwarelösung zur Leistungsbewertung virtueller Industriesteuerungen. VirtuBench erfasst wichtige Leistungsmetriken wie Echtzeitverhalten, Latenzzeiten und Auslastung. Durch den Einsatz einer Kombination aus physikalischen und softwarebasierten Benchmark- und Monitoring-Tools ermöglicht VirtuBench die Erfassung und Auswertung von Metriken an verteilten Messpunkten.

Die Lösung bietet folgende Vorteile: (1) Umfassende Analyse der Leistung virtueller Steuerungen und (2) Unterstützung bei der Einhaltung von Anforderungen an Automatisierungskomponenten. Erste Ergebnisse, die mit einer virtuellen Industriesteuerung von Phoenix Contact erzielt wurden, zeigen, dass die Testumgebung effektiv Leistungsmetriken virtueller Steuerungen erfassen und validieren kann.

In zukünftigen Arbeiten soll die Testumgebung um die Leistungsbewertung der Security erweitert werden. Ziel ist es, automatisiert ein Security-Lagebild zu erstellen, das unter anderem Schwachstellen in Softwarekomponenten, Update-Status und potenziell unsichere Konfigurationen umfasst. Die automatisierte Leistungsbewertung der Security für Komponenten gewinnt im Rahmen des EU Cyber Resilience Acts zunehmend an Bedeutung.

Literaturverzeichnis

[Ga24] Gaffurini, M.; Bellagente, P.; Depari, A.; Flammini, A.; Sisinni, E.; Ferrari, P.: Virtual PLC in Industrial Edge Platform: Performance Evaluation of Supervision and Control Communication. IEEE Transactions on Instrumentation and Measurement 73, S. 1–10, 2024.

[GH16] Goldschmidt, T.; Hauck-Stattelmann, S.: Software containers for industrial control. In: 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, 2016.

[GJF13] Gaj, P.; Jasperneite, J.; Felser, M.: Computer Communication Within Industrial Distributed Environment—a Survey. IEEE Transactions on Industrial Informatics 9 (1), S. 182–189, 2013.

[Li21] Liu, Y.; Lan, D.; Pang, Z.; Karlsson, M.; Gong, S.: Performance Evaluation of Containerization in Edge-Cloud Computing Stacks for Industrial Applications: A Client Perspective. IEEE Open Journal of the Industrial Electronics Society 2, S. 153–168, 2021.

[ON15] Otto, J.; Niggemann, O.: Automatic Parameterization of Automation Software for Plug-and-Produce. In: AAAI-15 Workshop on Algorithm Configuration (AlgoConf). Austin, USA, 2015.

[OVN18] Otto, J.; Vogel-Heuser, B.; Niggemann, O.: Automatic Parameter Estimation for Reusable Software Components of Modular and Reconfigurable Cyber-Physical Production Systems in the Domain of Discrete Manufacturing. IEEE Transactions on Industrial Informatics 14 (1), S. 275–282, 2018.

[Qu23] Queiroz, R.; Cruz, T.; Mendes, J.; Sousa, P.; Simões, P.: Container-based virtualization for real-time industrial systems—a systematic review. ACM Computing Surveys 56 (3), S. 1–38, 2023.

[Sc20] Scordino, C.; Savino, I. M.; Cuomo, L.; Miccio, L.; Tagliavini, A.; Bertogna, M.; Solieri, M.: Real-Time Virtualization For Industrial Automation. In: 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). 2020.

[Se14] Seo, K.-T.; Hwang, H.-S.; Moon, I.-Y.; Kwon, O.-Y.; Kim, B.-J.: Performance comparison analysis of linux container and virtual machine for building cloud. Advanced Science and Technology Letters 66 (105-111), S. 2, 2014.

[Sh16] Sharma, P.; Chaufournier, L.; Shenoy, P.; Tay, Y.: Containers and virtual machines at scale: A comparative study. In: Proceedings of the 17th international middleware conference. 2016.

[Si23] Sisinni, E.; Bellagente, P.; Depari, A.; Flammini, A.; Gaffurini, M.; Pasetti, M.; Rinaldi, S.; Ferrari, P.: Assessment of time performance of lightweight virtualization for edge computing applications. In: 2023 IEEE 19th International Conference on Factory Communication Systems (WFCS). IEEE, S. 1–4, 2023.

[SLV19] Sollfrank, M.; Loch, F.; Vogel-Heuser, B.: Exploring docker containers for time-sensitive applications in networked control systems. In: 2019 IEEE 17th International Conference on Industrial Informatics (INDIN). IEEE, 2019.

[So20] Sollfrank, M.; Loch, F.; Denteneer, S.; Vogel-Heuser, B.: Evaluating docker for lightweight virtualization of distributed and time-sensitive applications in industrial automation. IEEE Transactions on Industrial Informatics 17 (5), S. 3566–3576, 2020.

[St20] Struhár, V.; Behnam, M.; Ashjaei, M.; Papadopoulos, A. V.: Real-time containers: A survey. In: 2nd Workshop on Fog Computing and the IoT (Fog-IoT 2020). Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[STV18] Sollfrank, M.; Trunzer, E.; Vogel-Heuser, B.: A testbed for evaluating QoS of different classes of industrial Ethernet protocols based on raspberry Pi. In: 44th Annual Conference of the IEEE Industrial Electronics Society (IECON). IEEE, 2018.

[TMV18] Tasci, T.; Melcher, J.; Verl, A.: A container-based architecture for real-time control applications. In: IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). IEEE, 2018.

5G Performance Analysis for Material Handling in Manufacturing.

Gustavo Cainelli[1], Meik Kottkamp[2], Michelangelo Albanese[3], Nils Kranefeld[4], Lisa Underberg[1], and Cole Saunders[1]

Abstract: 5G continues to attract interest, particularly in the context of industrial applications using commercially available devices. The benefits of wireless communication in Industry 4.0 scenarios are well-known, but challenges remain in optimizing network parameters for specific use cases. Performance testing is crucial to understand the capabilities of 5G across each release. This paper presents performance testing across different test cases and devices to demonstrate the development and impact of 5G technology in a factory hall with intralogistics applications. Collaborations in this area aim to enhance understanding and application of 5G in industrial settings.

Keywords: 5G, intralogistics, performance evaluation.

1 Introduction

5G technology is continuously gaining attention, particularly in real-world industrial applications that utilize commercially available devices. As of 2024, the benefits of wireless communication in Industry 4.0 are well-known, including its potential to enhance flexibility, efficiency, and scalability in various industrial processes.

In the context of intralogistics, which involves the management and optimization of internal logistics processes within manufacturing environments, the role of wireless communication becomes even more critical. Effective material handling relies on the seamless integration of communication technologies with machinery and control systems, making it an ideal application area to explore the potential of 5G.

This paper aims to demonstrate the performance development of 5G technology and its impact on intralogistics applications, specifically within the DEMAG Research Factory in Wetter, Germany. The primary objective is to investigate how 5G can be optimized to meet the specific needs of material handling systems. By exploring the practical implications of 5G deployment in a real-world environment, this study provides valuable insights that can inform future implementations of wireless communication systems in manufacturing. Moreover, the findings from this research could serve as a benchmark for other industries looking to implement 5G in similar applications, thus contributing to the broader field of

1 Institut für Automation und Kommunikation e.V. , ICT & Automation, Werner-Heisenberg-Straße 1, 39106, Magdeburg, Germany, gustavo.cainelli@ifak.de; lisa.underberg@ifak.eu; cole.saunders@ifak.de
2 Rohde & Schwarz, meik.kottkamp@rohde-schwarz.com
3 Nokia, michelangelo.albanese@nokia.com
4 Demag Cranes and Components Gmbh, Department, Wetter, Germany, nils.kranefeld@demagcranes.com

industrial automation and communication.The performance testing presented in this paper were conducted in an 5G-ACIA endorsed testbed [5G23].

The paper is structured as follows: Section 2 presents some works that investigate the performance of 5G systems. Section 3 provides a description of the DEMAG Research Factory and the use case scenarios under investigation. Following that, the system under test and the test system are described in Section 4, outlining the specific 5G network configurations and measurement tools used. The measurement results are presented in Section 5. Finally, the paper concludes with a discussion of the key findings and their implications for future industrial applications of 5G in Section 6.

2 Related works

5G deployments ranged from non-standalone (NSA) to standalone (SA) solutions and included frequency bands from sub-6 GHz to millimeter waves. The studies measured a variety of parameters, including signal strength and other application-critical factors such as transmission time, response time, and packet error rates [Sa24].

Studies have been conducted on the network coverage and characteristics of 5G Non-Public-Networks (NPN) in both indoor and outdoor environments. These studies have been conducted in environments such as factories or warehouses [An22; Ca22; Da23; Ha23; Ho23; Ke22; Ko22; La22; Ly23], offices or laboratories [Ma22; Mo23; Ri21], and urban or outdoor settings [Ca22; Dí20; Fj22; Ma22].

Most measurements were performed using stationary user equipment (UEs) or measurement devices [Sa24]. The study in [Sa24] presented results from measurements in a process automation environment, where the network was tested both mobile and statically. Measurements of relevant parameters such as transmission time, update time, and packet loss rate were performed statically.

Mobile measurements were performed in a manufacturing environment [Da23], where throughput, response time, and packet loss were evaluated. In this scenario, the 5G network consisted of a single cell, so handover effects were not considered. The impact of handovers on throughput was investigated in a laboratory environment at [We22]. The results indicated that there was no significant reduction in throughput during handovers. To date, no measurements have been found that investigate the effects of handovers on transmission time, update time, and packet loss rate in process automation environments [Sa24].

3 Use case description

The use cases for the cranes at the Demag Research & Business Factory show in Figure 1 highlight the potential of 5G technology in improving intralogistics operations. The Demag's cranes have many modes as described below:

- Come to Me: The crane autonomously moves to the operator's position, facilitating ease of access and reducing manual intervention.
- Follow Me: The crane follows the operator autonomously, ensuring seamless coordination and minimizing operational delays.
- Go To: The crane moves autonomously to a target position selected by the operator, enabling precise and efficient material handling.
- Follow Machine: The crane autonomously tracks an automated guided vehicle (AGV) or a forklift truck, enhancing the integration of various automated systems within the facility.
- Restricted 3D Zones: Certain three-dimensional zones within the facility can only be accessed under specific conditions, such as reduced speed, ensuring safety and operational control.
- No Go Areas: Designated zones that must not be entered by the crane or AGV, preventing accidents and ensuring the integrity of the operational environment.
- Tandem Mode: Two cranes move synchronously, allowing for the coordinated handling of large or heavy materials, which is critical in complex manufacturing processes.

On these use cases, messages are exchanged between the cranes and other network components via 5G. During the performance analysis, the test system assessed the performance of the communication network, verifying if it meets the operational requirements of the different crane modes. The network's performance in managing these use cases is assessed through detailed measurements, which provide insights into its reliability and efficiency in a demanding industrial environment.

Fig. 1: Research Factory at Demag Cranes in Wetter, Germany and its innovative assist functions.

4 System under test and test system

4.1 System under test: 5G network

The system under test is a 5G NPN, specifically configured for industrial applications within the DEMAG Research Factory. The network is based on 5G Release 15 technology and operates within the N78 band, which is allocated for 5G in many regions worldwide. The network configuration parameters are carefully selected to optimize performance for the specific use cases tested in this environment. Key parameters of the testbed are detailed in Table 1 below:

Parameter	**Value**
Technology	5G Release 15
Band	N78
Frequency	3700-3720 MHz
Bandwidth	20 MHz
Number of cells	1
Carrier Frequency	3710.010 MHz
Subcarrier Spacing	30 kHz
TDD Ratio	UL 3 DL 7
Output Power	23 dBm
actCDrx	True
srPeriodicityMin	Slot 40
ulSchedTimeInterval	15

Tab. 1: 5G network configuration parameters.

The parameters *actCDrx*, *srPeriodicityMin*, and *ulSchedTimeInterval* are essential to improve the efficiency and performance of non-public 5G networks. Since these parameters are not widely known, their descriptions are provided below for better understanding.

- **actCDrx (Active Connected Discontinuous Reception)**: Manages power consumption in connected mode by allowing devices to alternate between active and inactive states, optimizing energy use without losing connectivity.
- **srPeriodicityMin (Scheduling Request Periodicity Minimum)**: Defines the minimum interval for sending scheduling requests, enabling quicker network response and supporting time-sensitive applications like URLLC with frequent resource allocation. The value slot40 for srPeriodicityMin means that scheduling requests can be sent every 40 time slots, approximately every 20 milliseconds, balancing frequency of requests with network responsiveness.
- **ulSchedTimeInterval (Uplink Scheduling Time Interval)**: Specifies the interval for uplink scheduling decisions, affecting latency and resource allocation granularity, crucial for efficient data transmission in high-demand scenarios. For ulSchedTimeInterval, a value of 15 means that uplink scheduling decisions are made every 15 slots.

4.2 5G user equipment

Several devices were utilised for the user equipment (UE) involved in the testing to ensure comprehensive coverage of different scenarios and use cases. These devices were selected based on their ability to support the required 5G features and their relevance to industrial applications. Below is an overview of the types of devices used:

- Telit Centerion Modem (MV-32): This modem card, compliant with 3GPP Release 16, represents the latest in 5G technology tailored for IoT (Internet of Things) applications. It supports both SA and NSA modes, making it ideal for testing in a diverse industrial environment. The MV-32 is particularly noted for its compact form factor and integrated eSIM, which simplifies deployment in various industrial equipment.
- Quectel Modem (RM500Q-GL): The RM500Q-GL is another 5G modem card used in the testing. It operates in the sub-6GHz band and is optimized for both industrial and eMBB (enhanced Mobile Broadband) applications. This device supports high-speed data transmission, with maximum downlink rates of 2.5 Gbps and uplink rates of 900 Mbps, making it suitable for the high-demand environment of the Demag Research Factory.
- QualiPoc Android with CrossCall Z5: This portable device was used to conduct service quality tests across various locations within the factory. It offers detailed insights into signalling and IP trace, which are crucial for analyzing the radio interface and optimizing the quality of experience in the mobile network.
- Siemens Scalance MU856 Router: It is an industrial-grade 5G router designed for reliable connectivity in harsh environments. It supports both SA and NSA 5G networks and complies with 5G Release 15. The router features Ethernet interfaces, ensuring seamless integration with industrial equipment in the Demag Research Factory.

4.3 Influencing quantities

Influencing quantities are parameters that describe conditions affecting the performance of a system under test, but they are not the primary focus of the test. These quantities are critical for accurately modeling and evaluating the performance of communication systems, as they represent external factors that can impact the outcomes. The three main influencing quantities used in this work are:

- **Transfer Interval (TI)**: It specifies the time interval between two initiations of data transmission at the source. For periodic transmissions, a specific time value is given, while for aperiodic transmissions, a mean value, standard deviation, and distribution function are used.
- **User Data Length**: It specifies the number of octets passed on for transmission at the reference interface of the source.

- **Distance Between Wireless Devices**: It specifies the physical distance between logical endpoints bridged by wireless data transmission. It is also represented by the position parameter.

4.4 Active measurements: performance parameters

Active measurements involve utilizing a user device authenticated and connected to the cellular network, allowing applications like voice calls or data sessions to operate. In compliance with [VD19], the following KPIs are used:

- **Message Loss Ratio (MLR):** This ratio of lost messages to total messages sent indicates the reliability of the communication system.
- **Transmission Time (TT):** Measures the time taken to transmit user data from source to target, crucial for assessing the system's responsiveness and real-time capabilities. The P95 percentile of TT is used to gauge performance, marking the maximum time that 95% of transmissions do not exceed.
- **Update Time (UT):** Measures the time between two received messages in the destination end point.
- **Response Time (RT):** It is the time interval from the initiation of a request at the reference interface of the service user to the identification of the response at the same interface. Essentially, it measures the time taken for packets to travel to a reflector unit in the network and back, also known as two-way latency [In23] or round-trip time (RTT).

4.5 Passive measurements: signal strength and quality indicators

Passive measurements are conducted using a network scanner that operates without establishing an active connection to the network. As a result, the scanner does not require authentication credentials, such as a SIM card. The network scanner detects RF signals transmitted by base stations across multiple configurable target frequency bands. These measurements primarily rely on broadcast and synchronization signals, which are consistently present within the network. The key parameters measured are:

- **SS Reference Signal Received Power (SS-RSRP):** SS-RSRP is the linear average of the power contributions (measured in Watts) from resource elements carrying secondary synchronization signals within the specified measurement frequency bandwidth. This measurement provides an indication of the network's coverage.
- **SS Signal-to-Noise and Interference Ratio (SS-SINR):** SS-SINR measures the quality of the signal by comparing the power of the desired signal to the combined power of interference and noise. A higher SINR indicates better signal quality.

4.6 Test system

The test system, designed to rigorously evaluate the 5G network's performance under various industrial use cases, incorporates advanced measurement tools and devices. A key component of this system is the Funk Transfer Tester (FTT), developed to analyze the time and error behavior of the 5G system from an application perspective. Operating on a FPGA, the FTT offers high-precision data generation and measurement, with accuracy down to less than 1 microsecond. This precision is vital for assessing the 5G network's suitability for time-sensitive applications, providing live measurement displays for real-time assessment and detailed analysis. The architecture of the test system including the FTT for transmition time measuremed is shwon in Figure 2.

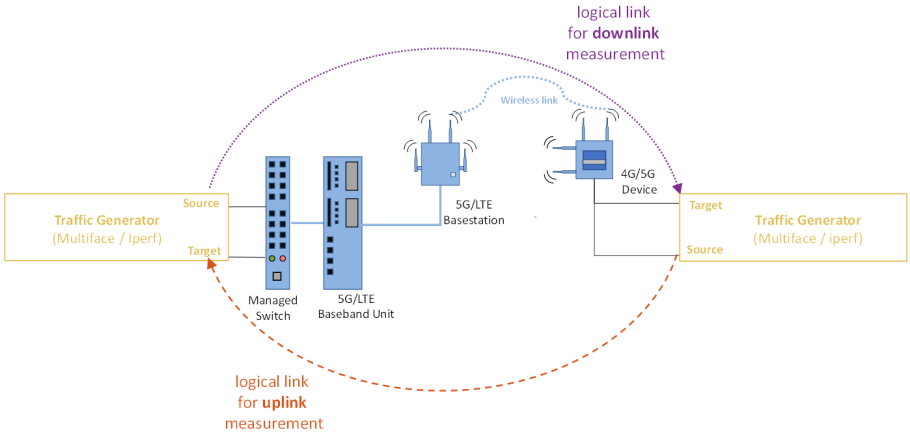


Fig. 2: Uplink and downlink performance testing of 5G network with ifak's test system.

Additionally, the test system integrates a range of R&S measurement tools, including:

- R&S TSMA6 Network Scanner with R&S ROMES4 Software: This system performs drive and walk testing, measuring all relevant mobile communication technologies, including 5G NR. It assesses coverage, signal quality, and other key performance indicators (KPIs) across the facility.
- Telit and Quectel Modems: These modems are controlled by the R&S ROMES4 Software to simulate industrial IoT traffic, enabling the testing of data throughput, latency, and reliability under typical use case conditions.
- QualiPoc Android with CrossCall Z5: This portable device provides service quality tests, offering detailed insights into the radio interface and the overall user experience within the network.

The architecture of the test system is described in Figure 3.

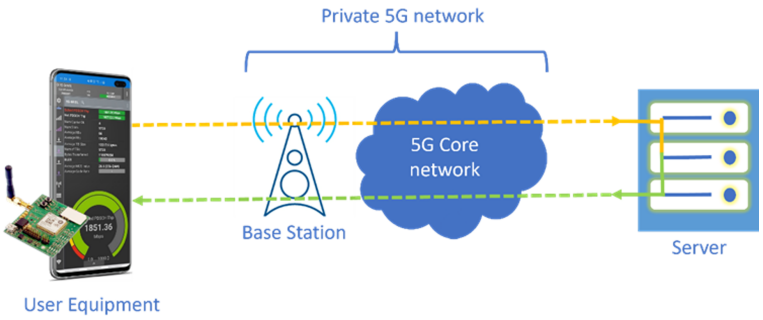


Fig. 3: Response time measurements on 5G network with R&S's test system.

The test devices communicate with a server connected to local breakout of the 5G core network. The server acts as counterpart for data rate tests (e.g., implementing Iperf) or latency tests (reflecting packets according to the ITU standardized test method described in Ref ITU-T G.1051).

Together, these tools form a robust test system that thoroughly evaluates the 5G network's performance, ensuring it meets the stringent requirements of industrial intralogistics and material handling applications.

5 Measurements and results

5.1 Comparison between LTE and 5G

Results and analysis of the measurement campaigns are presented, comparing LTE in the first measurement campaign (MC1) with 5G in the second campaign (MC2). Both campaigns included several test cases. In this section, we present downlink and uplink results, as illustrated in Table 2. The subsequent figures demonstrate the specific downlink results. Traffic in both cases followed a periodic pattern, with 30-byte packets being sent every 23 ms.

Tab. 2: Comparison between MC1 and MC2

TI [ms]	LL Direction	TT Min [ms]		TT P95 [ms]		UT Mean [ms]		UT SD [ms]	
		MC1	MC2	MC1	MC2	MC1	MC2	MC1	MC2
23	DL	5.1	3.2	7.6	4.7	23	23	2.1	2.7
23	UL	5.6	2.8	9.8	13.9	23	23	2.2	5

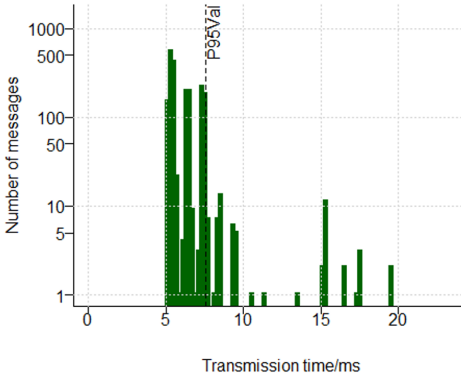


Fig. 4: First measurament campaing (LTE).

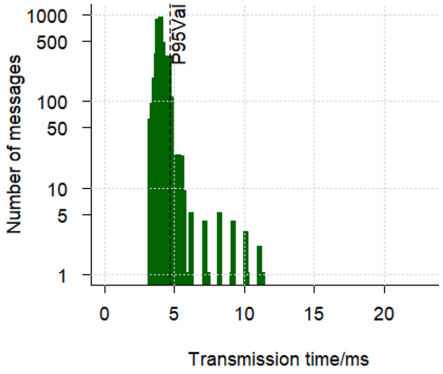


Fig. 5: Second measurement campaing (5G).

The comparison between LTE and 5G performance highlights significant improvements with 5G, particularly in reduced TT. In downlink, 5G achieved a minimum transmission time of 3.2 ms compared to 5.1 ms with LTE, and the 95th percentile decreased from 7.6 ms to 4.7 ms, indicating faster and more consistent performance. Uplink results show 5G's minimum TT dropped to 2.8 ms from 5.6 ms, though with a slight increase in variability for UT.

5.2 Influence of network parametrization

Results and analysis of the second measurement campaign focusing on the effects of network parameter adaptation are presented below. The tests carried out to evaluate the performance of the uplink and downlink configurations revealed significant differences between the two configurations tested.

Tab. 3: Network Parametrization

Test Case	actCDrx	srPeriodicityMin	ulSchedTimeInterval
TC01 (Default)	false	Slot40	15
TC02 (Adapted)	false	Slot10	4

Between the two configurations, the second presents the best performance in terms of transmission time in uplink. The P95 value was the lowest recorded, indicating a high concentration of transmission times around a low value. The results indicate that Configuration 2 is the most efficient in uplink without major effects in downlink.

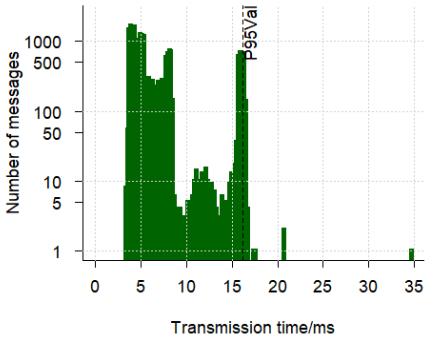


Fig. 6: Histogram of TT in uplink for TC01 (Default configuration).

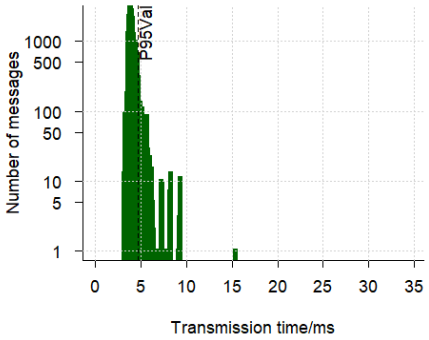


Fig. 7: Histogram of TT in downlink for TC01 (Default configuration).

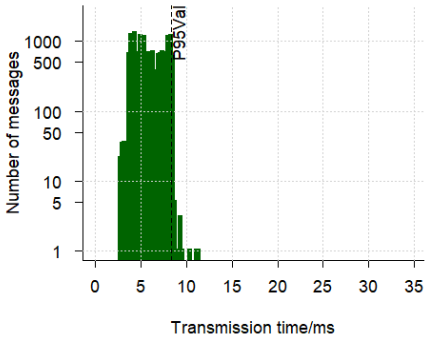


Fig. 8: Histogram of TT in uplink for TC02 (Adapted configuration).

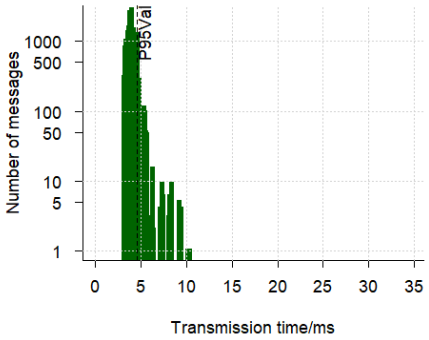


Fig. 9: Histogram of TT in downlink for TC02 (Adapted configuration).

The results demonstrate the impact of network parameter adjustments on 5G performance in industrial settings, particularly in uplink transmission. Among the two configurations tested, Configuration 2, which featured a reduced scheduling time interval and adapted settings, exhibited the best uplink performance, with the lowest 95th percentile TT(P95). This configuration resulted in a more concentrated distribution of TT around lower values, indicating improved network efficiency and responsiveness. The findings suggest that fine-tuning specific network parameters, such as scheduling intervals and activation settings, can significantly enhance the performance of 5G networks, particularly for time-sensitive applications.

5.3 Response time measurements

A number of different traffic patterns to select from are available. Table 4 below provides the traffic characteristics of the data services applied for the round trip measurements results.

Tab. 4: Overview of traffic profile characteristics for the interactivity test

Name	Transfer Interval (ms)	User data length (Bytes)	Data rate (kbps)	DL/UL Delay budget (s)	Max. Packet Error Rate (%)	Session duration (s)
Constant Low	8	100	100	2	2	2
Constant Medium	5	650	1,000	2	2	2
Constant High	0.77	1,450	15,000	2	2	2
Constant Long	6.67	250	300	2	2	30
I4.0	1.6	100	500	0.02	0.02	30

Figure 10 shows the ground floor plan corresponding to Figure 1, i.e., the test field of interest. Three main static positions were used for the latency measurements (P1, P2, and P3) as illustrated in the figure.

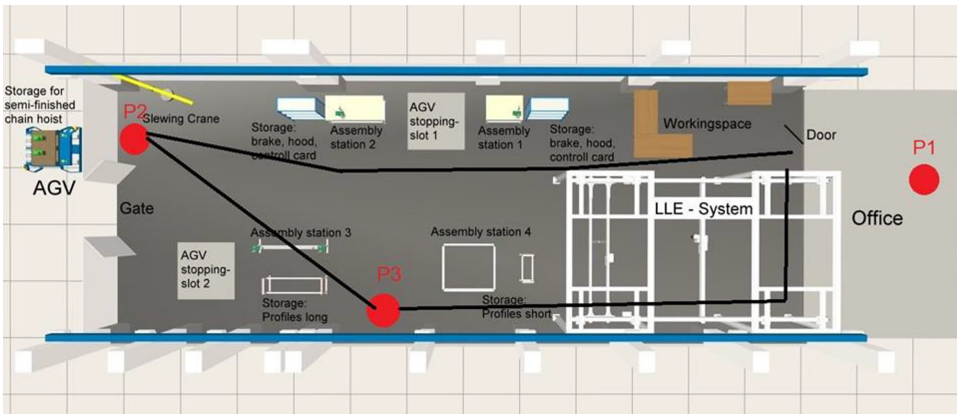


Fig. 10: Ground floor plan of the test field

The following table summarizes the measurement results for stationary tests at positions P1, P2, and P3. During all test sessions, the Packet Loss Ratio (PLR) was always 0%, indicating that no packets were lost during our measurement campaigns. As expected, RT values increase with traffic patterns requiring a higher data rate on the air interface. For the traffic profile I4.0 the delay budget is 20 ms while for the other profiles, it is 2s. The discarded packets (Disc. Ratio), which are those arriving later than the specified delay budget, are 0% for all traffic patterns except I4.0.

Tab. 5: Measurement results for stationary tests.

Position	Traffic Pattern	# Packets	PLR	Disc. Ratio	RT Median	PDV Median
P1	Low	250	0%	0%	16.4 ms	8.4 ms
	Med	400	0%	0%	20.7 ms	11.7 ms
	High	2600	0%	0%	23.25 ms	12.3 ms
	Long	4500	0%	0%	17.7 ms	10.1 ms
	I4.0	18750	0%	~35%	16.3 ms	10.9 ms
P2	Low	250	0%	0%	15.4 ms	7.2 ms
	Med	400	0%	0%	19.4 ms	9.1 ms
	High	2600	0%	0%	23.4 ms	11.8 ms
	Long	4500	0%	0%	16.1 ms	8.7 ms
	I4.0	18750	0%	~22.6%	15.7 ms	9.6 ms
P3	Low	250	0%	0%	16.1 ms	8.1 ms
	Med	400	0%	0%	20.0 ms	8.5 ms
	High	2600	0%	0%	21.4 ms	11.9 ms
	Long	4500	0%	0%	17.4 ms	9.6 ms
	I4.0	18750	0%	~34%	16.0 ms	10.6 ms

6 Conclusion

This study evaluated the performance of a non-public 5G network (3GPP Rel. 15) using commercial equipment on a 5G-ACIA endorsed testbed. The results showed robust network quality and efficient coverage at the physical and application layers, even when the user equipment was more than 100 meters away from the antennas. Data rate performance met the expected benchmarks, and the tests, including a continuous 24-hour session, reported no packet loss, highlighting the reliability of 5G in intralogistics complex operations in the industrial environment.

The comparative analysis between 5G and LTE showed significant improvements of 5G, especially in the reduction of transmission times and data reliability. Research has shown that 5G can efficiently support intralogistics operations, such as autonomous bridging functions and real-time data exchanges, even in challenging industrial scenarios. Different network configurations were tested, highlighting the need for adjustments to improve performance, especially in the uplink direction.

The findings suggest that network configuration settings, such as specific uplink configurations, are essential to maximizing 5G performance in industrial applications. Compatibility between the network and test equipment also showed improvements compared to previous campaigns, reinforcing the maturity of 5G for industrial environments. However, there is still room for further optimization, especially in time-sensitive applications.

The adoption of non-public 5G networks in industrial environments and the sharing of use cases show promising potential for the growth of the 5G ecosystem, encompassing both network infrastructure and connected devices. Progress in the optimization and adaptation of 5G networks will be crucial to expanding their application in manufacturing and other industrial sectors, cementing their role as a key technology for automation and operational efficiency.

Acknowledgement

This work was funded in parts by the project "Industrial Radio Lab Germany" under contract 16KIS1013 funded by the Federal Ministry of Education and Research, Germany.

References

[5G23] 5G-ACIA: 5G Performance Evolution for Material Handling in Manufacturing, 2023, URL: https://5g-acia.org/testbeds/5g-performance-evolution-for-material-handling-in-manufacturing/, visited on: 09/06/2023.

[An22] Ansari, J.; Andersson, C.; Bruin, P.; Farkas, J.; Grosjean, L.; Sachs, J.; Torsner, J.; Varga, B.; Harutyunyan, D.; König, N.; Schmitt, R.: Performance of 5G Trials for Industrial Automation. Electronics 11, p. 412, 2022, DOI: 10.3390/electronics11030412.

[Ca22] Caro, J. B.; Ansari, J.; Sachs, J.; de Bruin, P.; Sivri, S.; Grosjean, L.; König, N.; Schmitt, R. H.: Empirical Study on 5G NR Cochannel Coexistence. Electronics 11 (11), 2022, ISSN: 2079-9292, DOI: 10.3390/electronics11111676, URL: https://www.mdpi.com/2079-9292/11/11/1676.

[Da23] Damsgaard, S. B.; Segura, D.; Andersen, M. F.; Aaberg Markussen, S.; Barbera, S.; Rodríguez, I.; Mogensen, P.: Commercial 5G NPN and PN Deployment Options for Industrial Manufacturing: An Empirical Study of Performance and Complexity Tradeoffs. In: 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Pp. 1–7, 2023, DOI: 10.1109/PIMRC56721.2023.10293869.

[Dí20] Díaz Zayas, A.; Caso, G.; Alay, Ö.; Merino, P.; Brunstrom, A.; Tsolkas, D.; Koumaras, H.: A Modular Experimentation Methodology for 5G Deployments: The 5GENESIS Approach. Sensors 20 (22), 2020, ISSN: 1424-8220, DOI: 10.3390/s20226652, URL: https://www.mdpi.com/1424-8220/20/22/6652.

[Fj22] Fjodorov, A.; Masood, A.; Alam, M. M.; Pärand, S.: 5G Testbed Implementation and Measurement Campaign for Ground and Aerial Coverage. In: 2022 18th Biennial Baltic Electronics Conference (BEC). IEEE, pp. 1–6, 2022.

[Ha23] Hamidovic, D.; Hadziaganovic, A.; Muzaffar, R.; Bernhard, H.-P.: 5G Campus Network Factory Floor Measurements with Varying Channel and QoS Flow Priorities. In: 2023 International Wireless Communications and Mobile Computing (IWCMC). Pp. 1–6, 2023, DOI: 10.1109/IECON51785.2023.10311715.

[Ho23] Homayouni, S.; Paier, M.; Stangelmayer, G.; Kaipl, C.; Sulz, C.; Schweeger, T.; Rehak, J.: Design and Development of Private 5G Standalone Network for Vertical Industries. In: 2023 International Wireless Communications and Mobile Computing (IWCMC). Pp. 369–374, 2023, DOI: 10.1109/IWCMC58020.2023.10183014.

[In23] International Telecommunication Union: ITU-T Recommendation G.1051: Latency measurement and interactivity scoring under real application data traffic patterns, https://www.itu.int/rec/T-REC-G.1051, Accessed: 2024-09-10, 2023.

[Ke22] Kehl, P.; Ansari, J.; Jafari, M. H.; Becker, P.; Sachs, J.; König, N.; Göppert, A.; Schmitt, R. H.: Prototype of 5G Integrated with TSN for Edge-Controlled Mobile Robotics. Electronics 11 (11), 2022, ISSN: 2079-9292, DOI: 10.3390/electronics11111666, URL: https://www.mdpi.com/2079-9292/11/11/1666.

[Ko22] Kourtis, M.-A.; Oikonomakis, A.; Santorinaios, D.; Anagnostopoulos, T.; Xilouris, G.; Kourtis, A.; Chochliouros, I.; Zarakovitis, C.: 5G NPN Performance Evaluation for I4.0 Environments. Applied Sciences 12 (15), 2022, ISSN: 2076-3417, DOI: 10.3390/app12157891, URL: https://www.mdpi.com/2076-3417/12/15/7891.

[La22] Lackner, T.; Hermann, J.; Dietrich, F.; Kuhn, C.; Angos-Mediavilla, M.; Jooste, W.; Palm, D.: Measurement and comparison of data rate and time delay of end-devices in licensed sub-6 GHz 5G standalone non-public networks. Procedia CIRP 107, pp. 1132–1137, 2022, DOI: 10.1016/j.procir.2022.05.120.

[Ly23] Lyczkowski, E.: Wireless communication on the factory floor supporting agile production, PhD thesis, University of Koblenz-Landau, 2023.

[Ma22] Mallikarjun, S. B.; Schellenberger, C.; Hobelsberger, C.; Schotten, H. D.: Performance analysis of a private 5g sa campus network. In: Mobile Communication-Technologies and Applications; 26th ITG-Symposium. VDE, pp. 1–5, 2022.

[Mo23] Mondal, N.; Block, D.; Kroll, B.; Klingler, F.: Performance evaluation and application of real-time communication with 5G IIoT, Bauhaus-Universität Weimar, 2023, DOI: 10.25673/111641.

[Ri21] Rischke, J.; Sossalla, P.; Itting, S.; Fitzek, F. H.; Reisslein, M.: 5G campus networks: A first measurement study. IEEE Access 9, pp. 121786–121803, 2021.

[Sa24] Saunders, C.; Minne, A.; Müller, E.; Fuchs, C.; Kokowsky, M.; Gnad, A.; Albert, O.; Underberg, L.: 5G Campus Network in Process Automation: Measurement Campaigns and Use Case Demonstrations in TotalEnergies Raffinerie in Leuna. In (GmbH, V. W., ed.): Automation 2024. Accepted for publication, July 2024, VDI Verlag, 2024.

[VD19] VDI/VDE: VDI/VDE Guideline 2185: Radio Based Communication in Industrial Automation (Part 4) - Metrological performance rating of wireless solutions for industrial automation applications, tech. rep., VDI Association of German Engineers, VDE Association for Electrical, Electronic & Information Technologies, 2019, URL: https://www.vdi.de/en/home/vdi-standards/details/vdivde-2185-blatt-2-radio-based-communication-in-industrial-automation-management-of-the-coexistence-of-wireless-solutions.

[We22] Wei, Y. R.; Keshavamurthy, A. S.; Wittmann, R.; Zahonero, A. R.: A Standalone 5G Industrial Testbed Design Considerations for Industry 4.0. 2022 52nd European Microwave Conference (EuMC), pp. 884–887, 2022, URL: https://api.semanticscholar.org/CorpusID:253251372.

Evaluierung von Time-Sensitive Networking in OMNeT++ für die Simulation von IEC/IEEE 60802-konformen Netzwerken

Alexander Biendarra[1], Janis Albrecht[1]

Abstract: Die Zusammenführung von Informationstechnologie (IT) und operativer Technologie (OT) ermöglicht ein breites Spektrum an industriellen Anwendungen und sorgt dafür, dass erzeugte Daten großflächig für die Verarbeitung innerhalb von Unternehmen bereitgestellt werden können. Time-Sensitive Networking bietet mit seinen Mechanismen die Grundlage, um echtzeitfähige Netzwerke als Grundlage bereitzustellen. Die IEC/IEEE 60802 bietet die Rahmenbedingungen für eine einheitliche Umsetzung inklusive der Konfiguration solcher Netzwerke. Eine in Geräten nutzbare Umsetzung gibt es bisher nicht. Daher wird in diesem Beitrag der aktuelle Funktionsumfang der Implementierungen von TSN-Mechanismen in der Open Source-Simulationsumgebung OMNeT++ in Kombination mit dem Framework INET untersucht. Hierfür wird eine Evaluierung der TSN-Mechanismen durchgeführt. Die Evaluierung zeigt, dass ein IEC/IEEE 60802-konformes Netzwerk nicht realisiert werden kann. Gründe sind der derzeitige Stand der Implementierung und die zum Teil fehlende Interoperabilität der Simulations-Module.

Keywords: Time-Sensitive Networking (TSN), IT/OT-Konvergenz, Interoperabilität, IEC/IEEE 60802, Simulation von Netzwerken, OMNeT++, INET

1 Motivation

Ethernet Time-Sensitive Networking (TSN) bietet mit seinen Mechanismen die Möglichkeit der Realisierung von IT/OT-konvergenten Netzwerken [Sc22]. Um ein einheitliches Verständnis des in industriellen Netzwerken genutzten Funktionsumfangs von TSN sicherzustellen und einen interoperablen Ansatz der Konfiguration eines TSN-Netzwerks zu erreichen, wird zum gegenwärtigen Zeitpunkt der Profilstandard IEC/IEEE 60802 erstellt [II24]. Auch wenn inzwischen eine Vielzahl von TSN-Geräten am Markt erhältlich ist, werden diese über vom jeweiligen Hersteller bereitgestellte, individuelle Lösungen (bspw. Webinterfaces) konfiguriert. Um neue Konfigurationsansätze eines TSN-Netzwerks, wie in der IEC/IEEE 60802 definiert, schon frühzeitig evaluieren zu können, kann das Mittel der Simulation von Netzwerken genutzt werden. Eine für die Simulation von Ethernet-Netzwerken verfügbare Open Source-Lösung ist die Simulationsumgebung OMNeT++ [VH08] in Kombination mit dem Framework INET [Om24]. Hier stehen seit der Version INET 4.4 verschiedene Module zur Verfügung, die TSN-Mechanismen implementieren. Mithilfe dieser Module soll evaluiert werden, ob eine IEC/IEEE 60802-konforme Umsetzung eines TSN-Netzwerks möglich ist. Mithilfe eines solchen Simulationsmodells können Schwachstellen des Konfigurationsansatzes

[1] Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo,
{Vorname.Nachname}@iosb-ina.fraunhofer.de, https://www.iosb-ina.fraunhofer.de

aufgedeckt werden. Auch können die Erkenntnisse der Simulation genutzt werden, um Engineering-Werkzeuge für TSN-Netzwerke, die in der industriellen Automation eingesetzt werden sollen, um die für die Konfiguration von TSN-Netzwerken notwendigen Einstellungsmöglichkeiten zu erweitern. Automatische Konfigurationsansätze können ebenfalls auf diese Weise erprobt werden [ABJ23]. Des Weiteren ist es möglich mithilfe der Simulation, die Planung und Auslegung von TSN-Netzwerken zu unterstützen oder Überwachungskonzepte für die automatische Fehleranalyse zu entwickeln [FSJ23].

2 Stand der Technik

Dieses Kapitel gibt den Stand der Technik im Bereich Time-Sensitive Networking, der IEC/IEEE 60802 und der Simulation von Ethernet-Netzwerken wieder.

2.1 Time-Sensitive Networking

Time-Sensitive Networking (TSN) bezeichnet eine Menge von IEEE-Standards, die durch die Time-Sensitive Networking Task Group definiert werden. Im Fokus stehen Standards der Sicherungsschicht des ISO/OSI-Referenzmodells. Das Ziel der Standardisierung von TSN ist es, eine echtzeitfähige und robuste Kommunikation auf Basis von IEEE-Standards zu realisieren. TSN umfasst aktuell über 20 verabschiedete Standards oder Standarderweiterungen sowie 17 Standards oder Standarderweiterungen, die derzeit erarbeitet werden [Ts24]. Einer dieser Standards, der gemeinsam mit der IEC erarbeitet wird, ist ein Profil, das definiert, wie TSN innerhalb der industriellen Automation genutzt werden soll. Es trägt den Titel *IEEE/IEC 60802 Time-Sensitive Networking Profil for Industrial Automation* [II24].

Im Rahmen der IEC/IEEE 60802 werden fünf TSN-Mechanismen definiert, die für die industrielle Automation eingesetzt werden sollen. Die Mechanismen decken die Anforderungen der Zeitsynchronisation, die Reduzierung der Latenz in der Datenübertragung, die Redundanz von Kommunikationspfaden und der Limitierung von Best-Effort Datenverkehr ab.

Die Zeitsynchronisation nach IEEE 802.1AS-2020 [Ie20] wird genutzt, um ein einheitliches Zeitverständnis innerhalb des Netzwerks bereitzustellen. Sie kann genutzt werden, um die Kommunikation im Netzwerk aufeinander abzustimmen oder für Anwendungsfälle wie die Nachverfolgung von Ereignissen von Applikationen oder Produktionsschritten, um einen kausalen Zusammenhang herzustellen. Der Standard IEEE 802.1AS-2020 definiert das *generalized Precision Time Protocol (gPTP)* Profil, das eine Teilmenge des Funktionsumfangs des Standards IEEE 1588 PTPv2 nutzt.

Mit dem Mechanismus *Time Aware Traffic Shaping (TAS)* nach IEEE 802.1Q [Ie22] (IEEE 802.1Qbv) wurde ein Mechanismus definiert, der auf einem Zeitschlitzverfahren

beruht. Die Kommunikationsdaten werden dabei anhand einer Information im Ethernet-Frame, etwa dem Ethertype, in unterschiedlichen Zeitschlitzen übertragen. Die Steuerung der Kommunikation, wann welche Daten übertragen werden dürfen und für welche Dauer, wird innerhalb einer sogenannten *Gate-Control-List (GCL)* definiert. Diese wird zyklisch durchlaufen. Die Zeit, die für die GCL zugrunde gelegt wird, wird durch die Zeitsynchronisation nach IEEE 802.1AS-2020 im Netzwerk verteilt.

Ein weiteres Verfahren, um die Kommunikation zu priorisieren und die Latenz des Datenverkehrs zu reduzieren, ist Frame Preemption nach IEEE 802.1Q [Ie22] (IEEE 802.1Qbu) und IEEE 802.3br [Ie16]. Der Datenverkehr wird dafür in unterbrechbaren Datenverkehr (preemptable) und bevorzugten Datenverkehr (express) aufgeteilt. Die technische Umsetzung dafür erfolgt innerhalb der Sicherungsschicht des ISO/OSI-Referenzmodells und orientiert sich am Prinzip der Fragmentierung. Bevorzugter Datenverkehr ist hierbei in der Lage, unterbrechbaren Datenverkehr, sowohl von der Übertragung als auch während der Übertragung selbst, zu unterbrechen.

Mit dem Mechanismus *Frame Replication and Elimination for Reliablity* nach IEEE 802.1CB [Ie17a] wird das Thema der Robustheit von Kommunikationsverbindungen adressiert, indem die Kommunikation über redundante Pfade definiert wird. Für die technische Umsetzung werden Ethernet-Frames um einen 6 Byte langen *Redundancy-Tag* (R-Tag), der sich zwischen dem VLAN-Tag und dem Ethertype befindet, ergänzt.

Als fünfter TSN-Standard wird *Per-Stream Filtering and Policing* nach IEEE 802.1Qci [Ie17b] definiert. Mit dem Standard soll die Limitierung der Bandbreitennutzung und der Schutz von Speicherressourcen bei Überlastsituationen erreicht werden. Hierzu wird ein Verfahren angewandt, was durch das *Metro Ethernet Forum* (MEF)[Me13] standardisiert ist.

2.2 Profilstandard IEC/IEEE 60802

Um ein einheitliches Verständnis zu erlangen, wie TSN in der industriellen Automation eingesetzt werden soll, wird als gemeinsame Aktivität der IEC und der IEEE der Profilstandard *IEC/IEEE 60802 Time-Sensitive Networking Profil for Industrial Automation* [II24] erarbeitet. Aktuell liegt der Standard als Entwurf in der Version 2.4 vor. Ziel der Aktivität ist neben der Unterstützung der IT/OT-Konvergenz die Realisierung der parallelen Nutzung von Kommunikationsnetzwerken durch mehrere Industrial Ethernet Protokolle. Adaptierungen der Standards für PROFINET [LGAS20] und OPC UA [BW20] an die IEC/IEEE 60802 sind bereits verabschiedet oder in der Umsetzung.

Um eine große Zahl an Applikationen abdecken zu können, die auch durch verschiedene Industrial Ethernet Protokolle bedient werden, werden Übersetzungstabellen definiert. Jede Tabelle enthält acht verschiedene Verkehrsklassen von Kommunikation. Die Verkehrsklassen unterteilen sich in die isochrone Kommunikation, die zyklische synchrone Kommunikation, die asynchrone Kommunikation, Alarme und Ereignisse, Konfiguration und Diagnose, die Netzwerksteuerung und zwei Arten von Best-Effort

Datenverkehr. Damit eine parallele Nutzung des Netzwerks möglich ist, müssen die konkreten Verkehrsklassen des jeweiligen Protokolls durch die Übersetzungstabelle angeglichen werden.

Innerhalb der IEC/IEEE 60802 werden Geräte in Form einer *Industrial Automation Station* (IA-Station) definiert. Eine IA-Station zeichnet sich dadurch aus, dass sie eine oder mehrere funktionale Einheiten besitzt, die die Aufgabe einer Steuerung oder eines Endgerätes übernehmen. Innerhalb der funktionalen Einheiten wird die Anbindung an das Kommunikationsnetzwerk realisiert. Möglich ist die Ausprägung als End-Station mit einem oder mehreren Kommunikations-Schnittstellen ohne die Funktion der Weiterleitung von Ethernet-Frames oder als Bridge, mit mehr als einer Kommunikations-Schnittstelle mit der Funktion der Weiterleitung von Ethernet-Frames.

Des Weiteren definiert die IEC/IEEE 60802 zwei Konformitätsklassen, in die IA-Stationen eingeordnet werden können. Diese werden als die Konformitätsklasse A (Conformance Class A) und Konformitätsklasse B (Conformance Class B) definiert. Die Einordnung erfolgt auf Basis von zu erfüllenden Anforderungen an die IA-Station, die enthaltene End-Station Komponente und die enthaltene Bridge-Komponente. Ob eine IA-Station in die Konformitätsklasse A oder die Konformitätsklasse B eingeordnet wird, hängt maßgeblich von den Leistungsdaten des Gerätes und den unterstützten TSN-Mechanismen ab. Geräte der Konformitätsklasse A sind leistungsstärker und umfassen mehr TSN-Funktionalität als Geräte der Konformitätsklasse B.

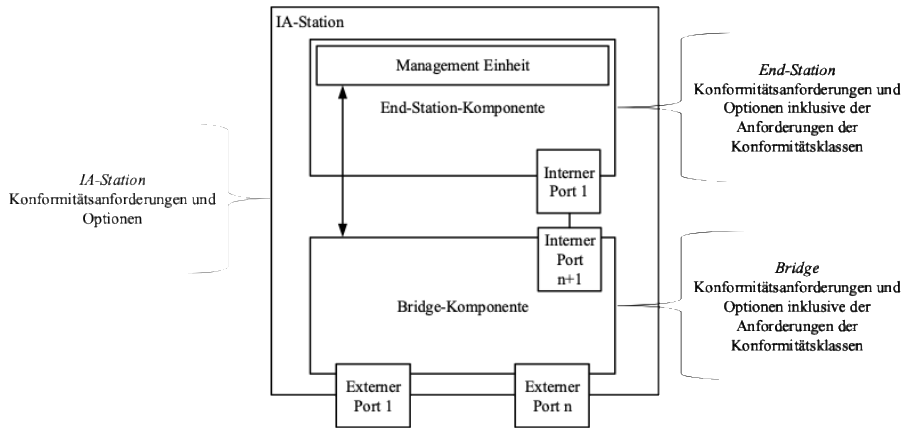


Abbildung 1: Aufteilung von Anforderungen an eine IA-Station (In Anlehnung an [II24])

Für die Konfiguration der IA-Stationen im Netzwerk soll ein zentralisierter Konfigurationsansatz genutzt werden. Dieser wird im Standard IEEE 802.1Qcc [Ie18] definiert. Die Informationen, die für die Konfiguration notwendig sind, werden über das *Network Configuration Protocol* (NETCONF) in Kombination mit dem Datenmodell YANG durchgeführt. [II24]

2.3 Simulation von Ethernet-Netzwerken

Für die Simulation von Ethernet-Netzwerken wird die diskrete, ereignisorientierte Simulation eingesetzt. Im Mittelpunkt dieses Ansatzes stehen Ereignisse, die unter anderem das Versenden eines Ethernet-Frames oder eine Zustandsänderung im System darstellen. Jedes Ereignis besitzt einen definierten Zeitpunkt, wann das Ereignis auftritt und Regeln, die den Ablauf der Simulation steuern. Das Auftreten eines Ereignisses kann einer Ereignisliste entnommen werden. In dieser werden Ereignisse, chronologisch, nach ihrem Auftrittszeitpunkt sortiert, abgelegt. [Rj91]

Eine Möglichkeit, Simulationsmodelle für Ethernet-Netzwerke zu erzeugen, bietet die Open Source-Simulationsumgebung OMNeT++ in Kombination mit dem Framework INET. OMNeT++ nutzt die Eclipse IDE und stellt die Basisfunktionalitäten für die Erzeugung von diskreten, ereignisorientierten Simulationsmodellen bereit. OMNeT++ folgt dabei einem hierarchischen Modellansatz. Simulationsmodelle werden in Form von Modulen organisiert. Auf der untersten Ebene wird das funktionale Verhalten in einem *Simple Module* abgebildet. Die Implementierung eines solchen Moduls erfolgt in der Programmiersprache C++. Werden mehrere *Simple Modules* miteinander verschaltet, entsteht ein *Compound Module*. Diese Module können wiederum miteinander verschaltet werden. Mit diesem Vorgehen kann ein beliebig komplexes System für die Simulation realisiert werden. Die oberste Ebene des so entstandenen Simulationsmodells wird als Netzwerk bezeichnet.[VH08]

Die Erweiterung INET stellt ein Framework bereit, das innerhalb von OMNeT++ für die Simulation von Kommunikationsprotokollen eingesetzt werden kann. Seit der Version 4.4 ist auch die Funktionalität von einzelnen TSN-Standards enthalten. [Om24]

3 Simulationsmodell

Wie in Abschnitt 2.3 erwähnt, wurde das INET-Framework mit der Version 4.4 um Implementierungen von TSN-Standards erweitert. Dies sind die Zeitsynchronisation, Time-Aware Shaping, Ethernet Frame Preemption, Frame Replication and Elemination for Reliablity und Per-Stream Filtering and Policing. Des Weiteren wurde, mit dem Mechanismus Cut-Through, ein Weiterleitungsverfahren für Ethernet Frames ergänzt, auf das in dieser Arbeit nicht weiter eingegangen wird, da es keinen TSN-Standard darstellt. Nicht durch die Simulation abgedeckt sind die Möglichkeiten der Konfiguration des Netzwerkes. Eine Implementierung von NETCONF, YANG oder dem zentralisierten Konfigurationsmodell aus dem Standard IEEE 802.1Qcc[Ie18] sind nicht vorhanden. Daher erfolgt die Konfiguration der TSN-Mechanismen in dieser Arbeit durch die von OMNeT++ bereitgestellten Möglichkeiten der Modulparametrierung.

3.1 Topologie des Simulationsmodells

Die Topologie, die dem Simulationsmodell zugrunde liegt, orientiert sich an der hierarchischen Struktur von industriellen Automatisierungsnetzwerken aus [II18]. Im Mittelpunkt steht das Fabrik-Netzwerk, das die IT- und OT-Komponenten miteinander verbindet. Im Bereich der OT unterteilt sich das Netzwerk weiter in die Ebenen der Produktionslinien, die eine oder mehrere Produktionszellen beinhalten, die wiederum eine oder mehrere Maschinen enthalten. Jede Ebene stellt eine eigene TSN-Domäne dar. Innerhalb jeder Domäne befinden sich Komponenten wie Steuerungen, Endgeräte mit unterschiedlichen Echtzeitanforderungen, Netzwerkinfrastrukturgeräte und Geräte, die dem Bereich der IT zugeordnet werden können. Die Kommunikation erfolgt sowohl innerhalb der TSN-Domäne als auch über die Grenzen der TSN-Domänen hinaus.

Übertragen auf die Topologie des Simulationsmodells, ergibt sich die in Abbildung 2 dargestellte Struktur des Netzwerkes.

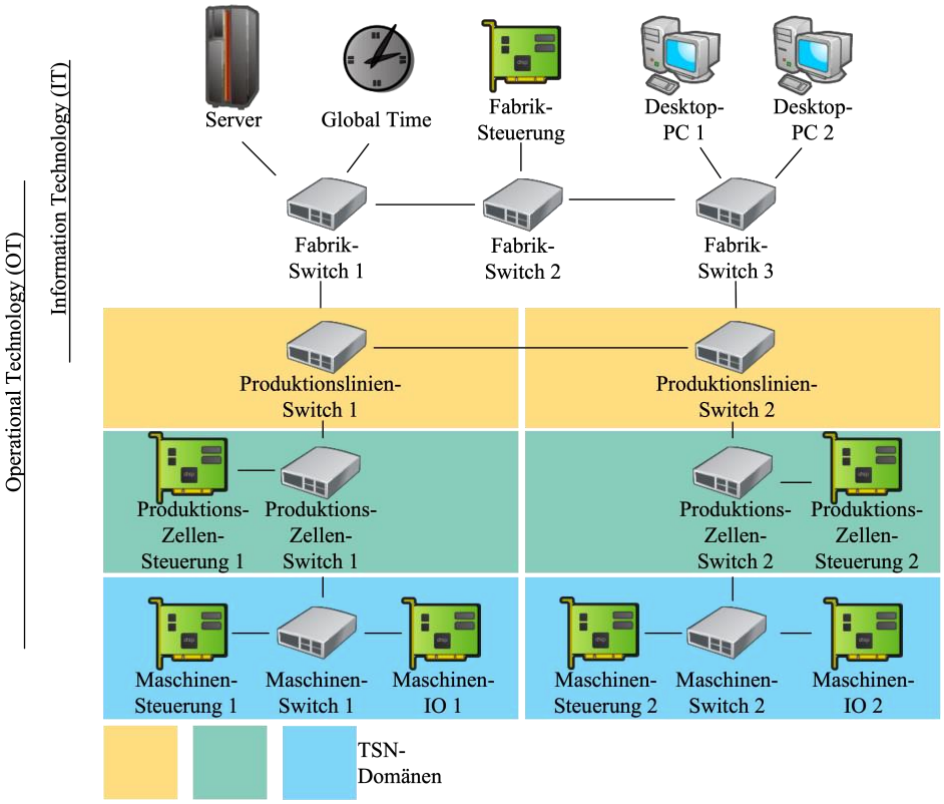


Abbildung 2: Topologie des Simulationsmodells

Innerhalb des Fabrik-Netzwerks werden ein Server, zwei Desktop-PCs, eine globale Zeitquelle (Global Time) und eine Fabrik-Steuerung angelegt. Alle Komponenten werden über Switches im Fabrik-Netzwerk miteinander verbunden. In der darunterliegenden Ebenen werden die Komponenten für die Produktionslinien, die Produktionszellen und die Maschinen angelegt. Jede der dargestellten Ebenen (ProduktionslinieX, ProduktionszelleX und MaschineX) sollen voneinander getrennte TSN-Domänen darstellen. Die Kommunikationsverbindungen repräsentieren eine Ethernet-Verbindung mit einer Datenrate von 1 Gbit/s.

3.2 Parametrierung des Simulationsmodells

Damit die einzelnen TSN-Mechanismen innerhalb des Netzwerkes evaluiert werden können und gleichzeitig der Aufwand der Parametrierung minimal bleibt, werden die zu testenden Funktionen in voneinander getrennten Abschnitten parametriert. Dadurch ist es möglich, Kombinationen aus Parametrierungen des Simulationsmodells in einem Simulationslauf miteinander zu verknüpfen.

Für die Zeitsynchronisation werden zwei Zeitdomänen parametriert. Eine für die Synchronisation der Working-Clock und eine für die Synchronisation der Global-Time. Die Rolle des *TimeTransmitter* für die Working-Clock Domäne übernimmt die Komponente *Global Time* (Abbildung 2). Für die Working-Clock Domäne ist der *TimeTransmitter* die Fabrik-Steuerung. Die Switches innerhalb der Topologie agieren in der Rolle eines *Time-Aware Relays*, die Steuerungen und Endgeräte (IOs) als *TimeReceiver*. [Ie24]

Die TSN-Funktion TAS soll auf der Maschinenebene innerhalb von *Maschine 1* (Abbildung 2) evaluiert werden. Hier wird eine Konfiguration, bestehend aus drei Zeitschlitzen, parametriert. Diese trennt Prozessdaten, Best-Effort Datenverkehr und Netzwerkmanagement-Datenverkehr voneinander ab.

Um die Funktion von Ethernet Frame-Preemption untersuchen zu können, wird die Funktion innerhalb des Fabrik-Netzwerks parametriert. Hier soll der Datenverkehr, der ausgehend von der Fabrik-Steuerung verteilt wird, dem Datenverkehr, der zwischen den IT-Komponenten ausgetauscht wird, bevorzugt behandelt werden.

Der TSN-Mechanismus Frame Replication and Elimination for Reliability soll innerhalb der Netzwerksegmente Fabrik und Produktionslinie (Abbildung 2) zum Einsatz kommen. Hier soll der Fabrik-Steuerung ein redundanter Pfad zur Verfügung gestellt werden, um mit den angeschlossenen Produktionszellen und Maschinen kommunizieren zu können.

Die TSN-Funktion Per-Stream Filtering and Policing wird innerhalb des Fabriknetzwerks (Abbildung 2) parametriert. Dafür wird Best-Effort Datenverkehr erzeugt, der sich über die Simulationszeit hinweg in der Datenrate verändert. Parallel dazu wird ein Prozessdatenaustausch von der Fabrik-Steuerung mit den Maschinen-Netzwerken parametriert.

```
#Referenz für die Zeitdomänen 0 = Gloabl Time, 1 = Working Clock
**.clock[0].referenceClock = "Globale_Zeitquelle.clock"
**.clock[1].referenceClock = "Fabrik_Steuerung.clock.clock[1]"

#Globale_Zeitquelle
*.Globale_Zeitquelle.gptp.typename = "MultiDomainGptp"
*.Globale_Zeitquelle.gptp.numDomains = 1
*.Globale_Zeitquelle.clock.typename = "SettableClock"
*.Globale_Zeitquelle.clock.initialClockTime = 1000s
*.Globale_Zeitquelle.gptp.domain[0].gptpNodeType = "MASTER_NODE"
*.Globale_Zeitquelle.gptp.domain[0].clockModule = "Globale_Zeitquelle.clock"
*.Globale_Zeitquelle.gptp.domain[0].syncInterval = 125ms
*.Globale_Zeitquelle.gptp.domain[0].masterPorts = ["eth0"]

# Fabrik_Steuerung
*.Fabrik_Steuerung.gptp.typename = "MultiDomainGptp"
*.Fabrik_Steuerung.gptp.numDomains = 2
*.Fabrik_Steuerung.clock.typename = "MultiClock"
*.Fabrik_Steuerung.clock.numClocks = 2
*.Fabrik_Steuerung.gptp.domain[0].gptpNodeType = "SLAVE_NODE"
*.Fabrik_Steuerung.gptp.domain[0].slavePort = "eth0"
*.Fabrik_Steuerung.gptp.domain[1].gptpNodeType = "MASTER_NODE"
*.Fabrik_Steuerung.gptp.domain[1].syncInterval = 125ms
*.Fabrik_Steuerung.gptp.domain[1].pdelayInitialOffset = 0.1ms
*.Fabrik_Steuerung.gptp.domain[1].pdelayInterval = 125ms
*.Fabrik_Steuerung.gptp.domain[1].masterPorts = ["eth0"]
```

Abbildung 3: Parametrierung des Simulationsmodells am Beispiel eines Code-Ausschnittes der Zeitsynchronisation

4 Evaluierung & Ergebnisse

Für die Evaluierung der einzelnen TSN-Mechanismen wurden mehrere Simulationsdurchläufe durchgeführt. Diese nutzen sowohl einzelne als auch eine Kombination aus mehreren TSN-Mechanismen. Im Folgenden soll auf die Durchführung der Simulationsdurchläufe zur Evaluierung der TSN-Funktionen eingegangen werden. Die Anforderungen, die für die Evaluierung herangezogen wurden, sind der IEC/IEEE 60802 [II24] zu entnehmen. Sie teilen sich auf in allgemeine Anforderungen an eine IA-Station wie die unterstützten Bandbreiten, Anforderungen an die einzelnen TSN-Mechanismen und Anforderungen, die sich speziell an eine der beiden im Kapitel 2.2 vorgestellten Konformitätsklassen richten.

Die Zeitsynchronisation nach IEEE 802.1AS-2020 [Ie20] konnte mit Einschränkungen IEC/IEEE 60802-konform evaluiert werden. Möglich ist die Konfiguration von zwei Zeitsynchronisationsdomänen mit den geforderten Sendeintervallen von jeweils 125 ms für die Synchronisation und die Messung zur Leitungslängenkompensation. Teilweise umgesetzt ist die Funktionalität, die im Standard IEEE 802.1AS-2020 [Ie20] definiert wurde. Es konnte beobachtet werden, dass der Fokus der Implementierung des Simulationsmodells für die Zeitsynchronisation auf der Umsetzung der Grundfunktionalität gelegen hat. Funktionen wie der *Best timeTransmitter Clock Algorithm* (BTCA), Announce- oder Signaling-Messages und der für die Synchronisation von mehreren Zeitsynchronisationsdomänen notwendige Dienst *Common Mean Link Delay Service* (CMLDS) sind nicht umgesetzt worden. Die Festlegung von Synchronisationspfaden erfolgt somit anhand von statisch festgelegten Pfaden durch das Netzwerk innerhalb der Parametrierung. [Ie24]

Die Funktionalität TAS nach IEEE 802.1Q [Ie22] (IEEE 802.1Qbv) konnte vollständig IEC/IEEE 60802 konform evaluiert werden. Es war möglich, eine Konfiguration des Mechanismus zu realisieren, der die Trennung von Echtzeitkommunikation, Netzwerk-Management Datenverkehr und Best-Effort-Datenverkehr ermöglichte. Zudem war es möglich, innerhalb der Simulation die Funktionalität von TAS parallel zur Zeitsynchronisation zu nutzen.

Das Ergebnis der Evaluierung der TSN-Funktion Frame Preemption nach IEEE 802.1Q[Ie22] (IEEE 802.1Qbu) und IEEE 802.3br [Ie16] zeigte, dass eine IEC/IEEE 60802-konforme Konfiguration des Simulationsmodells nicht möglich war. Dies zeigte sich in der Verarbeitung von Ethernet-Frames innerhalb der Bridge-Komponenten. Hier konnten die empfangenen Ethernet-Frames nicht an die entsprechenden Übertragungspfade für preemptable- oder express-Frames innerhalb der Medienzugriffsschicht zugeordnet werden. Sämtlicher eingehender Datenverkehr wurde dem express-Pfad zugeordnet (Abbildung 4), was dazu führte, dass Best-Effort Daten den Echtzeitdatenverkehr verzögerte. Die Ursache hierfür konnte dahin gehend identifiziert werden, dass die Information, die für die Unterscheidung der Daten notwendig ist, beim Verlassen des Senders entfernt wird. Die Kombination von Frame Preemption mit anderen TSN-Mechanismen konnte nicht durchgeführt werden, da die Implementierung derzeit inkompatibel zu anderen Implementierungen von TSN-Mechanismen ist.

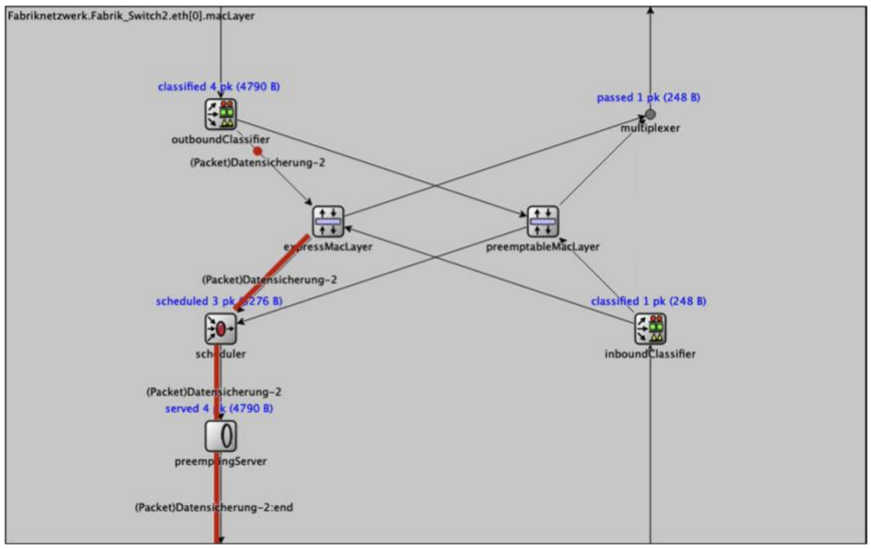


Abbildung 4: Übermittlung von Best-Effort Daten über den Express-Pfad in der MAC-Schicht

Als nächster TSN-Mechanismus wurde Frame Replication and Elimination for Reliability nach IEEE 802.1CB [Ie17a] evaluiert. Hier konnte ein konformes Verhalten zu den in der IEC/IEEE 60802 aufgeführten Anforderungen ermittelt werden. Die durch die

Konfiguration vorgegebenen Pfade wurden innerhalb der Topologie des Simulationsmodells durchlaufen. Die Erzeugung und Zusammenführung von *Member-Streams*, die auf den redundanten Pfaden übertragen werden, konnte, wie konfiguriert, analysiert werden. Das R-Tag, das für die Übertragung genutzt wird, wird innerhalb der Ethernet-Frames standardkonform eingesetzt. Eine parallele Nutzung mit anderen TSN-Standards konnte nicht evaluiert werden, da die derzeitige Implementierung nicht vorsieht, Ethernet-Frames, ohne das Vorhandensein eines R-Tags zu übertragen, sobald Frame Replication and Elimination for Reliability aktiv ist.

Zuletzt wurde die Konfiguration und Funktionalität des TSN-Mechanismus Per-Stream-Filtering and Policing nach IEEE 802.1Qci [Ie17b] auf Konformität zur IEC/IEEE 60802 geprüft. Hierfür wurde eine sich in Abhängigkeit der Simulationszeit veränderliche Kommunikationslast erzeugt, die der Verkehrsklasse Best-Effort zugeordnet wurde. Dieser Verkehrsklasse wird eine nutzbare Datenrate von 800 Mbit/s durch die Parametrierung zugesichert. Wird diese Grenze überschritten, wird ein für diese Situation vorgehaltener Vorrat an Tokens sukzessive verbraucht. Dieser Vorrat wird genutzt, um einen Puffer für vorübergehende Überlastsituationen bereitzustellen. Ist der Vorrat an Token aufgebraucht, werden die Pakete verworfen (Abbildung 5, eingekreister Bereich).

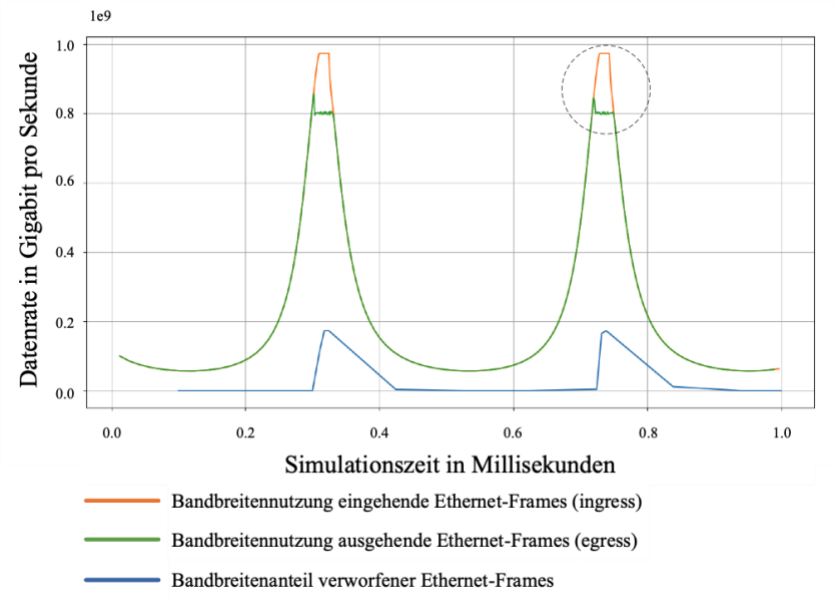


Abbildung 5: Limitierung von Best-Effort Datenverkehr durch Per-Stream Filtering and Policing

Somit wird erreicht, dass für die Übertragung von echtzeitkritischen Daten eine Bandbreite von bis zu 200 Mbit/s dauerhaft zur Verfügung steht. Fällt die genutzte Datenrate des Best-Effort Datenverkehrs wieder unter die Grenze von 800 Mbit/s, wird

der Vorrat an Token erneut aufgebaut und kann für die nächste Grenzüberschreitung als Puffer genutzt werden.

5 Fazit und Ausblick

In dieser Arbeit wurde untersucht, ob es möglich ist, IEC/IEEE 60802-konforme Simulationsmodelle mit der Simulationsumgebung OMNeT++ in Kombination mit dem Framework INET zu erstellen. Hierzu wurde ein Simulationsmodell implementiert, dessen Topologie sich an der Struktur orientiert, die in der IEC/IEEE 60802 für TSN-Netzwerke aufgeführt wird. Innerhalb dieser Topologie wurden die TSN-Mechanismen für die Zeitsynchronisation, Time-Aware Shaping, Ethernet Frame Preemption, Frame Replication and Elimination for Reliabilty und Per-Stream Filtering und Policing genutzt. Im Rahmen der Evaluierung wurde sowohl die Konfiguration als auch die korrekte Umsetzung der Funktionalität der TSN-Mechanismen analysiert. Das Ergebnis der Evaluierung ist, dass die Konfiguration und Nutzung der TSN-Mechanismen, so wie durch die IEC/IEEE 60802 gefordert, nur teilweise möglich ist. Es wurden Implementierungslücken im Bereich der Konfigurationsmöglichkeiten und der Funktionalität identifiziert werden. Die parallele Nutzung mehrerer TSN-Mechanismen konnte nur eingeschränkt genutzt werden.

Um die Simulation von IEC/IEEE 60802-konformen TSN-Netzwerken zukünftig durchführen zu können, müssen Arbeiten im Bereich der Erweiterung des Funktionsumfangs einzelner TSN-Module, der Interoperabilität der TSN-Standards im Simulationsmodell und der Möglichkeit der Konfiguration über NETCONF und YANG erfolgen.

Literaturverzeichnis

[ABJ23] Janis Albrecht, Alexander Biendarra. Jürgen Jasperneite. Increasing Ethernet TSN Multi-Protocol Interoperability by Algorithmic Configuration Merge, IEEE 21st Interational Conference on Industrial Informatics (INDIN), 2023

[BW20] Marvin Büchner, Sebastian Wolf. Plug & Work with OPC UA at the Field Level, Kommunikation in der Automation – KommA, 2020

[FSJ23] Tobias Ferfers, Sebastian Schriegel, Jürgen Jasperneite. Automated Root Cause Analysis in Time-Sensitive Networks based on Fault Models, International IEEE Symposium on Precision Clock Synchronization for Measurement, Control and Communication (ISPCS), 2023

[Ie16] IEEE Standard for Ethernet – Amendment 5: Specification and Management Parameters for Interspersing Express Traffic, IEEE 802.3br-2016). Institute for Electrical and Electronics Engineers (IEEE), 2016

[Ie17a] IEEE Standard for Local and metropolitan area networks – Frame Replication and Elimination for Reliability, IEEE 802.1CB-2017). Institute for Electrical and Electronics Engineers (IEEE), 2017

[Ie17b] IEEE Standard for Local and metropolitan area networks- Bridges and Bridged Networks – Amendment 28: Per-Stream Filtering and Policing, IEEE 802.1Qci-2017). Institute for Electrical and Electronics Engineers (IEEE), 2017

[Ie18] IEEE Standard for Local and Metropolitan Area Networks – Bridges and Bridged Networks, Amendment 31: Stream Reservation Protocol (SRP) Enhancements and Performance Improvements, IEEE 802.1Qcc-2018). Institute for Electrical and Electronics Engineers (IEEE), 2018

[Ie20] IEEE Standard for Local and Metropolitan Area Networks – Timing and Synchronization for Time-Sensitive Applications, IEEE 802.1AS-2020. Institute for Electrical and Electronics Engineers (IEEE), 2020

[Ie22] IEEE Standard for Local and Metropolitan Area Networks – Bridges and Bridged Networks, IEEE 802.1Q-2022. Institute for Electrical and Electronics Engineers (IEEE), 2022

[Ie24] IEEE Standard for Local and Metropolitan Area Networks – Timing and Synchronization for Time-Sensitive Applications, Amendment 1: Inclusive Terminology, IEEE 802.1ASdr-2024. Institute for Electrical and Electronics Engineers (IEEE), 2022

[II18] IEC/IEEE 60802 Use Cases IEC/IEEE 60802 V 1.3. International Electrotechnical Commission (IEC) and Institute for Electrical and Electronics Engineers (IEEE), 2018

[II24] IEC/IEEE 60802 Time-Sensitive Networking Profile for Industrial Automation (Draft-Version 2.4). International Electrotechnical Commission (IEC) and Institute for Electrical and Electronics Engineers (IEEE), 2024

[LGAS20] Gunnar Leßmann, Sergej Gamper, Janis Albrecht, Sebastian Schriegel. Skalierbarkeit von PROFINET over TSN für ressourcenbeschränkte Systeme. Kommunikation in der Automation – KommA (2020)

[Me13] MEF Technical Specification MEF 10.3, Ethernet Service Attributes Phase 3. The Metro Ethernet Forum (MEF), 2013

[Om24] OMNeT++ Archieved News, Time-Sensitive Networking (TSN) Introduced to INET, https://omnetpp.org/software/2022/06/15/tsn-released.html, Stand: 05.03.2024

[Rj91] Ray Jain, The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling. Wiley, 1991

[Sc22] Schriegel, S.: Kompatibilitätsverfahren für Profinet-Hardware mit Ethernet Time Sensitive Networking, Springer Vieweg, 2022

[Ts24] Time-Sensitive Networking (TSN) Task Group, https://1.ieee802.org/tsn/, 2024

[VH08] Varga A.; Hornig, R.: An Overview of the OMNeT++ Simulation Environment. In SIMUTOOLS 2008, 2008

Automatische Bewertung und Überwachung von Safety- & Security-Eigenschaften: Konzept, Informationsmodelle und Herausforderungen

Philip Kleen [1], Marco Ehrlich [2] und Sebastian Schriegel [3]

Abstract: Eine zukünftig wandelbare dynamische Produktion für kurzläufige und individuelle Produkte wird durch zunehmende Vernetzung, adaptive Maschinen und datengetriebene Optimierung realisiert. In diesem Beitrag wird eine automatisierte Überprüfung für Safety- und Security-Eigenschaften vorgestellt, an einer realen Produktion validiert und Handlungsbedarfe wie auch Mehrwerte abgeleitet. Jede Maschine bzw. Produktionseinheit wird mit Gefährdungen, Schnittstellen, Safety-Funktionen und Gegenmaßnahmen beschrieben. Die Komponenten werden durch interne Safety- und Security-relevante Eigenschaften und externe Formalisierungen aus den Community semantisch beschrieben. Es wird deutlich, dass es sich bei der Safety- und Security-Risikobewertung um einen datengetriebenen Prozess handelt, dem mit einem regelbasierten Algorithmus bzw. einem erhöhten Automatisierungsgrad gegenüber heutigem Vorgehen begegnet werden kann.

Keywords: Industrielle Automation, Modulare Produktion, Wandelbare Produktion, Safety, Security, Maschinensicherheit, Digitaler Zwilling, Asset Administration Shell (AAS)

1 Einleitung und Motivation

Die vergangenen Jahre haben gezeigt, dass der Bedarf an flexiblen und intelligenten Lösungen für die Maschinensicherheit steigt. Dies kann an Projekten und Ideen belegt werden, wie der ARENA 2036 [HA20], Manufacturing-X (Factory-X) [Sa23], das Fortschreiten der Industrie 4.0 und der Matrixproduktion [ac22]. Diese Ideen haben eins gemeinsam, eine wandelbare dynamische Produktion für kurzläufige und individuelle Produkte durch Vernetzung, adaptive Maschinen und datengetriebene Optimierung zu realisieren. Die derzeitigen Lösungen im Bereich Safety (funktionale Sicherheit) und Security (IT-Sicherheit) unterstützen derartige Produktionskonzepte nur bedingt, da diese auf Regularien und bewährten Technologien aufbauen.

Die heutige Durchführung von Analysen und Bewertungen der Maschinensicherheit werden manuell auf Basis von vielfältigen Informationen, welche meistens in Papierform zur Verfügung stehen, einmalig bei der Konstruktion der Maschine durchgeführt. Die Maschinensicherheit besteht allerdings zunehmend aus Aspekten der Safety und der Security,

1 Fraunhofer IOSB-INA, Digitale Infrastruktur, Maschinensicherheit, Campusallee 1, 32657 Lemgo, Deutschland, philip.kleen@iosb-ina.fraunhofer.de, https://orcid.org/0009-0004-4373-7544
2 Technische Hochschule Ostwestfalen-Lippe, inIT - Institut für industrielle Informationstechnik, Campusallee 6, 32657 Lemgo, Deutschland, marco.ehrlich@th-owl.de, https://orcid.org/0000-0003-1538-0547
3 Fraunhofer IOSB-INA, Digitale Infrastruktur, Campusallee 1, 32657 Lemgo, Deutschland, sebastian.schriegel@iosb-ina.fraunhofer.de, https://orcid.org/0000-0003-4753-6403

welche in der neuen EU-Maschinenverordnung (2023/1230) und dem Cyber Resilience Act (CRA, EU-Richtlinie 2022/2555) deutlich wird. Die beiden Domänen sind zum Teil von einander abhängig zum anderen unterscheiden sie sich durch unterschiedliche Stakeholder, Prozesse und Zeitpunkte bei der Durchführung.

Bereits heute ermöglichen Technologien die schnelle und automatische (Re-) Konfiguration und Anpassung von Maschinen auf veränderte Produktionsanforderungen. Bei einer Veränderung der Maschine ist aus Sicht der funktionalen Sicherheit immer zunächst zu klären, ob es sich um eine wesentliche Veränderung handelt. Im Bereich der IT-Sicherheit ist erneut eine Bedrohungsanalyse durchzuführen. Beide Analysen werden heute manuell von knappen Fachpersonal durchgeführt und verhindern einen wirtschaftlichen Betrieb von dynamischen Produktionsumgebungen mit beispielsweise Plug-and-Produce.

Im it's OWL Innovationsprojekt[4] wurden die Prozesse der Maschinensicherheit und zutreffende Normen analysiert und ein Konzept erarbeitet, welches die Safety- und Security-Eigenschaften des Produktionssystem automatisiert überwacht und bewertet. Mithilfe der Idee des Digitalen Zwillings und der Verwaltungsschale (VWS, engl. Asset Administration Shell (AAS)), als Ansatz für die Realisierung, wurde erarbeitet mit welchen Eigenschaften eine Maschine zur Überprüfung der funktionalen Sicherheit und Komponenten zur Analyse der IT-Sicherheit beschrieben werden müssen.

2 Problemstellung

Mit der Weiterentwicklung von Industrie 4.0 und den Möglichkeiten durch KI-basierte Optimierung und Rekonfiguration von industriellen Anlagen entsteht zunehmend die Möglichkeit modulare und dynamische Produktionssysteme einzusetzen. Dieser neue Art von Produkation zeichnet sich durch eine Effizienzsteigerung aus, indem mit steigender Flexibilität Rüstzeiten verkürzt und automatische Prozessoptimierung bzw. Rekonfiguration erreicht werden können. Außerdem ermöglicht diese kleinteilig organisierte Produktion eine Zusammenstellung von modularen Produktionssystemen basierend auf den speziellen Arbeitsschritten bzw. Fähigkeiten.

Ändert sich das Produktionssystem, bspw. durch neue Kombinationen von Maschinen, Produktionsparametern oder durch Self-X-Technologien, ist zu prüfen, ob diese Veränderungen bereits in der Risikobeurteilung vor der Inbetriebnahme berücksichtigt worden sind. So muss nach jeder (wesentlichen) Änderung auch immer geprüft werden, ob die grundsätzlichen Sicherheitsanforderungen der EU-Maschinenverordnung 2023/1230 gewährleistet werden. Dazu ist meistens eine erneute Identifizierung von Risiken (Gefährdungs- und Bedrohungsanalyse) erforderlich. Die Herausforderung ist die Evaluierung aller Sicherheitsfunktionen bzw. Schutzziele vollständig sicherzustellen, die sich aus der Variation von modularen und dynamischen Produktionsteilnehmern und -ablauf ergeben. Daher bedarf es einer Lösung,

4 https://its-owl.de/autos2

die mit den gesetzlichen Rahmenbedingungen einhergeht und die möglichen Auswirkungen von mangelnden Schutzeinrichtungen (Safety & Security) miteinbezieht, um die genannten Vorteile sicher umsetzen zu können [CG17].

Für die Analysen im Fall einer wesentlichen Änderung wird meistens ein interdisziplinäres Team von Safety-Experten benötigt. Diese führen die Analysen manuell mit Informationen aus papierbasierten Betriebsanleitungen, Datenblättern und Vorschriften wie auch mit dem eigenem Wissen und eigener Erfahrung durch. Dies erfordert einen hohen Aufwand, was zu einer klaren Diskrepanz zwischen der Dynamik von modularen Produktionssystemen und der notwendigen Sicherheit führt. Heutige standardisierte Prozesse sind manuell, statisch und benötigen domänenspezifisches Wissen. Außerdem müssen sie in einer konsistenten und zyklischen Weise durchgeführt werden, was der allgemeinen Anforderung der Anlagenverfügbarkeit für eine hohe Produktivität widerspricht. Die Herausforderungen im Bereich der Security sind die Adaption der vorhandenen Methoden zur Bedrohungsanalyse für die automatische Verarbeitung, die Abdeckung aller Lebenszyklusphasen (wie z.B. im Referenzarchitekturmodell Industrie 4.0 (RAMI4.0) definiert [eV20]) der modularen Produktionssysteme und eine erhöhte Automatisierung durch Software-basierte Werkzeuge [EEW22]. Die Behandlung von Software war auch schon immer eine Herausforderung im Bereich Safety, durch die Integration von Security aber wird diese noch verstärkt. Risikobeurteilungen müssen aktuell weiterhin manuell durchgeführt werden. Durch den Einsatz von Self-X-Technologien ist eine dynamisch automatisierte Lösung zur erneuten Bewertung der Sicherheit wünschenswert [Ri18].

3 Stand der Technik und Wissenschaft

Das grundsätzliche Vorgehen in der Maschinensicherheit hat sich auch mit der neuen EU-Maschinenverordnung nicht verändert und ist bereits in der Veröffentlichung auf der KommA 2020 im Stand der Technik vorgestellt worden [Eh22]. Im weiteren wird daher auf Themen eingegangen, die sich in den vergangen vier Jahren verändert oder neu entwickelt haben und im Zusammenhang mit dem Lösungskonzept und seiner Implementierung für die automatisierte Sicherheitsbetrachtungen (Safety & Security) relevant sind.

3.1 Safety

Die Maschinensicherheit bei Veränderungen im Produktionssystem automatisiert zu bewerten ist bereits mit verschieden Konzepten betrachtet und in der Lösungsfindung berücksichtigt worden.

Liggesmeyer u.a. stellten 2014 einen Ansatz mit Sicherheitsgarantien und -anforderungen für die Bewertung von Gefahren vor [LT14]. Die Idee war, dass jede Produktionseinheit (Modul, Maschine) sich durch Garantien und Anforderungen beschreibt. Ein Algorithmus

prüft, ob alle Anforderungen mit Garantien erfüllt werden. Ist dies der Fall ist das System sicher.

In mehreren Whitepapern der SmartFactory-KL wurden Ansätze mit Fehlerbäumen, Zertifikaten und Agentensystemen vorgestellt [Bu19; Mo20; Po18]. Parallel zu dieser Arbeit wurde ebenfalls ein Lösungskonzept mit dem Digitalen Zwilling und einem Knowledge Graph vorgestellt [Pf22]. Jedoch ist die Herangehensweise eine andere, als die hier entwickelte und implementierte.

Weiter gibt es Ansätze mit dem Modular Type Package (MTP) in der Prozesstechnik [Kl19; Kl20; PU17], jedoch weichen dort auch die Betrachtungen und Anforderungen von denen der diskreten Fertigung ab.

Gegenüber der vorausgegangen Veröffentlichung auf der KommA 2020 [Eh22] wurde eine neue EU-Maschinenverordnung 2023/1230 (MVO) verabschiedet, welche 2027 in Kraft tritt. Diese stärkt die Betrachtung von IT-Sicherheit im Zusammenhang mit Safety. Neben einer neuen Gliederung wird es auch ermöglicht, zukünftig Software als Safety-Bauteil in den Verkehr zu bringen. Dies ermöglicht grundsätzlich Ideen, wie das hier entwickelte Lösungskonzept. Dies ist jedoch zu relativieren, da Maschinen mit Sicherheitsbauteilen „*die sich vollständig oder teilweise selbstentwickelndem Verhalten unter Verwendung von Ansätzen des maschinellen Lernens, die Sicherheitsfunktionen gewährleisten*“ (kurz: KI-Ansatz) mit einem besonderen CE-Verfahren durch eine notifizierte Stelle geprüft werden müssen. In der gesamten Sicherheitsargumentation für dieses Lösungskonzept ist zu berücksichtigen, dass sich die Definition der (un-)vollständigen Maschine „*oder dazugehörige Produkte*“ und die Definition der „wesentlichen Änderung“ gegenüber der EU-Maschinenrichtline 2006/EG/42 verändert hat. Grundsätzlich berücksichtigt jedoch die neue MVO modulare und dynamische Produktionssysteme.

Ein weiterer besonderer Aspekt für das Lösungskonzept ist der Einsatz von Algorithmen (KI) bei der Analyse und Realisierung von funktional sicheren Überwachungsfunktionen (funktionale Sicherheit). Hier ist die Forschung und Standardisierung noch in der Entwicklung und Abstimmung. Bereits als technischen Report veröffentlicht ist die ISO/IEC TR 5469:2024 mit dem Titel: „*Artificial intelligence - Functional safety and AI systems*“. Besonders interessant ist hier in Abschnitt 6 die Tabelle 1 mit der Klassifizierung von KI im Zusammenhang mit der Nutzung. Weiter ist der Entwurf der technischen Spezifikation ISO/IEC TS 22440-1 mit dem Titel: „*Artificial intelligence - Functional safety and AI systems*“ des technischen Committee ISO/IEC JTC 1/SC 42 zu nennen. Durch die EU wurden etwas zur KI-Haftung und der AI-Act als „*Verordnung 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz*“ veröffentlicht.

Weiter wurde mit dem technischen Spezifikation IEC TS 63047:2023 mit dem Titel: „*Safety of machinery - Security aspects related to functional safety of safety-related control systems*“ das Kopplung von Safety und Security weiter betrachtet.

3.2 Security

Cyberkriminalität ist derzeit ein Wachstumsmarkt, dies ist schon fast täglich in den Medien oder in der steigenden Anzahl von Dienstleistungsanbietern in der Branche und Berichten der EU Agency for Cybersecurity, Lageberichten des Bundesamtes für Sicherheit in der Informationstechnik (BSI) und dem Lagebild des Bundeskriminalamt zu entnehmen [AET24; LSG23].

Eine weitere Herausforderung ist auch in dieser Branche der demografische Wandel und gleichzeitig der zunehmende Bedarf an Fachpersonal. Dies lässt die Nachfrage an vollautomatisierte IT/OT-Lösungen steigen [AET24].

Das Zusammenwirken und die Abhängigkeit von Security und Safety gewinnt durch die zunehmenden vernetzten Lösungen an Bedeutung. Im Zusammenhang mit der Maschinensicherheit wurde bisher nur die funktionale Sicherheit (Safety) verstanden. Durch die Vernetzung sind auch absichtliche Manipulationen von Safety-Maßnahmen möglich. Weiter ist auch die Software für die Safety-Komponenten zunehmend mit Security-Aspekten zu entwickeln. Zunehmend wird betrachtet, wie beide Safety- und Securityanalyseprozesse besser vereint und Synergien für eine gemeinsame Risikoprävention zu nutzen, z.B. gemeinsame Datenquellen und Informationsmodelle [AET24; Eh21].

Aus dem Cyber Resilience Act (CRA) und der Richtlinie über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der gesamten europäischen Union (NIS-2 Richtlinie) entstehen neue Herausforderungen, die in [AET24] beschrieben werden. Mit dem CRA soll ein besserer Schutz für Verbraucher und Unternehmen bezüglich Produkten entstehen. Weiter entsteht durch den CRA harmonisierte Vorschriften und mehr Transparenz hinsichtlich der Sicherheit bei Hard- und Software. Die regulatorischen Vorgaben für Betreiber gehen aus der NIS-2 Richtlinie hervor.

Der Security-Teil im hier präsentierten Lösungskonzept wurde bereits in verschieden Veröffentlichungen in Teilen präsentiert und diskutiert. Nachfolgend werden die Veröffentlichungen genannt, auf die das hier in Kapitel 5.2 vorgestellte Lösungskonzept beruht.

- Evaluation Concept for Prototypical Implementation towards Automated Security Risk Assessments [Eh23b]
- Evolution der IT/OT-Security durch modulare Anlagenkonzepte [AET24]
- Alignment of Safety and Security Risk Assessments for Modular Production Systems [Eh21]

3.3 Digitaler Zwilling

Die ursprüngliche Idee war es, eine integrierte Simulation eines realen Systems zu haben, um die Eigenschaften und das Verhalten in der digitalen Welt widerzuspiegeln und so die

Anwendung von simulationsbasiertem System Engineering zu ermöglichen. Die Verwirklichung des Digitalen Zwillings wurde ursprünglich etwa für das Jahr 2027 prognostiziert, aber wenn man sich das Thema und die aktuellen Entwicklungen jetzt ansieht, befinden wir uns mitten in den ersten stabilen Implementierungen des Digitalen Zwillings und der praktischen Anwendung. Daher gibt es derzeit verschiedene konzeptionelle und technologische Ansätze, die darauf abzielen, das Konzept des Digitalen Zwillings zu erfüllen.

Der erste Ansatz für eine industrielle Implementierung eines Digitalen Zwillings ist die Verwaltungsschale. Im Jahr 2016 startete die Plattform Industrie 4.0 aus Deutschland die I4.0 Initiative, RAMI 4.0 und AAS-Aktivitäten als Technologienetzwerk, bestehend aus Akteuren aus Unternehmen, Organisationen, Wissenschaft und Politik. Seit 2021 hat die Industrial Digital Twin Association (IDTA) mit Sitz in Deutschland die AAS-Entwicklungen als verantwortliche Instanz übernommen. Zu den großen Vorteilen gehören standardisierte Informationsmodelle auf der Basis von Teilmodellen für z.B. Typenschilder, Kontaktinformationen oder hierarchische Strukturen, die eine Stückliste ermöglichen, einheitliche Datenzugriffsmechanismen in einer organisierenden und aggregierenden Rolle über den gesamten Asset-Lebenszyklus.

Es folgten weitere Ansätze welche heute unter den Begriffen OPC UA der Open Platform Communications (OPC) Foundation, AutomationML (AML) vom gleichnamigen Verein und das Modular Type Package (MTP) der Profibus & Profinet International bekannt sind.

Für die interoperablen Implementierungen und die hier fokussierte automatische Überwachung von Safety- & Security-Eigenschaften werden entsprechende Informationen über das Asset benötigt. Für die Entwicklung des Lösungskonzepts wurden folgende Teilmodelle anhand von Informationen der IDTA betrachtet: Hierachical Sturctures, Handover Dokumentation, Reliability, Functional Safety, Security Engineering und das Teilmodell Vulnerability Management. Weiter wurde die Änderung der Spezifikation von V2 zu V3 berücksichtigt und umgesetzt.

3.4 Anwendungsfälle

Die nachfolgend beschriebenen Anwendungsfälle verdeutlichen zum einen die Herausforderung und dienen zum anderen dazu die Lösung zu evaluieren. Alle Anwendungsfälle haben ihren Ursprung im modularen Produktionsssystem der kundenindividuellen Produktion (CPS) der SmartFactoryOWL[5]. In einer modularen Produktionen müssen **Maschinen** zunehmend miteinander **kooperieren**, (re-)konfiguriert und neue hinzugefügt werden, um neue Produktionsanforderungen erfüllen zu können. Bei dieser Veränderung können sich Gefährdungen verändern oder neue entstehen. Diese Punkte sind zu prüfen, bevor die Produktion erneut gestartet wird. Ein weiterer Punkt ist, dass durch die Veränderung neue Komponenten und Dienste in das Netzwerk eingebracht werden und sich auch die

5 https://smartfactory-owl.de

Netzwerktopologie verändert. Auch dieser Punkt ist mit einer Bedrohungsanalyse im Sinne der IT-Sicherheit zu überprüfen. Konkret wird dies in der SmartFactoryOWL z.B. durch eine Maschine mit einer Lasereinheit und einer Maschine zur Absaugung von Dämpfen umgesetzt.
In den Lebenszyklusphasen Betrieb und Instandhaltung kommt es vor, dass **Komponenten** einen Defekt haben und ausgetauscht werden müssen. Der **Austausch** erfolgt aus eigener Erfahrung meist durch möglichst gleiche Bauteile. Gerade bei Safety-Komponenten, welche eine Sicherheitsfunktion gewährleisten müssen, ist viel Expertise notwendig, um Komponenten durch nicht exakt gleiche auszutauschen. Dabei ist sicherzustellen, dass die notwendige Architektur und Performance Level (PL) erfüllt werden. Häufig ist in Standard- als auch Sicherheitskomponenten Software enthalten. Es ist zu überprüfen, dass diese nicht manipuliert ist und den aktuellsten bzw. dem verbauten Stand hat.
Im Anwendungsfall „**dynamische Veränderung**“ wird das dynamische Produktionssystem z.B. für ein neues oder verändertes Produkt oder durch Algorithmen der Produktionsprozess verändert. Eine weitere Veränderung kann auch die Aktivierung oder Integration eines neuen Dienstes zur Datenübertragung durch ein Update sein. Diese Veränderungen können während der Lebenszyklusphase „Betrieb“ auftreten und bedürfen ebenfalls einer Analyse auf veränderte Gefahren, Risiken, Gefährdungen und Bedrohungen.

4 Anforderungen

In diesem Kapitel werden die Anforderungen und Herausforderungen aus dem Regelwerk im Bezug auf modulare dynamische Produktionssysteme hergeleitet und anhand der Anwendungsfälle konkretisiert. Zur Herleitung der Anforderungen sind zunächst zwei Normen zu nennen, für Safety die ISO 12100 und für Security die IEC 62443. Beide beschreiben Prozesse und Anforderungen, welche Analysen durchzuführen sind und leiten Handlungen ab, die das Risiko im Umgang mit einer Maschine auf ein vertretbares Minimum reduzieren. So lässt sich die erste Anforderung formulieren, dass das Vorgehen vom Lösungskonzept mit den heutigen Prozessen vergleichbar sein muss. Dies hat den Vorteil, dass die Lösung schneller anerkannt wird.

Bei genauerer Analyse der heutigen Vorgehensweise wird deutlich, dass viel Wissen bei der Gefärdungsanalyse notwendig ist. Dies beginnt mit physikalischen, chemischen Zusammenhängen über spezifisches Wissen über den Fertigungsprozess, den Umgang mit derartigen Maschinen und über das zu bearbeitende Werkstück und Materialien bis hin zu konkreten Einsatzort, Vernetzung der Maschine und verbauten Komponenten. Dieses Wissen fließt heute in die Beurteilungsprozesse ein und mündet in Sicherheitskonzepten. Diese können risikomindernde Maßnahmen sein oder eine konkrete Konfiguration von Netzwerkkomponenten und Diensten im Netzwerk zur Realisierung des Zone-and-Conudits-Ansatzes aus der IEC 62443. Zur automatisierten Überprüfung der Safety- & Security-Eigenschaften ist dieses Wissen, diese Zusammenhänge oder die ergriffenen Maßnahmen nachvollziehbar in Informationsmodellen oder Algorithmen semantisch zu beschreiben.

Das zu erarbeitende Lösungskonzept ist auf den derzeitigen Stand der Informations- und Operationstechnologien aufzubauen. Dabei ist insbesondere das Konzept des Digitalen Zwilling aus dem Stand der Technik zu berücksichtigen.

Die letzte Anforderung ist, dass das Vorgehen, die Implementierung die Ausführung einer automatisierten Überwachung von Safety- & Secruity-Eigenschaften den anerkannten Sicherheitsregeln entspricht, z.B. IEC 61508, EU-Maschinenverordnung 2023/1230 oder IEC 62443.

5 Lösungskonzept und Implementierung

Nachfolgend wird das Lösungskonzept vorgestellt, welches in dem it's OWL Innovationsprojekt AutoS² über etwa drei Jahre erarbeitet und umgesetzt wurde. Das Lösungskonzept wurde mit folgender grundsätzlicher Vorgehensweise und Beantwortung der Forschungsfragen erarbeitet [Eh22]:

1. Wie können die allgemeinen Informationen gesammelt werden, die für eine automatisierte Risikobewertung für den industriellen Bereich benötigt werden? (Information Collection)
2. Wie lässt sich ein semi-formales Informationsmodell für eine automatisierte Risikobewertung formalisieren? (Information Formalisation)
3. Wie lassen sich die Entscheidungen, die bei einer typischen manuellen Risikobewertung erforderlich sind, auf informierte Weise automatisieren? (Information Usage)
4. Wie lässt sich das Informationsmodell in einen digitalen Zwilling für eine automatisierte Risikobewertung integrieren? (Information Access)

Dabei wurde die Lösung immer wieder an den vorgestellten Anwendungsfällen und deren konkrete Realisierung im modularen Produktionssystem der SmartFactoryOWL diskutiert und evaluiert, siehe Abschnitt 6. Im Folgenden wird zunächst das Lösungskonzept für Safety und anschließend das für Security erläutert.

5.1 Safety

Zur Beantwortung der ersten Frage wurden vor allem die Normen ISO 12100 und ISO 13839 berücksichtigt. Durch Gespräche mit Experten und Maschinenbauern wurde das heutige Vorgehen analysiert. In einem Swimlane-Diagramm wurde festgehalten, wie der heutige normative Entscheidungsprozess aussieht und welche Stakeholder (Komponentenhersteller, Maschinenbauer, Integrator, Betreiber, Norm) mit welchen Informationen oder Anforderungen heute Entscheidungen in der Analyse getroffen werden. Jede Swimlane represäntiert einen Stakeholder ergänzt um den Prozess und die Produktionsumgebung. Mit diesem Vorgehen konnten die heutigen Informationsquellen identifiziert werden. Diese sind vielfältig

und hängen von der jeweiligen Entscheidung ab, z.B. Beeinflussung der Produktionsumgebung durch andere Maschinen, über das zubearbeitende Werkstück, über Schnittstellen der Maschine, über mögliche Gefährdungen, über Gefahren und das Safety-Experten-Wissen von Menschen über physikalische oder chemische Zusammenhänge.

Der nächste Schritt ist die Informationen zu formalisieren und in möglichst anwendungsneutrale Informationsmodelle zu überführen. Zur Erfüllung der Anforderungen wurde darauf geachtet, dass das Informationsmodell nicht zu komplex wird. Es wurde festgelegt, welche Informationen als Eigenschaften einer Maschine und welche in einem Algorithmus hinterlegt sind. Im Algorithmus wird zum einen Wissen zu physikalischen Zusammenhängen hinterlegt zum anderen die normativen Fragen/Entscheidungen aus der unterstützen Norm.

Eine Maschine wird im Wesentlichen durch die Eigenschaften beschrieben: Fähigkeit der Maschine, Schnittstellen, Gefährdungen, Safety-Funktionen und Gegenmaßnahmen. Dabei werden diese Informationen verknüpft. Die Safety-Funktionen verweisen auf die Gefährdungen, welche diese reduziert und Gegenmaßnahmen die sie realisieren. Weiter wird diese mit dem zu erreichenden Sicherheitsniveau (Performence Level oder Sicherheitsintegritätslevel) beschrieben und verknüpft die Komponenten, mit denen die Funktion realisiert ist. Weiter sind Merkmalskataloge für die Gefahren, Gegenmaßnahmen und Fähigkeiten von Maschinen definiert worden. Diese sind vergleichbar mit den Auflistungen in der ISO 12100. Die Beschreibung der Safety-Komponenten wurde mit dem AAS-Teilmodell Safety-Devices der IDTA realisiert, erweitert um den ProofTestInterval.

Die gesammelten und semantisch angereicherten Daten werden für zwei unterschiedliche Vorgehen genutzt. Zum Einen die Berechnung der Sicherheitsfunktion, z.B. wenn eine Komponente ausgetauscht wurde und die Andere zur Analyse der Gefahren und Gegenmaßnahmen. Die Sicherheitsfunktion kann nach den Regeln der ISO 13849 und ähnlich wie in SISTEMA[6] berechnet werden. Die dafür notwendigen Informationen zur Architektur kommen aus den oben beschriebenen und in Abbildung 1 gezeigten Teilmodell *Safety*.
Der zweite Teil ist die Analyse der Gefahren und Gegenmaßnahmen. Der entwickelte Algorithmus leitet aus den Informationen und der Struktur vom Informationsmodel einen Knowledge Graph ab. Jede Maschine wird mit diesen Informationen beschrieben und der Algorithmus bringt dies durch Auslesen der einzelnen Modelle zu einem Knowledge Graph für das gesamte Produktionsystem als Wissensbasis zusammen. Über gezielte Abfragen von diesem entsteht „Wissen“ über das modulare Produktionssystem. Diese Abfragen vom Algorithmus stellen eine Art regelbasiertes Wissen dar. Dies wurde manuell aus Normen und von Experten überführt. Weiter verfügt der Algorithmus über Zusammenhänge der Physik, z.B. bei welcher Laserleistung und Material entstehen Funken, Dämpfe, Reflexionen oder erwärmt sich das Material. Mit diesen und den Informationen über das Produktionssystem kann eine Analyse vorgenommen werden.

Das Informationsmodell wurde in ein eigenes Submodell *Safety* der Verwaltungsschale

6 https://www.dguv.de/ifa/praxishilfen/praxishilfen-maschinenschutz/software-sistema/index.jsp

Abb. 1: Ausschnitt des Teilmodells *Safety* zur Beschreibung der Lasermaschine aus dem AASX Package Explorer

überführt und für zwei Maschinen (Laser & Absaugung) aus den Anwendungsfällen mit konkreten Informationen mit dem AASX Package Explorer implementiert und auf dem AASX Server ausgeführt. Abbildung 1 zeigt das eigene Teilmodell für die Lasermaschine.

5.2 Security

Wie bereits in Kapitel 3.2 beschrieben ist dieser Teil bereits schon häufiger veröffentlicht und diskutiert worden. Die Methodik und Vorgehensweise in der Entwicklung erfolgte nach der gleichen, wie sie am Anfang von diesem Kapitel 5 beschrieben ist. Gemäß der IEC 62443-3-2 und dem in [Eh23a] gezeigten Fokus und Vorgehen wurde das Lösungskonzept erstellt [Eh24b]. Mit dieser Fokussierung wurde ebenfalls der Prozess mit seinen Entscheidungen und die dafür notwendigen Informationen mit einem Swimlane-Diagramm

analysiert. Mithilfe dieser Erkenntnisse wurde ein semi-formales menschenlesbares Informationsmodell erarbeitet. Für eine einheitliche semantische Beschreibung wurden z.B. Merkmale für Angreiferfähigkeiten und -Ressourcen für die Beschreibung eines Risikos mit Auswirkung, Komplexität und resultierendem Risiko definiert [Eh24b].

Wie im Safety-Konzept stellt jede Komponente die Informationen gemäß Informationsmodell bereit und ein Algorithmus sammelt diese und lässt einen Knowledge Graph entstehen. Dieser wird vom Algorithmus genutzt, um Fragen der Risikobeurteilung zu beantworten. Die Analyse nutzt dieses Wissen und ordnet die Komponenten auf die gleiche Weise zu einem Produktionsbreich oder Maschine, zugewiesene Zone oder Nutzung für eine funktional sichere Anwendung zu. Das Informationsmodell, siehe Abbildung 2, wurde als Teilmodell *Security* in der Verwaltungsschale realisiert und mit dem AASX-Server für jede Komponente ausgeführt. Dies wurde für die genannten Anwendungsbeispiele konkret für das CPS in der SmartFactoryOWL umgesetzt. Dabei wurde auch die Integrationsumgebung, bestehend aus dem Förderband und Schaltschrank, und die bereits erwähnten Maschinen zum Lasern und Absaugen entsprechend mit den Informationsmodellen, als Teilmodell *Security* in der VWS, erweitert und der Algorithmus exemplarisch ausgeführt. Als Ergebnis wurde zur Nachvollziehbarkeit ein Attest mit der Analyse und dem Ergebnis automatisiert als PDF generiert.

Die Implementierung sowohl vom Algorithmus als auch von den Teilmodell Security wurde bereits bei GitHub als open source Repository veröffentlicht[7]. Weiter wurde dieses Konzept und das zur Realisierung notwendige Teilmodell in das Interopera-Projekt mit dem Titel „Vulnerability Management“ und bei der NAMUR als „Security Engineering“ eingebracht. Ein wichtiger Baustein, um dieses Forschungsergebnis real zu implementieren ist, dass das Teilmodell und zugehörige Merkmale standardisiert werden. Dies war auch Bestandteil der Umsetzung des it’s OWL Innovationsprojekt AutoS². Weiter ist dieser Lösungsansatz bereits in vielen Veröffentlichungen und Vorträgen vorgestellt worden [Eh23a; Eh24b; Eh24c].

5.3 Safety & Security

Die beiden Lösungskonzepte greifen zusammen, in dem durch das Safety-Teilmodell deutlich wird, welche Safety-Komponenten welche Gefährdung reduzieren. Diese Informationen fließt in die Bedrohungsanlayse der Security-Analyse ein. Dieses Ergebnis könnte wiederum in der Risikobeurteilung der Safety-Analyse berücksichtigt werden.

6 Auswertung und Handlungsempfehlungen

Das heutige Vorgehen wurde analysiert und als Grundlage für die Identifizierung von Entscheidungen und dafür notwendigen Wissen und Informationen genutzt. Weiter kann

7 https://github.com/auto-s2/security-risk-assessment

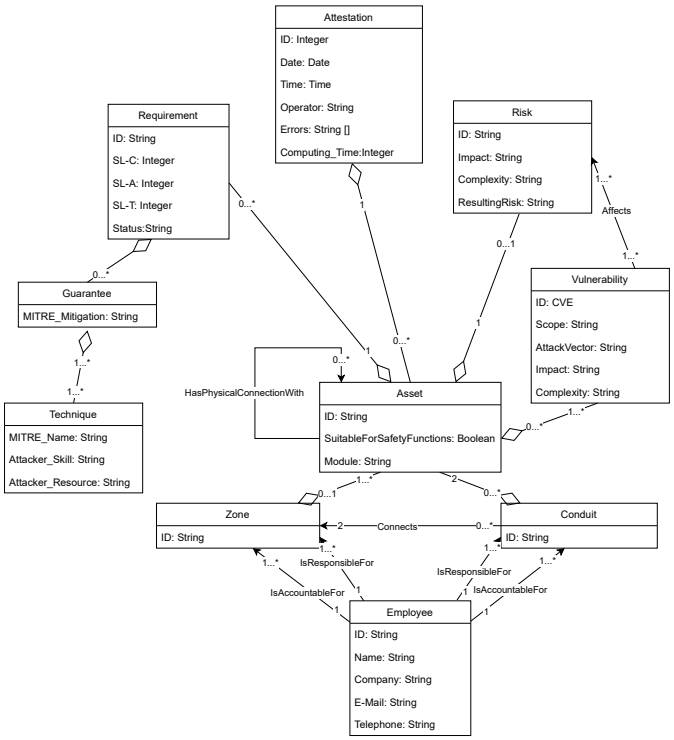


Abb. 2: Informationsmodell der Security-Eigenschaften einer Komponente

festgestellt werden, dass die Vorgehensweise bei der Implementierung und die Art und Weise wie der Algorithmus arbeitet dem heutigen Vorgehen gleicht. Dennoch entspricht die Implementierung und das gesamte automatisierte Verfahren (noch) nicht den anerkannten Regeln. Dies liegt zum einen daran, dass die Implementierung selbst nicht nach den anerkannten Methoden der Maschinensicherheit entspricht und zum anderen daran, dass diese Form der automatisierten Bewertung noch nicht anerkannt ist. Während des Projekts wurde jedoch die neue EU-Maschinenverordnung veröffentlicht, welche grundsätzlich Software als Sicherheitsbauteil in einer Sicherheitsfunktion ermöglicht. So eine Software könnte ggf. der Algorithmus aus dem Konzept sein.

Das Vorgehen und die Arbeitsweise der Umsetzung gleicht zwar den heutigen Prozessen, jedoch entspricht die Implementierung und Ausführung nicht den anerkannten sicheren Entwicklungsmethodiken der funktionalen Sicherheit. Weiter ist zu prüfen und zu klären, wie die funktional sichere Funktion der Automatisierung zur Überprüfung und Bewertung nachgewiesen werden kann bzw. nachzuweisen ist.

Das Vorgehen und Lösungskonzept hat eine mögliche Antwort auf die Frage gegeben: Wie

Security- und Safety-Eigenschaften (z.B. zur Bedrohungs- und Risikoanalyse) formalisiert werden können [Eh22]. Zwar sind die entwickelten formalen Modelle nicht mathematisch beweisbar, jedoch wurden semi-formale Modelle entwickelt. Diese definieren die notwendigen Eigenschaften einer Maschine oder Komponente und beschreiben diese mit einer gleichen Syntax. Durch die Modelle kann das Wissen über die Maschine wiederverwendet werden. Den zweite Teil bildet der Algorithmus, der hier nur exemplarisch für das Anwendungsbeispiel umgesetzt worden ist. Jedoch wurde auch damit gezeigt, wie semi-formal Wissen über physikalische und chemische Zusammenhänge wiederverwendbar gemacht werden kann. Aber auch Regeln aus Normen und Experten-Wissen in „Abfragen“ abgebildet und wiederverwendbar gemacht werden können. Die Umsetzung zeigte jedoch auch, dass dafür ausreichend Informationen über Produkt, Produktionsumgebung, Maschine und Komponente in einheitlicher Art und Weise vorliegen müssen.

Der gewählte Ansatz basiert auf dem Konzept von Digitalen Zwillingen und nutzt diese, um genügend Informationen zur Erstellung von Knowledge-Graphen zu bekommen. In der Implementierung hat sich gezeigt, dass jedoch das Konzept und die Umsetzung von Digitalen Zwillingen selbst noch in der Entstehung ist. Dies ist zum einen durch die Weiterentwicklung zu einer Version drei für AAS zu sehen. Es fehlten zunächst Werkzeuge zur Implementierung und Ausführung. Eine Abwärtskompatibilität war nicht gegeben. Zum anderen basiert es eben darauf, dass von jeden Asset eine implementierte Form des Digitalen Zwillings vorliegt. Dies war jedoch nicht für jede Komponente, die Produktionsumgebung, Maschine und das Produkt selbst der Fall. Hier musste sich mit Eigenentwicklungen für die exemplarische Umsetzung geholfen werden.

Weiter wurde eine Antwort auf die in [Eh22] hergeleiteten Forschungsfrage, *Welche Informationen müssen vom Produktionssystem, Infrastruktur und den Automatisierungskomponenten erfasst werden?*, gefunden. Das Lösungskonzept zeigt, welche Informationen von einer Maschine benötigt werden. Für die Analyse von Gefahren sind dies Schnittstellen, Sicherheitsfunktionen, Gegenmaßnahmen und Art der Maschine mit möglichen Gefahren, siehe Abbildung 1. Für die Analyse der Bedrohungslage konnte identifiziert werden, welche Informationen notwendig sind und beschrieben werden können z.B. welcher Security-Level (SL) erforderlich ist oder welche Schwachstellen bekannt sind, siehe Abbildung 2.

Das Lösungskonzept der Security-Analyse wurde bereits noch tiefgründiger analysiert und anhand einer genauen Anforderungsanalyse verifiziert [Eh24c]. Weiter wurde eine Vailidierung der grundsätzlichen Anforderungen: Sicherheitskonformität, Kopplung mit der funktionalen Sicherheit (Safety), Grad der Automatisierung, Wissenserhebung und Benutzerfreundlichkeit mit einer manuellen Vergleichsstudie durchgeführt [Eh24a].

7 Fazit und Ausblick

Es wird ein Lösungskonzept zur automatisierten Bewertung von Safety- & Security-Eigenschaften vorgestellt. Dazu wurden heutige Prozesse analysiert und Informationsmo-

delle mit den notwendigen Informationen für diese datengetriebene Analyse entwickelt. Jedoch handelt es sich bei der Implementierung um eine proprietäre Lösung. Diese ist jetzt in weiteren Gremien zu diskutieren und standardisierte Modelle und Merkmale zu entwickeln. Des Weiteren ist gerade für die funktionale Sicherheit der Algorithmus mit mehr Entscheidungen und somit Wissen weiterzuentwickeln. Dieser Aufwand bedarf jedoch gleichzeitig eine Antwort auf die Frage, wie derartige Konzepte entwickelt, getestet und ausgeführt werden müssen.

Danksagung

Dieser Beitrag wurde im Projekt „AutoS^2 - Automatische Bewertung und Überwachung von Safety & Security-Eigenschaften für Intelligente Technische Systeme“ im Rahmen des Technologie-Netzwerkes it‘s OWL mit Unterstützung des Landes Nordrhein-Westfalen gefördert.

Literaturverzeichnis

[ac22] acatech – Deutsche Akademie der Technikwissenschaften, Hrsg.: Umsetzung von cyberphysischen Matrixproduktionssystemen: Expertise des Forschungsbeirats der Plattform Industrie 4.0, 2022, URL: www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Matrixproduktion.pdf, Stand: 07. 03. 2024.

[AET24] Adamczyk, H.; Ehrlich, M.; Trsek, H.: Evolution der IT/OT-Security durch modulare Anlagenkonzepte. In (VDI Wissensforum GmbH, Hrsg.): 25. VDI-Kongress AUTOMATION: Automation 2024. VDI Verlag GmbH, Düsseldorf, 2024.

[Bu19] Burchardt, H.; Sprenger, M.; Horn, S.; Merx, J.; Schönhar, S.; Blügel, M.; Thielen, T.; Richter, D.; Varro, W.; Seidel, E.; Pfeifer, M.; Staub-Lang, P.: Smart Safety - Sicherheit in modularen Produktionsprozessen: Whitepaper SF-3.2: 04/2019, hrsg. von Technologie-Initiative SmartFactory KL e.V., Kaiserslautern, 2019, URL: https://smartfactory.de/wp-content/uploads/2019/03/Whitepaper_AG1_deutsch_042019.pdf, Stand: 17. 02. 2020.

[CG17] Carl, M.; Gondlach, K.: Sicherheit 2027: Konformitätsbewertung in einer digitalisierten und adaptiven Welt, Trendstudie des 2b AHEAD ThinkTank, 2017.

[EEW22] Eckhart, M.; Ekelhart, A.; Weippl, E.: Automated Security Risk Identification Using AutomationML-Based Engineering Data. IEEE Transactions on Dependable and Secure Computing 19 (3), S. 1655–1672, 2022, ISSN: 1545-5971, DOI: 10.1109/TDSC.2020.3033150.

[Eh21] Ehrlich, M.; Bröring, A.; Harder, D.; Auhagen-Meyer, T.; Kleen, P.; Wisniewski, L.; Trsek, H.; Jasperneite, J.: Alignment of safety and security risk assessments for modular production systems. e & i Elektrotechnik und Informationstechnik 138 (7), S. 454–461, 2021, ISSN: 0932-383X, DOI: 10.1007/s00502-021-00927-9.

[Eh22] Ehrlich, M.; Trsek, H.; Benk, S.; Harder, D.; Kleen, P.; Schriegel, S.; Jasperneite, J.: Automatische Bewertung und Überwachung von Safety & Security Eigenschaften: Strukturierung und Ausblick. In (Jasperneite, J.; Lohweg, V., Hrsg.): Kommunikation und Bildverarbeitung in der Automation. Bd. 14, Technologien für die intelligente Automation, Springer Berlin Heidelberg, Berlin, Heidelberg, S. 117–130, 2022, DOI: 10.1007/978-3-662-64283-2_9.

[Eh23a] Ehrlich, M.; Bröring, A.; Diedrich, C.; Jasperneite, J.: Towards automated risk assessments for modular manufacturing systems. at - Automatisierungstechnik 71 (6), S. 453–466, 2023, ISSN: 0178-2312, DOI: 10.1515/auto-2022-0098.

[Eh23b] Ehrlich, M.; Bröring, A.; Trsek, H.; Jasperneite, J.; Diedrich, C.: Evaluation Concept for Prototypical Implementation towards Automated Security Risk Assessments. In: 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, S. 1–4, 2023, DOI: 10.1109/ETFA54631.2023.10275455.

[Eh24a] Ehrlich, M.; Lukas, G.; Gebauer, L.; Trsek, H.; Jasperneite, J.; Kastner, W.; Diedrich, C.: Evaluation of an Automated Security Risk Assessment based on a Manual Reference. In (IEEE, Hrsg.): 29th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, 2024.

[Eh24b] Ehrlich, M.; Lukas, G.; Trsek, H.; Jasperneite, J.; Kastner, W.; Diedrich, C.: Method for Information and Process Modelling towards the Automation of Security Risk Assessments. In (IEEE, Hrsg.): 29th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, 2024.

[Eh24c] Ehrlich, M.; Lukas, G.; Trsek, H.; Jasperneite, J.; Kastner, W.; Diedrich, C.: Requirements Analysis for the Evaluation of Automated Security Risk Assessments. In (IEEE, Hrsg.): 20th International Conference on Factory Communication Systems (WFCS). IEEE, S. 1–4, 2024, DOI: 10.1109/WFCS60972.2024.10540830.

[eV20] e.V., V., Hrsg.: RAMI 4.0 und Industrie-4.0-Komponente: Mit dem von der Plattform Industrie 4.0 entwickelten „Referenzarchitekturmodell Industrie 4.0“ (RAMI 4.0) und der „Industrie-4.0-Komponente“ wurden zwei Werkzeuge geschaffen, um bestehende Standards und Technologien überschaubar zu machen. 2020, URL: http://industrie40.vdma.org/viewer/-/v2article/render/15557415, Stand: 22. 01. 2020.

[HA20] Hoßfeld, M.; Ackermann, C.: Der Forschungscampus ARENA2036. In (Bauernhansl, T.; Fechter, M.; Dietz, T., Hrsg.): Entwicklung, Aufbau und Demonstration einer wandlungsfähigen (Fahrzeug-) Forschungsproduktion. ARENA2036, Springer Berlin Heidelberg, Berlin, Heidelberg, S. 1–3, 2020, DOI: 10.1007/978-3-662-60491-5_1.

[Kl19] Klose Anselm; Bramsiepe Christian; Szmais Stefanie; Schäfer Christian; Krink Niclas; Welscher Wolfgang; Urbas Leon: Safety-Lifecycle of Modular Process Plants - Information Model and Workflow. In: 2019 4th International Conference on System Reliability and Safety (ICSRS). S. 509–517, 2019, DOI: 10.1109/ICSRS48664.2019.8987685.

[Kl20] Klose Anselm; Pelzer Florian; Barth Mike; Drath Rainer; Oehlert Rainer; León Silvia Vélez; Horch Alexander; Kotsch Christoph; Knab Jochen; Gut Bernhard; Manske Hartmut; Urbas Leon; Klose, A.; Pelzer, F.; Barth, M.; Drath, R.; Oehlert, R.; Leon, S. V.; Horch, A.; Kotsch, C.; Knab, J.; Gut, B.; Manske, H.; Urbas, L.: Distributed Functional Safety for Modular Process Plants. In (IEEE Xplore, Hrsg.): 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). Bd. 1, S. 1381–1384, 2020, DOI: 10.1109/ETFA46521.2020.9212037.

[LSG23] Lüder, A.; Steininger, H.; Goltz, D.: Quo vadis Automation? Trends für das Engineering von Automatisierungssystemen. at - Automatisierungstechnik 71 (1), S. 6–15, 2023, ISSN: 0178-2312, DOI: 10.1515/auto-2022-0102.

[LT14] Liggesmeyer, P.; Trapp, M.: Safety: Herausforderungen und Lösungsansätze. In (Bauernhansl, T.; ten Hompel, M.; Vogel-Heuser, B., Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik. Springer Fachmedien Wiesbaden, Wiesbaden, S. 433–450, 2014, DOI: 10.1007/978-3-658-04682-8{\textunderscore}21, URL: http://dx.doi.org/10.1007/978-3-658-04682-8_21.

[Mo20] Motsch, W.; David, A.; Pfeifer, M.; Harder, D.; Güntner, J.: Safety-Anforderungen an die digitale Maschinenrepräsentanz 2020: Whitepaper SF-3.3: 09/2020: smartFactoryKL, hrsg. von Technologie-Initiative SmartFactory KL e.V., Kaiserslautern, 2020, URL: https://smartfactory.de/wp-content/uploads/2020/09/SF_WhitePaper-082020_DE_WEB.pdf, Stand: 02. 09. 2024.

[Pf22] Pfeifer, M.; Harder, D.; Richter, D.; Reichenberger, K.; Neuschwander, B.; Schweiker, M.; David, A.; Rübel, P.; Heid, M.; Wagner, A.; Motsch, W.; Ruskowski, M.: Smart Safety – Das Konzept Knowledge Graph zur Umsetzung von Safety in Digitalen Zwillingen: Whitepaper SF-3.4: 04/2022: smartFactoryKL, hrsg. von Technologie-Initiative SmartFactory KL e.V., Kaiserslautern, 2022, URL: https://smartfactory.de/wp-content/uploads/2022/05/SF_Whitepaper_SmartSafety-WEB.pdf, Stand: 02. 09. 2024.

[Po18] Popper, J.; Blügel, M.; Burchardt, H.; Horn, S.; Merx, J.; Richter, D.; Varro, W.; Pfeifer, M.; Staub-Lang, P.: Safety an modularen Maschinen: Whitepaper SF-3.1: 04/2018, hrsg. von Technologie-Initiative SmartFactory KL e.V., Kaiserslautern, 2018, URL: https://smartfactory.de/wp-content/uploads/2018/04/SF_WhitePaper_Safety_3-1_DE_XS.pdf, Stand: 14. 10. 2019.

[PU17] Pfeffer Annett; Urbas Leon: HAZOP studies for engineering safe modular process plants. In (IEEE Xplore, Hrsg.): 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). S. 1–4, 2017, DOI: 10.1109/ETFA.2017.8247742.

[Ri18] Richter, D.: Sicherheit von vernetzten, modularen Industrieanlagen erfordert eine dynamische Sicherheitsarchitektur 4.0 (Safety & Security), Automatica 2018 – IT2Industry Forums, 2018, URL: http://fs-media.nmm.de/ftp/AUT/website/files/pdf/automatica_2018_presentation_DRichter_Sicherheit_von_vernetzten_Industrieanlagen.pdf, Stand: 19. 07. 2018.

[Sa23] Sauer, O.; Haller, M. L.; Wagner-Sardesai, S.; Henke, J.; Schmelting, J.; Meyer, T.; Kujath, M.; Seidel, H.; Kuhn, T.; Schnicke, F.; Harst, S.; Wenzel, K.: Manufacturing-X: Die Branche der Fabrikausrüster, 2023, DOI: 10.24406/publica-1386.

Herausforderungen in der Automatisierung von Security-Risikobeurteilungen für Operational Technology

Lisa Gebauer[1], Marco Ehrlich[1], Sebastian Wolf[2], Dimitri Harder[3], Luca Schäfer[4] und Henning Trsek[1]

Abstract: In der heutigen industriellen Praxis sind Information Technolgy (IT) und Operational Technolgy (OT) zunehmend stärker miteinander verflochten, wodurch Informationssicherheitsrisken eine Gefahr sowohl für IT- als auch für OT-Systeme darstellen. In Anbetracht der hoch dynamischen Bedrohungslandschaft im Bereich der Informations- und Datensicherheit (Security) ist der Schutz der gesamten IT/OT-Systemlandschaft eines Unternehmens oder einer Organisation daher unerlässlich. Assets müssen gegen eine kontinuierlich wachsende Anzahl an Bedrohungen geschützt werden, wofür ein ein angemessenes Security-Risikomanagement implementiert werden muss. Teil eines solchen Security-Risikomanagements ist eine effiziente und umfassende Methodologie für Security-Risikobeurteilungen. Heutzutage erfordert solche Risikobeurteilungen jedoch umfangreiche manuelle Tätigkeiten, die Verfügbarkeit von ausgebildeten Security-Experten, sowie erhebliche Ressourcen hinsichtlich Zeit und finanzieller Kosten. Zudem kann eine gleichbleibende Qualität der Ergebnisse solcher Risikobeurteilungen häufig nicht garantiert werden, da diese abhängig sind von den individuellen Kenntnissen und Erfahrungen der durchführenden Security-Experten. Ein Ansatz, diesen Problemen zu begegnen, ist es, Security-Risikobeurteilungen software-gestützt durchzuführen und Routine-Prozesse zu automatisieren. Die Nutzung von neuen Technologien aus dem Bereich der künstlichen Intelligenz (KI) und des maschinellen Lernens (ML) ist hierbei äußerst vielversprechend. Ein solcher Ansatz birgt ein großes Potenzial und zahlreiche Möglichkeiten, Security-Risikobeurteilungen effizienter und einheitlicher durchzuführen. Eine direkte Umsetzung wird jedoch aktuell noch durch verschiedene Herausforderungen erschwert. Insbesondere ein Mangel an standardisierten Daten in ausreichender Quantität und Qualität, welche für das Training von ML-Modellen benötigt werden, stellt eine Schwierigkeit dar. Diese Arbeit zeigt verschiedene Herausforderungen in der Automatisierung von Security-Risikobeurteilungen in der Industrie auf und legt dabei einen Fokus auf industrielle Komponenten und den OT-Bereich. Es werden zudem Lösungsansätze und weiterführende Forschungsrichtungen präsentiert, um die identifizierten Schwierigkeiten bei der praktischen Umsetzung zu bewältigen. Außerdem werden aktuelle Forschungsaktivitäten vorgestellt, welche die praktische Umsetzung eines Konzepts für automatisierte Security-Risikobeurteilungen für industrielle Komponenten zum Ziel haben.

Keywords: Security, Risikobeurteilungen, Operational Technology, Industrie, Automation

1 Institute Industrial IT (inIT), Technische Hochschule Ostwestfalen-Lippe, Lemgo, Germany, lisa.gebauer@th-owl.de; marco.ehrlich@th-owl.de; henning.trsek@th-owl.de
2 Weidmüller GmbH & Co KG, Detmold, Germany, sebastian.wolf@weidmueller.com
3 rt-Solutions.de GmbH, Cologne, Germany, harder@rt-solutions.de
4 Comma Soft AG, Bonn, Germany, luca.schaefer@comma-soft.com

1 Einleitung

Informationstechnologie (IT) und Betriebstechnik (OT) sind zunehmend stärker miteinander verwoben, wodurch die Zusammenarbeit verschiedener Partner unterstützt und der Zugriff auf geteilte Informationen erheblich vereinfacht wird [BS23a]. Dies birgt zahlreiche Vorteile, nicht zuletzt in der Fertigungsindustrie, in der beispielsweise Prozesse effizienter gestaltet werden können, indem benötigte Daten digital an die Produktion übertragen werden können, wodurch manueller Aufwand reduziert wird und Produktivität und Qualität verbessert werden können. Andererseits entstehen durch diese steigende Vernetzung von OT- und IT-Systemen auch neue Angriffsvektoren für Cyberkriminelle, weshalb alle Komponenten einer IT/OT-Landschaft angemessen gegen Bedrohungen geschützt werden müssen. Insbesondere in Anbetracht der hochdynamischen Bedrohungslage im Cyberraum [BS23b] ist dies eine große Herausforderung, da laufend neue Bedrohungen sowie Schwachstellen in Software-Produkten bekannt werden. Die gezielte Cyberkriminalität nimmt dabei stetig zu, es werden immer mehr kriminelle Handlungen im Cyberraum registriert und die Angreifer werden zunehmend professioneller und weltweit vernetzter [Bu20]. Dies sind besorgniserregende Entwicklungen, die Unternehmen jeder Branche und Größe betreffen und ein geeignetes Security-Risikomanagement unumgänglich machen.

Zusätzlich sind Unternehmen auch gesetzlich dazu verpflichtet, die Sicherheit ihrer IT/OT-Landschaften zu gewährleisten. Die NIS-2-Richtlinie[5] fordert beispielsweise, dass Unternehmen geeignete technische, operative und organisatorische Risikomanagementmaßnahmen im Bereich der Cybersicherheit umsetzen und angemessen auf Sicherheitsvorfälle reagieren können. Der Cybersecurity Act[6] fördert einen einheitlichen Zertifizierungsrahmen mit definierten Sicherheitsleveln für Produkte, Dienstleistungen und Prozesse aus dem Bereich der Informations- und Kommunikationstechnik (IKT). Der Cyber Resilience Act (CRA)[7] ist eine bevorstehende Regulierung, die sich mit Cybersicherheitsanforderungen für Produkte mit digitalen Elementen befasst und unter anderem Hersteller in die Pflicht nimmt, die Sicherheit von solchen Produkten bereits in der Entwicklungsphase zu berücksichtigen und über den gesamten Lebenszyklus zu verbessern.

Die geschilderten Umstände erfordern also, dass Unternehmen aktiv daran arbeiten, Security-Risiken zu identifizieren und geeignete Maßnahmen für den Schutz ihrer Assets umzusetzen. Die Implementierung eines umfassenden Security-Risikomanagements ist dabei unverzichtbar. Eine Schlüsselkomponente ist hierbei die Durchführung von Security-Risikobeurteilungen (engl. Security Risk Assessments) als Basis für die Umsetzung adäquater Maßnahmen. Diese Risikobeurteilungen müssen sowohl für die einzelnen Komponenten im Rahmen der Komponentenentwicklung durchgeführt werden, als auch für größere Systeme oder Anlagen im Rahmen des produktiven Betriebs. Dabei müssen die Risiken jedes Mal neu bewertet werden, wenn ein System oder eine Komponente geändert

5 https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32022L2555
6 https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32019R0881
7 https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex:52022PC0454

oder neu konfiguriert wird, was insbesondere durch die immer größere Popularität modularer Produktionsanlagen häufig vorkommen kann [Eh21].

Derzeit erfordert die Durchführung von Security-Risikobeurteilungen die Verfügbarkeit von ausgebildeten Security-Experten, die die Security der betreffenden Komponenten in Kooperation mit dem an der Komponentenentwicklung oder dem Systembetrieb beteiligten Personal beurteilen. Dabei kommt zumeist eine vordefinierte Methodologie zum Einsatz, um Komponenten auf Schwachstellen zu untersuchen, mögliche Bedrohungen und Risiken zu erarbeiten und geeignete Maßnahmen zu definieren. Aufgrund der Knappheit solcher Security-Experten auf dem Arbeitsmarkt ergibt sich hieraus jedoch ein nicht unerheblicher Bedarf an personellen und finanziellen Ressourcen. Insbesondere für kleine und mittlere Unternehmen (KMU) kann dies eine enorme Ressourcen-Belastung darstellen, da das benötigte Security-Fachwissen häufig nicht im eigenen Haus vorhanden ist und externe Security-Experten zu Rate gezogen werden müssen. Dieser Umstand ist Angreifern bekannt, wodurch KMU mit geringer in-house Security-Expertise besonders anfällig für Cyber-Bedrohungen sind [Fr22]. Der Lagebericht zur IT-Sicherheit in Deutschland für das Jahr 2023 [BS23b], welcher jährlich durch das Bundesamt für Sicherheit in der Informationstechnik (BSI) herausgegeben wird, unterstreicht, dass KMU immer häufiger von Cyber-Kriminellen angegriffen werden. Eine weitere Schwierigkeit ist in diesem Zusammenhang die hohe Varianz der Qualität der Ergebnisse, die stark abhängig sind von den individuellen Fachkenntnissen und Erfahrungen des durchführenden Experten. Es existieren zwar verschiedene allgemeine Normen für das Security-Risikomanagement, wie beispielsweise die ISO/IEC 27000-Reihe, die ISO 31000-Reihe oder auch die spezifischere IEC 62443, die sich mit der IT-Sicherheit von industriellen Netzen und Systemen befasst, praktische Richtlinien für die konkrete Durchführung von Risikomanagement-Prozessen fehlen in diesem Bereich jedoch noch weitgehend.

Insgesamt ist also zu erkennen, dass die dynamische Bedrohungslandschaft in Kombination mit häufigen Systemveränderungen, sowie ein signifikanter Fachkräftemangel in Kombination mit fehlenden Ressourcen in Unternehmen, dazu führen, dass Unternehmen in Hinblick auf die Durchführung von Risikobeurteilungen und die Implementierung geeigneter Gegenmaßnahmen vor nicht unerhebliche Schwierigkeiten gestellt werden. Dies betrifft insbesondere KMU, in welchen häufig kein oder wenig unternehmensinterne Security-Expertise vorhanden ist. Eine fehlende Standardisierung bei der praktischen Durchführung von Risikobeurteilungen führt zudem zu Ergebnissen von unterschiedlicher Qualität.

Die Automatisierung von Security-Risikobeurteilungen ist ein vielversprechender Ansatz, um den geschilderten Schwierigkeiten zu begegnen. Das vollautomatisierte Security-Risikomanagement steckt jedoch noch in den Kinderschuhen [HLL19]. Eine Teilautomatisierung von Routine-Prozessen, welche bei der Durchführung von Risikobeurteilungen wiederholt auftreten, ist jedoch zukunftsträchtig. Dabei ist der Einsatz von Software zur Unterstützung menschlicher Experten besonders interessant, da dieser weitgehend nahtlos in bereits existierende und erprobte Abläufe eingebunden werden kann, um so eine reibungslose Integration zu gewährleisten.

In diesem Beitrag wird ein konzeptioneller Ansatz zur software-basierten Unterstützung von Security-Risikobeurteilungen in der Industrie vorgestellt. Dieser hat zum Ziel, ausgewählte Prozesse in Security-Risikobeurteilungen mittels künstlicher Intelligenz (KI) zu automatisieren, sodass Security-Experten unterstützt und bei Routineaufgaben entlastet werden. Ein Hauptaugenmerk liegt in der hier vorliegenden Arbeit darauf, die Herausforderungen bei der praktischen Umsetzung des Ansatzes aufzuzeigen und dabei Lösungsansätze, sowie aktuelle und zukünftige Forschungsaktivitäten zu präsentieren, um so einen Leitfaden für die praktische Implementierung bereitzustellen. Dieser Beitrag fokussiert sich dabei auf Risikobeurteilungen, die in der Entwicklungsphase von industriellen Komponenten durchgeführt werden müssen. Eine Erweiterung des Ansatzes auf Risikobeurteilungen, die für den produktiven Betrieb von industriellen Systemen und Anlagen durchgeführt werden müssen, ist für einen späteren Zeitpunkt geplant.

2 Security-Risikobeurteilungen in der Entwicklung von industriellen Komponenten

Es existieren verschiedene allgemeine Normen im Security-Bereich, wie beispielsweise die ISO/IEC 27000-Serie[8], die eine Reihe von Standards beinhaltet, welche sich mit dem Management von Security beschäftigen. Der dazugehörige ISO/IEC 27005[9] Standard befasst sich dabei konkret mit dem Management von Security-Risiken. Zudem existiert mit der IEC 62443[10] eine wichtige Normenreihe für die Security im Bereich der Industrieautomatisierung, welche sich gezielt an Betreiber, Integratoren und Hersteller richtet.

Die ISO/IEC 27005 definiert ein Informationssicherheitsrisiko (engl. *information security risk*) die potenzielle Möglichkeit, dass eine Bedrohung (engl. *threat*) Schwachstellen (engl. *vulnerabilities*) eines Assets ausnutzt, sodass die Organisation, der das Asset gehört, einen Schaden erleidet. Dabei werden sowohl die Wahrscheinlichkeit, dass ein solcher Sicherheitsvorfalls eintritt, als auch die Auswirkungen im Falle des Eintretens eines solchen Vorfalls berücksichtigt. Ein allgemeiner Prozess für das Security-Risikomanagement wird in der ISO/IEC 27005 ebenfalls erläutert. Danach umfasst das Risikomanagement eine initiale Kontextfestlegung, eine Risikobeurteilung, die eine Risikoidentifizierung, -analyse und -bewertung umfasst, und schließlich eine Risikobehandlung zur Entschärfung der identifizierten Risiken. Die Bestandteile der Risikobeurteilung sind nach ISO/IEC 27005 im Folgenden kurz zusammengefasst.

Risikoidentifizierung (engl. *risk identification*) beschreibt den Prozess des Auffindens, Erkennens und Beschreibens von Risiken. Dies ist eine entscheidende Komponente im Prozess der Risikobeurteilung, da sie die Grundlage für alle nachfolgenden Risikomanagement-Aktivitäten bildet. *Risikoanalyse* (engl. *risk analysis*) beschreibt den Prozess der Bestimmung

8 https://www.iso.org/standard/iso-iec-27000-family
9 https://www.iso.org/standard/80585.html
10 https://www.dke.de/de/arbeitsfelder/industry/iec-62443-cybersecurity-industrieautomatisierung

verschiedener Arten von Risiken und des jeweiligen Risiko-Levels (engl. *risk level*). Dabei werden Ursachen und Quellen der Risiken ebenso betrachtet wie die Wahrscheinlichkeit, dass ein entsprechender Sicherheitsvorfall eintritt und die Wahrscheinlichkeit, dass dieser Auswirkungen auf die Organisation hat. Die mögliche Schwere der Auswirkungen wird ebenfalls untersucht. *Risikobewertung* (engl. *risk evaluation*) beschreibt den Prozess des Vergleichs der Ergebnisse der Risikoanalyse mit zuvor festgelegten Risikokriterien, die dazu dienen, zu bestimmen, ob ein bestimmtes Risiko akzeptabel ist oder durch geeignete Gegenmaßnahmen entschärft werden muss. Auf dieser Grundlage können den gefunden Risiken dann Prioritäten zugewiesen werden, sodass sie gemäß ihres Schweregrads behandelt werden können.

Im Hinblick auf die Security speziell im OT-Bereich stellt das US National Institute of Standards and Technology (NIST) einen Leitfaden zur Verfügung, der allgemeine Informationen zum OT-Security-Risikomanagement enthält. Unternehmen und Organisationen können diesen Empfehlungen folgen, müssen sich aber nicht unbedingt daran orientieren. Häufig werden stattdessen proprietäre Verfahren verwendet, was zu sehr unterschiedlichen Methodiken führen kann. Ein Komponentenhersteller beispielsweise muss alle relevanten Komponenten bereits in der Entwicklungsphase hinsichtlich der Security bewerten und dafür jeweils eigene Security-Risikobeurteilungen durchführen. In der Regel werden hierfür mehrere interne Workshops für jede zu bewertende Komponente organisiert, was insgesamt einen großen zeitlichen Aufwand bedeutet. An diesen Workshops nehmen zudem viele verschiedene Verantwortliche teil. Dazu zählen unter anderem Product Owner, Produktentwickler und technisches Personal, die jedoch meist nicht über das benötigte Security-Fachwissen verfügen, sowie interne oder externe Security-Experten, die wiederum meist nicht tiefgehend mit den zu bewertenden Produkten vertraut sind, weshalb hier zunächst eine Einarbeitung notwendig ist. In den Workshops werden die Komponenten üblicherweise in Hinblick auf ihre security-relevanten Schnittstellen untersucht und es wird bewertet, ob diese irgendwelchen bekannten Bedrohungen oder Bedrohungskategorien ausgesetzt sind oder sein könnten, um mögliche Security-Risiken zu ermitteln. Es wird außerdem betrachtet, ob und warum ein Angreifer motiviert sein könnte, eine solche Schnittstelle auszunutzen und diese Möglichkeit für einen Angriff zu nutzen. Um die möglichen Bedrohungen systematisch zu erfassen, werden häufig unterschiedliche Aspekte wie etwa Bedrohungstypen (engl. *threat agents*) und Bedrohungsaktionen einzeln betrachtet. Bezüglich eines möglichen Angriffs werden Auswirkungen, Exponiertheit, Ausnutzbarkeit und Wahrscheinlichkeit in der Regel qualitativ auf einer vordefinierten Skala numerisch bewertet, beispielsweise von 1 (vernachlässigbar) bis 5 (schwerwiegend). Die Einstufung gemäß dieser Skala hängt in der Regel von der Erfahrung und den individuellen Fachkenntnissen der durchführenden Security-Experten ab und ist damit nicht eindeutig. Der endgültige numerische Risikowert, der jedem gefundenen Security-Risiko zugewiesen wird, wird für gewöhnlich mathematisch mithilfe einer vordefinierten Formel bestimmt, die die numerischen Zwischenwerte für Auswirkungen, Exponiertheit, Ausnutzbarkeit und Wahrscheinlichkeit kombiniert. Unter Umständen werden hier auch Gewichtungsfaktoren einbezogen. Auf Basis dieser Bewertung können dann priorisiert Gegenmaßnahmen ermittelt werden.

Das Ergebnis eines solchen typischen manuellen Workshops zur Risikobeurteilung einer industriellen Komponente ist in der Regel eine Liste identifizierter potenzieller Bedrohungen und betroffener Schnittstellen mit den zugewiesenen Risikowerten, sowie möglichen Gegenmaßnahmen. Für gewöhnlich werden diese Prozesse und Ergebnisse in einer menschen-lesbaren und text-basierten Weise dokumentiert, beispielsweise mithilfe gängiger Office-Software und Text- oder Tabelleneditoren, was eine anschließende automatisierte Verarbeitung erschwert.

3 Verwandte Arbeiten

Die Automatisierung von Security-Risikobeurteilungen ist seit mehreren Jahren Gegenstand verschiedener Forschungsarbeiten mit unterschiedlichem Fokus. Dazu gehört beispielsweise die Verwendung von ML-Methoden zur Vorhersage von Malware-Angriffen [YB19] oder die Identifizierung von Netzwerkgeräten in Kombination mit einer automatisierten Firmware-Analyse, um aktuelle Risiken für IoT-Geräte abzuleiten [Os22]. Ein Framework für die Security-Beurteilung komplexer Active-Directory-Umgebungen unter Verwendung graphbasierter Methoden sowie ML-Verfahren wird ebenfalls vorgestellt [NC23]. Diese Ansätze fokussieren sich jedoch auf bestimmte Umgebungen und spezielle Risiken und sind nicht geeignet, das hier angestrebte Ziel der Unterstützung umfassender Risikobeurteilungen im OT-Bereich abzudecken. Weitere Ansätze konzentrieren sich explizit auf Security-Risikobeurteilungen im OT-Bereich, wobei insbesondere die integrierte Bewertung von Security und Safety, also der funktionalen Sicherheit, die den Schutz des Menschen vor der Maschine beschreibt, in den Fokus genommen wird. Für die Beurteilung der Security und Safety industrieller Kontrollsysteme (ICSs) wird beispielsweise die Verwendung von Bayesian Belief Networks (BBN) vorgeschlagen [BKS23]. Ein Threat-Modelling-Ansatz für die integrierte Beurteilung von Security und Safety wird in einer weiteren Arbeit präsentiert [HKS21]. Zudem werden Herausforderungen für die Absicherung von Cyber-Physischen Produktionssystemen (CPPS) in Hinblick auf Safety und Security aufgezeigt [Ho23]. Ein anderer Ansatz verwendet formale Wissensrepräsentationen von sicherheitsrelevanten Daten, um die technischen Daten eines Systems für eine automatische Identifizierung von Security-Risiken anzureichern und Angriffsgraphen zu generieren [EEW22]. Ein weiterer Ansatz zielt darauf ab, Security-Risikobeurteilungen für Industrie4.0-Fertigungssysteme zu automatisieren, wobei vier Hauptschritte benannt werden: die Informationssammlung, die Informationsformalisierung unter Verwendung eines semi-formalen Modells, die Informationsnutzung unter Verwendung von First Order Logic für die Extraktion von Expertenwissen und der Informationszugriff unter Verwendung digitaler Zwillinge [Eh23]. Diese Ansätze unterstreichen die Relevanz der Untersuchung von Möglichkeiten zur Automatisierung von Security-Risikobeurteilungen. Der in dieser Arbeit betrachtete Ansatz des Einsatzes von KI für die Unterstützung von Security-Risikobeurteilungen von industriellen Komponenten ist jedoch bislang weitgehend unerforscht, weshalb weitere Forschung nötig ist, um diesen Ansatz praktisch zu realisieren.

4 Design eines Software-Tools für die Unterstützung von Security-Risikobeurteilungen in der Industrie

In diesem Beitrag wird eine software-basierte Unterstützung von Security-Risikobeurteilungen [Ge24] in der Industrie als vielversprechender Ansatz vorgestellt, Routine-Aufgaben zu automatisieren und so Security-Fachpersonal zu entlasten, sodass Risikobeurteilungen insgesamt effizienter durchgeführt werden können und die Ergebnisse eine gleichbleibende Qualität erhalten. Konzeptionell besteht der Ansatz aus drei Komponenten. Zunächst muss eine geeignete Datenbasis als Grundlage für das Training eines KI-Modells vorhanden sein. Es müssen also Komponenten- und Systemdaten sowie relevante Informationen in Bezug auf Bedrohungen, Schwachstellen und mögliche Auswirkungen eines Cyber-Angriffs in ausreichender Menge und in einem maschinenlesbaren Format vorliegen, die semantisch angemessen miteinander verknüpft sind. Das KI-Modell bildet die zweite Komponente des Ansatzes und muss auf geeignete Art und Weise ausgewählt und trainiert werden, um beispielsweise mögliche Bedrohungen oder geeignete Gegenmaßnahmen für gefundene Risiken vorzuschlagen. Die dritte Komponente bildet der menschliche Nutzer, beispielsweise ein Security-Experte, der mit der Software über ein Benutzer-Interface interagiert und diese in bestehende Abläufe einbindet, also weiterhin essenziell bleibt für den Prozess der Risikobeurteilung und zu diesem Zeitpunkt nicht vollständig durch eine KI ersetzt werden kann oder soll. Abb. 1 zeigt das geplante Konzept für eine solche software-basierte Unterstützung von Security-Risikobeurteilungen.

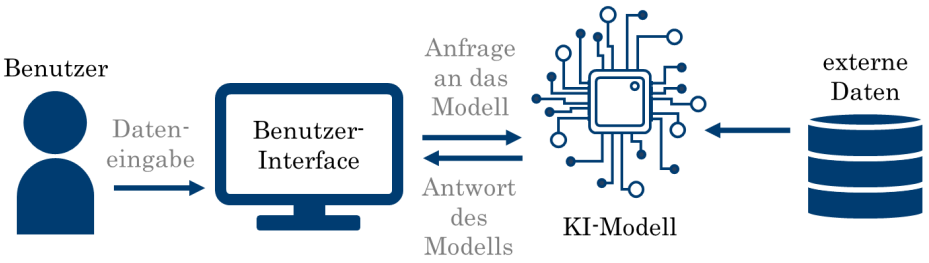


Abb. 1: Konzept für software-basierte Unterstützung von Security-Risikobeurteilungen

Es wird angenommen, dass eine Umsetzung dieses Konzepts zu einer erhöhten Effizienz bei der Durchführung von Risikobeurteilungen und damit zur Einsparung von Ressourcen beitragen kann. Für die praktische Umsetzung müssen jedoch einige Herausforderungen angegangen werden. Diese umfassen die Verfügbarkeit von System- oder Komponentendaten, die Integration von externen Daten zu Bedrohungen und Schwachstellen, die Datenaufbereitung für die maschinelle Weiterverarbeitung, eine geeignete Datenvorverarbeitung und eine angemessene Modellauswahl und -optimierung für bestmögliche Ergebnisse [Ge24].

5 Herausforderungen bei der Umsetzung von software-gestützten Security-Risikobeurteilungen in der Industrie

Im Folgenden werden die im letzten Kapitel identifizierten Herausforderungen beschrieben und es werden bestehende Lösungsansätze und aktuelle Forschungsarbeiten präsentiert, um die Umsetzbarkeit des hier vorgestellten Ansatzes zur Automatisierung von Security-Risikobeurteilungen im OT-Bereich zu zeigen. Zusätzlich wird im nächsten Kapitel ein exemplarisches Software-Mockup vorgestellt, welches das vorgestellte Konzept prototypisch implementiert.

Verfügbarkeit von Komponentendaten: Eine große allgemeine Herausforderung in der Automatisierung von Prozessen mithilfe von KI ist immer die Verfügbarkeit von geeigneten Daten, die repräsentativ genug sind, um das zugrunde liegende Problem zu beschreiben, da diese für das Training der Modelle unabdingbar sind. Im hier vorgestellten Ansatz zur Risikobeurteilung von industriellen Komponenten werden zunächst Daten benötigt, die die zu beurteilenden Komponenten möglichst genau beschreiben und dabei alle relevanten Informationen enthalten, die für die Beurteilung der Security erforderlich sind. Diese müssen in ausreichender Quantität und Qualität verfügbar sein. Die nötigen Informationen umfassen beispielsweise Schnittstellen, Datenflüsse, Hardware- und Software-Architekturen. Zusätzlich werden für das Training der KI-Modelle numerische Bewertungen benötigt hinsichtlich der Ausnutzbarkeit und Exponiertheit der Schnittstellen, der Wahrscheinlichkeit, dass eine exponierte Schnittstelle ausgenutzt wird, und der Schwere der möglichen Auswirkungen eines Angriffs. Solche Bewertungen hängen wie bereits erwähnt in der Regel von individuellen Experteneinschätzungen ab. Als Grundlage für die Generierung einer Datenbasis für den in dieser Arbeit vorgestellten Ansatz werden daher die Ergebnisse manueller Risikobeurteilungen verschiedener industrieller Komponenten zusammengetragen. Diese enthalten zum einen explizit relevante Informationen zu den Komponenten und ihren Schnittstellen und zum anderen implizit auch Informationen zur Ableitung der nötigen numerischen Bewertungen. Geeignet sind hier also vor allem reale Ergebnisse aus manuellen Workshops zur Risikobeurteilung, die von einem Komponentenhersteller zur Verfügung gestellt werden.

Integration von externen Daten: Zusätzlich zu den relevanten Komponentendaten aus den Unternehmen sollen auch Daten zu Bedrohungen und Schwachstellen aus öffentlichen Datenquellen einbezogen werden, um einerseits eine einheitliche Zuordnung und Namensgebung bezüglich Bedrohungen und Schwachstellen zu gewährleisten und andererseits rechtzeitig auf dynamische Veränderungen in der Bedrohungslandschaft reagieren zu können. In Hinblick auf Schwachstellen stehen beispielsweise die CVE[11] und NVD[12] als Datenquellen zur Verfügung. Für Informationen über Bedrohungen kann die MITRE ATT&CK[13]-Wissensdatenbank herangezogen werden, die eine Angriffsmatrix speziell für

11 https://www.cve.org
12 https://nvd.nist.gov
13 https://attack.mitre.org/matrices/ics/

industrielle Kontrollsysteme (ICS) enthält. Für die Kategorisierung von Bedrohungstypen kann die Intel Threat Agent Library (TAL) [Ca07] verwendet werden. Für die Abschätzung der Auswirkungen eines Sicherheitsvorfalls können unter Umständen öffentliche Berichte herangezogen werden, wie beispielsweise Veröffentlichungen des BSI [BS23b] oder von Verizon [Ve23]. Diese können jedoch lediglich eine Orientierungshilfe bieten, da eine Folgenabschätzung für jede Organisation sehr unterschiedlich ausfallen kann und daher individuell betrachtet werden muss. Daher werden externe Informationen zu den möglichen Auswirkungen von Angriffen in dieser Arbeit bisher nicht einbezogen, könnten aber im Zuge zukünftiger Forschungsaktivitäten berücksichtigt werden.

Datenaufbereitung: Um die genannten Daten zweckmäßig maschinell verarbeiten zu können, müssen sie zusammengeführt und in ein brauchbares Format gebracht werden. Die Datenaufbereitung in dieser Arbeit umfasst derzeit eine Standardisierung und Kategorisierung der verfügbaren Komponentendaten, um eine einheitliche Namensgebung sicherzustellen, und einen Abgleich mit externen Bedrohungsdaten. Die Ergebnisse realer manueller Risikobeurteilungen, die in Workshops bei einem Komponentenhersteller erarbeitet wurden, dienen als Grundlage für die Generierung eines Datenpools für das Training von KI-Modellen. Diese liegen in der Regel nur in tabellarischer Form in Office-Dateien vor und enthalten sowohl Freitext als auch unterschiedliche Begrifflichkeiten, die aber beispielsweise die gleichen Schnittstellen oder Bedrohungstypen beschreiben. Daher werden diese Daten in einem ersten Schritt von Fachleuten aus der Komponentenherstellung überprüft und so aufbereitet, dass sie kategorische Daten anstelle von Freitext enthalten und eine einheitliche Terminologie anstelle von unterschiedlichen Synonymen verwenden. Die Kategorien werden dabei aus der MITRE ATT&CK-Wissensdatenbank und aus der Intel TAL abgeleitet, um die Übereinstimmung mit etablierten öffentlichen Security-Datenbanken zu gewährleisten. Zur Unterstützung dieses Mapping-Prozesses wird ein großes Sprachmodell (engl. *Large Language Model* bzw. LLM) aus dem Bereich der generativen KI (engl. *Generative Artificial Intelligence* bzw. GenAI) eingesetzt, welches mit Retrieval-Augmented Generation (RAG) kombiniert wird. Das LLM kann die vorliegenden text-basierten heterogenen Daten verarbeiten. Durch die Kombination mit RAG können zudem externe Datenquellen, in diesem Fall eine Liste von Kategorien aus der ICS-Matrix der MITRE ATT&CK-Wissensdatenbank und aus der Intel TAL, eingebunden werden. Schließlich können die Daten in ein geeignetes Format gebracht werden. Welches Format im jeweiligen Anwendungsfall geeignet ist, hängt von den verwendeten KI-Modellen ab. Da in dieser Arbeit der Einsatz von LLMs anvisiert ist, wird eine text-basierte Form bevorzugt.

Datenvorverarbeitung: Damit ein KI-Modell erfolgreich auf den verfügbaren Daten arbeiten kann, ist eine geeignete Vorverarbeitung erforderlich. Dies kann beim Einsatz konventioneller Algorithmen aus dem Bereich des überwachten ML (engl. Supervised Learning) beispielsweise die Auswahl geeigneter Features für die Voraussage von bestimmten Kategorien sein. Zudem sollten Datensätze auf fehlende Daten überprüft werden und auf die relevanten Daten für den angestrebten Zweck zugeschnitten werden, um redundante Daten zu vermeiden. Es kann zudem in Betracht gezogen werden, die vorhandenen Daten

synthetisch zu erweitern, falls die Menge der zur Verfügung stehenden Daten nicht für das Modelltraining ausreicht. Synthetisch erzeugte Daten müssen dann jedoch sorgfältig evaluiert werden, insbesondere im Hinblick auf sensible Informationen, die möglicherweise sicherheitsrelevante oder geheime unternehmensinterne Daten enthalten. Im Falle des Einsatzes von LLMs, der hier bevorzugt wird, ist es ebenfalls wichtig, die Daten zu bereinigen und angemessen zu formatieren, sodass relevante Informationen in einem Format zur Verfügung stehen, das von den Modellen optimal verarbeitet werden kann, und keine überflüssigen Informationen enthalten sind.

Modellauswahl und -optimierung: Ein KI-Modell sollte unbedingt anhand der vorliegenden Datenbasis ausgewählt werden, da angenommen wird, dass die Daten, die für das Training eines Modells genutzt werden, einen größeren Einfluss auf die Qualität der Ergebnisse haben als die Auswahl eines konkreten Modells [JP20]. Dies unterstreicht die Wichtigkeit der Bewältigung der Herausforderungen, die in Bezug auf die Verfügbarkeit und Vorverarbeitung der Daten identifiziert wurden. Neben der initialen Auswahl des Modells ist auch eine Optimierung erforderlich, um eine hohe Vorhersage-Genauigkeit zu erreichen, damit die Ergebnisse wertvoll und vertrauenswürdig sind. In Bezug auf die Laufzeit sollte das Modell zudem effizient arbeiten, sowohl in Bezug auf die produktive Anwendung als auch hinsichtlich des Trainings, da die hochdynamische Natur der Bedrohungslandschaft schnelle Reaktionen auf neue Entwicklungen erfordert und eine Anpassung des Modells erforderlich machen kann. Diese und gegebenenfalls weitere Anforderungen fließen in die Wahl eines geeigneten Modells ein. Die Forschungsaktivitäten zur praktischen Umsetzung des hier vorgestellten Konzepts, die die Grundlage der hier vorliegenden Arbeit bilden, sind zum aktuellen Zeitpunkt noch nicht abgeschlossen und es wird kontinuierlich weiter an einer geeigneten Datenbasis gearbeitet. Aufgrund der Form und Menge der derzeitig zur Verfügung stehenden Daten wird hier jedoch die Verwendung von LLMs anvisiert, um den Risikobeurteilungsprozess zu unterstützen, da sich diese insbesondere bei geringen Datenmengen in unstrukturierter Form eignen. Die Einbindung weiterer Wissensquellen, wie etwa externer Daten aus öffentlichen Bedrohungsdatenbanken, mittels RAG ist ein weiterer Vorteil, welcher die Wahl von LLMs als KI-Komponente in der Umsetzung des hier vorgestellten Ansatzes motiviert. Zur Optimierung der Ergebnisse wird die relativ neue Disziplin des Prompt-Engineerings betrachtet, welches verwendet wird, um Anfragen an das LLM möglichst geschickt zu formulieren, um optimale Antworten zu erhalten.

6 Software-Prototyp

Es wird bereits an einer praktischen Umsetzung des vorgestellten Ansatzes für die softwaregestützte Risikobeurteilung von industriellen Komponenten geforscht. Dazu wird im Zuge aktueller Forschungsaktivitäten ein Software-Prototyp entwickelt, welcher web-basiert ist und die Programmiersprache Python nutzt, um eine nahtlose Integration von KI-basierten Automatisierungsansätzen und eine komfortable Bedienbarkeit zu ermöglichen. Die Software soll Komponentendaten durch Benutzereingaben erhalten und auf dieser

Basis sowohl potenzielle Bedrohungen als auch geeignete Gegenmaßnahmen vorschlagen können. Aktuell betrachtete Formen der Benutzereingabe sind Web-Formular-Eingaben und Datei-Uploads von text-basierten Dateien. Diese Eingaben werden dann im Hintergrund in einen entsprechenden Prompt umgewandelt, welcher an ein modernes LLM übergeben wird. Das LLM, welches auf einem sicheren Server läuft, verarbeitet diese Prompts unter Einbezug externer Datenquellen wie der MITRE ATT&CK ICS-Matrix und stellt eine Antwort bereit, welche dem Nutzer über die Web-Oberfläche als Text zur Verfügung gestellt wird. Zukünftige Entwicklungen sollen zudem ermöglichen, die Ergebnisse in Tabellenform, ähnlich zu den Ergebnissen manueller Risikobeurteilungen, darzustellen. Die Ergebnisse können schließlich von einem menschlichen Experten überprüft werden, wodurch gewährleistet ist, dass diese plausibel und korrekt sind. Durch den Einsatz dieses Software-Tools können langwierige Routine-Prozesse beschleunigt werden, wodurch erhebliche Zeit- und Ressourcen-Einsparungen erreicht werden können. Da das Tool Prozesse unterstützt, die derzeit bereits manuell durchgeführt werden, ist zudem eine reibungslose Integration in bestehende Arbeitsabläufe möglich.

7 Zusammenfassung und Ausblick

Security wird im Zuge der Integration von IT und OT immer wichtiger. Die software-basierte Unterstützung von Security-Risikobeurteilungen im OT-Bereich bietet dabei viele Möglichkeiten, die Effizienz bei der Durchführung der Risikobeurteilungen und die Konsistenz der Ergebnisse zu erhöhen. Fachpersonal kann so erheblich entlastet werden, wodurch Zeit und finanzielle Ressourcen eingespart werden können.

Diese Arbeit stellt ein Konzept zur software-basierten Unterstützung von Security-Risikobeurteilungen für industrielle Komponenten vor und identifiziert verschiedene Herausforderungen, die bei der praktischen Umsetzung des Ansatzes beachtet werden müssen. Diese betreffen zum einen die zugrundeliegenden Daten und zum anderen die technische Umsetzung mittels KI. Dieser Beitrag beschreibt diese Herausforderungen und Ein Mangel an strukturierten und maschinen-lesbaren Daten, die für das Training von Modellen geeignet sind, wird hier als größte Herausforderung identifiziert. Dokumentationen von Ergebnissen aus manuellen Risikobeurteilungen stehen jedoch als Grundlage für die Generierung einer geeigneten Datenbasis zur Verfügung und werden durch öffentlich verfügbare Informationen zu Bedrohungen und Schwachstellen ergänzt. Als KI-Komponente des vorgestellten Ansatzes werden LLMs in Kombination mit RAG als vielversprechend angesehen, um die verfügbaren Daten angemessen zu verarbeiten und nützliche Ergebnisse zu erzielen. Der vorliegende Beitrag stellt zudem eine vorläufige Software-Umsetzung vor, die derzeit im Rahmen des Forschungsprojekts SUSI[14] entwickelt wird. Zukünftige Projektaktivitäten werden die Software kontinuierlich weiterentwickeln.

14 https://www.init-owl.de/en/research/projects/detail/software-basierte-unterstuetzung-von-security-risikobeurteilungen-in-der-industrie/

Literaturverzeichnis

[BKS23] Bhosale, P.; Kastner, W.; Sauter, T.: Integrated Safety-Security Risk Assessment for Production Systems: A Use Case Using Bayesian Belief Networks. In: 2023 IEEE 21st International Conference on Industrial Informatics. 2023.

[BS23a] BSI: IT-Grundschutz-Kompendium, 2023.

[BS23b] BSI: The State of IT Security in Germany 2023, 2023.

[Bu20] Bundeskriminalamt: Bundeslagebild Cybercrime 2020, 2020.

[Ca07] Casey, T.: Threat Agent Library Helps Identify Information Security Risks, 2007.

[EEW22] Eckhart, M.; Ekelhart, A.; Weippl, E.: Automated Security Risk Identification Using AutomationML-Based Engineering Data. IEEE Transactions on Dependable and Secure Computing 19 (3), 2022.

[Eh21] Ehrlich, M.; Bröring, A.; Harder, D.; Auhagen-Meyer, T.; Kleen, P.; Wisniewski, L.; Trsek, H.; Jasperneite, J.: Alignment of Safety and Security Risk Assessments for Modular Production Systems. Elektrotech. & Informationstechnik 138 (7), S. 454–461, 2021.

[Eh23] Ehrlich, M.; Bröring, A.; Diedrich, C.; Jasperneite, J.: Towards automated risk assessments for modular manufacturing systems: Process analysis and information model proposal. at - Automatisierungstechnik 71 (6), 2023.

[Fr22] Franco, M. F.; Sula, E.; Huertas, A.; Scheid, E. J.; Granville, L. Z.; Stiller, B.: SecRiskAI: a Machine Learning-Based Approach for Cybersecurity Risk Prediction in Businesses. In: 2022 IEEE 24th Conference on Business Informatics (CBI). Bd. 01, 2022.

[Ge24] Gebauer, L.; Ehrlich, M.; Wolf, S.; Harder, D.; Schäfer, L.; Trsek, H.; Moriz, N.: Concept for Software-supported Automated Security Risk Assessments for Industrial Components. In: 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA). S. 1–4, 2024.

[HKS21] Hollerer, S.; Kastner, W.; Sauter, T.: Towards a Threat Modeling Approach Addressing Security and Safety in OT Environments. In: 2021 17th IEEE International Conference on Factory Communication Systems (WFCS). S. 37–40, 2021.

[HLL19] He, W.; Li, H.; Li, J.: Unknown Vulnerability Risk Assessment Based on Directed Graph Models: A Survey. IEEE Access 7, 2019.

[Ho23] Hollerer, S.; Brenner, B.; Bhosale, P. R.; Fischer, C.; Hosseini, A. M.; Maragkou, S.; Papa, M.; Schlund, S.; Sauter, T.; Kastner, W.: Challenges in OT Security and Their Impacts on Safety-Related Cyber-Physical Production Systems. In (Vogel-Heuser, B.; Wimmer, M., Hrsg.): Digital Transformation: Core Technologies and Emerging Topics from a Computer Science Perspective. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 171–202, 2023.

[JP20] Johnstone, M.; Peacock, M.: Seven Pitfalls of Using Data Science in Cybersecurity. Data Science in Cybersecurity and Cyberthreat Intelligence, 2020.

[NC23] Nebbione, G.; Calzarossa, M. C.: A Methodological Framework for AI-Assisted Security Assessments of Active Directory Environments. IEEE Access 11, 2023.

[Os22] Oser, P.; van der Heijden, R. W.; Lüders, S.; Kargl, F.: Risk Prediction of IoT Devices Based on Vulnerability Analysis. ACM Trans. Priv. Secur. 25 (2), 2022.

[Ve23] Verizon: 2023 Data Breach Investigations Report (DBIR), 2023.

[YB19] Yeboah-Ofori, A.; Boachie, C.: Malware Attack Predictive Analytics in a Cyber Supply Chain Context Using Machine Learning. In: 2019 International Conference on Cyber Security and Internet of Things (ICSIoT). 2019.

Edge-Management im Industrial Internet of Things

Interoperabilität basierend auf der Verwaltungsschale

Markus Rentschler[1], Xuan-Thuy Dang[2], Dr. Dominik Rohrmus[3], Dr. Andreas Graf Gatterburg[4]

Abstract: IIoT-Gerätemanagement bezieht sich auf die Bereitstellung, Registrierung, Konfiguration, Überwachung und Wartung von angeschlossenen Geräten in industriellen Netzwerken. IT-Dienstleister und Gerätehersteller bieten in der Regel entsprechende Dienste und Schnittstellen an, die jedoch in den meisten Fällen nicht miteinander kompatibel ("interoperabel") sind. An der sogenannten „Edge" zwischen der IT-Ebene und dem industriellen Netzwerk sind wichtige Funktionen und Datenprotokolle nicht standardisiert. Die Interoperabilität zwischen verschiedenen beteiligten Lösungen und Produkten über die architektonischen Schichten hinweg erfordert weitere Standardisierung. In diesem Beitrag wird der aktuelle Stand der Aktivitäten des Testbeds „Edge Management" des „Labs Network Industry 4.0 e.V." vorgestellt, welches Edge-Technologien herstellerübergreifend bezogen auf international standardisierte Softwareschnittstellen evaluiert.

Keywords: API, AAS, REST, Internet of Things, Interoperability, IIoT, Web of Things, W3C

1 Einleitung

Die Konzepte des „Industrial Internet of Things (IIoT)" und „Industrie 4.0" bzw. „Smart Manufacturing" verändern die Art und Weise, wie Daten in Kommunikationsnetzwerken gehandhabt, verarbeitet und geliefert werden. In diesem Zusammenhang hat sich der Begriff „IT/OT-Konvergenz" [1][2] und das Konzept des „Edge Computing" herauskristallisiert, welches als Paradigma und Netzwerkphilosophie darauf abzielt, Rechenkapazität so nah wie nötig an der Datenquelle zu platzieren, um neben anderen Vorteilen die Übertragungslatenz und die Bandbreitennutzung zu reduzieren. Eine anfängliche Motivation für das Edge-Computing-Paradigma war die Senkung der Betriebskosten durch Minimierung der Datenmenge, die an einem zentralen oder Cloud-basierten Standort übertragen und verarbeitet werden muss, und die Bereitstellung einer gewissen dezentralen „Offline-Fähigkeit". Aufgrund des Anstiegs der im Rahmen des IIoT generierten Datenmenge und Datenqualitätsanforderungen wird die Technologie in der Feldebene kontinuierlich vorangetrieben.

Die Relevanz des Themas wird auch unmittelbar an der Vielzahl der in jüngerer Zeit entstandenen Organisationen deutlich, die sich mit der Definition und Implementierung von Edge-Infrastrukturlösungen befassen, in der Regel initiiert von Produkt- oder Lösungsanbietern. In [3] wurden diesbezüglich 75 Edge-Computing-bezogene Aktivitäten identifiziert. Aus den bestehenden Lösungen ist erkennbar, dass sie oft ähnlichen

[1] ARENA2036 e.V., 70569 Stuttgart, markus.rentschler@arena2036.de
[2] Yacoub GmbH, 13355 Berlin, xuan-thuy.dang.ext@murrelektronik.de
[3] Labs Network Industrie 4.0 e.V., 10117 Berlin, dominik.rohrmus@siemens.com
[4] Hilscher Gesellschaft für Systemautomation mbH, 65795 Hattersheim, agatterburg@hilscher.com

Konzepten folgen, jedoch nicht auf interoperable Ansätze im Sinne einer Standardisierung fokussieren.

Hervorzuheben ist die jüngste Initiative „Margo - Edge Interoperabilität für industrielle Automatisierungsökosysteme“ [4], welche ebenfalls standardbasierte Interoperabilität als Ziel verfolgt.

Als eine führende Testbed-Organisation hat der vorwettbewerbliche und gemeinnützige deutsche Verein „Labs Network Industrie 4.0 e.V.“ (LNI 4.0) [5] im Jahr 2018 das Testbed „Edge Configuration“ eingerichtet und später neufokussiert in „Edge Management“ umbenannt. LNI 4.0 wurde 2015 zusammen mit der Deutschen „Plattform Industrie 4.0“ und dem „Standardization Council Industrie 4.0“ (SCI 4.0) [6] gegründet, das gemeinsam von DIN und DKE zur Unterstützung der Industrie 4.0-Standardisierungsaktivitäten getragen wird. Ein Ziel der Arbeitsergebnisse des LNI 4.0-Testbeds ist es, sicherzustellen, dass KMU (kleine und mittlere Unternehmen) einfach die Edge Technologie nutzen können, da KMU das Rückgrat der weltweiten Produktion bilden. Alle LNI 4.0 Testbeds arbeiten in einer neutralen Umgebung, die von öffentlichen Forschungsinstituten betrieben wird, und agieren vollständig vorwettbewerblich und neutral. LNI 4.0 kooperiert eng mit der „Open Industry 4.0 Alliance“ [7], einem internationalen Industriekonsortium, das Best Practices für die Anwendung von Standards für Interoperabilität entwickelt. Eines der Interessengebiete ist Edge Computing, wobei das LNI 4.0 Testbed „Edge Management“ die Standardisierung im Kontext der Interaktion zwischen der Edge-Management-, der Edge- und der Feldgeräteebene unterstützt (*Abbildung 1*). Da derzeit kein allgemein akzeptierter Standard mit diesem Fokus existiert, entwickelt das LNI 4.0 Testbed einen Vorschlag, welcher eine einheitliche Zugänglichkeit von Management-Anwendungen zu einer Vielzahl heterogener Assets (d.h. Geräte, Apps, etc.) ermöglicht. Die Ergebnisse und Erfahrungen des LNI 4.0 Testbeds werden den Standardisierungsgremien kontinuierlich zur Verfügung gestellt, um sie in die weitere Entwicklung von Standards einfließen zu lassen. In [8] berichtete das LNI 4.0 Testbed bereits über seine Aktivitäten zum Thema Edge-Konfiguration, also der Konfiguration der Interaktion zwischen Edge und der Edge-Management-Schicht, wie in Abbildung 1 dargestellt. Als Ergebnisse wurden verschiedene Aspekte der Edge-Konfiguration methodisch entwickelt, in Form eines „Business View“ [9], „Usage View“ [10], „Functional View“ und „Implementation View“.

Da bereits eine Reihe von proprietären Lösungen für bestehende IIoT-Gerätemanagementlösungen auf dem Markt sind, ist eine wichtige Anforderung, deren bereits existierenden APIs über ein einfaches Konzept und mit einem Minimum an Implementierungsaufwand in ein standardisiertes API-Konzept einbinden zu können, um von den Anbietern und Anwendern dieser bestehenden Lösungen akzeptiert zu werden. Dabei sollten möglichst viele bestehende offene Standards adaptiert werden.

Nachdem in der Testbed-Arbeitsgruppe anfänglich auf eine selbstbeschreibende REST-API basierend auf HATEOAS-Mechanismen [11] und WoT-Konzepten [12] hingearbeitet wurde, erfolgt nach einem Erkenntnisgewinn eine Umorientierung hin zur Anwendung der API-Definitionen der Verwaltungsschale (AAS) [13] und der Teilmodelle DNP [14], AID [15] und AIMC [16]. Im Kontext von anderen Projekten wie „TwinMap“ [17], „Catena-X“ [18] sowie „Verwaltungsschale für den Leitungssatz (VWS4LS)“ [19] wird ebenfalls dieser Ansatz genutzt.

Der Hauptbeitrag dieser Arbeit ist der Entwurf einer Architektur für eine API-gesteuerte Interaktion mit heterogenen Edge-Computing-Entitäten auf der Grundlage des offenen AAS-Standard (IEC 63278), insbesondere mit dem AAS Submodel-Template „Digital Nameplate" (IEC 61406-1/-2) zur Identifikation des jeweiligen Kommunikationspartners und den AID- und AIMC-Teilmodellen zur Veröffentlichung seiner Schnittstellen.

Der Beitrag ist wie folgt gegliedert:

- In Abschnitt 2 „Hintergrund" werden wesentliche Motivatoren für ein standardisiertes Edge-Management aufgeführt.

- Abschnitt 3 „Edge Architektur" beschreiben die von den LNI 4.0-Testbed-Partnern entwickelte Referenzarchitektur.

- In Abschnitt 4 „Edge Management API" wird auf der Grundlage der analysierten Konzepte verschiedener Standardisierungsorganisationen und der damit verbundenen offenen Standards die Bausteine für die Implementierung unseres Entwurfs einer API für das industrielle Edge-Management diskutiert.

- In Abschnitt 5 „LNI 4.0 Testbed Demonstrator" wird die Implementierung des Edge-Management-Testbeds beschrieben, unter Verwendung der vorgeschlagenen Modelle und API-Konzepte mit einem Anwendungsfall für die herstellerübergreifende Verwaltung und Interaktion von Edge-Assets.

- Der Abschnitt 6 „Fazit und Ausblick" schließt die Arbeit mit offenen Fragen und Vorschlägen zur weiteren Entwicklung der Edge-Interoperabilitätskonzepte ab.

2 Hintergrund

Neben den bereits in der Einleitung genannten technischen Vorteilen des Edge-Paradigmas, sind wesentliche Treiber für ein standardisiertes Edge Management im Kontext von Automatisierungstechnik und Industrie 4.0 vor allem Anwendungsfälle wie:

- Fertigungsaktivitäten sollen für eine effektive Interoperabilität nach außen hin transparent sein.
- Informationen über Modelle von Fertigungsaktivitäten und Artefakten (Assets) sollen digital dargestellt und über Informationsinfrastrukturen ausgetauscht werden können.
- Um eine intelligente Fertigung zu ermöglichen, sollen Technologien systematisch integriert werden können.
- Artefakte, die in der Fertigungsindustrie entstehen, sollen in verschiedenen Bereichen über den gesamten Lebenszyklus genutzt werden können.
- Bewertung der Nachhaltigkeit durch Erfassung des Umwelteinflusses, der durch Artefakte und Produktionstätigkeiten verursacht wird. Zu diesem Zweck ist die Einbeziehung von Umweltaspekten in die Modelle notwendig, welche in den herkömmlichen Produkt- und Prozessmodellen nicht hinreichend abgebildet werden. Die Berechnung des CO2-Fußabdrucks eines Produkts erfordert beispielsweise Informationen über die Menge der CO2-Emissionen, die mit den verwendeten Materialien dem Herstellungsprozess, dem aktuellen Betriebszustand, der Betriebsumgebung usw. zusammenhängen. Diese Beziehungen zwischen CO2-Emissionen und Herstellungs- sowie Einsatzbedingungen müssen gepflegt und transparent bereitgestellt werden.

3 Edge Architektur

Das LNI 4.0-Testbeds „Edge Management“ hat für seine Arbeit eine mehrschichtigen Architektur definiert, wie in *Abbildung 1* dargestellt.

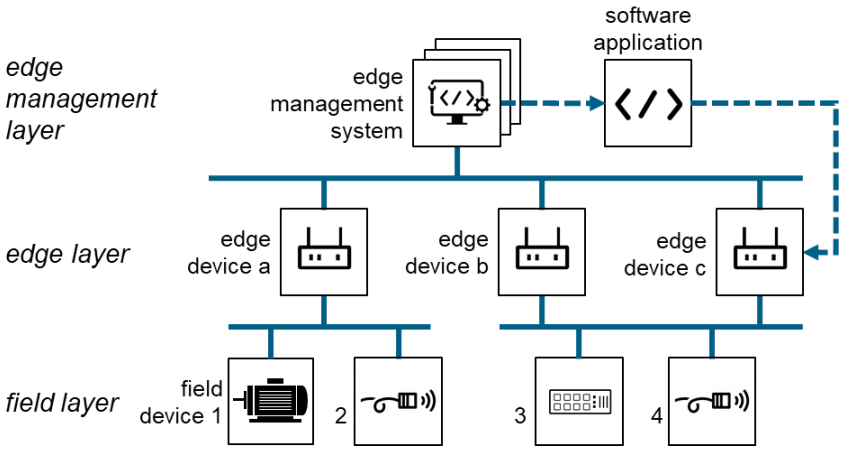


Abbildung 1: Architekturmodell [9][10]

Dabei lassen sich die folgenden grundlegenden Systemeinheiten identifizieren:

Feldgeräte sind physische Datenverarbeitungsressourcen mit oft deterministischen Kommunikationsfähigkeiten. Feldgeräte kommunizieren mit Edge-Geräten, können über Parameter konfiguriert werden und die Firmware eines Feldgeräts kann aktualisiert werden. Feldgeräte unterstützen nicht die Bereitstellung von Anwendungen.

Edge-Geräte sind physische Rechenressourcen mit Kommunikationsfähigkeiten und Edge-Laufzeiten, die auf dem Edge-Gerät eingesetzt werden können. Edge-Geräte können auch durch Parameter konfiguriert werden und die Firmware von Edge-Geräten kann aktualisiert werden. Edge-Geräte können mit Feldgeräten verbunden werden, und für jedes verbundene Feldgerät gibt es einen Datenendpunkt, der die Kommunikationsmöglichkeiten zwischen Feld- und Edge-Gerät darstellt. Diese Datenendpunkte können von einem Edge-Management-System konfiguriert werden.

Ein **Edge-Management-System** ist ein Softwareprogramm, das in einer IT-Infrastruktur eingesetzt wird. Ein Edge-Management-System kann Konfigurationsfunktionen für Feldgeräte und Edge-Geräte, einen Anwendungsspeicher zur Bereitstellung von Softwareanwendungen, Edge-Laufzeiten und Firmware sowie Konfigurations- und Bereitstellungsfunktionen bieten, die über den Anwendungsspeicher des Edge-Management-Systems bereitgestellt werden.

Software-Applikationen sind ausführbare Software-Programme, die auf einer Edge-Laufzeitumgebung oder einer IT-Infrastruktur bereitgestellt, ausgeführt und konfiguriert werden können. Die Software-Programme werden über den Anwendungsspeicher eines Edge-Management-Systems bereitgestellt. Diese Softwareanwendungen nutzen in der Regel die Informationen von angeschlossenen Feldgeräten, z. B. zur Datenanalyse. Grundlegende Anwendungen für den allgemeinen Gebrauch sind z.B. Asset Management [20] und Systemüberwachung [21].

4 Edge Management API

4.1 Grundlegende Dienste

In diesem Abschnitt werden kurz einige grundlegende Funktionen eines Edge-Management-Systems beschrieben, die im Wesentlichen **IT Asset Management (ITAM)** [22] Funktionalität abdeckt, um die Überwachung und Verwaltung von Hardware-Geräten, Software, Lizenzen, Zugriffsrechten usw. innerhalb des betrachteten IT/OT-Systems zu ermöglichen.

Detection & Discovery befasst sich mit Funktionen, die Informationen über neue oder geänderte Hardware- und Software-Entitäten lieferen.

Verwaltung von Zugriffsrechten ist notwendig zum Schutz vor unberechtigtem Zugriff auf Konfigurationsfunktionen, die auf Daten oder Parameter zugreifen. Diese Funktionen bieten eine Grundlage für die Zuweisung und Kontrolle von Zugriffsrechten für bestimmte Benutzer mit bestimmten Rollen.

Konfigurationsfunktionen ermöglichen die Manipulation eines Systems in einer Weise, damit es einen von externen Rollen vorgegebenen Zweck erfüllt, und das System dadurch zweckorientiert genutzt werden kann.

Parametrisierung umfasst Änderungen an jeder Art von Arbeitsparameter, um die Funktion der Entität zu verändern. Dies kann Kommunikationsparameter wie IP-Adressen einschließen.

Integritätsüberwachung ist verantwortlich für die Erkennung von Fehlfunktionen oder um unvorhergesehene äußere Einflüsse zu erkennen und zu signalisieren, die Auswirkungen auf das System haben können. Geeignete Ereignisbehandlungs-Verfahren müssen bei solchen möglicherweise problematischen Ereignissen ausgelöst werden.

4.2 IIoT-Protokollstandards

In industriellen IoT- und Machine-to-Machine (M2M) - Kommunikationsanwendungen sind OPC UA [23] sowie REST-basierte Schnittstellen [24] weit verbreitet. Eine REST-API definiert Endpunkte, die zur Adressierung der Interaktionen dienen. Die aufkommende AAS-Technologie definiert ebenfalls eine REST-Schnittstelle und einige ihrer Submodell-Definitionen beziehen sich auf „Web of Things (WoT)"-Architekturkonzepte, die vom „World Wide Web Consortium (W3C)" für die Interoperabilität verschiedener IoT-Plattformen und Anwendungsdomänen definiert wurden. Das Kernstück ist dabei die „Thing Description (TD)" [12], ein standardisiertes, maschinenlesbares Metadaten-Repräsentationsformat für die Entdeckung und Interpretation der Fähigkeiten eines „Thing" durch semantische Annotationen und HATEOAS [11] basierte Verknüpfungsmechanismen. In [25] wird gezeigt, wie dies die Interoperabilität über verschiedene IoT-Plattformen, Ökosysteme und Standards hinweg unterstützt.

„umati" [26] kann als eine führende Initiative genannt werden, die OPC UA in herstellerübergreifenden Anwendungsszenarien einsetzt.

Bezeichnenderweise hat auch die OPC Foundation eine Arbeitsgruppe angekündigt, die OPC UA um ein standardisiertes REST-Interface erweitern soll [27].

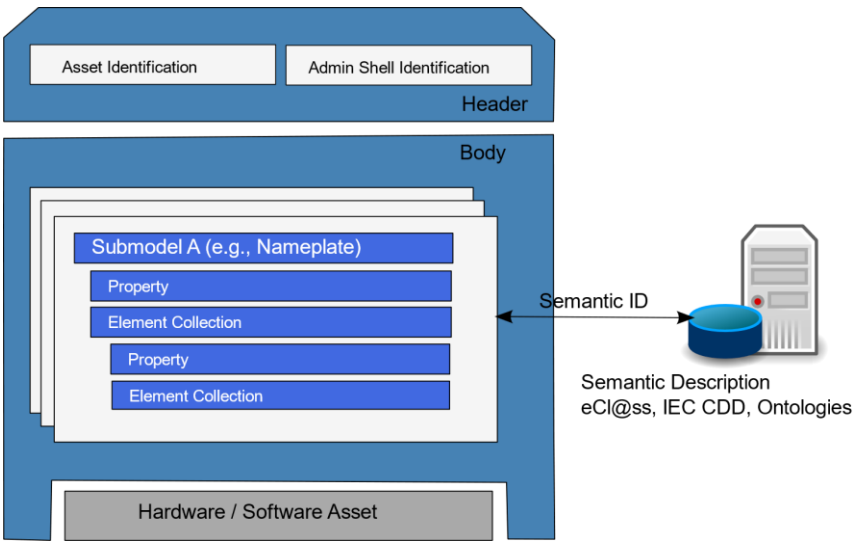
Es ist hierbei erwähnenswert, dass der OPC UA Standard in der Companion Specification „OPC 30270 Industry 4.0 Asset Administration Shell“ [28] bereits die Kommunikation von Asset-Informationen definiert. In einer kürzlich erschienenen Veröffentlichung [29] wurden die Kommunikationsfähigkeiten erörtert und Anwendungsszenarien für AAS-Typ-2-Implementierungen empfohlen.

4.3 Die Asset Administration Shell

Im Wesentlichen ermöglicht der AAS-Standard der Industrial Digital Twin Association (IDTA) die standardisierte und interoperable Darstellung des digitalen Zwillings von Hardware- und Software-Assets in industriellen Anwendungen. Eine AAS enthält u.a. Informationen über die Identifikationsdaten, den Lebenszyklus, die Funktionalitäten und die Betriebsdaten eines Assets. Im Kontext der RAMI4.0-Architektur [30] realisiert die AAS die sog. „I4.0-Komponente“. Die IDTA definiert auch Mechanismen für den Austausch von AAS entlang der Wertschöpfungskette und implementiert die Kommunikation zwischen I4.0-Komponenten.

Eine AAS-Struktur *(*

Abbildung 2) besteht aus einem Header-Teil, der Informationen zur Identifikation des AAS und seines Assets enthält, und einem Body-Teil, der alle anderen Eigenschaften und Funktionen in Bezug auf das Asset enthält. Bei den Eigenschaften handelt es sich um Schlüssel-Wert-Paare, die die Merkmale der Ressource beschreiben, während die Funktionen ihre Fähigkeiten und Operationen beschreiben. Die Eigenschaften und Funktionen werden in standardisierten Teilmodellen gruppiert, um den Austausch relevanter Asset-Informationen in den übergreifenden Prozessen oder zwischen den



Assets zu ermöglichen.

Abbildung 2: AAS-Struktur

Typische Beispiele für Teilmodelle sind Digitales Typenschild, Carbon Footprint, und Übergabedokumentation. Im Folgenden beschreiben wir kurz die Teilmodelle, die in

unserer Testumgebung für die herstellerübergreifende Interoperabilität im Industrial Edge eingesetzt werden:

Das Teilmodell **Digital Nameplate (DNP)** [12] bietet eine Vorlage zur Beschreibung von Asset-Nameplate-Informationen in der zugehörigen AAS. Die Asset-Informationen werden als Eigenschaften dargestellt, die auf der Grundlage von ECLASS- oder IEC Common Data Dictionary (CDD)-Wörterbüchern kategorisiert werden. Jede Eigenschaft im DNP hat dadurch eine semantische ID, die einen eindeutigen Austausch von Typenschildinformationen zwischen verschiedenen Partnern in der Wertschöpfungskette ermöglicht.

Das Teilmodell **Asset Interface Description (AID)** [13] spezifiziert ein Informationsmodell zur Beschreibung von Schnittstellen, die die Interaktion mit einem Asset-Dienst oder Asset-bezogenen Diensten ermöglichen. Dabei definiert AID die Datenendpunkte, um Daten vom Asset anzufordern oder Operationen des Assets auszulösen. Die W3C-TD-Struktur und Protokollbindungen wurden in die AID-Kernstruktur übernommen. Zusätzlich zu den Schnittstellenprotokollen können auch externe Deskriptoren eingebunden werden, z.B. GSD, GSDML, IO Device Description, WoT-TD, etc. In der Version 1 unterstützt AID vor allem die standardisierte Beschreibung von Schnittstellen auf der Basis der Modbus-, HTTP- und MQTT-Protokolle, in der zukünftigen Version soll auch das OPC-UA-Protokoll abgedeckt werden.

Das Teilmodell **Asset Interfaces Mapping Configuration (AIMC)** [14] ergänzt AID, indem es die Synchronisierung der über die beschriebenen Schnittstellen abgerufenen Daten mit einem bestimmten Ort innerhalb des AAS unterstützt. Dies geschieht durch die „*MappingSourceSinkRelation"*-Beschreibung in AIMC, die als Quelle einen im AID-Teilmodell definierten Asset-Datenpunkt mit einer Ziel-Property in einem beliebigen AAS-Teilmodell verknüpft.

AID und AIMC sind das geeignete Mittel für die generische Anbindung eines Assets an eine AAS, da sowohl die Metainformationen zu den jeweiligen Kommunikationstechnologien als auch die Endpunkte der Assetsignale abgebildet werden können. Eine diesbezügliche Anwendung wurde auch bereits beschrieben in [31]

5 LNI 4.0 Testbed Demonstrator

Der LNI 4.0 Edge Management Demonstrator zeigt eine prototypische vorwettbewerbliche Umsetzung. Die beteiligten industriellen KMU sowie Großunternehmen lösen gemeinsam die notwendigen Schritte für ein standardisiertes Edge Management, beginnend mit dem Onboarding von Edge-Geräten in ein Edge-Management-System. Die einheitliche Verwendung des AAS Teilmodell Templates „Digital Nameplate" (AAS DNP) durch die Geräte ermöglicht nicht nur deren gegenseitige Identifizierbarkeit, sondern auch die Versorgung mit weiteren Informationen. Verwaltungsschalenkonzepte wie die Asset Modellierung und der Datenaustausch über eine definierte API werden genutzt.

5.1 Architektur des Edge Management Testbed

Die Architektur wurde entworfen basierend auf den in Abschnitt 3 vorgestellten Edge Konzept. Dabei wurden auch Anforderungen von sog. „Brownfield"-Systemen

berücksichtigt, die durch proprietäre Kommunikationsprotokolle angebunden werden müssen (*Abbildung 3*).

Die generischen Schnittstellen zwischen den Edge-Management- und Edge-Ebene werden mittels einer AAS-Infrastruktur realisiert. Informationen aus den Edge-Geräten der verschiedenen Hersteller werden in entsprechenden Verwaltungsschalen aktualisiert über die REST API der AAS. Ein herstellerübergreifende Edge-Management-System kann so auf die Gerätinformationen zugreifen. Umgekehrt können Konfigurationen oder Befehle an die Geräte kommuniziert werden.

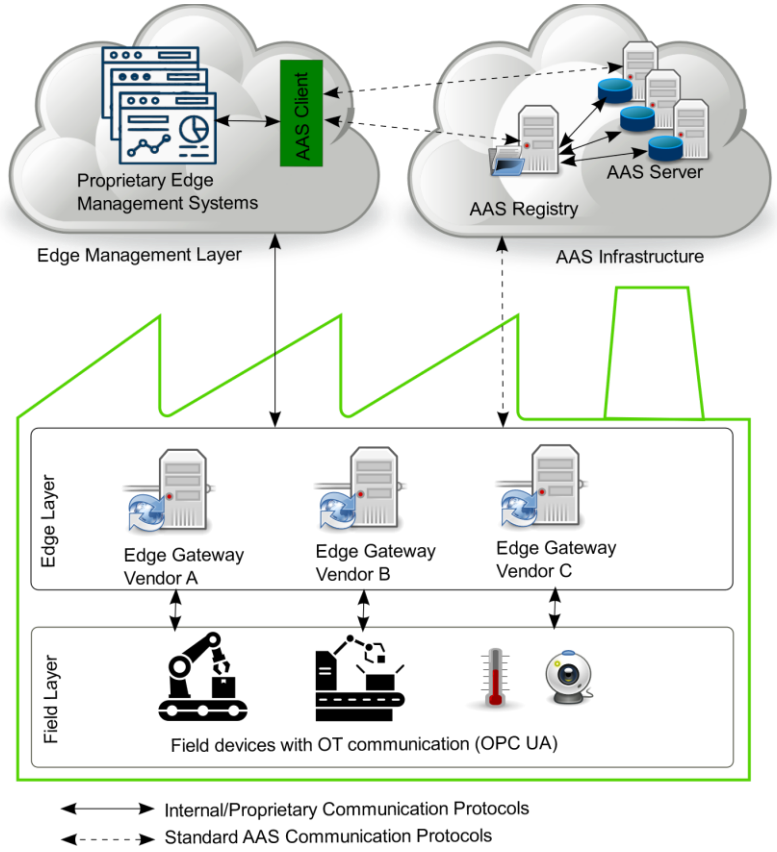


Abbildung 3: Architektur des Testbeds

5.2 Demonstration: Szenario und Umsetzung

Das Edge Management Szenario demonstriert die herstellerübergreifende Bereitstellung von Informationen von Edge-Geräten über Edge-Systeme verschiedener Hersteller. *Abbildung 4* zeigt zwei Szenarien des Edge-Management: Aktualisierung und Konfiguration von Edge-Gerät IP Adresse. Andere Informationen, z.B., Aktordaten, können in gleichartigen Szenarien angewendet werden.

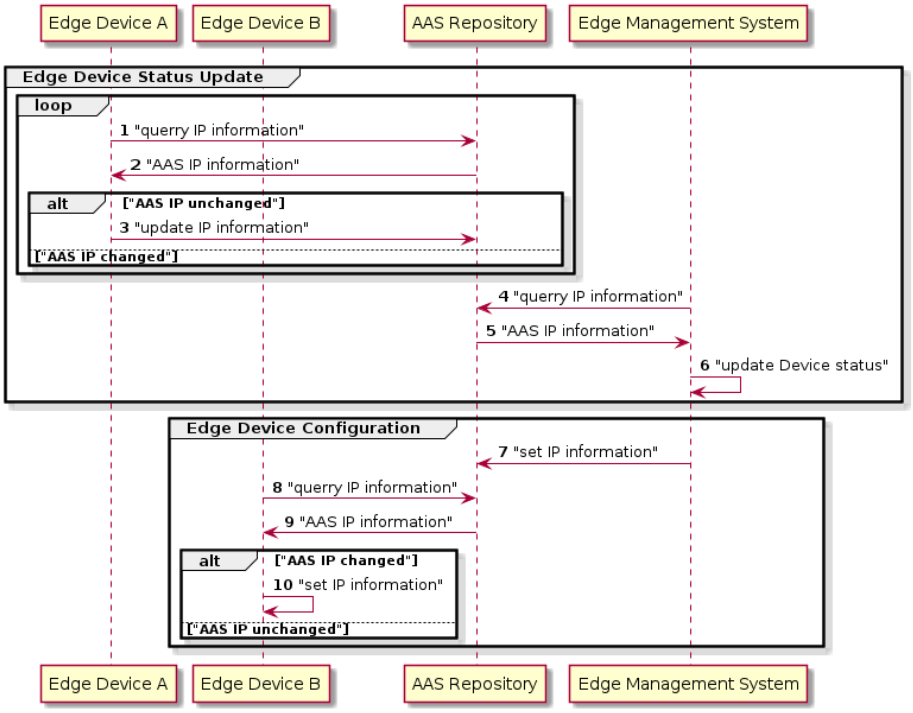


Abbildung 4: Kommunikations-Sequenzen für die Aktualisierung und Konfiguration von Edge-Geräten

In *Abbildung 4* sind folgende Szenarien abgebildet:

1. **IP-Konfiguration durch das Edge-Gerät:** Im Falle des Edge-Geräts A fragt dieses seine Verwaltungsschale im AAS Repository (d.h. den Digitalen Zwilling) nach seiner abgebildeten IP-Information (1). Falls die IP in der Verwaltungsschale abweicht, wird die lokale IP als aktueller angenommen und soll auch in der VWS aktualisiert werden (3). Nach der Aktualisierung fragt das EMS bei dem AAS Repository nach der IP des Geräts (4) und aktualisiert entsprechend die auf der GUI angezeigten Gerätdaten (6).
2. **IP-Konfiguration durch das EMS**: Im Falle des Edge-Gerät B wird die gewünschte IP durch den User über das EMS gesetzt und in der VWS gespeichert. Durch die periodische Synchronisation zwischen dem Edge-Gerät und dem AAS Repository (8) erkennt das Edge-Gerät, dass eine neuere IP in der VWS vorhanden ist und seine lokale IP entsprechend setzen (10).

Der LNI 4.0 Edge Management Demonstrator zeigt insbesondere die herstellerübergreifende Bereitstellung von Informationen von Edge-Geräten über Edge-Systeme verschiedener Hersteller. Insgesamt sind 12 namhafte Industrieautomatisierungsunternehmen im Demonstrator aktiv (*Abbildung 5*).

Abbildung 5: Testbed-Mitglieder

Der aktuelle Stand der Umsetzung im Demonstrator basiert auf der Verwaltungsschale, also der standardisierten Asset-Modellierung und den Austausch von Informationen über eine definierte API, zudem werden die Edge-Device-Informationen einheitlich in einer generischen Benutzeroberfläche (UI) dargestellt (*Abbildung 6*). Diese beinhaltet u.a.:

- IP-Adresse und andere Informationen zur Netzwerkkonfiguration

- Zustandsinformationen der verschiedenen Geräte wie z.B. Temperaturen der Elektronik, die Auslastung der Recheneinheit oder die Verfügbarkeit des Speichers auf Basis von AAS-Teilmodellen.

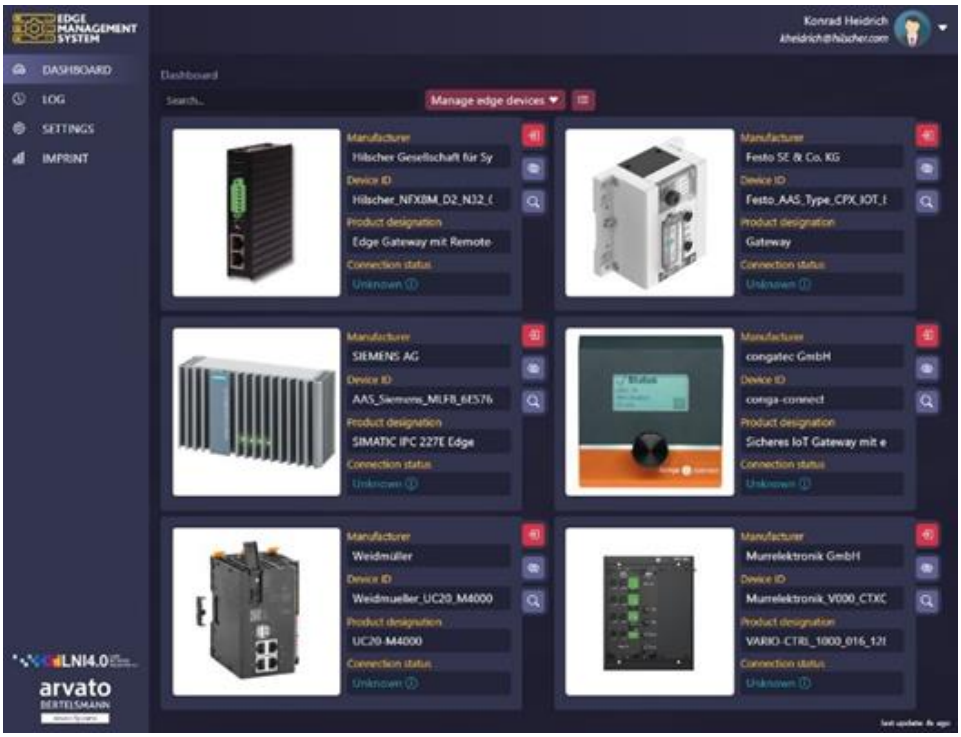


Abbildung 6: Prototypisches LNI 4.0 Dasboard

Die Implementierung basiert auf dem AAS-Server von IDTA-Version 3 und wird kontinuierlich weiterentwickelt. Ein Open-Edge-Device-Builder-Kit, das die Software Containerisierung mittels Docker-Technologie nutzt, wurde entwickelt, um die

Kommunikation über Herstellergrenzen hinweg so aufwandsarm wie möglich bereitzustellen. Zusätzlich wird die direkte Kommunikation über einen Message-Broker mit den integrierten Fähigkeiten der Edge-Geräte demonstriert.

Das System wird dauerhaft in der IT-Umgebung eines LNI 4.0 Mitglieds zur Verfügung gestellt, so dass der Demonstrator und die Edge Technologien ständig gewartet und optimiert werden können.

6 Fazit & Ausblick

Der LNI 4.0 Edge Management Demonstrator unterstützt die Vision, dass vollständige Interoperabilität für Geräte, Software-Onboarding, Softwarebereitstellung und Anwendungskonfiguration durch eine gemeinsame standardisierte Schnittstelle möglich ist.

Diese herstellerübergreifende Interoperabilität basierend auf dem AAS-Standard und den zugehörigen AAS Submodel-Templates kann die Hindernisse für Investitionsentscheidungen auf dem Markt beseitigen, da die Kunden sicher sein können, dass ihre installierte Gerätebasis auch in der Zukunft weiter interoperabel bleiben wird.

Der gezeigte Demonstrator bildet den Kern der LNI 4.0 Testbed-Aktivitäten und wird kontinuierlich weiterentwickelt, getestet und die Ergebnisse von LNI 4.0 kontinuierlich veröffentlicht werden.

7 Literaturverzeichnis

[1] D. Lewandowski, D. Pareschi, W. Pakos, and E. Ragaini, „Future of IoTSP – IT and OT Integration", in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, Barcelona, Spain, Aug. 2018, pp. 203–207

[2] M. Felser, M. Rentschler, and O. Kleineberg, „Coexistence Standardization of Operation Technology and Information Technology," *Proc. IEEE*, vol. 107, no. 6, pp. 962–976, Jun. 2019

[3] A. Hamm, A. Willner, and I. Schieferdecker, „Edge Computing: A Comprehensive Survey of Current Initiatives and a Roadmap for a Sustainable Edge Computing Development," 2019, doi: 10.48550/ARXIV.1912.08530.

[4] The Linux Foundation, „Margo - Edge interoperability for industrial automation ecosystems,", 2024. https://margo.org/

[5] German association „Labs Network Industrie 4.0", https://lni40.de

[6] German organization „SCI 4.0", https://www.sci40.com

[7] Swiss association „Open Industry 4.0 Alliance", https://openindustry4.com/news-und-presse/open-industry-40-alliance-and-labs-network-industrie-40-work-together-for-digital-transformation/

[8] M. Rentschler, D. Rohrmus, U. Löwen, A. G. Gatterburg, B. Vojanec, A. Willner, „Standardization of Edge Configuration", 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2020): 1147-1150.

[9] "LNI Testbed Edge Management – Business View", Labs Network Industrie 4.0 e.V., 2019, https://dev2.lni40.de/lni40-content/uploads/2021/03/BusinessView-LNI_Testbed-Edge-Management_V2.0.pdf

[10] “LNI Testbed Edge Management – Usage View”, Labs Network Industrie 4.0 e.V., 2020, https://lni40.de/wp-content/uploads/2023/02/UsageView-TestbedEdgeConfiguration_publish-10032020.pdf

[11] R. T. Fielding, „Architectural Styles and the Design of Network-based Software Architectures,” University of California, Irvine, 2000. https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

[12] „Web of Things (WoT) Thing Description 1.1”, W3C Editor's Draft, https://w3c.github.io/wot-thing-description/

[13] Industrial Digital Twin Association, „Specification of the Asset Administration Shell Part 2: Application Programming Interfaces – IDTA 01002-3-0”, https://industrialdigitaltwin.org/wp-content/uploads/2023/06/IDTA-01002-3-0_SpecificationAssetAdministrationShell_Part2_API_.pdf

[14] Industrial Digital Twin Association, „IDTA 02006 – Digital Nameplate for Industrial Equipment”, Oktober 2022, https://github.com/admin-shell-io/submodel-templates/tree/main/published/Digital%20nameplate/2/0

[15] Industrial Digital Twin Association, „IDTA 02017 – Asset Interfaces Description”, Januar 2024, https://github.com/admin-shell-io/submodel-templates/tree/main/published/Asset%20Interfaces%20Description/1/0

[16] Industrial Digital Twin Association, „IDTA 02027 – Asset Interfaces Mapping Configuration”, Juni 2024, https://github.com/admin-shell-io/submodel-templates/tree/main/published/Asset%20Interfaces%20Mapping%20Configuration/1/0

[17] „TwinMaP - Digitalisierung für eine branchenübergreifende Vernetzung”, https://twinmap.de/

[18] „Digital Twins in Catena-X”, Catena-X-Standard CX-0002, https://catenax-ev.github.io/docs/standards/CX-0002-DigitalTwinsInCatenaX

[19] „VWS4LS - Verwaltungsschale für den Leitungssatz”, ARENA2036 e.V., https://arena2036.de/de/vws4ls

[20] M. Wang, J. Tan, Y. Li, „Design and implementation of enterprise asset management system based on IOT technology,” in *IEEE International Conference on Communication Software and Networks (ICCSN)*, Chengdu: IEEE, Jun. 2015, pp. 384–388.

[21] M. D. Unde, Prasad. S. Kurhe, „Web based control and data acquisition system for industrial application monitoring,” in *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai: IEEE, Aug. 2017, pp. 246–249.

[22] ISO/IEC 19770 Standard series, ITAM, https://www.itam.org/itam-best-practices-standard/

[23] M. Ghazivakili, C. Demartini, C. Zunino, „Industrial data-collector by enabling OPC-UA standard for Industry 4.0,” in *14th IEEE International Workshop on Factory Communication Systems (WFCS)*, Imperia, Italy: IEEE, Jun. 2018, pp. 1–8.

[24] F. Aydemir and F. Basciftci, „Application of HATEOAS Principle in RESTful API Design,” 2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)

[25] H. K. Pakala, Kazeem. O. Oladipupo, S. Käbisch, C. Diedrich, „Integration of asset administration shell and Web of Things,” KommA 2021, http://dx.doi.org/10.25673/39570

[26] German association „Universal machine technology interface”, https://umati.org, https://www.umati.app/

[27] OPC Foundation, „OPC UA REST Subgroup Launched“, https://opcconnect.opcfoundation.org/2023/03/opc-ua-rest-subgroup-launched/

[28] OPC Foundation, „OPC 30270 Industry 4.0 Asset Administration Shell”, https://opcfoundation.org/developer-tools/documents/view/273

[29] R. Drath *et al.*, „Diskussionspapier - Interoperabilität mit der Verwaltungsschale, OPC UA und AutomationML”, https://www.automationml.org/about-automationml/publications/discussionpaper-2023/

[30] DIN SPEC 91345:2016, „Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)”, Apr. 2016.

[31] A. Schließmann, M. Stolze, M. Riedl, T. Schroeder, „Standardisierte Maschinenanbindung an ein Produktionsleitsystem über die Asset Administration Shell”, KommA 2023, http://dx.doi.org/10.25673/111742

Impressum

15. Jahreskolloquium
Kommunikation in der Automation
(KommA 2024)

06. November 2024 • Lemgo

OPEN-Book

ISBN: 978-3-9818463-5-5
DOI: https://doi.org/10.25644/jnac-vp34

Herausgeber:
Jürgen Jasperneite, Lemgo
Ulrich Jumar, Magdeburg

Institut für industrielle Informationstechnik
Technische Hochschule Ostwestfalen-Lippe

Campusallee 6
D-32657 Lemgo
Telefon: +49 5261 702-2400
Internet: www.init-owl.de | www.jk-komma.de