INSTITUT FÜR AUTOMATION UND KOMMUNIKATION E.V. MAGDEBURG
AN-INSTITUT DER OTTO-VON-GUERICKE-UNIVERSITÄT MAGDEBURG

KOMMA 2025

Kommunikation in der Automation

TAGUNGSBAND

11./12.11.2025 | 16. JAHRESKOLLOQUIUM

« KOMMUNIKATION IN DER AUTOMATION »

LISA UNDERBERG, JÜRGEN JASPERNEITE (HRSG.)

■ EINE KOOPERATION VON: ifak inIT TH OWL

■ UNTERSTÜTZT VON: VDE ITG GESELLSCHAFT FÜR INFORMATIK

Impressum

16. Jahreskolloquium

« KOMMUNIKATION IN DER AUTOMATION» (KOMMA 2025)

11./12.11.2025 • Magdeburg

ISBN 978-3-948749-60-6

http://dx.doi.org/10.25673/121097

HERAUSGEBER

Lisa Underberg, Magdeburg
Jürgen Jasperneite, Lemgo

Institut für Automation und Kommunikation e.V.
An-Institut der Otto-von-Guericke-Universität Magdeburg
Werner-Heisenberg-Straße 1
39106 Magdeburg

Telefon: +49 (0)391 990140
Fax: +49 (0)391 9901590
Internet: www.ifak.eu

Vorwort

Tagungsband des 16. Jahreskolloquiums
Kommunikation in der Automation – KommA 2025

Vor 15 Jahren fand in Lemgo das erste Jahreskolloquium „Kommunikation in der Automation – KommA“ statt. Die damals von den beiden Instituten, dem inIT der Hochschule Ostwestfalen-Lippe in Lemgo und dem ifak e.V. in Magdeburg, initiierte Veranstaltungsreihe findet bis heute gute Resonanz. Die Veranstaltungsreihe ist also von Kontinuität geprägt, die auch nach der Staffelstabübergabe der Leitung des ifak weitergeführt wird. Das 1991 vom Feldbusexperten Prof. Peter Neumann in Magdeburg initiierte ifak wurde bis 2005 von ihm geleitet. Er gab anschließend den Staffelstab an Prof. Ulrich Jumar weiter, dessen ruhestandsbedingte Nachfolge in der Leitung des ifak Prof. Lisa Underberg antrat. Die bewährte Kooperation der Fachkolleginnen und -kollegen aus Lemgo und Magdeburg hat – nicht nur im Rahmen des Jahreskolloquiums KommA – weiterhin Bestand.

In den letzten Jahren, sogar Jahrzehnten, haben sich Automatisierung und Kommunikation längst zu Querschnitts und Schlüsseltechnologien entwickelt, die in immer mehr Bereichen – von intelligenten Energienetzen über vernetzte Mobilität bis hin zu digitalen Gesundheitssystemen – zentrale Funktionen übernehmen. Ihr Beitrag ist dabei oft unsichtbar, doch unbestritten: Sie bilden das unsichtbare Rückgrat moderner Wirtschaft und Gesellschaft. Das Jahreskolloquium richtet den Blick gezielt auf diese unsichtbaren Schlüsseltechnologien, und macht die aus Anwendungssicht verborgenen Technologien sichtbar. Damit schafft es die Basis für weiterführende wissenschaftliche, methodische und praxisorientierte Ansätze, die anschließend in den vielfältigen Anwendungsfeldern ihre volle Wirkung entfalten. Eine besondere Rolle spielen dabei Funkkommunikationssysteme, die zum einen durch ihre Nachrüstbarkeit und wartungsarme Handhabung die Effektivität bestehender Anwendungsfelder erhöhen können und zum anderen durch ihre Flexibilität neue Anwendungsfelder, insbesondere im Kontext von Mobilität, erschließen können. Der besonderen Rolle widmet sich der Industrial Radio Day, welcher zum fünften Mal ausgerichtet wird, und in diesem Jahr in Magdeburg in Verbindung mit der KommA stattfindet.

Zum 16. Jahreskolloquium Kommunikation in der Automation, das am 11./12. November 2025 in Magdeburg am ifak stattfand, finden Sie hier den elektronischen Tagungsband. Damit trotz der kurzen Tagungsdauer dieser kleinen Veranstaltung viele Themen Berücksichtigung finden konnten, gehören neben den Vorträgen zusätzlich Poster zum Programm, die ebenfalls als Vollmanuskripte im Tagungsband enthalten sind. Die Tagungsleiter wünschen eine interessante Lektüre!

Prof. Dr. Lisa Underberg

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

Prof. Dr. Jürgen Jasperneite

inIT - Institut Industrial IT TH Ostwestfalen-Lippe, Lemgo

Inhaltsverzeichnis

KommA 2025 – Kommunikation in der Automation

http://dx.doi.org/10.25673/121097

Virtuelle Steuerungen
Evolution, Stand der Technik und Anforderungen an die industrielle Kommunikation

Prof. Dr. René Simon[1]

Abstract: Virtuelle Steuerungen (Virtual PLC, vPLC) sind ein weiterer Schritt bei der Digitalisierung der Produktion sowie der damit verbundenen OT/IT-Integration. Die Steuerungsfunktionalität wird dabei räumlich und organisatorisch von der Maschine / Anlage separiert. Diese aktuelle Entwicklung beruht auf einer Reihe von Technologien, z. B. der industriellen Kommunikation, die in den vergangenen Jahren erfolgreich etabliert worden sind. Der Vortrag beschreibt die bisherige Entwicklung, den Stand der Technik sowie Herausforderungen für die zukünftige Weiterentwicklung von virtuellen Steuerungen. Dabei werden neue Anforderungen an die Kommunikation bzgl. Verfügbarkeit und Echtzeitfähigkeit gestellt.

[1] Hochschule Harz, Friedrichstraße 57-59, 38855 Wernigerode, Deutschland
rsimon@hs-harz.de

http://dx.doi.org/10.25673/121097

Dezentrale Authentifizierung und Zugangskontrolle in OPC UA mittels Self-Sovereign Identity und Decentralized Identifier

Alex Korn#, Duy Lam Tran* und Matthias Riedl*

#IoTIQ , Emilienstraße 13, 04107 Leipzig, alex@iotiq.de

*ifak, Werner-Heisenberg-Str. 1, 39106 Magdeburg, { duylam.tran|matthias.riedl}@ifak.eu

Abstract: In der Industrie 4.0 und dem Internet der Dinge (IoT) gewinnen Machine-to-Machine-Netzwerke (M2M) zunehmend an Bedeutung, wobei die Sicherheit und Flexibilität der Kommunikationsprotokolle eine zentrale Rolle spielen. Traditionelle, zentralisierte Authentifizierungsmethoden wie Benutzername/Passwort oder Public-Key-Infrastructure (PKI) stoßen in dezentralisierten Umgebungen an ihre Grenzen. Dieser Beitrag untersucht die Ablösung dieser zentralisierten Authentifizierungsverfahren durch einen dezentralen Ansatz, der auf Self-Sovereign Identity (SSI) zur Authentifizierung. Ziel dieser Arbeit ist es, die Integration dieser Technologien in das OPC UA-Protokoll zu analysieren, um die Sicherheit und Flexibilität von M2M-Netzwerken zu erhöhen und eine zukunftsfähige Lösung für die Authentifizierung und Kommunikation in Industrieanwendungen zu bieten.

1 Einleitung und Problemdarstellung

In der Industrie 4.0 und dem IoT nehmen M2M-Netzwerke eine immer bedeutendere Rolle ein. Diese Netzwerke bestehen aus einer Vielzahl vernetzter Geräte, die autonom kommunizieren, Daten austauschen und Entscheidungen treffen. Die Sicherheit und Flexibilität der Kommunikationsprotokolle sind dabei von entscheidender Bedeutung, um eine zuverlässige und vertrauenswürdige Interaktion zwischen diesen Geräten zu gewährleisten. In traditionellen, zentralisierten Authentifizierungssystemen wie der Verwendung von Benutzernamen und Passwörtern oder PKI stoßen diese Verfahren jedoch schnell an ihre Grenzen, insbesondere in dezentralen Umgebungen wie IoT-Netzwerken, in denen zahlreiche Geräte agieren, die keine feste, zentrale Identität haben [1].

Blockchain, eine Datenbank, in der Daten dezentral über den Konsens der Teilnehmer eines Peer-Netzwerks verwaltet werden, passt gut zur Architektur eines IoT-Netzwerks [2]. Während es viele Forschungsrichtungen zur Anwendung von Blockchain in M2M-Netzwerken gibt, konzentriert sich diese Arbeit auf den Aspekt der Authentifizierung und Zugriffsverwaltung. Um eine dezentrale Authentifizierung zu erreichen, wird der Ansatz der SSI verwendet, der es jedem Gerät oder jeder Entität ermöglicht, die Kontrolle über seine Identität und die damit verbundenen Informationen zu behalten [3]. Begleitend zur SSI-Architektur verwenden wir Decentralized Identifier (DID) zur Identifizierung der Teilnehmer im Netzwerk.

Die in dieser Arbeit entwickelte Lösung wird dann mit Open Platform Communications Unified Architecture (OPC UA) implementiert. OPC UA wird in der Industrie, insbesondere im Zusammenhang mit Industrie 4.0, häufig als Kommunikationsprotokoll verwendet [4]. Angesichts der wachsenden Bedeutung von OPC UA muss dessen Sicherheitsaspekt entsprechend berücksichtigt werden. Das Ziel besteht darin, SSI/DID für die Authentifizierung zwischen Client und Server zu verwenden und dem Client je nach seiner Rolle bestimmte Zugriffsrechte zu gewähren.

2 Vorgeschlagene Architektur

Eine digitale Identität wird durch einen eindeutigen DID beschrieben. Informationen zu einer Identität werden in Form von Verifiable Credentials (VCs) dargestellt. Diese überprüfbaren Zertifikate können beliebige identitätsbezogene Angaben enthalten, beispielsweise das Alter einer Person, den höchsten Bildungsabschluss oder eine Mitgliedschaft in einer Organisation.

Die zugrunde liegende Architektur (Abbildung 1) folgt dem sogenannten Triangle of Trust, das aus den drei Rollen Issuer, Holder und Verifier besteht. Der Issuer ist eine Instanz (Person, Organisation oder System), die VCs ausstellt und kryptografisch signiert. Der Holder bewahrt diese VCs in einem Wallet, erzeugt daraus Zero-Knowledge-Proofs und stellt diese bei Bedarf einem Verifier zur Verfügung. Der Verifier fordert entsprechende Nachweise an und prüft, ob die vorgelegten VCs von einem vertrauenswürdigen Issuer ausgestellt wurden und die gewünschten Kriterien erfüllen. [5]

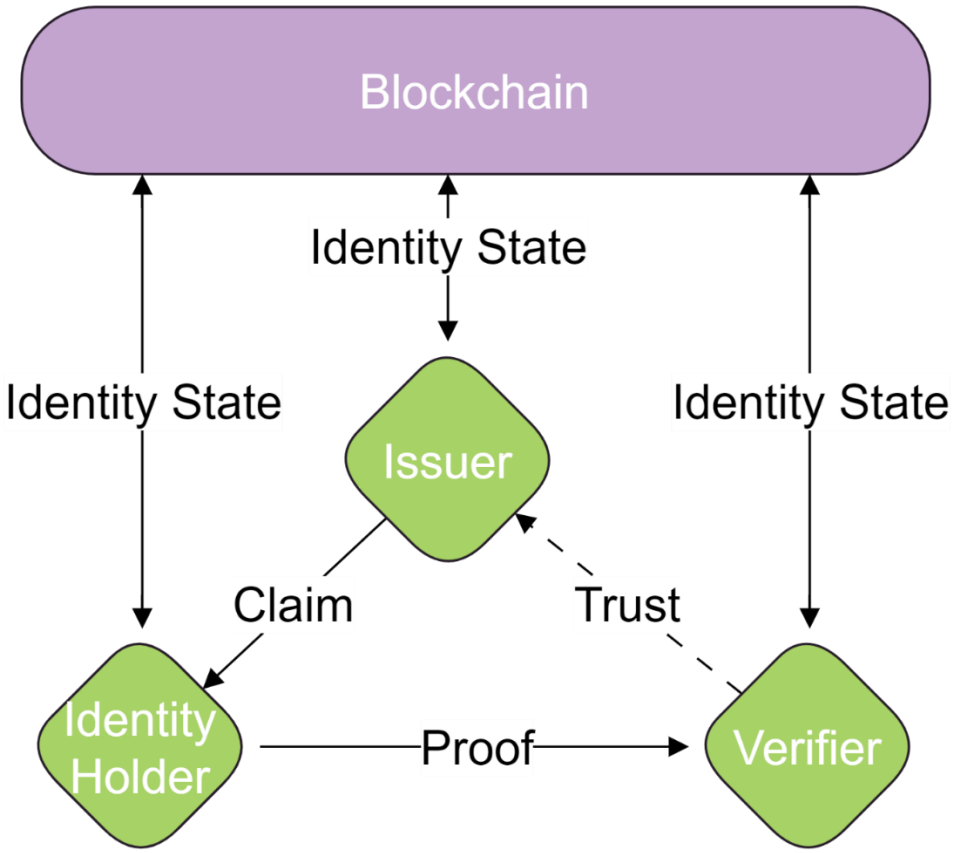


Abbildung 1: Vorgeschlagene Systemarchitektur auf Basis des Vertrauensdreiecks

Ein Credential umfasst dabei die notwendigen Daten, um nachzuweisen, dass eine bestimmte Aussage (Claim) von einer spezifischen Identität ausgestellt und von einer anderen Identität gehalten wird. Es setzt sich aus dem eigentlichen Claim und dem zugehörigen kryptografischen Proof zusammen.

Die Blockchain dient als dezentrales, manipulationssicheres Register, in dem ausschließlich öffentliche Informationen zur Validierung von Credentials gespeichert werden. Anstelle personenbezogener Daten wird lediglich ein kryptografischer Hash des aktuellen Identitätszustands (Identity State) auf der Blockchain abgelegt, der mit Zeitstempel und Block verknüpft ist [6].

Der Identity State entspricht der Wurzel eines Hash-Baums (Merkle Tree)(Abbildung 2) [7], der sich aus den drei Hashes Claims Tree Root, Revocation Tree Root und Roots Tree Root zusammensetzt.

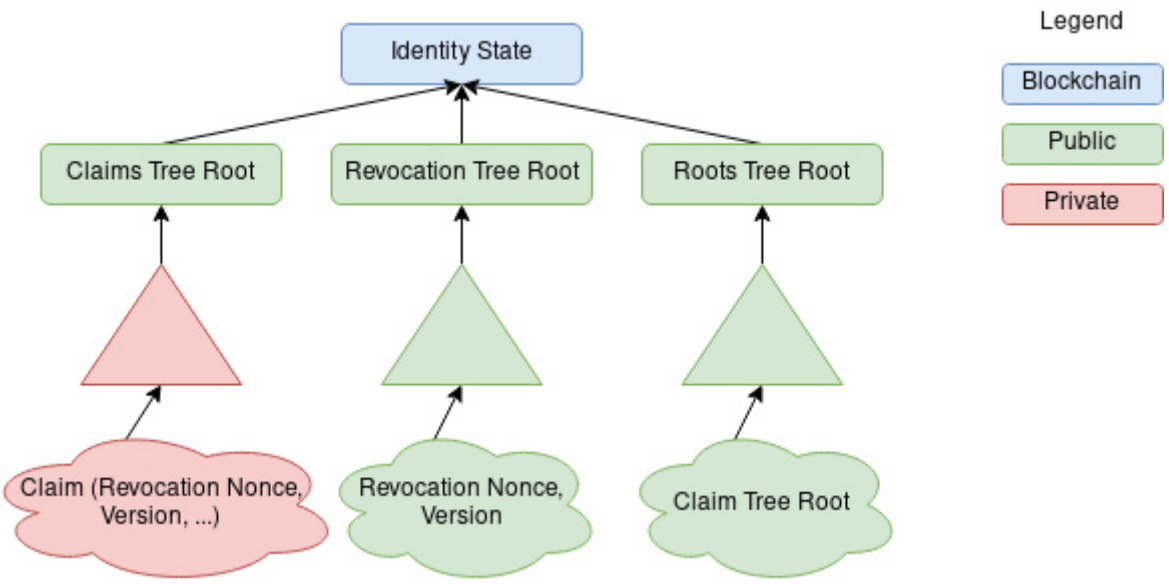


Abbildung 2: Identitätszustands-Hashbaum

Zum Widerruf eines Credentials fügt der Issuer eine zufällige Zahl (Nonce) in den Revocation Tree ein, berechnet den Identity State neu und veröffentlicht diesen. Voraussetzung hierfür ist, dass sowohl die Wurzel-Hashes als auch die Revocation- und Roots Trees öffentlich einsehbar sind. Auf dieser Grundlage kann der Holder den Merkle Tree eigenständig rekonstruieren, nachweisen, dass ein Credential nicht widerrufen wurde, und gleichzeitig bestimmte private Claims verbergen [6].

Um den aktuellen Revocation Status eines Credentials zu ermitteln, muss der Holder den Issuer nach dem neuesten Stand befragen. Um die Abhängigkeit von einem einzelnen Issuer zu vermeiden, können mehrere Reverse Hash Services (RHS) [8] eingesetzt werden, die die relevanten Informationen replizieren und so im Falle eines Ausfalls alternative Abfragewege bereitstellen. Nach einem Widerruf kann der Verifier mithilfe der drei Wurzel-Hashes überprüfen, ob der vom Issuer veröffentlichte State weiterhin gültig ist und ob das Credential noch als vertrauenswürdig anzusehen ist. Lässt der Holder die Nonce aus dem Revocation Tree weg, verändert sich der Wurzel-Hash und damit auch der Identity State. In diesem Fall vergleicht der Verifier den abgeleiteten Identity State mit dem veröffentlichten Wert, erkennt die Abweichung und verweigert den Zugang.

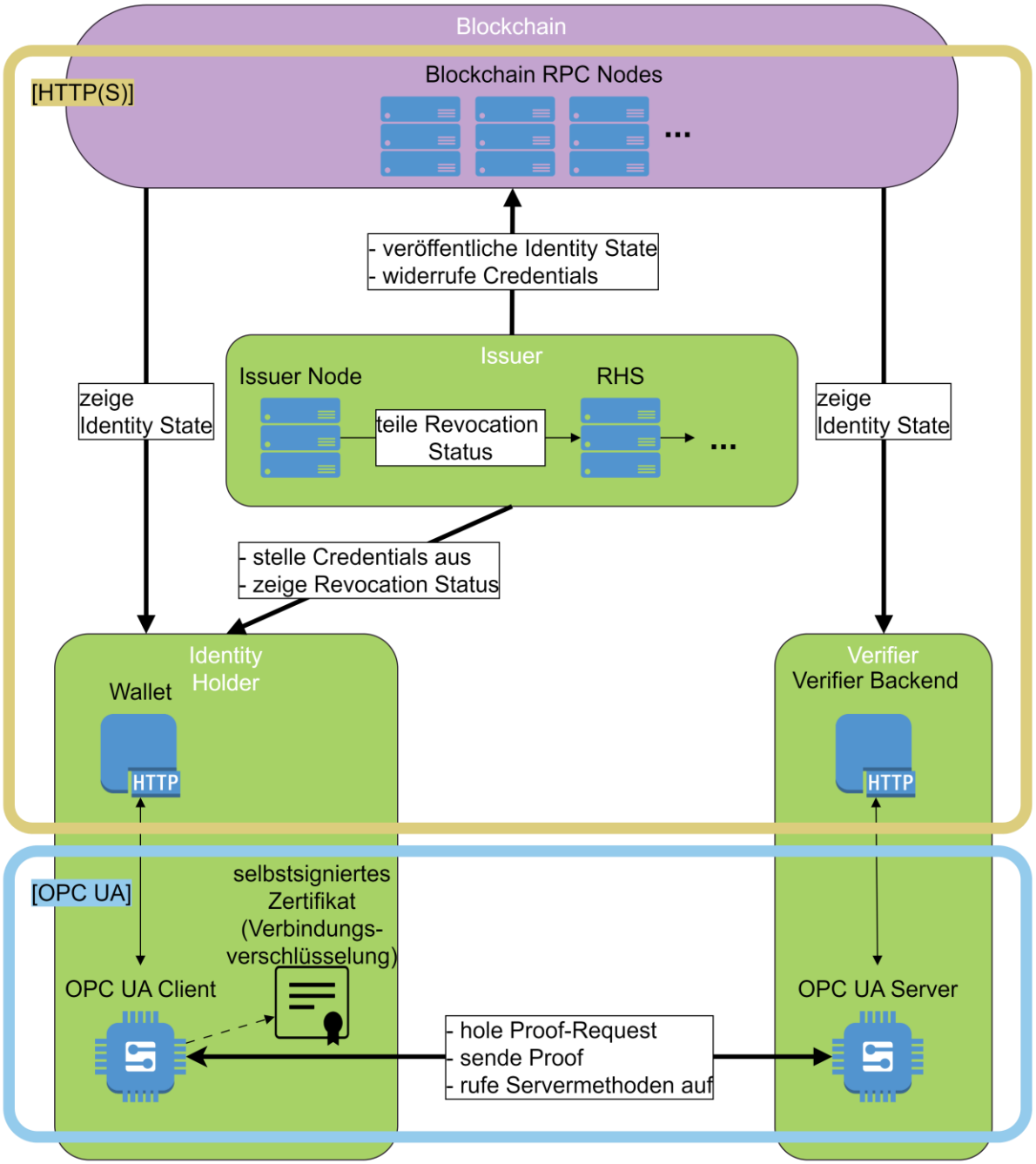


Abbildung 3: Detaillierte Architektur mit SSI und OPC UA

3 Umsetzung

Für die Umsetzung wird die C-Bibliothek open62541 v1.4.6 [9] verwendet, die auf dem Standard IEC 62541 (OPC UA) basiert. OPC UA ist ein etabliertes Protokoll für industrielle Kommunikation. Es setzt auf ein asynchrones Nachrichtenprotokoll über sichere Kommunikationskanäle auf, definiert ein Typsystem mit binärer und XML-Kodierung, stellt ein objektorientiertes Metamodell zur Informationsmodellierung bereit und umfasst 37 Standardservices zur Interaktion mit serverseitigen Informationsmodellen [10].

Für die Authentifizierung von Benutzern unterstützt OPC UA verschiedene UserIdentityTokens, darunter Benutzername/Passwort, X.509 v3 Zertifikate und JSON Web Tokens (JWT). Session-lose

Dienste können zudem ein AccessToken nutzen, das von einem AuthorizationService bereitgestellt wird, dabei ist ein verschlüsselter Kommunikationskanal erforderlich [11]. In der open62541-Bibliothek sind allerdings session-lose Services wie OPC UA HTTPS und SOAP-HTTP nicht implementiert [12].

Die Protokollarchitektur erlaubt drei SecurityModes: None, Sign und SignAndEncrypt. Auch wenn im Modus None keine Verschlüsselung eingesetzt wird, muss stets ein logischer Kanal mit eindeutiger SecureChannel-ID bestehen [13]. In den Modi Sign und SignAndEncrypt erfolgt die Sicherung der Kommunikation mithilfe von X.509-Zertifikaten. Vor dem Verbindungsaufbau erzeugen Client und Server jeweils selbstsignierte Zertifikate, die primär der Transportverschlüsselung dienen, jedoch nicht als vertrauenswürdig gelten. Das Vertrauen wird mit SSI sichergestellt. Der anschließende Handshake folgt einem standardisierten Ablauf gemäß der Spezifikation (Abbildung 4).

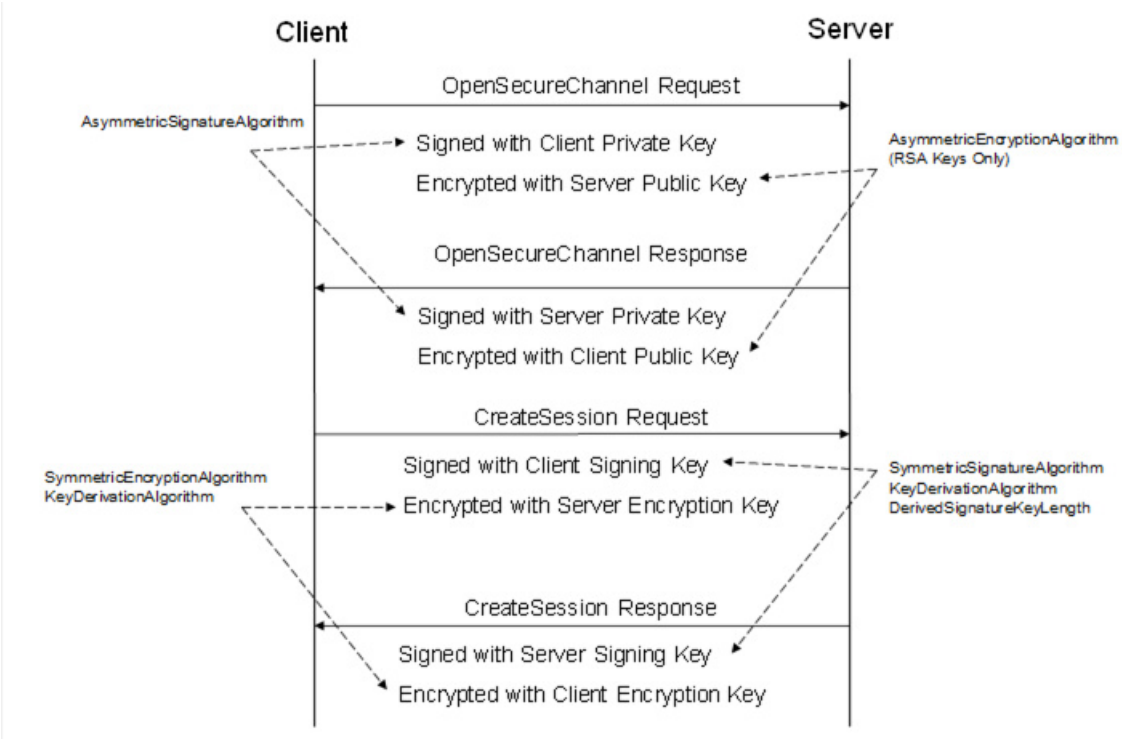


Abbildung 4: Standard-Handshake für OPC UA Secure Channel [13]

Für die Integration von SSI stellt sich eine besondere Herausforderung: Im Gegensatz zum SSI-Verfahren auf HTTP-Basis wie in iden3-Login-Workflow abgebildet (Abbildung 5) muss der OPC UA-Client bereits beim Verbindungsaufbau authentifiziert werden, wodurch ein dynamischer Austausch von Proof-Requests (generate ZK request) und Proofs (callback ZK proof) mit dem Server nicht möglich ist.

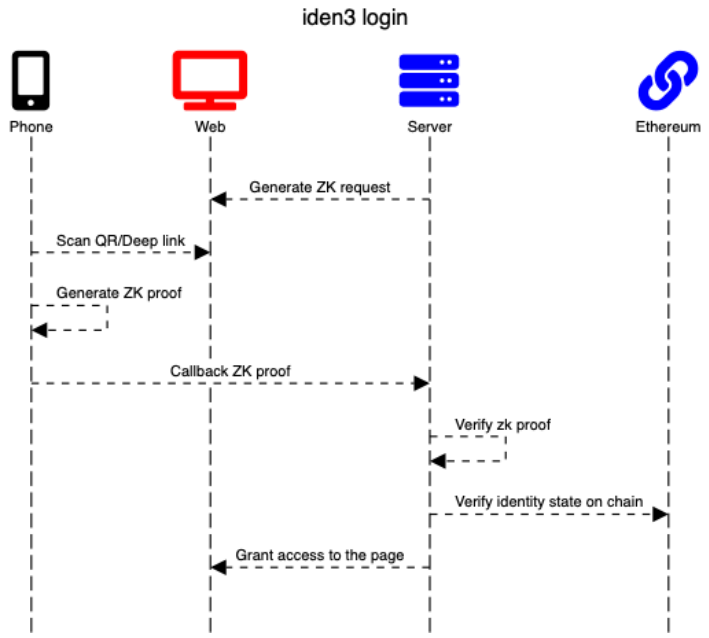


Abbildung 5: iden3-Login-Workflow [14]

Während eine Umsetzung mit statischen Proof-Requests denkbar wäre, scheitert ein flexibles SSI-Verfahren an dieser Stelle ohne tiefgreifende Änderungen der Bibliothek.

Eine alternative wäre die Nutzung von Authorization Service [15], der die SSI-basierte Authentifizierung mit Rückgabe eines AccessTokens implementiert und der Client diesen AccessToken nutzt, um sich mit dem Server, der die eigentliche Geschäftslogik umsetzt zu verbinden. Diese Architektur würde jedoch die Systemkomplexität erhöhen.

Aus praktischen Gründen wurde daher eine Lösung auf Basis des bestehenden AccessControl-Plugins von open62541 gewählt. Hierbei werden eigene Callbacks implementiert, die vor jedem Methodenzugriff prüfen, ob ein Client bereits einen gültigen Proof erbracht hat. Auf diese Weise kann der Zugriff solange eingeschränkt werden, bis eine erfolgreiche Authentifizierung stattgefunden hat. Ein Nachteil dieser Methode besteht darin, dass Clients zunächst eine gültige Verbindung zum Server aufbauen können, wodurch bei unzureichender Absicherung kritische Methoden unbeabsichtigt freigegeben oder Denial-of-Service-Angriffe durch eine Vielzahl von Verbindungen ermöglicht werden könnten. Bei Nutzung der zuvor beschriebenen Variante mit einem Authorization Service würde dieser Nachteil wesentlich eingeschränkt.

Für die Umsetzung des SSI-Mechanismus wird das Framework PrivadoID (vormals PolygonID) genutzt, das auf dem iden3-Protokoll basiert. Die Rollen Issuer (Issuer Node) [16], Verifier (Verifier Backend) [17] und Holder (Wallet) [18] sind durch entsprechende SDKs abgebildet in der vorgeschlagenen Architektur. Holder und Verifier sollen untereinander ausschließlich über OPC UA kommunizieren. Für den Zugriff auf die Blockchain, konkret das Polygon-Netzwerk, werden externe RPC Node Provider wie Alchemy [19] oder Infura [20] eingesetzt, die HTTPS-API-Endpunkte für Blockchain-Abfragen bereitstellen. Alternativ ist der Betrieb eigener Nodes möglich, jedoch mit hohen Ressourcenanforderungen und zusätzlichem Wartungsaufwand verbunden.

Zur Umsetzung einer an den iden3-Login-Workflow angelehnten Authentifizierung im OPC UA-Server wurden zwei Methoden implementiert: getProofRequest und sendProof. Diese Methoden leiten die Anfragen an ein Verifier-Backend weiter, das wiederum über HTTP(S) und einen RPC Node die

Gültigkeit der vorgelegten VCs überprüft. Bis zur erfolgreichen Präsentation eines gültigen VCs verhindert das im Server integrierte AccessControl-Plugin sämtliche Zugriffe auf weitere Funktionalitäten. Ausgenommen hiervon sind lediglich die Authentifizierungsmethoden getProofRequest und sendProof. Der angepasste Authentifizierungsablauf ist in der Abbildung 6 dargestellt.

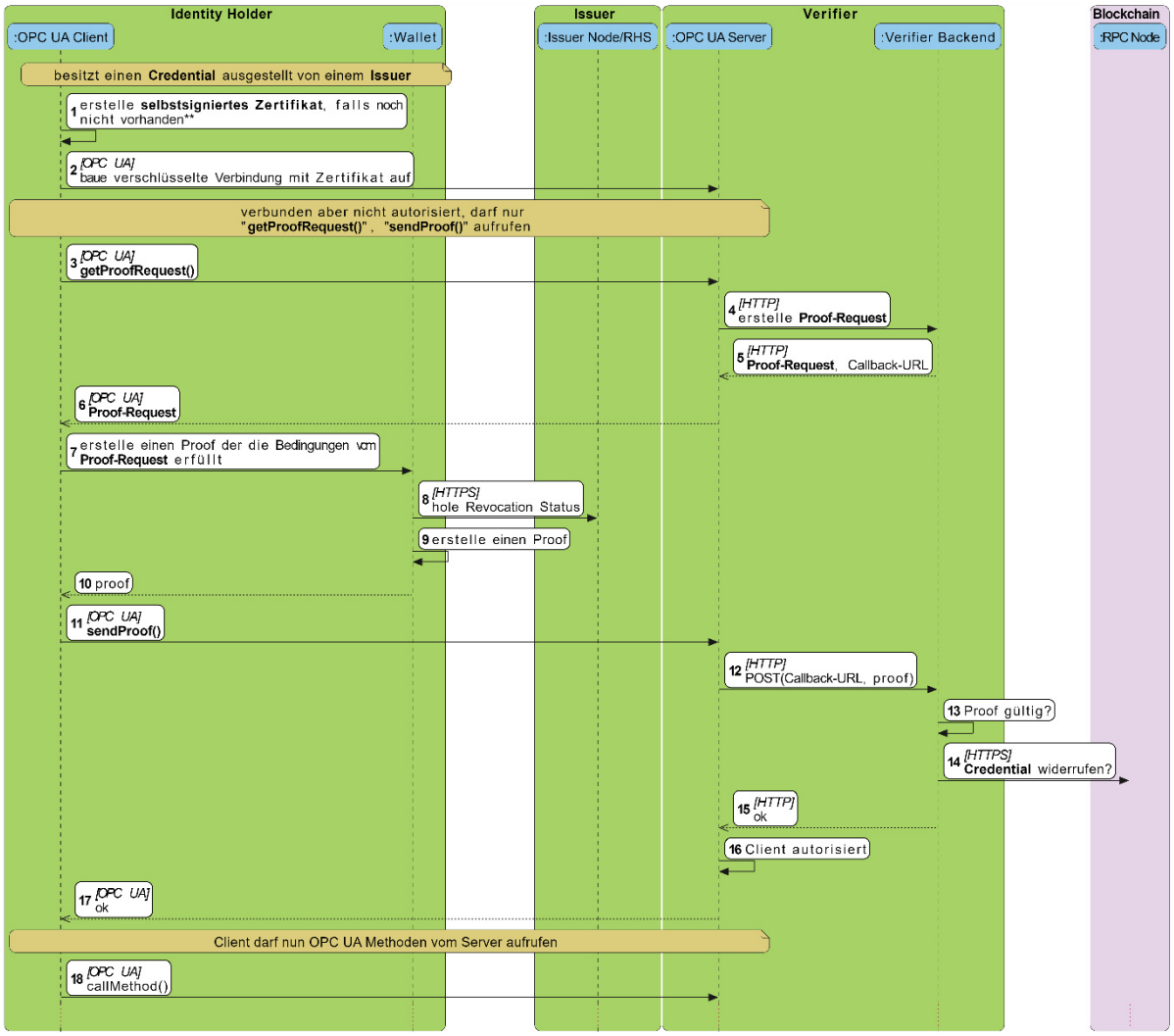


Abbildung 6: Authentifizierungsablauf

4 Bewertungskonzept

Die vorgeschlagene Architektur mit vielen Komponenten erfordert ein intensives Bewertungskonzept. In diesem Abschnitt werden die allgemeinen Hauptaspekte für unsere Bewertung aufgeführt.

Ausstellung von Verifiable Credentials (VCs): Der Issuer verwendet seinen privaten Schlüssel, um die Nachweisen zu signieren, während sein öffentlicher Schlüssel vom Verifier verwendet wird, um die Signaturen der Nachweisen zu validieren. Diese Tabelle listet die Kriterien auf um die Authentizität des Issuer und die Integrität der Nachweise sicherzustellen.

Kriterien	Beschreibung	Erfolgsmetriken

VC-Formatkonformität	Sicherstellen, dass VC den W3C-Standards entsprechen	Das Format der Nachweise ist korrekt (JSON-LD, JSON-Schema)
Sicherheit der Ausstellerschlüssel	Sicherstellen, dass die privaten Schlüssel des Ausstellers sicher verwaltet werden	Die Schlüssel werden sicher gespeichert und nicht weitergegeben
Integrität der Nachweise	Sicherstellen, dass die Nachweise kryptografisch signiert und verifizierbar sind	Die Nachweise sind durch Dritte verifizierbar (Validierung der Ausstellersignatur)

Vorlage von VCs und Überprüfung von Proofs: Wenn der Holder eines Berechtigungsnachweises diesen dem Verifier vorlegen möchte, verwendet er einen Zero-Knowledge Proof (ZKP), der die Gültigkeit seines Berechtigungsnachweises belegt, ohne sensible Daten preiszugeben. Diese Tabelle listet die Kriterien auf um sicherzustellen, dass ZKPs ordnungsgemäß generiert werden und vom Verifier validiert werden können. Die Privatsphäre des Holders sollte auch während der Vorlage von Ansprüchen gewährleistet sein.

Kriterien	Beschreibung	Erfolgsmetriken
Proofkonstruktion	Korrektheit von ZKPs sicherstellen	ZKPs sind korrekt aufgebaut und geben keine sensiblen Daten preis
Proofverifizierung	Überprüfbarkeit von ZKPs sicherstellen	ZKPs sind gültig (mit aktueller Implementierung, ohne Blockchain)
Datenschutzgarantien	Sicherstellen, dass ZKPs keine sensiblen Daten preisgeben	ZKPs schützen die Privatsphäre, bei der Überprüfung der Ansprüche werden keine Informationen preisgegeben

Berechtigungswiderruf über Blockchain: Wenn ein Credential widerrufen wird, wird diese Änderung in der Blockchain über den Identity State widergespiegelt, sodass der Verifier die Gültigkeit des Credentials abfragen und überprüfen kann. Diese Tabelle listet die erforderlichen Kriterien für den Widerruf von Credentials auf.

Kriterien	Beschreibung	Erfolgsmetriken
Verfügbarkeit des Identity State	Sicherstellen, dass der Identity State in der Blockchain verfügbar ist	Der Identity State ist öffentlich zugänglich und Abfragen liefern korrekte Statusmeldungen
Aktualisierung des Identity State	Sicherstellen, dass der Identity State auf dem neuesten Stand ist	Der Identity State wird nahezu in Echtzeit aktualisiert
Integrität des Identity State	Sicherstellen, dass der Identity State nicht manipuliert werden kann	Der Identity State kann nicht durch nicht autorisierte Komponenten verändert werden
Leistung von Sperrungsabfragen	Bewertung der Leistung von Sperrungsabfragen	Die Abfrage von Sperrungen ist für den Verifier schnell und effizient, auch wenn die Blockchain öffentlich ist

Off-Chain-Datenverwaltung: In dieser Arbeit werden die Credentials der Benutzer außerhalb der Blockchain verwaltet und in einer privaten Datenbank gespeichert. Diese Tabelle listet die Kriterien auf um Datenintegrität, Sicherheit und Zugriffskontrolle zu gewährleisten.

Kriterien	Beschreibung	Erfolgsmetriken
Schutz privater Daten	Sicherstellen, dass die Off-Chain-Datenspeicherung verschlüsselt und sicher ist	Off-Chain-Daten werden verschlüsselt und sicher gespeichert
Zugriffskontrolle	Sicherstellen, dass der Zugriff auf Berechtigungsdaten korrekt eingeschränkt ist	Nur autorisierte Benutzer können auf Berechtigungsdaten zugreifen oder diese ändern

Vertrauensmodelle & PKI: Diese Arbeit basiert teilweise auf einem Vertrauensmodell mit PKI. Diese Tabelle listet die Kriterien auf um die Vertrauenswürdigkeit des Issuers und eine sichere Schlüsselverwaltung zu gewährleisten.

Kriterien	Beschreibung	Erfolgsmetriken
Vertrauenswürdigkeit des Issuers	Vertrauensmodell für Issuer überprüfen	Der mit dem Issuer verbundene öffentliche Schlüssel ist vertrauenswürdig und wurde nicht manipuliert
Schlüsselverwaltung	Bewertung der Verteilung von Schlüsseln und Credentials	Öffentliche Schlüssel und Credentials von Issuer werden sicher an Verifier und Benutzer verteilt

Datenschutz und Sicherheit: Diese Tabelle listet die Kriterien auf um die insgesamt gesicherte Interaktion zwischen den Komponenten im System und die Zugriffskontrolle auf der Grundlage des vom Benutzer vorgelegten Nachweises zu gewährleisten.

Kriterien	Beschreibung	Erfolgsmetriken
API-Sicherheit	Sicherstellen, dass APIs frei von Schwachstellen sind (z. B. SQL-Injection (SQLi), Cross-Site-Scripting (XSS))	API-Endpunkte sind vor gängigen Angriffen geschützt
Sichere Kommunikation	Sicherstellen, dass der Datenaustausch zwischen Komponenten vor Man-in-the-Middle-Angriffen geschützt ist	Alle Kommunikationskanäle sind gesichert und verschlüsselt
Zero-Knowledge-Integrität	Überprüfen, ob ZKP-Systeme die Privatsphäre während der Überprüfung von Credentials wahren	Keine unbeabsichtigte Datenlecks während der Überprüfung von Nachweisen
OPC UA Sitzungsbindung	Sicherstellen, dass Sitzungen an sichere Kanäle gebunden sind	Sitzungsbindung ist aktiviert
OPC UA-Sitzungs- und Wiederverbindungsrichtlinien	Sicherstellen, dass der OPC UA-Server so konfiguriert ist, dass er gegen DDoS-Angriffe geschützt ist	Die Parameter werden entsprechend konfiguriert, um eine Erschöpfung der Ressourcen zu verhindern

		(maximale Sitzungsanzahl, maximale Kanäle, Sitzungszeitlimit, Wiederverbindungsrate ...)
Einschränkungen für OPC UA Methodenaufrufe	Validierung des Ergebnisses der Autorisierung für Methodenaufrufe	Der Client darf nur Methoden auf dem Server aufrufen, die den vorgelegten Nachweisen entsprechen
OPC UA Variable Schreibberechtigungen	Validierung des Ergebnisses der Autorisierung für variable Schreibberechtigungen	Nur Clients mit bestimmten Nachweisen können in bestimmte Variablen schreiben

Einhaltung von Standards:

Kriterien	Beschreibung	Erfolgsmetriken
OPC UA-Spezifikationskonformität	Sicherstellen, dass die Implementierung mit OPC UA Teil 2 konform ist	Die Implementierung entspricht dem OPC UA-Standard
IEC 62443-Konformität	Sicherstellen, dass die Architektur mit den industriellen Cybersicherheitsstandards übereinstimmt	Zonen/Kanäle werden definiert

5 Bedeutung des Beitrags & zukünftige Arbeiten

Ziel dieser Arbeit ist es, die Auswirkungen der Integration von SSI/DID und Blockchain-basierten Identitätslösungen in das OPC UA-Protokoll zu untersuchen und eine zukunftssichere, sichere und flexible Lösung für die Authentifizierung und Autorisierung in industriellen Anwendungen vorzustellen. Diese Technologien bieten das Potenzial, die Sicherheit von M2M-Netzwerken erheblich zu erhöhen und eine nahtlose Kommunikation zwischen Geräten zu ermöglichen, ohne auf zentrale Instanzen angewiesen zu sein.

Als erste Stufe der Implementierung konzentriert sich diese Arbeit nur auf die Nutzung der SSI/DID-Architektur für die Authentifizierung, während das Vertrauensmodell zwischen Issuer und Verifier weiterhin auf dem traditionellen PKI-System basiert. Die Bereitstellung von Schlüsseln und VC war ebenfalls kein Schwerpunkt dieser Arbeit. Diese Einschränkungen eröffnen Möglichkeiten für weitere Verbesserungen in zukünftigen Projekten, in denen Blockchain auch für den Authentifizierungsprozess verwendet wird. In dieser Arbeit wurde OPC UA auch nicht tief in die SSI/DID-Architektur integriert. Der OPC-UA-Server akzeptiert TLS-Verbindungen von allen Clients, und die Authentifizierung erfolgt erst nach dem Aufbau der Verbindungen. Dies ist ebenfalls ein Punkt, der in Zukunft weiter untersucht werden muss.

Die Forschungsarbeit, die diesem Beitrag zugrunde liegt, wird vom Bundesministerium für Forschung, Technologie und Raumfahrt der Bundesrepublik Deutschland (BMFTR) im Rahmen des Projekts SUSTAIN gefördert.

6 Literatur

[1] Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. *Security, Privacy and Trust in Internet of Things: The Road Ahead*. Computer Networks, 76, 146-164, 2015. http://dx.doi.org/10.1016/j.comnet.2014.11.008

[2] Lo, S. K., Liu, Y., Chia, S. Y., Xu, X., Lu, Q., Zhu, L. Analysis of Blockchain Solutions for IoT: A Systematic Literature Review. IEEE Access, vol. 7, pp. 58822-58835, 2019.

[3] https://de.wikipedia.org/wiki/Selbstbestimmte_Identität

[4] Mühlbauer, N., Kirdan, E., Pahl, M. & Carle, G. Open-Source OPC UA Security and Scalability. 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 2020, pp. 262-269, 2020

[5] https://docs.privado.id/docs/introduction/#core-concepts-of-privado-id-verifiable-credentials-identity-holder-issuer-and-verifier

[6] https://docs.iden3.io/protocol/spec/#definitions

[7] https://de.wikipedia.org/wiki/Hash-Baum

[8] https://docs.privado.id/docs/issuer/features/#reverse-hash-service

[9] https://www.open62541.org/

[10] https://www.open62541.org/doc/v1.4.6/index.html

[11] https://reference.opcfoundation.org/Core/Part2/v105/docs/5.2.3

[12] https://github.com/open62541/open62541/blob/v1.4.6/FEATURES.md

[13] https://reference.opcfoundation.org/Core/Part6/v105/docs/6.1

[14] https://docs.iden3.io/imgs/login.png

[15] https://reference.opcfoundation.org/Core/Part4/v105/docs/6.2

[16] https://docs.privado.id/docs/category/issuer/

[17] https://docs.privado.id/docs/category/verifier/

[18] https://docs.privado.id/docs/category/wallet

[19] https://www.alchemy.com/

[20] https://www.infura.io/

http://dx.doi.org/10.25673/121097

Asynchronitätstolerante Datenintegration für IIoT-Sensoren durch Industrie 4.0 Asset Administration Shells (AAS)

Frank Hilbert [1], Santiago Soler Perez Olaya [1] und Martin Wollschlaeger [1]

Abstract: Der Einsatz von KI und maschinellem Lernen verändert Produktionsprozesse grundlegend und definiert neue Anforderungen an die Datenqualität. Die Integration heterogener IIoT-Sensoren bleibt jedoch schwierig: Ressourcenbeschränkungen, asynchrone Datenströme und fehlende Standardisierung beeinträchtigen die Qualität. Besonders Retrofit-Sensoren liefern oft lückenhafte und unsynchronisierte Daten. Digitale Zwillinge, insbesondere Asset Administration Shells (AAS), versprechen Interoperabilität. In dieser Arbeit wird das AAS-Submodell „Time Series Data“ um Mechanismen für Datenqualität, Kompensationsverfahren und virtuelle Sensorik erweitert und ein Ansatz präsentiert, der eine standardkonforme Dokumentation asynchroner Daten ermöglicht. Durch transparente Qualifier wird somit eine Grundlage für robuste, KI-taugliche Datenintegrationen gelegt.

Keywords: Verwaltungschale (AAS), Idustriellen IoT, Künstliche Intelligenz

1 Einleitung

Künstliche Intelligenz und maschinelles Lernen prägen die Industrie 4.0 durch Optimierung, vorausschauende Wartung und autonome Fertigung [2, 15]. Grundlage sind konsistente, hochwertige Datenströme. IIoT-Sensoren erfassen solche Betriebsparameter, stoßen in Retrofit-Szenarien aber an Grenzen: eingeschränkte Ressourcen führen zu asynchronen, unvollständigen Daten [10].

Klassische Zeitreihenmethoden wie ARMA-Modelle setzen äquidistante Abtastung voraus [3], während die Realität stark unterschiedliche Abtastraten und Datenlücken zeigt. Zwar sind Konnektivitätsfragen weitgehend gelöst [5], doch Standards wie AAS fehlen Mechanismen für asynchrone Zeitreihen und deren Qualitätsbewertung.

Diese Arbeit untersucht daher: *Wie lassen sich heterogene IIoT-Sensordaten über AAS standardisiert so dokumentieren, dass KI-taugliche Datensätze entstehen, trotz asynchroner Abtastung und Ressourcenlimitierungen?*

Lösungsansatz: Erweiterung des AAS Time Series Data Submodells durch Qualifier für Datenqualität, Kompensationsverfahren und virtuelle Sensorik. Damit entsteht eine standardisierte Brücke zwischen Sensoren und KI-Systemen.

1 Technische Universität Dresden, 01062 Dresden , Germany ,
frank.hilbert@tu-dresden.de, https://orcid.org/0000-0000-0000-0000;
santiago.soler_perez_olaya@tu-dresden.de, https://orcid.org/0000-0003-1247-4492;
martin.wollschlaeger@tu-dresden.de, https://orcid.org/0000-0002-4646-4455

2 Stand der Technik

Der Stand der Technik zur Datenintegration in IIoT-Umgebungen lässt sich grob in drei Forschungsfelder gliedern:

1. Konzepte der Datenqualität
2. Verfahren zur Bereinigung und Kompensation fehlerhafter Daten
3. Digitale Zwillinge mitsamt Asset Administration Shell (AAS) als Standardframework

Im Folgenden werden diese Teilbereiche vorgestellt und die bestehende Forschungslücke identifiziert.

Die Bewertung von Datenqualität basiert auf den Dimensionen Genauigkeit, Relevanz, Vollständigkeit, Konsistenz und Zugänglichkeit [15, 12]. Neuere Ansätze ergänzen IoT-spezifische Metriken wie Inter-Arrival-Time-Regularität oder Ausreißererkennung [4]. Batini et al. [2] kombinieren subjektive und objektive Verfahren und liefern damit eine Grundlage für IIoT-Anwendungen.

Data Cleaning umfasst vor allem Ausreißerbehandlung und Interpolation fehlender Werte [16]. Verfahren wie PCA oder ANN sind weit verbreitet. Studien zeigen Vorteile von Kriging für räumlich korrelierte Daten [6] und kubischer Spline-Interpolation für hochauflösende Zeitreihen [14]. Echtzeitfähige Streaming-Verfahren adressieren die Ressourcenbeschränkungen typischer IoT-Systeme [11].

Die Asset Administration Shell (AAS) bildet das digitale Kernmodell der Industrie 4.0 [7]. Das Time Series Data Submodell [9] definiert Formate und Strukturen für Zeitreihendaten, jedoch ohne systematische Beschreibung von Datenqualität. Neuere Submodelle adressieren Nachhaltigkeit und CO_2-Fußabdrücke [8]. Tools wie Eclipse BaSyx sind für Echtzeitintegration besonders geeignet [13].

Trotz bekannter Datenqualitätsmetriken fehlt ihre Integration in das AAS-Ökosystem. Aktuelle Qualifier wie ValueOriginQualifier und ValueQualityQualifier reichen für asynchrone Zeitreihen nicht aus. Insbesondere fehlen:

- Semantische Beschreibung von Abtastraten und -methoden (äquidistant vs. event-driven),
- Dokumentation von Kompensations- und Interpolationsverfahren,
- Nachverfolgung von Transformationen der Datenqualität.

Diese Lücke adressiert die Arbeit durch eine Erweiterung des Time Series Data Submodells.

3 Problemstellung

Ausgehend vom Stand der Technik wird deutlich, dass trotz vorhandener Konzepte zu Datenqualität, Kompensationsverfahren und digitalen Zwillingen zentrale Herausforderungen ungelöst bleiben. Insbesondere zeigen sich Defizite bei der Integration heterogener Sensorlandschaften, den Ressourcenbeschränkungen von IIoT-Geräten sowie bei der semantischen Beschreibung im AAS-Modell. Die folgenden Teilaspekte verdeutlichen diese Problemstellung.

3.1 Herausforderung I: Technische Heterogenität der Sensorsysteme

IIoT-Sensoren besitzen sehr unterschiedliche Abtastraten: Vibrationssensoren arbeiten im Bereich von 100 Hz, Temperatursensoren oft nur bei 0,1–1 Hz. Hinzu kommt der Unterschied zwischen äquidistanter Abtastung (konstante Intervalle) und Synchronisierung mehrerer Sensoren (zeitgleiche Werte).

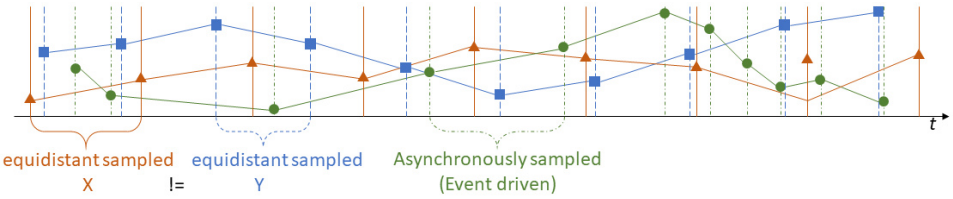


Abb. 1: Sampling-Asynchronität in IIoT-Sensoren: Abtastraten und deren Einfluss auf Korrelationen.

Verbindungsabbrüche oder Energiemanagement führen zusätzlich zu Datenlücken. Klassische Analysemethoden wie Fourier- oder Korrelationsanalysen setzen jedoch konstante Raten und synchrone Zeitstempel voraus, sodass ohne Kompensation verfälschte Ergebnisse entstehen.

3.2 Herausforderung II: Ressourcenbeschränkungen in IIoT-Umgebungen

Retrofit-Sensoren basieren oft auf Mikrocontrollern mit geringer Rechenleistung und sehr begrenztem Speicher. Batteriebetrieb erfordert lange Laufzeiten und striktes Energiemanagement. Dadurch können größere Datenmengen kaum lokal verarbeitet werden.

Die Folge: Sensoren beschränken sich auf reines Rohdaten-Forwarding, während Verarbeitungsschritte (Filterung, Interpolation, Aggregation) in leistungsfähigeren Systemen stattfinden müssen. Dies erzeugt Bursts, Latenzen und erhöht die Anforderungen an Gateways und Netzwerke.

3.3 Herausforderung III: Fehlende Semantik im Asset Administration Shell

Während aktuelle AAS-Qualifier elementare Qualitätsdimensionen adressieren, fehlt weiterhin die semantische Beschreibung von Abtastmethoden, Kompensationsverfahren und Transformationen – eine Lücke, die im folgenden Lösungsansatz gezielt geschlossen wird.

Es fehlen insbesondere:

- Dokumentation der Abtastmethoden (äquidistant, event-driven, on-change),
- standardisierte Beschreibung eingesetzter Kompensationsverfahren,
- Nachverfolgung abgeleiteter Werte und Datenqualitätstransformationen.

Auch die Definition virtueller Sensoren (z.B. Kombination von Strom- und Spannungssensor zu einem Leistungssensor) wird bisher nicht formal unterstützt. Dadurch entstehen Interoperabilitätsprobleme und eingeschränkte Nachvollziehbarkeit.

4 Lösungsansatz

Der Ansatz erweitert das AAS Time Series Data Submodellauf zwei Ebenen. Erstens, es zieht eine klare Trennung zwischen unverarbeiteten Rohdaten und abgeleiteten Zeitreihen, indem die Segmente in RawDataSegment und DerivedSegment geteilt werden. Zweitens, es ergänzt die Annotation der Daten durch neue Qualifier, die Datenqualität (*QualityQualifier*) und angewandte Methoden (*MethodQualifier*) transparent dokumentieren.

Die Architektur umfasst drei Schichten, siehe Abbildung 2:

1. Ressourcenbeschränkte IIoT-Sensoren liefern nicht-synchrone Rohdaten,
2. ein ID-IIoT-Gateway bündelt, puffert und synchronisiert diese Daten,
3. die AAS-Runtime verwaltet RawData- und DerivedSegments.

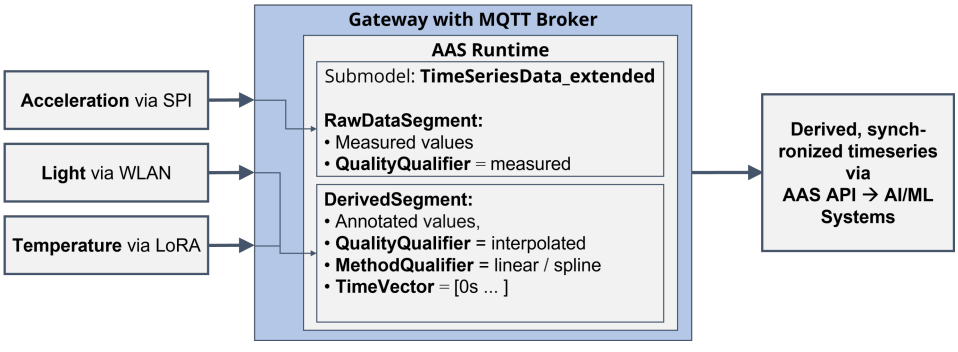


Abb. 2: Architekturübersicht: Gateway mit AAS Runtime für asynchronitätstolerante Datenintegration.

Das Gateway [5] entkoppelt die verfügbarkeit der Kommunikation mit den Sensoren von der Verfügbarkeit der Daten. Dafür dient sie als Zwishenpuffer zwischen den Sensoren und, in diesen Fall, die AAS-Runtime. Dadurch kann die AAS-Runtime jederzeit die Werte der Sensoren erreichen. Auf AAS Ebene dokumentieren **RawDataSegments** Originalwerte inklusive nativer Zeitstempel, während **DerivedSegments** synchronisierte Daten basierend auf einem definierten Zeitvektor sowie angewandten Interpolations- oder Extrapolationsverfahren bzw. virtuelle Sensorwerte enthalten.

4.1 Erweiterte Qualifier

Das bestehende AAS-Schema wird mit zwei Qualifier, QualityQualifier in Tabelle 1 und MethodQualifier in Tabelle 2. Die Kombination von beide Qualifier erlaubt eine ausführliche und maschinenfreundliche Annotierung der abgeleitete Werte. Dadurch kann Belastbarkeit der Daten jederzeit evalueirt werden. Eine Rückverfolgung der Herkunft der Daten ist auch möglich.

Tab. 1: Übersicht der QualityQualifier

Herkunft	**Syntax (QualityQualifier)**	**Besonderheiten**
Gemesenen Wert	measured	direkt von Sensoren
Interpoliert	interpolated	Wert zw. 2 gemesenen Werten
Extrapoliert	extrapolated	Geschätzten künftigen Wert
Gefiltert	filtered	Durch ARMA Angepassten Werte
Virtuellen Wert	virtual	Virtual Sensor Wert

Tab. 2: Übersicht der MethodQualifier mit Parametern

Methode	**Syntax (MethodQualifier)**	**Parameter / Besonderheiten**
Lineare Interpolation	linear	keine Parameter
Kubischer Spline	spline3-<r>	Randbedingung r $\in \{nat, clamp, ...\}$
ARMA	ARMA-<p>-<q>	Ordnungen für AR p und für MA q
PCHIP	PCHIP	keine Parameter
Virtual Sensor	virtual(<Expr>)	Arith. Komb. gemessener Werte

4.2 Implementierung der Datenverarbeitungspipeline

Das Gateway verarbeitet eingehende Daten indem Rohwerte gespeichert werden, abgeleitete Werte berechnet werden und auf Nachfrage vom AAS-Runtime, derivedSegment Werte errechnet werden. Die DerivedSegment Werte sind dann mit passenden Qualifiers versehen, sodass sowohl die qualität der Daten als auch die angewandte Rechnungen nachverfolgt werden können. Tabelle 3 bietet einen Beispielset.

Tab. 3: Beispiel: DerivedSegment Zeitreihe mit Qualifiern

Timestamp	Value	MethodQualifier	QualityQualifier
0s	21.3	–	measured
1s	22.0	spline3-nat	interpolated
2s	23.8	–	measured
3s	25.2	–	measured
4s	25.8	ARMA-2-1	extrapolated

5 Use Case und Evaluation

Zur Demonstration und Validierung des Ansatzes wurde ein repräsentatives Szenario aus der Lebensmittelindustrie simuliert.

5.1 Szenario: Intelligente Kühlkettenüberwachung

Gerade im Retrofit-Kontext, in dem kostengünstige Sensoren mit stark variierenden Abtastraten eingesetzt werden, zeigt der erweiterte Qualifier-Ansatz seine Vorteile durch standardkonforme Dokumentation und Synchronisation.

Das Szenario umfasst vier heterogene Retrofit-Sensoren: einen Temperatursensor (NTC-Thermistor, 0,1 Hz, LoRa), einen Feuchtigkeitssensor (kapazitiver Sensor, 0,2 Hz, WLAN), einen Türstatussensor (magnetischer Reed-Kontakt, event-driven, GPIO) und einen Vibrationssensor (MEMS-Accelerometer, 10 Hz, I²C).

Diese Konfiguration repräsentiert typische IIoT-Herausforderungen: drastisch unterschiedliche Abtastraten (Faktor 100), verschiedene Abtastmethodiken (kontinuierlich vs. event-driven) und variable Übertragungscharakteristika durch Funktechnologie.

Temperaturdaten werden nominell alle 10 Sekunden übertragen, können jedoch Verzögerungen von bis zu 3 Minuten aufweisen. Der Vibrationssensor überträgt kontinuierlich RMS-Werte alle 100 ms, zeigt aber bei Netzwerküberlastung Burst-Verhalten. Der event-driven Türsensor erzeugt azyklische Zeitstempel.

5.2 Implementierung und Simulation

Die Umsetzung erfolgte durch eine Erweiterung des TimeSeriesData-Submodells in einer AAS-Laufzeitumgebung, wodurch die vorgeschlagenen Qualifier und Datenstrukturen integriert wurden.

Für das Kühlkettenszenario wurde ein Zeitvektor mit 1 Hz Auflösung definiert. Das ID-IIoT-Gateway transformierte die asynchronen Eingangsdaten wie folgt:

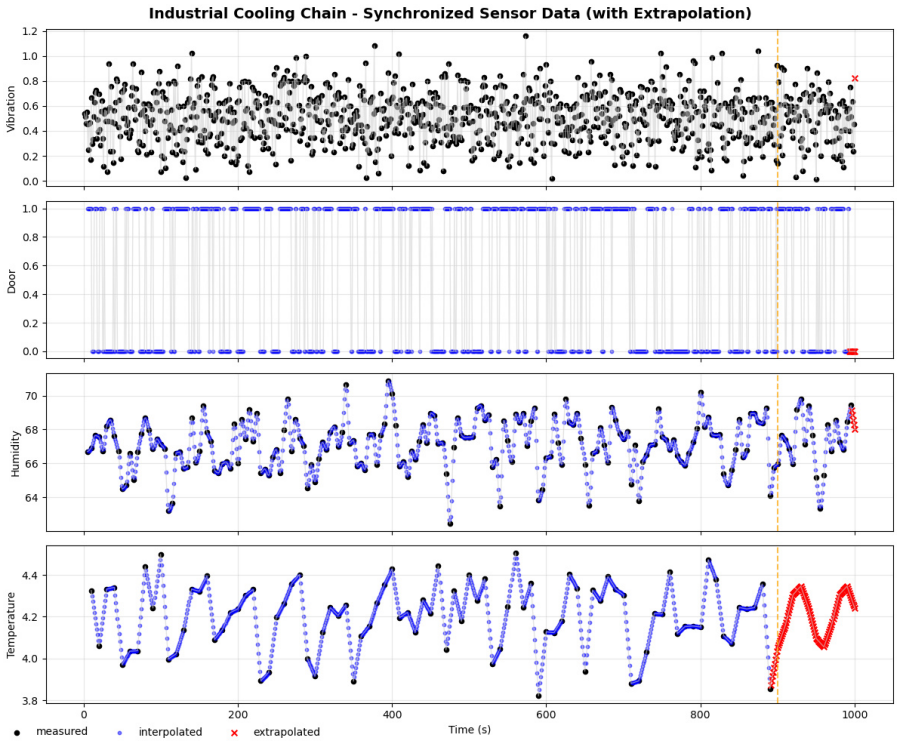


Abb. 3: Simulationsergebnisse: synchronisierte Sensordaten mit dokumentierter Datenqualität.

- **Temperaturdaten**: lineare Interpolation, Extrapolation von Lücken bis 0,5 Minuten.
- **Feuchtigkeitsdaten**: PCHIP-Interpolation für glatte Übergänge.
- **Türsensor**: Zero-Order Hold, Ereignisse als Marker dokumentiert.
- **Vibrationsdaten**: Rohwerte als RawDataSegment erhalten.

Die Python-Implementierung umfasst mehrere Kernkomponenten:

Sensorklassen Jeder Sensor wurde als Klasse modelliert, die Abtastrate, Übertragungsprotokoll und Verzögerungsverhalten abbildet. Der `TemperatureSensor` simuliert LoRa-Verzögerungen bis 180 Sekunden, der `HumiditySensor` WLAN-Latenzen bis 5 Sekunden. Der `DoorSensor` arbeitet ereignisgetrieben (event-driven) mit zufälligen Zustandsänderungen, während der `VibrationSensor` hochfrequente I²C-Daten liefert.

AAS-Erweiterung Die Klasse `AASTimeSeriesExtended` verwaltet RawDataSegments und synchronisiert diese zu DerivedSegments. Die Methode `synchronize_data()` erzeugt den Zielzeitvektor, füllt fehlende Werte per Interpolation bzw. Extrapolation und dokumentiert jeden Punkt mit Quality- und MethodQualifier.

Interpolation und Extrapolation Für kontinuierliche Variablen wird standardmäßig lineare Interpolation eingesetzt, für Feuchtigkeit PCHIP, für Ereignissensoren ZOH. Extrapolation erfolgt modellbasiert über ARMA (Temperatur: ARMA(2,2), Feuchtigkeit: MA(3)); bei unzureichender Datenbasis kommt ZOH als Fallback zum Einsatz.

Störungsszenarien Über eine Gesamtdauer von 1000 Sekunden wurden Netzwerkausfälle, zufällige Übertragungsverzögerungen und Burst-Ereignisse simuliert. In den letzten 10% der Laufzeit wurde zusätzlich der Ausfall des Temperatursensors modelliert, um die Extrapolationsfähigkeit zu prüfen.

Export und Visualisierung Am Ende wurden drei CSV-Dateien erzeugt: (1) synchronisierte Daten mit Qualifiern, (2) die ursprünglichen Messzeitpunkte und (3) die tatsächlich empfangenen Daten (mit Verzögerungen). Die Visualisierung hebt die QualityQualifier farblich hervor (schwarz: *measured*, blau: *interpolated*, rot: *extrapolated*).

Abbildung 3 zeigt den synchronisierten Datensatz der vier Sensoren (Vibration, Tür, Feuchtigkeit, Temperatur) über 1000 Sekunden. Die orange gestrichelte Linie markiert den Beginn des simulierten Ausfalls des Temperatursensors. Im oberen Diagramm (Vibration) sind Messungen nahezu lückenlos synchronisiert. Der Türsensor zeigt erwartetes binäres Verhalten, dargestellt per ZOH. Die Feuchtigkeit wurde mittels PCHIP-Interpolation in ein gleichmäßiges 1-Hz-Raster überführt. Im Temperaturverlauf treten zum Ende extrapolierte Werte auf, die durch das ARMA(2,2)-Modell erzeugt wurden.

Die Ergebnisse verdeutlichen drei Aspekte: (1) heterogene, asynchrone Datenströme lassen sich in einen synchronisierten Zeitvektor integrieren, (2) alle Transformationen werden durch Qualifier transparent dokumentiert, und (3) Sensorausfälle können modellbasiert überbrückt werden, ohne die Konsistenz des Datensatzes zu verlieren.

5.3 Beispielhafte Definition virtueller Sensoren

Über die synchronisierten Sensordaten können virtuelle Sensoren als abgeleitete Größen innerhalb des erweiterten AAS-Modells definiert werden. Dies dient als Beispiel, um die Erweiterbarkeit des Ansatzes zu verdeutlichen.

Ein **Kondensationsrisiko-Indikator** könnte Temperatur- und Feuchtigkeitswerte kombinieren, beispielsweise mittels Magnus-Formel [1]: `CondensationRisk = (Humidity / HumidityMax) * exp((17.27 * Temperature) / (Temperature + 237.3))`.

Ein **Systemstabilitäts-Index** könnte verschiedene Merkmale wie Temperaturvarianz, Vibrationsanomalien und Türöffnungsfrequenzen zu einem qualitativen Kennwert aggregieren.

Diese Beispiele zeigen, wie durch den `VirtualSensorDefinitionQualifier` komplexere Indikatoren transparent und AAS-konform beschrieben werden können.

5.4 Validierung der Datenqualität und Nachvollziehbarkeit

Die Simulation demonstrierte die Effektivität der erweiterten AAS-Qualifier. Im Temperatursensor-Ausfallszenario (letzte 10% der Simulationszeit) dokumentierte das System automatisch die betroffenen Zeitbereiche als „extrapoliertïm QualityQualifier und vermerkte die verwendete ARMA-Methode im MethodQualifier.

Die Synchronisation auf den gemeinsamen Zeitvektor ermöglichte konsistente Korrelationsanalysen. Ereignisse wie Türöffnungen mit nachfolgendem Temperatur- und Feuchtigkeitsanstieg konnten präzise als zusammenhängende Muster identifiziert werden.

Die Simulation validierte sowohl die technische Machbarkeit als auch den praktischen Nutzen der erweiterten AAS-Semantik für die Integration asynchroner IoT-Sensordaten.

6 Diskussion

Die Erweiterung des AAS-Submodells integriert erstmals Konzepte zur Datenqualität in eine standardisierte Industrie 4.0-Architektur. Vorteile sind vor allem:

- **Standardkonformität**: Erweiterungen bleiben innerhalb der AAS-Spezifikation und sind interoperabel.
- **Transparenz**: Quality- und MethodQualifier sichern Data Lineage, Nachvollziehbarkeit und regulatorische Compliance.
- **Ressourcenschonung**: Verarbeitung wird ins Gateway verlagert, Sensoren bleiben energieeffizient.
- **Migrationsfähigkeit**: RawData- und DerivedSegments können parallel genutzt werden, wodurch bestehende Anwendungen erhalten bleiben.

Die höhere Rechenlast im Gateway erfordert leistungsfähige Hardware bei großen Sensormengen. Interpolationsverfahren sind methodenabhängig, was zu Glättung oder künstlichen Oszillationen führen kann. Da das AAS-Verfahren nur die *Beschreibung*, nicht aber die

algorithmische Implementation standardisiert, besteht die Gefahr von Inkonsistenzen zwischen Implementierungen. Für virtuelle Sensoren fehlt eine klare Validierungsmethodik, da physische Referenzen nicht existieren.

Andere Arbeiten adressieren Datenqualität meist rein algorithmisch, oft proprietär und ohne Standardintegration (z.B. Echtzeitkorrektur-Frameworks [11]). Edge-Computing-Ansätze optimieren Latenz, dokumentieren aber keine Datenherkunft. Kommerzielle Digital Twin Plattformen (z.B. Azure, AWS) unterstützen Zeitreihen, jedoch ohne AAS-kompatible Qualifier. Der hier vorgestellte Ansatz kombiniert Standardkonformität mit transparenter Datenverarbeitung.

Zentrale Entwicklungsfelder sind:

- Standardisierung der neuen Qualifier im IDTA-Prozess,
- Unsicherheitsmodelle (z.B. Bayes'sche Intervalle) für Interpolation und Extrapolation,
- datengetriebene Auswahl optimaler Verfahren via Machine Learning,
- Skalierungsstudien mit tausenden Sensoren zur Bewertung von Performanz und Interoperabilität,
- Integration der Qualifier in ML-Feature-Engineering-Pipelines.

7 Fazit und Ausblick

Die Arbeit erweitert das AAS Time Series Data Submodell um Mechanismen zur semantischen Beschreibung asynchroner IIoT-Zeitreihen. Kernelemente sind die Trennung von RawData- und DerivedSegments sowie neue Qualifier für Datenqualität, Methoden und virtuelle Sensoren. Damit wird erstmals eine standardkonforme Dokumentation von Interpolation, Extrapolation und Datenherkunft (Data Lineage) ermöglicht.

Der Ansatz adressiert speziell Retrofit-Szenarien, in denen günstige IIoT-Sensoren in bestehende Anlagen integriert werden. Die Auslagerung komplexer Verarbeitung ins Gateway verlängert Batterielaufzeiten und erleichtert den Einsatz energiearmer Sensorknoten. Durch Standardkonformität bleibt Interoperabilität mit bestehenden AAS-Landschaften erhalten und unterstützt eine schrittweise Migration von Legacy-Systemen. Die hohe Transparenz schafft Vertrauen für KI-Anwendungen in sensiblen Bereichen (z.B. Pharma, Automobil, Luftfahrt).

Zukünftige Arbeiten sollten die vorgeschlagenen Qualifier formell im IDTA-Prozess standardisieren. Vielversprechend sind auch:

- Integration probabilistischer Unsicherheitsmodelle für interpolierte und extrapolierte Daten,
- adaptive, ML-basierte Methodenauswahl für Kompensationsverfahren,

- Skalierungsstudien in großindustriellen IIoT-Umgebungen,
- Anbindung automatisierter Feature-Engineering-Pipelines für KI.

Damit wird eine Grundlage geschaffen, um asynchronitätstolerante, KI-taugliche Datenintegration als festen Bestandteil zukünftiger Industrie 4.0-Architekturen zu etablieren.

Literatur

[1] O. A. Alduchov und R. E. Eskridge. „Improved Magnus Form Approximation of Saturation Vapor Pressure“. In: *Journal of Applied Meteorology and Climatology* 35.4 (1996), S. 601–609. DOI: 10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2.

[2] Carlo Batini u. a. „Methodologies for data quality assessment and improvement“. In: *ACM computing surveys* 41.3 (2009), S. 1–52. DOI: 10.1145/1541880.1541883.

[3] George EP Box und David A Pierce. *Distribution of residual autocorrelations in autoregressive-integrated moving average time series models*. University of Wisconsin. Department of Statistics, 1970.

[4] Novoneel Chakraborty u. a. „Privacy-Preserving Data Quality Assessment for Time-Series IoT Sensors“. In: *2024 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*. 2024, S. 51–57. DOI: 10.1109/IoTaIS64014.2024.10799255.

[5] F. Hilbert, S. Soler Perez Olaya und M. Wollschlaeger. „ID-IIOT – Ein servicebasierter Ansatz für das Management von eingeschränkt erreichbaren IIoT-Geräten“. In: *Automation 2019*. ISBN print: 978-3-18-092351-2, ISBN online: 978-3-18-102351-8. Düsseldorf, Germany: VDE Verlag GmbH, 2019, S. 61–74. DOI: 10.51202/9783181023518-61.

[6] Nynke Hofstra u. a. „Comparison of six methods for the interpolation of daily, European climate data“. In: *Journal of Geophysical Research: Atmospheres* 113.D21 (2008). DOI: 10.1029/2008JD010100.

[7] Industrial Digital Twin Association (IDTA). *Details of the Asset Administration Shell - Part 1: Metamodel*. Techn. Ber. IDTA-01001-3-0-1. IDTA, Juni 2024. URL: https://industrialdigitaltwin.org/wp-content/uploads/2024/06/IDTA-01001-3-0-1_SpecificationAssetAdministrationShell_Part1_Metamodel.pdf.

[8] Industrial Digital Twin Association (IDTA). *IDTA Submodels Overview*. https://industrialdigitaltwin.org/en/content-hub/submodelsd. GitHub Organization. 2024. (Besucht am 30. 12. 2024).

[9] Industrial Digital Twin Association (IDTA). *Submodel Template Time Series Data*. Techn. Ber. IDTA-02008-1-1. IDTA, März 2023. URL: https://industrialdigitaltwin.org/wp-content/uploads/2023/03/IDTA-02008-1-1_Submodel_TimeSeriesData.pdf.

[10] Amine Karkouch u. a. „Data quality in internet of things: A state-of-the-art survey“. In: *Journal of Network and Computer Applications* 73 (2016), S. 57–81. DOI: 10.1016/j.jnca.2016.08.002.

[11] Anja Klein und Wolfgang Lehner. „Real-time data quality sensors for sensor networks“. In: *Proceedings of the Third International Workshop on Data Management for Sensor Networks*. 2009, S. 1–6. DOI: 10.1145/1599776.1599778.

[12] Leo L Pipino, Yang W Lee und Richard Y Wang. „Data quality assessment“. In: *Communications of the ACM* 45.4 (2002), S. 211–218. DOI: 10.1145/505248.506010.

[13] John Smith u. a. „Comparative Analysis of Asset Administration Shell Tools for Industrial Digital Twins“. In: *Sensors* 25.7 (2025), S. 1978. DOI: 10.3390/s25071978.

[14] Petr Stepanek. „Interpolation techniques used for data quality control and calculation of technical series: an example of Central European daily time series“. In: *Időjárás* 115.1-2 (2011), S. 87–98.

[15] Diane M Strong, Yang W Lee und Richard Y Wang. „Beyond accuracy: what data quality means to data consumers“. In: *Journal of management information systems* 12.4 (1996), S. 5–33. DOI: 10.1080/07421222.1996.11518099.

[16] Fei Tao u. a. „Sensor data and information fusion to construct digital-twins virtual machine tools for cyber-physical manufacturing“. In: *Procedia manufacturing* 45 (2020), S. 86–91. DOI: 10.1016/j.promfg.2020.04.078.

http://dx.doi.org/10.25673/121097

Untersuchung der Updatefähigkeit von PROFINET IRT-Feldgeräten mit gPTP nach IEC/IEEE 60802

Kevin Heubacher[1] und Alexander Biendarra[1]

Abstract: Time Sensitive Networking (TSN) sind eine Reihe für die industrielle Automation relevanter Ethernet Bridging Standards. Das Profil IEC/IEEE 60802 TSN Profile for Industrial Automation [4] selektiert TSN Standards und stellt Anforderungen an die Hardware, Software und Konfiguration von Feldgeräten. Teil dieser Anforderungen ist die Unterstützung von mehreren Zeitdomänen für das Zeitsynchronisationsprotokoll generalized Precision Time Protocol (gPTP). Unter Berücksichtigung der Unterstützung von mehreren Zeitdomänen wird in dieser Arbeit die PROFINET IRT-Spezifikation analysiert, um Hindernisse bei einer Portierung von gPTP auf PROFINET IRT-Feldgeräte nach IEC/IEEE 60802 zu identifizieren und Lösungskonzepte zu entwickeln. Eine bestehende gPTP-Implementierung wird mit diesen Ansätzen modifiziert und auf einem PROFINET IRT-Feldgerät in Betrieb genommen. Abschließend hat eine Messung ergeben, dass die Anforderungen der IEC/IEEE 60802 bezüglich der maximalen zeitlichen Abweichung zwischen PTP Grandmaster und PTP End Instanz für die Domänen Working Clock und Global Time eingehalten werden.

Keywords: Time Sensitive Networking (TSN), 802.1AS - Timing and Synchronization, PROFINET, IEC/IEEE 60802 TSN Profile for Industrial Automation

1 Motivation

Time Sensitive Networking (TSN) sind eine Reihe von Ethernet Bridging Standards, die Ethernet mit einer deterministischen Konnektivität ausstatten sollen, wie sie in Anwendungsfällen der Automobilindustrie und industrieller Automation notwendig sind [3]. Um die Nutzung von TSN Standards im engeren Rahmen für die industrielle Automation zu definieren, arbeiten die IEEE und IEC gemeinsam an der Standardisierung des IEC/IEEE 60802 TSN Profile for Industrial Automation [4]. Basierend auf den Anwendungsfällen in der industriellen Automation selektiert das Projekt notwendige TSN Standards und stellt Anforderungen bezüglich der verwendeten Hardware und deren Konfiguration.
Einer dieser TSN Standards ist IEEE 802.1AS Timing and Synchronisation for Time-Sensitive Applications, in dem das Protokoll gPTP (generalized Precision Time Protocol) für die Zeitsynchronisation definiert wird [2]. Der Standard umfasst die Auswahl der Zeitquelle (Time Transmitter), den Transport von Zeitinformationen und das Messen von zeitlichen Verzögerungen auf dem Kommunikationsmedium. Eine der Anforderungen der IEC/IEEE 60802 ist die Unterstützung von mehreren im IEEE 802.1AS Standard definierten Zeitdomänen. Zu diesen Zeitdomänen gehören die „Working Clock“, die „Global Time“ und je eine redundante Domäne. Mit Working Clock wird eine gemeinsam etablierte

1 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Germany, kevin.heubacher@iosb-ina.fraunhofer.de; alexander.biendarra@iosb-ina.fraunhofer.de

Zeitbasis zum zeitlich koordinierten Ausführen von Anwendungen und Austausch von Daten bezeichnet. Die Global Time bezeichnet eine Zeitbasis, die die koordinierte Weltzeit (UTC) bereitstellt. Mit ihrer Hilfe wird die Nachverfolgung von Ereignissen (Sequence of Events) ermöglicht.
Viele der TSN-Standards erfordern eine Umsetzung in Hardware, was oft die Implementierung neuer Kommunikations-Chips zur Folge hat. Die Migrationszeit neuer Technologien mit einer Dauer von ca. 10 bis 20 Jahren ist lang [6], jedoch befinden sich Bestandsgeräte in aktiven Produktionsanlagen, die zumindest zu Teilen eine Hardwareunterstützung bereitstellen. Durch geeignete Konzepte für diese Bestandsgeräte kann ein wertvoller Beitrag im Bereich des Investitionsschutzes und der Nachhaltigkeit von Bestandsanlagen geleistet werden. PROFINET IRT-Feldgeräte müssen beispielsweise das Zeitsynchronisationsprotokoll PTCP (Precision Time Control Protocol) ausführen und sind mit entsprechender Hardware ausgestattet.

1.1 Problemstellung

Der IEEE 802.1AS Standard definiert funktionale Anforderungen für gPTP: Er bestimmt das Format der ausgetauschten Nachrichten und wie enthaltene Informationen verarbeitet werden. Der IEC/IEEE 60802 Standard definiert Anforderungen an die Performance von gPTP-Implementierungen und der ausführenden Hardware. Um bestehende Profinet IRT Feldgeräte mit gPTP in TSN-fähigen Netzwerken zu betreiben, ist es notwendig, die Anforderungen der IEEE 802.1AS und IEC/IEEE 60802 zu erfüllen. Dafür müssen hardware-spezifische Herausforderungen identifiziert und Lösungen entwickelt werden.

1.2 Struktur des Papers

Das Paper ist in 5 weitere Kapitel unterteilt: Kapitel 2 analysiert die Anforderungen der beiden Standards und identifiziert, dass Profinet IRT-Hardware Zeitstempel in empfangene Frames schreibt. In Kapitel 3 werden Lösungen für die überschriebenen Felder konzipiert. Eine gPTP-Implementierung mit diesen Lösungen wird auf einem Profinet IRT-Gerät in Betrieb genommen und auf die Einhaltung der maximalen zeitlichen Abweichung überprüft. Der Ansatz dafür ist in Kapitel 4 beschrieben. In Kapitel 5 werden die Ergebnisse der Messung analysiert. Abschließend wird in Kapitel 6 das Paper zusammengefasst.

2 Analyse der Anforderungen

Die IEC/IEEE 60802 legt Metriken für die Ausführung von gPTP und unterstützender Hardware fest, die von konformanten Geräten eingehalten werden müssen. Ein Teil der Anforderungen richtet sich an die verwendeten Uhrwerke. Für diese wird zum Beispiel

die maximale relative Abweichung der Nominalfrequenz von verwendeten Quarzkristallen festgelegt. Die Größe der erlaubten Abweichung basiert auf dem Typ der PTP Instanzen, die das Gerät unterstützt. Für die Ausführung von gPTP werden unter anderem maximale Verzögerungen zwischen dem Empfang und der Weiterleitung beziehungsweise Beantwortung von Sync Message und Peer Delay Messages definiert. Ein Großteil dieser Anforderungen dient dem Zweck, die maximal erlaubte zeitliche Abweichung von PTP End Instanz zur PTP Grandmaster Instanz über bis zu 99 PTP Relay Instanzen einzuhalten. Die Abbildung 1 visualisiert diesen Synchronisationspfad. Die Variablen t_m und t_s stellen jeweils die Uhrzeiten der PTP Grandmaster Instanz und der PTP End Instanz dar. Eine weitere Anforderung

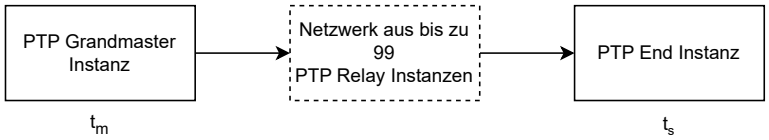


Abb. 1: Mögliche Anzahl an PTP Relay Instanzen zwischen PTP Grandmaster Instanz und PTP End Instanz

ist die Unterstützung von zwei gPTP Domänen, der Working Clock und die Global Time. Für die Working Clock Domäne gilt die Anforderung $|t_s - t_m| \leq 1\mu s$ und für die Global Time Domäne gilt $|t_s - t_m| \leq 100\mu s$.
Die Profinet Spezifikation setzt auch für PTCP das Ziel von einer maximalen Abweichung von $1\mu s$, allerdings ohne eine Angabe zur Anzahl der synchronisierten Geräte (4.2.6.4 in [5]). Für die verwendeten Quarzkristalle gibt es ebenfalls keine Anforderungen. Jede Umsetzung der Profinet Spezifikation als Gerät könnte daher mit Hardware ausgestattet sein, die diese Anforderungen nur teils oder gar nicht erfüllt. Aus diesem Grund können keine allgemeinen Aussagen zur Erfüllung der IEC/IEEE 60802 Anforderungen von Profinet IRT-Geräten gemacht werden. Das in dieser Arbeit verwendete Profinet IRT Gerät wird deshalb auf die maximale zeitliche Abweichung für Working Clock und Global Time Domäne überprüft. Skalierend mit der Anzahl an Ports und unterstützten Domänen benötigt eine gPTP-Implementierung zudem eine bestimmte Menge an Arbeitsspeicher und Rechenkapazität.
Für die Realisierung von gPTP nach IEEE 802.1AS ist es notwendig, dass die unterstützende Hardware der Software Informationen bereitstellt und von der Software gesteuert werden kann. Die verwendeten Uhrwerke müssen zum Beispiel in ihrer Geschwindigkeit und Zeit stellbar sein. Darüber hinaus ist die Zeitstempelung von ein- und ausgehenden Nachrichten erforderlich (8.4.3 in [2]). Hier hat eine Analyse der Profinet Spezifikation ergeben, dass Profinet IRT-Geräte Empfangszeitstempel in die empfangenen Frames schreiben können und sie darüber der Software zur Verfügung stellen. Eingehende gPTP Messages, für die ein Zeitstempel notwendig ist, werden somit teilweise von den Empfangszeitstempeln überschrieben. Im folgenden Abschnitt werden die überschriebenen Felder der gPTP Messages identifiziert und Lösungskonzepte entwickelt, um die teilweise überschriebenen gPTP Messages trotzdem standardgerecht verarbeiten zu können.

3 Lösungskonzepte für überschriebene gPTP Message Felder

Eine mögliche Inkompatibilität mit den funktionalen Anforderungen der IEEE 802.1AS ist, dass Profinet IRT Geräte Zeitstempel von empfangenen Frames direkt in diese Frames schreiben. Für diesen Zweck enthält die SDU eines PTCP Synchronisationsframes zwei leere 32-Bit-Felder mit den Namen `reserved_1` und `reserved_2` (4.4.1.2 in [1]). Tabelle 1

	Profinet Header	**Profinet SDU (PTCP Header)**						
Feld	FrameID	reserved_1	reserved_2	Delay_10ns	SequenceID	Delay_1ns_Byte	Padding	Delay1ns
Größe in Bits	16	32	32	32	16	8	8	32

Tab. 1: Das Frame Format eines PTCP Synchronisationsframes. Die dargestellten Felder folgen direkt auf den Ethernet-Header.

stellt die Felder eines PTCP Synchronisationsframes nach dem Ethernet-Header dar. Die Zeitstempelung eines Frames wird basierend auf dem Wert im `FrameID` Feld ausgelöst. Ist es möglich, den `FrameID` Wert zu konfigurieren, mit dem bei Empfang eine Zeitstempelung ausgelöst wird, dann ist auch eine Zeitstempelung von gPTP Messages möglich. Jedoch würden diese Zeitstempel mehrere Felder der gPTP Message überschreiben. Die Abbildung 2 stellt die Felder des gPTP-Headers vor und nach der Zeitstempelung dar. Die ersten

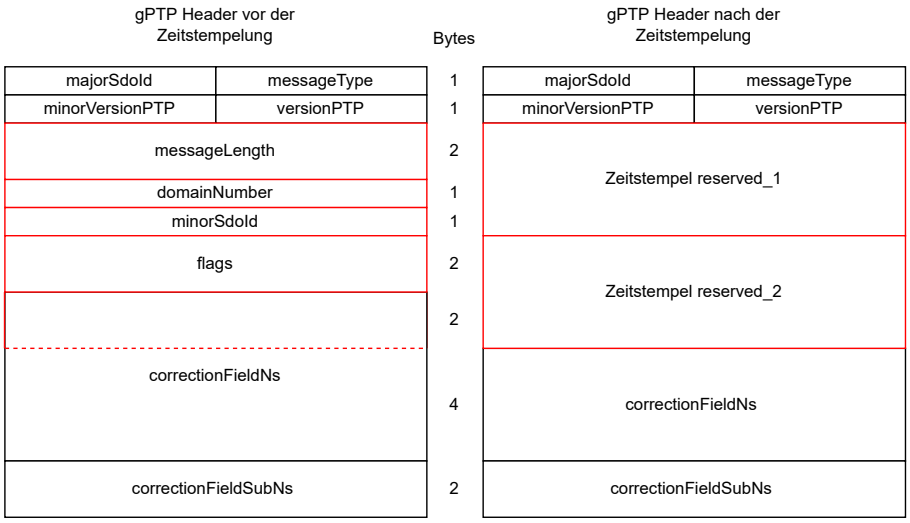


Abb. 2: Felder des gPTP Headers vor und nach der Zeitstempelung von Profinet IRT Hardware.

beiden Bytes der Message sind nicht vom Überschreiben betroffen, weil ihre Position mit der des `FrameID` Feldes überlappt. Die Felder `messageLength`, `domainNumber`, `minorSdoId` und `flags` werden vollständig von den Zeitstempeln überschrieben. Der Wert im Feld `messageLength` ist die Größe der gPTP Message in Bytes, das heißt vom Beginn des gPTP Headers bis zum letzten Feld der Message (11.4.2.3 in [2]). Eine `domainNumber` ordnet die gPTP Message einer von 255 möglichen Domänen zu (8.1 in [2]). Das Feld `minorSdoId`

ist eine Erweiterung des Feldes `domainNumber`, das für gPTP jedoch immer den Wert 0 hat. In dem Bitarray `flags` können je nach Typ der gPTP Message verschiedene Flags gesetzt werden (10.6.2.2.8 in [2]). Hier ist ausschließlich das Bit 1, was als `twoStepFlag` bezeichnet wird, relevant. Zusätzlich werden die beiden höchstwertigen Bytes des Feldes `correctionFieldNs` überschrieben. Dieses Feld enthält einen Korrekturwert für Zeitinformationen. Da diese Informationen leitend für die Verarbeitung von empfangenen Frames sind, müssen Lösungen entwickelt werden, um die verlorenen Informationen aus dem Gerät bekannten Kontext wiederherzustellen.
Zeitstempel sind nur für eine Teilmenge aller gPTP Messages notwendig, die als „Event Message“ bezeichnet werden. Event Messages sind Sync, Pdelay_Req und Pdelay_Resp Messages (11.3.2.1 in [2]). In der Tabelle 2 sind die Datentypen und möglichen Werte der überschriebenen Felder für die jeweilige Event Message aufgelistet. Die Event Messages

	Datentyp	Sync	Pdelay_Req	Pdelay_Resp
messageLength	UInteger16	34 oder 44	44	44
domainNumber	UInteger8	0 bis 255	0	0
minorSdoId	UInteger8	0	0	0
flags	Octet2	Bit 1: TRUE oder FALSE, Rest 0	Bit 1: TRUE, Rest 0	Bit 1: TRUE, Rest 0
correctionFieldNs	Integer64	-2^{47} bis $2^{47}-1$	0	0

Tab. 2: Datentyp und mögliche Werte der überschriebenen Felder von Event Messages.

Pdelay_Req und PDelay_Resp haben für jedes der überschriebenen Felder jeweils nur einen möglichen Wert. Das heißt, die enthaltenen Werte der Felder sind im Vorhinein bekannt und können in Software ohne Informationsverlust rekonstruiert werden. Der Rest dieses Abschnitts widmet sich deshalb den Ansätzen zur Rekonstruktion der Sync Message Informationen.
Anhand des `twoStepFlag` wird zum Empfang einer Sync Message entschieden, wie diese verarbeitet wird. Hat das `twoStepFlag` den Wert FALSE, ist die `messageLength` 44 Bytes und mit den in der Message enthaltenen Informationen wird die Uhrzeit des Empfängers geregelt. Hat das `twoStepFlag` den Wert TRUE, ist die `messageLength` 34 Bytes und der Empfänger erwartet eine darauf folgende Follow_Up Message, bevor die Uhrzeit des Empfängers geregelt werden kann. Die Felder der Sync Message sind: gPTP Header, `originTimeStamp` und bei gesetztem `twoStepFlag` gibt es ein zusätzliches Feld `Follow_Up_TLV` (11.4.3.1 in [2]). Aus dem `Follow_Up_TLV` ergeben sich die zusätzlichen 10 Bytes der `messageLength`. Das Feld `originTimestamp` enthält die Uhrzeit des Synchronisationsmasters zum Zeitpunkt des Sendens der Message. Ist das `twoStepFlag` der Sync Message jedoch gesetzt, ist der Wert des `originTimestamp` genullt und diese Information wird in der Follow_Up Message erwartet. Anhand der im `originTimestamp` enthaltenen Informationen kann also entschieden werden, ob `twoStepFlag` gesetzt ist. Hierfür würde man bei Empfang einer Sync Message das `originTimestamp` Feld auslesen und je nach Wert die Felder `flags` und `messageLength` rekonstruieren. Diese Methode ist fehleranfällig, da nicht garantiert werden kann, dass gPTP-Implementierungen das Feld `originTimestamp` nullen. Eine alternative Methode wäre, die Verarbeitung einer Sync Message basierend auf der Gesamtlänge des Ethernet

Frames zu basieren, falls diese Information zur Verfügung steht.
Das Feld `domainNumber` ordnet die Zeitsynchronisationsinformationen der Sync Message einer bestimmten Domäne und somit auch einem bestimmten Uhrwerk des Empfängers zu. Um diese Zuordnung auch bei Überschreibung des Feldes zu ermöglichen, muss über ein anderes Feld der Sync Message die Domäne eindeutig identifiziert werden. Ein geeigneter Kandidat hierfür ist das `clockIdentity` Feld. Die `clockIdentity` ist im Header jeder gPTP Message enthalten und identifiziert das Uhrwerk beziehungsweise die PTP Instanz des Absenders eindeutig im Netzwerk (6.4.3.6 in [2]). Eine PTP Instanz (und somit eine `clockIdentity`) kann nur einer einzelnen Domäne angehören, wodurch eine Rekonstruktion der `domainNumber` durch eine bekannte `clockIdentity` ermöglicht wird. Bevor Sync Messages ausgetauscht werden, legt der BMCA anhand von ausgetauschten Announce Messages fest, welche Ports Sync Messages senden (Time Transmitter Port) und welche sie empfangen (Time Receiver Port). Löst der Empfang einer Announce Message an einem Port die Rekonfiguration zu einem Time Receiver Port aus, kann die Kombination aus `domainNumber` und `clockIdentity` der Announce Message in einer Tabelle hinterlegt werden. Beide Felder sind in der Announce Message vorhanden, weil diese nicht zeitgestempelt wird. Empfängt der Port daraufhin eine Sync Message, kann die `domainNumber` über die Tabelle und die `clockIdentity` der Sync Message rekonstruiert werden.
Die Abbildung 3 stellt eine Topologie aus 3 Geräten mit 2 aktiven gPTP-Domänen dar. Das

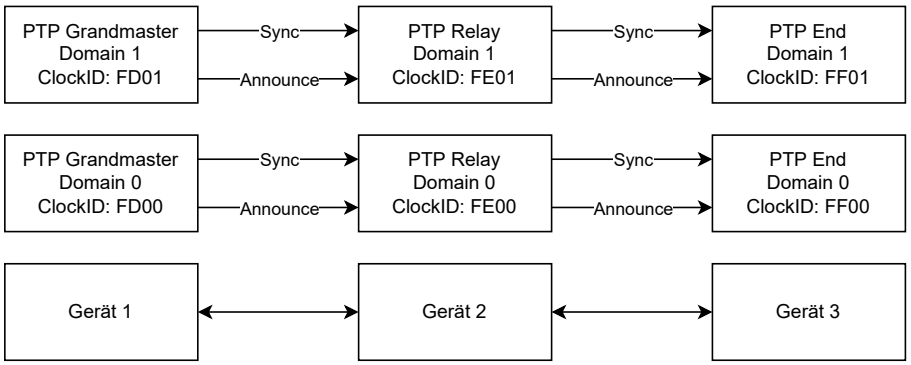


Abb. 3: Toplogie aus 3 Geräten mit jeweils 2 aktiven gPTP Instanzen. Das Gerät 1 führt PTP Grandmaster Instanzen für die Domänen 0 und 1 aus.

Gerät 1 führt die PTP Grandmaster Instanzen für beide Domänen aus. Gerät 2 empfängt Sync und Announce Messages der beiden Instanzen von Gerät 1. Durch das Verarbeiten der Announce Messages kann Gerät 2 empfangene Sync Messages über eine Tabelle der richtigen Domäne zuordnen. Über die gleiche Methode kann auch das Gerät 3 empfangene Sync Messages von Gerät 2 zuordnen. Die Zuordnungen der beiden Geräte sind in der Tabelle 3 dargestellt.
Das Feld `correctionFieldNs` ist ein vorzeichenbehaftetes 64 Bit Feld, das in zwei Bereiche aufgeteilt ist: Die höchsten 48 Bit stellen den ganzzahligen Anteil dar und die restlichen

16 Bit sind der Nachkommastellen-Bereich. Der ganzzahlige Anteil hat mit 48 Bit normalerweise einen Wertebereich von $[-2^{47}, 2^{47} - 1]$. Durch das Überschreiben der beiden höchstwertigen Bytes wird das Vorzeichen entfernt und der interpretierbare Wertebereich wird auf $[0, 2^{32} - 1]$ eingeschränkt. Für Sync Messages ist das `correctionFieldNs` je-

clockIdentity	domainNumber
FD00	0
FD01	1

clockIdentity	domainNumber
FE00	0
FE01	1

Tab. 3: Zuordnung der clockIdentity und domainNumber für empfangene Sync Frames der Geräte 2 (links) und 3 (rechts).

doch immer positiv, da der enthaltene Wert die verstrichene Zeit in Nanosekunden seit Aufnahme des Zeitstempels `originTimestamp` darstellt. Das heißt, es wird lediglich eine Einschränkung des Wertebereichs von $[0, 2^{47} - 1]$ auf $[0, 2^{32} - 1]$ verursacht. Die maximale interpretierbare verstrichene Zeit seit Aufnahme des Zeitstempels `originTimestamp` liegt somit bei $4,295$ Sekunden. Unter Betrachtung der Anforderungen sollte dieser Wert jedoch nie erreicht werden. Der Weiterleitungsprozess einer Sync Message beinhaltet die Addition der Leitungsverzögerung zum Absender und die im Gerät verstrichene Zeit auf das Feld `correctionFieldNs`. Diese beiden Zeiträume sind durch Anforderungen eingeschränkt. Die im Gerät verstrichene Zeit ist auf 15 Millisekunden (6.2.3 in [4]) und die Leitungsverzögerung auf 800 Nanosekunden beschränkt (11.2.2 in [2]). Über einen Synchronisationspfad mit 99 PTP Relay Instanzen würde dies 99 Weiterleitungen jeder Sync Message bedeuten, was einem Zeitraum von $1,485$ Sekunden entspricht.
Diese Lösungsansätze wurden in einer gPTP-Implementierung für ein Profinet IRT-Feldgerät realisiert, um die funktionalen Anforderungen der IEEE 802.1AS zu erfüllen. Im nächsten Abschnitt wird das Profinet IRT-Feldgerät auf die Einhaltung der maximalen zeitlichen Abweichung nach IEC/IEEE 60802 überprüft.

4 Ermittlung der zeitlichen Abweichung

Es gibt mehrere mögliche Ansätze, die Synchronisationsgenauigkeit von Geräten zu ermitteln. Ein analytischer Ansatz wäre, durch mathematische Modellierung die verursachte Ungenauigkeit zu approximieren. Eine weitere Methode ist die Simulation durch Modellierung der verwendeten Komponenten in Software. Eine Eigenschaft dieser Verfahren ist, dass sie unabhängig von der Inbetriebnahme der untersuchten Geräte ausführbar sind. Das heißt, durch die Implementierung des Protokolls verursachten Ungenauigkeiten können nicht erfasst werden. Deshalb wird diesen Methoden eine praktische Methode zur Ermittlung der zeitlichen Abweichung vorgezogen. Hierzu werden mehrere in Hardware und Software identische Profinet IRT Geräte (DUTs) zu einer Linientopologie verbunden und mittels der gPTP Implementierung synchronisiert. Die Abweichung der PTP End Instanz zur PTP Grandmaster Instanz $|t_s - t_m|$ wird mithilfe von PPS-Signalen ermittelt. Ein PPS-Signal wird zum Sekundenübergang des jeweiligen Gerätes in Form eines Rechteck-Impulses

ausgegeben. Diese Impulse werden von einem Oszilloskop erfasst und die zeitliche Verzögerung zwischen den Impulsen wird ermittelt. Die zeitliche Verzögerung der Impulse von PTP End Instanz zur PTP Grandmaster Instanz ergibt einen Messwert für $|t_s - t_m|$. Die Abbildung 4 stellt den Messaufbau dar. Insgesamt wird über 30 PTP Relay Instanzen,

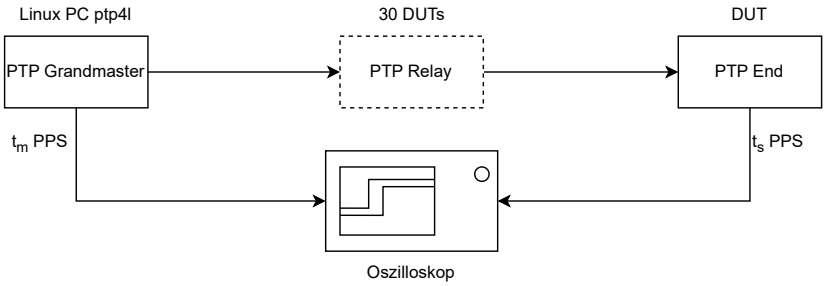


Abb. 4: Messaufbau zur Ermittlung der maximalen zeitlichen Abweichung. Ein ptp4l ausführender Linux PC agiert als Grandmaster und synchronisiert insgesamt 31 DUTs über eine Linientopologie

die jeweils von einem DUT ausgeführt werden, synchronisiert. Ein Linux-PC, der ptp4l ausführt, ist PTP Grandmaster der beiden Domänen. Für die Messung der Working Clock Domäne wird die PTP End Instanz von einem DUT ausgeführt, dies ist für die Global Time Domäne nicht möglich, weil das verwendete Uhrwerk keine Hardwareunterstützung für die Ausgabe eines PPS-Signals hat. Deshalb wird für die Messung der Global Time Domäne ein zweiter ptp4l ausführender Linux-PC als PTP End Instanz verwendet.
Um eine Aussage bezüglich einer Einhaltung der maximalen zeitlichen Abweichung zu treffen, ist es notwendig, mit den über 30 PTP Relay Instanzen ermittelten Messwerten die zeitliche Abweichung über 99 PTP Relay Instanzen zu approximieren. In einem Synchronisationspfad bestehend aus gleichartigen Geräten kann von einem linearen Verhältnis zwischen der maximalen zeitlichen Abweichung und der Anzahl der Geräte ausgegangen werden [7]. Die maximale zeitliche Abweichung zur PTP Grandmaster Instanz Δt_{max} lässt sich somit als Produkt der Anzahl der PTP Relay Instanzen N und deren maximal verursachtem Zeitfehler $t_{relay,error}$ approximieren:

$$\Delta t_{max} \approx N * t_{relay,error} \tag{1}$$

Bei bekanntem N und Δt_{max} kann dann der Beitrag einer einzelnen PTP Relay Instanz zum maximalen Zeitfehler berechnet werden mit $\frac{\Delta t_{max}}{N} \approx t_{relay,error}$. Der vom PTP Grandmaster und PTP End Instanz verursachte Beitrag zu Δt_{max} wird ebenfalls in $t_{relay,error}$ berücksichtigt, dadurch wird der Beitrag einer PTP Relay Instanz zum maximalen Zeitfehler überschätzt.
Im folgenden Abschnitt wird mithilfe der Ergebnisse des Messaufbaus ein Wert für $t_{relay,error}$ berechnet, um zu evaluieren, ob die Anforderungen der IEC/IEEE 60802 für die jeweilige Domäne eingehalten werden.

5 Analyse der Ergebnisse

Die Tabelle 4 beinhaltet die vom Oszilloskop gemessene Standardabweichung, Mittelwert, Min und Max Werte der zeitlichen Abweichung für beide Domänen über die 30 PTP Relay Instanzen. Die Daten wurden über eine Betriebsdauer von 5 Stunden erfasst, was einer Anzahl von 18000 Messwerten entspricht. Mit den maximalen Abweichungen von 133.05

Domäne	Mittelwert	Min	Max	Standardabweichung
Working Clock	41.335 ns	-73.60 ns	133.05 ns	25.532 ns
Global Time	359.30 ns	244.30 ns	473.60 ns	28.406 ns

Tab. 4: Messergebnisse der Berechnung des Abstands der PPS-Signale von PTP Grandmaster Instanz und PTP End Instanz

ns und 473.60 ns über 30 PTP Relay Instanzen ergibt sich nach Formel 1 ein $t_{relay,error}$ von ungefähr 4.5 ns und 16 ns für Working Clock und Global Time Domänen. Über die geforderten 99 PTP Relay Instanzen würde dadurch eine maximale zeitliche Abweichung von $|t_s - t_m| \approx 445.5ns < 1\mu s$ für die Working Clock Domäne und $|t_s - t_m| \approx 1.584\mu s < 100\mu s$ für die Global Time Domäne entstehen. Das bedeutet, dass für beide Domänen die Anforderungen erfüllt werden. Die Abweichungen der Global Time Domäne sind jedoch sehr viel größer als die der Working Clock Domäne. Die Qualität der verwendeten Uhrwerke und die Genauigkeit ihrer Zeitstempel sind gleich, deshalb wird diese Ungenauigkeit wahrscheinlich durch Details in der Umsetzung von gPTP verursacht. Das untersuchte Profinet IRT Gerät wurde nicht für die Unterstützung von mehreren Domänen gebaut, wodurch die in Software verfügbaren Informationen zu diesem Uhrwerk begrenzt sind.
Es gibt jedoch auch externe Einflüsse, die sich auf die Genauigkeit der Synchronisation auswirken, die in dem Messaufbau nicht berücksichtigt werden. Während der Messung wurden ausschließlich gPTP Messages übertragen, sonst gab es keinen Datenverkehr. Zusätzlicher Datenverkehr kann dazu führen, dass das Verarbeiten von Sync Messages verzögert und somit die Ungenauigkeit des Korrekturfeldes steigt. Auch Umwelteinflüsse, wie Temperaturänderungen, haben einen Einfluss auf die Genauigkeit. Der Messaufbau wurde bei einer Raumtemperatur von 20°C betrieben.

6 Zusammenfassung und Ausblick

In diesem Paper wurden die Anforderungen der IEEE 802.1AS und IEC/IEEE 60802 analysiert und Herausforderungen bei einem Update von Profinet IRT-Feldgeräten mit gPTP identifiziert. Die Analyse hat ergeben, dass gPTP Messages teilweise von Zeitstempeln überschrieben werden können. Für jedes der überschriebenen Felder wurden Lösungskonzepte entwickelt, um trotzdem eine Verarbeitung nach IEEE 802.1AS Standard zu ermöglichen. Eine gPTP Implementierung für ein Profinet IRT-Feldgerät wurde mit diesen Lösungen erweitert und dann wurde das Gerät auf die Einhaltung der maximalen zeitlichen Abweichung von PTP Grandmaster zu PTP End Instanz über 99 PTP Relay Instanzen geprüft. Dafür wurde mithilfe eines Oszilloskops die maximale zeitliche Abweichung über

30 PTP Relay Instanzen ermittelt und anschließend wurde das Ergebnis auf 99 Instanzen approximiert. Die Anforderung der zeitlichen Abweichung wurde für beide Domänen erfüllt. Eine allgemeine Aussage zur Erfüllung der Anforderungen von Profinet IRT-Geräten kann jedoch nicht getroffen werden, da die Umsetzungen der Profinet Spezifikation auf unterschiedlichen Hardwarearchitekturen basieren können. Für die Erfüllung der IEC/IEEE 60802 Anforderungen sind vor allem die Qualität der verwendeten Quarzkristalle, Anzahl der Uhrwerke, Zeitstempelmechanismus und die Informationsabrufbarkeit und Steuerbarkeit dieser Hardwarekomponenten in Software relevant. In zukünftiger Arbeit könnte eine genauere Ermittlung des Beitrags einer einzelnen PTP Relay Instanz zur maximalen zeitlichen Abweichung ermittelt werden. Hierfür definiert die IEC/IEEE 60802 Ansätze mit ausführlicheren Messaufbauten.

Literatur

[1] *IEC FDIS 61158-6-10*. PROFIBUS Nutzerorganisation. Mai 2023. URL: https://www.profibus.de/downloads/profinet-specification.

[2] Institute of Electrical and Electronics Engineers. *IEEE: Standard for Local and Metropolitan Area Networks–Timing and Synchronization for Time-Sensitive Applications*. 2020. URL: https://standards.ieee.org/ieee/802.1AS/7121/ (besucht am 15. 04. 2025).

[3] Institute of Electrical and Electronics Engineers: Time-Sensitive Networking (TSN) Task Group. 2025. URL: https://1.ieee802.org/tsn/ (besucht am 15. 04. 2025).

[4] International Electrotechnical Commission and Institute of Electrical and Electronics Engineers. *IEC/IEEE: 60802 TSN Profile for Industrial Automation*. 2025. URL: https://1.ieee802.org/tsn/iec-ieee-60802/ (besucht am 15. 04. 2025).

[5] *Profiles for decentralized periphery*. PROFIBUS Nutzerorganisation. Mai 2023. URL: https://www.profibus.de/downloads/profinet-specification.

[6] Sebastian Schriegel. *Kompatibilitätsverfahren für Profinet-Hardware mit Ethernet Time Sensitive Networks*. 978-3-662-64742-4. Springer Vieweg, 2022.

[7] Sebastian Schriegel, Lukasz Wisniewski. *Investigation in automatic determination of time synchronization accuracy of PTP networks with the objective of Plug-and-Work*. Sep. 2014. URL: https://ieeexplore.ieee.org/document/6948525 (besucht am 15. 04. 2025).

http://dx.doi.org/10.25673/121097

Integration von Digitalen Zwillingen und PLCs

Softwaretechnische Perspektiven für die Industrie 4.0

Nico Braunisch [1], Santiago Soler Perez Olaya [2], Uwe Schmidt [3], Martin Wollschlaeger [4]

Abstract: Die Eröffnung des vollen Potenzials von Industrie 4.0 erfordert ein nahtloses Zusammenwirken zwischen modernen speicherprogrammierbaren Steuerungen (SPSen) und der transformativen Datenstrukturierung der digitalen Zwillinge von Industrie 4.0 in ihrer interoperablen Form der Asset Administration Shell (AAS). In diesem Beitrag wird die Softwareentwicklung in modernen SPS-Umgebungen für digitale Zwillinge der Industrie 4.0 untersucht. Zu diesem Zweck wird durch eine qualitative Analyse, wie AAS-Instanzen als digitale Zwillinge in SPSen unter den Aspekten des Software-Engineerings implementiert werden können, durchgeführt. Es wird dabei nicht nur die Leistungsfähigkeit eines Programms des digitalen Zwillings, sondern auch der Softwareentwicklungsaufwand, sowie dessen Korrektheit, Wartbarkeit und Evolvierbarkeit betrachtet. Darüber hinaus wird der aktuelle Stand der Technik untersucht und praktische Lösungen für die Implementierung einer AAS, im Hinblick auf spezifische SPS-Entwicklungsumgebungen, mit Ratschlägen für Entwickler gegeben.

Keywords: Industrie 4.0, I4.0, Digitaler Zwilling, SPS, Asset Administration Shell

1 Einleitung

Die Verbindung von Industrie 4.0 (I4.0) und speicherprogrammierbaren Steuerungen (SPSen) prägt zunehmend die industrielle Automatisierung. Sie markiert den Übergang zu einer Ära, die durch Digitalisierung, Interoperabilität und datengestützte Prozesse bestimmt wird.

Digitale Zwillinge stellen virtuelle Repliken physischer Anlagen, Prozesse oder Systeme dar und sind zu zentralen Elementen moderner Steuerungs- und Überwachungskonzepte geworden. Sie verbessern das Situationsbewusstsein durch Echtzeitdaten, ermöglichen simulationsgestützte Optimierungen und fördern fundierte Entscheidungen [PeNW17;

[1] TU Dresden, Professur für Prozesskommunikation, Nöthnitzer Str. 46, 01062 Dresden, nico.braunisch@tu-dresden.de, https://orcid.org/0009-0000-2432-5529
[2] TU Dresden, Professur für Prozesskommunikation, Nöthnitzer Str. 46, 01062 Dresden, santiago.soler_perez_olaya@tu-dresden.de, https://orcid.org/0000-0003-1247-4492
[3] TU Dresden, Professur für Prozesskommunikation, Nöthnitzer Str. 46, 01062 Dresden, uwe.schmidt1@tu-dresden.de, https://orcid.org/0009-0001-8899-2504
[4] TU Dresden, Professur für Prozesskommunikation, Nöthnitzer Str. 46, 01062 Dresden, martin.wollschlaeger@tu-dresden.de, https://orcid.org/0000-0002-4646-4455

WeZo18; CaSa20]. In I4.0-Initiativen sind digitale Zwillinge somit Schlüsselbausteine vernetzter Ökosysteme, die SPSen, Sensoren, Aktoren und IoT-Komponenten zusammenführen [LWZL22], die in Abb. 1 dargestellt ist.

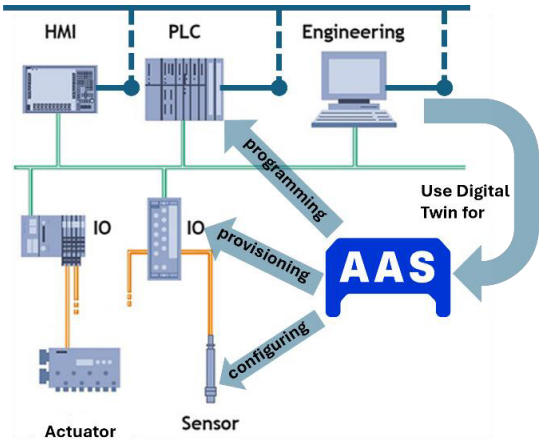


Abb. 1: Interaktion des Digitaler Zwillings (AAS) mit einer SPS-Umgebung

Parallel entwickelt sich die SPS-Programmierung von prozeduraler Logik hin zu höheren Programmiersprachen. Die aktuelle Norm IEC 61131-3 (3. Edition) integriert Konzepte objektorientierter Programmierung, während Plattformen wie TwinCAT 3 [Beck00] darüber hinaus Hochsprachen wie C++ unterstützen. Viele Systeme verharren jedoch bei älteren Editionsständen und damit eingeschränkter Modularität, was die Wiederverwendung von Quelltext erheblich einschränkt [Somm15].

Vor diesem Hintergrund gewinnt die Asset Administration Shell (AAS) als interoperabler Standard für digitale I4.0-Zwillinge zunehmend an Bedeutung [BBBO21]. Um ihre Implementierung zu erleichtern, wurden verschiedene Software Development Kits (SDKs) entwickelt, die insbesondere die Modellierung und Serialisierung von AAS-Strukturen unterstützen. Ziel dieses Beitrags ist daher eine qualitative, softwaretechnisch fundierte Analyse, wie AAS-basierte Zwillinge auf modernen wie auch auf Legacy-SPSen implementiert werden können.

2 Verwandte Arbeiten

Die Überschneidung von AAS in I4.0 und moderner SPS-Entwicklung bringt zwei unterschiedliche Perspektiven zusammen, um die vertikale Integration und den interoperablen Informationsaustausch vom Shopfloor zur vernetzten Welt zu fördern.

2.1 Moderne SPS-Entwicklung

Die heutige Entwicklung von SPSen orientiert sich zunehmend an modernen Methoden und Technologien, die Effizienz und Flexibilität steigern [VoFS15]. Soft-SPSen markieren dabei eine Abkehr von hardwarezentrierten Ansätzen und laufen vollständig softwarebasiert auf Mehrzweckplattformen [LaSt19]. Sie erleichtern Wartung, senken Hardwarekosten und ermöglichen die Integration mit anderen Softwaresystemen, wodurch sie eine Grundlage für I4.0-Initiativen bieten [IPJP24].

Traditionell wurden SPSen mit Low-Level-Sprachen wie Kontaktplan oder strukturiertem Text programmiert. Aktuelle Trends verlagern den Fokus jedoch auf Hochsprachen wie C++ oder Python, die mit ausdrucksstarker Syntax und umfangreichen Bibliotheken Entwicklung und Wartbarkeit verbessern [YaLP24; ZoSt19]. Damit eröffnen sich auch Möglichkeiten für den Einsatz von maschinellem Lernen und Datenanalyse.

SPSen entwickeln sich so zu zentralen Komponenten digitaler Zwillinge, die Echtzeitüberwachung, vorausschauende Wartung und datenbasierte Entscheidungen ermöglichen. Entsprechend erweitern sie ihre Rolle im I4.0-Paradigma und schaffen ein vernetztes, flexibles Ökosystem [CJXI23].

2.2 Digitale Zwillinge der I4.0

In I4.0 verändern digitale Technologien die industriellen Prozesse hin zu einem intelligenten, vernetzten Ökosystem. Zentrale Rolle spielt dabei die Interoperabilität, die verschiedene Systeme verbindet. Digitale Zwillinge repräsentieren physische Einheiten in virtueller Form und ermöglichen datenbasierte Entscheidungen und Prozess-optimierungen.

Die Asset Administration Shell (AAS) gilt inzwischen als Standard für I4.0-Zwillinge. Sie gewährleistet den interoperablen Austausch zwischen Informationsmodellen, Domänen und Stakeholdern. 2016 wurde sie als DIN SPEC 91345 und 2020 als IEC 63278 normiert [BMLU19]. Im AAS-Metamodell beschreibt die Klasse AssetAdministrationShell ein Asset und seine Submodelle. Eine Übersicht zentraler Elemente ist in Abb. 2 dargestellt [Deta22, S.44].

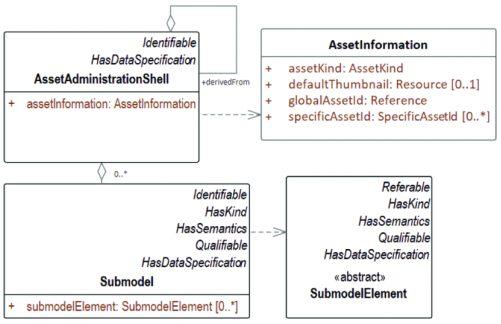


Abb. 2: Ausschnitt der Metamodell-Übersicht der Asset Administration Shell [Deta22, S.44]

Die Plattform Industrie 4.0 veröffentlichte ein technologieneutrales UML Informationsmodell [Deta22] sowie Serialisierungen (JSON, XML, RDF) und eine HTTP-Schnittstellendefinition [Deta21]. Submodelle bestehen aus Submodel-Element-Objekten, die Eigenschaften, Wertebereiche oder semantische Verweise enthalten. Diese Verweise sichern die maschineninterpretierbare Kommunikation und ermöglichen damit den automatisierten Datenaustausch.

Unterschieden werden drei Interaktionsmuster: Austausch über Dateien (Typ 1), REST-Schnittstellen (Typ 2) und Peer-to-Peer-Kommunikation auf Basis der I4.0-Sprache (Typ 3) [MSVU21; BeDi19].

2.3 Software Development Kits für I4.0 Digital Twins, die für moderne PLCs geeignet sind

Die Wiederverwendung von Quelltexten ist ein zentrales Paradigma moderner Softwareentwicklung, etwa in der objektorientierten [BMEY07] und funktionalen Programmierung [AbSu96]. Hochsprachen wie C++, Python oder C# ermöglichen dies und erleichtern damit auch die Entwicklung digitaler Zwillinge. Um die Implementierung der Asset Administration Shell (AAS) zu vereinfachen, stehen verschiedene Software Development Kits (SDKs) zur Verfügung, die insbesondere die Modellierung des AAS-Metamodells übernehmen [BLWR23].

Drei Open-Source-Initiativen sind besonders relevant: **Eclipse BaSyx** mit SDKs für Python, Java und C# (C++ ist nicht gepflegt, Rust in Entwicklung) [Ecli25a], **AAS4J** mit einem stark auf das Builder-Muster fokussierten Java-SDK [Ecli25b] sowie **aas-core** [Aas-25]. Letzteres verfolgt einen generativen Ansatz, bei dem ein formalisiertes AAS-Metamodell in vereinfachtem Python als Grundlage für SDKs in vielen Sprachen dient, darunter C++, C#, Java, Python (inkl. MicroPython), Go und TypeScript [BRLW23].

Diese Ansätze entlasten Entwickler, sichern Konsistenz mit dem AAS-Metamodell und fördern Korrektheit, Anpassungsfähigkeit und Wartbarkeit.

2.4 Für AAS-SDKs geeignete SPS-Plattformen

Eine Auswahl relevanter SPS-Plattformen wird in [GWLW19] vorgestellt, ausgewählt nach der Fähigkeit zur offenen Softwareentwicklung in Hochsprachen [LiLi11]. **TwinCAT** von Beckhoff bietet als Echtzeit-Automatisierungssystem eine enge Integration in Visual Studio und unterstützt neben IEC 61131-3 auch C++ und C# [Beck00]. **Sigmatek** stellt ein modulares, echtzeitfähiges System mit IEC 61131-3 sowie ANSI-C und Python bereit und legt Wert auf benutzerfreundliche Entwicklungswerkzeuge. **CODESYS** zeichnet sich durch Hardwareunabhängigkeit, breite Herstellerkompatibilität und Unterstützung von IEC 61131-3, C/C++ und Python aus; zusätzlich steigern Protokolle wie OPC UA und MQTT die Interoperabilität. **PLCnext** von Phoenix Contact kombiniert Echtzeitsteuerung mit Hochsprachen wie C++, C# und Python sowie Cloud-Anbindung. **ctrlX** AUTOMATION von Bosch Rexroth basiert auf einem Linux-Echtzeitsystem, erlaubt App-basierte Programmierung und IoT-Integration und unterstützt IEC 61131-3, C++, Python und JavaScript.

Alle Plattformen eröffnen durch Mehrsprachigkeit und Offenheit günstige Voraussetzungen für die Integration von AAS-SDKs und fördern so die Umsetzung digitaler Zwillinge in Industrie-4.0-Infrastrukturen.

2.5 Verknüpfung von I4.0 Digital Twin und SPS

In der Literatur existieren mehrere Ansätze zur Implementierung digitaler Zwillinge in Verbindung mit SPSen. Cavalieri et al. stellen ein Modell der Asset Administration Shell (AAS) zur Repräsentation von IEC 61131-3-Programmen vor und zeigen, wie damit das Lebenszyklusmanagement verbessert werden kann [CaSa20]. Cheong et al. untersuchen den Einsatz digitaler Zwillinge zur Prozessverbesserung und betonen Vorteile wie Echtzeitüberwachung und Rückkopplung, weisen aber auf die Limitierungen von Kontaktplänen hin [CJXI23]. Schäfer et al. schlagen zur Synchronisierung von SPSen mit AAS den PackML-Standard vor, um Konfigurationsaufwände zu reduzieren, und verwenden Eclipse BaSyx als Infrastruktur [SSKL23]. Wenger und Zoitl entwickeln ein Framework zur Anbindung von SPSen an ihre AAS zur automatischen Gerätekonfiguration [WeZo18]. Li et al. zeigen das Potenzial von digitalen Zwillingen zur Echtzeitüberwachung in der additiven Fertigung und heben die Vorteile höherer Programmiersprachen und Datenanalyse hervor [LWZL22].

Während diese Arbeiten spezifische Implementierungen behandeln, hebt sich der vorliegende Beitrag dadurch ab, dass er die Entwicklung digitaler Zwillinge für SPSen erstmals systematisch aus softwaretechnischer Perspektive analysiert.

3 Methodik

Der Beitrag untersucht qualitativ die Besonderheiten der Softwareentwicklung für digitale Zwillinge auf SPSen, schlägt Ansätze vor und bewertet diese. Schwerpunkte sind

Korrektheit, Effizienz, Wartbarkeit und Anpassungsfähigkeit [Somm15]. Sicherheit und Schutzaspekte bleiben künftigen Arbeiten vorbehalten. Analysiert werden unterschiedliche Konfigurationen der Programmierunterstützung.

3.1 Verwendung bestehender AAS-SDKs

Für die Entwicklung digitaler Zwillinge auf Basis der Asset Administration Shell (AAS) stehen mittlerweile verschiedene Software Development Kits (SDKs) bereit (vgl. 2.3). Da SPS-Entwicklungsumgebungen gängige Hochsprachen unterstützen (siehe 2.4), können diese SDKs weitgehend direkt genutzt werden. Ihre Verbreitung und Offenheit erhöhen die Wahrscheinlichkeit der Fehlerentdeckung; zusätzliche Testverfahren wie Fuzzing steigern die Korrektheit [BLWR23; OtKl22; BROS24]. Der generative Ansatz von aas-core [BRLW23] erleichtert zudem die Weitergabe von Metamodellkorrekturen. Entwickler profitieren durch höhere Entwicklungsgeschwindigkeit, da Metamodell- und Infrastrukturoperationen (z. B. De-/Serialisierung) abgedeckt sind. Entwurfsmuster wie Builder und Visitor [GHJV94] verbessern Lesbarkeit und Wartbarkeit, fördern Erweiterbarkeit etwa bei neuen Visualisierungen und erleichtern die Anpassungsfähigkeit. Einschränkungen bestehen jedoch durch die Echtzeitanforderungen vieler SPSen, die Laufzeitumgebungen verhindern und damit SDKs wie AAS4J, BaSyx C# oder aas-core C# ausschließen. Alternative Umgebungen wie MicroPython [GeCo25] und TinyGo [Cont25a] sind nur eingeschränkt kompatibel. Auch C++ wird auf manchen Plattformen unvollständig unterstützt (z. B. fehlen in TwinCat 3 PLC Smartpointer, [Beck00]), was die Nutzung von aas-core-cpp verhindert. Praktisch verfügbare Optionen beschränken sich derzeit auf aas-core SDKs für MicroPython und Go [Aas-25], deren tatsächliche Einsetzbarkeit auf SPSen jedoch noch untersucht werden muss.

3.2 Anpassung

Wenn SDKs nicht direkt einsetzbar sind, können sie an die jeweilige SPS-Plattform angepasst werden. Manuelle Ergänzungen fehlender Funktionalitäten (z. B. Smartpointer) sind für die meisten Teams nicht realistisch. Ebenso ist eine händische Überarbeitung der SDKs, etwa zur Entfernung von Typannotationen, aufgrund des Umfangs unpraktikabel [BRLW23]. Automatisierte Ansätze wie Semantic Patching [BHGR24] sind praktikabler: Der Quelltext wird mittels Syntaxbaum analysiert und algorithmisch angepasst, etwa zur Ableitung von aas-core-micropython aus aas-core-python [Cont25b]. Diese Methode ist jedoch nur bei lokal begrenzten Änderungen effektiv; für tiefgreifende Anpassungen, wie das Ersetzen von Smartpointern in C++, ist sie ungeeignet.

3.3 Erzeugung

Wenn bestehende SDKs nicht angepasst werden können, muss ein neues SDK erzeugt werden. Anstelle einer manuellen Neuentwicklung wird ein generativer Ansatz

vorgeschlagen, bei dem ein SDK-Generator den Quelltext für die jeweilige SPS-Umgebung erzeugt (Abb. 3). Dieses Verfahren wird bereits erfolgreich in aas-core eingesetzt [BRLW23]. Ausgangspunkt ist ein formalisiertes Metamodell in vereinfachtem Python, das in eine Intermediate Representation (IR) überführt wird. Diese abstrahiert von Syntax und Textdarstellung, erfasst jedoch alle wesentlichen Elemente wie Klassen, Einschränkungen, Eigenschaften und Methoden. Anschließend übersetzt der Generator die IR algorithmisch in Quelltext, typischerweise über Textvorlagen.

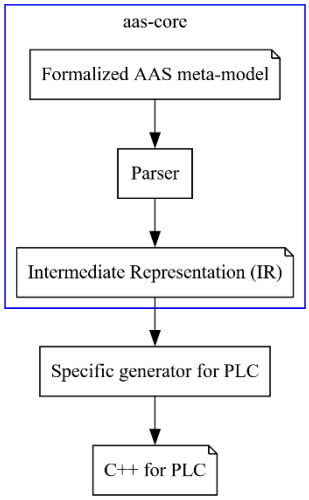


Abb. 3: Der vorgeschlagene Ablauf zur Erstellung eines AAS-SDKs
für eine SPS am Beispiel für C++

über verschiedene Versionen des Metamodells hinweg gewährleistet. Dies steht im Einklang mit bestehenden Forschungsarbeiten, die über die Herausforderungen der Wartbarkeit handgeschriebener digitaler Zwillinge für SPSen angesichts Der Entwicklungsaufwand für einen Generator entspricht mindestens dem eines SDKs, ist in der Praxis jedoch kürzer: aas-core berichtet von einer um den Faktor 3–5 reduzierten Entwicklungszeit [BRLW23]. Zudem lassen sich Parser und IR-Module wiederverwenden. Durch die algorithmische Übersetzung ist der generierte Quelltext konsistent mit dem Metamodell, wodurch eine ganze Klasse von Inkonsistenz-Fehlern vermieden wird. Korrekturen am Metamodell werden automatisch in den Quelltext übernommen, was Wartbarkeit und Anpassungsfähigkeit langfristig erhöht. Damit wird auch die Herausforderung adressiert, die in [SSKL23] für handgeschriebene digitale Zwillinge beschrieben wird: ihre schwierige Anpassung an sich stetig ändernde AAS-Metamodelle.

3.4 Erzeugung für IEC 61131-3 < 3. Ausgabe

Die IEC 61131-3 (Std. 2013) definiert die klassische Programmierumgebung für SPSen und umfasst fünf Modi: 1) Strukturierter Text, 2) Anweisungsliste, 3) Kontaktplan, 4)

Funktionsblockdiagramm und 5) Ablaufdiagramm. Vor der dritten Ausgabe erlaubte sie jedoch nur geringe Modularität. Quelltext konnte zwar in prozedurale Blöcke gegliedert werden, bot aber kaum Unterstützung für komplexe Datenstrukturen. Objektorientierung (Verknüpfung von Operationen mit Daten, Kapselung, Subtypen [BMEY07; LiWi94]) und Generizität (parametrisierte Typen [MuSt89]) fehlten vollständig. Die Programmierung blieb prozedural, Wiederverwendung auf einzelne Prozeduren mit fixierten Argumenttypen beschränkt.

Dies erschwert die Abbildung des AAS-Metamodells erheblich, da dieses objektorientierte Konzepte wie Mehrfachvererbung nutzt. Die semantische Diskrepanz zum prozeduralen Code ist entsprechend groß. Eine Lösung bietet wiederum die Generierung: Schon bei aas-core-golang konnte trotz fehlender Vererbung in Go eine konsistente Abbildung erzeugt werden [BRLW23]. Interfaces und Methoden werden dupliziert, sodass das AAS-Metamodell trotz Sprachbeschränkungen vollständig abgebildet werden kann.

Für IEC 61131-3 wird vorgeschlagen, neben Datenstrukturen auch Casting-Methoden und Polymorphismus zu generieren. Polymorphismus kann über CASE-Anweisungen oder Funktionszeiger implementiert werden. Erste Experimente mit diesem Ansatz zeigen Machbarkeit, auch wenn weiterführende Untersuchungen notwendig sind. Bisher traten keine offensichtlichen Probleme auf.

3.5 Fahrplan für die Evaluation

Um zu testen, ob eine AAS-SDK-Implementierung mit der SPS-Umgebung übereinstimmt und ob sie korrekt ist, wird vorgeschlagen, dem Testplan zu folgen, der sich an dem in [MHGK22] für AAS Typ 1 bzw. [MHGR24] für AAS Typ 2 vorgestellten Testkonzept orientiert. Beim Testen von AAS Typ 1 liegt der Schwerpunkt auf der De-/Serialisierung von AAS-Instanzen in und aus JSON und XML sowie auf der Validierung der AAS-Metamodellbeschränkungen. Während zu erwarten ist, dass die Validierungstests unproblematisch sind, gibt es SPSen, welche die De-/ Serialisierung von und nach JSON sowie XML nicht unterstützen. In einem solchen Fall wird vorgeschlagen, eine benutzerdefinierte *String*-Darstellung in das generierte SDK zu integrieren.

4 Schlussfolgerung und zukünftige Arbeit

In diesem Beitrag wurden verschiedene Aspekte der Softwareentwicklung von digitalen I4.0-Standard-Zwillingen für SPSen qualitativ bewertet. Dabei wurde sich auf Asset Administration Shells (AAS), als eine Standardrealisierung eines digitalen I4.0-Zwillings, konzentriert. Es wurden verschiedene Ansätze, wie die Verwendung bestehender AAS-SDKs, deren Anpassung oder die Generierung neuer SDKs, abhängig von der Unterstützung der SPS-Plattform, analysiert. Wichtig ist, dass diese Analyse zeigt, dass AAS bereits auf SPS-Plattformen realisiert werden kann, welche MicroPython oder TinyGo unterstützen. Praktische Tests stehen dabei noch aus. Dies ist ein enormer Vorteil

für die Softwareentwickler, da sie bei der Erstellung digitaler I4.0-Zwillinge alle Vorteile der Wiederverwendung von Quelltexten nutzen können (schnellere Entwicklung, bessere Korrektheit, langfristige Wartbarkeit und Evolvierbarkeit). Andererseits wird für den Fall, dass eine SPS-Plattform die vorhandenen SDKs nicht unterstützt, vorgeschlagen einen alternativen Ansatz zu verwenden, der die Erstellung neuer AAS-SDKs für die spezifischen SPS-Plattformen beinhaltet.

Zukünftig sollen die Herausforderungen bei der Entwicklung eines AAS-SDK für ältere Varianten der IEC 61131-3 (d.h. vor der 3. Ausgabe) durch eine konkrete Implementierung des entsprechenden Generators untersucht werden. Darüber hinaus werden die Experimente auch Tests auf tatsächlichen SPS-Plattformen umfassen, von denen erwartet wird, dass sie die Durchführbarkeit des in diesem Beitrag vorgeschlagenen Testfahrplans bestätigen.

Literaturverzeichnis

[Aas-25] *aas-core-works*. URL https://github.com/aas-core-works/. - abgerufen am 2025-02-02

[AbSu96] ABELSON, H. ; SUSSMAN, G. J.: *Structure and Interpretation of Computer Programs*. 2nd. Aufl. : MIT Press, 1996

[BBBO21] BADER, SEBASTIAN ; BELLINGHAUSEN, VANESSA ; BOSS, BIRGIT ; OTHERS, MANY: *What is the Asset Administration Shell from a technical perspective?*, 2021

[Beck00] BECKHOFF AUTOMATION GMBH & CO. KG: *TwinCAT 3 C++: Programming Guide*. 1.17.

[BeDi19] BELYAEV, ALEXANDER ; DIEDRICH, CHRISTIAN: *Specification „Demonstrator I4.0-Language“ v2.0*, 2019

[BHGR24] BRAUNISCH, NICO ; HEPPNER, SEBASTIAN ; GARMAEV, IGOR ; RISTIN, MARKO ; MINY, TORBEN ; KLEINERT, TOBIAS ; VAN DE VENN, HANS WERNHER ; WOLLSCHLAEGER, MARTIN: Software Evolution of I4.0 Digital Twins with Semantic Patching. In: *Industrial Electronics (ISIE)*, 2024

[BLWR23] BRAUNISCH, N. ; LEHMANN, R. ; WOLLSCHLAEGER, M. ; RISTIN, M. ; VAN DE VENN, H. W. ; OTTO, B. ; KLEINERT, T.: Maturity Evaluation of SDKs for I4.0 Digital Twins. In: *Emerging Technologies and Factory Automation (ETFA)*, 2023

[BMEY07] BOOCH, G. ; MAKSIMCHUK, R. A. ; ENGLE, M. W. ; YOUNG, B. J. ; CONALLEN, J. ; HOUSTON, K. A.: *Object-Oriented Analysis and Design with Applications*. 3rd. Aufl. : Addison-Wesley, 2007

[BMLU19] BOSS, B. ; MALAKUTI, S. ; LIN, S. ; USLÄNDER, T. ; CLAUER, E. ; HOFFMEISTER, M. ; STOJANOVIC, L.: *Digital Twin and Asset Administration Shell Concepts and Application in the Industrial Internet and Industrie 4.0* : Industrial Internet Consortium, Plattform Industrie 4.0, 2019

[BRLW23] BRAUNISCH, N. ; RISTIN, M. ; LEHMANN, R. ; WOLLSCHLAEGER, M. ; VAN DE VENN, H. W.: Empowering Industry 4.0 with Generative and Model-Driven SDK Development. In: *Industrial Electronics Society (IECON)*, 2023

[BROS24] Braunisch, Nico ; Ristin, Marko ; Otto, Björn ; Schanely, Phillip ; van de Venn, Hans Wernher ; Wollschlaeger, Martin ; Kleinert, Tobias: Automatic Fuzzing of Asset Administration Shells as Digital Twins for Industry 4.0. In: *Automation Science and Engineering (CASE)*, 2024

[CaSa20] Cavalieri, Salvatore ; Salafia, Marco Giuseppe: Asset Administration Shell for PLC Representation Based on IEC 61131–3. In: *IEEE Access* Bd. 8 (2020), S. 142606–142621

[CJXI23] Cheong, Alexander Chee Hon ; Jie, Kok Fu ; Xian, Jerome Ignatius Yuen Yi ; Ibrahim, Zunaidi: Digital Twin in Manufacturing by Using Programmable Logic Controller (PLC). In: *AIP Conference Proceedings*, 2023

[Cont25a] Contributors, TinyGo: *TinyGo: A Go Compiler for Small Places*. URL https://tinygo.org/. - abgerufen am 2025-02-02

[Cont25b] Contributors, AAS Core Works: *AAS Core 3.0 MicroPython*. URL https://github.com/aas-core-works/aas-core3.0-micropython. - abgerufen am 2025-02-02

[Deta21] *Details of the Asset Administration Shell - Part 2 V1.0-RC02* : Federal Ministry for Economic Affairs and Climate Action (BMWK), 2021. — Backup Publisher: Plattform Industrie 4.0

[Deta22] *Details of the Asset Administration Shell - Part 1 V3.0-RC02* : Federal Ministry for Economic Affairs and Climate Action (BMWK), 2022. — Backup Publisher: Plattform Industrie 4.0

[Ecli25a] Eclipse Foundation: *Eclipse BaSyx: Open Source Platform for Next Generation Automation*. URL https://eclipse.dev/basyx/. - abgerufen am 2025-02-02

[Ecli25b] Eclipse Foundation: *Eclipse AAS4J*. URL https://projects.eclipse.org/projects/dt.aas4j. - abgerufen am 2025-02-02

[GeCo25] George, D. ; contributors: *MicroPython*. URL https://micropython.org/. - abgerufen am 2025-02-02

[GHJV94] Gamma, E. ; Helm, R. ; Johnson, R. ; Vlissides, J.: *Design Patterns: Elements of Reusable Object-Oriented Software* : Addison-Wesley, 1994

[GWLW19] Geng, Yangyang ; Wang, Yi ; Liu, Wenwen ; Wei, Qiang ; Liu, Ke ; Wu, Haolan: A survey of industrial control system testbeds. In: *IOP Conference Series: Materials Science and Engineering* Bd. 569, IOP Publishing (2019), Nr. 4, S. 042030

[IPJP24] Ingale, Amey ; Patil, Sakshi ; Jambhale, Sakshi ; Patil, Sneha: Transforming Industrial Control: The Evolution and Future Trends of PLC Systems. In: *Creative Research Thoughts (IJCRT)* (2024)

[LaSt19] Langmann, Reinhard ; Stiller, Michael: The PLC as a Smart Service in Industry 4.0 Production Systems. In: *Applied Sciences* (2019)

[LiLi11] Liang, Quan ; Li, Li: The Study of Soft PLC Running System. In: *Procedia Engineering* Bd. 15 (2011)

[LiWi94] Liskov, B. ; Wing, J. M.: A Behavioral Notion of Subtyping. In: *Principles of Programming Languages (POPL)*, 1994

[LWZL22] Li, Zhongkai ; Wang, Yu ; Zhang, Yusheng ; Liu, Yuchen ; Zhang, Yuru ; Zhang, Yuru: Towards a digital twin for real-time monitoring of robotic additive manufacturing. In: *Procedia CIRP* Bd. 107 (2022), S. 1486–1491

[MHGK22] Miny, Torben ; Heppner, Sebastian ; Garmaev, Igor ; Kleinert, Tobias ; Ristin, Marko ; Van De Venn, Hans Wernher ; Otto, Björn ; Meinecke, Karsten ; u. a.: Semi-Automatic Testing of Data-Focused Software Development Kits for Industrie 4.0. In: *Industrial Informatics (INDIN)*, 2022

[MHGR24] Miny, Torben ; Heppner, Sebastian ; Garmaev, Igor ; Ristin, Marko ; Otto, Björn ; Braunisch, Nico ; Jacoby, Michael ; Kleinert, Tobias ; u. a.: Towards a Test Framework for Reactive (Type 2) Asset Administration Shell Implementations. In: *Emerging Technologies and Factory Automation (ETFA)*, 2024

[MSVU21] Miny, T. ; Stephan, G. ; Vialkowitsch, J. ; Usländer, Th.: *Functional View of Asset Administration Shell in an Industrie 4.0 System Environment*, 2021

[MuSt89] Musser, D. R. ; Stepanov, A. A.: Generic Programming. In: *Symbolic and Algebraic Computation (ISSAC)*, 1989

[OtKl22] Otto, B. ; Kleinert, T.: A Flow Graph based Approach for controlled Generation of AAS Digital Twin Instances for the Verification of Compliance Check Tools. In: *Industrial Electronics Society (IECON)*, 2022

[PeNW17] Pethig, Florian ; Niggemann, Oliver ; Walter, Armin: Towards Industrie 4.0 compliant configuration of condition monitoring services. In: *2017 ieee 15th international conference on industrial informatics (indin)* : IEEE, 2017, S. 271–276

[Somm15] Sommerville, Ian: *Software Engineering*. 10. Aufl. : Pearson, 2015

[SSKL23] Schäfer, Stephan ; Schöttke, Dirk ; Kämpfe, Thomas ; Lachmann, Oliver ; Zielstorff, Aaron: Synchronizing Devices Using Asset Administration Shells. In: *Innovative Intelligent Industrial Production and Logistics*, 2023

[VoFS15] Vogel-Heuser, Birgit ; Fay, Alexander ; Schaefer, Ina: Evolution of software in automated production systems: Challenges and research directions. In: *Systems and Software (J. Syst. Softw.)* Bd. 110 (2015)

[WeZo18] Wenger, Monika ; Zoitl, Alois: Connecting PLCs with their Asset Administration Shell for Automatic Device Configuration. In: *Industrial Informatics (INDIN)*, 2018

[YaLP24] Yao, Kai-Chao ; Lin, Cheng-Lung ; Pan, Chih-Hsuan: Industrial Sustainable Development: The Development Trend of Programmable Logic Controller Technology. In: *Sustainability* Bd. 16 (2024), Nr. 14

[ZoSt19] Zoitl, Alois ; Strasser, Thomas: High-Level Programming Languages for Industrial Automation Systems. In: *Industrial Informatics (INDIN)* (2019)

http://dx.doi.org/10.25673/121097

Lightweight CNN-Augmented SLAM for Smart Factories: A ROS 2 Framework for Mapping and Loop Closure

Krithiga Ramesh[1], Maxim Friesen[1], Tullio Facchinetti[2], and Lukasz Wisniewski[1]

Abstract: Mobile robots are vital for autonomous operations in modern manufacturing. Advances in LiDAR, odometry, and vision sensors have improved their perception and navigation in complex environments. A core requirement for such autonomy is Simultaneous Localization and Mapping (SLAM), enabling robots to estimate their position while building an environmental map. Smart factories pose unique challenges for SLAM like movable machinery, reflective surfaces, varying lighting, and human activity often degrade performance, causing incomplete maps and unreliable loop closure. Single-modality systems, such as LiDAR-only or vision-only SLAM, fail under feature-poor or low-light conditions. We propose MobileNet-augmented RTAB-Map, a lightweight, CNN-enhanced multi-sensor SLAM framework for CPU-only deployment. It integrates a MobileNet-V3-Small semantic filter into RTAB-Map's visual front end to suppress dynamic objects and retain static features. The system achieves real-time performance ($\geq$ 20 Hz) on a standard Intel i7-6600U CPU, improving mapping accuracy and robustness in the living lab, SmartFactoryOWL.

Keywords: SLAM, ROS 2, Loop Closure, Lightweight CNN

1 Introduction

Modern industrial environments demand navigation stacks that remain robust to human motion, movable entities, reflective machinery, and reconfigurable layouts. Multi-sensor SLAM (e.g., LiDAR+RGB-D) can stabilize localization. However, dynamic elements still corrupt front-end data association and loop closure, leading to degraded maps and increased drift. Recent progress in semantic and dynamic SLAM uses deep networks to detect and suppress dynamic regions before pose optimization, improving robustness in cluttered scenes [1] [2]. However, many approaches depend on computationally large, high-parameter networks or GPU acceleration. These are misaligned with real-time, CPU-only deployments typical of cost-constrained robots.

In this work, we propose a lightweight CNN-augmented SLAM front-end designed for a remote edge PC (Intel i7-6600U, CPU-only) that controls a mobile robot via Wi-Fi 5. We chose a MobileNet-family architecture due to its depthwise separable convolutions and squeeze-and-excite blocks, which deliver competitive accuracy at low latency on CPUs. Evidence from diverse robotics tasks, including dynamic filtering [3], robot control stacks [4], [5], and even human-robot interaction (HRI), confirms that MobileNet variants provide efficient perception, strengthening our choice for a CPU-only ROS 2 ecosystem.

1 Technische Hochschule Ostwestfalen-Lippe, Institut für Industrielle Informationtechnik, Campusallee 6, 32657 Lemgo, Deutschland, krithiga.ramesh@th-owl.de; maxim.friesen@th-owl.de; lukasz.wisniewski@th-owl.de
2 Università di Pavia, Faculty of Engineering, Via Adolfo Ferrata 5, 27100 Pavia, Italy, tullio.facchinetti@unipv.it

The key contributions of this paper are:

- A MobileNet-based static and dynamic masking front-end is proposed, integrated with a multi-sensor SLAM framework (RTAB-Map(F)) to reduce loop-closure errors and ghost artifacts while preserving keypoint density.
- A CPU-only, ROS 2-compatible implementation that maintains $\geq 20\,\text{Hz}$ without GPU.
- A tailored evaluation in a factory-like indoor environment showing improvements in Absolute Trajectory Error (ATE), Relative Pose Error (RPE), Loop-closure count (LC), map coverage (MC) and CPU Resources.

2 State of the Art

Recent state-of-the-art SLAM systems bridge traditional geometric mapping with learning-based perception to achieve robust localization in complex, dynamic environments. This section provides a thorough review of the literature relevant to this research.

2.1 CNN-Based SLAM in Dynamic Environments

Traditional SLAM systems assume static environments [6], causing pose errors when moving objects corrupt feature tracks. Geometric outlier rejection and temporal filtering partially mitigate these errors but fail in environments with frequent human activity or mobile equipment. CNN-augmented SLAM addresses this limitation by learning to detect and mask dynamic regions before pose optimization.

Zhao et al. [7] introduced a visual SLAM method that applies CNN-based segmentation to exclude dynamic elements, stabilizing localization in complex indoor scenes. Wu et al. [8] expanded this approach by enabling SLAM operation in dynamic industrial spaces through the real-time filtering of human motion and reflective surfaces. Chen et al. [9] addressed the sensitivity of SLAM to illumination variations on factory floors, achieving improved loop-closure stability. Building on these developments, Xu et al. [10] incorporated an enhanced semantic detection stage that reinforces front-end feature matching via selective feature weighting.

For LiDAR-based systems, Dynamic LiDAR Odometry and Mapping (DLOAM) employs CNNs to detect and remove dynamic objects from 3D point clouds, thereby improving the robustness and accuracy of odometry estimation in urban and factory environments [11]. Furthermore, multi-modal fusion approaches that combine LiDAR, Inertial Measurement Unit (IMU), and attention mechanisms have been shown to mitigate drift under challenging conditions such as sensor occlusion, rapid motion, and feature-sparse areas [12]. In semantic mapping, researchers like Chen et al. have advanced object-level representations to embed

individual entity semantics into SLAM maps [13], while Zhang et al. demonstrate that semantic SLAM with RGB-D data can maintain mapping under occlusions and poor visibility in rescue scenarios [14]. Taken together, these works support the consensus that CNN-based segmentation and dynamic-object filtering are now foundational components in modern SLAM systems [15] that work best for the aimed smart factory use case.

2.2 Temporal and Recurrent Learning in Visual Odometry

While CNNs enhance spatial perception in SLAM, maintaining temporal coherence across frames is critical for consistent motion estimation [16]. To address this, recurrent neural architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks have been incorporated to learn motion continuity directly from sequential visual data.

Zheng et al. [17] proposed ATFVO (Attentive Tensor-Compressed Flow-based Visual Odometry). This method integrates optical-flow cues with LSTM encoding to generate smoother trajectories and improve temporal consistency. Kang et al. [18] tackled the overfitting challenge in deep visual odometry by introducing dropout-based temporal regularization, enhancing generalization across diverse motion patterns. Alam et al. [19] demonstrated that temporal CNN layers can emulate recurrent behavior in monocular visual odometry, enabling efficient operation on low-power mobile robots. Extending this idea, Zhao et al. [20] fused RGB-D and inertial inputs within a joint CNN–RNN architecture, mitigating drift under intermittent sensor failures.

Beyond ego-motion estimation, advances in temporal modeling for human-motion prediction indirectly contribute to dynamic-scene SLAM. Li et al. [21] achieved real-time motion forecasting from RGB video, while Zhao et al. [22] trained robots to anticipate human trajectories in collaborative environments. Similarly, LSTM-based activity-recognition frameworks [23], [24] have shown that temporal models can efficiently infer motion states on embedded CPUs, an insight that informs temporal smoothing and dynamic-object masking strategies in SLAM pipelines.

Collectively, these works demonstrate that integrating temporal learning mechanisms, either through explicit recurrent layers or implicit temporal encoding, enhances the stability and adaptability of SLAM systems operating in real-world, dynamic environments.

2.3 Lightweight CNN Architectures for Real-Time Robotics

Although semantic SLAM improves the robustness of localization and mapping in dynamic environments, its practical deployment on embedded or edge devices is often constrained by high computational demand. In other words, most deep semantic models are too resource-intensive to run in real time on low-power processors. To overcome this limitation, researchers

have proposed lightweight convolutional neural network (CNN) architectures such as ENet (Efficient Neural Network), Fast-SCNN (Fast Semantic Segmentation Convolutional Neural Network), and MobileNet, which maintain reasonable accuracy while significantly reducing computation and memory requirements.

Patel et al. [25] benchmarked recent real-time semantic-segmentation algorithms (ENet, Fast-SCNN, ESPNet) and reported that Fast-SCNN provides the best balance between accuracy and inference speed for robotic vision. Luo et al. [1] integrated Fast-SCNN into FD-SLAM, adding DeepFill v2 inpainting to restore occluded map regions and demonstrating significant accuracy gains over ORB-SLAM3 on TUM datasets. Rai et al. [26] developed InPosNet, a context-aware network with a MobileNetV3-inspired encoder, providing compact scene descriptors suitable for CPU-level indoor relocalization. Zhou et al. [2] presented a MobileNetV3-based feature-pyramid network with attention that reduced parameters by 65 % while maintaining descriptor stability—an important factor for keypoint-sparse industrial environments.

Practical robotics implementations confirm these efficiencies. Wang et al. [4] combined MobileNetV3 detection with ORB-SLAM2 for pipeline-inspection robots, achieving near-real-time operation on Jetson-class hardware. Chang et al. [5] used MobileNetV3-SSD in a low-cost mobile garbage-collection robot under ROS, showing consistent detection latency below 30 ms on Raspberry Pi. Warohma et al. [3] applied MobileNetV3 to speaker recognition for voice-based navigation, illustrating that the same architecture scales well to multi modal inputs. Additional TinyML studies confirm edge-CPU feasibility: Mobilenet-SSD v2 [27], Tiny-YOLO optimizations [28], [29], and benchmarking across heterogeneous CPUs [30], [31] all report sub-30 ms inference on standard desktop processors.

These studies substantiate the choice of MobileNetV3 as a practical choice for real-time, CPU-based deployment, particularly effective for lightweight semantic pre-processing in SLAM front-ends.

2.4 Research Gap and Problem Definition

Despite considerable advances in CNN-augmented SLAM, limited efforts have focused on integration within ROS 2 Framework. Most existing implementations remain ROS 1-centric and rely heavily on GPU acceleration, restricting deployment on industrial PCs with constrained computational resources. While some vision-based inspection robots and mapping frameworks employ neural modules, they typically omit node-level CPU profiling and direct connectivity to graph-based fusion back-ends such as RTAB-Map(F). Frameworks like InPosNet and FD-SLAM demonstrate progress toward modular integration [26]. However, both operate primarily on standalone datasets rather than complete ROS 2 navigation pipelines.

The review of literature emphasizes three key findings relevant to this study. First, CNN-based segmentation and dynamic-object filtering are critical for maintaining SLAM stability in non-static environments. Second, lightweight architectures such as MobileNetV3 and Fast-SCNN have demonstrated real-time inference capabilities on CPU-based systems, making them suitable for resource-constrained robotic deployments. Third, despite these advancements, there remains a clear lack of ROS 2–native, CPU-only SLAM front-ends explicitly tailored for mobile robots.

In the SmartFactoryOWL [32], a living lab offering a 2000 square-meter industry-close environment, transient entities such as humans and forklifts frequently corrupt visual feature tracks and loop-closure proposals. Therefore, a front-end capable of preserving static features while reliably suppressing dynamic regions under strict real-time, CPU-only constraints is essential.

To address this gap, we designed and integrated a MobileNet-based binary masking module into RTAB-Map(F) within the ROS 2 Humble framework. Our approach aims to:

1. Sustain keypoint density for stable visual odometry (VO),
2. Filter visual correspondences passed to the multi-sensor fusion back-end (RTAB-Map(F)),
3. Enhance trajectory and mapping accuracy, quantified through ATE, loop-closure count, and map coverage without requiring GPU acceleration.

The quantitative evaluation relies on standard SLAM accuracy and consistency metrics. Absolute Trajectory Error (ATE) measures the global drift between estimated and ground-truth trajectories, while Relative Pose Error (RPE) captures short-term drift between consecutive poses. Loop-closure (LC) indicating the number of successful landmark detections during a run. Map coverage (MC) quantifies the spatial consistency and completeness of the reconstructed map across successive updates, reflecting the continuity of the environment representation.

3 Experimental Setup

The previous benchmarking study done at the SmartFactoryOWL [33] identified **RTAB-Map(F)** as the most stable SLAM configuration by comparing all the standard SLAM error metrics across different sensor modalities. Building on this work, we introduce a **MobileNet-augmented RTAB-Map(F)**, designed to enhance mapping robustness in such dynamic industrial environments.

RTAB-Map(F) [34] integrates LiDAR and visual features within a graph-based optimization framework under ROS 2 Humble. However, it remains susceptible to dynamic feature contamination from humans and equipment. To mitigate this, a lightweight MobileNet-V3-Small network was embedded as a **pre-filter** in the visual front-end. The CNN produces

binary static and dynamic masks at 20 Hz, ensuring real-time performance without GPU acceleration on an Intel i7-6600U CPU-only system

3.1 MobileNet Model and Training

The MobileNet-V3-Small model was initialized with COCO-2017 segmentation weights and fine-tuned on a custom SmartFactoryOWL dataset developed for this study. Each RGB frame was manually annotated with binary semantic classes specific to this work: *static* (infrastructure, walls, stations) and *dynamic* (humans, forklifts, reflective zones, and other mobile robots). Although reflective zones are physically static, they were labeled as dynamic, as their appearance changes with illumination and camera motion often lead to false feature correspondences during SLAM [35]. The dataset was split 75/17/8 for training, validation, and testing.

Transfer learning with the first four backbone blocks frozen enabled stable adaptation to factory lighting variations and metallic textures. We employed the Adam optimizer (learning rate 1×10^{-4}, weight decay 1×10^{-5}) and trained for 25-epochs with early stopping using a patience-based criterion on validation-IoU. Inputs were resized to 512×288 and normalized using MobileNetV3's internal pre-processing layer. Data augmentation, brightness jitter, horizontal flips, Gaussian noise, random affine transformations (±3°), and light motion blur, further improved robustness to reflections and motion artifacts. The final model achieved an average **binary map IoU of 0.87** (dynamic vs. static) on the validation split.

3.2 Temporal Mask Stabilization

To prevent mask flickering between frames, temporal smoothing was applied:

$$M_t = \alpha M_{t-1} + (1 - \alpha)\hat{M}_t, \qquad \alpha = 0.6$$

where $\hat{M}_t$ is the current CNN prediction. Both M_t and $\hat{M}_t$ are per-pixel probability masks in $[0, 1]$. The smoothing coefficient was empirically set to $\alpha = 0.6$, providing an optimal trade off between temporal stability and responsiveness. Higher values introduced latency in adapting to new motion cues, whereas lower values increased inter-frame mask flickering. After smoothing, a fixed threshold $\tau = 0.5$ binarizes the mask. A lightweight **Pyramidal Lucas-Kanade optical-flow** consistency check propagates masks during dropped or skipped frames, maintaining temporal continuity crucial for loop-closure stability.

3.3 LiDAR-Visual Fusion

The CNN affects only the visual feature-extraction pipeline of RTAB-Map(F). Filtered ORB keypoints from static regions contribute to the visual odometry constraint, while

LiDAR scan-matching constraints are added in parallel. Both modalities feed into the same factor-graph optimizer using synchronized timestamps (`'ApproximateTime'` policy). Direct feature-to-edge correspondences are not enforced. Instead, visual outliers are removed upstream, which improves loop-closure consistency without altering LiDAR-only matching or the `'g2o'` back-end.

3.4 Feature Density Monitoring

To avoid tracking degradation from over-filtering, the system maintains a minimum feature-density threshold of 400-keypoints per frame. This value corresponds to the median visual feature count observed in baseline RTAB-Map(F) runs. If the filtered density drops below this threshold, the ORB extractor automatically lowers the `'fastThreshold'` parameter and temporarily increases the `'nFeatures'` cap to recover visual coverage in low-texture regions. Once density stabilizes, default parameters are restored, ensuring consistent visual odometry and map continuity in factory corridors and cluttered zones.

3.5 Computational Profiling

Computational profiling was performed on an Intel i7-6600U CPU during manual teleoperation mapping sessions in the SmartFactoryOWL environment. CPU utilization was recorded using `'htop'`, while per-node latency was estimated from ROS 2 topic statistics (`'ros2 topic hz'`) and callback timing logs. Comparable latency margins for sensor callbacks in ROS 2 Humble are reported in [36], supporting the adopted 20 Hz operating rate as a stable real-time configuration. Each experiment lasted approximately three minutes and was repeated three times to average load variations. The input frame rate was configured to 20 Hz while ensuring synchronization with the LiDAR input (10–15 Hz). Higher frame rates offered no measurable improvement in mapping accuracy but increased latency and CPU load beyond 80%, thus 20 Hz was adopted as an optimal trade-off for stable SLAM performance.

Tab. 1: Per-node computational profile of the proposed system.

Node / System	**CPU Load (%)**	**Avg. Latency (ms)**
RTAB-Map Front-End	38 ± 3	22
MobileNet CNN	20 ± 2	25
Graph Optimizer	17 ± 2	8
Total System	**75 ± 4**	**47 ± 3**

Baseline RTAB-Map(F) consumed approximately 65 % total load; thus, the CNN increased utilization by about 12 pp while maintaining real-time ($\geq$20 Hz) performance. As summarized in Table 1, the proposed MobileNet-augmented RTAB-Map(F) preserves real time feasibility with the CNN module contributing around one fourth of the total computational

load. In addition to mean latency, we observed low jitter (below 10 ms across all nodes) and no timing outliers exceeding the 50 ms frame interval required for 20 Hz operation, confirming stable real-time performance. The complete processing flow of the proposed MobileNet-augmented RTAB-Map(F) system is summarized in Figure 1 illustrating the sequence from data collection to final graph optimization and performance evaluation.

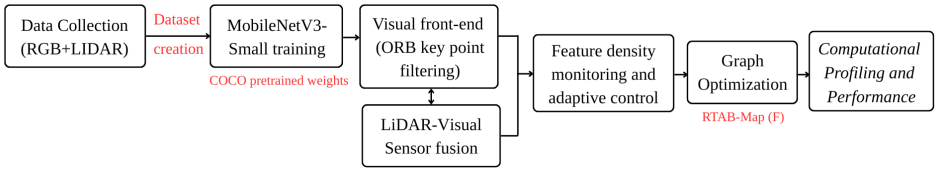


Fig. 1: Workflow of the proposed MobileNet-augmented RTAB-Map(F) SLAM pipeline

3.6 Dataset Development

To evaluate the proposed CNN-augmented SLAM system, a dedicated dataset was collected across four representative SmartFactoryOWL zones (A–D) covering stations, machinery, and human-activity corridors. Each run recorded approximately 800 RGB frames (512 × 288 px) were extracted at 1 fps to avoid temporal redundancy.

All frames were manually annotated in CVAT using two binary classes: *static* (infrastructure, walls, workstations, machinery) and *dynamic* (humans, forklifts, reflective metallic zones, movable carts). The dataset was divided into 75 % train, 17 % validation, and 8 % test splits, balanced across all zones to capture variations in lighting and texture.

The MobileNetV3-Small network was fine-tuned on this dataset before integration into the RTAB-Map(F) front-end. Metrics reported in Section 4 were computed on trajectories recorded within the same zones, ensuring a consistent evaluation domain between training and deployment.

4 Results

This section reports the quantitative metrics and qualitative observations comparing the proposed MobileNet-augmented RTAB-Map(F) with the baseline RTAB-Map(F) in the SmartFactoryOWL.

4.1 Quantitative Metrics

Across all quantitative metrics, the proposed configuration achieved reductions of approximately 12.5% in Absolute Trajectory Error and 20% in Relative Pose Error compared to the

Tab. 2: Comparison of RTAB-Map(F) Benchmark vs. MobileNet-Augmented Variant in SmartFactoryOWL

Method	**ATE (m)↓**	**RPE (m)↓**	**LC (count)↑**	**MC (%)↑**	**CPU (%)**
RTAB-Map(F)$_{\text{baseline}}$	0.08 ± 0.03	0.05 ± 0.01	8 ± 2	92 ± 3	65 ± 4
Proposed MobileNet-RTAB-Map(F)	**0.07** ± **0.02**	**0.04** ± **0.01**	**10** ± **2**	**94** ± **2**	**77** ± **4**

‡ Baseline data adapted from prior SLAM benchmark [33].

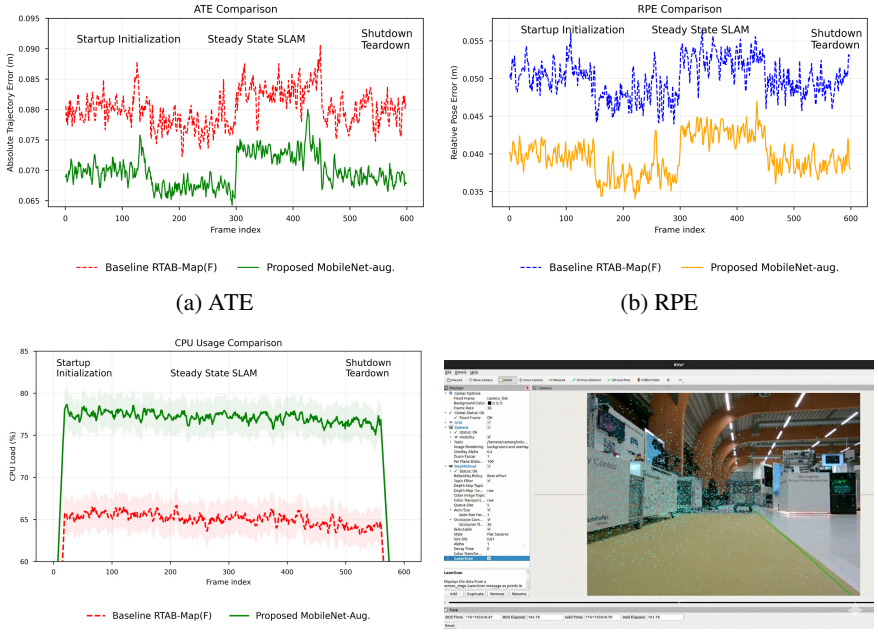
baseline, as summarized in Table 2, indicating improved motion estimation stability and reduced drift accumulation. Loop-closure count increased from 8 to 10 on average, and map coverage improved by around 2 percentage points, confirming enhanced place recognition and reduced ghost artifacts in human-active and reflective zones. Relocalization success also improved, suggesting better long-term spatial coherence under dynamic conditions. CPU utilization increased by about 12 percentage points (from 65% to 77%) due to the additional MobileNet front-end, yet the system sustained real-time operation at 20 Hz with latency remaining within the real-time threshold. These results were consistent across three independent runs, with no statistically significant variation ($p > 0.05$), validating the reproducibility and robustness of the observed improvements.

4.2 Qualitative Observations

The RViz visualization in Fig. 3(d) illustrates a cleaner and more stable reconstruction of the SmartFactoryOWL environment using the MobileNet-augmented RTAB-Map(F). The output preserves sharp object boundaries and reduces ghosting near reflective or dynamic areas, yielding a coherent and visually consistent map. These improvements highlight the model's enhanced scene understanding and real-time robustness, complementing the quantitative accuracy gains.

5 Discussion and Limitations

The proposed approach enhances RTAB-Map(F) performance in dynamic indoor environments with minimal computational cost. Nevertheless, the model is limited to binary static/dynamic classification and does not capture fine-grained semantics like DynaSLAM or Deep-VO. An important extension would be integrating real-time dynamic obstacle detection and motion-aware responses, such as temporarily halting navigation or re-planning trajectories when moving entities are detected. This could be achieved by combining the CNN-based segmentation with a lightweight multi-object tracking method such as Deep SORT [37], which facilitates persistent object identities and velocity estimation, enabling the system to react adaptively to dynamic obstacles during SLAM. Future work will incorporate



(a) ATE (b) RPE

(c) CPU Load (d) Qualitative Mapping

Fig. 2: Performance comparison of baseline and MobileNet-augmented RTAB-Map(F), showing improved accuracy ((a)–(b)), real-time scalability ((c)), and enhanced visual mapping quality ((d)) in SmartFactoryOWL.

lightweight temporal attention to handle partially moving objects and extend evaluation to larger statistical samples.

6 Conclusion

This study demonstrated that integrating a lightweight MobileNetV3-Small front-end into RTAB-Map(F) enhances SLAM performance in dynamic industrial settings. The proposed system reduced Absolute Trajectory Error by 12.5% and improved loop closure reliability by 25%, while maintaining real-time feasibility at approximately 77% CPU utilization. The resulting maps were more consistent, with clearer boundaries and fewer ghost artifacts in reflective zones. These findings confirm that robust LiDAR-visual fusion is achievable on standard edge PCs without GPU acceleration. Overall, the MobileNet-augmented configuration provides a practical path toward resilient and adaptive SLAM in smart factory environments.

References

[1] Luo *et al.*, "Fd-slam: A semantic slam based on enhanced fast-scnn dynamic region detection and deepfillv2-driven background inpainting," *IEEE Access*, 2023.

[2] Y. Zhou, Z. Gao, F. Bu, and L. Zhao, "A lightweight feature point detection model based on feature pyramid and attention," in *Proc. IEEE ICETCI*, 2024.

[3] A. M. Warohma, H. Hindersah, and D. P. Lestari, "Speaker recognition using mobilenetv3 for voice-based robot navigation," in *Proc. IEEE ICAICTA*, 2024.

[4] C. Wang *et al.*, "Research on intelligent inspection robot based on neural network and visual slam technology," in *Proc. IEEE ICEACE*, 2023.

[5] Z. Chang, L. Hao, H. Tan, and W. Li, "Design of mobile garbage collection robot based on visual recognition," in *Proc. IEEE AUTEEE*, 2020.

[6] M. Henein, J. Zhang, R. Mahony, and V. Ila, "Dynamic slam: The need for speed," *arXiv preprint arXiv:2002.08584*, 2020, Preprint.

[7] Z. Zhao and X. Zhang, "Research of visual slam in dynamic environment using convolutional neural network," *IEEE Access*, 2022.

[8] Q. Wu and S. Zhou, "Deep learning-driven real-time slam algorithm for dynamic scenarios," *IEEE Access*, 2022.

[9] H. Chen and L. Wang, "Deep learning based visual slam algorithm for indoor dynamic scene," *IEEE Access*, 2022.

[10] T. Xu *et al.*, "Slam algorithm based on improved semantic detection," *Sensors*, 2022.

[11] L. Liu *et al.*, "Dloam: Real-time and robust lidar slam system based on cnn in dynamic urban environments," in *Proc. IEEE ICRA*, 2022.

[12] Y. Hu and M. Li, "Imu and lidar odometry based on cnn and rnn self-supervised learning and attention mechanism," *IEEE Access*, 2022.

[13] L. Chen *et al.*, "Representing objects in map building using a semantic slam scheme," *IEEE Access*, 2021.

[14] J. Zhang *et al.*, "Rgb-d based semantic slam framework for rescue robot," *Proc. IEEE ICAR*, 2021.

[15] Wang *et al.*, "A survey of deep learning application in dynamic visual slam," *IEEE Access*, 2023.

[16] Valente *et al.*, "An lstm network for real-time odometry estimation," *arXiv Preprint arXiv:1902.08536*, 2019.

[17] H. e. Zheng, "Atfvo: An attentive tensor-compressed lstm model with optical flow features for monocular visual odometry," in *Proc. IEEE ICCV Workshops*, 2023.

[18] L. Kang and Y. Sun, "Overfitting reduction of pose estimation for deep learning visual odometry," *IEEE Access*, 2022.

[19] S. Alam and A. Singh, “Monocular visual odometry in mobile robot navigation,” *IEEE Robotics and Automation Letters*, 2021.

[20] Y. Zhao and J. Chen, “Slam for robotic navigation by fusing rgb-d and inertial data in recurrent and convolutional neural networks,” *Sensors*, 2023.

[21] Q. Li and K. Zhou, “Real-time human motion forecasting using an rgb camera,” in *Proc. IEEE ICRA*, 2023.

[22] X. Zhao *et al.*, “Teaching robots to predict human motion,” in *Proc. IEEE IROS*, 2022.

[23] L. Zhou and P. Xu, “Human activity detection employing full-type 2d blazepose estimation with lstm,” *IEEE Access*, 2023.

[24] H. Feng and Q. Liu, “Embedded fall-detection method based on mediapipe and lstm neural network,” *IEEE Sensors Journal*, 2023.

[25] D. Patel and R. Mehta, “A comparative study of recent real-time semantic segmentation algorithms for visual semantic slam,” *IEEE Access*, 2023.

[26] Rai *et al.*, “Inposnet: Context aware dnn for visual slam,” in *Proc. IEEE IRI*, 2023.

[27] H. Qin and T. Zhang, “Mobilenet-ssdv2: An improved object detection model for embedded systems,” *IEEE Access*, 2022.

[28] Z. Li and J. Zhou, “Tiny yolo optimization oriented bus passenger object detection,” *IEEE Access*, 2022.

[29] S. Park and D. Kim, “Embedded real-time pedestrian detection system using yolo optimized by lnn,” *IEEE Access*, 2022.

[30] H. Kim and S. Lee, “Toward real-time object detection on heterogeneous embedded systems,” *IEEE Transactions on Industrial Electronics*, 2023.

[31] A. Pico *et al.*, “Real-time performance benchmarking of tinyml models in embedded systems,” *IEEE Access*, 2023.

[32] SmartFactoryOWL, *Smartfactoryowl – an industry 4.0 research and demonstration platform*, `https://smartfactory-owl.de/?lang=en`, Accessed: October 6, 2025.

[33] K. Ramesh, M. Friesen, T. Facchinetti, and L. Wisniewski, “Multi-sensor SLAM in smart factories: A comparative study on LiDAR and visual techniques using the ROS2 framework,” in *Proceedings of IECON 2025*, IEEE, Madrid, Spain, 2025.

[34] M. Labbé and F. Michaud, “Rtab-map as an open-source lidar and visual slam library for large-scale and long-term online operation,” *Journal of Field Robotics*, 2019.

[35] S. Pyykölä, “Non-lambertian surfaces and their challenges for visual slam,” *IEEE Open Journal of the Computer Society*, 2024. DOI: 10.1109/OJCS.2024.3419832.

[36] X. Yang *et al.*, “Ros 2 communication latency and synchronization analysis for real-time robotic systems,” *IEEE Access*, vol. 11, pp. 123 456–123 465, 2023.

[37] Wojke *et al.*, “Simple online and realtime tracking with a deep association metric,” in *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2017.

http://dx.doi.org/10.25673/121097

Current Developments in Data Spaces

A Case Study on Interoperability Challenges and Solutions

Alfred Barnard[1], Cheng Xin[2] and Mario Thron[3]

Abstract: Data spaces are emerging as a key paradigm for sovereign, secure, and decentralized data sharing, particularly in industrial contexts. While foundational technologies like the Eclipse Data Connector (EDC) provide a common protocol, the landscape is fragmenting into large, domain-specific ecosystems such as Catena-X, which, to ensure security and trust, have developed highly structured frameworks. This necessary standardization, however, presents challenges for interoperability with external systems. This paper addresses the critical challenge of interoperability between these disparate data space implementations. We present a practical analysis based on experiments from the ROBOT-X project, detailing efforts to connect a bespoke EDC instance with both another custom connector and the highly-structured Tractus-X ecosystem. The findings highlight that protocol alignment is insufficient; true interoperability hinges on solving challenges in identity management, credential verification, and governance. We propose a methodical, tool-driven approach using resources like the Tractus-X SDK and its compatibility test suite to bridge these gaps, providing a tangible roadmap for achieving a federated "network of data spaces."

Keywords: Data Spaces, Eclipse Data Connector (EDC), Interoperability, Industrial Automation

1 Introduction

The concept of sovereign, interoperable data spaces represents a paradigm shift for industrial collaboration, promising a future of secure and efficient data exchange across company boundaries. Driven by initiatives like the European Data Strategy [EC22], these ecosystems are designed to unlock the economic and social potential of data while upholding principles of data sovereignty. However, as the landscape matures, a critical challenge is emerging from a success: the proliferation of multiple, independent data space ecosystems. While major initiatives like Catena-X [CX24] create powerful, standardized environments for specific industries, they also introduce specific governance and technical

[1] ifak, Automation & Digitalisierung, Werner-Heisenberg-Str. 1, 39106 Magdeburg, alfred.barnard@ifak.eu
[2] ifak, Automation & Digitalisierung, Werner-Heisenberg-Str. 1, 39106 Magdeburg, cheng.xin@ifak.eu
[3] ifak, Automation & Digitalisierung, Werner-Heisenberg-Str. 1, 39106 Magdeburg, mario.thron@ifak.eu

requirements. While essential for creating a trusted environment, this presents a learning curve for external parties and makes ad-hoc interoperability a complex challenge. [BI23].

This paper addresses the crucial, practical problem of interoperability between disparate data space implementations. We argue that achieving seamless cross-ecosystem communication requires more than just shared open-source protocols like the Eclipse Data Connector (EDC) [EF24a]; it demands a concerted effort to align on identity management, credential verification, and complex onboarding processes.

Drawing on practical experiments from the ROBOT-X project, this paper provides a hands-on analysis of these interoperability challenges. We present two key case studies: a successful but limited connection between two bespoke EDC-based data spaces, and the far more complex endeavour of integrating with the highly structured, Tractus-X-based data space. Through this analysis, we demonstrate a clear, tool-driven methodology for achieving compatibility, highlighting the critical role of resources like the Tractus-X SDK and its official compatibility test suite [EF24b]. This work aims to provide a tangible roadmap for developers and architects working to bridge the gaps between today's emerging data spaces.

2 The Data Space Concept and Core Frameworks

A Data space is a governed ecosystem that facilitates secure and interoperable data sharing between participants based on a foundation of trust and shared rules, as defined by organizations like the International Data Spaces Association (IDSA) [IDSA24]. Unlike traditional data-sharing platforms that often rely on a central intermediary, a datas pace operates on a decentralized model where participants retain control over their data [Ca24]. This principle of data sovereignty is a cornerstone of the European Data Strategy [EC22], which is legally underpinned by regulations such as the Data Governance Act (DGA) and the Data Act [EC24].

At the heart of many modern data spaces lies Eclipse Dataspace Components, which is an open-source framework for creating and operating sovereign, interoperable data spaces [EF24a]. Central to this is the Eclipse Data Connector (EDC). This is the first part where confusion sets in, the Eclipse Data Connector (EDC) is a part of the bigger framework of Eclipse Dataspace Components. In this paper we will be referring primarily to the EDC. The EDC provides the core components for enforcing data usage policies and facilitating

secure data transfers, but it is not a monolithic, out-of-the-box solution. Instead, it serves as a foundational technology that can be adapted and extended to suit specific needs.

This flexibility has led to the rise of major, industry-specific data space implementations. A leading example is Catena-X, the automotive industry's data space, which aims to create a uniform standard for information and data exchange throughout the automotive value chain [CX24]. The open-source software developed for this ecosystem is published under the project name Tractus-X [EF24b]. While based on the EDC, Tractus-X adds a significant layer of governance, specific APIs, and mandatory processes—such as a network-wide identity and onboarding system—to ensure trust and standardization at scale. The tension between the flexibility of a bespoke EDC and the rigid standardization of an ecosystem like Tractus-X is central to the interoperability challenge this paper explores.

3 The Interoperability Landscape: Federation vs. Fragmentation

The long-term vision for data spaces extends beyond isolated, domain-specific ecosystems. The ultimate goal is to create a "network of data spaces"—a federated environment where an industrial manufacturer, for example, can seamlessly consume logistics data from a mobility data space and sustainability data from a circular economy data space [BI23]. This vision of seamless federation promises to unlock exponential value by enabling cross-sectoral innovation [De24].

3.1 The Ideal: A Federated Network of Data Spaces

In a fully federated ecosystem, participants would be able to:

- **Discover Data Across Networks:** A participant in one data space could easily find and query the asset catalogues of another.
- **Establish Trust On-Demand:** Identity and credentials issued by one data space would be verifiable and accepted by others, creating a web of trust. Emerging concepts, such as a common identity wallet [TRX25], aim to solve this by allowing participants to hold and present verifiable credentials that are valid across multiple data spaces.
- **Negotiate Contracts Seamlessly:** The process of policy negotiation and contract creation would follow a standardized protocol, such as the Dataspace Protocol [DP24], regardless of the specific data spaces involved.

This ideal relies on a high degree of technical and semantic interoperability, built upon shared standards and open protocols. However, the current reality is more fragmented.

3.2 The Reality: The Challenge of Integrating Governed Ecosystems

Large-scale data spaces like Catena-X/Tractus-X are designed to ensure a high degree of security, trust, and governance for all members. To achieve this, they have necessarily introduced specific rules and technical requirements that every participant must follow. This structured approach is highly effective for participants within the ecosystem but creates a clear distinction between internal and external systems, making direct integration from the outside a significant engineering task that requires careful planning.

This fragmentation arises from several key factors:

- **Centralized Onboarding and Identity:** To ensure all participants are trusted entities, Tractus-X requires a formal onboarding process where each organization is vetted and assigned a unique Business Partner Number (BPN) [EF24b]. Identity is typically managed through a central Identity Provider, and credentials must conform to the ecosystem's specific standards. An external connector without a BPN or the correct credential format is inherently untrusted.
- **Specific Implementations for Trust:** To guarantee network-wide trust and functionality, Tractus-X provides a specific and well-defined implementation of the EDC. It mandates specific APIs, data models (e.g., for Asset Administration Shells), and communication workflows [EF24b]. A generic, bespoke EDC may not support these specific extensions, causing communication failures even if both use the same core protocol.
- **Governance and Legal Frameworks:** Participation often requires agreement to a common legal framework and rulebook that governs data usage, liability, and compliance. This creates a high barrier to entry for casual or ad-hoc data exchange.

There are a variety of "MX-Topic Groups" [MXTG24]. In some of these topic groups, the topic of data space interoperability is discussed especially in the groups "Data Exchange Technologies" and "Identification of participants in and across data spaces". The outcomes of these topic groups are not yet published at the time of writing this paper but some preliminary outcomes can be discussed. Some of these outcomes are based on the "MX-Port Concept". This concept structures the technical data exchange stack in five layers, as shown in Fig.1.

	Layer	Purpose
L5	Discovery	. . . is used to find business partners, data assets (e.g. devices) or business applications.
L4	Access & Usage Control	. . . is used to ensure, that data providers can define the data access and usage as well as restrict the access and usage of the provided data.
L3	Gate	. . . is used to exchange data in a uniform way.
L2	Converter	. . . provides the semantic model for the data to be exchanged.
L1	Adapter	. . . enables any business application to use the MX-Port.

Fig. 1: Five layers of data exchange, https://factory-x.org/wp-content/uploads/MX-Port-Concept-V1.00.pdf

Further in this concept paper, two configurations are expanded upon. Specifically, the Catena-X and other EDC based data spaces are represented by the “Hercules” configuration, shown in Fig.2.

	Layer	MX-Port “Hercules”[3]
L5	Discovery	Data Space Protocol / Decentral Claims Protocol[4]
L4	Access & Usage Control	
L3	Gate	AAS - REST
L2	Converter	AAS Submodels[5]
L1	Adapter	Application specific

Fig. 2: MX-Port configuration "Hercules", https://factory-x.org/wp-content/uploads/MX-Port-Concept-V1.00.pdf

According to this configuration cross-data space communication is possible only if the L5 and L4 layers are similar. In this paper, similarity implies that both data spaces share the same technology, i.e. EDC.

As a result of the fragmentation, the current landscape is not a single, open network but a collection of highly structured, semi-isolated ecosystems. The work, therefore, lies in building the bridges between these "walled gardens," which is the focus of the practical experiments detailed in the next section.

4 Experimental Interoperability in the ROBOT-X Project

To investigate the practical challenges of cross-data space interoperability, we conducted a series of experiments using the ROBOT-X project. Our goal was to connect our bespoke, EDC-based datas pace with two distinct external systems: another non-standard EDC implementation and a highly-structured Tractus-X data space.

4.1 Case Study 1: Connecting to a Bespoke EDC Data Space

Our first experiment involved connecting the ROBOT-X connector to another EDC-based data space developed by Connect Software. This scenario represents a common case of two independent projects using the same open-source foundation but without a shared governance framework.

- **Setup:** Both data spaces were running independent instances of the Eclipse Data Connector. Neither had a shared Identity Provider or a common rulebook.
- **Method:** To enable initial communication, we had to deliberately lower the security posture. The identity and credential verification mechanisms on both connectors were disabled. This forced the EDCs into a mode of implicit trust, where any connecting party was assumed to be valid.
- **Results:** With trust verification disabled, our ROBOT-X connector successfully communicated with the target EDC. We were able to query its data asset catalog and view the available assets, confirming that the basic discovery protocols were compatible**.**
- **Limitations:** While a technical success, this approach is not viable for a production environment. Disabling security controls defeats the core purpose of a sovereign data space. The next logical steps—policy agreement and contract negotiation—failed because there was no trusted mechanism to manage identities or enforce rules. This experiment clearly demonstrated that a shared protocol is insufficient without a shared trust anchor.

4.2 Case Study 2: Integrating with the Tractus-X Ecosystem

Our second, more ambitious experiment focused on connecting our ROBOT-X connector to a Tractus-X data space, in collaboration with the ROX project. This case study highlights the challenges of integrating with a mature, highly-governed ecosystem.

- **Problem Analysis:** Our initial integration efforts helped us understand the specific requirements of the Tractus-X ecosystem. As expected for a governed data space, our generic connector could not immediately join. The connection

was not established due to two primary factors, which represent the standard security and governance checks of the network:

1. **Missing Identity:** Our connector lacked a Business Partner Number (BPN), the mandatory credential issued during the formal Tractus-X onboarding process.
2. **Technical Incompatibility:** The Tractus-X EDC implementation is more "opinionated" than our flexible connector, expecting specific API calls and data formats that we did not natively support.

- **Proposed Solution and Current Work**: Our strategy is not to discard our connector but to enhance it to become "Tractus-X friendly." This involves a systematic process of adaptation guided by the open-source tools provided by the Tractus-X community [EF24b].

  1. First, we are leveraging the Tractus-X SDK. This SDK acts as a middleware layer, providing a standardized REST API that abstracts the complexities of direct interaction with Tractus-X services. By integrating its "Connector KIT," we can adapt our application logic to communicate with the Tractus-X EDC without needing to perform a complete architectural overhaul.
  2. Second, we are using the Tractus-X EDC Compatibility Tests. This official GitHub repository provides a Docker-based test suite to validate a connector's compliance with the Tractus-X specifications. By integrating these tests into our development workflow, we can methodically identify and fix incompatibilities. Our current work is focused on successfully passing this entire test suite, which serves as a critical milestone for achieving verifiable compatibility and being accepted into the Tractus-X network.

5 The End-to-End Data Sharing Workflow

The ultimate goal of achieving interoperability is to enable the seamless, end-to-end data sharing workflow as envisioned by the EDC architecture. The challenges detailed in our experiments directly impact the feasibility of this workflow across different data spaces. The workflow, consists of three main phases:

1. **Asset Management:** A data provider first publishes a data asset, defines a set of usage policies (e.g., restricting use to a specific purpose or time frame), and creates a contract definition that links the asset to the policy. This phase is typically straightforward within a single data space but becomes complicated when the asset needs to be discoverable by external participants who may not understand the local catalog's structure.

2. **Contract Negotiation:** A data consumer, having discovered an asset, requests a contract. This initiates an automated negotiation process. The provider offers a contract, and if the consumer agrees to the terms, a formal contract agreement is signed and stored. This phase is the most significant hurdle in cross-data space communication, as it requires a shared understanding of policies and, crucially, a trusted identity mechanism to validate that both parties are who they claim to be. Our first case study failed at this stage due to a lack of trust.
3. **Secure Data Transfer:** With a valid contract agreement in hand, the consumer requests the data transfer. The provider's EDC verifies the contract and issues a temporary authorization token. The consumer then uses this token to access the data directly from the provider's backend system via the data plane. This final phase is typically standardized, but it is entirely dependent on the successful completion of the preceding negotiation phase.

6 Discussion: Navigating Towards a Federated Future

Our experiments reveal a fundamental tension in the current data space landscape: the relationship between the robust standardization offered by large ecosystems like Tractus-X and the flexibility of bespoke implementations. The standardization within Tractus-X is a powerful enabler of security and trust at scale. A natural consequence of this robust governance is that it requires a dedicated integration effort from external parties, in contrast to more flexible, ad-hoc connections. Conversely, independent EDC deployments offer agility but struggle to establish trust with other systems, limiting their reach.

The path to a federated future lies not in choosing one model over the other, but in building the bridges between them. Our work suggests that this is an engineering challenge that can be solved with a methodical approach. Open-source tools developed by the major ecosystems themselves, such as the Tractus-X SDK and compatibility test suites, are the key enablers for this integration. They provide a clear, verifiable path for external systems to become "good citizens" in a governed data space without requiring them to adopt the entire technology stack from the outset. This "adapter" approach appears more practical than waiting for a single, universal standard to emerge.

Further steps and protocols enabling cross-data space communication will become clearer as the MX-Topic groups make their findings and results public.

7 Conclusion and Future Work

Data spaces hold immense promise for the future of industrial automation, but their true potential will only be unlocked through interoperability. This paper has provided a practical, hands-on analysis of the critical challenges hindering the creation of a federated

"network of data spaces." Through case studies in the ROBOT-X project, we have demonstrated that while shared protocols like the EDC are a necessary foundation, they are not sufficient. The primary hurdles are rooted in differences in governance, identity management, and technical specialization between ecosystems.

Our work presents a clear, tool-driven methodology for bridging these gaps, centered on adapting external connectors to meet the requirements of governed data spaces like Tractus-X. By leveraging official resources such as the Tractus-X SDK and compatibility tests, we have defined a concrete path toward achieving verifiable interoperability.

Our future work will proceed along three main tracks:

1. **Achieve Full Compatibility:** Our immediate goal is to successfully pass the complete Tractus-X EDC Compatibility Test suite, leading to a live, end-to-end data exchange with a Tractus-X participant.
2. **Implement Complex Policy Negotiations:** Once basic connectivity is established, we will focus on testing and implementing more complex, cross-data space policy and contract negotiations.
3. **Investigate Decentralized Identity:** As a long-term solution to the trust problem, we will explore the integration of decentralized identity solutions and verifiable credentials, aligning with the emerging concept of a common identity wallet to move beyond the limitations of centralized onboarding.

References

[BI23] Barbara IoT: Federated Data Spaces: A key piece for scaling innovation in the Industry. Barbara IoT Blog, 2023. URL: https://www.barbara.tech/blog/industrial-data-spaces-the-importance-of-data-in-the-industry

[Ca24] Capgemini: Data spaces mark a new era in European data sharing. Capgemini Research, 2024. URL: https://www.capgemini.com/insights/expert-perspectives/data-spaces/

[CX24] Catena-X Automotive Network: Official Website and Documentation. Catena-X e.V., 2024. URL: https://catena-x.net/en/

[De24] Deshmukh, R. A. et al.: Challenges and Opportunities for Enabling the Next Generation of Cross-Domain Dataspaces. CEUR Workshop Proceedings, 2024. URL: https://ceur-ws.org/Vol-3705/short12.pdf

[DP24] Dataspace Protocol: The Dataspace Protocol. AWS Prescriptive Guidance, 2024. URL: https://docs.aws.amazon.com/prescriptive-guidance/latest/strategy-building-data-spaces/dataspace-protocol.html

[EC22] European Commission: A European strategy for data. Shaping Europe’s digital future, 2022. URL: https://digital-strategy.ec.europa.eu/en/policies/strategy-data

[EC24] European Commission: The Data Act. Shaping Europe's digital future, 2024. URL: https://digital-strategy.ec.europa.eu/en/policies/data-act

[EF24a] Eclipse Foundation: Eclipse Dataspace Connector. Eclipse Foundation Projects, 2024. URL: https://projects.eclipse.org/projects/technology.dataspaceconnector

[EF24b] Eclipse Foundation: Eclipse Tractus-X. Eclipse Foundation Projects and GitHub Repository, 2024. URL: https://projects.eclipse.org/projects/automotive.tractusx

[IDSA24] International Data Spaces Association: Official Website. IDSA e.V., 2024. URL: https://internationaldataspaces.org/

[TRX25] PoC Multi-Dataspace Credentials Support (Issuance & Holding), https://github.com/eclipse-tractusx/sig-release/issues/1477

[MXTG24] MX – Topic Groups, https://mx-guidanceboard.org/topic-groups/

[MXPC25] Factory-X introduces the MX-Port concept, 2025. https://factory-x.org/factory-x-introduces-the-mx-port-concept/

http://dx.doi.org/10.25673/121097

Leistungsbewertung von DECT NR+ für industrielle Anwendungen

Technologieauswahl, -inbetriebnahme und Einbettung in Lernsituationen

Timo Siekmann[1], Björn Kroll[2], Denis Gustin[3]

Abstract: Dieser Beitrag untersucht die Leistungsbewertung der neuen Funktechnologie DECT NR+ im Kontext industrieller Anwendungen. Als seit 2021 von der ITU anerkannter 5G-Standard bietet DECT NR+ durch sein weltweit verfügbares Frequenzspektrum im Bereich um 1,9 GHz sowie seine nicht-zellulare Mesh-Topologie eine kosteneffiziente Alternative zu 5G-Campusnetzen oder WLAN-Systemen. Ziel der Arbeit ist es, DECT NR+ anhand definierter KPIs – insbesondere Latenz, Datenrate und Interferenzrobustheit – in Indoor- und Outdoor-Szenarien zu evaluieren. Dafür wurden Messungen mit DECT NR+-fähigen Entwicklungsboards durchgeführt und mit etablierten Technologien wie WLAN 6 und 5G NR verglichen. Die Ergebnisse zeigen ein hohes Potenzial der Technologie hinsichtlich Reichweite und Energieeffizienz, jedoch Einschränkungen bei Mobilität und Geräteverfügbarkeit. Insgesamt wird DECT NR+ als vielversprechende Ergänzung im Spektrum industrieller Funklösungen bewertet, insbesondere für zeitkritische IoT- und Automatisierungsanwendungen.

1 Einleitung

Ein Ziel von Industrie 4.0 ist die automatisierte Fertigung individueller Produkte. Bereits heute werden auch für kleine Losgrößen Automatisierung und Roboter eingesetzt. Stationäre Roboter werden dabei zunehmend durch AGVs (Autonomous Guided Vehicles) sowie AMRs (Autonomous Mobile Robots) ergänzt. Diese Art von freibeweglicher Robotik erfordert eine durchgehende drahtlose Kommunikation. [1]

Die Integration von drahtlosen Kommunikationssystemen ist von Herausforderungen geprägt, wie zum Beispiel fehlenden Maschinenschnittstellen, IT/OT-Integrationsproblemen, Wartungs- und Bedienbarkeit, dem Gefahrenpotenzial für Interferenzen mit anderen Systemen sowie zuletzt von hohen Investitionskosten (Capital Expenditures, CapEx) und Betriebskosten (Operating Expenses, OpEx). [2] Insbesondere infrastrukturbasierte Systeme wie 5G-Campusnetze oder controllerbasierte WLAN-

[1] Fraunhofer IOSB-INA, Digitale Infrastrukturen, Campusallee 1, 32657 Lemgo, timo.siekman@iosb-ina.fraunhofer.de
[2] Fraunhofer IOSB-INA, Digitale Infrastrukturen, Campusallee 1, 32657 Lemgo, timo.siekman@iosb-ina.fraunhofer.de
[3] Fraunhofer IOSB-INA, Digitale Infrastrukturen, Campusallee 1, 32657 Lemgo, timo.siekman@iosb-ina.fraunhofer.de

Systeme sind komplex und teuer. Hinzu kommen Herausforderungen für System- und Komponentenhersteller, ihre Produkte in einem weltweiten Lieferantennetzwerk zu vertreiben, was oftmals spezielle Ländervarianten, wie zum Beispiel für die EU, die USA und Asien, erfordert, um diverse Anforderungen von Funkregulierungsbehörden wie der FCC oder der BNetzA gerecht zu werden und letztendlich die Geräte zertifizieren zu lassen. Vor diesen Herausforderungen suchen Forscher, Entwickler und Hersteller weltweit kontinuierlich nach Lösungen. Dadurch steigt die Produktvielfalt an drahtloser Kommunikationstechnologie. Dies wird durch einen großen Anstieg neuer Technologien und Produktvarianten untermauert. Eine dieser neuen Technologien ist DECT NR+. DECT NR+ ist seit 2021 ein von der International Telecommunication Union (ITU) anerkannter 5G-Standard.

Neben der Ernennung zum 5G-Standard ist DECT aufgrund des weltweit verfügbaren Frequenzspektrums interessant. [3] Hervorzuheben ist, dass DECT NR+ anders als die Mobilfunktechnologien aus der 3GPP-Standardisierung kein zellular basiertes Kommunikationssystem ist. Daraus ergibt sich, dass DECT ohne ortsfeste (Mobilfunk-) Infrastruktur arbeitet und eine Mesh-Kommunikation zwischen den einzelnen Geräten aufbaut. Dadurch leitet sich ein geringer Gesamtsysteminvest ab. [4]

1.1 Ziel, Vorgehen und Struktur dieses Beitrages

Das Ziel des Papers ist es, die Technologie DECT NR+ in der Landschaft der drahtlosen Kommunikationssysteme für Schmal- und Breitbandkommunikation in der OT einzuordnen und anhand definierter KPIs mittels Berechnungen und Messungen zu evaluieren. Dazu werden Messungen im Indoor-Produktionsumfeld sowie im bebauten Outdoor-Umfeld durchgeführt. Der Fokus liegt dabei auf Schlüsselparametern wie Latenz, Durchsatz und Interferenzrobustheit in Indoor- und Outdoor-Szenarien. Zu Beginn wird die Technologie DECT NR+ mit den bereits etablierten Technologien in diesem Umfeld verglichen und bewertet.

Im Kapitel 2 wird der technologische Stand von DECT NR+ vorgestellt. In Kapitel 3 wird zunächst eine theoretische Einstufung von DECT NR+ mittels technologieunabhängiger KPIs wie Spektrum, Sendeleistung, Netzwerktopologie, Zugriffsverfahren sowie Mechanismen zur Garantie von Echtzeit und Robustheit dargestellt. Anschließend wird die Testumgebung beschrieben. Diese besteht aus zwei DECT NR+-fähigen Boards, inklusive PHY, MAC und Network Layer Integration, sodass mit Hilfe eines Traffic Generators QoS (Quality of Service) Parameter für eine Leistungsbewertung gemessen werden können. Diese QoS-Parameter bestehen aus Datenrate, Latenz, Jitter sowie Framelossrate. Zudem sollen die Messungen unter dem Einfluss verschiedener Reichweiten ausgeführt werden, um die theoretischen Reichweiten empirisch im industriellen Einsatz zu belegen.

In Kapitel 4 werden die Messergebnisse mit anderen Technologien wie WLAN und 4G/5G-Campusnetzen verglichen. Abschließend wird ein Fazit zum aktuellen technischen

Stand sowie zur Eignung für industrielle Anwendungen gegeben, das einen Ausblick auf weitere Forschungsfragen bietet.

2 Stand der Technologie

3 DECT NR+

3.1 Theoretische Einstufung von DECT NR+

Im Folgenden wird DECT NR+ mit Wi-Fi, Bluetooth und dem 5G NR-Standard verglichen, da DECT NR+ Use Cases in dem Bereich adressiert, die mit diesen Technologien bedient werden sollen. Die vier Technologien werden miteinander verglichen, da sie die häufigsten verwandten Technologien in der drahtlosen Kommunikationstechnik für Automatisierungsprotokolle sind.

Parameter	**Bluetooth (BLE)**	**DECT NR+**	**WLAN 6**	**5G NR (Campusnetz)**
Center Frequency	2,4 GHz	1990 MHz	2,4 GHz, 5GHz, 6 GHz	3,7 GHz
Freq. Bandbreite [max]	2 MHz (BLE)	13,8 MHz	160 MHz	100 MHz
Kanalzugriff	FHSS	TDMA/ FDMA	OFDMA	OFDMA
Duplexing Verfahren	TDD	TDD	TDD	TDD

TX Power	Claas 1: 20 dBm Claas 2: 4 dBm Claas 3: 0 dBm	Class 1: 23 dBm Class 2: 19 dBm Class 3: 10 dBm	@2.4 GHz 20 dBm @5 GHz 23 dBm @5 GHz 23 dBm (indoor)	*Class 1: 29 dBm *Class 2: 26 dBm *Class 3: 23 dBm
Datenrate	2 Mbit/s	3,4 Mbit/s	9608 Mbit/s	~ 1000 Mbits
Topologie	Punkt zu Punkt, Broadcast, Mesh	Punkt zu Punkt, Stern, Mesh	Punkt zu Punkt, Stern, Mesh	Zentral gesteuert, Sidelink

[5] [6] [7] [8] Tabelle 1: Drahtlose OT Technologien im Vergleich

Im Vergleich mit den in der Tabelle aufgeführten Technologien fällt auf, dass die Technologie DECT NR+ mit 20 MHz die geringste Bandbreite aufweist. Jedoch sticht hervor, dass die Frequenz im Sub-2-GHz-Bereich liegt und sich dadurch gut eignet, Daten über längere Distanzen energiearm zu übertragen. Die Frequenzbandbreite ist nach Bluetooth LE die geringste und ordnet sich damit im Bereich der Schmalbandkommunikationen ein.

Die Technologie DECT NR+ nutzt das Frequenzband zwischen 1980 MHz und 2000 MHz. Dieser Frequenzblock ist in einzelne Bänder unterteilt. Durch die Nutzung des DECT-Bands können die Geräte nahezu weltweit eingesetzt werden, sodass ein Hersteller von Kommunikationstechnik nicht mehrere Länderversionen produzieren und warten muss. Maschinen- und Anlagenhersteller können ihre Maschinen unabhängig vom Zielland ausrüsten. Zudem ist die DECT-Frequenz nicht von Gleichkanalstörern gestört, wie sie im WLAN-Bereich bei 2,4 GHz und 5 GHz durch konkurrierende Technologien oder Netzwerke vorkommen. Die Technologie DECT NR+ ist für die Koexistenz mit den aktuell noch verwendeten DECT-Telefonen geeignet. Mechanismen wie LBT (Listen Before Talk) reduzieren das Potenzial von Funkinterferenzen zwischen den Technologien.

3.2 Messaufbau

Der Messaufbau umfasst zwei Boards mit einem Raspberry Pi 4 Computing-Modul, das auf einer Platine mit einem nRF9161 verbaut ist. Das Betriebssystem des Raspberry Pi ist ein Linux OS, welches mit seinen Host-Daemons die IP-Einstellungen parametrisiert und die Applikationen startet. Das Dev Board stammt von RFmondial und beinhaltet

grundlegende Software-Stack-Implementierungen für MAC und IP sowie Beispielapplikationen. Da Nordic Semiconductor der einzige Chiphersteller für DECT NR+ ist, der aktuell Dev Boards oder Chips vertreibt, und mit Last Mile Semiconductor nur ein weiterer Chiphersteller angekündigt wird, ist die Implementierung der Protokollstacks im Vergleich zu etablierten Technologien auf einem geringeren Technology Readiness Level.

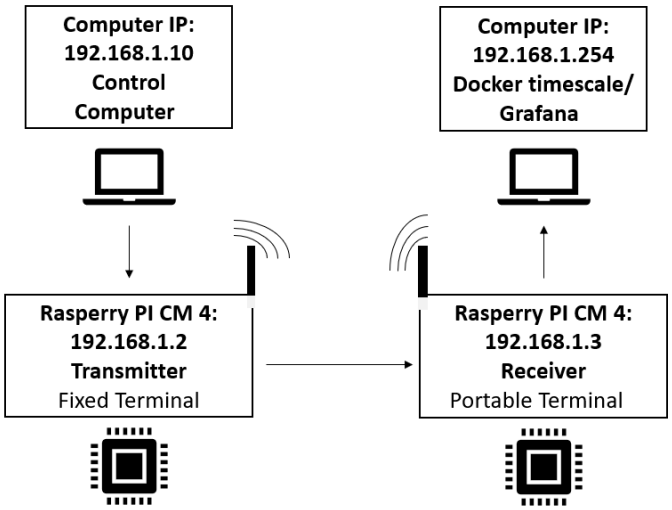


Abbildung 1: Aufbau DECT NR+ Testumgebung

Der Messaufbau umfasst zwei DECT NR+-Modems, wobei ein Modem im FT (Fixed Terminal) und das andere im PT (Portable Terminal) Modus agiert. Das FT-Board sendet die Payload von 70 Bytes an den nRF9161-Chip, welcher das DECT-Band 1 im LBT-Vorgang abhört. Sobald der Chip keinen Pegel größer als -60 dBm registriert, wird das Paket über die Luftschnittstelle an das PT-Board geschickt.

In dem Messframework der Firma rfmondial ist auch eine Grafana- und Timescale-DB-Docker-Anwendung enthalten, die die Parameter SNR (Signal-to-Noise-Ratio), RSSI (Relative Signal Strength Indicator) sowie die Packet Error Rate auswertet. Der aktuelle Stand der Boards ermöglicht die unidirektionale Kommunikation, sodass keine Messungen mit einem Industrial Ethernet-Protokoll wie PROFINET IO durchgeführt werden können. Jedoch kann mittels des angepassten SDK (Software Development Kit Payload) zwischen den Boards übertragen werden, um erste Erkenntnisse zur Packet Error

Rate in Abhängigkeit von der Reichweite in industriellen Umgebungen sowie im Freien zu gewinnen.

Der Messaufbau verbindet die beiden Boards mit einem Switch. Über Ethernet-Kommunikation können die Boards von einem Computer, der ebenfalls mit dem Switch verbunden ist, per Remote-Verbindung konfiguriert werden. Auf dem Raspberry Pi Computing-Modul werden Konfigurationsdateien im YAML-Format parametrisiert und an den Nordic nRF9161 gesendet. Der DECT NR+ bootet neu und übernimmt dann die Konfiguration.

3.3 Messablauf

Die Messungen wurden sowohl indoor als auch outdoor durchgeführt. Dabei wurde der Transmitter als mobiles Gerät genutzt und örtlich bewegt, während der Receiver ortsgebunden war. Mittels der Datenansicht im Grafana-Dashboard sowie den Datenbankeinträgen konnten die Messwerte beobachtet werden.

Die Messung indoor wurde mit einer Sendeleistung von 16 dBm durchgeführt. Die Outdoor-Messungen wurden mit einer Sendeleistung von 16 dBm sowie 19 dBm durchgeführt, wobei 19 dBm aktuell den maximalen Leistungspegel des nRF9161-Chips darstellt. Durch technische Limitierungen ist der im Standard angegebene Leistungspegel von 23 dBm derzeit nicht realisierbar.

Aufgrund der Vielzahl an Möglichkeiten in den Kombinationen der unterschiedlichen Parameter wurden Annahmen getroffen, die den Einstellungen eines späteren industriellen Nutzers entsprechen könnten. Die genutzten Parameter sind wie folgt:

- Payloadgröße: 70 Bytes
- Protokoll: UDP
- DECT Bandnummer: 1
- LBT: off
- Packet Length Type: 1
- Packet Length m1: 2
- Modulation and Coding Scheme: 2 (QPSK)
- Frame Subslots 48
- Frame Subslots ULDL 48

3.4 Messergebnisse

Die ersten Testmessungen zeigen deutlich, dass die vom Hersteller angegebenen Limitierungen und Probleme in der Mobilität der Geräte vorhanden sind. Dabei ist zu beobachten, dass bei einer Entfernungsänderung in Schrittgeschwindigkeit zwischen den beiden nRF9161-Boards ein deutlicher Paketverlust bis fast zum totalen Paketverlust auftritt. Sobald die Boards örtlich gebunden sind, lässt der Paketverlust nach und fällt in

Abhängigkeit vom SNR und RSSI auf 0 zurück. Eine weitere Beobachtung im Indoor-Szenario ist der geringe Pfadverlust zwischen dem Sender und Empfänger.

Im Outdoor-Szenario wurde ein besserer Pfadverlust erreicht, als es mittels empirischer Modelle berechnet wurde. Laut dem Modell aus 3GPP TR 38.901 V15.1.0, das auch zuvor für 5G-basierte Feldtests genutzt wurde.[9]

Laut dem Rural Macro Modell beläuft sich der errechnete Pfadverlust, wie im Folgenden dargestellt, auf 95 dB im LOS (Line of Sight)-Szenario. Der gemessene Pfadverlust, berechnet durch die Differenz zwischen Sendeleistung und RSSI-Wert am Empfänger, beträgt im Worst Case:

Sendeleistung Indoor: 12 dBm

Entfernung Indoor: 25m

Pfadverlust Indoor LOS Messung:

$$PL_{LOS\_Indoor\_Messung} = 12\ dBm\ (Sendeleistung) - (-48\ dBm) Empfangspegel$$

$$PL_{LOS\_Messung} = 60\ dB$$

PL Berechnung:

$$PL_{LOS} = 31.84 + 21.50 \log_{10}(d_{3D}) + 19.00 \log_{10}(f_c)$$

$$PL_{LOS} = 31.84 + 21.50 \log_{10}(30m) + 19.00 \log_{10}(1.95\ GHz)$$

$$PL_{LOS} = 67\ dB$$

Daraus ergibt sich im Indoor Szenario eine Abweichung von 7 dB, um die die realen Messwerte besser sind als die Berechnung. Dadurch kann die Entfernung zwischen den DECT NR+ Geräten, in dem mit einer geringen Paketverlustrate empfangen werden kann, hingegen der Simulation erhöht werden.

Outdoor Messung:

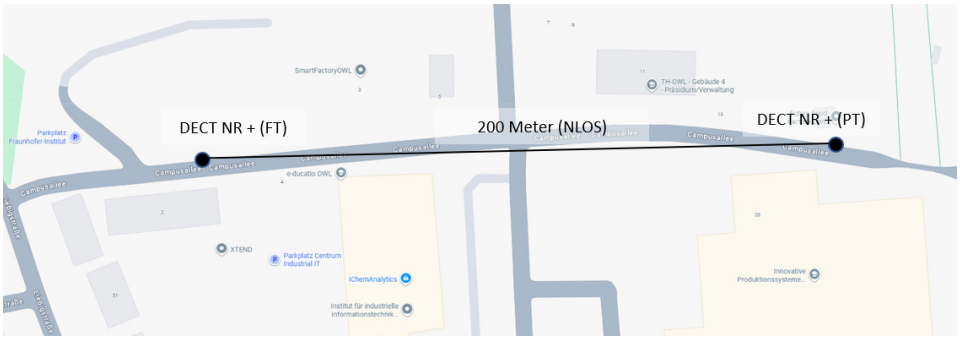


Abbildung 2: Outdoor-Messbereich vor der SmartFactoryOWL, Lemgo

Im Outdoorbereich wurde eine maximale Distanz von 200 Meter erreicht. Die Berechnung ergibt hier einen Pfadverlust von 153 dB.

$$PL'_{RMa-NLOS} = 161.04 - 7.1log_{10}(W) + 7.5log_{10}(h) - \left(24.37 - 3.7\left(\frac{h}{h_{BS}}\right)^2\right)log_{10}(h_{BS}) + \left(43.42 - 3.1log_{10}(h_{BS})\right)(log_{10})(d_{3D}) - 3) + 20log_{10})(f_c) - 3.2(log_{10})(11.75h_{UT}))^2 - 4.97)$$

$$PL'_{RMa-NLOS} = 161.04 - 7.1log_{10}(W) + 7.5log_{10}(1.5m) - \left(24.37 - 3.7\left(\frac{1.5m}{1.5m}\right)^2\right)log_{10}(1.5m) + \left(43.42 - 3.1log_{10}(1.5m)\right)(log_{10})(200m) - 3) + 20log_{10})(1.95\ Ghz) - 3.2(log_{10})(11.75 * 1.5m))^2 - 4.97)$$

Der Pfadverlust beträgt 138 dB.

Die Messungen hingegen zeigen einen maximalen Pfadverlust von

$$PL_{NLOS\_Outdoor\_Messung} = 16\ dBm\ (Sendeleistung) - (-97\ dBm) Empfangspegel$$

$$PLN_{LOS\_Outdoor\_Messung} = 113\ dB$$

In dieser Betrachtung wird die Abweichung um ein Vielfaches deutlich zwischen der Berechnung und der realen Messung im Outdoor NLOS.

Bei maximaler Distanz hat die Oneway Paketverlustrate 7 % betragen und die maximale Oneway Latenz lag bei 20 ms. Hingegen die maximale Latenz bei 0 % Packetverlustrate bei 10,058 ms lag. Die Entfernung hat 100 Meter betragen und der RSSI Wer am Empfänger lag konstant bei -70 dBm. Das zeigt eine deutliche Vervielfachung der Latenz sowie der Packetverlustrate und untermauert, dass das DEV Board mit der aktuellen PCB sowie Antennen Konfiguration die Receiver Sesitivity nicht ausschöpfen kann. Diese wird mit -99,7 dBm in dem Standard angegeben. [10]

4 Zusammenfassung und Ausblick

DECT NR+ stellt eine vielversprechende Funktechnologie dar, die insbesondere durch die Nutzung von weltweit nahezu einheitlich verfügbarem, kostenfreiem, aber reguliertem 20-MHz-Spektrum hervorsticht. Diese Ressource garantiert eine stabile Nutzbarkeit und reduziert gleichzeitig das Risiko von Gleichkanalstörungen sowie Interferenzen durch konkurrierende Technologien, wie sie im ISM-Band bei 2,4 GHz oder im UNII-Band bei 5 GHz auftreten. Darüber hinaus integriert DECT NR+ fortschrittliche Mobilfunktechniken, die niedrige Latenzen ermöglichen. Ein Beispiel hierfür ist HARQ, durch das fehlerhafte Übertragungen innerhalb von nur 417 µs auf PHY-Ebene erneut zugestellt werden können. Ergänzt wird dies durch die Unterstützung von bis zu 8×8 MIMO gemäß ETSI-Standard, was eine hohe Robustheit und Flexibilität bei der Datenübertragung eröffnet.

Demgegenüber stehen jedoch signifikante Einschränkungen: Zum einen existieren bislang nur wenige technische Implementierungen sowie eine sehr begrenzte Anzahl an Herstellern, darunter vor allem Nordic Semiconductor und Last Mile Semiconductor. Zum anderen erschweren bestehende Patente auf zentrale Chipfunktionen sowie die starke Marktdurchdringung etablierter Standards wie IEEE 802.11 (WLAN) oder 3GPP-5G eine breitere industrielle Adaption.

4.1 Ausblick

Sollte es gelingen, DECT NR+ in marktreife Produkte vollständig zu integrieren, könnte der Standard eine strategisch wichtige Nische besetzen, die Verbindung zwischen drahtlosen LPWAN-Technologien mit sehr niedrigen Datenraten und großflächiger Reichweite einerseits sowie hochperformanten WAN-Systemen mit höherer Datenrate andererseits. Mit seiner Kombination aus garantierter Frequenzressource, niedriger Latenz und moderater Datenrate ist DECT NR+ prädestiniert für industrielle Anwendungen, zeitkritische Kommunikationsszenarien und IoT-Umgebungen, die bisher von bestehenden Standards nicht optimal abgedeckt werden. Der zukünftige Erfolg hängt jedoch maßgeblich von einer weiteren Verbreitung der Hardware, der Überwindung

patentbedingter Markteintrittsbarrieren und der Etablierung eines breiteren Hersteller-Ökosystems ab.

5 Literaturverzeichnis

[1] A.-L. e. a. Kampen, „Case study of AGV in Industry 4.0 environments–an evaluation of wireless communication protocols,“ IEEE, Prague, Czech , 2022.

[2] C. e. a. Kalalas, „echno-economic study on capillary networks and cellular technologies for machine-to-machine communications.,“ KTH Royal Institute of Technology, Stockholm, Sweden, 2014.

[3] D. Forum, „DECT Forum,“ [Online]. Available: https://www.dect.org/dect-technology. [Zugriff am 20 03 2025].

[4] S. M. I. B. J. ,. J.-X. C. Lorenzo Vangelista, „DECT NR+: Unveiling the Essentials of a new non-cellular 5G Standard for Verticals,“ VDE, Osnabrueck, 2022.

[5] B. SIG, „ Bluetooth SIG,“ 2025. [Online]. Available: https://www.bluetooth.com/de/learn-about-bluetooth/tech-overview/. [Zugriff am 09 08 2025].

[6] N. Semiconductor, „Nordic Semiconductor DECT NR+,“ Nordic Semiconductor, 2025. [Online]. Available: https://docs.nordicsemi.com/bundle/ps_nrf9151/page/dect.html. [Zugriff am 06 09 2025].

[7] Intel, „Intel Wi-Fi,“ [Online]. Available: https://www.intel.de/content/www/de/de/gaming/resources/wifi-6.html. [Zugriff am 23 08 2025].

[8] N. Al-Falahy und O. Y. Alani, „IEEE Xplore,“ 02 01 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7839836. [Zugriff am 16 07 2025].

[9] B. R. Timo Siekmann, „Feldtest eines vernetzten autonomen Einschienenfahrzeugs,“ Universität Halle, Magdeburg, 2023.

[10] A. Anttonen, P. Karhula, M. Lasanen und M. Majanen, „publications.vtt.fi,“ 01 01 2021. [Online]. Available: https://publications.vtt.fi/julkaisut/muut/2021/VTT-R-00367-21.pdf. [Zugriff am 10 09 2025].

http://dx.doi.org/10.25673/121097

Datenredunktionsstrategien bei Feldbusübertragungen wie Profinet über Funk erhöhen die Anlagenverfügbarkeit

Autor Thomas Schildknecht

Vorstand Schildknecht AG

Abstrakt

Steuerungen kommunizieren über Feldbusse wie Profinet mit dezentralen Komponenten wie Antriebe und dezentrale Io-Peripherie. Typischerweise erfolgt dies bei einer Kabelverbindung in einem einstellbaren Zeittakt, die Profinet Aktualisierungszeit. Bei Kabelverbindungen wird diese Zeit üblicherweise auf 1ms eingestellt. Wird statt einer Kabelverbindung jedoch eine Funkverbindung, wie z.B. WLAN, 5G Campusnetz oder Bluetooth, zur Übertragung verwendet muss diese Zeit durch try-and-error angepasst werden. Einen grossen Einfluss auf die Funkübertragungsparameter wie Latenz haben dabei die spezifischen Eigenschaften und gesetzlichen Vorgaben der jeweiligen Funktechnologie. Für eine stabile Feldbuskommunikation erhöht man bei einer Funkverbindung die PN Aktualisierungszeit typisch auf 32 oder 64ms. Letztlich führt dies zu einer Reduzierung der Kommunikationslast, auf Kosten einer Verschlechterung der Echtzeitfähigkeit und der Notwendigkeit einer Änderung im Anlagenprojekt. Es bedarf also einer Softwareänderung im Vergleich zu einer Kabelverbindung.

Ein anderer Lösungsansatz soll hier vorgestellt werden. Vergleichbar mit der Datenreduktion bei Audio, bekannt als MP3, wird hier ein Ausfiltern der eingehenden redundanten Feldbustelegramme realisiert. Dadurch kann erreicht werden daß die Feldbus Aktualisierungszeit wie bei einer Kabelverbindung erhalten bleibt und keine Projektänderungen durchgeführt werden muss, was die Akzeptanz für Funklösungen bei den SPS Programmierern erhöht. Diese Lösungsvariante soll hier näher beschreiben werden.

Ausgangslage Transparente Feldbusübertragung über ein Funkmedium

Die folgende Skizze zeigt den zeitlichen Ablauf einersFeldbustelegramms wie Profinet über einen Funkkanal. Die Vertikale Achse ist die Zeit. Dauert die Übertragung bei einer Kabelverbindung nur Nanosekunden, benötigt ein Telegramm über Funk im Bereich von 100us bis 10ms ist also um Faktor mindestens 10000 länger. Man kann die Zeit vom Aussenden von der SPS bis zum Empfang beim Remote-IO als Übertragungszeit der Funkstrecke bezeichnen.

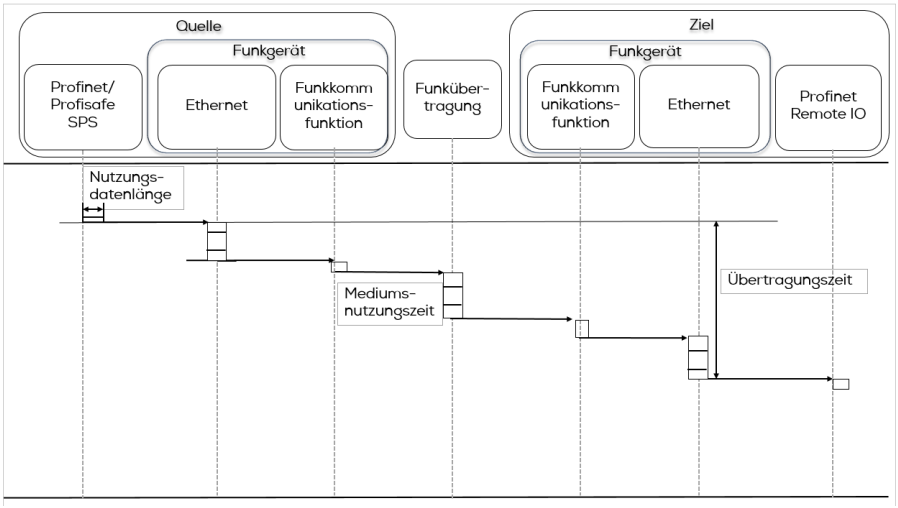


Bild 1 Telegrammaufbau bei einer Funkübertragung (Quelle VDI Richtlinie 2185-4)

Profinet Telegramme werden im Takt des PN Sendezeitintervalls wiederholt, typisch 1ms. Die Herausforderung in der Praxis ist es nun das Sendeintervall so gering wie möglich zu wählen um eine hohe Applikationsfrequenz (Reakionszeit) zu ermöglichen. Da die Latenz der Funkübertragung jedoch keine feste Grösse ist und einer Vielzahl von Einflüssen unterliegt, kann die Einstellung eine gerade noch passenden Sendeintervalls langwierig und fehleranfällig sein. Wählt man das Sendeintervall niedrig, läuft man Gefahr dass die Funkstrecke mit der Datenübertragung zeitlich nicht mehr folgen kann. Dies führt zu einem Busfehler und somit zu einem Anlagenstillstand, oft von Minuten. Man muss also eine Einstellung für das Sendeinterfall finden die einen Kompromissdarstellt zwischen Anzahl der ungewollten Anlagenstillstände und der erreichbaren Applikationsfrequenz, also wie lange des dauert bis ein Ausgang am Remote-Io geschaltet wird oder Eingang in der SPS abgebildet wird.

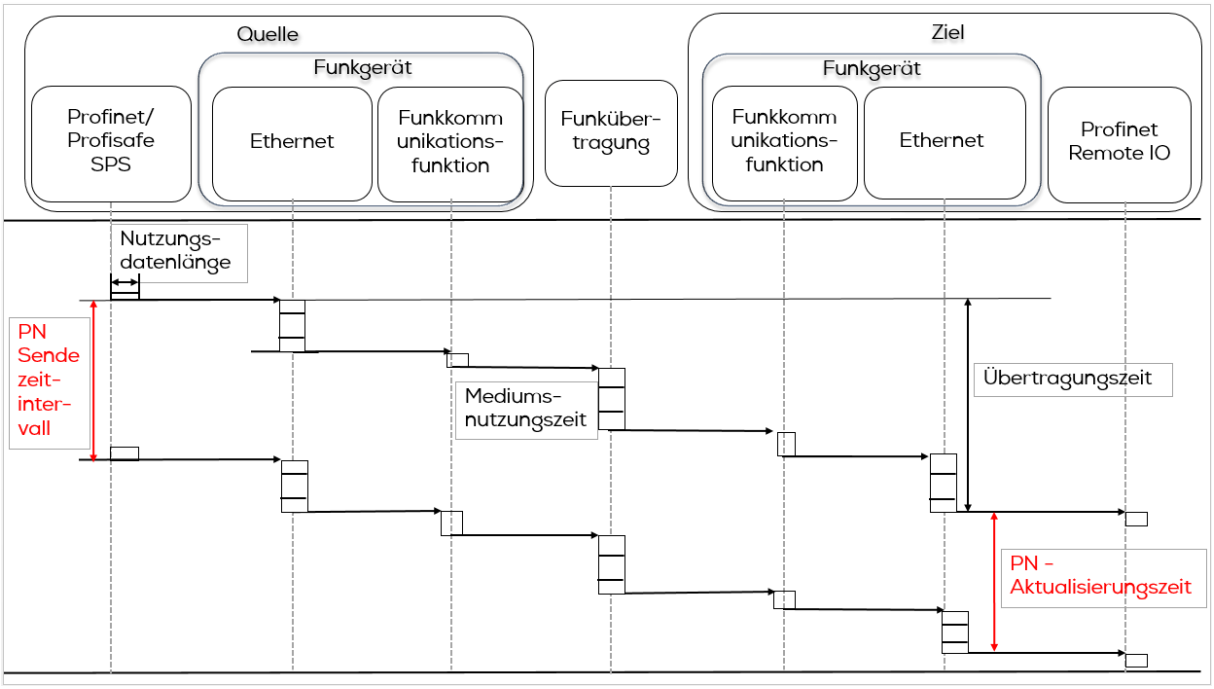


Bild 2 Telegrammaufbau bei einer Funkübertragung (Quelle VDI Richtlinie 2185-4)

Erschwerend kommt bei der Suche nach einer für die Anwendung noch akzeptablen Aktualisierungszeit hinzu, daß die Medien Nutzungszeit nicht konstant ist. Dies führt zu einem Jitter um den Mittelwert der Übertragungszeit. Es gibt Funktechnologien wo dieser Jitter typisch nur einige wenige Millisekunden beträgt, es gibt jedoch auch Funktechnologien die einen inakzeptabel hohen Jitter erzeugen können.

Dies hängt wiederum mit den gesetzlichen Vorgaben für die Nutzung des Frequenzbandes ab. Breitbandfunktechnologien wie WLAN müssen regulatorisch ein Listen-before-talk Prinzip anwenden (CSMA/CA - Carrier Sense Multiple Access/Collision Avoidance). Dies kann dazu führen, dass eine undefiniert lange Zeit der Sender ein Telegramm zurückhält und damit verzögert wenn der eigenen Receiver eine Energie anderer Sender auf seinem Kanal feststellt. Bei WLAN sind Jitter bis in den 1000ms Bereich messbar. In der Praxis kann eine solche Aktualisierungszeit von den Steuerungskomponenten nicht aufgefangen werden und führt zu einem Profinetfehler und damit zu einem Anlagenstillstand.

Bessere Echtzeitfähigkeiten haben hier 5G Campusnetze und Bluetooth. Beides wurde bestätigt im Forschungsprojekt 5GANG.

Quelle https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/5gang

Im Abschlussbericht KoMe DFAM wurde die Profnet Kommunikation von 10 Teilnehmern über WLAN untersucht. Bei diesem Beispiel ist die Aktualisierungszeit auf 5ms eingestellt. Jeder der 10 Teilnehmer wird mit einer Sicherheitsreserve von 4.6ms ausgestattet um Profnet Fehlerauslösungen zu vermeiden. Bei 10 Teilnehmern führt das zu einer Roundtripzeit von 45ms im best case Fall. Zu bedenken ist daß es bei jeder einzelnen Übertagung zu einer Latenz von 1000ms kommen kann.

Tabelle 7-13: PROFINET IO Kommunikation mit WLAN

Zeit [ms]	Kanal	Frequenz [MHz]	Übertragungsrichtung	Paketlänge [Byte]	Paketlänge [µs]	Pakettyp
0,000	7	2442	Access Point → Client 1	185	142	Nutzdaten
0,152	7	2442	Client 1 → Access Point	29	50	ACK
0,230	7	2442	Client 1 → Access Point	185	142	Nutzdaten
0,382	7	2442	Access Point → Client 1	29	50	ACK
5,000	7	2442	Access Point → Client 2	185	142	Nutzdaten
5,152	7	2442	Client 2 → Access Point	29	50	ACK
5,230	7	2442	Client 2 → Access Point	185	142	Nutzdaten
5,382	7	2442	Access Point → Client 2	29	50	ACK
10,000	7	2442	Access Point → Client 3	185	142	Nutzdaten
10,152	7	2442	Client 3 → Access Point	29	50	ACK
10,230	7	2442	Client 3 → Access Point	185	142	Nutzdaten
10,382	7	2442	Access Point → Client 3	29	50	ACK
15,000	7	2442	Access Point → Client 4	185	142	Nutzdaten
15,152	7	2442	Client 4 → Access Point	29	50	ACK
15,230	7	2442	Client 4 → Access Point	185	142	Nutzdaten
15,382	7	2442	Access Point → Client 4	29	50	ACK

Zeit [ms]	Kanal	Frequenz [MHz]	Übertragungsrichtung	Paketlänge [Byte]	Paketlänge [µs]	Pakettyp
…	…	…	…	…	…	…
45,000	7	2442	Access Point → Client 10	185	142	Nutzdaten
45,152	7	2442	Client 10 → Access Point	29	50	ACK
45,230	7	2442	Client 10 → Access Point	185	142	Nutzdaten
45,382	7	2442	Access Point → Client 10	29	50	ACK
…	…	…	…	…	…	…
128,000	7	2442	Access Point → Client 1	185	142	Nutzdaten
128,152	7	2442	Client 1 → Access Point	29	50	ACK
128,230	7	2442	Client 1 → Access Point	185	142	Nutzdaten
128,382	7	2442	Access Point → Client 1	29	50	ACK
…	…	…	…	…	…	…
173,000	7	2442	Access Point → Client 10	185	142	Nutzdaten
173,152	7	2442	Client 10 → Access Point	29	50	ACK
173,230	7	2442	Client 10 → Access Point	185	142	Nutzdaten
173,382	7	2442	Access Point → Client 10	29	50	ACK

Bild3 Profinet Kommunikation über WLAN

Quelle: Abschlussbericht KoMe DFAM Kognitive Mediumszugangsalgorithmen für industrielle Funkanwendungen

Ein anderer Lösungsansatz

Ein anderer Lösungsansatz hat sich in der Praxis seit nunmehr 25 Jahren in mehreren Tausend Anwendungen bewährt. Dazu werden die Funkmodule um 2 in Soft -und Hardware umgesetzte Bausteine erweitert: Eine Datenbank für Profinet Telegramme und eine Vorverarbeitungseinheit die entscheidet welche Telegramme gesendet werden müssen.

Ein erster Vorteil dieses Konzeptes ist, daß die Profinet Telegramme mit der für Kabelübertragungen typischen Aktualisierungszeiteinstellung von 1ms gesendet werden können, es also keiner Änderung des SPS Projektes bedarf. Voraussetzung ist daß die Datenbank schnell genug ist um die Daten mit dieser Aktualisierungszeit aufzunehmen was einen entsprechend leistungsfähigen Microkontroller bedingt sowie ein echtzeitfähiges Betriebssystem.

Die zweite und wichtigste Aufgabe muss die Vorverarbeitseinheit erfüllen. Sie muss entscheiden ob ein Profinet Telegramm neue Daten enthält und gegebenfalls dieses Telegramm an die Funkstrecke übergeben werden soll. Profinet sendet typisch ein Abbild des Ausgangssendespeichers im eingestellten Aktualisierungsintervall. Dadurch kommt es zu einem sehr hohen Redundanzanteil der Daten. Dies spielt bei einer Kabelverbindung keine Rolle, hat jedoch bei einer 10000 fach weniger performanten Funkübertragung einen entscheidenden negativen Effekt. Nach unserer Erfahrung sind in typischen wireless automation Anwendungen bis zu 80% redundante Daten.

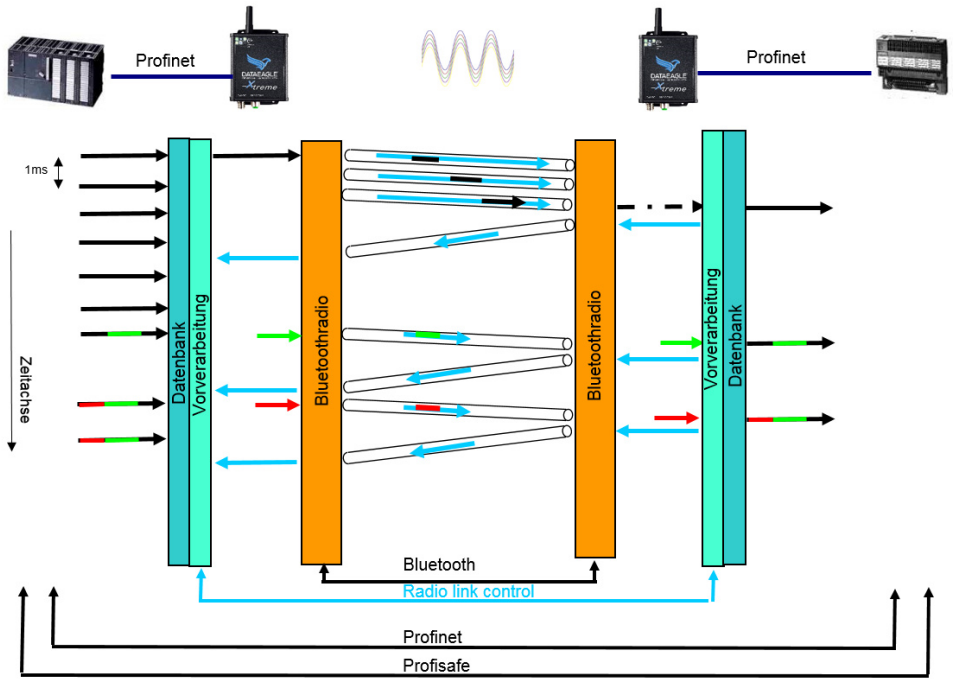


Bild4 Aufbau der Funklösung mit Datenbank und Vorverarbeitungsschicht

Beschreibung der Funktion der Vorverarbeitung auf der SPS Seite

Das Bild4 zeigt auf der linken Seite die PN Telegramme die hier als farbige Pfeile dargestellt sind. Gleiche Farbe soll denselben Dateninhalt darstellen. Die Vorverarbeitung übergibt der Funkstrecke das erste PN Telegramm (schwarz)vollständig, verhindert jedoch die Weiterleitung der folgenden gleichen Telegramme solange bis sie eine Datenänderung feststellt. (schwarz-grün).

Es wurde neben der reinen Sendeunterdrückung weitere Algorithmen zur Datenreduktion umgesetzt Eine Besonderheit kann sein, daß nur wenige Bytes eines Telegramms, bei Profinet bis zu 1000 Byte, sich ändern. In diesem Fall werden z.B. auch nur diese Änderungen übertragen.

Beschreibung der Funktion der Vorverarbeitung auf der Remote IO-Seite

Die Ausgabe Seite nach der Funkstrecke hat andere Aufgaben der Telegrammbearbeitung als die Senderseite. Eine der Aufgaben kann es sein die in unterschiedlichen Übertragungszyklen der Funkübertragung eingegangenen Telegrammfragmente zu einem Profinet Telegramm wieder zusammenzusetzen und dann auf der Remote-IO Seite über die Kabelschnittstelle auszugeben. Ändern sich nur wenige Bytes der Daten werden diese mit den schon übertragenen PN Daten aktualisiert.

Interne Kommunikation zwischen Sender und Empfänger

Es hat sich als sehr vorteilhaft gezeigt wenn eine zusätzliche interne Kommunikationstruktur zwischen den Geräten und den Software Modulen gibt. Diese interne Kommunikation wird als *Radio-link-control* bezeichnet. Hier werden Status Informationen über den Zustand der Funkstrecke und der Kommunikation ausgetauscht und verarbeitet. Die hellblauen Pfeile sollen dies bildlich darstellen. So ist es möglich, dass die Module jederzeit über den aktuellen Status der Telegrammverarbeitung informiert sind. Telegrammzähler und Zeitüberwachungen können so feststellen, ob die letzte Übertragung erfolgreich war. Diese Information wird verwendet, um Funkwiederholungen und Telegramm Wiederholungen gegebenenfalls auszulösen.

Filterzeit

Die Algorithmen der Vorverarbeitung arbeiten innerhalb einer vom Anwender definierbaren Zeit der *Filterzeit*. Diese Zeitüberwachung ist typisch für 50ms bis 5000ms aktiv. Läuft die Filterzeit ab ohne dass eine Übertragung erfolgreich war, wird ein Busfehler ausgelöst, äquivalent zu einer Kabelunterbrechung. Diese Funktion ist wichtig um bei einer längeren Funkstörung die Feldbuskommunikation gezielt zu beenden. In der SPS wird dazu der Softwarebaustein OB 86 angesprochen mit dem die Anwendung dann auf einen solchen Fehlerzustand reagieren kann.

Einfluss der Funktechnologie auf den Feldbus

Sehr grossen Einfluss auf die Anwendung und die Anzahl der Profinet Busfehler und somit Anlagenstillstände hat die eingesetzte Funktechnologie. Im Forschungsverbundprojekt 5GANG

wurden das Verhalten von 3 Funktechnologien WLAN, 5G Campusnetz Release 15, und Bluetooth untersucht unter der exakt selben Applikation (wireless Profinet/Profisafe mit einem Remote-IO) in Bezug auf Latenz vermessen. Es hat sich dabei gezeigt, daß WLAN deutlich längere Funkunterbrechungen erzeugt als Bluetooth. Über 5G Campusnetze kann abschliessend noch keine Aussage getroffen werden da bisher keine Langzeit Erfahrungen in praktischen Anwendungen vorliegen.

Ergebnisse 5GANG Forschungsprojekt und Anwendbarkeit der Redundanzstrategie

Im 5GANG Projekt konnte erfolgreich messtechnisch nachgewiesen werden daß die vorgestellte Datenredundanzstrategie deutlich stabilere Feldbuskommunikationen ermöglicht im Vergleich ohne diese.

Eine weitere Erkenntnis ist das jede der getesteten Funktechnologien WLAN, 5G Campusnetz und Bluetooth davon profitiert.

http://dx.doi.org/10.25673/121097

Ontology-Driven Extraction of Field Device Semantics Using Large Language Models

Victor Chavez [1], Jörg Wollert[2]

Abstract: The commissioning and deployment of field devices requires expertise in standards and communication protocols to translate domain-specific knowledge into higher-level applications. This integration typically involves loading a device description file into a vendor-specific programming tool to setup a device´s parameters and assign its application data within an integrated development environment. Due to the semi-structured format of device description files, their interpretation still requires cross-referencing unstructured data sources such as datasheets and operation manuals. A promising opportunity to reduce engineering effort is the use of Large Language Models (LLMs) to extract a formal semantic model based on ontologies. In this paper, we present an ontology-driven approach to automate the interpretation of field device documents using LLMs. Building on our previous work, the Industry 4.0 Field Device (I40FD) ontology, we instruct an LLM to generate Turtle-based RDF assertions for a field device guided by SHACL constraints. Our contributions include: a pipeline for processing field device documentation, formal semantic validation using Semantic Web technologies, and an evaluation with commercial devices based on Modbus, IO-Link, and CANopen.

1 Introduction

The standardization of field device technologies led to industrial organizations to develop device description languages that address the heterogeneity of accessing device information through the application layer of a specific communication protocol. Integrated Development Environments (IDEs) enable automation systems to interpret the semi-structured format of these files (e.g., XML) and provides engineers with a user interface to commission and deploy field devices. Although Device Description Languages (DDLs) enable seamless integration with software tools, the semantic interpretation of DDLs remains primarily a manual process that requires expertise in communication protocols, domain-specific concepts from the field device standard, and cross-referencing of unstructured data sources (e.g., datasheets, operation manuals).

Device profiles partially mitigate this problem by defining a common set of functions (e.g., motor controller, inclinometer, encoder) allowing IDE vendors to provide tools or libraries with a common interface to access a standardized conceptualization. Nevertheless, the semantics of device profiles are embedded in the standard and only partially represented in semi-structured DDL files. As a result, existing IDEs and software solutions rely on engineers with domain expertise to interpret device profiles and DDLs by cross-referencing documentation.

Figure 1 illustrates a simplified workflow for commissioning and deploying field devices. First, a device description file is retrieved for a field device; afterwards, it is loaded into a vendor-specific IDE. The IDE dynamically generates a configuration interface based on

[1] Institut für angewandte Automation und Mechatronik, Fachhochschule Aachen, Mechatronik und Eingebettete Systeme, Goethestr. 1, 52062 Aachen, chavez-bermudez@fh-aachen.de, https://orcid.org/0000-0001-6419-1641

[2] Institut für angewandte Automation und Mechatronik, Fachhochschule Aachen, Mechatronik und Eingebettete Systeme, Goethestr. 1, 52062 Aachen, wollert@fh-aachen.de https://orcid.org/0000-0001-7576-1339

the device description and provides menus to use process data and configuration parameters. The IDE relates the Application Data Objects (AOs) of the device to the internal symbolic representation of the communication interface of an industrial PC or Programmable Logic Controller (PLC). The semantic interpretation of the AOs is not part of this process and mostly focuses on mapping primitive data types to the programming language, displaying human-readable text associated with the AOs, and supporting addressing mechanisms (e.g., index, subindex). Subsequently, the commissioning engineer interprets the purpose of the AOs by cross-referencing vendor documentation that explains the use of the AOs and any configuration required by the implemented communication protocol. Finally, the engineer manipulates the AOs through the IDE and develops the application. In the absence of a formal semantic model (i.e., ontologies), field device integration remains a labor-intensive manual process that relies on experts to interpret semi-structured and unstructured sources.

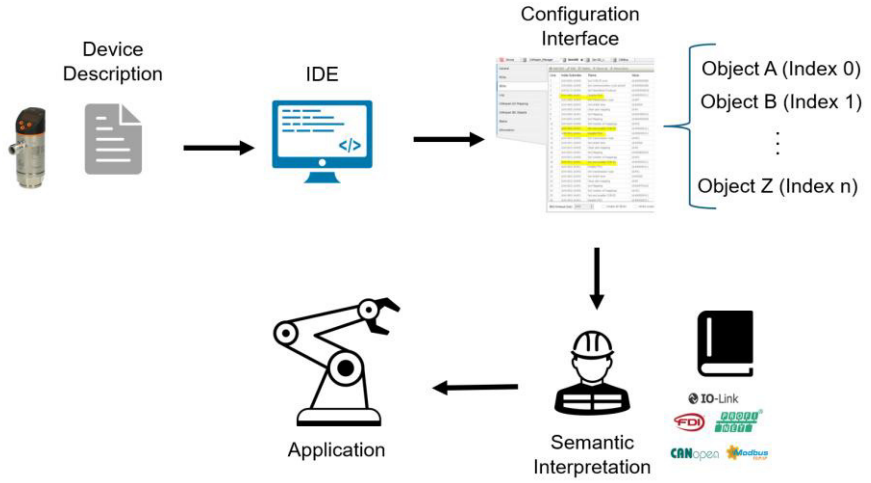


Figure 1 Commissioning and deployment workflow for field devices.

Large Language Models (LLMs) offer a promising way to bridge the gap between manual interpretation and the generation of structured information. However, deploying LLMs introduces several challenges [1], [2], such as hallucinations and non-deterministic results which are unacceptable for industrial applications. To solve this issue an active field of research is the integration of Semantic Web Technologies (SWT) to validate the output of LLMs [3]–[5]. In this paper we present an LLM ontology-driven extraction of field device semantics to address the semantic gap that unstructured data sources and their interpretation represent. Based on the Industry 4.0 Field Device (I40FD) ontology [6], [7], we guide an LLM to align datasheets and operation manual with a generic vocabulary that defines device description files, application data, and their intrinsic relationship. Leveraging SWT, we infer high-level capabilities and validate the LLM assertions. We evaluate the approach with three protocols: Modbus, CANopen, and IO-Link. The remainder of the paper is organized as follows. Section 2 surveys related work; Section 3 summarizes the I40FD semantics; Section 4 details our methodology; Section 5 presents the evaluation and results; Section 6 discusses implications; Section 7 concludes and outlines future work.

2 Related Work

The integration of SWT and ontologies with LLMs is an active area of research. Lippolis et al. [8] investigated the capability of LLMs to generate OWL ontology drafts directly from requirements expressed as user stories and competency questions. The results indicate that LLMs can produce models that satisfy engineering requirements and support novice ontology engineers. The authors identified recurring issues such as multiple domains and ranges, and missing namespace declarations. Their results demonstrate the capability of LLMs to develop ontology drafts that can shorten the initial modeling phase.

Sadruddin et al. [9] recognized the difficulty of representing and reasoning over scientific processes with natural language. Their work presents Schema-miner-pro, a human-in-the-loop framework that extracts structured information from scientific manuscripts. Two manufacturing procedures are presented: atomic layer deposition and atomic layer etching where process parameters and outputs are aligned to the Quantity Units and Dimensions and Types (QUDT) ontology [10]. The LLM is tasked with generating a JSON schema and then iteratively refining it under expert guidance. The agent system grounds the schema to QUDT properties via lexical heuristics, and when no direct match is found, it applies vector similarity search to identify ontology embeddings corresponding to schema terms.

In an Industry 4.0 (I4.0) centric approach, Xi et al. [11] design an LLM agent system to process technical datasheets and produce Asset Administration Shell (AAS) submodels. The method extracts technical properties into a “semantic node”, a predefined schema with fields including name, conceptual definition, affordance, value, value type, unit, and source. The LLM follows the semantic node schema to populate the AAS submodels. Through the AAS property *semanticID* a Retrieval Augmented Generation (RAG) search embeds ECLASS dictionary entries and retrieves the closest match for grounding. Evaluation was conducted on 2,259 technical property samples from 20 automation components, where human annotators assessed semantic clarity and correctness. The approach achieved an effective generation rate of 62–79%, demonstrating that a significant portion of manual modeling effort can be redirected from creation to validation.

3 Field Device Semantics

The objective our of work is to automate the extraction of structured information from vendor documentation to produce field device semantic models. We leverage the I40FD ontology to define a field device in standard-specific terms aligned with a generic vocabulary that captures the intrinsic relationship between application data, metadata about values (e.g., limits, scaling), and QUDT units. For example, consider an IO-Link sensor that measures degrees Celsius. According to the IO-Link specification, process data is transmitted over the Process Data Input channel, and the unit code representing degrees Celsius is encoded as the integer value 1001. With the I40FD ontology, we can assert these facts in Turtle RDF, as presented in Listing 1.

```
@prefix i40fd:   <http://w3id.org/iaam/i40fd#> .
@prefix :        <http://w3id.org/iaam/example#> .

:SensorA  i40fd:hasDeviceDescription          :IODD_1
:IODD_1   i40fd:hasApplicationDataObject      :PD_1
:PD_1     i40fd:hasApplicationDataElement     :AE_1
:AE_1     i40fd:hasUnitCode     1001.
```

Listing 1 Demonstrative RDF Turtle representation of an IO-Link sensor with the I40FD ontology.

This RDF representation is similar to what is contained in a device description file, with the difference that it is semantically enriched with OWL and SWRL rules that enable automatic entailments and generalization of the device's capabilities. When running these assertions in the I40FD framework [6], we entail that the individual *PD_1* carries cyclic process data transmitted from the sensor to the IO-Link controller. Moreover, since the unit code is aligned with the QUDT ontology — using an IO-Link units vocabulary[3] — it is inferred that the device has a temperature measurement capability, realized through an Applcation Data Element (AE) *AE_1*, and an ECLASS-based classification (see Listing 2).

```
@prefix i40fd:   <http://w3id.org/iaam/i40fd#> .
@prefix iof:     <https://spec.industrialontologies.org/ontology/core/Core/>
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix bfo:     <http://purl.obolibrary.org/obo/> .
@prefix device:  <http://w3id.org/iaam/i40fd/vocab/device#>.

:SensorA iof:hasCapability        :cap_1;
         i40fd:hasClassification  device:SensorTemperature.
:cap_1   rdf:type                 i40fd:SensingCapability;
         bfo:0000054              :proc1.  # realizedBy
:proc1   rdf:type                 i40fd:CommunicationProcess;
         iof:prescribedBy         :AE_1 .
```

Listing 2 Inference example for an IO-Link sensor with the I40FD ontology.

The main advantage of the I40FD ontology over semi-structured DDLs is that semantic interpretation is embedded in OWL axioms and SWRL rules. Additional context can be defined by aligning capabilities and application data with other ontologies, enabling reuse and extension, which is not possible with semi-structured files.

4 Methodology

We developed a pipeline to extract field device semantics aligned with the I40FD ontology and supported by the natural language interpretation of documentation with LLMs. The LLM takes as input a system message that defines the role of an ontology expert whose task is to generate RDF Turtle assertions based on user documentation. The system message embeds a Shape Constraint Language (SHACL) shape that defines AOs, AEs, and their relationship to QUDT units, and data types. Additionally, we dynamically load protocol-specific shapes that constraint how specific field-device terms and their communication protocols represent terms with respect to the I40FD ontology. We instruct the LLM with recognition rules to align units from the QUDT vocabulary[4], prefixes, and common quantity types. The input message for the LLM consists of documentation from

[3] http://w3id.org/iaam/i40fd/units/iolink
[4] https://qudt.org/3.1.5/vocab/unit

a specific field device product. We limit the token input of the message by extracting only relevant pages from technical characteristics of the device and the operation manual related to the communication protocol and their interpretation. This approach allows us to focus on the interpretation of concepts related to the I40FD ontology and increase the accuracy of results by filtering non-relevant information (e.g., mechanical drawings, device certificates). Figure 2 presents an overview of the LLM system and input message part of the prompt.

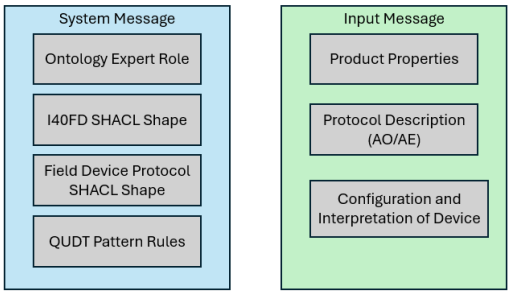


Figure 2 Overview of system and input message for the LLM.

Product documentation is mostly provided in PDF format. We developed a simple extraction script in python that leverages the python library *PyMuPDF4LLM* [12] to convert PDFs to markdown string representation. As presented in Figure 3, the LLM pipeline consists of the following four steps. First, the product documentation is converted to markdown to generate the input message. Secondly, a validation step checks the Turtle syntax and SHACL shapes against the LLM output. Lastly, if the validation is successfully, we run the Turtle RDF against the I40FD inference engine to discovery the capabilities of the device and their realization through the annotated AEs (e.g., units, limits, scaling).

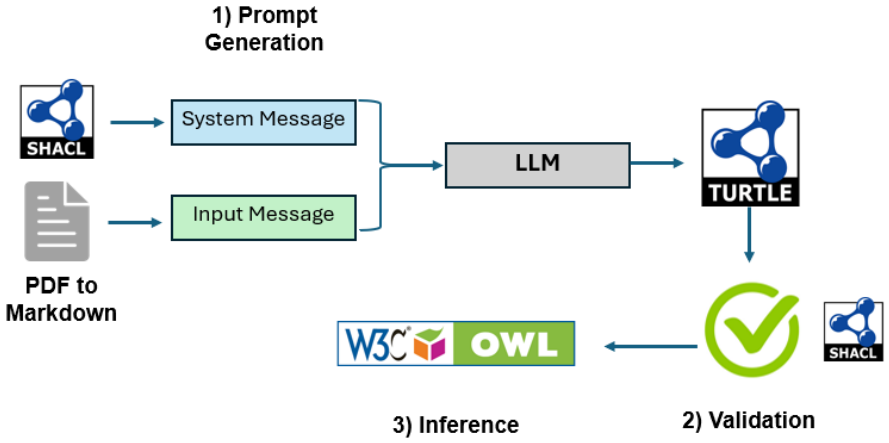


Figure 3 LLM-based pipeline for RDF Turtle generation, validation and OWL inference.

5 Evaluation

We tested the implementation of our pipeline with 15 commercial products of three field device standards: Modbus, IO-Link and CANOpen, each consisting of 5 products. An online Git repository is provided with the implementation of the pipeline[5]. The goal of our evaluation is to verify the feasibility of LLM generated RDF Turtle assertions guided by SHACL shapes, the interpretation of field device documentation with respect to the I40FD ontology and the validation of the inferred capabilities. The following LLM providers were used: OpenAI, Google and Mistral AI.

5.1 RDF Turtle and SHACL Validation

In Table 1, the results obtained towards validation of RDF Turtle syntax and SHACL are presented. In general, the smallest models (*gpt-5-nano*, *gemini-2.5-flash-lite*, *mistral-small-2506*) tended to have more errors, with exception of the *mistral-large-2411* which generated more errors than its smaller counterparts. The main source of errors of RDF Turtle generation were the assertion of integer values as string literals (e.g., *0x20^xsd:integer*) and trailing whitespace before the declaration of an individual. With respect to SHACL constraints, some LLMs struggled to instantiate CANOpen Service Data Objects (SDOs) and their respective subentries, SDO indexes, and subindexes. Metadata from IO-Link Process Data Input was missing such as bit offset or bit length.

Table 1 Summary of RDF Turtle and SHACL validation errors.

	OpenAI		
	gpt-5	gpt-5-mini	gpt-5-nano
Turtle	0 %	0 %	20 %
SHACL	0 %	0 %	26 %
	Google		
	gemini-2.5-pro	gemini-2.5-flash	gemini-2.5-flash-lite
Turtle	0 %	6 %	40 %
SHACL	0 %	6 %	53 %
	Mistral		
	large-2411	medium-2508	small-2506
Turtle	26 %	6 %	13 %
SHACL	33 %	13 %	20 %

5.2 Capability inference results

The RDF Turtle assertions generated by the LLM were run against the I40FD inference engine[6] to verify its soundness based on the metadata of the AEs (units, scaling, limits) and the correct interpretation of AOs (e.g., SDOs, PDIn, Modbus registers). We prepared an initial assessment by checking how many RDF assertions generated by LLMs enabled the inference engine to entail capability assertions for each field device. In Table 2 we summarize our findings per LLM provider, and field device protocol.

[5] https://git.fh-aachen.de/iaam/semantics/i40fd-llm-gen

[6] https://git.fh-aachen.de/iaam/semantics/i40fd-inference-engine

Table 2 LLM output generation that produced inference results for 15 distinct field devices per provider.

	OpenAI		
	gpt-5	gpt-5-mini	gpt-5-nano
Modbus	100 %	100 %	40 %
IO-Link	60 %	60%	20 %
CANOpen	20 %	80 %	0 %
Total	**60 %**	**80 %**	**20 %**
	Google		
	gemini-2.5-pro	gemini-2.5-flash	gemini-2.5-flash-lite
Modbus	100 %	100 %	80%
IO-Link	100 %	100 %	80%
CANOpen	40 %	40 %	20 %
Total	**80 %**	**80 %**	**60 %**
	Mistral		
	large-2411	medium-2508	small-2506
Modbus	100 %	60%	100 %
IO-Link	100 %	40 %	0 %
CANOpen	40 %	60%	20 %
Total	**80 %**	**53 %**	**40 %**

We observed that most LLMs had difficulties producing inference results for the CANOpen protocol, producing inference in average 35 percent of the time. In contrast Modbus presented inference results 86 percent of the time, followed by IO-Link with 62 percent. In general, larger models were able to produce inference results between 60 and 80 percent of the time, whereas smaller models generated inference results between 20 and 60 percent with *gemini-2.5-flash-lite* producing the most and *mistral-small-2506* the least.

5.3 Capability Soundness

To verify the soundness of the RDF assertions, we reviewed the PDF documentation and datasheets supplied to the models to build the ground truth. First, we identified AOs that represent either a measurement or control signal. Then, we selected four relevant properties of the I40FD ontology that the LLM should generate: QUDT units, range, scaling information, and addressing mechanism (i.e., subindex, index for IO-Link and CANOpen; Modbus register address). Based on this information we generated a list of 51 expectations across all devices to identify the total amount of matched capabilities, their coverage, and the amount of devices that were modelled correctly. In Table 3 we summarize our findings.

We observed that large models such as *gpt-5*, *gemini-2.5-pro* and *mistral-large-2411* matched at least 68 percent of the 51 expected capabilities. The medium and small models however had difficulties in matching capabilities with less than half being correct. The lower and medium LLMs from each provider had better results with the Modbus and IO-Link protocol. In general, less than half of the products per protocol were matched, with the exception of *gemini-2.5-pro*.

Table 3 LLM ontology soundness comparison for the inference field device capabilities.

Provider	LLM	Capability coverage	Matched Devices (%)			
			IO-Link 5	CANOpen 5	Modbus 5	Total % 15
OpenAI	gpt-5	24/35 (68%)	40	20	40	33
	gpt-5-mini	28/40 (70%)	40	40	20	33
	gpt-5-nano	11/17 (64%)	0	0	20	6
Google	gemini-2.5-pro	22/37 (59%)	60	20	20	33
	gemini-2.5-flash	17/37 (45%)	40	0	20	20
	gemini-2.5-flash-lite	4/16 (25%)	0	0	20	6
Mistral	large-2411	23/39 (59%)	40	20	40	33
	medium-2508	8/16 (50%)	20	20	20	20
	small-2506	17/27 (63.0%)	0	0	20	6

6 Discussion

The results of our evaluation demonstrate the possibility to use LLMs to generate RDF Turtle assertions for the field device domain. However, the inference results and soundness contrast highlight variability of reliability among LLMs. We noticed that the quality of results improved when manually curating the PDF files by parsing them through a multimodal LLM to extract information more reliably. For example, one device documenta stated a range limit as 2^{60} while our PDF parser extracted 260. We analyzed the parsed documentation of those devices with empty inference and notice a consistency in information loss which caused the LLMs to perform badly. Additionally, for families of devices stated in a datasheet, some LLMs chose the wrong metadata properties due to the lack of guidance on how to process multiple values for a single device.

The matching of QUDT units was mostly correct, with the exception of units not described in the prompt**.** For an accurate matching of units, a secondary step such as the one proposed by Sadruddin et al. [9], could increase the soundness of units. Range properties and limits varied in some cases where the LLM set a unit as per the datasheet but different than the AO. For instance, if a voltage is described in millivolts in a Modbus register but the datasheet specifies volts, the range was mistakenly asserted using volt units.

In summary, we find that to increase the accuracy of the pipeline additional parsing for edge cases are required (e.g., multiple devices), report of contradictions in the documentation, and an LLM capable PDF parser that can extract accurately the documentation. We position this work as complementary to DDLs by enriching their semantic interpretation and automating ontology population, thereby enabling an interoperable inference of high-level capabilities.

7 Conclusions

This paper presented a novel LLM-based semantic extraction pipeline for field devices with the I40FD ontology. In comparison to semi-structured DDLs, ontologies provide the intrinsic relationships between communication-specific AOs and AEs. The semantic formality of ontologies enables automatic inference and entailment with description logics. In contrast, traditional approaches with DDLs only interpret the low-level concepts of an industrial protocol and a human-readable user interface to manipulate the data. The most time-consuming part remains semantic interpretation.

Our results have shown that RDF generation for field devices with the I40FD ontology is feasible. The extraction was majorly successful on the Modbus and IO-Link protocol. The larger LLM variants proved to be the most suitable for extraction. Nevertheless, we observed a lower ontological soundness. The main causes we identified are ambiguity in documentation, PDF extraction quality and contradictions in manuals. As future research work, we will investigate assisted-tooling methods that enable domain experts to review the results and trace cross-references from documentation. Moreover, we plan to increase the catalog of devices to identify the limitations of LLMs and increase the alignment to the I40FD ontology with methods such as vector similarity search, and GraphRAG.

References

[1] R. Asgari, M. Moradi, K. Yan, D. Colwell, and M. Samwald, "A Critical Review of Methods and Challenges in Large Language Models," *Computers, Materials & Continua*, vol. 82, no. 2, pp. 1681–1698, 2025, doi: 10.32604/cmc.2025.061263.

[2] X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, "A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges," *Vicinagearth*, vol. 1, no. 1, p. 9, Oct. 2024, doi: 10.1007/s44336-024-00009-2.

[3] A. Soularidis, K. Kotis, M. Lamolle, Z. Mejdoul, G. Lortal, and G. Vouros, "LLM-Assisted Generation of SWRL Rules from Natural Language," in *2024 International Conference on AI x Data and Knowledge Engineering (AIxDKE)*, IEEE, Dec. 2024, pp. 7–12. doi: 10.1109/AIxDKE63520.2024.00008.

[4] J. Moncada-Ramirez, J.-L. Matez-Bandera, J. Gonzalez-Jimenez, and J.-R. Ruiz-Sarmiento, "Agentic Workflows for Improving Large Language Model Reasoning in Robotic Object-Centered Planning," *Robotics*, vol. 14, no. 3, p. 24, Feb. 2025, doi: 10.3390/robotics14030024.

[5] F. M. Soares *et al.*, "Exploring a Large Language Model for Transforming Taxonomic Data into OWL: Lessons Learned and Implications for Ontology Development," *Data Intell*, vol. 7, no. 2, pp. 265–302, Jun. 2025, doi: 10.3724/2096-7004.di.2025.0020.

[6] V. Chavez and J. Wollert, "An Industry 4.0 Semantic Framework for the Inference of Generic Field Device Capabilities," *IEEE Open Journal of the Industrial Electronics Society*, vol. 6, pp. 868–882, May 2025, doi: 10.1109/OJIES.2025.3571319.

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

[8] A. S. Lippolis *et al.*, “Ontology Generation Using Large Language Models,” in *Lecture Notes in Computer Science*, vol. 15718, Cham: Springer, 2025, pp. 321–341. doi: 10.1007/978-3-031-94575-5_18.

[9] S. Sadruddin *et al.*, “Schema-Miner Pro: Agentic AI for Ontology Grounding over LLM-Discovered Scientific Schemas in a Human-in-the-Loop Workflow,” *Semantic Web Journal*, 2025.

[10] QUDT.org, “Quantities, Units, Dimensions and Types.” Accessed: Sep. 01, 2025. [Online]. Available: https://qudt.org/

[11] Y. Xia, Z. Xiao, N. Jazdi, and M. Weyrich, “Generation of Asset Administration Shell With Large Language Model Agents: Toward Semantic Interoperability in Digital Twins in the Context of Industry 4.0,” *IEEE Access*, vol. 12, pp. 84863–84877, 2024, doi: 10.1109/ACCESS.2024.3415470.

[12] “PyMuPDF Documentation.” Accessed: Sep. 28, 2025. [Online]. Available: https://pymupdf.readthedocs.io/en/latest/pymupdf4llm/

http://dx.doi.org/10.25673/121097

A Communication Control Co-Design Approach in Industrial Automation

Owis Alsabbagh [1,2], Rafiul Islam[1], and Lisa Underberg[1,2]

Abstract: This paper aims to explore the preconditions of Communication-Control Co-Design (CoCoCo) for wireless networked control systems. Rather than focusing on averting the worst-case scenarios and employing decoupled designs that often result in over-provisioning resources and performance constraints, this study analyzes the performance trade-offs in a minimal industrial wireless application scenario. A simulation-based study employing OMNeT++, INET, and Simu5G assesses performance under varying channel and load conditions, focusing on packet transmission time, update time, and packet loss ratio. This approach reveals cross-layer sensitivities that influence control outcomes. The proposed approach aims to address three practical challenges: the scarcity of representative case studies, the modeling complexity of intertwined subsystems, and operational risk in time-critical settings. Its goal is to identify operating conditions in which adaptive co-design policies can provide measurable improvements. The contributions encompass a representative minimal use case, which has been investigated under various parameter profiles and evaluated according to the performance metrics: transmission time (TT), update Time (UT) and packet loss ratio (PLR). The findings reveal the impact of specific requirements on performance metrics, with some requirements exerting a more substantial influence than others. For instance, the transfer interval (TI) demonstrates a more pronounced effect than factors such as power or bandwidth, which will be examined in greater detail subsequently.

Keywords: CoCoCo, Industrial Automation, 5G, Industry 4.0

1 Introduction

The rapid evolution of industrial automation toward Industry 4.0 has ushered in an era of intelligent, networked systems that demand real-time data exchange and decision-making. Wireless Networked Control Systems (WNCS) have emerged as a key enabler of flexible, scalable deployments. However, unpredictable delays and packet losses pose serious challenges to maintaining control stability and performance [Ta19]. The CoCoCo approach is a holistic solution that addresses these issues by jointly optimizing control algorithms and communication protocols, rather than treating them as separate entities. By accounting for network-induced delays, packet losses, and resource constraints within the control law, and simultaneously adapting communication schedules based on control requirements, CoCoCo can optimize timing behavior, enhance reliability, and improve overall efficiency. The foundation of the CoCoCo concept lies in utilizing advanced control theories, such as linear,

[1] Otto-von-Guericke-University, Faculty of Electrical Engineering and Information Technology, Univeristätsplatz 2, 39106 Magdeburg, Germany, oalsabba@st.ovgu.de, https://orcid.org/0009-0000-0555-7500; rafiul.islam@ovgu.de; lisa.underberg@ifak.eu

[2] ifak , Industrial Communication, Werner-Heisenberg-Str. 1, 39106 Magdeburg, Germany, oalsabba@st.ovgu.de, https://orcid.org/0009-0000-0555-7500; lisa.underberg@ifak.eu

quadratic, Gaussian, and predictive packetized control, in conjunction with communication techniques, including event-triggered updates, adaptive scheduling, and semantic encoding of control-relevant information.In time-sensitive applications, such as robotics and smart manufacturing, CoCoCo principles ensure stability through dynamic adaptation to channel conditions and control objectives. Despite the theoretical promise of these principles, their practical implementation necessitates the development of scalable models for nonlinear, time-varying plants, the creation of robust algorithms for large-scale deployments, and the validation of these methods in real-world scenarios, taking into account factors such as fading, interference, and security constraints. The advent of emerging technologies, such as 5G/6G network slicing, edge computing, and machine learning, presents novel opportunities to circumvent these limitations by facilitating context-aware resource allocation and predictive adaptation.

This paper offers a brief overview of CoCoCo principles, incorporating recent advancements in industrial automation and WNCS. The main contribution is the presentation of a representative minimal scenario systematically assessed under various parameter profiles and evaluated using key metrics, such as transmission time, update time, packet loss ratio, and the OMNeT++, INET, and Simu5G frameworks.The paper is structured as follows, section 2 reviews related work and foundational theories. Section 3 details the scenario, methodology, and simulation setup. Section 4 presents a discussion of the results focused on communication-control trade-offs. Finally, Section 5 concludes the paper and outlines remaining challenges for the robust and adaptive deployment of industrial CoCoCo.

2 Related Work

The convergence of communication and control in NCSs for industrial automation has gained significant attention due to the stringent requirements of Industry 4.0, edge computing, and the Industrial Internet of Things (IIoT). Conventional industrial systems were dependent on wired networks with predetermined schedules. However, contemporary applications necessitate low latency, high reliability, and flexible resource allocation. The CoCoCo paradigm addresses these demands by jointly optimizing control performance and communication resources to mitigate delays, packet loss, and bandwidth constraints, rather than treating control algorithms and communication protocols independently. Recent research has demonstrated that CoCoCo can significantly enhance NCS performance by dynamically adapting control parameters and communication schedules in response to network conditions and control objectives [PAJ11; QFY23; Ta20; Zh19].

At the core of CoCoCo lies the understanding that communication metrics—such as packet loss and transmission times—exert a direct influence on quality of control (QoC), and that control demands have a significant impact on network resource requirements. Independent designs frequently result in overprovisioning or unmet control specifications. Control engineers may prioritize high reliability, potentially leading to the excessive utilization of network resources. Conversely, network designers may neglect control stability requirements,

potentially compromising the system's integrity [QFY23; Zh19]. CoCoCo frameworks are designed to address the competing needs of system components by coordinating resource allocation, scheduling, and adaptive control strategies. These frameworks aim to maintain system stability under conditions of uncertain communication and to preserve quality of service (QoS) in time-sensitive applications such as robotics and industrial process control [Ay24; Li14; Ta19].

The theoretical foundation of CoCoCo is rooted in established control and communication theories. Linear quadratic Gaussian (LQG) control and linear time-variant (LTI) models have been extended to account for limited data rates and channel noise, enabling the design of optimal controllers under communication constraints [CL10; Ta19]. Nonlinear control methodologies, including PID variants customized for autonomous guided vehicles (AGVs), have been investigated to address time-varying plant dynamics in conjunction with adaptive communication protocols [Fr23]. Event-triggered control schemes have been implemented in wide-area cyber-physical systems to reduce network congestion by transmitting updates only upon the occurrence of specified events. These schemes have been shown to decrease resource usage without compromising stability [Bh21; Li14]. Predictive control methodologies, encompassing packetized predictive control (PPC), entail the integration of sequences of future control commands within each packet to counteract delays and losses, though they introduce analytical intricacy [Ta19; Zh19].

In the domain of communication, meticulous channel modeling—incorporating Rayleigh and Rician fading—facilitates the evaluation of minimum data rate requirements for maintaining system observability and stability. QoS frameworks are designed to integrate latency, packet loss, and reliability metrics to guide dynamic scheduling and resource allocation that align with control priorities [Wa23]. Novel medium access control (MAC) protocols, spanning CSMA/CA and TDMA variants, are optimized for reduced delay and enhanced reliability in control-centric scenarios [CL10; PAJ11]. Coding strategies for short block lengths and scheduling algorithms for multi-loop control systems further illustrate how communication techniques can be co-designed with control objectives [Ta20; Zh19]. In recent developments, semantic communication has come to the fore, placing emphasis on the transmission of information pertinent to control rather than raw data. This enhancement in efficiency is achieved through the implementation of context-aware, task-oriented encoding methods. [Gi24].

Nonetheless, several challenges impede the widespread adoption of CoCoCo. The modeling of abstractions that capture the interaction between control dynamics and communication networks remains challenging when attempting to scale beyond simplified linear systems, particularly in the context of nonlinear, time-varying, and heterogeneous industrial processes [Ta19; Zh19]. Theoretical limits on stability under short-packet transmissions and information aging phenomena in predictive control demand further investigation [De21; Ta20]. The co-design of complex systems poses significant challenges during implementation. The integration of control and communication elements within a system results in an increase of system variables, a complexity of diagnosis, and the necessity of co-validated

trials to ensure real-world performance. This assertion is supported by the literature, as evidenced by the seminal work cited in [Zh19]. In the context of large-scale deployments, concerns regarding scalability intensify, particularly in scenarios where the computational and communication loads must be balanced across multiple devices [Ei22]. In the context of constrained real-time systems, it is imperative to address the issues of security and privacy. The existing solutions designed for long-packet networks may not be suitable for control loops that require minimal overhead. This issue is discussed in detail ins [Zh19].

The advent of emerging technologies has emerged as a potential solution to these limitations. The integration of 5G/6G network slicing and edge computing enables fine-grained resource partitioning for control tasks; however, joint management of wireless and computational resources remains an open problem [Ay24]. The application of artificial intelligence and machine learning techniques, including deep reinforcement learning for channel-aware control policies and Koopman operator methods for linearizing complex dynamics, has seen a marked increase in recent studies. These techniques are employed to predict channel conditions and derive adaptive co-design strategies [Ei22; Gi24; Me22; QFY23]. Semantic communications and Koopman autoencoders have the potential to extract and transmit the most salient features of system states, thereby reducing communication overhead while preserving control fidelity [Ay24; Gi24]. The real-time adaptation of control sampling intervals, AGV velocities, and packet delivery ratios is an active area of research. These parameters are based on the current network state. The objective of this research is to maintain stability and performance under dynamic conditions [Ei19; Ta20]. It is imperative to validate CoCoCo models through real-world experiments to bridge the gap between simulation and practice. This ensures that theoretical gains translate into operational benefits [CL10; Fr23].

In summary, the CoCoCo paradigm signifies a substantial shift towards the integration of communication and control systems, which can enhance performance and optimize resource utilization in the domain of industrial automation. The future success of this paradigm is contingent upon the development of scalable modeling frameworks, a more profound theoretical comprehension of co-design boundaries, reliable implementation approaches, and practical validation to unlock the complete benefits of CoCoCo in next-generation NCSs.

3 Methodology

3.1 Scenario Description

A practical example of the use case can be observed in a production line comprising a variety of robots, each specialized in distinct tasks such as welding, screwing, and cutting. Each robot sends and receives updates at varying rates and levels of priority. For example, welding robots require immediate and low-latency responses to ensure reliable performance,

whereas screwing robots can tolerate slightly longer delays. The incorporation of diverse communication needs into a unified, fixed user equipment (UE) framework ensures the system's simplicity while maintaining sufficient intricacy for the analysis of performance compromises and the validation of communication-control strategies inspired by CoCoCo in real-world industrial settings.

However, instead of describing a single process that accurately represents a real-world scenario, this study employs a minimal scenario to demonstrate the use case. This approach draws inspiration from two sources: common practices in performance testing and the principles of CoCoCo. As demonstrated in figure 1, a typical rectangular factory floor is represented, with a 5G base station (gNB) situated centrally and linked to a central controller via a router. Throughout the duration of the simulation, all UE nodes maintain a fixed position, with the evaluation concentrated on a single UE (UE0), that is positioned in the top left corner (0,0) and representing the worst case distance wise. The simplicity of this configuration allows straightforward result interpretation, while the subsequent section defines requirements derived from pragmatic values published in [3r24].

As illustrated in figure 1, the factory floor layout exhibits a uniform distribution of 30 UEs across 5 rows and 6 columns. For diffrent numbers of UEs, the devices will be distributed in a similar manner to the uniform distribution, with the distribution starting from the top left corner. For instance, 20 UE devices are distributed across 4 rows and 5 columns, while 10 UEs are distributed across 2 rows and 5 columns. This configuration ensures the perpetual presence of UEs in each of the 4 corners of the array. The configuration of the floor layout imposes the limitation on the distance, ensuring that its effects are consistent across experimental iterations.

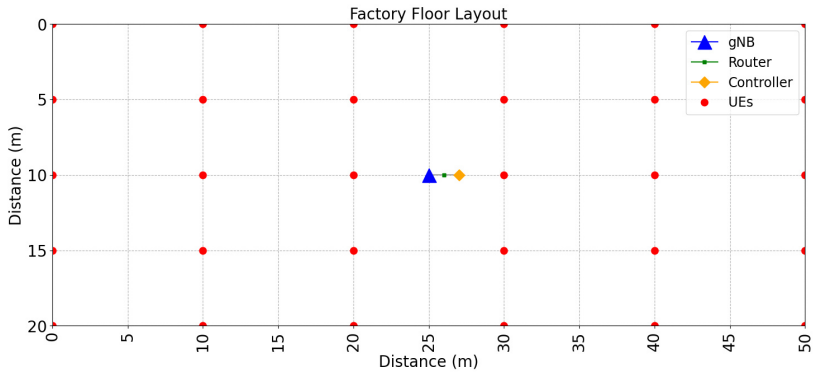


Fig. 1: Layout diagram of the factory floor with 30 UEs

In accordance with the minimal scenario, it is imperative to note that the requirements'

profiles are applied to all UEs. Throughout the duration of the simulation, each UE perpetually transmits and receives data while adhering to its assigned requirement profile. These profiles encapsulate key parameters such as packet size, transfer interval, and available bandwidth, ensuring that every UE operates under a consistent, yet representative workload. The application of uniform requirements across all UE nodes ensures the maintenance of the evaluation's focus on the application reliability represented by UE0, thereby eliminating superfluous variability.

3.2 Implementation

For the simulation, a specific group of profiles is established by adjusting certain parameters that originate either from the control system or the communication network. The first profile shows a network that is experiencing significant stress, indicated by its large amount of UEs and frequent updates. In order to minimize variance and isolate the impact of these stress factors, the message length, the UE-gNB distance are maintained at constant levels, 50 B and $\approx$ 53 m respectively. As for the downlink transfer interval, it is set to be equal to the uplink TI. By maintaining the aforementioned variables constant, any observed changes in performance metrics can be attributed primarily to the modified parameters.

Profile Nr.	Bandwidth (MHz)	TxPower (gNB/UE) (dBm)	Transfer Interval (ms)	Number of UE N
1	50	23/23	0.5	30
2	50	23/23	0.5	20
3	50	23/23	1	20
4	100	30/23	1	10
5	100	23/23	0.5	30
6	100	30/23	0.5	30
7	50	23/23	3	30
8	50	23/23	7	30
9	50	23/23	10	20

Tab. 1: Requirement profiles

The simulation is conducted for each profile presented in table 1. Each profile combines values for bandwidth, transmit power, TI, and number of UEs.These profiles are structured to evaluate different traffic densities, power configurations, and update frequencies. The downlink, originating from the controller and passing through the base station to the UEs, and the uplink, going in the opposite direction, are both analyzed. This analysis is facilitated through the utilization of OMNeT++ along with the modules of Simu5G and INET, whose interoperability and modular design enable the establishment of connections and necessary measurements. In addition, each profile is simulated for a total of 100,000 messages. The messages are each sent in a single packet, and they are evaluated based on three primary metrics: PLR, TT, and UT.

The packet loss ratio is defined as:

$$PLR = 1 - \frac{N_{\text{received}}}{N_{\text{sent}}} \tag{1}$$

with N being the number of packets sent or received.

As for the transmission time; it is defined as the time measured from the point when a message or a packet is handed from the application layer interface of the source application device, until the same message/packet is received at the application layer interface of the target application device[3r24].

Furthermore, the update time is the time between the reception of two consecutive messages/packets at the application layer's interface to the target application device [3r24].

4 Results and Discussion

4.1 Analysis of the transmission time

The histograms in figure 2 show distributions of the transmission time for UE0 for selected profiles and in the uplink direction.

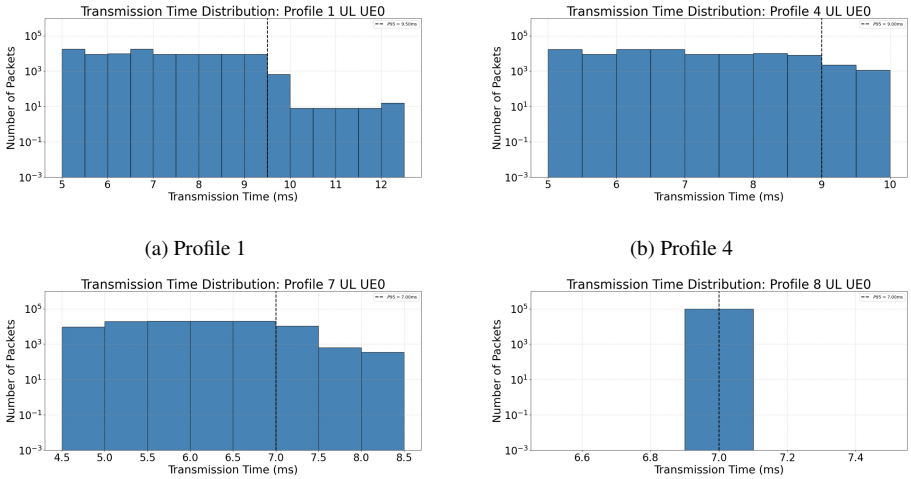
The distribution of TT values in profile 1 predominantly lies within the range of 6 to 9.5 ms, with a minority of data points extending between 10 and 12.5 ms (see Figure 2a). Similar to profile 1, the TT values in profile 4 are largely confined to approximately 6–9.5 ms (see Figure 2b). In contrast, profile 7 exhibits a more restricted range, with the majority of TT values falling between 5 and 7 ms (see Figure 2c).

Profile 8 is characterized by a single bin histogram, with the mean and P95 values being equal to each other (see figure 2d).

The data as illustrated in figures 2a,2b and 2c appears to be discretely categorized in 0.5 ms increments, reflecting the slot-based scheduling system in NR communications. Subcarrier spacing of 30 kHz yields slots of 0.5 ms. However, the small data sets outside of the main cluster[3] suggests irregular slow transmission instances. The distribution in its totality demonstrates a distinct propensity for widespread grouping over multiple value ranges, suggesting the occurrence of retransmission events.

Furthermore, the absence of clustering outside the main body in profiles 4 and 7 indicates low variance, which is further confirmed by the small min-to-max range of profile 7.

[3] Clusters in a histogram are groupings of data points located close together on the number line. They represent a high concentration of data in a specific range, often appearing as one or more bars with high frequencies, or "peaks".



(a) Profile 1

(b) Profile 4

(c) Profile 7

(d) Profile 8

Fig. 2: Histograms of transmission time (UL) from diffrenet profiles

As for profile 8, the presence of a single value for all packets, which is confirmed upon analyzing the data set. This outcome indicates that the system transitioned to an ideal behavior, where each packet possesses the same TT.

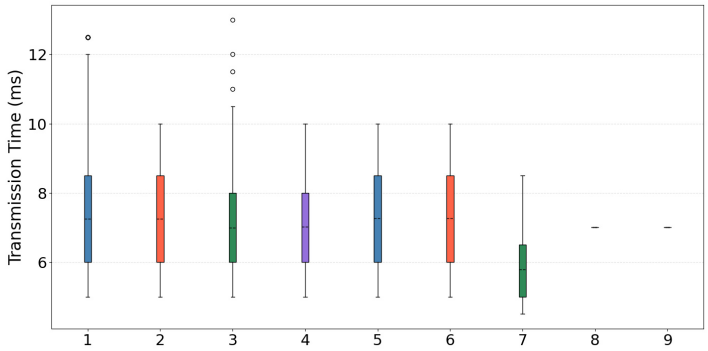


Fig. 3: Transmission time (UL) boxplot diagrams for all profiles

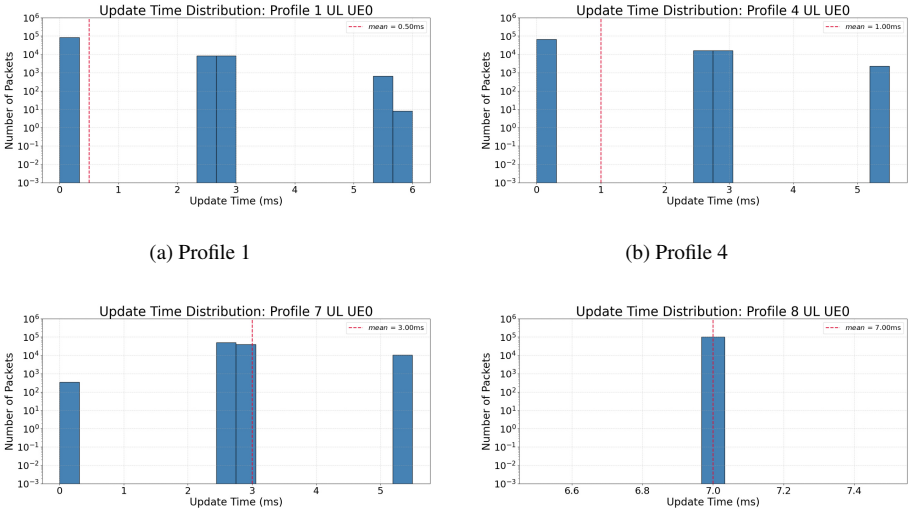
The boxplots of TT values across all profiles are compared, as illustrated in figure 3. A notable decrease in the range, in which the majority of the TT values fall, is visible transitioning from profile 1 to profile 4. A similar trend is noted when transitioning from profile 6 to profile 4. Furthermore, as illustrated by profiles 7 through 9, a substantial alteration in the TI significantly influences the range of TT values, so that even an ideal

behavior is noted in profiles 8 and 9, where all packets arrive within 7 ms.

Furthermore, it is noteworthy that the downlink profiles exhibit nearly identical TT distributions, with medians occurring near 2.5 ms, narrow boxes, and only a few outliers around 3 - 4 ms.

4.2 Analysis of the update time

The update times in profile 1 cluster closely around the mean, with 95% of values ranging from 0 to 3 ms and a small subset extending beyond this primary distribution (see figure 4a). Profile 4 exhibits similar characteristics, with a mean near 3 ms, but shows a narrower range of update times limited to 0 to 5.5 ms (see figure 4b). Likewise, profile 7's values are concentrated around the mean, mostly spanning from 2.5 to 3 ms (see figure 4c). In contrast, profile 8 displays a distinct pattern, appearing as a single-bin histogram centered at 7 ms (see figure 4d).



(a) Profile 1

(b) Profile 4

(c) Profile 7

(d) Profile 8

Fig. 4: Histograms of update time (UL) from diffrenet profiles

Across all selected profiles, the mean values are equivalent to the TI, which is an expected behavior of the UT. However, the number of histogram bins, outside of the main cluster around the mean values, and their heights in profiles 1, 4, and 7 indicate the degree of variance in the UT values. As the variance of the data decreased during the transition from profiles 1 to 4 and subsequently to profile 7, a substantial reduction in the number or heights

of the aforementioned bins is evident. Finally, the observation of the single bin histogram of profile 8 signifies the absence of variance and the presence of a uniform value across all packets, a conclusion that is further confirmed upon analysis of the data set.

A thorough examination of the UL update time distributions by profile, as illustrated in figure 5, reveals that they exhibit low central values with extended upper tails and the presence of outliers for early profiles. Additionally, a discernible tight cluster is observed near 2.8–3 ms in profile 7, and near-constant updates are noted at approximately 7 ms (profile 8) and 10 ms (profile 9).The DL update-time boxplots demonstrate medians equivalent to the TI for Profiles 1–6, exhibiting minimal variability. Subsequently, there is a discrete shift to approximately 3 milliseconds (profile 7), 7 ms (profile 8), and 10 ms (profile 9), with minimal dispersion, suggesting a predominantly monotonic pattern across profiles.

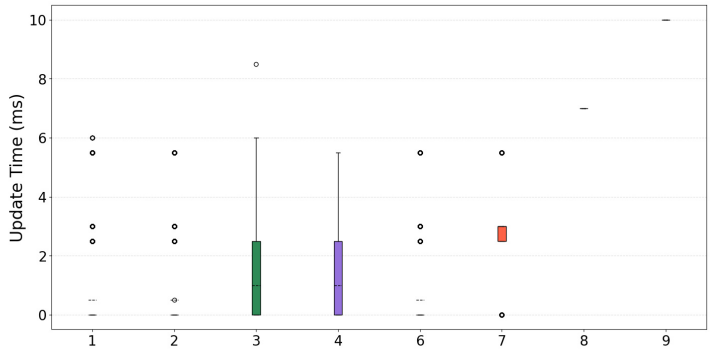


Fig. 5: Graph comparing the boxplots update time (UL) values across profiles

4.3 Analysis of the packet loss ratio

As depicted in figure 6, the bar chart compares the uplink and downlink packet loss ratio across all profiles, demonstrating a consistent higher UL PLR compared to DL and a discernible decline in both as the traffic interval increases in profiles 7–9. The presence of spikes is evident in Profiles 2 and 5, where brief 0.5 ms intervals, characterized by numerous UEs, result in the most significant UL loss. Conversely, DL exhibits a lower yet analogous trend.

Profiles 7–9, which extend the transfer interval from 3 to 10 ms, demonstrate an order-of-magnitude reduction in PLR relative to profiles 1–6. This suggests that decreasing the offered load enhances reliability.

The enhancement of gNB transmission power from 23 dBm to 30 dBm yields a mixture of advantages and disadvantages, as illustrated in figure 6. profile 6 exhibits a reduced UL PLR in comparison to profile 5. Conversely, profile 4 demonstrates a comparable performance

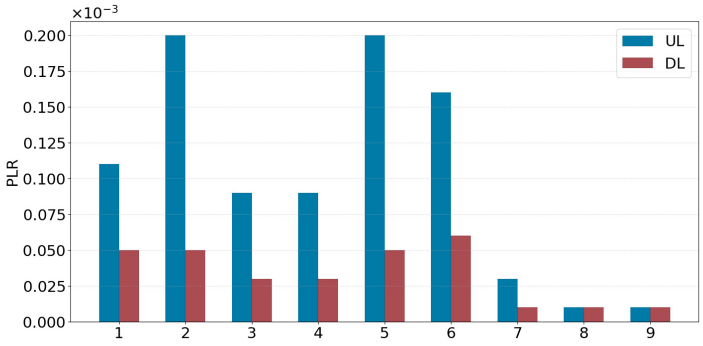


Fig. 6: Bar graph comparing PLR values across profiles

to profile 3, despite the augmented power level, indicating that the load capacity may be a more significant factor than the power enhancement in these specific configurations.

The offered traffic load and contention demonstrate that short 0.5 ms intervals with numerous UE devices (30) in profiles 1–6 elevate buffer pressure and scheduling demand, thereby increasing PLR in comparison to longer intervals in profiles 7–9 with equivalent or fewer UEs.

The interplay between bandwidth and power is a complex phenomenon that warrants further investigation. A transition from 50 to 100 MHz has been observed to result in a reduction of congestion in certain scenarios. However, when the interval is of an extremely short nature and the number of UEs is high, the efficacy of this adjustment is constrained. To mitigate this limitation, it is necessary to pair this adjustment with either elevated gNB power or a reduction in load, as evidenced by the comparative analysis of profiles 3–6. The impact of the number of UEs is evident in scenarios with 30 UEs, which demonstrate higher PLR compared to those with 20 UEs at comparable intervals and bandwidths. This observation signifies that multiplexing pressure is a predominant factor contributing to losses.

5 Conclusions

This paper investigates the preconditions and performance trade-offs of CoCoCo in WNCS through a simulation-based study, addressing the need for representative case studies and cross-layer analysis. The adoption of a simulation methodology employing OMNeT++, INET, and Simu5G was instrumental in the investigation of a minimal industrial scenario, which models critical aspects of wireless communication that affect control performance under varying network profiles. The evaluation process entailed a meticulous examination of pivotal metrics, encompassing TT, UT and PLR. This comprehensive approach yields insights into performance patterns and trade-offs, thereby illuminating the potential of

CoCoCo to enhance reliability and latency-aware control. These simulation results align with patterns previously observed in real-world studies [UCN22].Nevertheless, these frameworks currently abstract certain physical and dynamic complexities inherent in industrial networks, which limits the direct applicability of results to more complex or variable real-world settings. In light of these findings, the subsequent logical steps entail the augmentation of scenario complexity to encompass mobility, interference diversity, and heterogeneous packet sizes or message lengths, in conjunction with plant-aware triggering and advanced HARQ/scheduler telemetry to enhance the localization of communication tail events. Subsequent validation through testbed experiments is essential to establish a direct mapping from network QoS to control-level QoC, demonstrating measurable gains of co-design under dynamic industrial conditions.

References

[3r24] 3rd Generation Partnership Project (3GPP): Service requirements for cyber-physical control applications in vertical domains; Stage 1 (Release 19), 3GPP TS 22.104, version V19.2.0, (C) 2024 3GPP Organizational Partners; Release 19, Sophia Antipolis, Valbonne, France: 3GPP Technical Specification Group Services and System Aspects, 2024.

[Ay24] Ayan, O. et al.: Enabling Communication and Control Co-Design in 6G Networks, 2024, arXiv: 2402.16462 [cs.NI], https://arxiv.org/abs/2402.16462.

[Bh21] Bhatia, L. et al.: Control Communication Co-Design for Wide Area Cyber-Physical Systems. ACM Trans. Cyber-Phys. Syst. 5 (2), 2021, https://doi.org/10.1145/3418528.

[CL10] Chamaken, A.; Litz, L.: Joint design of control and communication in wireless networked control systems: A case study. In: Proceedings of the 2010 American Control Conference. Pp. 1835–1840, 2010.

[De21] Deng, L. et al.: Joint design of control policy and network scheduling policy for wireless networked control systems: Theory and application. Information Sciences 575, pp. 563–586, 2021, https://www.sciencedirect.com/science/article/pii/S0020025521006149.

[Ei19] Eisen, M. et al.: Control Aware Radio Resource Allocation in Low Latency Wireless Control Systems. IEEE Internet of Things Journal 6 (5), pp. 7878–7890, 2019.

[Ei22] Eisen, M. et al.: Communication-Control Co-design in Wireless Edge Industrial Systems. In: 2022 IEEE 18th International Conference on Factory Communication Systems (WFCS). Pp. 1–8, 2022.

[Fr23] Frotzscher, A. et al.: Self-Organization and Collaboration in Robotic Manufacturing Systems. In. Pp. 1–6, 2023.

[Gi24] Girgis, A. M. et al.: Semantic and Logical Communication-Control Codesign for Correlated Dynamical Systems. IEEE Internet of Things Journal 11 (7), pp. 12631–12648, 2024.

[Li14] Li, J. et al.: Communication and control co-design for wireless sensor networked control systems. In: Proceeding of the 11th World Congress on Intelligent Control and Automation. Pp. 156–161, 2014.

[Me22] Melnyk, S. et al.: Wireless Industrial Communication and Control System: AI Assisted Blind Spot Detection-and-Avoidance for AGVs. In: Proceedings of the 3rd International Conference on Innovative Intelligent Industrial Production and Logistics - ETCIIM. INSTICC, SciTePress, pp. 307–313, 2022.

[PAJ11] Park, P.; Araújo, J.; Johansson, K. H.: Wireless networked control system co-design. In: 2011 International Conference on Networking, Sensing and Control. Pp. 486–491, 2011.

[QFY23] Qiao, Y.; Fu, Y.; Yuan, M.: Communication–Control Co-Design in Wireless Networks: A Cloud Control AGV Example. IEEE Internet of Things Journal 10 (3), pp. 2346–2359, 2023.

[Ta19] Tayade, S. et al.: Cloud Control AGV over Rayleigh Fading Channel - The Faster The Better. In: SCC 2019; 12th International ITG Conference on Systems, Communications and Coding. Pp. 1–6, 2019.

[Ta20] Tayade, S. et al.: Impact of Short Blocklength Coding on Stability of an AGV Control System in Industry 4.0. In: 2020 IEEE International Conference on Communications Workshops (ICC Workshops). Pp. 1–6, 2020.

[UCN22] Underberg, L.; Cainelli, G.; Neugebauer, T.: Performance Evaluation of a Campus Network emphasizing the ICT-OT-relation. In. Pp. 637–650, 2022.

[Wa23] Wang, Z. et al.: QoS-Oriented Sensing–Communication–Control Co-Design for UAV-Enabled Positioning. IEEE Transactions on Green Communications and Networking 7 (1), pp. 497–511, 2023.

[Zh19] Zhao, G. et al.: Toward Real-Time Control in Future Wireless Networks: Communication-Control Co-Design. IEEE Communications Magazine 57 (2), pp. 138–144, 2019.

http://dx.doi.org/10.25673/121097

Einrichtung und Optimierung von virtualisierten Steuerungen für die Nutzung mit Echtzeit-Anwendungen

Lukas Schreckenberg[1], Sergej Gamper[1], Alexander Biendarra[1]

Abstract: Virtualisierte Steuerungen (vPLC) halten Einzug in die OT, um Rechenkapazitäten von dezentralen Hardware-PLCs an dynamisch orchestrierbare Server zu verlagern. Frühere Arbeiten mit Virtualisierungslösungen zeigten prinzipielle RT-Tauglichkeit. Aktuelle vPLC-Produkte bestätigen die Machbarkeit. Dabei erreichen sie jedoch nur minimale Zykluszeiten von 2–4 ms in der Kommunikation und bleiben damit gegenüber performanten Hardware-PLCs zurück. Ziel ist daher die Identifikation und Bewertung von Optimierungsoptionen für niedrigere Zykluszeiten und verbesserten Determinismus. Aufbauend auf existierenden vPLC-Produkten und vorliegenden Messungen werden Host- und vPLC-seitige Optimierungsoptionen betrachtet: eXpress Data Path (XDP) zur Umgehung des Network-Stacks, Single-Root I/O Virtualization, sowie präzise Zeitsynchronisation über IEEE 802.1AS-2020. Eine praxisnahe Testumgebung mit echtzeitfähigem Linux, TSN-Switch und leistungsfähiger COTS-Plattform wird skizziert. Als Bewertungsmethoden werden Roundtrip-Messungen mit IEC-61131-3-Programmen als auch Möglichkeiten zur Jitter-Analysen vorgestellt. Erste Befunde zeigen zwei zentrale Herausforderungen bei XDP im Container: die Notwendigkeit von Host-Root-Rechten und die Einschränkung auf einen AF_XDP-Socket. Diese Limitierungen gefährden derzeit XDP-basierte vPLC-Optimierungen in Container-Architekturen.

Keywords: virtuelle Steuerung, niedrige Latenz, Zeitsynchronisation, Ethernet TSN, XDP

1 Motivation

Virtualisierungstechniken, die in der IT-Welt und dort vorwiegend im Cloud-Umfeld weit verbreitet sind, halten derzeit in Form von virtual Programmable Logic Controller (vPLC) auch Einzug in die OT-Domäne. In Kombination mit Ethernet Time Sensitive Networking (TSN) und konvergenten IT/OT-Ethernet-Netzwerken soll es so ermöglicht werden, die benötigten Rechenkapazitäten von dezentralen Steuerungen an den Maschinen zu dynamisch orchestrierten und skalierbaren Edge-Servern zu verschieben. Gleichzeitig sollen die geforderten Rahmenbedingungen von Realtime Anwendungen, wie niedrige Zykluszeiten und Determinismus in der Kommunikation, weiter erfüllt werden. In einer Marktanalyse geht das Marktforschungsunternehmen IoT Analytics von einem Marktanteil von Soft- und vPLC von ca. 25% im Jahr 2030 aus [Ra24]. Es wurde bereits zuvor gezeigt, dass sowohl bei der Nutzung von Docker Containern [So21] als auch bei der Verwendung von virtuellen Maschinen (VM) [Ja22] die Möglichkeit besteht, die geforderten Latenzen und Jitter in der Kommunikation zu erreichen. Dazu wurden jedoch vereinfachte Implementierungen ohne echte vPLC-Software-Stacks und ohne industrielle Kommunikationsprotokolle wie PROFINET untersucht oder mit EtherCAT, ohne die

[1] Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Germany, e-mail: {firstname.lastname}@iosb-ina.fraunhofer.de

Nutzung des TSN-Profils [Et18], ein nicht TSN fähiges Protokoll eingesetzt. Mittlerweile sind vPLC-Produkte jedoch frei auf dem Markt verfügbar. Eine Untersuchung der Phoenix Contact vPLC-Lösung bestätigte die bisherigen Ergebnisse für eine solche vPLC mit einer Kommunikation über ein Ethernet TSN fähiges Protokoll [Sp24]. Allerdings wurden hier ausschließlich Zykluszeiten bis $4\,ms$ ohne weitere Optimierungen und ohne deterministisches Sendeverhalten betrachtet. In einer Untersuchung einer Siemens vPLC-Lösung konnte eine Zykluszeit von $2\,ms$ ohne Frameverlust erreicht werden [Ga25]. Diese erreichten Zykluszeiten sind zwar schon für viele Anwendungsszenarien ausreichend, aber immer noch schlechter als performante klassische Hardwaresteuerungen und damit, je nach Nutzungsszenario, zu langsam oder ungenau. Der nächste Schritt ist daher die Ermittlung von Optimierungsoptionen und Untersuchung von vPLC-Lösungen auf niedrigere Zykluszeiten und damit die Nutzungsmöglichkeiten in anspruchsvollen Szenarien. In diesem Papier sollen verschiedene dieser Optimierungsoptionen sowohl der vPLC selbst als auch des Host-Systems betrachtet werden. Des Weiteren werden Messmethoden zur Evaluierung vorgestellt.

2 Virtuelle Steuerungstechnik

Virtualisierung von Steuerungstechnik entkoppelt die PLC-Laufzeit von dedizierter Hardware und betreibt sie als softwarebasierte PLC in Containern oder virtuellen Maschinen auf Standard-IT-Hardware, auch Commercial-off-the-shelf (Cots) Hardware genannt, an der Maschine, an der Edge, im Rechenzentrum oder in der Cloud. Kernkonzepte sind modulare, containerisierte Laufzeiten oder virtuelle Maschinen, Orchestrierung (z. B. automatisiertes Starten und Skalieren), softwaredefinierte I/O-Anbindung über industrielle Protokolle und Gateways (z.B. PROFINET, OPC UA, MQTT) sowie Echtzeitunterstützung durch RT-Erweiterungen des Hosts. Dies ermöglicht neben dem produktiven Betrieb auch digitale Zwillinge, Simulation, Test und CI/CD-basierte Inbetriebnahme mit identischen Images über den gesamten Lebenszyklus. Zentrale Verwaltung, Versionierung, Monitoring und Security-Isolation (Mandantentrennung) sind integraler Bestandteil. Gegenüber konventioneller Steuerungstechnik bietet Virtualisierung schnellere Skalierung und Rollout, geringere Hardwarebindung und Stillstandszeiten, effizientere Nutzung von Ressourcen, vereinfachte Updates und Wartung sowie beschleunigte Entwicklung und Tests bei gleichzeitiger Verbesserung von Verfügbarkeit und Flexibilität. Herausforderungen sind dabei die geringere Performance und der schlechtere Determinismus durch die zusätzliche Abstraktionsschichten und geteilten Ressourcen.

Grundsätzlich gibt es bei der Verwendung von Virtualisierungstechniken die Abwägung zwischen der Nutzung von Containern oder VMs. Dabei haben Container den Vorteil, dass sie leichtgewichtiger sind als VMs, da sie den Kernel des Hostsystems mitbenutzen, bzw. für der Kernel ist der Container nur ein mittels verschiedener Mittel isoliert laufender Prozess. Dadurch entfällt bei Containern die Abstraktionsschicht des Hypervisors.

Eine weitere Isolierung kann durch die Verwendung von User Namespaces erzeugt werden. Hierbei werden die Ids von Usern umgemappt, sodass beispielsweise der Root Users des Containers auf dem Host-System als normaler User erscheint.

Ein Nachteil der Containerisierung ist trotz dieser Mittel, dass die Isolation zwischen den Containern untereinander und dem Host nicht so stark ist wie bei VMs, welche vollkommen gekapselt sind und einen eigenen Kernel haben. Dadurch können sie auch voll auf Kernelfunktionen zugreifen, die bei Containern dem Host vorbehalten sein können.

Es muss also immer abgewogen werden, welche Aspekte wichtiger sind oder ob es Blocker gibt. Im Folgenden werden verschiedene am Markt verfügbare vPLC-Lösungen vorgestellt.

2.1 Beckhoff virtualPLC

Eine Lösung für den Einsatz von virtuellen Steuerungen im Bereich EtherCAT bietet TwinCAT für Linux. Es ermöglicht die ressourceneffiziente, virtualisierte Verteilung von TwinCAT-Anwendungen als Container, sowohl lokal auf der Steuerung als auch im Rechenzentrum. Durch den modularen Aufbau lassen sich einzelne TwinCAT-Komponenten auf mehrere Container verteilen und durch das Automation Device Specification Protokoll über MQTT (ADS-over-MQTT) netzwerkübergreifend verbinden. Auch eine Anbindung eines virtualisierten Engineering-Systems ist möglich. Die vPLC erweitert die Einsatzmöglichkeiten ins lokale Rechenzentrum, wodurch Steuerungs-, Simulations- und Testaufgaben von der Maschine bzw. dem Steuerungs-PC ausgelagert werden können. Besonders interessant ist dies für Applikationen mit geringeren Echtzeitanforderungen. Die Kommunikation mit dezentralen I/Os erfolgt über den EtherCAT-Koppler EK1000, der geswitchte und geroutete IP-Netzwerke unterstützt. [Be25]

2.2 CODESYS Virtual Control SL

CODESYS Virtual Control SL ist ein vollständig virtualisiertes Steuerungssystem, das IEC-61131-3-kompatible SPS-Funktionalität unabhängig von der zugrunde liegenden Hardware bietet. Es abstrahiert dabei die SoftPLC Lösung von CODESYS weiter und lässt sich flexibel in Containern, virtuellen Maschinen oder Hypervisor-Umgebungen betreiben und skaliert je nach verfügbarer Rechenleistung. Die Verwaltung mehrerer Instanzen erfolgt über Linux-Skripte, Orchestrierungstools wie Kubernetes oder perspektivisch über den CODESYS Automation Server. Der Zugriff auf die Feldebene wird über virtualisierte Netzwerktechnologien wie VLAN sowie industrielle Ethernet-Protokolle realisiert. Die Lösung reduziert den Aufwand für Hardware, Verdrahtung und Wartung und unterstützt eine zentrale, flexible Steuerungsarchitektur. Ihre Sicherheitsarchitektur basiert auf isolierten Steuerungsinstanzen mit leistungsfähiger Interkommunikation. [CO25a] Durch die Variante CODESYS Virtual Safe Control SL ist es auch möglich Anwendungen, die

eine nach IEC 61508 SIL3 zertifizierte Steuerungen erfordern, in einer virtuellen Umgebung zu betreiben. [CO25b]

2.3 Siemens SIMATIC S7-1500V

Siemens stellt mit der SIMATIC S7-1500V eine vPLC zur Verfügung, die als App auf einem Industrial Edge Device (IED) im Siemens Industrial Edge Ökosystem genutzt werden kann. Bei der App handelt es sich um eine Docker Container Lösung. Als IED kann entweder ein physikalisches IED von Siemens, eine mit einem Siemens Betriebssystem bespielte Cots Hardware oder eine VM mit einem Siemens Industrial Edge Virtual Device Image betrieben werden. Die vPLC kann wie eine klassische Hardware-S7-1500 im TIA Portal projektiert werden. Sie ermöglicht schnelle, flexible Skalierung und parallele Instanzen für Entwicklung, Test, Simulation und produktive Anwendungen mit moderaten Echtzeitanforderungen. Funktionen und Diagnose verhalten sich weitgehend, wie bei der Hardware-SPS, sodass bestehendes Know-how und Projekte nutzbar bleiben. Vorteile sind geringere Hardwareabhängigkeit, zentrale Verwaltung, einfache Updates und die Integration in moderne IT-/Container-Workflows. [Si25] Auch hier gibt es mit der SIMATIC S7-1500V F eine Variante, die für funktionale Sicherheit zertifiziert ist. [Si24]

2.4 Phoenix Contact Virtual PLCnext Control

Virtual PLCnext Control von Phoenix Contact ist eine vPLC, bei der die Steuerungsfunktionalität in einem Container ausgeführt wird. Dadurch können mehrere virtuelle Steuerungsinstanzen auf einer gemeinsamen Hardwareplattform betrieben werden, was zu höherer Ressourcenauslastung und geringeren Kosten führt. Die Plattform unterstützt eine flexible Skalierung durch dynamische Zuweisung von Rechenleistung und Speicher und bietet Portabilität von Applikationen über unterschiedliche Systeme hinweg. Funktionen klassischer PLCnext-Hardware wie Firewall, Zertifikatsmanagement und Feldbusunterstützung sind ebenfalls integriert. Der Datenaustausch mit anderen Systemen erfolgt über etablierte Industriestandards wie OPC UA und PROFINET. Zielanwendungen liegen vor allem im Bereich Edge-Computing und in Rechenzentren, mit Fokus auf geringe Latenzen, hohe Flexibilität und IT/OT-Konvergenz. Die Lösung berücksichtigt zudem moderne Cybersecurity-Anforderungen durch Containerisolation und den kombinierten Einsatz von IT- und OT-Sicherheitsmechanismen. [Ph25]

3 Optimierungsoptionen für virtuelle Steuerungstechnik

Im Folgenden werden die verschiedenen untersuchten Technologien erläutert, die in diesem Papier erwogen werden, um niedrigere Zykluszeiten in der Kommunikation mit der vPLC zu erreichen.

3.1 eXpress Data Path (XDP) und XDP Zerocopy

Die Technologie eXpress Data Path (XDP) ist eine Linux-Kernel-Funktionalität, die wiederum die extended Berkley Packet Filter (eBPF) Funktionalität des Linux-Kernels nutzt. Die grundlegende Funktion von eBPF ist es an bestimmten Stellen im Kernel, sogenannten Kernel-Hooks, kleine Programme laden zu können, die dann in die Kernelabläufe an diesen Stellen eingreifen können. Der XDP-Hook befindet sich im Ethernet-Empfangspfad und wird direkt nach dem Treiber und vor dem Linux Network/IP-Stack durchlaufen. In einem XDP-Programm wird jedes durchlaufende Frame gefiltert und dann darauf aufbauend, verworfen, an den Network/IP-Stack weitergeleitet oder umgeleitet. Als Ziel der Umleitung kann entweder ein anderes Interface oder ein AF_XDP-Socket dienen. Der AF_XDP-Socket wird in einer Applikation geöffnet und über ihn kann die Applikation direkt auf die Frames zugreifen und der Networkstack wird umgangen. Je nach Anwendungsfall kann so z.B. das Herausfiltern von schädlichem Denial-of-Service-Traffic mit weniger Prozessorlast durchgeführt werden oder die verringerte Verarbeitungsdauer kann genutzt werden, um bei zeitkritischem Traffic die Latenz zu senken. [Hø18]

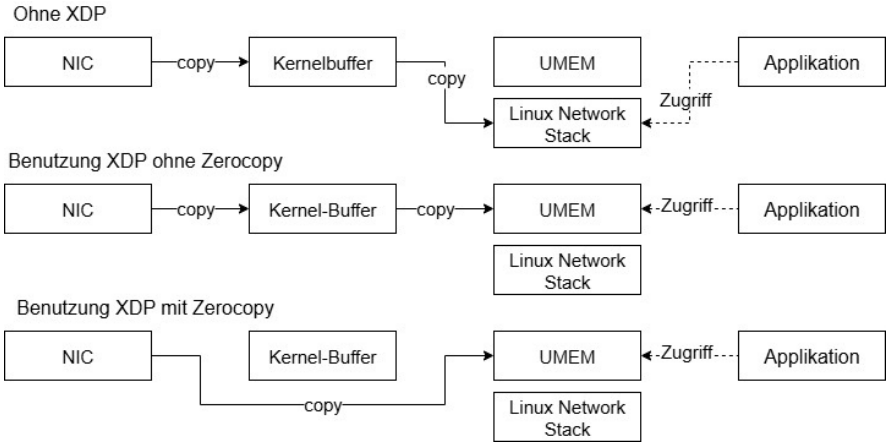


Abb. 1: Schematische Abläufe bei Benutzung von XDP

Im Kontext von XDP müssen auch die Hardware Empfangs- und Sendequeues der Netzwerkadapter betrachtet werden, da ein AF_XDP-Socket fest einer dieser Queues zugeordnet wird. Die Queues werden genutzt, um Traffic-Priorisierung sowohl in Empfangs- als auch in Senderichtung vornehmen zu können.

Eine weitere Senkung der Latenz kann durch die Nutzung von XDP-Zerocopy erreicht werden. Hierbei wird auf das Kopieren des Frames in den Kernel-Buffer verzichtet und er wird stattdessen direkt in den UMEM genannten Zugriffs-Buffer abgelegt. Die

verschiedenen Ablaufswege sind schematisch in Abb. 1 dargestellt. Der Zerocopy-Mechanismus kann für jeden AF_XDP-Socket einzeln aktiviert werden.

Die Nutzung XDP und XDP-Zerocopy verspricht zum einen auf dem Host als auch im Container Verbesserungen in der Übertragungsperformance. Auf dem Host ist das Weiterleiten der Frames direkt an das Interface des Containers das Ziel und im Container die Weiterleitung an den AF_XDP-Socket der vPLC-Applikation.

3.2 Single Root I/O Virtualization and Sharing (SR-IOV)

SR-IOV ist eine Technologie, die es ermöglicht ein physikalisches PCIe Gerät in mehrere virtuelle Geräte, virtuelle Funktion (VF) genannt, aufzuteilen. Diese VF, können dann an die Virtualisierungslösung durchgereicht und von ihr verwendet werden, als ob sie exklusiv auf ein physikalisches Gerät zugreifen kann.

Bei manchen Netzwerkkarten gibt es zusätzlich noch die Option, dass eine VF auf mehrere Queues zurückgreifen kann, sodass dadurch auch in der Virtualisierungslösung eine Traffic-Steuerung stattfinden kann. Ein Nachteil von SR-IOV ist, dass es sich um eine Technologie handelt, die hauptsächlich auf leistungsfähigen Server-Netzwerkadaptern implementiert ist und deswegen nur eingeschränkt verfügbar ist. SR-IOV ist ein Standard der Peripheral Component Interconnect Special Interest Group (PCI-SIG), dem Standardisierungsgremium von PCI und PCIe. [In11]

3.3 Zeitsynchronisation

Im Bereich der Steuerungstechnik wird die Synchronisation nach dem Standard IEEE 802.1AS-20202 durchgeführt. Dabei wird noch Unterschieden, ob nur die Hardwareuhr des Ethernet Netzwerkadapter synchronisiert ist oder auch die Hardwareuhr des Systems. Bei der Betrachtung der Zeitsynchronisation im Virtualisierungsumfeld gibt es neben der Synchronisation des Hosts noch die Synchronisation einer VM zu beachten. Die Synchronisation eines Containers benötigt keine gesonderte Betrachtung, da dieser den Kernel und damit die Systemuhr des Hosts mit nutzt.

Intel® Time Coordinated Computing

Das Feature Intel® Time Coordinated Computing (TCC) ist eine Funktionalität von Intel Edge Computing Plattformen. Hier werden der Prozessor und die Verarbeitung von Aufgaben des Hosts auf Zuverlässigkeit und niedrigen Latenz optimiert. [In24] Eine dieser Optimierungen ist PCIe PTM, das im folgenden Absatz beschrieben ist.

PCIe Precision Time Measurement (PTM)

PTM ist ein weiterer von der PCI-SIG herausgegebener Teilstandard. Er beschreibt, wie auf einem zeitsynchronisierten System, die Ungenauigkeit der Synchronisation auf dem

letzten Teilabschnitt, verursacht von Abweichungen der Uhren des Ethernet-Netzwerkadapters und des Prozessors, herausgerechnet und kompensiert werden kann. [St24]

Zeitsynchronisation einer VM

Da sich eine VM, im Gegensatz zu einem Container, mehr wie ein eigenständiges System verhält, also auch einen eigenen Kernel besitzt, muss die Software-Systemuhr der VM explizit synchronisiert werden. Da hier durch die Abstraktionsschichten zusätzliche Ungenauigkeiten auftreten und nur eine Softwareuhr genutzt wird, die funktionsbedingt auch von der Systemlast abhängt, ist die Synchronisierungsgenaugigkeit im Vergleich zu einer Hardwareuhr sehr gering und liegt anstatt weniger Nanosekunden in der Größenordnung von hunderten Mikrosekunden [On18].

Nutzung von PTP_KVM

Bei der Verwendung von Linux mit KVM als Hypervisor kann das Problem abgeschwächt werden, in dem der PTP_KVM Mechanismus genutzt wird. Hier wird auf dem Host eine virtuelle PTP Uhr angelegt, die aus der VM direkt abgefragt werden kann, sodass zusätzlicher Netzwerkprotokoll Overhead vermieden wird. Dadurch verbessert sich die Synchronsationsgenauigkeit deutlich. [Ji20]

4 Testumgebung

4.1 Hardwareumgebung

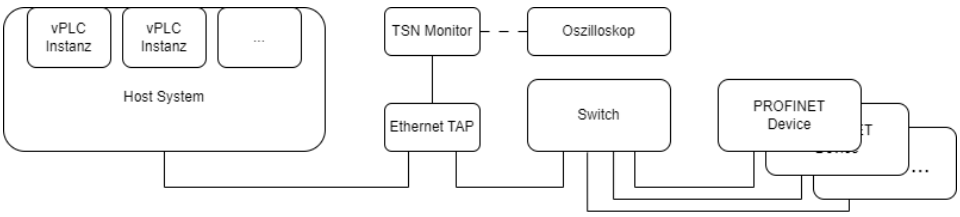


Abb. 2: Schematische Darstellung des Setups

Der vorgeschlagene Testaufbau ist in Abb. 2 dargestellt. Als Host System sollte ein leistungsfähiger Cots Server genutzt werden, bei dem die zuvor angesprochenen Funktionen unterstütz werden. Für das Sammeln erster Erfahrungswerte kam beispielsweise folgende Konfiguration zum Einsatz. Es wurde ein Intel i9-12900E x86_64 Prozessor als Basis genutzt, da diese Plattform die zuvor angesprochene TCC-Funktionalität unterstützt. Als Hauptspeicher sind 64 GB verbaut. Der verwendete Netzwerk Adapter ist ein Intel E810, mit Unterstützung für SR-IOV und vielen verfügbaren Queues.

Der Switch sollte TSN fähig sein und die Synchronisation nach IEEE 802.1AS-2020 unterstützen.

4.2 Softwareumgebung

Das Betriebssystem des Host System muss echtzeitfähig sein und die erläuterten Features unterstützen. Bei dem in Abschnitt 4.1 angesprochenen System kam beispielsweise ein Debian Linux 12 mit dem in den Paketquellen verfügbaren echtzeitfähigen PREMPT_RT-Kernel zum Einsatz.

4.3 Messmethoden

Roundtrip Messung mittels IEC 61131-3 Programm

Um die Performance einer Anwendung im Feld bestmöglich nachstellen zu können, sollte für die Performancemessung die Roundtrip-Kommunikation zwischen der vPLC und einem PROFINET Device und wieder zurück unter Nutzung von RT-Kommunikation untersucht. Dazu kann beispielsweise sowohl ein IEC 61131-3 Programm für die vPLC als Controller und ein IEC 61131-3 Programm für das PROFINET Device erstellt werden. Das Programm auf der vPLC schreibt einen Counter-Wert in die Nutzlast des PROFINET Frames und speichert den dazugehörigen Sendezeitstempel ab. Das Programm auf dem Device spiegelt diese Nachricht nun und schickt sie zurück an die vPLC. Hier werden die Frames anhand des Counterwertes erkannt und der Empfangszeitstempel wir mit dem Sendezeitstempel verglichen. Wie in Abb. 3 gezeigt, setzt sich der ermittelte Wert dabei aus den Verarbeitungszeiten in der vPLC und dem Device als auch durch die beiden Übertragungsstrecken zusammen.

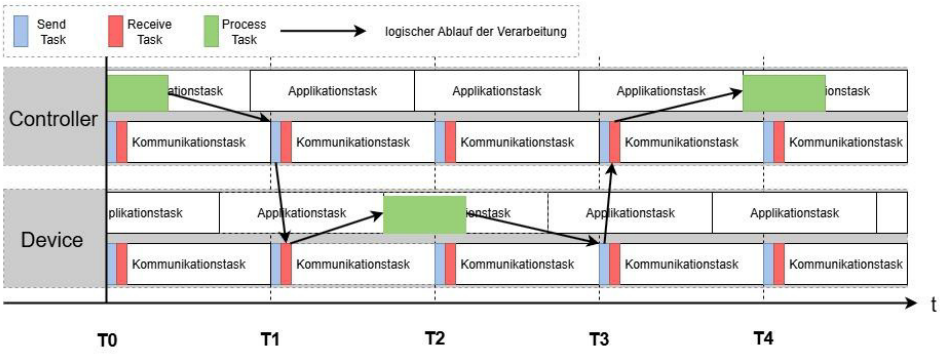


Abb. 3: Ablauf der Roundtrip Kommunikation

Jittermessung mit dem TSN-Monitor

Um die Genauigkeit des Sendeverhalten der vPLC zu untersuchen, kann der Jitter der PROFINET Frames mit dem TSN-Monitor vermessen werden.

Der TSN-Monitor ist eine FPGA basierte, vom Fraunhofer IOSB-INA selbst entwickelte, Messlösung, mit der vorkonfigurierte Frames in Echtzeit auf einer Ethernet-Leitung erkannt und dann als Signal ausgegeben werden. Dieses Signal kann dann mit einem Oszilloskop visualisiert und so die Zykluszeit und der Jitter eines zyklischen Ethernet Protokolls wie PROFINET ausgewertet werden.

TSN Box und TSN Tools von TSN Systems

Eine Alternative zum TSN-Monitor ist das Testsystem von TSN Systems. Hierbei handelt es sich bei der TSN Box um eine Lösung für die Aufzeichnung von Netzwerkverkehr auf FPGA Basis und bei den TSN Tools um eine Auswertungssoftware, in der über eine GUI der Traffic gefiltert, untersucht und dann auch grafisch dargestellt werden kann. [TS25]

5 Diskussion und Bewertung der Lösungsarchitekturen und Optimierungsoptionen

Bei der Untersuchung der Lösungsarchitekturen und Optionen sind zwei Herausforderungen identifiziert worden, die mit der Verwendung von XDP in Containern zusammenhängen. Das erste Problem ist, dass es nicht möglich war, XDP im Container in Betrieb zu nehmen, wenn in diesem der Root User in einen User Namespace gemappt war. Es werden also Host Rootrechte im Container benötigt. Zusammen mit den generell erhöhten Rechten, die benötigt werden, um XDP als Kernel-Mechanismus zu benutzen, wurde so die Isolierung so stark reduziert, dass sie effektiv nicht mehr bestand.

Die zweite Herausforderung ist es, innerhalb eines Containers mehr als eine Netzwerk Queue zu nutzen. Sowohl bei der Verwendung von virtuellen Netzwerkadaptern mit der Unterstützung von mehreren Queues als auch bei der Benutzung einer VF mit mehreren konfigurierten Queues konnte nicht mehr als ein AF_XDP-Socket geöffnet werden. Ob es sich dabei um eine Einschränkung des XDP-Mechanismus oder eine noch fehlerhafte Konfiguration handelt, konnte im Rahmen des Papiers nicht geklärt werden.

Falls die Ursache für diese Probleme nicht gefunden und behoben wird, ließen sich vPLCs auf Basis von Containern und der Nutzung von XDP nur über Umwege oder überhaupt nicht umsetzen. Und damit wäre ein möglicher Weg für die Optimierung von vPLCs versperrt.

6 Ausblick

Die Gründe für die Probleme bei der Nutzung von XDP in einem Container sollten gefunden und gelöst werden.

In einem nächsten Schritt sollten die betrachteten Optimierungsmöglichkeiten implementiert und die Wirksamkeit auf die Latenz und den Jitter in der Kommunikation überprüft werden. Weiterhin sollten dann auch die Skalierbarkeit und die Veränderung der Performance der Lösungen unter zusätzlicher System- und Netzlast untersucht werden.

7 Literaturverzeichnis

[Be25] Beckhoff Automation GmbH & Co. KG: Virtual PLC. https://www.beckhoff.com/de-de/produkte/automation/virtual-plc/, Stand: 01.10.2025.

[CO25a] CODESYS GmbH: CODESYS Virtual Control SL. https://de.codesys.com/produkte/runtime/virtual-control-sl/, Stand: 01.10.2025.

[CO25b] CODESYS GmbH: CODESYS Virtual Safe Control SL. https://de.codesys.com/produkte/safety/virtual-safe-control-sl/, Stand: 01.10.2025.

[Et18] EtherCAT und TSN: die perfekte Ergänzung, 2018.

[Ga25] Gaffurini, M. et al.: Characterizing the Real-Time Communication Performance of Virtual PLC in Industrial Edge Platform. IEEE Open Journal of Instrumentation and Measurement 4, S. 1–11, 2025.

[Hø18] Høiland-Jørgensen, T. et al.: The eXpress data path. ACM, S. 54–66, 2018.

[In11] Intel: PCI-SIG SR-IOV Primer, 2011.

[In24] Intel: Intel® Edge Real-Time Get-started Guide, 2024.

[Ja22] Javier Perez, D. et al.: How Real (Time) Are Virtual PLCs?: 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, S. 1–8, 2022.

[Ji20] Jianyong Wu: Enable ptp_kvm for arm64. https://lwn.net/Articles/830643/.

[On18] Ong Boon Leong, Anil Kumar, Usman Sarwar: A comparative analysis of Precision Time Protocol in native, virtual machines and container-based environments for consolidating automotive workloads, 2018.

[Ph25] Phoenix Contact GmbH & Co. KG: Virtual Control. https://www.phoenixcontact.com/de-de/produkte/sps-steuerungen-und-i-o/virtual-control, Stand: 01.10.2025.

[Ra24] Raghav Kadian: Virtual PLCs: Can they become the industry norm by 2030? https://iot-analytics.com/virtual-plcs/, Stand: 25.02.2025.

[Si24] Siemens AG: Safety First: Jetzt auch im virtuellen Umfeld. https://blog.siemens.com/de/2024/12/safety-first-jetzt-auch-in-der-virtuellen-umgebung/?stc=wwdi134614, Stand: 01.10.2025.

[Si25] Siemens AG: SIMATIC S7-1500V. https://www.siemens.com/de/de/produkte/automatisierung/systeme/industrie/sps/simatic-s7-1500/virtual-plc.html, Stand: 01.10.2025.

[So21] Sollfrank, M. et al.: Evaluating Docker for Lightweight Virtualization of Distributed and Time-Sensitive Applications in Industrial Automation. IEEE Transactions on Industrial Informatics 5/17, S. 3566–3576, 2021.

[Sp24] Specht, F. et al.: VirtuBench: Leistungsbewertung für virtuelle Industriesteuerungen. In (Technische Hochschule Ostwestfalen-Lippe; Jasperneite, J.; Jumar, U. Hrsg.): Kommunikation in der Automation Beiträge des Jahreskolloquiums KommA 2024. Institut für industrielle Informationstechnik - inIT der Technischen Hochschule Ostwestfalen-Lippe, S. 171–248, 2024.

[St24] Stanton, K. B. et al.: Precision Time in the Last Centimeters for Distributed Applications: 2024 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS). IEEE, S. 1–6, 2024.

[TS25] TSN Systems GmbH: Fortschrittliche TSN-Lösungen für Ihre industrielle Vernetzung. https://tsn.systems/produkte/industrial/.

http://dx.doi.org/10.25673/121097

Nutzung von Small-Scale Kommunikationsprotokollen für ressourcenbeschränkte Geräte mit Single-Pair Ethernet Anbindung

Mike Neelen[1], Alexander Biendarra[1], und Arndt Schübel[2]

Abstract:

Dieses Paper behandelt einen neuen Ansatz, wie in industriellen Anlagen Sensoren und Aktoren mit einer Steuerung verbunden werden können. Hierbei wird das aus dem Automotive Bereich kommende und auf Google Protobuf basierte Remote Control Protocoll (RCP) verwendet. Der Ansatz basiert auf der Virtualisierung von speicherprogrammierbaren Steuerungen sowie die dazu gehörigen IOs und dem Transport der IO Daten über das RCP. Dieses RCP kann auf Schicht 2 oder 3 des ISO/OSI Modells verwendet werden und arbeitet nach dem Client-Server-Prinzip. Es beinhaltet einen Scheduler, um Tasks abzuarbeiten und unterstützt verschiedene IOs wie beispielsweise GPIO, UART und I^2C. Das RCP kann in Kombination mit der SPE-Technologie 10BASE-T1S verwendet werden, um mehrere Sensoren bzw. Aktoren in einer Automatisierungsanlage mit Ethernet auszustatten und diese via half-duplex Verbindung an einem Multidrop Bus anzuschließen, was eine durchgängige Ethernet-basierte Kommunikation ermöglicht.
In diesem Paper wird eine Beispielsarchitektur vorgeschlagen, welche aus zwei RCP-Server und einem Client besteht.

Keywords: SPE, Small-Scale Kommunikationsprotokoll, Industrielle Kommunikation, Sensorkommunikation, Ressourcenbeschränkte Geräte

1 Einleitung

Die Integration von Single Pair Ethernet (SPE) in ressourcenbeschränkte Geräte findet immer mehr Anklang in verschiedensten Industrien. SPE hat den Vorteil, dass vorher genutzte Steckerbauformen wie beispielsweise M12 oder M8, verwendet werden können und die Spannungsversorgung auf denselben Aderpaar übertragen werden kann. Die SPE-Integration ermöglicht also eine direkte Ethernet-Kommunikation mit den Geräten bei gleichbleibender Bauform. [Mo21]
Dieser Ansatz hat den Vorteil, dass das gesamte Produktionsnetz von Aktoren über speicherprogrammierbare Steuerungen (SPS) bis zum Sensor direkt über Ethernet kommuniziert, was neue Möglichkeiten für den Datenaustausch und die Diagnose der Geräte bietet.[Se25]

1 Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Germany, mike.neelen@iosb-ina.fraunhofer.de; alexander.biendarra@iosb-ina.fraunhofer.de
2 onsemi, Einsteinring 28, 85609 Aschheim-Dornach, Germany, Arndt.Schuebel@onsemi.com

1.1 Motivation

In Kombination mit virtualisierten Steuerungen und IO-Endgeräten bietet die Integration von SPE in Endgeräten einen neuen Ansatz der Kommunikation und Diagnose in der Automatisierungstechnik, welcher in diesem Papier behandelt werden soll. Dazu wird ein Ethernet-basiertes Kommunikationsprotokoll benötigt, das einen geringen Ressourcenfußabdruck besitzt, womit der Datenaustausch zwischen Sensoren und Steuerungen realisiert werden kann [Sc20]. Eine aus dem Automotive-Bereich kommende Möglichkeit ist das Remote Control Protocol (RCP), welches in der Open Alliance vorgestellt wurde [Gu24]. Abbildung 1 zeigt das zu verwendende Konzept mit den IOs in einem Multidrop SPE Netzwerk. Hierbei wird ein Modul an die IOs angeflanscht, welches SPE-fähig ist und ein RCP Stack implementiert hat, womit die IOs Daten remote austauschen können.

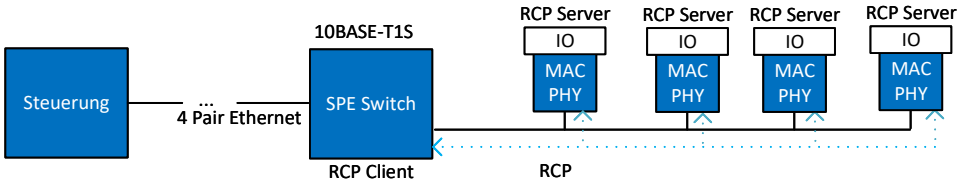


Abb. 1: Netzwerkarchitektur Konzept im Multidrop Verfahren

Aufgrund der Neuheit des RCPs sind aktuell nur Muster für Chips verfügbar, welche das RCP unterstützen. In diesem Paper wird der Ansatz daher analytisch diskutiert.

1.2 Gliederung des Papers

Das Paper startet mit dem Stand der Technik im Bereich von Single Pair Ethernet und Small-Scale Kommunikationsprotokollen. Anschließend erfolgt eine Beschreibung der Funktionsweise des eingesetzten *Remote Control Protocolls*, der heterogenen Kommunikationsarchitektur und des Netzwerklayouts. Abschließend wird ein Ausblick gegeben, welche Schritte für die Nutzung von Small-Scale Kommunikationsprotokollen auf Sensoren mit einer SPE-Schnittstelle notwendig sind.

2 Stand der Technik

Dieses Kapitel beschreibt die in dem Paper verwendeten Technologien. Hierbei wird kurz erklärt, welche Übertragungstechnologien in dem theoretischen Aufbau zu finden sind und wie das RCP aufgebaut ist. Aus umfangslimitierenden Gründen werden nicht alle Technologien betrachtet.

2.1 Single Pair Ethernet

Single Pair Ethernet (SPE) hat das Ziel, konventionelles Ethernet mit vier Aderpaaren auf ein Aderpaar zu reduzieren. SPE befindet sich in dem ISO/OSI Schichtenmodell auf Schicht 1 (Physikalische Schicht). Hierbei gibt es verschiedene Standards des Institute of Electrical and Electronics Engineers (IEEE), welche die verschiedenen Transmissionstechnologien beschreiben. SPE kann in halb-duplex oder full-duplex mit einer Bandbreite von 10 MBit/s bis 25 GBit/s genutzt werden. Jede dieser Transmissionstechnologien hat einen eigenen Anwendungsfall: Von 10 Mbit/s für Sensorik in der Prozessindustrie über längere Strecken bis zum Multimedia-System eines Autos mit Audio und Videoübertragung. [Se25]

2.1.1 IEEE 802.1cg 10BASE-T1S

Eine der SPE-Übertragungsstechnologien, beschrieben in IEEE 802.3cg, ist 10BASE-T1S [IE20a]. Mit dieser SPE-Technologie ist es möglich, halb-duplex mit mehreren Teilnehmern an einem geteilten Medium, namentlich *Multidrop*, zu kommunizieren (vgl. Abbildung 1). Das full-duplex-Verfahren von 10BASE-T1S wird in diesem Paper nicht betrachtet, da dieses in der Praxis meistens durch 10BASE-T1L ersetzt wird. Die maximale Anzahl der Teilnehmer ist laut IEEE-Standard bis zu 8 Knoten, welche mit einer Gesamtkabellänge von 25 Metern verbunden sind. Abbildung 2 zeigt den Aufbau der Technologie.

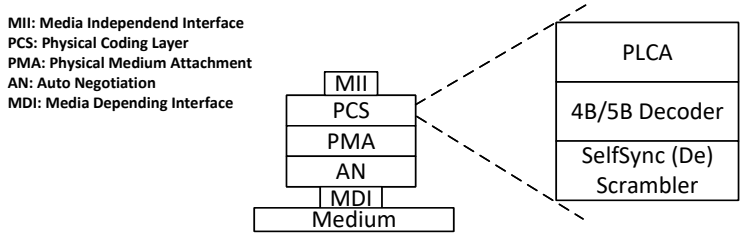


Abb. 2: 10BASE-T1S PHY Mechanismen

Die Codierung im PCS ist eine 4B/5B-Codierung, was zusätzliche Steuersignale ermöglicht. Um eine Kollision der Teilnehmersignale zu verhindern, wird das *Physical Layer Collision Avoidance* (PLCA) Verfahren angewendet. Hierbei wird jedem Teilnehmer eine eindeutige PLCA-ID-Nummer zugewiesen. Ein Teilnehmer darf dann nur senden, wenn seine PLCA-ID mit der gegenwärtigen Transmit Opportunity-Nummer übereinstimmt. Um den Mechanismus zu starten wird ein Beacon Signal vom PHY des Koordinators verschickt. Danach wird ein Zähler (*CurrentID*) mit der eigenen ID verglichen, woraufhin der Node mit der dementsprechenden ID eine Möglichkeit hat, ein *Commit* Signal zu verschicken und ein Frame wird versendet. Das Gerät mit der jeweiligen ID hat dann die Möglichkeit, innerhalb einer TO (transmit opportunity) einen oder mehrere (Burst) Frames zu verschicken. Nachdem die Frames verschickt wurden oder eine TO-Zeit überschritten ist, wird der NodeID Counter

wieder erhöht und geprüft, ob dieser über der angegebenen Maximalzahl an Nodes liegt. Ist dies der Fall, wird der Zähler zurückgesetzt und der Kontroller startet wieder von vorne mit dem *Beacon*-Signal. Damit dieses Prinzip funktioniert, werden die Netzwerkteilnehmer vorher dementsprechend konfiguriert. D.h. jeder Teilnehmer bekommt eine *Node ID*, die *MaxNodeID* und die *TO Zeit* übermittelt. Abbildung 3 zeigt den PLCA-Mechanismus für n+1 Geräte. Hierbei repräsentieren die grünen Kästen ein versendetes Frame der jeweiligen Node und das orangene Kästchen mit dem B ein Beacon Signal.

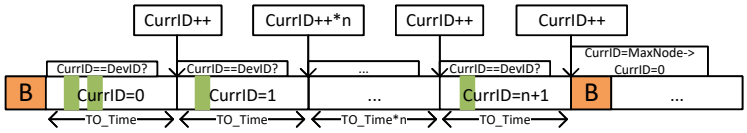


Abb. 3: PLCA Mechanismus

Die Steuersignale für das PLCA-Verfahren werden im PHY über der 4B/5B De/Encoder Ebene generiert. In Kombination mit dem selbst-synchronisierenden Scrambler entstehen außerdem genug Imparität, um im physikalischen Signal einen Gleichanteil zu verringern und eine Taktrückgewinnung zu ermöglichen. Bei der Übertragung der Daten auf der Leitung wird das Differential Manchester Encoding (DME) verwendet. Die Autonegotiation wird in diesem Paper nicht behandelt.[IE20a]

2.2 IEEE 802.1AS generalized Precision Time Protokoll

Das generalized Precision Time Protocol (gPTP) ist standardisiert in IEEE 802.1AS-2020. Im Vergleich zum Network Time Protokoll (NTP) ist hier die Synchronisation der Teilnehmer genauer möglich, da die Signallaufzeit in der Kommunikationsstrecke mit einberechnet wird und die Zeitstempel in Hardware generiert werden sobald das Frame den PHY verlässt. Für die Berechnung der Signallaufzeit werden zwischen den Teilnehmern Nachrichten ausgetauscht, welche Zeitstempel beinhalten. [IE20b]
Eine für das RCP und 10BASE-T1S im Multidrop-Modus relevante Änderung ist IEEE 802.1ASds, welches die Kompatibilität von gPTP mit halb-duplex Übertragungsverfahren beinhaltet. IEEE 802.1ASds ändert die Art, wie Delay-Request Nachrichten behandelt werden. In einem Multidrop-Netzwerk können Delay-Request Nachrichten nicht wie bei gPTP üblich per Multicast versendet werden. Dies hat den Hintergrund, dass bei Multicast-Nachrichten mehrere Teilnehmer auf den Delay-Request antworten und somit die Signallaufzeiten nicht eindeutig bestimmbar sind. Aus diesem Grund erlaubt ASds das Versenden von Delay-Request und -Response Nachrichten mittels Unicast-Adressen. [Hu21]
Auf kurzen Leitungslängen, welche bei 10BASE-T1S üblich sind, kann jedoch auch auf die Delay Request / Response Methode verzichtet werden und eine Abschätzung, basierend auf der halben zur Verfügung stehenden Leitungslänge, herangezogen werden.

2.3 IO-Link

Ein verbreitetes Protokoll, welches einen ähnlichen Anwendungsfalle in industriellen Automatisierungsanlagen einnimmt, ist das IO-Link-Protokoll. Das unter IEC 61131-9 standardisierte IO-Protokoll besteht aus einem Point-to-Point-Master-Slave-Prinzip. Bei IO-Link wird ein M12 Stecker mit 3 bzw. 4 Polen verwendet. und die Übertragungsgeschwindigkeit variiert zwischen 4,8 kbit/s und 230,4 kBit/s. Es werden Daten und Energie von dem IO-Link Master an die IO-Link Geräte verteilt. [DI22]

Im Vergleich zu IO-Link ist das RCP direkt per Ethernet mit anderen Teilnehmern verbunden und hat eine Mindestübertragungsgeschwindigkeit von 10 Mbit/s aufgrund von 10BASE-T1S.

3 Remote Control Protocol (RCP)

Das Remote Control Protocol wurde 2024 in der Open Alliance im Tech Committee TC18 vorgestellt [PBB24]. Es basiert auf Google Protokollbuffern (Protobuf) und ist auf Layer 2 bzw. Layer 3 des ISO/OSI Modells umsetzbar.
Mit dem RCP können verschiedene Standard IOs angesprochen werden. Dies beinhaltet GPIOs, SPI, UART, aber auch die Konfiguration der Netzwerktechnik wie das PLCA oder die Synchronisation sind möglich. Das Protokoll funktioniert nach dem Client-Server Prinzip. Abbildung 4 zeigt den Aufbau des RCP-Servers. Dieser dient als Hardwareschnittstelle zwischen dem RCP und der anzusprechenden Hardware.

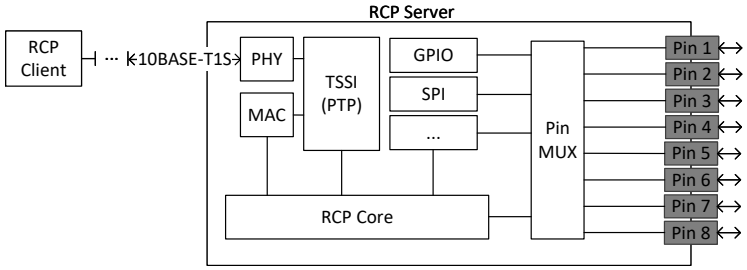


Abb. 4: RCP-Server Blockschaltdiagram

Jeder Pin auf der rechten Seite des Blockschaltdiagramms kann eine eigene Schnittstelle, also beispielsweise ein GPIO oder eine UART, darstellen. Der RCP-Core beinhaltet einen Scheduler, welcher die verschiedenen Aktionen plant. Es wird darauf hingewiesen, dass das volle Protokoll nicht detalliert innerhalb dieses Papers beschrieben werden kann und dementsprechend eine Umfang angemessene Detailtiefe gewählt wurde. Das RCP besteht im Wesentlichen aus Direktiven und Reports.

3.1 Direktiven im RCP

Jede Direktive hat eine eindeutige Sequenz-ID, welche nachher am RCP-Device zur Zuordnung aber auch für das Scheduling genutzt wird. Die Direktiven im RCP unterteilen sich in verschiedene Arten: Konfigurationen, Aktionen, Abfragen und der Stornierung einer Aktion. Jedes Frame kann mehrere Direktiven einer Art beinhalten.

Konfigurationen
An einem RCP-fähigen Gerät können verschiedene IO-Module angeschlossen werden. Die jeweiligen Geräte können durch einen Konfigurationsbefehl konfiguriert werden. Dies beinhaltet die Netzwerkkonfiguration oder die Konfiguration des IO-Moduls selbst. In beiden Fällen wird außerdem eine Handle-ID, ein Freischaltsignal und die Geschwindigkeit in Hertz mitgeteilt. In diesem Paper werden die Optionen für GPIOs dargestellt. Weitere Schnittstellen, wie beispielsweise, *SPI* und *UART* sind in Planung/möglich.

Aktionen
Nach der Konfiguration können Aktionen an die jeweiligen Teilnehmer gesendet werden. Diese unterscheiden sich wieder im jeweiligen IO-Modul des Gerätes. Jede Aktion beinhaltet einen Header, in dem der Ausführzeitpunkt, die Periodizität, die Anzahl der Wiederholungen und ob der Befehl quittiert oder unquittiert durchgeführt werden soll, steht.

Abfragen
Neben den Aktionen ist es auch möglich, den Status des jeweiligen Moduls abzufragen. Dafür sind im Protokoll mehrere Abfragen pro Frame möglich. Eine Abfrage beinhaltet beispielsweise die aktuelle Konfiguration des Gerätes, Geräte ID oder aber den Netzwerkstatus. Auf eine Abfrage folgt immer eine Antwort.

Stornierung
Da das RCP-Model mit einem Scheduler arbeitet, welcher verschiedene Aktionen (einmalig oder zyklisch) einplant, ist es auch möglich, falsch geplante Aktionen bzw. nicht mehr benötigte Aktionen zu stornieren. Die hierfür verfügbare Option beinhaltet das Stornieren eines Befehls auf Basis der Action ID, welche einzigartig für jede Aktion ist und beim Versenden des Befehls im Frame enthalten ist.

3.2 Reports im RCP

Während die Direktiven im RCP vom Client zum Server geschickt werden, sind die Reports vom Server an den Client gerichtet. Hierbei beinhalten sie neben der Sequenznummer die Resultate einer Konfiguration, Aktion oder Query. Alternativ können auch Benachrichtigung

geschickt werden. Des Weiteren sind herstellerspezifische Reports möglich, worauf hier jedoch nicht weiter eingegangen wird. Jedes Frame kann mehrere Reports einer Art beinhalten.

Konfigurationsresultat
Wenn ein Gerät im RCP-Netzwerk konfiguriert wird, wird dementsprechend das Resultat eben jener Konfiguration vom RCP-Server verschickt, solange nicht die Option "*quiet*"verwendet wird. Hierbei wird wieder je nach Konfiguration das Ergebnis mitgeteilt. Beispielsweise ob jene Konfiguration erfolgreich war oder ggf. das zu konfigurierende Gerät nicht erkannt wird.

Resultate einer Aktion
Aktionen, welche nicht mit der *"quiet"* Option abgeschickt werden, resultieren in einem Report, welcher vom RCP-Server an den Client verschickt wird. Ein Aktionsresultat beinhaltet einen Fehlercode, eine optionale Fehlerbeschreibung, die ID der Aktion sowie einen Rückgabewert, welcher je nach Aktion unterschiedlich ist.

Resultate einer Abfragen
Bei einer Abfrage wird immer ein Resultat vom RCP-Server erwartet. Am Anfang der Nachricht befindet sich zunächst der Status der Abfrage, worauf das Resultat der Anfrage selbst folgt. Hierbei werden zusätzliche Informationen vom RCP-Server verschickt, welche abhängig von der Anfrage sind.

Benachrichtigung
Eine Benachrichtigung wird vom RCP-Server verschickt, wenn dieser beispielsweise eine Änderung des Status eines GPIO's feststellt. Dieses Ereignis wird dann mit einem Handle, einer Sequenznummer, einem Zeitstempel und den Daten an den RCP-Client verschickt. Diese Benachrichtigung kann beispielsweise einen Pegelwechsel in einem Sensor beinhalten.

4 Ansatz und theoretische Umsetzung

Das RCP kann als neuer Ansatz für die Automatisierungsindustrie verwendet werden. Hierbei ist die Idee, speicherprogrammierbare Steuerungen und die dazugehörigen Input-/Output-Module virtuell auszuführen und das Ausgangssignal der vIOs über ein RCP an ein reales Gerät zu kommunizieren. Dadurch kann bereits bestehende Software in neuer Hardware, welche beispielsweise zentral in einem Unternehmen auf einem Server sitzen kann, wiederverwendet werden.

4.1 Annahmen zur Kommunikation zwsichen Client und Server

Für die Ethernet-Kommunikation mit ressourcenbeschränkten Geräten werden verschiedene Annahmen getroffen:
Aufgrund des geringen Datendurchsatzes von Sensoren und deren verwendeten Netzwerk-Topologie in Produktionsanlagen wird die SPE-Technologie 10BASE-T1S in der Multidrop-Topologie eingesetzt. Für die Übertragung zwischen dem Sensor und der virtualisierten Steuerung bzw. dem virtualisierten IO-Endgerät wird ein heterogenes Netzwerk verwendet, welches verschiedene Datenübertragungsraten (10 Mbit/s - 10 Gbit/s) sowie verschiedene Übertragungsmedien (Single Pair Ethernet, Standard Ethernet, Lichtwellenleiter, ggf. eine 5G-Funktstrecken) beinhaltet.
Um Anwendungsfälle wie beispielsweise die Nachverfolgung von Ereignissen (Sequence-of-Events, SoE) bedienen zu können oder aber die Kommunikationssteuerung und den RCP-Scheduler zu ermöglichen, werden die Kommunikationsteilnehmer mithilfe von einem PTP-Profil synchronisiert. Für die Kommunikation zwischen dem Sensor und einem virtualisierten IO-Endgerät soll ein Small-Scale-Kommunikationsprotokoll eingesetzt werden, welches in Abschnitt 3 beschrieben wurde.
Die Kommunikation zwischen virtueller SPS und virtueller IO wird nicht betrachtet. Die Security des Netzwerkes wird nicht betrachtet, ist jedoch in solch einem Netzwerk in der Realität zwingend notwendig. Dies hat den Hintergrund, das eine Ethernet Kommunikation über einen Funk Kanal und ggf. das Internet durch externe Einflüsse manipuliert werden kann.
Es wird außerdem angenommen, dass der RCP-Client, welcher als Umsetzer vom vPLCs Kommunikationsprotokoll zum RCP dient, das RCP mit Standard Ethernet (1000BASE-T) als Übertragungsmedium unterstützt.

4.2 Theoretischer Einsatz in einem Beispielnetzwerk

Das RCP soll es ermöglichen, unabhängig von der in der Produktion verwendeten speicherprogrammierbaren Steuerung und dem Kommunikationsprotokoll die IO-Daten zu übermitteln.

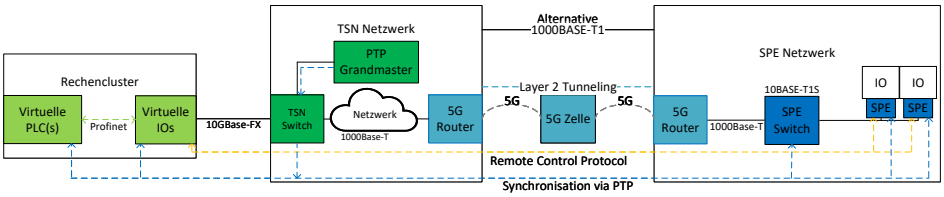


Abb. 5: Einsatz des RCP in einem heterogenen Kommunikationsnetzwerk

Abbildung 5 zeigt eine Beispielarchitektur mit zwei RCP-Servern, welche mit einer Steuerung als RCP-Client verbunden sind. Die GPIOs werden in dem Beispiel als Eingänge betrachtet und mit dem RCP-Client durch eine Konfigurationsnachricht eingestellt. Weitere

Konfigurationen, welche mit dem RCP geschehen, sind die Konfiguration vom PLCA, der Synchronisation über PTP, sowie die Konfiguration des Pins selbst. Bei einer Änderung des Signals kann dann eine Notification an den Client geschickt werden, welcher auf einem Pegelwechsel bei den Sensoren hinweist. Diese Notification ist mit einem Zeitstempel versehen; d.h. eine Auswertung von zeitlichen Differenzen könnte hier auch außerhalb eines SPS-Programms im RCP-Client erfolgen. Es wird ein Umsetzer auf dem RCP-Client benutzt, welcher dafür benötigt wird, das RCP in die jeweilige IO Information für die virtuelle IO umzuwandeln, die dann wiederum das Ergebnis an die virtuelle SPS übermittelt. Alternativ kann angedacht werden, dass der RCP-Teil sich rein im SPE-Teil des Netzwerkes befindet und im SPE Switch eine Implementierung des Profinet zu RCP Umsetzers befindet.

5 Zusammenfassung und Ausblick

Dieses Paper beschreibt den Einsatz eines neuen Protokolls aus dem Automotive Bereich, welches für die Virtualisierung von Steuerungen und IOs einsetzbar ist, um IOs anzusteuern und abzufragen. Dabei wurde auf den Aufbau des Protokolls eingegangen und ein Anwendungsfall des RCPs in der Industrie theoretisch beschrieben. Tabelle 1 zeigt die Vor- und Nachteile für den neuen Kommunikatiosansatz.

Tab. 1: Vor- und Nachteile des Ansatzes

Vorteile	Nachteile
- Höhere Datenraten möglich (Im Vergleich zu IO-Link) - Volle Ethernet Kommunikation bis zum Sensor möglich - Etabliertes Medium (Ethernet) - Geringere Verdrahtungsaufwand im Vergleich zu Buskopplern - Steuerungshersteller unabhängige Lösung	- Rekonfigurationen im Multidrop aufwendig - Security Konzept und Mechanismen müssen dementsprechend bedacht werden. - Höhere Kosten pro Sensor durch zusätzliche Hard-und Software

Das RCP ist unabhängig von den jeweiligen Steuerungsprotokollen und wäre dementsprechend für Protokolle wie CC-Link, Profinet oder Ethercat mittels eines Umsetzers verwendbar.
Es ist angedacht, den theoretischen Ansatz, welcher in diesem Paper beschrieben wurde praktisch zu verfolgen und eine dementsprechende Umsetzung unter Realbedingungen zu evaluieren.

Literaturverzeichnis

[DI22] DIN/EN: Speicherprogrammierbare Steuerungen - Teil 9: Schnittstelle für die Kommunikation mit kleinen Sensoren und Aktoren über eine Punkt-zu-Punkt-Verbindung (IEC 61131-9:2022); Deutsche Fassung EN IEC 61131-9:2022. S. 1–82, 2022.

[Gu24] Gumersindo Veloso, A. N., 2024, URL: https://opensig.org/tech-committee/tc18-remote-control-protocol/.

[Hu21] Huh, W.: 802.1AS use cases in 10B-T1S and improvement. 2021.

[IE20a] IEEE: Standard for Ethernet - Amendment 5: Physical Layer Specifications and Management Parameters for 10 Mb/s Operation and Associated Power Delivery over a Single Balanced Pair of Conductors. Std 802.3cg-2019 (Amendment to IEEE Std 802.3-2018), S. 1–256, 2020, DOI: 10.1109/IEEESTD.2020.8982251.

[IE20b] IEEE: Standard for Local and Metropolitan Area Networks–Timing and Synchronization for Time-Sensitive Applications. Std 802.1AS-2020 (Revision of IEEE Std 802.1AS-2011), S. 1–421, 2020, DOI: 10.1109/IEEESTD.2020.9121845.

[Mo21] Moritz, F.: Benefits of SPE for sensors in automation, 2021, URL: https://singlepairethernet.com/wp-content/uploads/2021/07/SPE-Benefits-for-Sensors-in-automation9.pdf.

[PBB24] Propp, A.; Beruto, P.; Brtan, F.: RCP Baseline Proposal. S. 1–49, 2024.

[Sc20] Schriegel, S.; Biendarra, A.; Kobzan, T.; Leurs, L.; Jasperneite, J.: Ethernet TSN Nano Profil - Migrationshelfer vom industriellen Brownfield zum Ethernet TSN-basierten IIoT, de, 2020, DOI: 10.1007/978-3-662-59895-5_13.

[Se25] Seereiner, S.; Lohhöfener, A.; Muyshondt, H.; Moritz, F.; Luskind, Y.; Bilz, F.; Kindermann, T.; Radau, M.; Perrot, A.; Graf, S.; et al.: Single pair ethernet (SPE) system architecture, 2025.

http://dx.doi.org/10.25673/121097

Umsetzung des BSI-Bausteins OPS.bd.1.31 Precision Time Protocol für die Sektorenkopplung

Alexander Biendarra[1]

Abstract: Die Sektorenkopplung wird als Herausforderung und wirkungsvolles Hilfsmittel zur Erreichung der Energiewende und Klimaneutralität angesehen. Ein Problem, welches bei der Nutzung von erneuerbaren Energien auftritt, ist die volatile Erzeugung. Um eine schnelle Reaktionszeit und einen präzisen Abgleich von vorhandener und benötigter Energie zu erreichen, ist unter anderem die Messung des Energienetzzustandes mit einer hochgenauen Zeitsynchronisation innerhalb der Energieerzeugungsnetze wie auch in den Bereichen der Energienutzung notwendig. Im Bereich der industriellen Automation, in dessen Bereich auch Anlagen der Energieerzeugung fallen, soll in zukünftigen Ethernet-Time-Sensitive Networking (TSN) Netzwerken die Zeitsynchronisation nach IEEE 802.1AS-2020 zum Einsatz kommen. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) stellt mit dem Baustein „OPS.bd.1.31 Precsion Time Protocol“ eine Auflistung von Absicherungen zur Verfügung, wie die Zeitsynchronisation gegen Gefährdungslagen abgesichert werden kann. Im Rahmen einer Fallstudie wird in diesem Beitrag die Umsetzbarkeit des BSI-Bausteins für das Zeitsynchronisationsprotokoll IEEE 802.1AS-2020 erarbeitet. Während der Großteil der Anforderungen erfüllt werden kann, werden unter anderem im Bereich der Authentifizierung, Authorisierung noch offene Punkte identifiziert.

Keywords: BSI, Energienetzautomation, Industrielle Automatisierungstechnik, OPS.bd.1.31, Precision Time Protocol (PTP), Sektorenkopplung, Time-Sensitive Networking (TSN), Zeitsynchronisation

1 Motivation

Die Sektorenkopplung wird als Herausforderung und wirkungsvolles Hilfsmittel zur Erreichung der Energiewende und Klimaneutralität angesehen. Bis dato getrennte Netze für Strom, Gas und Wärme sollen direkt mit den Verbrauchsbereichen Haushalt, Gewerbe, Industrie und Mobilität verknüpft werden, um die durch insbesondere Wind und Sonne gewonnene Energie effizient nutzen zu können [Ga24]. Ein Problem, welches bei der Nutzung von erneuerbaren Energien auftritt, ist die volatile Erzeugung. Betreiber von Energieerzeugungsanlagen müssen sicherstellen, dass die eingespeiste Energie das Energieübertragungsnetz nicht durch Leistungsspitzen überlastet. Um dies zu gewährleisten, muss ein dauerhafter, hochpräziser Abgleich zwischen dem Energiebedarf bzw. der Netzkapazität und der Erzeugung durchgeführt werden [De24]. Dies erfordert die Vernetzung der energieerzeugenden Sektoren (Windenergie, Photovoltaik, Wasser- bzw. Gezeitenkraftwerke, Erdwärme) mit den energieverbrauchenden (Industrie, Haushalt, Elektromobilität, Nah- und Fernverkehr). Dies resultiert in einer OT-OT-Kopplung.

[1] Fraunhofer IOSB-INA, Digitale Infrastrukturen, Campusallee 1, 32657 Lemgo, Deutschland, alexander.biendarra@iosb-ina.fraunhofer.de

Um eine schnelle Reaktionszeit und einen präzisen Abgleich vorhandener und benötigter Energie zu erreichen, ist unter anderem die Messung des Energienetzzustandes mit einer hochgenauen Zeitsynchronisation innerhalb der Energieerzeugungsnetze wie auch in den Bereichen der Energienutzung notwendig. Hier werden Genauigkeiten von ±1 µs gefordert, um die Netzeinspeisung zu steuern [De24]. In Kanada, wo bereits ein höherer Ausbaugrad im Bereich der erneuerbaren Energien im Vergleich zu Deutschland besteht [De19], wird für die Erfassung von Ereignissen wie Fehleinspeisungen, eine Genauigkeit von ± 100µs gefordert [Al25]. Diese wird durch das im Standard IEEE Std. 1588-2019 definierte Precision Time Protocol in der Version 2 (PTPv2) [IE19] umgesetzt. Eingesetzt wird es sowohl in zukünftigen TSN-Netzwerken für die industrielle Automatisierungstechnik als auch in der Energienetzautomation, die ebenfalls in den Energieerzeugungsanlagen durch TSN-Netzwerke umgesetzt sein können (OT-OT-Kopplung). Je nach Anwendungsbereich kommt eine bestimmte Ausprägung von PTPv2 in Form eines definierten Profils zum Einsatz. Gleichzeitig bietet es aber auch ein großes Potenzial für Angriffsszenarien. Dadurch, dass die Energieerzeugung zur kritischen Infrastruktur (KRITIS) zählt, muss diese zum Schutz der Bevölkerung besonders geschützt werden. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) stellt mit dem Baustein „OPS.bd.1.31 Precision Time Protocol“ [Bu22] eine Auflistung von Absicherungen zur Verfügung, wie die Zeitsynchronisation in Gefährdungsanlagen abgesichert werden kann. Dadurch, dass zukünftige Ethernet TSN-Netzwerke das Zeitsynchronisations-Profil gPTP (IEEE 802.1AS-2022 [IE20]) nutzen, muss sichergestellt werden, dass die Anforderungen auch mit diesem Profil erfüllt werden können.

Der Beitrag behandelt im Kapitel 2 den Stand der Technik im Bereich der Sektorenkopplung, der relevanten Zeitsynchronisationsprotokolle, der IEC/IEEE 60802 und des BSI-Grundschutzkompendiums. Kapitel 3 widmet sich dem Thema des BSI-Bausteins OPS.bd.1.31 Precision Time Protocol. In Kapitel 4 wird in einer Fallstudie die Umsetzbarkeit des BSI-Bausteins validiert. Kapitel 5 schließt den Beitrag mit dem Fazit und einem Ausblick ab.

2 Stand der Technik

Im Folgenden wird der Stand der Technik im Bereich der Sektorenkopplung, von Protokollen für die Zeitsynchronisation, des Profil-Standards IEC/IEEE 60802 und der BSI-Bausteine dargelegt.

2.1 Sektorenkopplung

Sektorenkopplung bezeichnet die gezielte Vernetzung der Energiesektoren Strom, Gas und Wärme mit den Verbrauchsbereichen Haushalt, Gewerbe, Industrie und Mobilität, um so erneuerbare Energien in mehreren Bereichen zu nutzen und Treibhausgasemissionen

nachhaltig zu senken. Die Ziele, die mit der Sektorenkopplung erreicht werden sollen, sind die Dekarbonisierung, die Flexibilisierung und effizientere Nutzung des Energiesystems und eine bessere Integration von erneuerbaren Energien. [Ga24]

2.2 Zeitsynchronisationsprotokolle

Um eine zuverlässige Interaktion zwischen den an der Sektorenkopplung beteiligten Bereichen der Energieerzeugung und der Energienutzung zu realisieren, ist es zum einen notwendig, schnelle Reaktionszeiten zu ermöglichen und zum anderen präzise Messwerte von Fehlern und Phasenlagen zu erkennen und das Versorgungsnetz stabil zu steuern. Sollten Ereignisse in Form von Unterversorgungen oder Ausfällen im Energienetz eintreten, müssen diese zu ihrem Entstehungsort zurückverfolgt werden können. Um dies zu realisieren, wird eine hochgenaue Zeitsynchronisation benötigt. Für diese Art von verteilten Systemen, die mit einer hochgenauen Zeit synchronisiert werden müssen, wird die Zeitsynchronisation nach IEEE 1588 PTPv2 [IE19] eingesetzt. PTPv2 ermöglicht es, die Zeit sowohl in Schicht 2- als auch in Schicht 3-Kommunikationsnetzwerken zu verteilen. Das Prinzip der Verteilung der Zeit beruht darauf, dass ein Zeitgeber (Time Transmitter) die Zeit an die Teilnehmer im Netzwerk (Time Receiver) verteilt. Netzwerkinfrastrukturkomponenten verbreiten die Zeit im Netzwerk, wobei PTPv2 selbst keine Anforderungen daran stellt, ob Verzögerungen wie interne Weiterleitungszeiten kompensiert werden oder nicht.
Um die Vielzahl an Parametrierungsmöglichkeiten im Bereich von PTPv2 zu vereinheitlichen und die Interoperabilität der Kommunikationsteilnehmer, die die synchronisierte Zeit benötigen, zu verbessern, wurden Profile eingeführt. Diese Profile werden in Anwendungsbereichen wie dem Telekommunikationsbereich, im Bereich von Energienetzen oder aber der industriellen Kommunikationstechnik eingesetzt.
Das für die industrielle Automation und TSN relevante Profil von PTPv2 ist das generalized Precision Time Protocol (gPTP), das im Standard IEEE 802.1AS-2020 [IE20] definiert ist. gPTP nutzt für die Synchronisation die Schicht 2 Multicast-Kommunikation. Für die Kompensation der Laufzeit der Nachrichten auf dem Medium wird der PDelay-Mechanismus genutzt. Weiterhin definiert das Protokoll Sendeintervalle für die verschiedenen Nachrichten und schließt die Nutzung von nicht gPTP-fähigen Geräten im Synchronisationspfad des Netzwerks aus.
Für die Synchronisation der Zeit im Bereich der Energienetzautomation kommen mehrere Profile zum Einsatz. Neben dem Telecom-Profile kommen das IEC61850-9-3 Utility Profile [IE16] und das IEEE C37.238 Power Profile [IE17] zum Einsatz. Während das Telcom-Profile zumeist auf die Verteilung der Zeit über das Zeitformat IRIG-B und somit auf eine zur Ethernet-basierten Kommunikation parallele Verkabelung setzt, wird für das Utility Profile und das Power Profile die Schicht 2 Ethernet-Kommunikation mit dem PDelay-Mechanismus zur Lautzeitkompensation eingesetzt.[PT19] Damit sind die beiden zuletzt genannten Profile sehr ähnlich zu dem in der industriellen Kommunikationstechnik eingesetzten Variante mit dem Standard IEEE 802.1AS-2020. [IE19][IE20][IE16][IE17]

2.3 IEC/IEEE 60802 TSN Profile for Industrial Automation

Der Profilstandard IEC/IEEE 60802 [IE25] verfolgt das Ziel, eine einheitliche Nutzung von TSN in der industriellen Automation sicherzustellen. Der unter dem Titel IEC/IEEE 60802 Time-Sensitive Networking Profile for Industrial Automation sich in der laufenden Standardisierung befindliche Profilstandard soll zum Ende des Jahres 2025 verabschiedet werden. Innerhalb der IEC/IEEE 60802 werden neben den eigentlichen TSN-Standards und deren Parametrierung ein Architekturmodell für TSN-Geräte, zwei Konformitätsklassen, eine Möglichkeit der Konfiguration, Vorgaben für den Bereich der Security und Leistungsanforderungen der daraus resultierenden Kommunikationsteilnehmer definiert. Teil davon ist auch die Nutzung der Zeitsynchronisation mit IEEE 802.1AS-2020 im Bereich von IT/OT-Kommunikationsnetzen.

2.4 Das BSI IT-Grundschutz-Kompendium

Das IT-Grundschutz-Kompendium des Bundesamtes für Sicherheit in der Informationstechnik (BSI) behandelt das Thema der Informationssicherheit. Dabei besteht das Kompendium aus sogenannten IT-Grundschutz-Bausteinen, die für ein bestimmtes Themengebiet relevante Sicherheitsaspekte beleuchten. In jedem IT-Baustein wird zunächst auf die elementaren Gefährdungen und im Anschluss auf die Sicherheitsanforderungen eingegangen. Insgesamt sind die IT-Bausteine in zehn Bereiche aufgeteilt. Der Themenbereich der Zeitsynchronisation PTP wird im IT-Grundschutz-Baustein OPS (Operations) auf der Ebene der Prozessbausteine behandelt. [Bu23]

3 Der BSI-Baustein OPS.bd.1.31 Precision Time Protocol

Mit dem BSI-Baustein OPS.bd.1.31 Precision Time Protocol [Bu22] wird das Ziel verfolgt, Gefährdungen für das Protokoll aufzuzeigen, diesen entgegenzuwirken und ein Mindestmaß an Sicherheit zu gewährleisten. Als Geltungsbereich für den Baustein sieht das BSI jedes Kommunikationsnetzwerk an, in dem PTP eingesetzt wird. Des Weiteren soll beachtet werden, dass weitere Systembausteine des BSI Berücksichtigung finden sollen, die sich sowohl auf aktive als auch auf passive Komponenten des Kommunikationsnetzwerkes beziehen. Im Baustein selbst wird explizit auf den Systembaustein 1.1 Allgemeine Server und auf den Baustein NET 3.1 Router und Switches verwiesen.
Der BSI-Baustein beschäftigt sich des Weiteren mit der Gefährdungslage, also Angriffsszenarien, die typisch dafür sind, wie PTP-Kommunikation gestört werden kann und Anforderungen aus Sicht des sicheren Betriebs, die sich sowohl auf die Parametrierung als auch auf den Betrieb der Kommunikationsteilnehmer im Netzwerk beziehen. [Bu22]

3.1 Gefährdungslagen

Für die Darstellung von Angriffsszenarien und Schwachstellen, die bei der Nutzung von PTP in Kommunikationsnetzwerken vorliegen können, wird auf sogenannte typische Szenarien eingegangen, die die Nutzung der Zeitinformation der synchronisierten Uhren beeinträchtigen können. Innerhalb des Bausteins werden 14 Gefährdungslagen beschrieben, die die Themen der Manipulation von Paketen oder PTP-Teilnehmern, das Mithören der PTP-Kommunikation, die Übernahme von Funktionen im PTP-Netzwerk wie die des PTP-Grandmasters, die Manipulation oder direkte Beeinflussung von Sende- und Übertragungsverzögerungen, das gezielte Verwerfen von Paketen, die Störung durch Überlastsituationen auf dem Übertragungsmedium, die Ausnutzung von Protokollschwachstellen oder die Manipulation der Konfiguration der PTP-Teilnehmer betreffen. [Bu22]

3.2 Anforderungen des BSI-Bausteins

Die Anforderungen, die der BSI-Baustein für eine sichere PTP-Kommunikation fordert, gliedern sich in drei Kategorien: Basis-Anforderungen, Standard-Anforderungen und Anforderungen bei erhöhtem Schutzbedarf. Basis-Anforderungen sind mit Umsetzungsaufforderung *Müssen* verknüpft und somit in der Umsetzung zwingend umzusetzen. Standard-Anforderungen stellen genau wie die Basis-Anforderungen den aktuellen Stand der Technik dar und sollten somit umgesetzt werden. Im Falle der Anforderungen bei erhöhtem Schutzbedarf sollen Maßnahmen ergriffen werden, die über den aktuellen Stand der Technik hinausgehen. Ob die Einhaltung dieser Anforderungen für das jeweilige Kommunikationsnetz und die darin befindlichen Kommunikationsteilnehmer eingehalten werden muss, soll fallspezifisch anhand einer Risikoanalyse ermittelt werden.
Die Basis-Anforderungen fokussieren sich auf die Planung des PTP-Netzwerkes, die Konfiguration bzw. Parametrierung der PTP-Komponenten und PTP-Endgeräte sowie die Auswahl eines PTP-Grandmasters. Im Bereich der Standard-Anforderungen werden das Thema der Dokumentation des PTP-Netzwerkes, der Einsatz von PTP-Sicherheitsverfahren, die Festlegung der Dienstgüte (Quality of Service, QoS) und die Definition von Boundary-Clock Master Interfaces behandelt. Diejenigen Anforderungen, die im Falle erhöhter Schutzanforderungen gelten, decken den Bereich von Verschlüsselungsalgorithmen, des automatischen Monitorings und der automatischen Überwachung der Konfiguration und Kommunikation, die Authentifizierung von PTP-Kommunikationsteilnehmern, Robustheit, Verfügbarkeit, Notfallkonzepte und den Zugriff auf das Netzwerk aus der Ferne ab. Die nachstehende Tabelle listet die Anforderungen der drei Kategorien auf. Zur Einordnung der Anforderungen in die einzelnen Kategorien wird die aus dem BSI-Baustein genutzte Bezeichnung genutzt (Basis-Anforderungen (B), Standard-Anforderungen (S) und Anforderungen bei erhöhtem Schutzbedarf (H)).

Bezeichnung	Kategorie	Anforderung
BB.A1	B	Planung eines PTP-Netzwerkes
BB.A2	B	Sichere Basiskonfiguration der eingesetzten PTP-Komponenten
BB.A3	B	Sichere Basiskonfiguration der eingesetzten PTP-Endgeräte
BB.A4	B	Auswahl eines PTP-Grandmasters
BB.A5	S	Betriebsdokumentation
BB.A6	S	Einsatz von PTP-Sicherheitsverfahren
BB.A7	S	Einsatz von Quality-of-Service (QoS)
BB.A8	S	Definiton von Boundary-Clock Master Interfaces
BB.A9	H	Einsatz von Verschlüsselungsalgorithmen
BB.A10	H	Automatisiertes Monitoring der PTP-Konfiguration
BB.A11	H	Automatisierte Überwachung des Datenverkehrs
BB.A12	H	Aufnahme von PTP-Endpunkten in das Authentisierungs- und Autorisierungsverfahren
BB.A13	H	(D)DoS-Schutz
BB.A14	H	Hochverfügbarkeit
BB.A15	H	Notfdallkonzept
BB.A16	H	Abschottung der Netzwerkumgebung
BB.A17	H	Ausschluss von Transparent-Clocks
BB.A18	H	Aufbau einer Testumgebung

Tab. 1: Übersicht der Anforderungen

4 Fallstudie

Im Rahmen einer Fallstudie soll untersucht werden, ob die Anforderungen des BSI-Bausteins OPS.bd.1.31 für ein konvergentes, hybrides Kommunikationsnetzwerk, in dem für die Zeitsynchronisation das generalized Precision Time Protocol eingesetzt wird, umgesetzt werden können. Die Grundlage dafür stellt ein Testnetzwerk in der SmartFactoryOWL in Lemgo dar, das das Themengebiet der konvergenten, hybriden Kommunikation mit Ethernet-TSN veranschaulicht. Das Testnetzwerk ist in Abbildung 1 dargestellt.
Das Testnetzwerk besitzt einen OT-Anteil und einen IT-Anteil, die den Aspekt der konvergenten Kommunikation abdecken. Für die Kommunikation selbst kommen die Übertragungsmedien Full-Duplex-Ethernet (100Base-TX, 1000Base-T, 10GBase-CX4), Single-Pair Ethernet (SPE) (10Base-T1L), Advanced Physical Layer (APL) (10Base-T1L) und 5G zum Einsatz, was die hybride Komponente des Demonstrators darstellt. Die Datenraten im Netzwerk befinden sich im Bereich von 10 Mbit/s bis zu 10 Gbit/s. Im Bereich der OT-Kommunikation teilt sich der Demonstrator auf die Ebenen der speicherprogrammierbaren Steuerungen, der

Netzwerkinfrastruktur und der Endgeräte auf. Als Industrial-Ethernet-Protokolle kommen Profinet, CC-Link, OPC UA Publisher/Subscribe und EtherCAT mit der Kommunikationsart Open Mode zum Einsatz. Auf der IT-Ebene kommt ein Videostream einer Kamera unter Verwendung des GigE-Protokolls zum Einsatz. Innerhalb des Netzwerkes wird die Zeitsynchronisation nach IEEE 802.1AS-2020 (gPTP) bereitgestellt. Hierfür wird ein Zeitserver genutzt, der als Zeitreferenz GNSS nutzt. Außerdem kommt auf der Netzwerkinfrastrukturebene die TSN-Funktion Frame Preemption zum Einsatz. Die 5G-Kommunikationsstrecke bindet ein weiteres Testnetzwerk an, der das Thema der bildbasierten Positionssteuerung eines Linearantriebes und Predictive Maintenance behandelt. Diese Applikationen werden von einer virtualisierten Steuerung gesteuert, die für die Kommunikation sowohl das TSN-Kommunikationsnetz als auch die 5G-Kommunikationsverbindung nutzt.
Im Nachfolgenden soll evaluiert werden, ob die Anforderungen aus Kapitel 3.2 innerhalb des Testnetzwerks umgesetzt werden können.

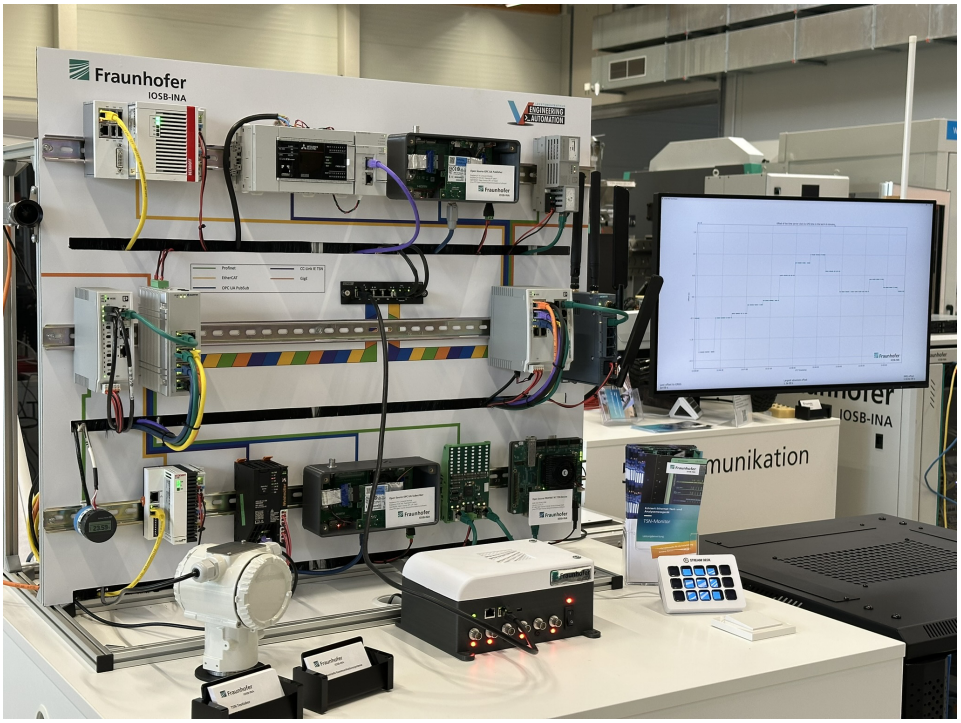


Abb. 1: Testnetzwerk IT/OT-Konvergenz und Echtzeitkommunikation mit Ethernet TSN in der SmartFactoryOWL

4.1 Umsetzung von Basis-Anforderungen

Die Anforderungen BB.A1 - BB.A4 stammen aus der Kategorie der Basis-Anforderungen. Diese sehen vor, dass eine Planung des Netzwerkes zugrunde liegen soll, dass eine sichere Basiskonfiguration von PTP-Komponenten und Engeräten vorliegt und ein PTP-Grandmaster ausgewählt wird. Im Bereich der Planung des PTP-Netzwerkes gilt es festzulegen, welches Gerät im PTP-Netzwerk die Rolle des PTP-Grandmasters einnimmt, wie sichergestellt wird, dass dieser diese Rolle auch im Normalfall übernimmt und wie das Verhalten des PTP-Netzwerkes im Fehlerfall ist. Außerdem wird festgelegt, welches Transportprotokoll genutzt wird und wie lange ein Ausbleiben von Zeitinformationen tolerierbar ist. Abbildung 2 zeigt anhand eines Blockschaltbildes die Topologie und Rollenverteilung der Komponenten im Testnetzwerk. Dargestellt ist die Umsetzung der Anforderungen BB.A1 - BB.A3. Im Normalbetrieb erfolgt die Vorgabe der Zeitinformation durch den im Netzwerk vorhandenen Zeitmaster. Die Festlegung, dass der Zeitmaster im Normalfall auch als gPTP-Grandmaster ausgewählt wird, wird durch den BTCA-Parameter Priority 1 realisiert. Dieser wird in Anlehnung an den Standard IEEE 802.1AS-2020 auf den Wert 246 gesetzt, der aussagt, dass es sich bei dem Gerät um einen zentralen und kritischen Teil des Netzwerkes handelt, von dem nicht erwartet wird, dass er abgeschaltet wird. Analog dazu, werden die Werte der anderen gPTP-Teilnehmer auf einen aus Sicht der Entscheidungsfindung des BTCA schlechteren, also zahlenmäßig höheren Wert, gesetzt. Für Geräte, die nicht an der Auswahl zum Grandmaster bzw. Time-Transmitter teilnehmen sollen, wird ein Wert von 255 eingestellt. Für die Übertragung der Zeitinformation wird durch IEEE 802.1AS-2020 die Schicht-2-Kommunikation des ISO/OSI-Referenzmodells mit einer LLDP-Multicast-Adresse vorgesehen. Der Zeitmaster erhält seine Zeit über einen GNSS-Empfänger. In diesem Fall wird GPS genutzt. Im Falle des Verlustes oder einer Störung des GNSS-Signals ist der Zeitmaster mit einem Quarz-Oszillator ausgerüstet, der nachdem er 24 Stunden auf die GNSS-Zeit synchronisiert war, innerhalb von einem Tag um ±65 µs von der GNSS-Zeit wegdriftet. Somit können die derzeitigen Anforderungen, die eine Genauigkeit von ±100 µs fordern, eingehalten werden. [Bu22][IE20][IE24]

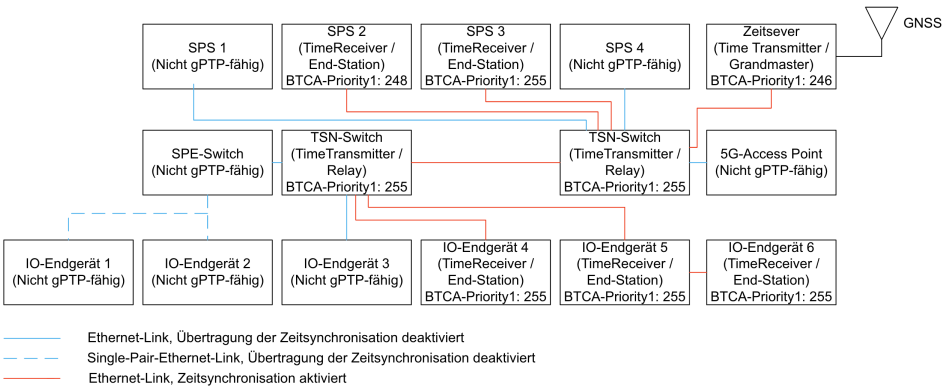


Abb. 2: Blockschaltbild der PTP-Parametrierung im Technologiedemonstrator

4.2 Umsetzung von Standard-Anforderungen

Die Anforderungen BB.A5 - BB.A8 decken die Standard-Anforderungen ab. Für das Testnetzwerk und die technische Umsetzung sind die Anforderungen BB.A6 und BB.A7 relevant. Die Erstellung einer Betriebsdokumentation (BB.A5) unterliegt keinen technischen Anforderungen. Die Anforderung BB.A6 kann derzeit unter der Nutzung von gPTP nicht erfüllt werden. Während der PTP-Standard in der Version aus dem Jahr 2019 innerhalb des Annex P [IE19] Sicherheitsmechanismen für die Authentifizierung definiert, definiert der Standard IEEE 802.1AS-2020, dass dieser Teil von PTP nicht unterstützt wird. Die Anforderung BB.A7, die für die Übertragung der Zeitinformation fordert, dass Quality-of-Service Maßnahmen ergriffen werden, dass die Zeitsynchronisation priorisiert im Netzwerk übertragen wird. Dies ist durch die IEC/IEEE 60802 für zukünftige TSN-Netzwerke sichergestellt. Bei einer konformen Nutzung wird die Zeitsynchronisation mit der höchsten Priorität im Netzwerk übertragen. Anforderung BB.A8 bezieht sich auf die Parametrierung von Kommunikations-Interfaces bei Boundary-Clocks. Während die Zeitsynchronisation mit PTPv2 mehrere Uhrenmodelle definiert, definiert gPTP zwei Ausprägungen: die End-Station und das Relay. Zusätzlich definiert gPTP den booleschen Parameter syncLocked, der eine Aussage darüber trifft, wie die Zeitinformation weitergeleitet wird. Ist der Wert TRUE, sendet das Relay die Zeitinformation so schnell wie möglich weiter und verhält sich bezogen auf die Uhrenmodelle von PTPv2 wie eine P2P-Transparent Clock. Ist der Wert FALSE, wird das Sendeintervall für Synchronisations-Nachrichten eingehalten. Dies entspricht dann dem Uhrenmodell einer Boundary Clock. Da innerhalb des Technologiedemonstrators der Wert für syncLocked in allen PTP-Instanzen den Wert TRUE führt, ist die Anforderung BB.A8 nicht relevant. [IE19][IE20]

4.3 Umsetzung von Anforderungen mit erhöhtem Schutzbedarf

Die Anforderungen mit erhöhtem Schutzbedarf umfassen die Anforderungen BB.A9 - BB.A18. Anforderung BB.A9 wird innerhalb des Testnetzwerks nicht umgesetzt. Keine der im Testnetzwerk eingesetzten Komponenten unterstützt eine Verschlüsselung oder Funktionen wie MACSec oder IPSec. Genau wie im Falle der aktuell noch fehlenden Authentifizierung wird die Verschlüsselung mit neuen TSN-Geräten auch in der Breite verfügbar sein. Die Anforderungen BB.A10 und BB.A11 behandeln das automatisierte Monitoring der PTP-Konfiguration und die automatisierte Überwachung des Datenverkehrs. Diese Anforderungen befinden sich aktuell in der Umsetzung innerhalb des Technologiedemonstrators. Neue Ansätze und Konzepte, wie das Monitoring von Parametrierungen oder aber die Überwachung von Kommunikationsdaten mithilfe künstlicher Intelligenz unterstützt werden können, werden im Rahmen des Fraunhofer Leistungszentrums Engineering-Automation erarbeitet. [Wi25] Für die Überwachung der Kommunikationsdaten wurde zur Veranschaulichung ein Mechanismus in das Testnetzwerk integriert, der die PTP-Kommunikation in ihrer Dienstgüte auf ein Minimum reduziert, jedoch die Synchronisation nicht abbricht. Hierfür werden für eine definierte Zeitspanne Synchronisations-Frames in einem Maße verworfen,

dass der Mechanismus zur Erkennung von ausbleibenden Synchronisations-Frames dies nicht erkennt. Die automatisierte Auswertung erkennt die Anomalie innerhalb der gPTP-Kommunikation und meldet diese dem Netzwerkbetreiber. Zur Visualisierung wird das in Abbildung 3 gezeigte Dashboard genutzt. Der aktuell genutzte Ansatz wird innerhalb einer TSN-Endstation genutzt. Zukünftig ist aber auch der Einsatz auf Infrastrukturkomponenten wie TSN-Switches möglich.

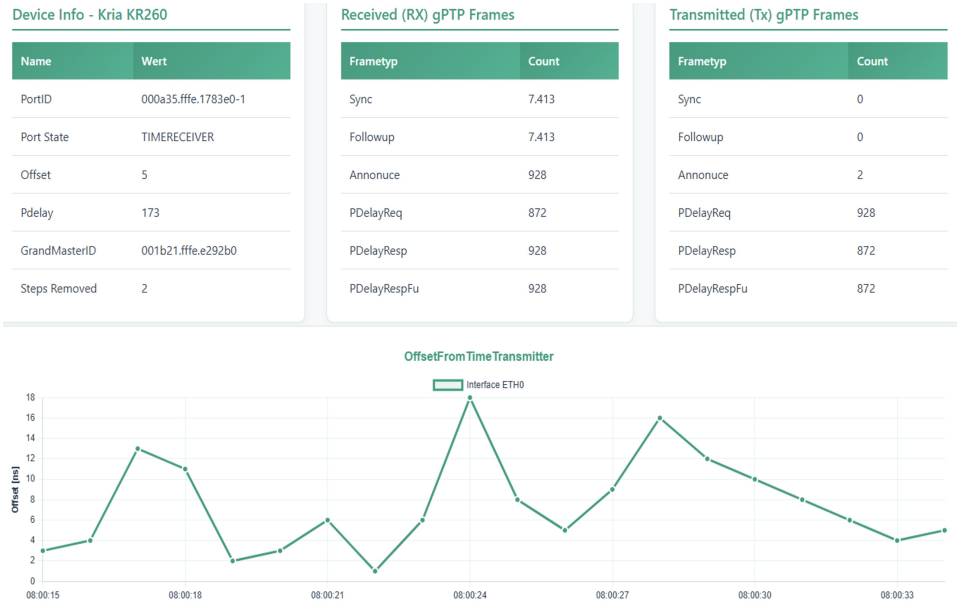


Abb. 3: Visualisierung der automatisierten Überwachung der PTP-Kommunikation

Die Anforderungen BB.A12 und BB.A13 sind innerhalb des Testnetzwerks nicht umgesetzt. Wie bereits für die Anforderung BB.A6 gilt, dass Authentifizierung seitens der IEEE 802.1AS-2020 nicht unterstützt wird. Maßnahmen zum Schutz vor (D)DoS-Attacken sind im aktuellen Stadium des Demonstrators nicht ausgeprägt. Die Anforderung BB.A14 ist durch die Nutzung des BTCA-Algorithmus und die Übernahme der Zeitsynchronisation durch einen weiteren Zeitmaster umgesetzt. Ein Notfallkonzept (BB.A15) ist nicht vorhanden. Je nach Applikation, die über zukünftige TSN-Kommunikationsnetzwerke betrieben wird, sind jedoch verschiedene Konzepte vorstellbar, die je nach Fehlersituation innerhalb der Zeitsynchronisation umgesetzt werden können. Anforderung BB.A16 wurde umgesetzt, indem für alle Kommunikationsschnittstellen, die nicht für die Zeitsynchronisation genutzt werden, die Funktion von gPTP deaktiviert ist. Anforderung BB.A17, die den Ausschluss von Transparent-Clocks fordert, ist nicht ohne Weiteres umsetzbar. Wie bereits im Rahmen von Anforderung BB.A8 erläutert, verhält sich gPTP wie eine Transparent Clock, sobald die Sendeintervalle innerhalb des Netzwerkes auf einen identischen Wert eingestellt sind (syncLocked). Abschließend fordert Anforderung BB.A18 eine Testumgebung, um Anpassungen

innerhalb des Produktivnetzwerkes im Vorfeld auf ihre Auswirkungen untersuchen zu können. Das Testnetzwerk selbst kann als eine solche Testumgebung angesehen werden. Im Falle von Topologien mit großen Mengengerüsten oder großen geografischen Ausdehnungen, kann auch die Simulation von Kommunikationsnetzwerken als geeignetes Mittel eingesetzt werden. Für die Simulationsumgebung OMNeT++ und das Erweiterungsframework INET sind bereits heute Simulationsmodelle vorhanden, die die Zeitsynchronisation nach IEEE 802.1AS umsetzen und als Basis für eine Testumgebung dienen können. [BA24]

5 Ausblick und Fazit

Der vorliegende Beitrag hat gezeigt, das eine Umsetzung des BSI-Bausteins OPS.1.31 Precision Time Protocol anhand eines Testnetzwerkes für die IT/OT-Konvergenz und die Echtzeitkommunikation mit Ethernet-TSN realisiert werden kann. Während ein Großteil der Anforderungen erfüllt werden, besteht im Bereich der Authentifizierung und Authorisierung noch Handlungsbedarf. Hier befindet sich noch eine Umsetzungslücke, die weder durch den eigentlichen IEEE-Standard für die Zeitsynchronisation, noch durch die IEC/IEEE 60802 abgedeckt wurde. Soll TSN und somit auch gPTP den Weg in zukünftige Kommunikationsnetzwerke finden, muss hier nachgebessert werden. Auch im Bereich der Uhrenmodelle muss im Weiteren eine Bewertung durchführt werden, ob durch die Nutzung von Transparent-Clocks im Bereich der Synchronisation mit gPTP eine grundsätzlich Schwachstelle besteht. Die aktuell nicht umgesetzte Verschlüsselung wird in den kommenden Jahren mit der Verbreitung von Technologien wie MACSec sich weiter etablieren. Gleichzeitig wird auch der Bedarf nach Lösungen im Bereich des automatisierten Monitorings und der automatisierten Überwachung, wie er in Kapitel 4.3 wieder steigen, wenn die gesamte Kommunikation verschlüsselt wird und eine Analyse der Daten mit einem erhöhten Aufwand möglich ist.

Literaturverzeichnis

[Al25] Alberta Electric System Operator (AESO): Connection Requirements for Inverter-Based Resources, 2025.

[BA24] Biendarra, A.; Albrecht, J.: Evaluierung von Time-Sensitive Networking in OMNeT++ für die Simulation von IEC/IEEE 60802-konformen Netzwerken, 2024.

[Bu22] Bundesamt für Sicherheit in der Informationstechnik (BSI): OPS.bd.1.31 Precision Time Protocol, 2022.

[Bu23] Bundesamt für Sicherheit in der Informationstechnik (BSI): IT-Grundschutz-Kompendium, 2023.

[De19] Deutsches statistisches Bundesamt: Erneuerbare Energien decken ein Fünftel des globalen Verbrauchs, 2019, https://www.destatis.de/DE/Themen/Laender-Regionen/Internationales/Thema/umwelt-energie/energie/ErneuerbareEnergienGlobal.html, Stand: März. 2025.

[De24] der Deutschen Bundesrepublik, B.: Entwurf eines Gesetzes zur Änderung des Energiewirtschaftsrechts im Bereich der Endkundenmärkte, des Netzausbaus und der Netzregulierung, 2024.

[Ga24] Gabriel, P. et al.: Standards und Normen für die digitale Sektorenkopplung in Deutschland und Europa. 2024.

[IE16] IEC/IEEE: Communication networks and systems for power utility automation-Part-9-3: Precision time protocol for power utility automation, 2016.

[IE17] IEEE: IEEE Standard Profile for Use of IEEE 1588 Precision Time Protocol in Power System Applications, 2017.

[IE19] IEEE: IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, 2019.

[IE20] IEEE: IEEE Standard for Local and Metropolitan Area Networks - Timing and Synchronization for Time-Sensitive Applications, 2020.

[IE24] IEEE: IEEE Standard for Local and Metropolitan Area Networks - Timing and Synchronization for Time-Sensitive Applications, Amendment 1: Inclusive Terminology, 2024.

[IE25] IEC/IEEE: IEC/IEEE 60802 Time-Sensitive Networking Profile for Industrial Automation, Draft 3.4, IEC/IEEE 60802 Time-Sensitive Networking Profile for Industrial Automation, Draft 3.4, 2025.

[PT19] PTP Telecom Profile in Power Grids, 2019, https://blog.meinbergglobal.com/2019/01/23/ptp-telecom-profile-in-power-grids/, Stand: Sep. 2025.

[Wi25] Windmann, S. et al.: NetPilot – Towards LLM-Assisted Configuration of Hybrid TSN/5G Networks. In: Emerging Technologies and Factory Automation (ETFA) 2025. 2025.

http://dx.doi.org/10.25673/121097

Eclipse Dataspace Components: A Security Framework for Decentralized Industrial Data Sharing

Cheng Xin
ifak[1], OVGU[2]
Cheng.Xin@ifak.eu

Alfred Barnard
ifak
Alfred.Barnard@ifak.eu

Abstract: Modern industrial systems increasingly depend on secure and sovereign data exchange across organizational boundaries. Traditional approaches such as peer-to-peer sharing or centralized platforms fail to balance flexibility, scalability, and data ownership. To address these challenges, **data spaces** have emerged as a decentralized paradigm that enables trusted and interoperable data sharing among autonomous participants. This paper introduces the **Eclipse Dataspace Connector (EDC)** as an open-source implementation of this concept. The paper first reviews the evolution of industrial data sharing, from isolated peer-to-peer systems to federated and policy-driven data spaces. It then details the EDC architecture which covering its **Control Plane**, **Data Plane**, **Identity Hub**, and **policy framework**, and explains how decentralized claims, contract negotiation, and policy enforcement ensure privacy and trust. Finally, it compares two major ecosystem implementations, **Catena-X** and **Pontus-X**, showing how each extends the EDC blueprint through distinct trust and governance mechanisms: BPN-based federation in Catena-X and blockchain-anchored self-sovereign identity in Pontus-X. The analysis highlights that EDC provides a flexible, extensible foundation for realizing both industry-specific and cross-domain data spaces.

Keywords: Data Spaces, EDC, Decentralized Claims Protocol, Industrial Data Sharing.

1. Introduction

Data has become a critical asset in modern industrial systems. These systems continuously collect data from sensors and control devices, and use it to optimize operations. Beyond local control, the collected data is increasingly forwarded to other applications for analysis and decision support. This process of transferring or exchanging data is generally referred to as **data sharing**. Data sharing may occur within a single enterprise, e.g., between production, logistics, and maintenance systems, or across organizational boundaries.

1.1. Traditional Peer-to-Peer Sharing

Historically, data sharing followed a **peer-to-peer** model, in which data is exchanged directly between two entities. This approach works relatively well within one organization, where security policies and trust relationships are pre-established. However, when sharing data between different companies, the process becomes considerably more complex: both parties must negotiate technical interfaces, data formats, and, most importantly, trust.

[1]ifak - Institut für Automation und Kommunikation e. V., Germany
[2]Otto-von-Guericke-University Magdeburg, Germany

Typical implementations include file transfer via FTP or APIs exposed by the data provider. In this model, the data provider is also the data holder, responsible for hosting the data and maintaining the service interface. The peer-to-peer model offers flexibility and full control to the data provider, but it also suffers from several drawbacks:

- Lack of control over data usage after transfer [1], [2].
- Absence of standardized interfaces, making integration on the consumer side difficult [3].
- Limited scalability when many participants are involved [4].

1.2. Centralized Data Platforms

To overcome these limitations, industries adopted **centralized data platforms** as a generalized solution for sharing data. Here, all relevant data is collected in a central repository that offers standardized access interfaces to external consumers [5], [6].

In this case, the data provider is usually not the data holder. Providers must upload or publish their data to the platform operator, who then manages access for multiple consumers [7]. While this centralization simplifies integration and management, it introduces new challenges, that is, providers lose direct control over their data, and the platform operator directly gains ownership [2]. Furthermore, centralized systems face privacy and governance concerns, since the platform operator must process identity and authorization information about both data providers and consumers.

1.3. Federated Data Sharing

To regain control and data sovereignty, **federated data sharing** architectures were introduced. A federated system acts as a meta–data management framework that coordinates access to distributed datasets without requiring central storage [8], [9]. Instead of hosting data, it manages **metadata**, i.e., semantics describing the meaning, location, and ownership of data sources.

When a consumer requests access, the federation establishes a direct connection to the data provider. The provider remains the data holder, while access policies and metadata are managed centrally. Although this approach preserves data locality, traditional federated systems often lack standardized mechanisms for cross-organizational trust, usage control, and contract enforcement. It also lacks a standardized and effective way to scale across companies and domains.

1.4. Data Spaces: Toward Trusted and Sovereign Sharing

Data spaces represent the latest stage in the evolution of industrial data sharing. Their purpose is to enable **trusted**, **sovereign**, and **interoperable** data exchange among multiple, autonomous participants [10], [11].

In essence, data spaces combine the strengths of earlier data-sharing paradigms while addressing their key limitations. Unlike traditional **peer-to-peer** approaches, they rely on standardized interfaces and protocols that simplify integration and reduce the need for customization. At the same time, they avoid the drawbacks of **centralized data platforms** by keeping data within the provider's own domain, that is, only controlled access is granted, never ownership transfer. Compared to **federated data sharing** systems, data spaces take a further step toward decentralization, that is, both data and policy enforcement remain local,

trust, identity, and contractual agreements are established through common standards and interoperable services (respecting Data Governance Act (DGA) [12]).

A data space consists of loosely coupled components that can be deployed independently and interoperate through well-defined standards such as the **Dataspace Protocol** (DSP) [13]. This decentralized architecture enables scalability across organizations of different sizes while maintaining interoperability. Data transfer within a data space is secured end-to-end through encryption and policy enforcement mechanisms.

Furthermore, data spaces rely on **decentralized identity and credential management**. Participants, whether providers or consumers, can manage their credentials locally or delegate verification to trusted third-party identity services. This design eliminates the need for a central authority to store sensitive identity information, thereby improving privacy and reducing compliance risks.

The **EDC** is an open-source framework designed to facilitate secure, decentralized sharing of industrial and business data [14]. It is one of the core building blocks of a data space, capable of acting in multiple roles, including **data provider**, **data consumer**, **credential issuer**, or **verifier**. Well-fitted for industrial and automation domains, the EDC implements a modular and extensible security framework that allows organizations to focus on data functionality and interoperability rather than low-level security integration.

Its decentralized architecture enables scalable deployment based on business requirements and infrastructure capacity. It supports extensibility, allowing users to adapt the verification workflow to their own security policies and pipelines. Organizations can dynamically update or extend these policies at runtime, integrate additional validation mechanisms, and tailor security checks to fit specific needs.

1.5. Contributions

This paper makes the following contributions:

- Provides a detailed introduction to the EDC architecture, including its core components and decentralized trust mechanisms.
- Compares the EDC-based approach with ecosystem implementations such as **Catena-X** [15] and **Pontus-X** [16], illustrating how each extends the EDC blueprint through different governance and identity frameworks (BPN-based federation versus blockchain-anchored self-sovereign identity).
- Presents a generalized industrial architecture based on EDC and analyzes its advantages over traditional peer-to-peer, centralized, and federated data-sharing systems.

2. Architecture Overview

2.1. Core Components

The EDC comprises several core components: **Identity Hub**, **Catalog Service**, **Control Plane**, and **Data Plane** [10]. The **Identity Hub** manages decentralized identity and credential exchange. It issues **self-issued tokens** and provides **verifiable credentials** that represent the participant's identity and claims. The **Control Plane** controls data exchange operations, i.e., catalog discovery, contract negotiation, and transfer initiation, while ensuring compliance

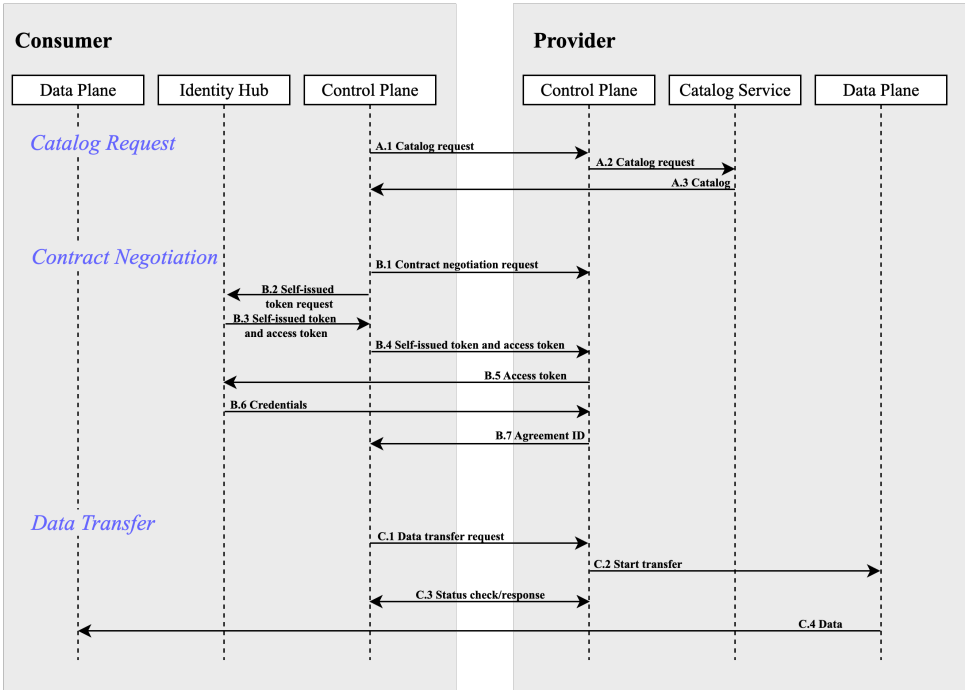


Figure 1: Sequence diagram illustrating the interaction between a consumer and a provider during catalog discovery, contract negotiation, and data transfer in an EDC-based data space.

with data usage policies. Finally, the **Data Plane** handles the actual data transmission between participants, enforcing secure, encrypted, and policy-constrained transfers.

These components work together to ensure secure, auditable, and data exchange between organizations. Figure Figure 1 illustrates a typical interaction between two participants: a **consumer** and a **provider**, each operating their own EDC instance with the aforementioned components.

A complete data-sharing workflow typically consists of three main stages:

1. **Catalog Request (A)**: The consumer's Control Plane sends a catalog query to the provider's Control Plane, which forwards the request to its Catalog Service. The Catalog Service returns a list of available data assets, including metadata and usage policies.
2. **Contract Negotiation (B)**: The consumer selects a desired asset and initiates a contract negotiation with the provider. During this process, the consumer's Control Plane requests a **self-issued token** and an **access token** from its Identity Hub (steps **B.2**–**B.3**). These tokens are used to authenticate the consumer and enable secure credential exchange. The provider's Control Plane validates the received tokens (**B.4**–**B.5**) and retrieves the consumer's credentials for verification. Once both sides agree on the contract terms, an **agreement ID** is issued, marking the start of a trusted relationship.
3. **Data Transfer (C)** Following successful contract negotiation, the consumer's Control Plane initiates a data transfer request (**C.1**). The provider's Control Plane triggers its Data Plane to prepare and transmit the requested data (**C.2**–**C.4**). All transfers are encrypted and governed by the policies defined by the provider.

Among these stages, steps **B.2**–**B.5** are particularly critical because they demonstrate **decentralized claims presentation**. By relying on **self-issued tokens** rather than a

centralized identity provider, each participant retains control over its own credentials and authentication process. This design enhances privacy and sovereignty, that is, the consumer determines when and how its credentials are shared, without exposing them to a central authority.

3. Identity and Trust

3.1. Contract and Policy

Contract negotiation is one of the most essential mechanisms within the EDC. It governs how data access is granted and under what conditions data usage is permitted. This is achieved through **policies**, which define the rules, permissions, and obligations that apply to each data exchange.

Unlike traditional centralized authorization systems, the activation and deactivation of policies in EDC are managed **locally**. This decentralization offers both flexibility and security, that is, each participant maintains full control over its own access policies without relying on a central authority. Consequently, organizations can enforce their governance rules independently while still adhering to the standardized contract negotiation process defined by the DSP.

The validation of these policies occurs during the **B.6–B.7** phase in the sequence diagram referenced in Figure 1. At this stage, the provider validates the consumer's credentials and evaluates the applicable usage policies before granting access to the requested data.

Figure Figure 2 describes the internal structure of policy-related entities as implemented in the EDC framework. A **ContractNegotiation** leads to a **ContractAgreement**, which references a

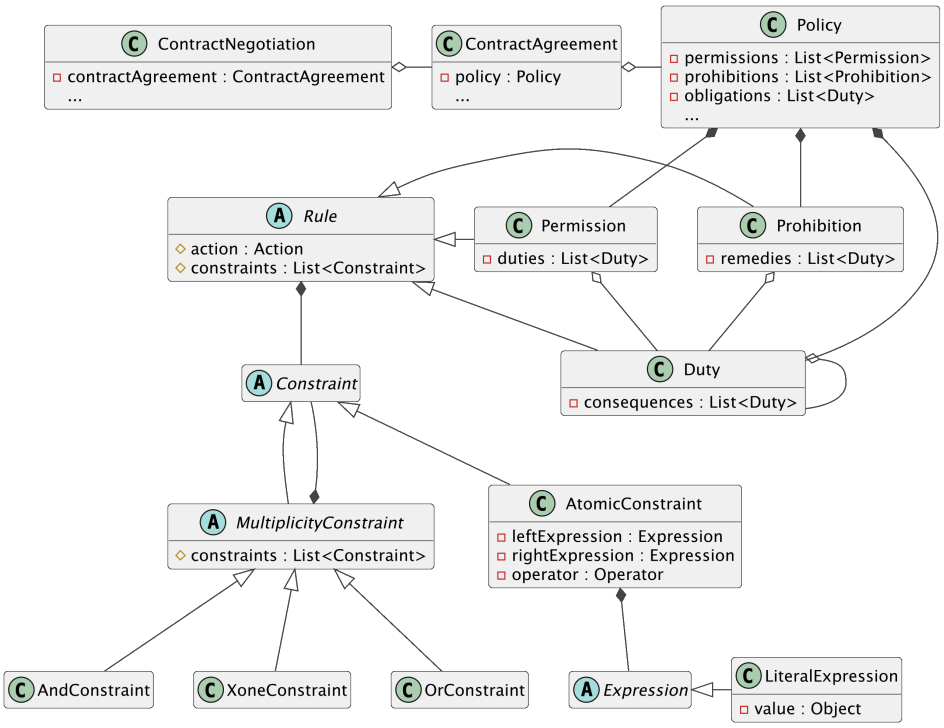


Figure 2: Class diagram of policy-related entities in the EDC implementation, adapted from https://github.com/eclipse-edc/Connector/tree/release/0.14.0 .

Policy that governs the data usage terms. A **Policy** consists of multiple **Permissions**, **Prohibitions**, and **Duties**:

- A **Permission** defines allowed actions and includes **Duties** (e.g., obligations to acknowledge or delete data after use).
- A **Prohibition** specifies forbidden actions and lists **Remedies** that describe corrective measures when violations occur.
- A **Duty** defines mandatory actions, such as notifying the provider or deleting data after expiration, and contains **Consequences** for non-compliance.

Each of these rule entities (**Permission**, **Prohibition**, **Duty**) is based on one or more **Constraints**. The **Constraint** class is recursively defined, enabling complex logical relationships such as **AndConstraint**, **OrConstraint**, and **XoneConstraint** (exclusive one). This recursive structure allows policies to express fine-grained, composable conditions. For example, it can specify that data access is permitted **only if** multiple claims or contextual parameters are satisfied.

At the lowest level, the **AtomicConstraint** class binds a left-hand and right-hand **Expression** with a logical **Operator** (e.g., EQUALS, GREATER_THAN, IS_A, HAS_PART). The **Expression** may take the form of a **LiteralExpression**, whose **value** represents an evaluated attribute, such as a credential claim obtained from the **Identity Hub** service. In this context, a **LiteralExpression** acts as a concrete verification input, enabling dynamic policy evaluation based on verifiable credentials.

3.2. Claims Model

In a data space, maintaining **security** and **privacy** fundamentally depends on effective **claims management**. The EDC achieves this through the **Decentralized Claims Protocol (DCP)** [17], which defines how claims are issued, exchanged, and verified among participants.

Within DCP, four principal roles are defined: **Issuer**, **Subject**, **Holder**, and **Verifier**. Each role performs specific actions, as summarized in Table 1.

In the EDC ecosystem, these roles correspond to specific services that handle identity and credential verification:

- The **Consumer's Identity Hub** functions as the **Holder**, responsible for managing and presenting credentials.
- The **Consumer's Control Plane** acts as the **Subject**, requesting and asserting credentials during negotiation.
- The **Provider's Control Plane** serves as the **Verifier**, validating the received credentials and applying access policies accordingly.
- The **Issuer** is typically an external authority or organization that is not part of the EDC instance itself.

Through this decentralized claims model, EDC enables participants to establish trust dynamically, that is, without relying on a centralized identity provider, while preserving sovereignty over their credentials and data-sharing relationships.

Role	Action Name	Description
Issuer	Issue Claim (Credential)	Issuing a claim to represent a verifiable credential.
Issuer	Revoke Claim	Revoking a previously issued claim when it becomes invalid.
Issuer	Amend Claim	Updating an existing claim.
Holder	Assert Claim	Presenting a claim to another participant as part of an interaction.
Holder	Store Claim	Storing credentials in one or more credential repositories for later use.
Holder	Retrieve Claim	Selecting and preparing appropriate credentials to present to a **verifier**.
Verifier	Verify Claim	Ensuring that a claim has not been altered and remains valid.
Verifier	Validate Claim	Checking whether a verified claim satisfies a specific policy or usage condition.

Table 1: Possible actions in the DCP.

4. Industrial Systems Architecture with EDC

The above figure illustrates a typical hierarchy of an industrial system built upon the EDC. This architecture is organized into multiple layers, each serving a distinct role in enabling secure and sovereign data exchange. At the top, the **Industrial Applications** layer hosts enterprise systems, analytics tools, digital twins, and other domain-specific applications. These applications consume and process data provided through the data space without needing to manage security or identity mechanisms directly. Beneath it lies the **EDC Connector** layer, representing the participant's integration point with the data space. This layer implements the **DSP**, which governs data transfer, policy enforcement, and contractual control between participants. The next layer, **Trust & Governance**, ensures that all data

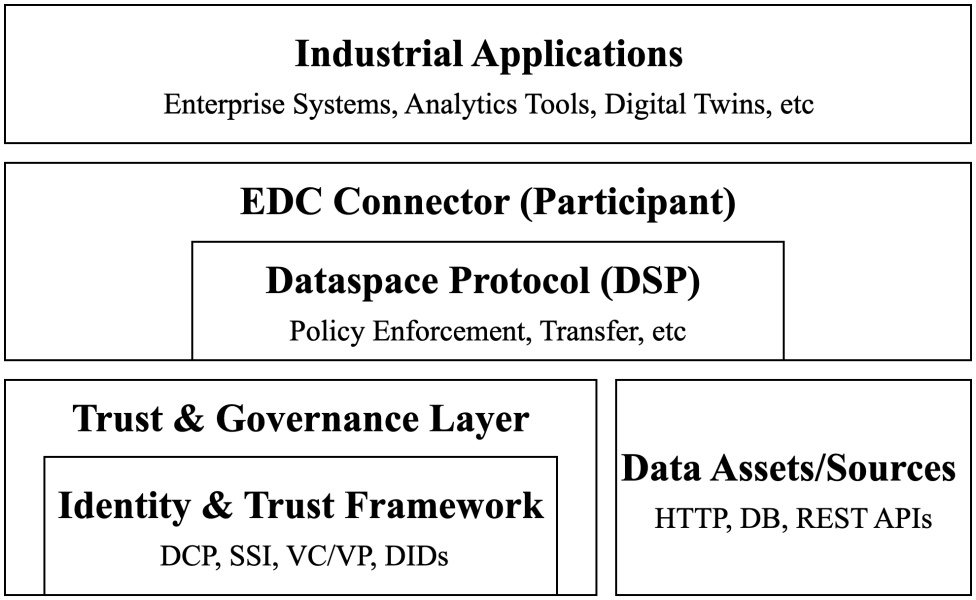


Figure 3: A typical hierarchy of an industrial system built upon the EDC.

exchanges occur under verifiable and compliant conditions. It incorporates an *Identity & Trust Framework* that supports decentralized credentials, including **Decentralized Identifiers (DIDs)** [18], **Verifiable Credentials (VCs)** or **Verifiable Presentations (VCs)** [19], and **Self-Sovereign Identity (SSI)** [20] principles. This enables participants to establish trust without relying on a centralized authority. At the foundation are the **Data Assets or Sources**, such as databases, HTTP endpoints, and REST APIs, which provide the actual data to be shared.

For engineers, the lower layers, i.e., policy enforcement, identity, and credential management, remain abstracted. They only need to focus on implementing functional logic at the application layer, such as issuing a catalog request or connecting to a selected dataset. The EDC framework transparently handles negotiation, authentication, and secure data transfer. Similarly, data providers simply expose their data and define the appropriate access policies, while the EDC ensures that all identity verification and contract enforcement occur automatically.

For data space designers, the hierarchy framework shown in Figure 3 serves as a blueprint for constructing interoperable and secure data spaces. Different ecosystems such as **Catena-X** and **Pontus-X** adopt this same foundational architecture, but they extend or specialize individual layers according to their governance and identity models.

In Catena-X, the **Trust & Governance** layer is expanded with a formal **Federation and Identity Management** system. Participants are registered and identified by **Business Partner Numbers (BPNs)** issued through the **Business Partner Data Management (BPDM)** service. These BPNs are embedded in verifiable credentials that form the trust foundation of the Catena-X ecosystem. At the **application layer**, Catena-X further integrates **Digital Twin** and **Asset Administration Shell (AAS)** components, enabling standardized representation and interaction with industrial assets. This tight integration of governance, identity, and semantic interoperability ensures traceability and compliance across the automotive value chain.

In Pontus-X, the **Identity & Trust Framework** evolves toward a **blockchain-anchored self-sovereign identity** model [10]. Instead of relying on a central registry, participant identifiers are published on a distributed ledger, providing immutable and verifiable trust anchors. The VCs and VPs are exchanged off-chain, but their validity and revocation are verified through blockchain-based registries. Pontus-X also extends the **Trust & Governance** layer with **Federation Services** and **Smart Contract-based Governance**, allowing federations to manage participant onboarding, compliance, and policy attestation transparently. This design supports open, cross-domain data spaces and aligns closely with the **GAIA-X** trust framework.

In summary, both **Catena-X** and **Pontus-X** inherit the same EDC-based architecture:
- **Catena-X** emphasizes industry-specific governance, identity assurance via BPNs, and asset semantics (AAS, Digital Twin).
- **Pontus-X** emphasizes decentralized, blockchain-based trust infrastructure and open federation interoperability.

Together, they exemplify how the EDC blueprint can be adapted to realize both **governed industrial ecosystems** and **open federated data spaces**.

5. Conclusion

The EDC provides a robust and extensible foundation for secure, scalable, and decentralized data sharing in industrial ecosystems. By separating trust and identity management from application logic, it enables organizations to collaborate and exchange data without sacrificing sovereignty or privacy. Its modular architecture allows each participant to retain full control over data, credentials, and policies while still achieving interoperability through the **DSP**.

The study demonstrates how the EDC's design principles are realized in large-scale implementations. **Catena-X** applies EDC to an industry-governed automotive ecosystem, introducing **BPNs**, **Digital Twin**, and a formal federation governance model. In contrast, **Pontus-X** extends the same architecture toward a **blockchain-anchored** and **self-sovereign identity** framework that supports open, cross-domain federations. Together, they illustrate the adaptability of EDC across distinct trust models, that is, from centrally governed to fully decentralized data ecosystems.

As industrial digitization accelerates, data sovereignty and trust will remain central challenges. EDC's open, standardized, and policy-driven architecture positions it as an important part of the next generation of data sharing systems, bridging technological interoperability with verifiable governance across industrial systems.

Bibliography

[1] O. for Economic Co-operation and Development, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use Across Societies*. Organisation for Economic Co-Operation, Development, 2019.

[2] T. Devriendt, P. Borry, and M. Shabani, "Factors that influence data sharing through data sharing platforms: A qualitative study on the views and experiences of cohort holders and platform developers," *PLoS One*, vol. 16, no. 7, p. e254202, 2021.

[3] I. M. Putrama and P. Martinek, "Heterogeneous data integration: Challenges and opportunities," *Data in Brief*, vol. 56, p. 110853, 2024.

[4] A. Rosenthal and L. Seligman, "Scalability issues in data integration," in *Proceedings of the AFCEA federal database conference*, 2001.

[5] G. W. Hruby, J. McKiernan, S. Bakken, and C. Weng, "A centralized research data repository enhances retrospective outcomes research capacity: a case report," *Journal of the American Medical Informatics Association*, vol. 20, no. 3, pp. 563–567, 2013.

[6] A. Gieß and A. Hutterer, "The Future of Data Management: A Delimitation of Data Platforms, Data Spaces, Data Meshes, and Data Fabrics," *Information Systems and e-Business Management*, pp. 1–27, 2025.

[7] S. A. Azcoitia and N. Laoutaris, "A survey of data marketplaces and their business models," *ACM SIGMOD Record*, vol. 51, no. 3, pp. 18–29, 2022.

[8] M. E. van Genderen, M. Cecconi, and C. Jung, "Federated data access and federated learning: improved data sharing, AI model development, and learning in intensive care," *Intensive Care Medicine*, vol. 50, no. 6, pp. 974–977, 2024.

[9] J. Casaletto, A. Bernier, R. McDougall, and M. S. Cline, “Federated analysis for privacy-preserving data sharing: a technical and legal primer,” *Annual review of genomics and human genetics*, vol. 24, no. 1, pp. 347–368, 2023.

[10] V. Siska, V. Karagiannis, M. Drobics, and others, “Building a dataspace: Technical overview,” *Gaia-X Hub Austria*, 2023.

[11] V. Karagiannis, G. Simhandl, D. ž. Ignjatović, B. Bürger, and M. Drobics, “The role of data spaces in the Internet of Things,” *e+ i Elektrotechnik und Informationstechnik*, vol. 142, no. 3, pp. 221–227, 2025.

[12] G. Carovano and M. Finck, “Regulating data intermediaries: The impact of the Data Governance Act on the EU's data economy,” *Computer Law & Security Review*, vol. 50, p. 105830, 2023.

[13] P. Koen, M. Kollenstart, J. Marino, and others, “Dataspace Protocol 2025-1-RC4.” [Online]. Available: https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/

[14] M. Neubauer *et al.*, “Architecture for manufacturing-X: Bringing asset administration shell , eclipse dataspace connector and OPC UA together,” *Manufacturing Letters*, vol. 37, pp. 1–6, 2023.

[15] F. Schöppenthau *et al.*, “Building a digital manufacturing as a service ecosystem for Catena-X,” *Sensors*, vol. 23, no. 17, p. 7396, 2023.

[16] S. Forte, F. Gast, T. Dickopf, C. Sturmer, and M. Weber, “Gaia-X based Data Ecosystem Offerings in the Context of Data Management Systems,” in *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2024, pp. 1–6.

[17] A. Bertagnolli, P. Koen, M. Kollenstart, and others, “Eclipse Decentralized Claims Protocol v1.0-RC4.” [Online]. Available: https://eclipse-dataspace-dcp.github.io/decentralized-claims-protocol/v1.0-RC4/

[18] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, and C. Allen, “Decentralized Identifiers (DIDs) v1.1.” [Online]. Available: https://www.w3.org/TR/did-1.1/

[19] S. Manu, L. Dave, C. David, and H. Ivan, “Verifiable Credentials Data Model v2.0.” [Online]. Available: https://www.w3.org/TR/vc-data-model-2.0/

[20] A. Preukschat and D. Reed, *Self-sovereign identity*. Manning Publications, 2021.

http://dx.doi.org/10.25673/121097

Secure Engineering of Ethernet TSN Networks for Resource Constrained Industrial Devices

Janis Albrecht[1], Kevin Heubacher[1], Alexander Biendarra[1]

Abstract:

In this paper a TSN-configuration scenario for a resource-constrained embedded device is investigated. The IEC/IEEE 60802 TSN Profile for Industrial Automation defines the use of NETCONF/YANG for device configuration and suggests the use of a NETCONF proxy to configure resource-constrained devices. A concept for how such a proxy may work is defined, using COSE as the message format to carry CBOR encoded configuration information for resource-constrained devices. The COSE message is communicated over the Ethernet protocol.

Keywords: Resource-constrained Device, NETCONF, COSE

1 Introduction

Ethernet Time Sensitive Networking (TSN) offers a toolbox of standards to enable real-time communication in the same network infrastructure as information technology. TSN-mechanisms such as IEEE 802.1Qbu (Frame Preemption) or IEEE 802.1Qbv (Time Aware Shaping), and others, utilize frame prioritization in a time-synchronized network [1]. The IEC/IEEE 60802: Time Sensitive Networking Profile for Industrial Automation aims to reduce heterogeneity in the configuration and utilization of TSN standards by creating a top-down TSN profile for the industrial automation domain. One key aspect for the adoption of TSN standards by the industrial automation industry is the required complex configuration efforts in future TSN-enabled networks. For this purpose, the IEC/IEEE 60802 specification mandates the fully centralized TSN configuration model in combination with the NETCONF protocol for remote network management, YANG data models and Transport Layer Security (TLS) for authentication [2]. NETCONF is a connection-oriented, session based remote network management protocol, succeeding the Simple Network Management Protocol (SNMP). Its messages are encoded in utf-8 XML data format. The utilization of NETCONF and YANG introduces overhead in the form of required static and dynamic memory. This originates from the required transport protocol, commonly a TCP/IP stack, from the XML encoding requiring decoding, encoding and buffering of data, as well as the Network Management Datastore Architecture (NMDA) that is used to store multiple YANG data models for different purposes at once. In addition, using TLS for authentication requires additional computational resources of field devices. This poses a challenge for resource-constrained embedded devices (RCED) as they are limited in memory capacity and computational resources. The IEC/IEEE 60802

acknowledges this and proposes a solution by using a NETCONF proxy. The proxy is intended to run on a more capable device and connects to RCEDs to configure the devices in an implementation specific manner. At the time of writing, there is no specification or suggestion on how the proxy communicates with the RCEDs and how the security functions are to be implemented. The motivating question is therefore: **How can a secure, resource-saving remote management connection between a NETCONF proxy and RCEDs be implemented in accordance with IEC/IEEE 60802?** A promising candidate is the CBOR Object Signing and Encryption (COSE) protocol.

This paper is structured as follows: Chapter 2 shows a state of the art with focus on network configuration protocols and COSE. In chapter 3 a concept for a NETCONF proxy is defined. The paper is concluded by a short summary and required future work.

2 State of the Art Network Configuration Protocols

2.1 Remote Service Interface

The Remote Service Interface (RSI) is a configuration protocol defined in [3] and used for acyclic services (non–time-critical tasks such as parameterization, diagnostics, or commissioning) in Profinet systems. It allows an initiator to call a service on the responder. Usually, the initiator is a controller and the responder an IO device. RSI is introduced as an Application Service Element (ASE), that uses Ethernet as its transport protocol.

2.2 COSE Protocol

Concise Binary Object Representation (CBOR) is a data format defined in [4]. CBORs underlying data model is an extension of the JSON data model. Data is encoded to or decoded from a byte string carrying encoded data items. The first three bits of an encoded data item identify its type. There are 8 possible types: signed and unsigned integers, byte strings, text strings, arrays, tags, floating-point numbers and a map of data pairs. Each data pair consists of a key and its value.

The CBOR Object Signing and Encryption (COSE) protocol defines a way of carrying CBOR encoded information securely [5]. For this purpose, it specifies how to create and process signatures and encryption using CBOR for serialization. COSE messages are encoded into a CBOR array. The different types of COSE messages are identified with a tag at the beginning of the message. The array after the tag always contains 3 fields: protected headers, unprotected headers, and a payload. Protected headers are encoded as a map and wrapped in a byte string. Unprotected headers are also encoded as a map but not wrapped in a byte string. Both fields contain security parameters. Payload encoding is defined by the application. Subsequent fields are dependent on the message type.

Communication between RCED and NETCONF Proxy would be realized with the help of cose-sign1 messages. This message type is identified with tag 18. In addition to the 3 static fields, it contains a signature field, in which the payload signature is placed to validate the payload integrity. The signature is encoded in a byte string. The contents of a message can be described in the extended CBOR diagnostic notation (defined in [6]). Using that notation, a cose-sign1 message can be written as follows [5]:

```
18(
     [
       / protected / h'a10126' / {\ alg \ 1:-7 \ ECDSA 256 \} /,
       / unprotected / {/ kid / 4:'11'},
       / payload / 'This is the content.',
       / signature / h'8eb33e4ca31d1c465ab05aac34cc6b23d58fef5c083106c4d25a
       91aef0b0117e2af9a291aa32e14ab834dc56ed2a223444547e01f11d3b0916e5a4c3
       45cacb36'
     ]
   )
```

First, the message type is identified via tag 18. Rounded brackets indicate that the following information is part of that tag. Square brackets indicate that information is encoded as a CBOR array. Comments are enclosed by slashes. Protected header parameters are wrapped in a CBOR byte string, which are prefixed by “h” for hexadecimal notation, followed by their contents enclosed by single quotation marks. Here, the protected header contains the “alg” parameter. It indicates what algorithm is used for security functions. In this case, ECDSA 256 was used to generate the signature. Inside the unprotected header is a “kid” parameter, which signals the receiving device which key to use with the ECDSA 256 algorithm to validate the signature. It is encoded as a CBOR map, which is indicated by curly brackets. Following it are payload and signature. In this example, the payload is encoded as a text string.

2.3 NETCONF/YANG

The IEC/IEEE 60802 uses the communication protocol NETCONF in combination with YANG modules for device configuration. A NETCONF client can perform remote procedure calls (RPCs) to retrieve, upload and edit configuration information of network devices that contain a NETCONF server. RPCs are encoded in XML and use <rpc> and <rpc-reply> elements to frame NETCONF requests and responses. If a client sends an <rpc> request to a server, an <rpc-reply> response is expected. Both elements contain a mandatory “message-id” field, to associate a request with a response. If an error occurs during the processing of a request, the response contains an <rpc-error> element. If no data is returned and no error occurred, the response contains the <ok> element.

NETCONF servers execute operations that are contained in <rpc> request. The supported operations are defined in 6.3.2.2 of [2]. For example, a request can contain a <get-config> or <edit-config> operation. A <get-config> operation causes the server to respond with the complete or partial configuration information of a configuration datastore. The <edit-

config> operation causes the server to edit all or parts of the configuration information in a configuration datastore.

NETCONF supports one or more configuration datastores. A datastore contains all the necessary configuration information of a device. A <running> datastore is always present and contains the devices currently active configuration. In addition to the <running> datastore, the IEC/IEEE 60802 specifies that <candidate> and <startup> datastores can also be used. The <candidate> datastore holds configuration information that can be edited without impacting the device's current configuration information. Configuration information contained in a <candidate> datastore can be activated and moved to the <running> datastore with the <commit> operation. The <startup> datastore initializes the configuration of a device when it boots.

Configuration data is modeled using the YANG data modeling language. YANG modules define a hierarchy of data that NETCONF operations can be performed on. The YANG modules used for device configuration in IEC/IEEE 60802 are listed in table 20 of [2].

2.4 Overview of configuration protocols

Table 2 provides an overview of the configuration protocols discussed in this chapter. It compares each of the protocols in regard to some of their aspects.

Protocol / Aspect	RSI	NETCONF	Netconf Proxy Concept with COSE
Domain	Industrial Automation with Profinet	Telecommunications, enterprises and increasingly industrial automation	IEC/IEEE 60802 based TSN-enabled networks
Primary Purpose	Acyclic communication in Profinet networks	Remote device configuration and state information	Configuration of RCEDs
Transport Protocol	Ethernet Protocol	TCP/IP	Ethernet Protocol
Standard Origin	IEC FDIS 61158-6-10	IETF (RFC 6241)	IETF (RFC 8152)
Data Encoding	Binary	XML	CBOR
Security Model	No security	Security via SSH	Security via TLS
RCED Resource Requirements	Low (no TCP/IP Stack, no security)	High (XML parsing, TCP/IP stack, security services)	Medium (CBOR encoding/decoding, security services)

Tab.2: Comparison of configuration protocols

3 NETCONF Proxy Concept

The information exchange necessary for a NETCONF proxy to operate can be divided into three distinct tasks:

1. Communication between NETCONF client and server
2. Translation between NETCONF requests and RCED configuration protocol
3. Communication between NETCONF proxy and RCED

Figure 2 depicts the communication relation between the three devices with COSE chosen as the RCED's configuration protocol. Communication between NETCONF client and server is mostly defined in IEC/IEEE 60802. However, some open issues

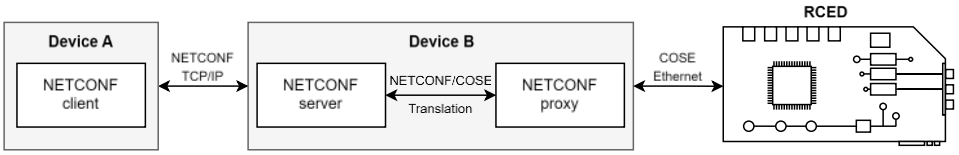


Fig.1: Logical representation of information exchange between Device A with NETCONF client, Device B with NETCONF server and proxy, and RCED

exist. For example, it is not defined how the NETCONF client learns of a present proxy and what devices it represents. Furthermore, both NETCONF client and server need a way to differentiate between the proxy's configuration information and the configuration information of each of the devices it represents. Solutions to these problems are proposed in [7]. The source also proposes a solution for RCEDs to find proxies that represent them via LLDP.

To retrieve and edit the configuration of RCEDs it is necessary that the NETCONF proxy translates between NETCONF messages and the chosen RCED's configuration protocol. Regardless of the chosen configuration protocol, every implementation of the translation process needs to determine what subset of NETCONF requests require a translation in the first place. Given that there is a mechanism by which the configuration information target can be determined, those XML based requests then need to be translated into the protocol that configures the RCED and addressed to that specific device.

NETCONF operation	Operation description	RCED communication
get	Retrieves device state information and running configuration	Required if <running> or <startup> datastore is targeted
get-config	Retrieves all or part of a configuration datastore	Required if <running> or <startup> datastore is targeted

edit-config	Loads all or part of a configuration to the specified configuration datastore	Required if <running> or <startup> datastore is targeted
copy-config	Creates or replaces a configuration datastore	Required if <running> or <startup> datastore is targeted
commit	Copies configuration in <candidate> datastore to <running> configuration	Required
discard-changes	Reverts contents of the <candidate> datastore to contents of the <running> datastore	Not necessary
delete-config	Deletes a configuration datastore	Not used in IEC/IEEE 60802
lock	Locks configuration datastore from access from other NETCONF clients	Not necessary
unlock	Releases configuration lock	Not necessary
close-session	Requests for termination of NETCONF session	Not necessary
kill-session	Forces termination of NETCONF session	Not used in IEC/IEEE 60802

Tab.1: A list of NETCONF operations and their descriptions. For each operation it is determined if communication with the RCED is necessary

It is assumed that the communication between the NETCONF server inside the NETCONF proxy and the NETCONF client would function as defined in [8]. That means a NETCONF proxy would answer inbound <rpc> requests as usual with a <rpc-reply>. Dependent on the operations contained in an <rpc> request and targeted datastore it may be necessary to communicate with the RCED. Table 2 lists possible NETCONF operations and determines if communication with the RCED is necessary.

Whenever configuration information inside the <running> and <startup> datastores is requested or edited, communication with the RCED is necessary. Configuration edits need to be communicated, because the RCED needs to be informed on what its configuration is going to be, so that it can reconfigure itself. Configuration requests need to be communicated, because the NETCONF proxy needs to be made aware of the RCED's current configuration. Figure 2 displays the necessary steps for processing a NETCONF request inside a device that also acts as a NETCONF proxy.

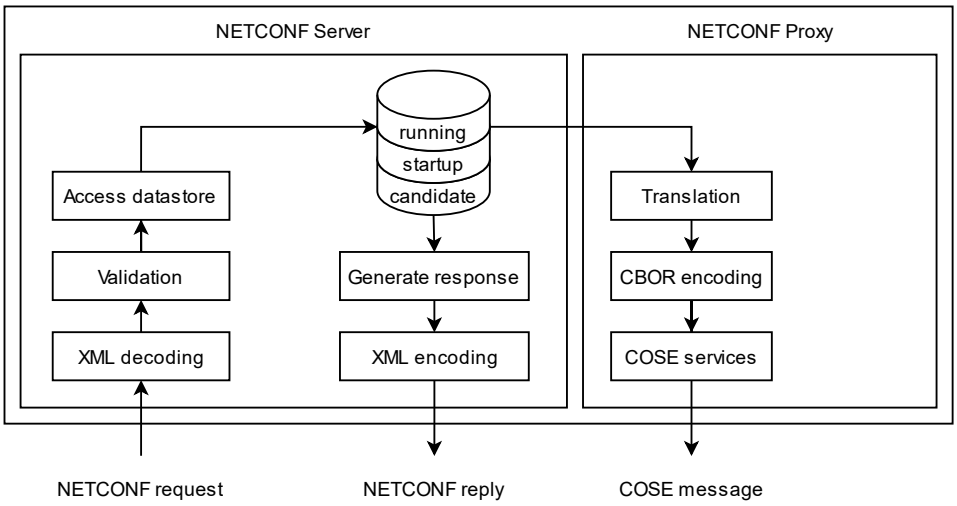


Fig.2: NETCONF request processing steps of a device that also acts as a NETCONF proxy with COSE chosen for RCED communication

Incoming well-formed NETCONF requests are decoded into an in-memory data tree representation of the contained XML. Then, the request is tested for validity. It has to be verified if the NETCONF client user is authorized to perform contained operations or access the targeted datastores. Furthermore, contained configuration information is compared to the YANG module definition. If the validity checks are successful, the targeted datastore is accessed and configuration data is either edited or retrieved. Once the datastore queries are complete, a NETCONF reply is generated to inform the client of the successful NETCONF request processing. Should any of the steps fail, the NETCONF server replies with an error.

If the NETCONF request targets a <running> or <startup> datastore and its processing is successful, a message needs to be sent to the RCED that had its configuration datastores edited. Building that message requires information on what data has been edited and the new values for that data. In addition to that, it is necessary to include the NETCONF requests message id, so that a reply by the RCED can be traced back to the original request by the NETCONF proxy. To identify the targeted configuration information and operation, the YANG module name and NETCONF operation is also included. The resulting output would then be CBOR encoded and set as the payload of a COSE message.

The Information exchange between NETCONF proxy and RCED would be based on layer 2 and the COSE protocol. To enable the processing and constructing of frames carrying configuration information, it is necessary to define a shared message syntax. This includes:

1. An EtherType value inside the Ethernet Header
2. A message format for the COSE payload

3. A mapping of the received COSE payload to status and configuration information of the receiving device

The value for EtherType can be any of the unused EtherTypes, for example 0x8FFF. Its purpose is to differentiate an Ethernet frame carrying a CBOR encoded byte string from other Ethernet frames. The frame handler of the respective device can then hand the frame to the application that deals with processing the COSE message. Figure 3 displays the information contained in a cose-sign1 message and the information in its payload.

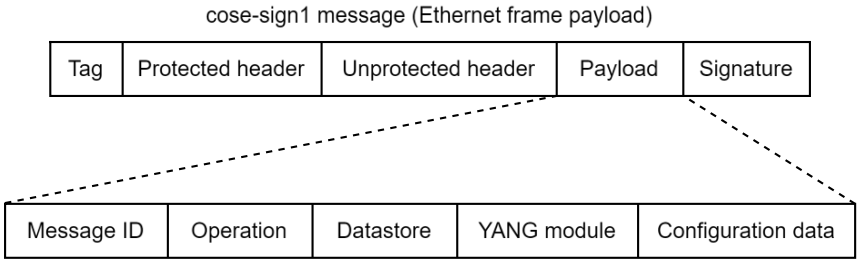


Fig.3: Fields of a cose-sign1 message and the format of its payload for RCED configuration

The fields of a cose-sign1 message are detailed in chapter 2. It is used to provide data integrity via signature generation. How the required keys are exchanged is out of scope of this paper. Following is a simple example of an XML-based NETCONF request and the resulting COSE message that a NETCONF proxy based on this concept sends to an RCED. The edited configuration information is based on the YANG module ieee1588-ptp-tt. The <rpc> request has a message-id value of 10 and performs an <edit-config> operation that targets the <running> datastore. Contained configuration information is inside the <config> element. The ptp instance with index 1 is assigned values for priority1 and priority2 fields.

```
<rpc message-id="10" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>
    <config xmlns="urn:ieee:std:1588:yang:ieee1588-ptp-tt">
      <ptp>
        <instances>
          <instance>
            <instance-index>1</instance-index>
            <priority1>245</priority1>
            <priority2>250</priority2>
          </instance>
        </instances>
      </ptp>
    </config>
  </edit-config>
</rpc>
```

The NETCONF proxy processes the shown NETCONF request and updates the <running> datastore with the contained configuration information. If processing is successful, the proxy sends the following COSE message via Ethernet protocol to the RCED that had its <running> datastore updated. Protected header, unprotected header and signature are identical to the example used in chapter 2 and are only included for completeness. The payload of a COSE messages is encoded into a CBOR array, each of the array fields semantic is shown in figure 3. Configuration data inside the payload is encoded into a CBOR map.

```
18(
     [
       / protected / h'a10126',
       / unprotected / {4:'11'},
       / payload / [10, 'edit', 'running', 'ieee1588-ptp-tt', {'instance-
       index':1,'priority1':245, 'priority2':250}],
       / signature / h'8eb33e4ca31d1c465ab05aac34cc6b23d58fef5c083106c4d25a
       91aef0b0117e2af9a291aa32e14ab834dc56ed2a223444547e01f11d3b0916e5a4c3
       45cacb36'
     ]
   )
```

4 Summary and Future Work

This paper investigated a TSN-configuration scenario of a resource-constrained embedded device. In chapter 2, the configuration protocols NETCONF and RSI were explored and compared with the use of COSE for configuration. Then, a basic concept for a NETCONF proxy, using COSE as the configuration protocol for RCEDs was defined. Afterwards, it is determined what NETCONF requests require communication with the RCED and how a received request is processed by a NETCONF proxy. Finally, a payload structure for the COSE message is defined and a simple example of how an XML-based request may be translated into a CBOR-based payload is shown.

In future work, the mapping of YANG modules to configuration information inside exchanged COSE messages may be defined. In the example they were included via the name of a configuration element encoded as a string and the associated value. In reality, the configuration information carried in NETCONF request and COSE messages will be larger in size, potentially exceeding maximum frame size. In that case a fragmentation mechanism is required. Furthermore, it hasn't been defined if and how COSE messages received by RCEDs are answered, which will be necessary for answering <get-config> requests.

5 References

[1] Institute of Electrical and Electronics Engineers (IEEE), 2025. [Online]. Available: https://1.ieee802.org/tsn/. [Zugriff am 24 04 2025].

[2] IEC and IEEE, „IEC/IEEE 60802 Time-Sensitive Networking Profile for Industrial Automation - Draft 3.2," 2025.

[3] PROFIBUS Nutzerorganisation, „IEC FDIS 61158-6-10," 2023.

[4] Internet Engineering Task Force (IETF), „RFC 8949," 2020. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8949. [Zugriff am 25 04 2025].

[5] Internet Engineering Task Force (IETF), „RFC 8152," 2017. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8152. [Zugriff am 24 04 2025].

[6] Internet Engineering Task Force (IETF), „RFC 8610," 2019. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8610. [Zugriff am 25 04 2025].

[7] M. Hantel, M. Kießling, C. Mangin, M. Stanica und G. Steindl, 18 10 2024. [Online]. Available: https://www.ieee802.org/1/files/public/docs2024/60802-steindl-et-al-extended-onboarding-model-1024-v03.pdf. [Zugriff am 24 04 2025].

[8] Internet Engineering Task Force (IETF), „RFC 6241," 2025. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc6241. [Zugriff am 25 04 2025].

[9] Internet Engineering Task Force (IETF), „RFC 8342," 2025. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8342. [Zugriff am 24 04 2024].

http://dx.doi.org/10.25673/121097

Messung und Evaluierung der Mobilfunkversorgung für die Konnektivität von UAS im BVLOS-Flugbetrieb

Andreas Hecker[1], Timon Slowik[2], Ana Belén Martínez[1] und Gerhard Fettweis[1]

Abstract: Der sichere Betrieb unbemannter Luftfahrtsysteme (Unmanned Aircraft System, UAS) während teilautonomer oder missionsbasierter Flüge im BVLOS (Beyond Visual Line Of Sight) erfordert eine permanente Kontroll- und Eingriffsfähigkeit und damit eine kontinuierliche Datenübertragung. Mobilfunknetze stellen hierfür eine flächendeckend vorhandene Lösung dar. Die zugehörige Planung, Messung und Optimierung erfolgt klassisch auf 1,5 m über Grund. Zugehörige Ausbreitungsmodelle wurden über Jahrzehnte angepasst. Zur Netzabdeckung und Signalgüte im unteren Luftraum (Very Low Level Airspace, VLL) bis 150 m über Grund existieren zwar veröffentlichte Erfahrungswerte, stellen aber noch keine flächendeckende, systematische Datenlage dar, insbesondere nicht in ländlichen Gebieten. Dies erschwert eine verlässliche Planung und Risikobewertung von Flügen in diesen Höhenlagen. Im Rahmen des BMDV-Projekts InnoDCon (Innovative Drone Connectivity) fanden erstmals im VLL umfangreiche, flächendeckende Messflüge auf einem Erprobungsgelände in der Lausitz mit einer Gesamtfläche von 28 km^2 statt. Mittels der erhobenen Daten wurden die Kommunikationsverbindungen hinsichtlich ihrer anwendungsbezogenen Eignung bewertet. Ergebnisse zeigen trotz schlechterer Signalqualitäten in höheren Lagen die prinzipielle Umsetzbarkeit der Anwendungen, wenngleich zur Stabilitätssicherung der Verbindungen auf Redundanzkonzepte gesetzt werden sollte.

Keywords: Unbemannte Luftfahrtsysteme (UAS), BVLOS-Flugbetrieb, Konnektivität im unteren Luftraum (VLL), Mobilfunkmessung, 5G- und LTE-Netze, U-Space-Integration

1 Einleitung

Die Nutzung unbemannter Luftfahrtsysteme (Unmanned Aircraft Systems, UAS) hat in den letzten Jahren erheblich zugenommen. Zum 31. Dezember 2024 waren in Deutschland über 700.000 UAS-Betreiber beim Luftfahrt-Bundesamt (LBA) registriert. 627 Betriebsgenehmigungen wurden in der sogenannten speziellen Kategorie erteilt, die Flugbetriebe mit erhöhtem Risiko, etwa über besiedeltem Gebiet, mit einer Flughöhe über 120 m über Grund (Above Ground Level, AGL) oder außerhalb der Sichtweite des Piloten (Beyond Visual Line of Sight, BVLOS) umfasst.

Für den sicheren Betrieb von UAS im BVLOS-Modus ist eine kontinuierliche Kommunikationsverbindung zwischen Luftfahrzeug und Bodenstation zwingend erforderlich. Nur durch eine stabile Datenübertragung können Steuerkommandos, Telemetriedaten und

[1] TU Dresden, Vodafone Stiftungsprofessur für Mobile Nachrichtensysteme, Barkhausenbau, 01069 Dresden, Deutschland, andreas.hecker@tu-dresden.de; ana_belen.martinez_monton@tu-dresden.de; gerhard.fettweis@tu-dresden.de

[2] AEF - Autonom Elektrisch Fliegen gGmbH, Flugplatz Kamenz, Zum Tower 6, 01917 Kamenz, Deutschland, timon.slowik@aef.aero

Sicherheitsmeldungen zuverlässig ausgetauscht werden. Klassische Fernsteuerungssysteme stoßen hierbei an ihre physikalischen und funktionalen Grenzen, da ihre Reichweite auf Sichtverbindung beschränkt ist und keine durchgängige Netzabdeckung gewährleistet werden kann. Öffentliche Mobilfunknetze bieten demgegenüber auch für überregionale BVLOS-Operationen eine flächendeckende, standardisierte und sichere Kommunikationsinfrastruktur.

Kommerzielle Mobilfunknetze werden für bodennahe Anwendungen (in etwa 1,5 m über dem Boden) geplant und optimiert. Die Kalibrierung der zugrunde liegenden Funkausbreitungsmodelle basiert auf entsprechenden Messdaten, die im Rahmen von Walk- und Drive-Tests erhoben werden. Für den unteren Luftraum – den sogenannten Very Low Level Airspace (VLL) bis 150 m AGL – existieren bislang nur vereinzelte Erfahrungswerte, jedoch keine systematisch erhobene, flächendeckende Datenbasis. Dies erschwert sowohl die Risikoanalyse als auch die Einsatzplanung für Drohnenflüge im BVLOS-Betrieb, insbesondere in ländlichen Regionen mit geringer Basisstationsdichte.

Im Rahmen des vom Bundesministerium für Digitales und Verkehr (BMDV) geförderten Projekts InnoDCon (Innovative Drone Connectivity) wurden erstmals umfassende Messflüge im VLL durchgeführt, um die Eignung bestehender Mobilfunknetze für UAS-Anwendungen zu bewerten. Das Erprobungsgelände „AERO-Lausitz“ in Ostsachsen diente dabei als Testgebiet. Ziel war die Erfassung und Bewertung von Netzabdeckung, Signalqualität und Datenraten im Kontext typischer UAS-Anwendungsfälle. Die gewonnenen Ergebnisse bilden eine Grundlage für die Weiterentwicklung sicherer, verlässlicher Kommunikationsstrukturen im europäischen U-Space.

Der Beitrag ist wie folgt aufgebaut: Abschnitt 2 gibt einen Überblick über den aktuellen Stand der Forschung und die relevanten 3GPP-Standardisierungsaktivitäten. Abschnitt 3 fasst die betrachteten Anwendungsfälle zusammen. Abschnitt 4 beschreibt das Messkonzept, die eingesetzte Methodik und das Untersuchungsgebiet der durchgeführten Feldversuche. Die Messergebnisse werden in Abschnitt 5 hinsichtlich anwendungsrelevanter Schwellenwerte und Leistungsanforderungen bewertet. Abschnitt 6 fasst die wichtigsten Erkenntnisse zusammen und gibt einen Ausblick auf zukünftige Forschungsarbeiten im Kontext der U-Space-Integration.

2 Stand der Forschung

Mobilfunknetze gelten aufgrund ihrer flächendeckenden Verfügbarkeit, etablierten Sicherheitsmechanismen und standardisierten Dienstarchitekturen als vielversprechende Kommunikationsinfrastruktur für UAS. Da bestehende Netze jedoch ursprünglich für bodennahe Anwendungen konzipiert wurden, ergeben sich beim Einsatz von UAS neue Herausforderungen hinsichtlich Ausbreitung, Interferenzmanagement und Mobilität. In den letzten Jahren haben Standardisierungsgremien sowie die Forschungsgemeinschaft eine Vielzahl von Arbeiten vorgelegt, die diese speziellen Anforderungen adressieren.

Das 3rd Generation Partnership Project (3GPP) hat die Unterstützung von UAS in mehreren Technical Reports (TRs) und Technical Specifications (TSs) thematisiert. Der Bericht [TR36.777] legte den Grundstein für die Untersuchung des Interferenzverhaltens von Aerial User Equipments und zeigte, dass die Sichtbarkeit mehrerer Basisstationen in höheren Luftlagen zu einem Abfall des Signal-to-Interference-plus-Noise-Ratio (SINR) führt. Aufbauend darauf definieren [TR22.829; TR23.754; TR23.755] Anforderungen und Systemkomponenten für die Unterstützung von UAS in Mobilfunknetzen, die in verbindlichen Spezifikationen wie [TS22.125; TS23.255; TS29.255; TS29.256] umgesetzt wurden. Diese Dokumente legen Rahmenbedingungen für Identifikation, Autorisierung, Quality-of-Service (QoS)-Parameter und Schnittstellen zu externen *U-Space*-Diensten fest.

Empirische Untersuchungen zur Konnektivität zu einem unbemannten Luftfahrzeug (Unmanned Aerial Vehicle, UAV) lassen sich grob in zwei Ansätze einteilen: (i) aktive Endgerätetests und (ii) passive, scannerbasierte Messungen. Aktive Verfahren nutzen handelsübliche Endgeräte mit spezialisierter Software (z. B. TEMS, QualiPoc), um reale Datenverbindungen und Leistungsparameter zu erfassen. Passive Messungen erfassen dagegen Broadcast- und Signalisierungsdaten über Netzwerkscanner (z. B. Rohde & Schwarz TSMA-Serie). Frühere Studien konzentrierten sich häufig auf einzelne Netzbetreiber und LTE-Technologien [Am17; Li19], während neuere Arbeiten Multi-Operator- und 5G-Umgebungen einbezogen [Br24; Gh23; GNA21; Ka23; Sä20; Si23]. Die zentrale Bewertungsgrundlage bilden dabei die Key Performance Indicators (KPIs) Empfangsleistung des Referenzsignals (Reference Signals Received Power, RSRP), SINR, Latenz, Datendurchsatz sowie Verbindungsstabilität.

Unabhängig von der Methode zeigen alle Arbeiten konsistente Trends: Mit zunehmender Flughöhe steigt in der Regel der RSRP, da Abschattungen entfallen und Sichtverbindungen dominieren. Gleichzeitig verschlechtert sich jedoch das SINR, da mehrere benachbarte Zellen sichtbar werden und Interferenz verursachen. Dieses Verhalten beeinflusst unmittelbar Durchsatz, Latenz und Handover-Stabilität [Ca19; Wa24].

Mit der Einführung von 5G rückten Untersuchungen zu Multi-Operator-Szenarien und Netzarchitekturen in den Vordergrund. Studien wie [Ba24] zeigen, dass 5G-Netze durch neue Frequenzbänder und Beamforming zwar höhere Spitzenraten ermöglichen, jedoch auch empfindlicher auf Interferenz und Zellwechsel reagieren. Der Einsatz mehrerer Netzbetreiber kann die Ausfallsicherheit erhöhen, setzt jedoch regulatorische und technische Koordination voraus [Br24; GNA21].

Standardisierung und empirische Untersuchungen zeigen damit, dass Mobilfunknetze prinzipiell für den Einsatz in BVLOS-UAS geeignet sind, wobei offene Fragen bestehen bleiben. Während sich die meisten Studien auf urbane Umgebungen konzentrieren, sind flächendeckende, systematische Daten für den VLL Airspace in ländlichen Gebieten bislang kaum verfügbar. Zudem kombinieren nur wenige Untersuchungen aktive und passive Verfahren unter identischen Bedingungen, wodurch Vergleichbarkeit und Validität der KPIs eingeschränkt bleiben.

3 UAV-Anwendungsfälle

Die Kommunikation zwischen einem UAV und der Bodenstation kann je nach Missionsprofil unterschiedliche Anforderungen an Latenz, Datenrate und Zuverlässigkeit stellen. In den 3GPP-Spezifikationen, insbesondere [TS22.125], werden diese Anforderungen über definierte Use Cases und zugehörige KPIs beschrieben. Drei Hauptanwendungsfälle werden für die Bewertung der Messergebnisse in Abschnitt 4 betrachtet: Command & Control (C2), Telemetrie und Video-Streaming.

C2-Verbindungen dienen der Übertragung sicherheitskritischer Steuer- und Kontrollkommandos zwischen UAV und Fernpilot. Für diese Kategorie ist eine extrem hohe Zuverlässigkeit bei minimaler Latenz erforderlich, um eine kontinuierliche Steuerbarkeit zu gewährleisten. Telemetrie umfasst den regelmäßigen Austausch von Status- und Sensordaten des UAV, beispielsweise Position, Geschwindigkeit, Batteriestatus oder Missionsfortschritt. Diese Datenpakete sind klein, erfordern jedoch eine periodische und stabile Übertragung. Der Anwendungsfall Video-Streaming stellt die höchsten Anforderungen an Bandbreite und Stabilität. Für Echtzeit-Übertragungen (z. B. visuelle Inspektion, Search-and-Rescue) werden in der Regel Datenraten zwischen <1 (periodic photos) und 9 Mbit/s (1080p Video) benötigt, abhängig von Kompression, Auflösung und Bildfrequenz. 8K-Video würde sogar 100 Mbit/s bedeuten.

4 Messkampagne im U-Space Lausitz

Die im Rahmen des Projekts *InnoDCon* durchgeführte Messkampagne hatte das Ziel, die tatsächliche Leistungsfähigkeit kommerzieller Mobilfunknetze für den BVLOS-Betrieb von UAS unter realen Bedingungen zu untersuchen und die Mobilfunkversorgung im Höhenbereich von bis zu 150 m AGL flächendeckend zu erfassen. Die Messungen fanden auf dem Erprobungsgelände *AERO-Lausitz* in der Oberlausitz (Ostsachsen, Deutschland) statt (s. Abbildung 1). Das Gelände verfügt über eine Fläche von rund 28 km^2 und dient als dedizierter Testbereich für BVLOS-Flüge. Eine Erweiterung auf etwa 500 km^2 wurde beantragt und befindet sich im Genehmigungsprozess beim Luftfahrt-Bundesamt (LBA). Der Bereich ist dabei von ländlichen Regionen mit geringer Basisstationsdichte geprägt. Dies stellt die dort typische Herausforderung für die UAS-Konnektivität dar.

Für die Messflüge kam ein automatisiert betriebenes UAS des Typs *Nokia Drone Networks (NDN)* zum Einsatz, das für BVLOS-Betrieb ausgelegt ist. Die Drohne verfügte über eine maximale Nutzlast von 3 kg und wurde mit Messequipment ausgestattet, das eine Kombination aus aktiven und passiven Messverfahren bietet, um ein vollständiges Bild der Netzabdeckung und -qualität zu gewinnen. Als passives Messsystem ein Netzwerkscanner *Rohde & Schwarz TSMA6B* mit GPS und Breitbandantenne zur Erfassung der Mobilfunkversorgung über die Auswertung der kontinuierlich ausgesendeten Broadcast- und Signalisierungsinformationen. Die Messung umfasste Kennwerte wie RSRP, SINR, Zell-IDs und die Anzahl sichtbarer Zellen pro Technologie (2G/GSM, 4G/LTE, 5G/NR SA

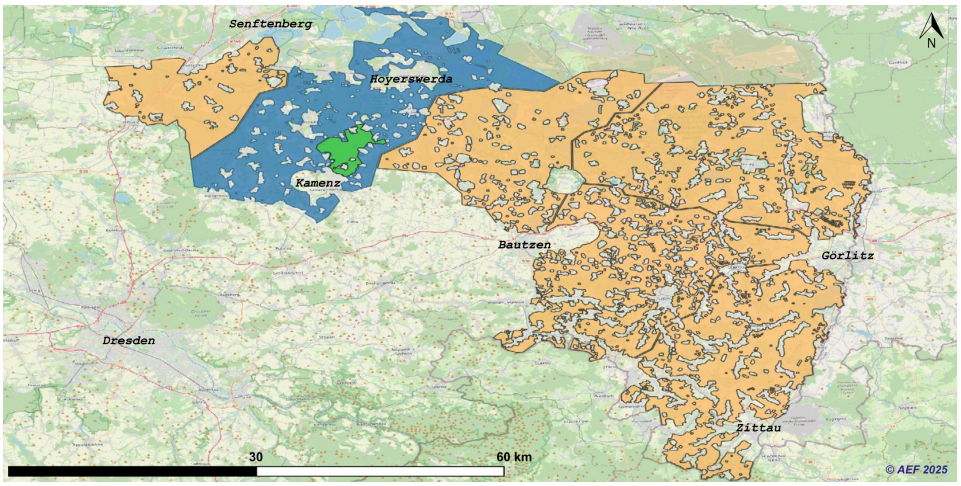


Abb. 1: Untersuchungsgebiet im Raum Kamenz/Hoyerswerda. In Grün: das zur Projektzeit genehmigte Testfeld (28 km^2); in Blau: das beantragte Erweiterungsgebiet (500 km^2) mit voraussichtlicher Genehmigung Ende 2025; in Orange: das vollständig angestrebte U-Space-Gebiet in Sachsen (3000 km^2) mit Zielzeitraum 2030+.

und NSA) und Betreiber. Die Energieversorgung des TSMA erfolgte über den Drohnenakku. Als aktives Messsystem wurde ein Smartphone *Samsung S23* mit installierter Messsoftware *QualiPoc* verwendet, ebenfalls von Rohde & Schwarz. Über eine handelsübliche SIM-Karte erfolgte eine Verbindung zum ausgewählten Netz von Vodafone. Zur Latenzerfassung wurden periodische Ping-Tests ausgeführt sowie iPerf-Tests zur Bestimmung der maximal erreichbaren Datenrate. Bezüglich der in Abschnitt 3 beschriebenen Anwendungsfälle wurde die UL-Richtung für die Tests ausgewählt. Die Kombination beider Verfahren erlaubt eine direkte Korrelation zwischen physikalischen Signalparametern und anwendungsrelevanten Performance-Kennzahlen. Die gesammelten Daten bilden die Grundlage für eine quanti-

Abb. 2: Nokia-Drohne mit Netzwerkscanner TSMA6B als Payload.

Tab. 1: Übersicht der Messparameter im Projekt InnoDCon

Parameter	**Wert / Beschreibung**
Testgebiet	AERO-Lausitz, 28 km^2 (ländlich)
Flughöhe	100 m AGL
Fluggeschwindigkeit	10 $m\,s^{-1}$
Messsystem	TSMA6B (Scanner), S23 + QualiPoc (aktiv)
Technologien	2G (GSM), 4G (LTE), 5G (NR SA/NSA)
Betreiber	Multi-Operator (inkl. Campusnetz)
Messsoftware	SmartMonitor, SmartAnalytics
Messgröße	RSRP, SINR, Datenrate, Latenz

tative Bewertung der Netzqualität und die Überprüfung von Schwellenwerten für die in Abschnitt 3 beschriebenen UAV-Anwendungsfälle.

Die Steuerung und das Monitoring der Messsysteme erfolgten über die Software *R&S SmartMonitor*. Dabei wurden beide Messgeräte als Remote-Probes registriert, wodurch eine Live-Überwachung von Signalparametern, Flugstatus und Messfortschritt in Echtzeit möglich war. Während das Smartphone mit der Qualipoc-Lösung seine eigene Mobilfunkverbindung für die aktive Messung hatte,erfolgte die Verbindung der Drohne Während des Fluges zum Internet über eine Mobilfunkverbindung, die über LAN an den Netzwerkscanner weitergereicht wurde. Dadurch konnten sowohl der Scanner als auch die Drohne selbst remote gesteuert und überwacht werden. Alle aufgezeichneten Messdaten wurden nach Abschluss der Flüge auf einen zentralen FTP-Server übertragen und mit *R&S SmartAnalytics* offline ausgewertet. Mit diesem Prozess wurden zwei wesentliche Aspekte in der Luftraum-Vermessung abgedeckt: das Monitoring während des Flugbetriebs mit der zugehörigen Visualisierung kurzzeitiger KPIs sowie die Auswertung anhand langfristig aufgenommener Messdaten.

Die resultierende Flugzeit war auf etwa 25-30 Minuten pro Mission begrenzt, sodass das gesamte Gebiet durch mehrere Teilflüge erfasst werden musste. Die einzelnen Flugrouten wurden vorab halb-automatisiert geplant durch Angabe des abzudeckenden Bereichs, der Rastergröße und der Fluggeschwindigkeit. Jede Teilmission umfasste eine konstante Flughöhe von 100 m AGL, eine Fluggeschwindigkeit von 10 $m\,s^{-1}$, eine horizontale Rasterplanung zur gleichmäßigen Abdeckung des 28 km^2-Gebiets sowie vergleichbare Wetterbedingungen über alle Messtage (windarm, trocken). Die daraus entstandenen Missionsdateien wurden auf das UAS hochgeladen und autonom abgeflogen.

Tabelle 1 fasst die wesentlichen Parameter der Messkampagne zusammen.

5 Auswertung der Messergebnisse

Die Auswertung bestätigte bereits bekannte Effekte aus der Literatur: Während die Mobilfunkversorgung am Boden optimiert ist, steigt in größeren Höhen die Zahl der gleichzeitig sichtbaren Zellen, was zu erhöhter Interferenz führt (Abb. 3). Trotz dieses deutlich geringeren mittleren SINR konnten in über 95 % der abgeflogenen Fläche Datenraten von mehreren Mbit/s erzielt werden (Abb. 4). Die im Rahmen der Messkampagne erzielten Latenzzeiten lagen typischerweise zwischen 40 und 80 ms. Verbindungsabbrüche durch Handover-Fehler wurden nicht festgestellt. Insgesamt zeigte sich eine gute LTE- und 5G-Netzabdeckung. 2G als Rückfallebene fand dagegen nicht statt.

Die Messergebnisse zeigen, dass die im VLL erzielten Netzparameter für die meisten operationellen UAS-Anwendungen ausreichend sind. Die Datenraten ermöglichen C2- und Telemetrie-Anwendungen sowie komprimierte Videoübertragung bis zu 3 MMbit/s nahezu flächendeckend. Auch unter erhöhter Interferenz in 100 m Flughöhe blieben vollständige Verbindungsabbrüche aus. Zukünftige Messkampagnen über das wachsende Areal des U-Space Lausitz wird weitere Erkenntnisse bringen, die als Grundlage für die Planung zukünftiger BVLOS-Flüge sowie für die Integration von Mobilfunkdiensten in den europäischen *U-Space* dienen können. Dabei steht die nahtlose Integration von Mobilfunk-KPIs in die Dienste von U-Space-Service-Providern (USSP) und Common-Information-Service-Providern (CISP) im Fokus. Die Messdaten bieten eine wertvolle Datenbasis zur Kalibrierung von Funkausbreitungsmodellen, welche die Abschätzung der Mobilfunkversorgung innerhalb von UAS-Missionsplanungssoftware verbessern und damit eine netzqualitätsabhängige Routenoptimierung ermöglichen kann.

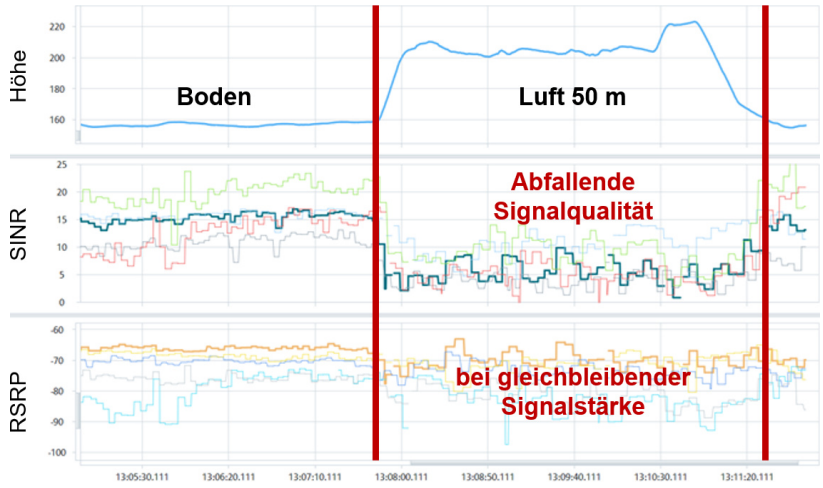


Abb. 3: Abnehmende Signalqualität in höheren Luftlagen infolge erhöhter Interferenz durch Mehrfachsichtverbindungen.

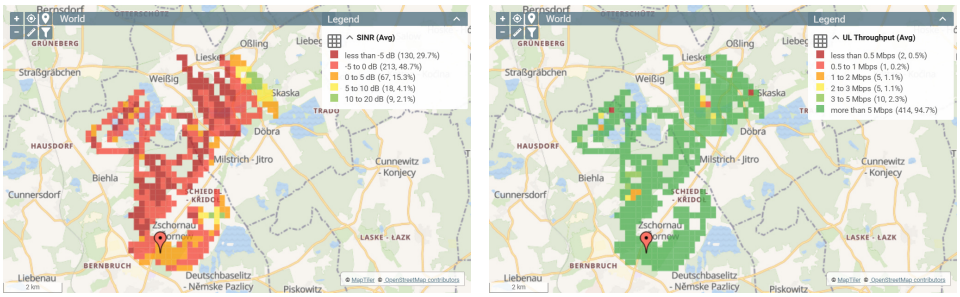


Abb. 4: Mittlerer Signal-Stör-Abstand (SINR) und erzielte Datenraten am Beispiel des Vodafone-Netzes bei 100 m AGL.

Weiterhin lassen sich für die UAV-Anwendungsfälle die lokal realisierbaren Schwellenwerte ableiten, die mit den standardisierten Größen bzw. den technisch möglichen verglichen werden können. Tabelle 2 fasst für jeden dieser Anwendungsfälle beispielsweise die ermittelten Werte zusammen, die im vermessenen Flugbereich entnommen werden konnten.

Insgesamt lässt sich derzeit festhalten, dass ein sicherer BVLOS-Betrieb über öffentliche Mobilfunknetze prinzipiell realisierbar ist, d. h. öffentliche Mobilfunknetze können die funktionalen Anforderungen für C2- und Telemetrie-Anwendungen im BVLOS-Betrieb erfüllen. Für datenintensive Anwendungen wie Video-Streaming ist die Netzqualität im VLL größtenteils ausreichend, jedoch stark von lokalen Bedingungen abhängig. Die Kombination von LTE und 5G ermöglichte eine weitgehend unterbrechungsfreie Datenübertragung. Allerdings bleibt die Interferenz in größeren Höhen ein begrenzender Faktor, der sich insbesondere auf Videoanwendungen auswirkt. Insbesondere zur Absicherung kritischer Missionsphasen (Start, Landeanflug, Notfallmodus) wird daher empfohlen, auf redundante Kommunikationslösungen zu setzen: parallele Nutzung mehrerer Mobilfunkbetreiber (Multi-MNO), Integration privater 5G-SA-Campusnetzes, und optionale Satelliten-Backupverbindungen für Notfallkommunikation. Für Telemtrie-Verbindungen beispielsweise lässt sich die Dienstgüte durch die Nutzung von QoS-Class Identifiers (QCIs) gemäß 3GPP-Spezifikation [TS23.255] priorisieren, ohne dass die Netzlast signifikant erhöht wird.

Tab. 2: Anwendungsfälle und empfohlene Schwellenwerte für UAV-Konnektivität

Anwendungsfall	**Latenz**	**Datenrate**	**Zuverlässigkeit**	**Bewertung**
Command & Control (C2)	<100 ms	<0.5 Mbit/s	>99.999 %	**hoch**
Telemetrie	<500 ms	0.1–1 Mbit/s	>99.9 %	**mittel**
Video-Streaming	0.5–1 s	2–5 Mbit/s	>99.5 %	**variabel**

6 Zusammenfassung

Die im Rahmen des Projekts *InnoDCon – Innovative Drone Connectivity* durchgeführten Messungen der Netzparameter im VLL zeigen, dass öffentliche Mobilfunknetze grundsätzlich geeignet sind, eine zuverlässige Kommunikationsbasis für UAS im BVLOS-Betrieb zu bilden. Die Messungen bestätigen das für UAV-Kommunikation typische Verhalten: höhere RSRP-Werte aufgrund freier Sichtverbindungen, jedoch reduzierter SINR infolge verstärkter Interferenz durch Mehrfachsichtverbindungen. UAV-Anwendungen wie C2- und Telemetrie-Dienste können zuverlässig betrieben werden, während für Video-Streaming Datenraten von mehreren Mbit/s lediglich mit lokaler Variabilität erreichbar waren. Für kritische Anwendungen werden daher zusätzliche Redundanz- und Sicherheitsmechanismen empfohlen. Künftige Untersuchungen können dazu Multi-MNO-Szenarien, Campusnetze sowie hybride 5G-/SAT-Lösungen analysieren, um die Kommunikationssicherheit zu erhöhen. Die Messdaten können zur Validierung von U-Space-Architekturen gemäß EASA-Vorgaben und zur Verbesserung der Routenoptimierung beitragen, indem sie Vorhersagemodelle für die Verbindungsqualität im VLL verbessern und helfen, Abdeckungslücken automatisiert zu identifizieren.

Literaturverzeichnis

[Am17] Amorim, R. M. D.; Mogensen, P. E.; Sørensen, T. B.; Kovács, I.; Wigard, J.: Pathloss Measurements and Modeling for UAVs Connected to Cellular Networks. In: Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC-Spring). IEEE, S. 1–6, 2017, DOI: 10.1109/VTCSpring.2017.8108204.

[Ba24] Barrios-Muñoz, R.; Bernabé, M.; López-Pérez, D.; Gomez-Barquero, D.; Quintanilla-García, I.: Evaluating 5G Networks for U-Space Applications: Insights from Dense Urban Measurement Campaign. CoRR, Preprint, 2024, arXiv: 2411.09666 [cs.NI].

[Br24] Braunfelds, J.; Jakovels, G.; Murans, I.; Litvinenko, A.; Senkans, U.; Rumba, R.; Onzuls, A.; Valters, G.; Lidere, E.; Plone, E.: Experimental Study on LTE Mobile Network Performance Parameters for Controlled Drone Flights. Sensors 24 (20), Published 14 October 2024., S. 6615, 2024, DOI: 10.3390/s24206615, URL: https://doi.org/10.3390/s24206615.

[Ca19] Cai, X.; Rodríguez-Piñeiro, J.; Yin, X.; Wang, N.; Ai, B.; Pedersen, G. F.; Yuste, A. P.: An Empirical Air-to-Ground Channel Model Based on Passive Measurements in LTE. IEEE Transactions on Vehicular Technology 68 (2), S. 1140–1154, 2019, DOI: 10.1109/TVT.2018.2886961.

[Gh23] Gharib, M.; Hopkins, B.; Murrin, J.; Koka, A.; Afghah, F.: 5G Wings: Investigating 5G-Connected Drones Performance in Non-Urban Areas. In: 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Presented October 2023, S. 1–6, 2023, DOI: 10.1109/PIMRC56721.2023.10294063, URL: https://doi.org/10.1109/PIMRC56721.2023.10294063.

[GNA21] Gharib, M.; Nandadapu, S.; Afghah, F.: An Exhaustive Study of Using Commercial LTE Network for UAV Communication in Rural Areas. In: 2021 IEEE International Conference on Communications (ICC) Workshops. Presented at the ICC Workshops, May 31 – June 4, 2021, S. 1–6, 2021, DOI: 10.1109/ICCW50388.2021.9473547, URL: https://doi.org/10.1109/ICCW50388.2021.9473547.

[Ka23] Kasurinen, T.: LTE Mobile Network Technical Feasibility for Unmanned Aerial Vehicle BVLOS Operations in a Rural Test Area. In: Proceedings of the 13th SESAR Innovation Days. Paper no. 28, Seville, Spain, 2023, URL: https://www.sesarju.eu/sites/default/files/documents/sid/2023/Papers/SIDs%5C_2023%5C_paper%5C_28%20final.pdf.

[Li19] Lin, X.; Wiren, R.; Euler, S.; Sadam, A.; Määttänen, H.-L.; Muruganathan, S. D.; Gao, S.; Wang, Y.-P. E.; Kauppi, J.; Zou, Z.; Yajnanarayana, V.: Mobile-Network Connected Drones: Field Trials, Simulations, and Design Insights. IEEE Vehicular Technology Magazine 14 (3), S. 115–125, 2019, DOI: 10.1109/MVT.2019.2917363, URL: https://doi.org/10.1109/MVT.2019.2917363.

[Sä20] Säe, J.; Wirén, R.; Kauppi, J.; Torsner, J.; Andreev, S.; Valkama, M.: Reliability of UAV Connectivity in Dual-MNO Networks: A Performance Measurement Campaign. In: IEEE International Conference on Communications Workshops (ICC Workshops). S. 1–5, 2020, DOI: 10.1109/ICCWorkshops49005.2020.9145477.

[Si23] Singh, R.; Jepsen, J. H.; Ballal, K. D.; Nwabuona, S.; Berger, M.; Dittmann, L.: An Investigation of 5G, LTE, LTE-M and NB-IoT Coverage for Drone Communication Above 450 Feet. In: 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). S. 370–375, 2023, DOI: 10.1109/WoWMoM57956.2023.00066.

[TR22.829] 3GPP: TR 22.829: Enhancement for Unmanned Aerial System (UAS), Technical Report TR 22.829, Release 16, 3rd Generation Partnership Project (3GPP), 2018, URL: https://www.3gpp.org/dynareport/22829.htm.

[TR23.754] 3GPP: TR 23.754: Study on Support of Unmanned Aerial System (UAS), Technical Report TR 23.754, Release 16, 3rd Generation Partnership Project (3GPP), 2019, URL: https://www.3gpp.org/dynareport/23754.htm.

[TR23.755] 3GPP: TR 23.755: Study on Application Layer Support for Unmanned Aerial Systems (UAS), Technical Report TR 23.755, Release 17, 3rd Generation Partnership Project (3GPP), 2020, URL: https://www.3gpp.org/dynareport/23755.htm.

[TR36.777] 3GPP: TR 36.777: Enhanced LTE Support for Aerial Vehicles, Technical Report TR 36.777, Release 15, 3rd Generation Partnership Project (3GPP), 2017, URL: https://www.3gpp.org/dynareport/36777.htm.

[TS22.125] 3GPP: TS 22.125: Unmanned Aerial System (UAS) Support in 3GPP, Technical Specification TS 22.125, Release 16, 3rd Generation Partnership Project (3GPP), 2020, URL: https://www.etsi.org/deliver/etsi_ts/122100_122199/122125/16.04.00_60/ts_122125v160400p.pdf.

[TS23.255] 3GPP: TS 23.255: System Architecture for Unmanned Aerial System (UAS) Support, Technical Specification TS 23.255, Release 17, 3rd Generation Partnership Project (3GPP), 2021, URL: https://www.3gpp.org/dynareport/23255.htm.

[TS29.255] 3GPP: TS 29.255: UAS Network Function (UAS NF) Services, Technical Specification TS 29.255, Release 17, 3rd Generation Partnership Project (3GPP), 2021, URL: https://www.3gpp.org/dynareport/29255.htm.

[TS29.256] 3GPP: TS 29.256: UAS Service Enabler Function (SEF) Services, Technical Specification TS 29.256, Release 17, 3rd Generation Partnership Project (3GPP), 2021, URL: https://www.3gpp.org/dynareport/29256.htm.

[Wa24] Warczek, J.; Kozuba, J.; Marcisz, M.; Pamuła, W.; Dyl, K.: Research on Mobile Network Parameters Using Unmanned Aerial Vehicles. Sensors 24 (17), S. 5526, 2024, DOI: 10.3390/s24175526, URL: https://doi.org/10.3390/s24175526.

http://dx.doi.org/10.25673/121097

Aktuelle Messergebnisse der dynamischen Eigenschaften von Funkkanälen in der Industrie

Zusammenfassung der Zwischenergebnisse aus dem Forschungsprojekt DynChanEm

Olaf Albert, André Gnad [1]

Abstract: Dieser Beitrag beschreibt die neuesten Arbeiten des ifak auf dem Gebiet des Messens, der Analyse und der Synthese von industriellen Funkkanälen. Das Projekt Dynamic Channel Emulator (DynChanEm), in dessen Rahmen die vorgestellten Arbeiten durchgeführt wurden, verfolgt das Ziel, die dynamischen Veränderungen des Funkkanals zu messen und perspektivisch in Abhängigkeit vom momentanen Zustand der Funkübertragung zu reproduzieren. Es werden aktuelle Messungen in funktechnisch-anspruchsvoller Umgebung präsentiert und ein Demonstrator-Ansatz vorgestellt, der die dynamischen Eigenschaften des Funkkanals auch abhängig vom Zustand des Funkprotokolls und so eine detailliertere Analyse von Verbindungsabbrüchen ermöglicht

Keywords: Industrial Radio Channel, Funkkanalmessungen, Kanalemulation, 5G NR, WLAN

1 Einleitung

Durch die steigenden Anforderungen an Flexibilität, Mobilität und einen geringen Installationsaufwand erhöht sich die Anzahl der drahtlosen Verbindungen in der Industrie laufend. Weitere Faktoren dafür sind die gute Verfügbarkeit von preiswerten Funkkommunikationslösungen aus dem IT-Bereich, die zunehmende Vertrautheit der zuständigen Mitarbeiter mit Funkkommunikationslösungen und deren Vertrauen in die Reife der Funklösungen. So kommt es, dass nicht nur die Anzahl der eingesetzten Funkverbindungen steigt, sondern auch immer anspruchsvollere Kommunikationsverbindungen durch Funklösungen abgedeckt werden. Immer häufiger verwendete autonome Fahrzeuge (inklusive Teleoperation) führen zusätzlich zu einem erhöhten Datenaufkommen zwischen den vernetzten Geräten. Dies alles resultiert in höheren Anforderungen bei der Funkübertragung [Rauc20].

Der industrielle Funkkanal zeichnet sich durch einige Besonderheiten aus [Rapp89], [Düng18]: Die vielen vorhandenen Metallflächen verursachen zahlreiche dominante Reflexionen und durch die im Vergleich zur Büroumgebung großen Bewegungen existiert eine hohe Dynamik im industriellen Funkkanal. Das führt dazu, dass die Funksysteme ständig die Parameter ihrer Synchronisationsalgorithmen und des *Channel Equalizers* umfassend anpassen müssen.

[1] ifak, Abt. Information und Kommunikation, Werner-Heisenberg-Str. 1, 39106 Magdeburg, [olaf.albert, andre.gnad]@ifak.eu, https://orcid.org/0000-0000-0000-0000

Hinzu kommt, dass die Anforderungen an die Zuverlässigkeit und Echtzeitfähigkeit der Kommunikation in der Industrie strikten Anforderungen unterworfen sind. Um zum Beispiel einen reibungslosen eng getakteten Ablauf in der Fertigung zu gewährleisten, müssen die Daten vollständig, fehlerfrei und rechtzeitig bei den entsprechenden Empfängern vorliegen, um schnelle Regelungen mit Ist-Größen zu versorgen, Teile am Fließband bereitzustellen und tonnenschwere Lasten ferngesteuert präzise und sicher durch Anlagen zu bewegen.

Moderne Funksysteme sind durch die implementierten Korrekturmechanismen bereits heute sehr robust. Sowohl Pegeleinbrüche im Funkkanal als auch – in begrenzterem Maße – Störsignale können die Übertragung aus Sicht der Applikation bis zu einer bestimmten Grenze kaum beeinflussen.

Um dies zu erreichen, werden essentielle Parameter der Funkkommunikation angepasst. Als Beispiel können hier genannt werden

- Die Kodierung, die für die Verteilung der Nutzdaten über der Übertragungsbandbreite verantwortlich ist und weitere Informationen einfügt, die die Reparatur fehlerhafter Daten ermöglicht,
- Die Modulationsart, die ein Optimum zwischen ausreichender Robustheit der Übertragung an das Signal-Rausch-Verhältnis und der Übertragungszeit (Datenrate) einstellt,
- Die Verwendung verschiedener Signalpfade bis einschließlich zusätzlicher Antennen, um entweder die am besten geeigneten auszuwählen (Antenna Diversity) oder durch Algorithmen verschiedene Übertragungspfade zu separieren und so eine teilweise parallelisierte Datenübertragung mit entweder höherer Datenrate oder größerer Zuverlässigkeit der Datenübertragung durch mehr Redundanz (MIMO) zu ermöglichen.

Zum Einstellen dieser Parameter durch die Funkkomponenten ist die Kenntnis des Funkkanals grundlegend. Verschiedene Funksysteme arbeiten hier mit unterschiedlichen Methoden. Es werden bekannte Signale gesendet, die auch gestört leicht erkennbar sind (Trainingssymbole, Pilotträger) und über Zeit und Frequenz verteilt in den Datenstrom eingefügt werden. Eine andere Methode ist das explizite Anfordern von Kanalmessungen und das Mitteilen des Ergebnisses an den Kommunikationspartner, dass bei Veränderungen im Funkkanal stattfindet.

Das Projekt DynChanEm verfolgt das Ziel, diese dynamischen Veränderungen des Funkkanals zu messen und perspektivisch in Abhängigkeit vom Zustand des Funkprotokolls diese dynamischen Funkkanäle zu reproduzieren.

Es werden zugrunde liegende aktuelle Messungen in funktechnisch-anspruchsvoller Umgebung präsentiert und ein Demonstrator-Ansatz vorgestellt, der die dynamischen

Eigenschaften des Funkkanals auch abhängig vom Zustand des Funkprotokolls verändern kann und so eine detailliertere Analyse von Verbindungsabbrüchen ermöglicht.

2 Stand der Technik

Kanalmessungen werden durchgeführt, seit es Funkkommunikation gibt. Ist dieser nicht unerhebliche Aufwand immer noch gerechtfertigt?

Es gibt mehrere Gründe, die diesen Aufwand immer noch als sinnvoll erscheinen lassen:

- Es werden Kommunikationsabbrüche in dynamischen Funkkanälen beobachtet, die noch nicht einwandfrei erklärt werden können.
- Die Funksysteme entwickeln sich ständig weiter. Um die neuen Eigenschaften simulieren zu können werden Funkkanalmodelle benötigt, die auf die neuen Systeme z. B. in Bandbreite und zeitlichen Verhalten zugeschnitten sind. Alte Modelle sind hier oft zu begrenzt oder zu ungenau. Zum Beispiel verwenden heutige Funksysteme oft Modulationsarten, die sehr viele Subträger mit einer Kombination aus Amplituden- und Phasenmodulation benutzen (genannt OFDM). Diese Systeme verhalten sich bei Pegeleinbrüchen im Funkkanal (*Fading*) robuster als frühere Funksysteme, die zur Datenübertragung lediglich einen einzelnen Träger mit Amplituden- oder Frequenzmodulation benutzt haben.
- Die Hersteller der integrierten Komponenten geben nur selten und ungern ihr *Core Knowhow* zu den verwendeten Algorithmen preis. Wenn man Funksysteme also deswegen als Black-Box testen möchte, sind Geräte erforderlich, die vordefinierte und reproduzierbar Funkkanäle erzeugen können. Diese Geräte werden Emulatoren genannt. Und für diese Emulatoren benötigt man Trainingsdaten, die für Funksysteme eine Herausforderung - also grenzwertige - Funkkanäle darstellen. Ein weiterer Vorteil ist, dass die Emulatoren mit überschaubarem Aufwand auf neue Funksysteme erweitert werden können.

2.1 Funkkanalmessungen

Forschungen in diesem Gebiet befassen sich mit unterschiedlichen Aspekten des industriellen Kanals. Ein Schwerpunkt liegt darauf, Modelle für industrielle Kanäle zu erstellen. Daher werden Kanalstatistiken entweder allgemein oder für eine Vielzahl definierter Kommunikationsstandards untersucht [Croo17, Rapp89, Zhan18].

Es werden verschiedene Messmethoden eingesetzt, etwa Vektor-Signal-Generatoren als Sender und Echtzeit-Spektrumanalysatoren als Empfänger [Bloc15], Software definierte Radios [Burm21], Vektor-Netzwerk-Analysatoren [Zhan18] oder Puls-Modulationen auf Frequenzträgern [Rapp89]. Der Kompromiss zwischen diesen Methoden besteht meist in der Messgeschwindigkeit gegenüber dem Dynamikbereich der Messergebnisse. Andere

Forschungsgruppen verwenden dedizierte drahtlose Kommunikationssysteme [Halp10] und können so das Verhalten eines bestimmten Kommunikationsstandards detaillierter untersuchen.

Die Forschung erstreckt sich zudem über verschiedene Frequenzbänder: von 1,3 GHz in [Rapp89] über alle Kommunikationsfrequenzbänder bis zu 6 GHz [Burm21, Holf16], Millimeter-Wellen-Bänder zwischen 20 GHz und 30 GHz bis hin zu Infrarotlicht [Düng18, Koym15]. Der Grund dafür ist, dass zum Zeitpunkt der Veröffentlichung häufig neue Kommunikationsstandards im Fokus stehen, die aufgrund der höheren verfügbaren Bandbreite und damit der höheren möglichen Datenrate an Bedeutung gewinnen.

Übersichten, die verschiedene Messkampagnen zusammenfassen, finden sich beispielsweise in [Düng18]. Diese Arbeiten bieten einen guten Überblick über die durchgeführten Forschungen.

In vorhergehenden Projekten wurde erkannt, dass statische Kanäle für die gegenwärtigen Funksysteme keine besondere Herausforderung darstellen. Bis zu einem minimal empfangbaren mittleren Empfangspegel regelt das Funksystem die Parameter nach und etwas darunter bricht die Funkverbindung komplett ab.

Daher wurde der neue Messaufbau so gewählt, dass er dynamische reproduzierbare Messungen erlaubt und so auch den Test der Regelmechanismen der Funksysteme ermöglicht.

2.2 Funkkanalemulation

Heute verfügbare digitale Kanalemulatoren basieren auf der statistischen Anwendung von Funkkanälen. Dies bewirkt neben einer großen absoluten Testabdeckung jedoch eine relativ geringe Wahrscheinlichkeit für das Auftreten von Fehlern. Weiterhin sind diese digitalen Emulatoren sehr komplex, also kostenintensiv in Anschaffung und Nutzung.

Der hier verfolgte Ansatz unterscheidet sich dahin gehend, dass die Realisierung des Hochfrequenz-relevanten Teils als gesteuerte parallele analoge Pfade erfolgt. Diese Pfade können in Übertragungsdämpfung und -phase gesteuert werden. Ziel ist nicht die für die Entwicklung notwendige große absolute Testabdeckung, sondern der systematische Test von in der Industrie auftretenden kritischen Kanälen.

Weiter enthält der Aufbau Pegeldetektoren, mit denen das Senden der angeschlossenen Geräte detektiert werden kann, um so in vom Funkprotokoll abhängigen, definierten Momenten den emulierten Funkkanal zu verändern und so das Testobjekt zum Nachregeln anzuregen.

3 Beschreibung der Messmethode und des Messaufbaus

Die verwendete Methode besteht aus zwei Teilen: erstens einem kommerziellen WLAN-Funksystem (IEEE 802.11n, Bandbreite = 40 MHz, Kanal = 159 [5,8 GHz]), das durch weitere Messtechnik beobachtet wird und zweitens dem Funkkanal-Messsystem, das die Dämpfung und Phase des Funkkanals ermittelt. Da die Veränderungen im Funkkanal bei der Messung reproduzierbar auftreten, lassen sich beide Teile des Tests zeitlich trennen.

Dieser Aufbau wurde von uns Channel Co-Measurement genannt, um auf die mehrfache Messung desselben Funkkanals mit unterschiedlichen Methoden hinzuweisen.

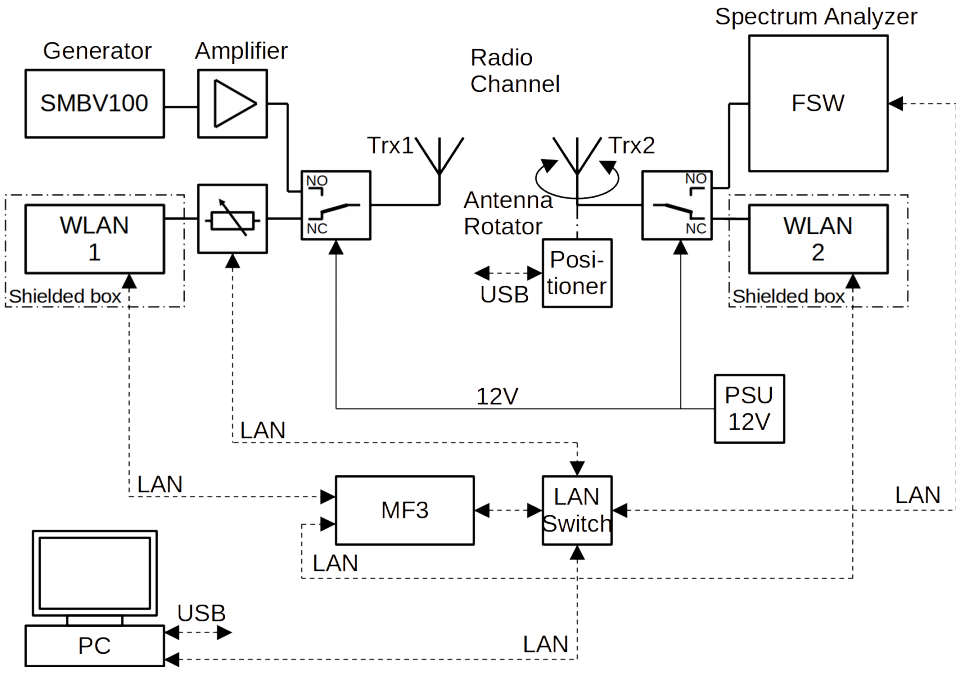


Abb. 1: Blockschaltbild des Aufbaus für die Channel-Co-Measurements

Abbildung 1 zeigt den prinzipiellen Aufbau dafür. In der Mitte ist der Funkkanal dargestellt mit den Antennen für beide Seiten. Auf der Seite des Transceivers 2 (Trx2) befindet sich der Aufbau zum Rotieren der Antennenposition auf einem 0,5 m-langen Arm (siehe auch Abbildung 2 unten). Der kabelgebundene Teil des HF-Signalpfades wird zwischen Funkkanal-Messsystem (oben) und WLAN-System (unten) mittels HF-Relais und 12 V-Spannung umgeschaltet.

Das Messsystem besteht aus einem Vektorsignalgenerator, der eine Zadoff-Chu-Folge mit einer Bandbreite angepasst auf das WLAN-Signal erzeugt. Um den Dynamikbereich zu erweitern, ist dem Generator ein Verstärker nachgeschaltet, der die Sendeamplitude

anhebt. Auf der Empfangsseite rechts befindet sich ein Echtzeit-Spektrumanalysator, der das empfangene Signal entsprechend schnell abtasten und speichern kann. Je Messung entstehen 220 Mio. komplexe Abtastwerte. Nicht dargestellt sind die Leitungen, die das Triggersignal verteilen sowie die 10 MHz-Referenz-Verbindung zwischen Generator und Spektrum Analyzer.

Der untere Teil enthält auf beiden Seiten das WLAN-Kommunikationssystem. Auf der Seite des Transceivers Trx1 ist zusätzlich ein variabler HF-Abschwächer im Signalpfad, um die Signalstärke in den kritischen Bereich verschieben zu können. Im kritischen Bereich entsteht bereits eine nennenswerte Anzahl von Fehlern, ohne dass die Verbindung komplett abreißt. Datenquelle und -senke werden vom ifak-eigenen Messsystem Funk Transfer Tester mit der Hardware-Komponente Multiface (MF3) gebildet. Dieses Messsystem ermittelt die Anzahl der übertragenen, verlorenen oder sonst beschädigt empfangenen UDP-Pakete.

Die Steuerung übernimmt ein PC, der sich mit dem Operator abgesetzt in einem anderen Raum befinden kann, um Einflüsse auf die Messung durch die Bewegungen des Operators zu vermeiden.

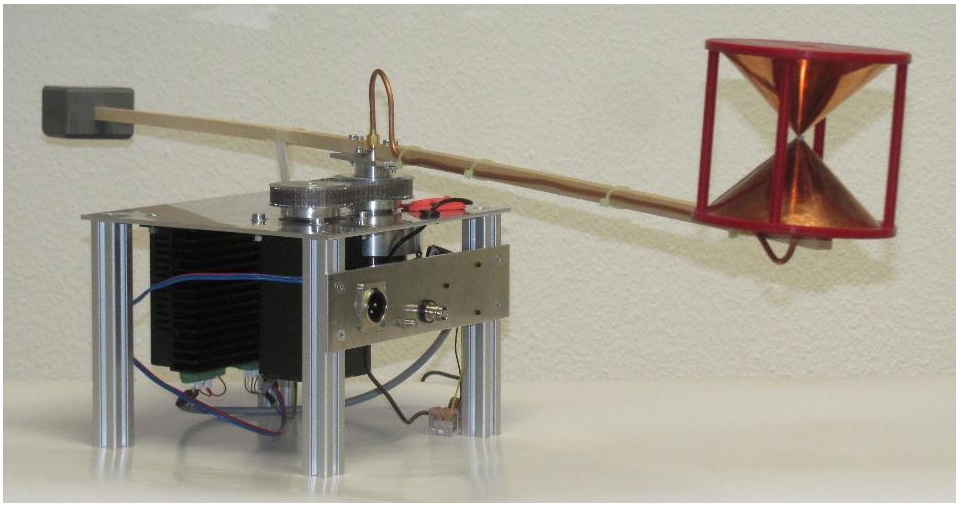


Abb. 2: Foto des Antennenrotators mit Bicone-Breitbandantenne und HF-Drehkupplung in der Achse

4 Beschreibung des Emulator-Prototypen

Kernstück des Prototyps sind mehrere Hochfrequenzpfade, die sich in Dämpfungs-amplitude und -phase einstellen lassen, um die Fadingeinbrüche an den gewünschten Stellen im Frequenzbereich mit der geplanten Tiefe platzieren zu können. Diese

Einstellungen müssen in wenigen Mikrosekunden anpassbar sein, um die für die Untersuchung notwendige Dynamik zu erreichen.

Um das Prinzip testen zu können, wurde die Anzahl der Pfade auf drei reduziert. Für komplexere Kanaleinstellungen mit vielen Fadingeinbrüchen werden hier mehr Pfade in einem Gerät verwendet werden müssen.

Um die Pfade in Betrag und Phase einstellen zu können, werden kürzlich auf dem Markt erschienene Hochfrequenz-IC verwendet, die für das *Antenna Beam Steering* (Schwenken des Antennenrichtdiagramms) entwickelt wurden. Mit Hilfe des *Beam Steering* werden schnell elektronisch schwenkbare Antennenkeulen realisiert, um zum Beispiel im Mobilfunk die räumliche Kanalkapazität des Netzes zu erhöhen, in dem mit jedem Nutzer über einen separaten Beam kommuniziert wird.

Dieser IC vom Typ ADAR4002 (Analog Devices Inc.) wird über einen seriellen Bus angesteuert, der mittels einer Taktrate von bis zu 100 MHz betrieben werden kann. So können die Eigenschaften der Pfade in sehr geringer Zeit umgestellt werden.

Die folgende Abbildung 3 zeigt ein Foto der Hauptbaugruppe des Prototyps: oben der 8-Bit-µController (ATMega) darunter die drei HF-Pfade mit je einem ADAR4002 in der jeweiligen Mitte des Pfades (kleine schwarze Rechtecke). Oben links befindet sich der USB-Anschluss zur Steuerung.

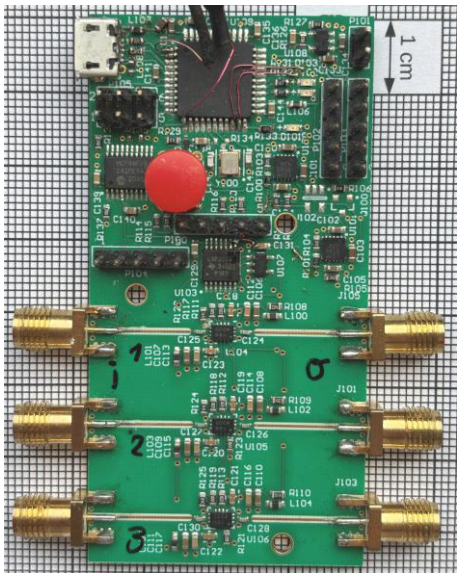


Abb. 3: Foto der Hauptbaugruppe des Prototyps mit µController und drei ADAR4002

Diese Hauptbaugruppe ist Teil des Prototypaufbaus mit den drei genannten HF-Pfaden. Die Abbildung 4 zeigt das vereinfachte Gesamt-Blockschaltbild des Prototyps. Dort sind die Funktionsblöcke der Hauptbaugruppe in hellgrün markiert.

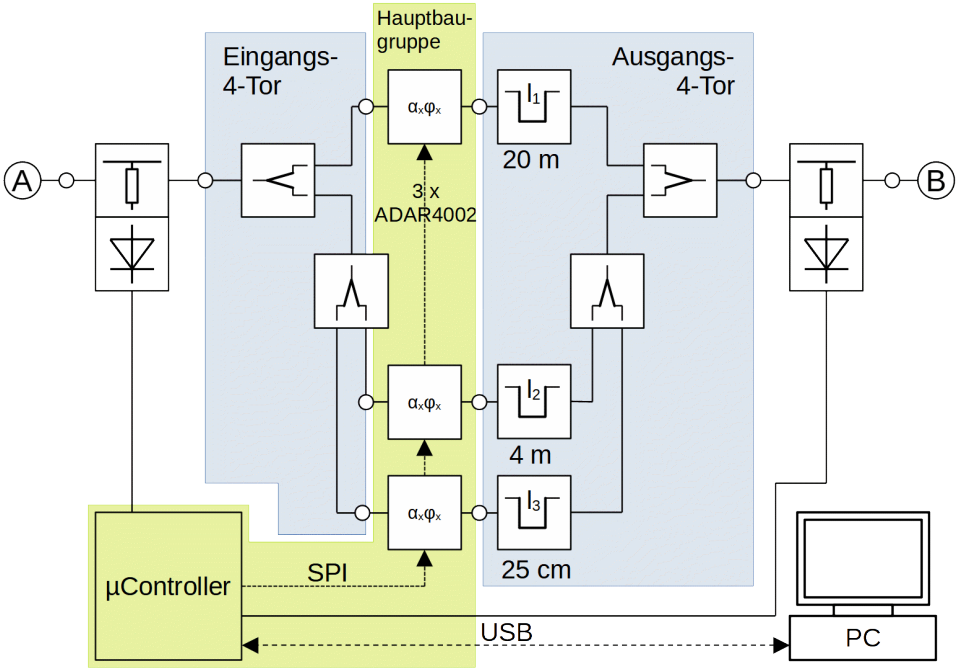


Abb. 4: Vereinfachtes Gesamt-Blockschaltbild des Prototyps

Zusätzlich zu den Komponenten der Hauptbaugruppe besteht der Prototyp aus:

- Zwei HF-Detektor-Baugruppen, die an den jeweiligen Außenanschlüssen (A und B) des Prototyps den HF-Pegel messen und somit das Senden der jeweiligen Seite detektieren,
- Insgesamt vier HF-Leistungsteilern, die die HF-Signale in drei HF-Pfade aufteilen und wieder zusammenführen,
- drei unterschiedlich langen Kabeln (l_1, …, l_3), die gemeinsam mit den ADAR4002-IC die HF-Pfade nachbilden und
- dem Steuer-PC, der den µController konfiguriert.

5 Vorläufige Ergebnisse

Die Arbeiten am Projekt laufen derzeit noch und die Auswertung der Daten ist deshalb noch nicht abgeschlossen. Dieses Kapitel zeigt deshalb den aktuellen Stand der Dinge wie er zum gegenwärtigen Zeitpunkt verfügbar ist.

Bei den Messungen wurde darauf Wert gelegt, dass die Auswertung an die Möglichkeiten und Verfahren realer Kommunikationssysteme angelehnt ist und so möglichst praktisch relevante Resultate erzielt.

5.1 Dynamische Messungen in kontrollierter Umgebung

Die folgende Abbildung 5 zeigt das Blockschaltbild der dynamischen Messung in kontrollierter Umgebung. In dieser gab es nur die Bewegungen des Antennenrotators.

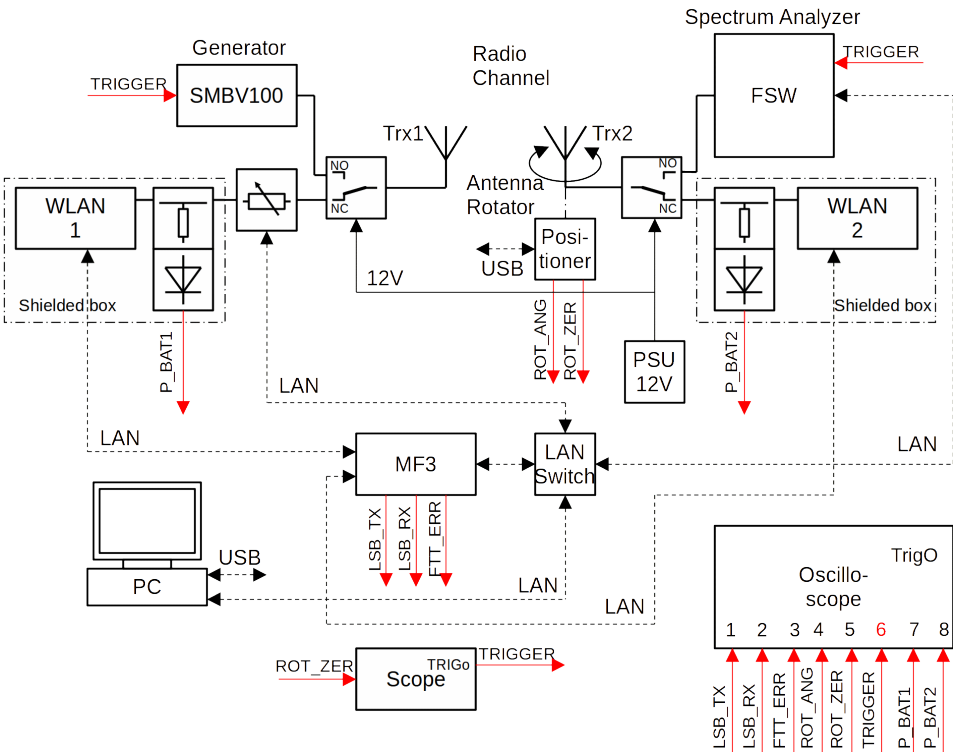


Abb. 5: Blockschaltbild des Messaufbaus, hervorgehoben in rot die analogen Messkanäle des 8-Kanal-Oszilloskops

Gegenüber Abbildung 1 ist das Blockschaltbild um ein 8-kanaliges Oszilloskop (unten rechts) sowie um Detektoren an den Antennenanschlüssen der WLAN-Router ergänzt. Das Oszilloskop fasst als analoge Messung sowohl die digital-analog-gewandelten Ausgaben des Multifaces wie auch alle anderen rot gekennzeichneten Analogsignale in einer Messung mit hoher zeitlicher Auflösung (10 MSamples/s) und großer Speichertiefe (200 MSamples) zusammen. Die Detektoren zeigen das Senden des jeweiligen WLAN-Routers (zu sehen in Abbildung 5 als Kurven *p_bat1* und *pbat2*) an.

Durch die Ausgabe der 4 LSBs der Sequenznummer auf Sende- sowie auf Empfangsseite und die Ausgabe des FTT-Fehlercodes (Abbildung 6: Kurven *ftt_lsb_tx*, *ftt_lsb_rx* und *ftt_err*) lassen sich leicht Übertragungsprobleme erkennen und ein Bezug zur parallel abgespeicherten Log-Datei des Funk-Transfer-Testers - mit deren niedrigerer zeitlichen Auflösung - herstellen. Zusätzlich werden die aktuelle Position des Antennenrotators als winkelproportionale Spannung und der Durchgang durch die Null-Position (Abbildung 6: Kurven *rot_angle* und *trigger*) durch die Oszilloskop-Messung erfasst.

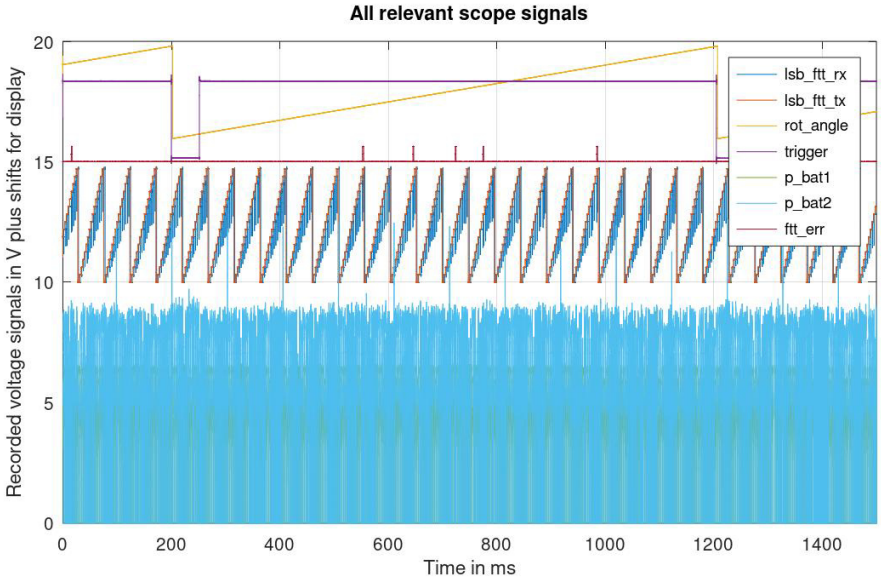


Abb. 6: Oszilloskop-Messung mit allen erfassten Signalen

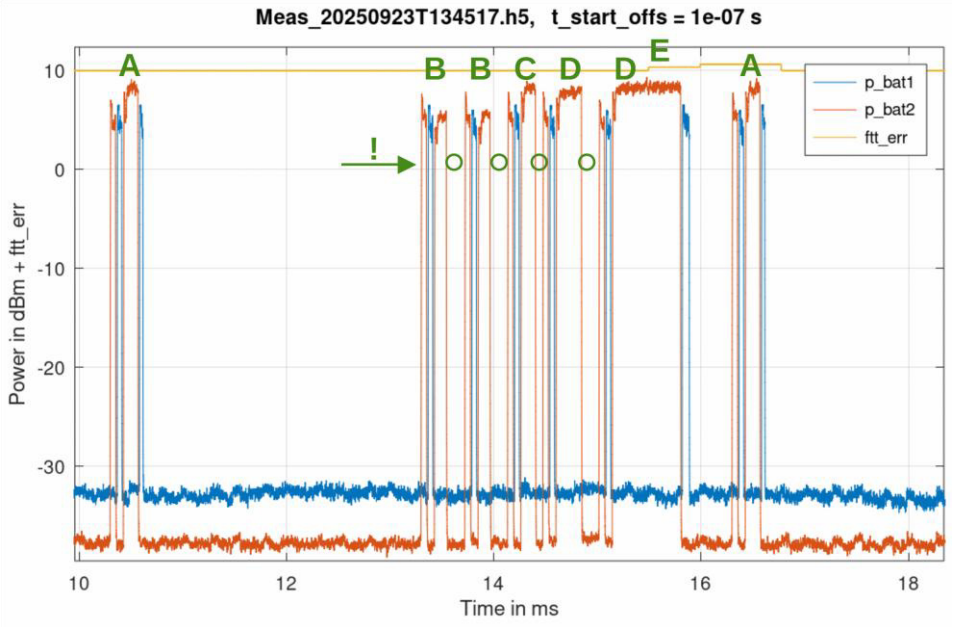


Abb. 7: Oszilloskop-Messung mit Sendepulsabfolge mit Kommunikationsproblemen

In Abbildung 7 sind die Detektor-Signale beider WLAN-Router dargestellt und das FTT-Fehler-Signal.

Die Messung zeigt folgende Kommunikationsabschnitte: Bei Position A erfolgt die Datenübertragung (von Bat2 zu Bat1) fehlerfrei, während bei Position B die Bestätigung von Bat1 fehlt (Kreise bei „→ !“), bei C die Sendeleistung von Bat2 erhöht wird und sich bei D die Burstlänge - vermutlich durch eine Modulation mit weniger Bit je Symbol - verlängert. An Position E ist zu sehen, dass ein FTT-Fehler erkannt wurde (*ftt_err* > 0), die Nachricht wurde nicht innerhalb der vorgegebenen Zykluszeit von t_{cycle} = 3 ms übertragen.

Abbildung 8 zeigt die Auswertung der FTT-Logfiles verschiedener Messungen mit Bezug auf den momentanen Drehwinkel des Antennenrotators. Die einzelnen Kurven unterscheiden sich beim Parameter Drehgeschwindigkeit des Antennenrotators v in U/min. Der Kanal blieb bei allen Messungen sonst konstant (ca. 4 m, NLOS) und war durch den variablen Abschwächer (siehe Abbildungen 1 und 4) so in der Dämpfung eingestellt, dass die Fehlerrate sichtbar anstieg.

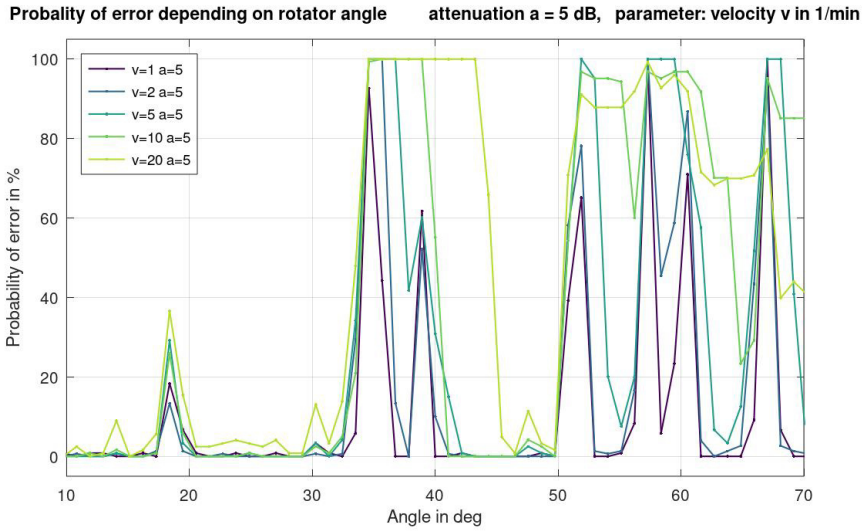


Abb. 8: Auswertung der Fehlerrate bestimmt aus dem FTT-Logfile mit Bezug auf den aktuellen Drehwinkel des Antennen-Rotators (Ausschnitt 0º … 90º zur besseren Sichtbarkeit)

Zu sehen ist die Häufung der Fehler bei bestimmten Drehwinkeln. Die Fehlerrate nimmt mit höherer Drehgeschwindigkeit, also dynamischerer Änderung des Funkkanals, zu. Dies zeigt, dass die Änderungsgeschwindigkeit zumindest im Bereich hoher bis kritischer Kanaldämpfung einen wesentlichen Einfluss auf die Fehlerrate hat.

Ist die Kanaldämpfung deutlich kleiner, ist die Auswirkung nicht so stark und bei höhere Kanaldämpfung kommt keine Nachrichtenübertragung mehr zustande.

5.2 Messungen im Hochregallager

Hochregallager sind extreme Beispiele für industrielle Funkkanäle. Sie sind räumlich groß, bestehen aus überwiegend parallelen Metallstrukturen, die Reflektoren für das HF-Signal bilden und enthalten sich schnell bewegende Bediengeräte ($v_{bg} \leq 10$ km/h) in schmalen Gängen, die für den Transport der Paletten zu den einzelnen Lagerplätzen zuständig sind.

Die Messung erfolgte in einem Hochregallager, das 41 m hoch ist, 120 m lang und über 9 Gänge verfügt. Abbildung 9 zeigt die Situation vor Ort in Fotos.

Abb. 9: Fotos des Hochregallagers (oben: Außenansicht, unten, v. l.: Gang, Senderaufbau, Empfängeraufbau)

Nicht in Abbildung 9 zu sehen ist eine Lichtschranke, die in dem Moment einen Triggerimpuls auslöst, in dem das Regalbediengerät eine bestimmte Position zwischen Sende- und Empfangsantenne passiert. Die Position der Lichtschranke wurde gewählt in der Annahme, dass sich hier der Funkkanal deutlich ändert (s_{Trig} = 18 m). Sender- und Empfängerantenne waren ca. 30 m voneinander entfernt. Der Empfänger Rx befand sich am Anfang des Ganges (s_{Rx} = 0 m, s_{Tx} = 30 m).

Im Hochregallager waren aufgrund der Bewegungen durch den Betrieb in den anderen Gängen keine Channel-Co-Measurements realisierbar, da sich zum einen der Funkkanal

fortwährend änderte und zum anderen auch die Bewegung des Regalbediengerätes nicht hochreproduzierbar war, obwohl immer dieselben Lagerorte angefahren wurden.

Praktische Effekte wie Störungen des Triggerimpulses durch den über Schienen zugeführten Antriebsstrom des Bediengerätes bildeten zusätzliche Herausforderungen während der Messung.

Abbildung 10 zeigt als Beispiel für ein Messergebnis den logarithmierten Betrag der Kanalantwort über der Frequenz während der Bewegung. Besonders bemerkenswert sind die zeitlich periodischen Änderungen, die durch die Bewegung im stark echobehafteten Raum über einen längeren Zeitraum entstehen. Ein Effekt, der bei diesen Messungen erstmalig zu sehen war.

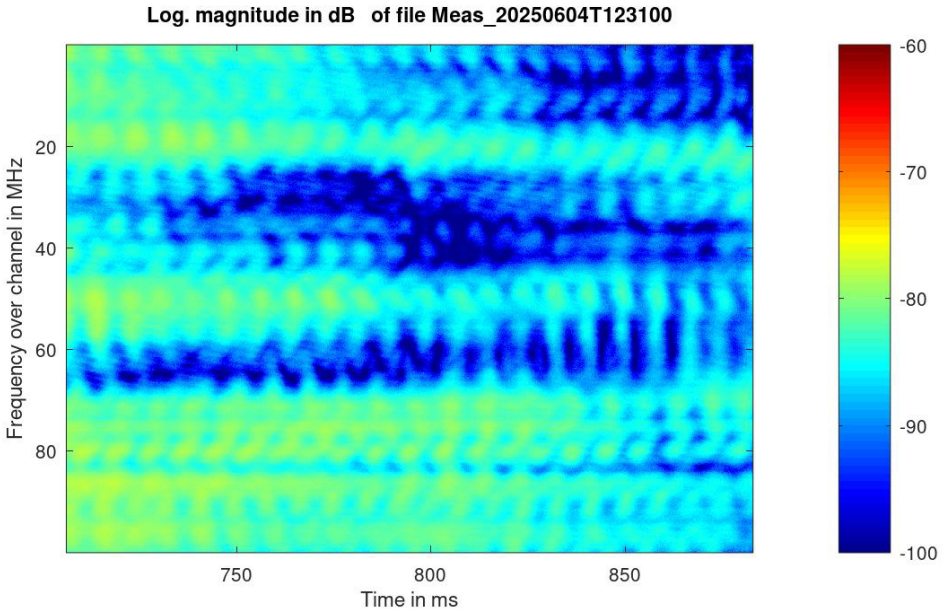


Abb. 10: Log. Betrag der Kanalantwort

Neu bei diesen Messungen ist, dass die Auswertung der Kanalantwort nicht nur für den Betrag erfolgte, sondern auch für die Phase (siehe Abbildung 11). Wie zu erkennen ist, ändert sich in bestimmten Frequenzbereichen die Phase innerhalb von 1 ms um bis zu ±45° (rote und blaue Bereiche in Abbildung 12).

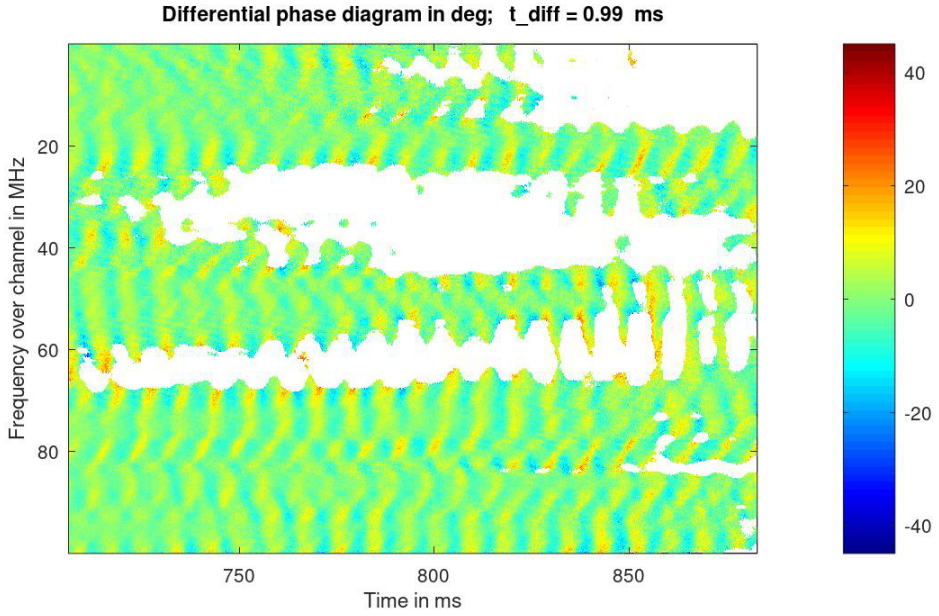


Abb. 11: Phasenänderung der Kanalantwort zwischen zwei Messungen im Abstand von 1 ms

Um auswertbare Daten zu erhalten, ist es notwendig den *Carrier Frequency Offset* und den *Sampling Time Offset* zu korrigieren. Beide würden beim verwendeten Messverfahren sonst dafür sorgen, dass aufeinanderfolgende Messungen auch bei sonst konstantem Kanal Sprünge und Drehungen der Phasenverläufe zeigen. Zusätzlich muss die Phasendarstellung an Stellen ausgeblendet werden, wo der Signalpegel zu gering wird (hier threshold = -90 dB). Das sonst dargestellte Rauschen der Phase von ($-\pi \ldots +\pi$) würde eine optische Beurteilung oder weitere Verarbeitung behindern. Dies geschieht in Abbildung 11 durch ersetzen der Messwerte mit weißer Fläche.

Es wird davon ausgegangen, dass das Funksystem eine erste Demodulation durchführen konnte, also auch eine Korrektur der Einflüsse des Funkkanals initial durchführt. Es interessiert dann die Änderung des Funkkanals nach einer festgelegten Zeit, die zum Verlust der Synchronisation oder Demodulierbarkeit führte. Die Messungen werden deshalb als Differenzen zu vorhergehenden Messungen dargestellt. So ist besonders leicht ersichtlich, bei welcher Frequenz und welcher Zeit sich der Kanal verändert hat.

In Abbildung 11 fallen wie beim Betrag die zeitlich periodischen starken Phasenänderungen auf. Diese Strukturen bilden wohl die Ursache für die störungsanfällige Kommunikation in Hochregallagern.

5.3 Messungen des Emulator-Prototypen

Zur Inbetriebnahme des Prototyps wurden alle Komponenten mit Hilfe von Netzwerkanalysator-Messungen charakterisiert. Weitere NWA-Messungen dienen als Berechnungsgrundlage für die Fitting-Algorithmen, die die gemessenen Funkkanäle durch Prototyp-Einstellungen nachbilden werden. Diese Arbeiten sind für die kommenden Monate geplant.

Erste Ergebnisse zeigen, dass die Anpassung der SMA-Anschlüsse aufgrund der 4-lagigen Leiterplatte (ein Core aus FR-4 mit beidseitiger Auflage aus HF-Substrat Ro4350B) einer Verbesserung bedürfen, da hier frequenzabhängig ein Reflexionsfaktor von S11max = >-8 dB zu messen ist. Die Isolation zwischen den Pfaden hingegen ist ausreichend hoch, um die Verkopplung der Pfade nicht in den Fitting-Algorithmen berücksichtigen zu müssen. Die Pfade 1, 2 und 3 können also separat gemessen und als Tabellen gespeichert werden (3×2^{14} S2P-Daten).

Die Einstellgeschwindigkeit der HF-Pfade der ADAR4002 entspricht den Datenblattangaben und bietet somit keine erwähnenswerten Neuigkeiten. Die drei IC können an einem SPI-Bus innerhalb von 1 µs umkonfiguriert werden.

Wie durchgeführte Simulationen zeigen, ist das geplante Vorgehen mit dem separaten Messen des Eingangs-Viertors (siehe Abbildung 4), dem Ausgangs-Viertor, dass auch die Kabel enthält, und dem Einfügen der Zweitor-Parameter (S2P) der drei Pfade der ADAR-4002-Baugruppe realisierbar.

Gegenwärtig wird auch an den Algorithmen gearbeitet, die die Parametersätze der ADAR4002-Baugruppe so bestimmen, dass der Prototyp die gleiche Kanalantwort zeigt, wie die Messung. Dieses Fitting zwischen Prototyp-HF-Parameter und gemessenen Daten ist anspruchsvoll – insbesondere bei begrenzter Rechenleistung und verfügbarer Zeit für die Berechnung der Einstellungen.

6 Zusammenfassung und Ausblick

Die durchgeführten Messkampagnen zeigen deutlich den Einfluss der dynamischen Kanaländerungen auf Funksysteme wie WLAN (Absatz 5.1) und das Beispiel der Messungen im Hochregallager (Absatz 5.2) verdeutlicht die Relevanz der Untersuchungen insbesondere für schwierige industrielle Kanäle.

Geplant ist in den kommenden Wochen die Implementierung des Emulator-Kanal-Fittings passend zu den Messungen und die Tests auf die Wirkung der WLAN-Kommunikation.

Folgeprojekte mit Partnern zur Entwicklung des Emulators als marktgängiges Gerät sind angestrebt und Kontakte bestehen bereits aus vorhergehenden Projekten.

Der Fokus liegt hierbei auf der Gruppe von Anwendern, die Funksystemkomponenten mit überschaubarem Aufwand auf Eignung bezüglich kritischer HF-Pfade testen möchten.

Danksagungen

Die Forschungsarbeit, die diesem Beitrag zugrunde liegt, wird vom Bundesministerium für Wirtschaft und Energie der Bundesrepublik Deutschland (BMWE; FKZ: 49VF240002) gefördert.

Für die Messungen im Hochregallager wurde uns von der OvGU Magdeburg, Prof. Lurz, ein Leistungsverstärker zur Verfügung gestellt.

Die Hochfrequenz-IC wurden uns von der Analog Devices Inc., Herrn Horvath, zur Verfügung gestellt.

Die Messungen im Hochregallager wurden durch die EGLO Logistikzentrum Deutschland GmbH, Magdeburg, Herrn Matthias Schröder, Ronny Roitzsch und andere ermöglicht und tatkräftig unterstützt.

Der deutsche Messgerätehersteller Rohde & Schwarz, in Person Herr Falko Jirka, unterstützte die Arbeiten durch zeitnahe und unentgeltliche Bereitstellung eines leistungsfähigen 8-kanaligen Oszilloskops MXO58.

All den genannten möchte ich für die bereitgestellten Mittel, Möglichkeiten, Zeit und Mühen danken! Auch den beteiligten und unterstützenden Kolleginnen und Kollegen gilt mein Dank für Ihre Hilfe und ihren Support!

Literaturverzeichnis

[Bloc15] Dimitri Block u. a., Wireless Channel Measurement Data Sets for Reproducible Performance Evaluation in Industrial Environments, IEEE, 2015

[Burm21] Friedrich Burmeister u. a., Measuring Time-Varying Industrial Radio Channels for D2D Communications on AGVs, IEEE Wireless Communications and Networking Conference, 2021.

[Croo17] Ramona Croonenbroeck u. a., Measurements for the Development of an Enhanced Model for Wireless Channels in Industrial Environments, IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2017

[Düng18] Düngen, M., Hansen, T., Croonenbroeck, R., a. o. Channel measurement campaigns for wireless industrial automation, at - Automatisierungstechnik, Germany, https://doi.org/10.1515/auto-2018-0052, 2018.

[Halp10] Daniel Halperin a. o., Predictable 802.11 Packet Delivery from Wireless Channel Measurements, SIGCOMM'10, August 30–September 3, New Delhi, India, 2010

[Holf16] Bernd Holfeld, Radio Channel Characterization at 5.85 GHz for Wireless M2M Communication of Industrial Robots IEEE Wireless Conference and Networking Conference (WCNC 2016) Track 1: PHY and Fundamentals, 2016

[Koym15] Ozge Koymen, a. o., Indoor mm-Wave Channel Measurements: Comparative Study of 2.9 GHz and 29 GHz, 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 2015

[Rauc20] Editor Lutz Rauchhaupt Anforderungsprofile für eine zuverlässige drahtlose Kommunikation in der Industrie, BMBF Förderprogramm IKT 2020, https://industrial-radio-lab.eu/download/805/, 2020.

[Rapp89] Rappaport, T. S. Characterization of UHF Multipath Radio Channels in Factory Buildings. Transactions on Antennas and Propagation, Vol. 37, No 8. 1989.

[Zhan18] Quilong Zhan. Understanding the Temporal Fading in Wireless Industrial Networks: Measurements and Analyses, IEEE, 10th International Conference on Wireless Communications and Signal Processing (WCSP), Hangzhou. China, doi: 10.1109/WCSP.2018.8555681, 2018

http://dx.doi.org/10.25673/121097

Einsatzbeispiele für ein Werkzeug zur Untersuchung der Zuverlässigkeit industrieller Kommunikationssysteme

Sarah Willmann[1], Marco Meier[2] und André Gnad[3]

Abstract: In diesem Beitrag werden Messergebnisse vorgestellt, die aufzeigen, dass ein Test zur Bewertung der Zuverlässigkeit industrieller Kommunikationssysteme zu umfangreichen Erkenntnissen über mögliche funktionale Einschränkungen der jeweiligen Implementierung führt. Kleine Stichproben können schon aussagekräftige Ergebnisse über das grundsätzliche Verhalten einer Kommunikationslösung liefern. Für eine Bewertung der Zuverlässigkeit sind jedoch immer Langzeitmessungen erforderlich, um eine inhaltsreiche Datengrundlage für die statistische Auswertung zu haben sowie die Ursachen für die unzureichende Funktionsfähigkeit ermitteln zu können. Weiterhin treten sporadische Ereignisse, die die Kommunikation negativ beeinflussen, in kleinen Stichproben oft nicht zu Tage.

Keywords: Testsystem, Zuverlässigkeitsbewertung von industriellen Kommunikationssystemen, Performancemessungen, anwendungsbezogene Kenngrößen, physikalische Kenngrößen

1 Einleitung

Die Zahl der Einsätze von Funkkommunikationslösungen in der Industrie steigt stetig. Die industrielle Umgebung ist für gewöhnlich für Funkkommunikation besonders anspruchsvoll. Belegen nun vermehrt Funkteilnehmer das Funkspektrum kommt es schnell zu gegenseitiger Beeinflussung und die Zuverlässigkeit der Datenübertragung nimmt ab. Um vorab die Realisierbarkeit von Funklösungen in industriellen Anlagen zu untersuchen, können Performancemessungen zur Bewertung der Zuverlässigkeit des jeweiligen Kommunikationssystems durchgeführt werden. Dabei wird untersucht, ob die Kommunikationstechnologie die Anforderungen an Zuverlässigkeit sowie hinreichend geringe Verzögerungen bei großer Teilnehmerzahl und -dichte erfüllt. Andere wichtige Fragestellungen betreffen die Funkabdeckung, die Kompatibilität zu Feldbussystemen bzgl. des Zeitverhaltens (Kommunikationszyklus) oder der Einfluss der Übertragungsrichtung. Dabei ist zu beachten, dass die Anforderungen der verschiedenen Anwendungen wie zum Beispiel Krane, fahrerlose Transportsysteme, Werkstückträger, Service-Roboter, Industrieroboter oder Logistikfahrzeuge stark variieren können.

[1] Institut für Automation und Kommunikation, Industrielle Kommunikation, Werner-Heisenberg-Straße 1, 39106 Magdeburg, sarah.willmann@ifak.eu
[2] Institut für Automation und Kommunikation, Industrielle Kommunikation, Werner-Heisenberg-Straße 1, 39106 Magdeburg, marco.meier@ifak.eu
[3] Institut für Automation und Kommunikation, Industrielle Kommunikation, Werner-Heisenberg-Straße 1, 39106 Magdeburg, andre.gnad@ifak.eu

Um das Zeit- und Fehlerverhalten derartiger Kommunikationslösungen während des Entwicklungsprozesses, im Vorfeld des Einsatzes in einer realen Anwendung, oder auch während des Betriebes bewerten zu können, ist ein hochspezialisiertes Testsystem erforderlich. Mit dem verwendeten Testsystem kann das Kommunikationsverhalten verteilter Anwendungen emuliert werden. Dabei können verschiedenste Datenverkehrsmodelle gezielt parametriert und am realen Kommunikationssystem erprobt werden. Der parallele Einsatz realer Kommunikationsgeräte zur Generierung von zusätzlichem Datenverkehr (Grundlast) ist ebenfalls möglich. So werden sowohl für Entwickler als auch für Anwender aussagekräftige Messwerte geliefert.

Konkret wird auf folgende Aspekte eingegangen:

- Welche Fragestellungen können beantwortet werden, wenn das Fehlerverhalten ausgewertet wird?
- Welche Erkenntnisse können aus Langzeitmessungen im Vergleich zu Stichproben gewonnen werden?
- Gibt es Einflüsse verschiedener Endgeräte auf die Performance?
- Welche Aussagekraft haben anwendungsbezogene Kenngrößen (z. B. Übertragungszeit) und physikalischen Kenngrößen (z. B. Signalstärke) in Bezug auf die Funktionsfähigkeit der Anwendung?
- Welche Herausforderungen bringen die jeweiligen Schnittstellen der Endgeräte mit sich?

2 Grundlagen

Zur Bewertung der Funkkommunikation können verschiedene Kenngrößen herangezogen werden. Zunächst können die Kenngrößen ermittelt werden die sich aus der rein physikalischen Übertragung ergeben. Beim Mobilfunk sind folgende Kenngrößen typisch.

- RSRP Reference Signal Received Power
- RSRQ Reference Signal Received Quality
- RSSI Received Signal Strength Indicator
- SNIR Signal-to-noise-plus-interference ratio

Diese Kenngrößen können zum Beispiel mit Messgeräten von Rohde & Schwarz [RS25], Vivavi [VI25] oder ExFo [EX25] ermittelt werden. Die Kenngrößen beziehen sich auf die Signalstärke oder die Signalqualität bezogen auf die vorhandene gemessene Leistung oder auf einem Referenzsignal. Diese Kenngrößen bewerten allerdings nicht die Performance, die für die Anwendung aussagekräftig ist. Für die Anwendung relevante Kenngrößen beinhalten auch den Einfluss der Kommunikationslösung. Diese Kenngrößen sind in

[Me17] und [Me20] vorgestellt. Ein Konzept zur Zustandsbewertung unter der Nutzung anwendungsbezogener Kenngrößen wird hier vorgestellt [WKR17]. Dabei werden die zu bewertenden Kenngrößen normiert, gewichtet und zu einem Klarzustandswert zusammengefasst. Um aus dieser Zustandsbewertung Handlungen abzuleiten, wurde zur Verwendung der anwendungsspezifischen Kenngrößen ein Konzept zum Koexistenzmanagement [Wi19] vorgestellt.

3 Testkonzepterweiterung

[Gn17] beschreibt ein Testkonzept zur Bewertung des Zeit- und Fehlerverhaltens von industriellen Kommunikationssystemen. Das zu bewertende Kommunikationssystem wird durch eine verteilte Testanwendung umklammert. Ein Teil der verteilten Testanwendung generiert periodisch Nachrichten, die über das zu bewertende Kommunikationssystem an einen anderen Teil der Testanwendung überragen werden. Eine Nachricht kann aus einem oder mehreren Paketen bestehen. Die physische Anbindung an das zu bewertende System erfolgt über ein speziell entwickeltes Messgerät, das Multiface. In dem Testkonzept kamen bisher vorrangig serielle Schnittstellen zur Erzeugung und Verarbeitung von Echtzeitdatenverkehr zur hochauflösenden Ermittlung der Kenngrößen zur Anwendung. Die stetig fortschreitende Digitalisierung stellt jedoch höhere Anforderungen an Funkkommunikationssysteme, die mit klassischen seriellen Protokollen nicht erfüllbar sind:

- Hohe Datenraten (>100 Mbit/s, je nach Anforderungen auch bis zu 1 Gbit/s)
- Globale Erreichbarkeit
- Inhärente IT-Security Maßnahmen

In der industriellen Kommunikation wird als Netzwerktechnologie vorrangig Ethernet als Schnittstelle zu Funkkommunikationsgeräten eingesetzt. Auch für diese Funkkommunikationssysteme ist die Einhaltung der Kenngrößen nach [Me17] und [Me20] essentiell, so dass eine Erweiterung des Testkonzeptes notwendig ist, die die Grundlage für ein Testen von aktuellen WiFi- oder Mobilfunklösungen schafft.

In einem ersten Schritt wurde das Multiface um eine weitere Bezugsschnittstelle für Ethernet erweitert. Um eine größtmögliche Flexibilität zur Anbindung unterschiedlichster Kommunikationssysteme zu gewährleisten, wurde eine FPGA basierte Implementierung des MAC gewählt. Somit können potentiell auch Funkgeräte untersucht werden, die bestimmte sonst notwendige Merkmale der [Ie22b] nicht unterstützen, ohne dass das Multiface modifiziert werden muss.

Bisher konnte sich im Testsystem lediglich auf die Definition der Nutzdaten fokussiert werden. Bei Ethernet-Kommunikation kommen jedoch nahezu immer unterschiedliche Protokolle entlang der Schichten des ISO/OSI -Referenzmodells zum Einsatz. Um eine korrekte Adressierung der Nachrichten zu gewährleisten, ist spezielle Kenntnis über die

jeweiligen Protokolle (Ethernet, IP, UDP, TCP, …) notwendig. Damit Nutzer des Testsystems mit begrenzten Kenntnissen dieser Protokolle dennoch Tests konfigurieren können, wurde ein sogenannter PacketBuilder für den FTT Manager konzipiert, der nur die unbedingt notwendigen Protokollelemente zur Adressierung abfragt. Darüber hinaus wurden praxisorientierte und dem Stand der Technik entsprechende Werte für Paketelemente angenommen. In besonderen Fällen, in denen für einen speziellen Test von diesen Werten abgewichen werden muss, ist auch eine manuelle Konfiguration des gesamten Paketinhaltes vorgesehen. Es wird zwischen verschiedenen Typen von Paketelementen unterschieden, um beim Empfang einer Nachricht bzw. eines Paketes gezielte Auswertung vornehmen zu können. Das Resultat der Auswertung für den Fall einer inhaltlich falschen Nachricht ist eines der folgenden:

- Fehler bei der Adressierung
- Längenfehler
- Inhaltliche Fehler
- Fehler in Prüfsummen

Dies ermöglicht eine gezielte Diagnose des Kommunikationsverhaltens.

Die oben genannte Forderung nach hohen Datenraten lässt sich aufgrund des Overheads der unterschiedlichen OSI Schichten nur mit großen Datenlängen pro Nachricht erfüllen, da hier der relative Anteil des Overheads an der Paketlänge sinkt. Längere Nutzdaten ermöglichen das konsekutive Senden von mehreren Paketen, ohne jeweils eine Bestätigung der Gegenseite abzuwarten. Die Pakete können in mit dem kleinstmöglichen Sendeabstand übertragen werden (Interpacket Gap, 12 Byte bei Ethernet). Diese Art der Kommunikation wird im Folgenden als Streaming bezeichnet. Beispiele für Streaming sind die Videoübertragung via RTP (Real-time Transport Protocol, Übertragung erfolgt über UDP) sowie die Datenübertragung über HTTP (Hypertext Transfer Protocol, Übertragung erfolgt über TCP). Die Anforderungen an die Kommunikation nach [Me17] und [Me20] müssen auch hierfür eingehalten werden, um bspw. eine Fernsteuerung basierend auf Kamerabildern durchführen zu können. Damit auch diese Arten der Kommunikation modelliert und getestet werden können, wurde die Möglichkeit der Nutzung eines dedizierten IP- sowie TCP- und UDP Stacks vorgesehen. Das UDP Streaming ist so konzipiert, dass es die korrekte Reihenfolge der Datagramme auf Empfangsseite verifiziert. Das Ausbleiben von einzelnen Paketen führt bei den meisten UDP basierten Streamingprotokollen nicht zu einem Abbruch der Verbindung, jedoch ist die Information hierüber entscheidend für die Bewertung der Qualität der Kommunikationsverbindung. Eine unvollständige Nachricht wird in diesem Testkonzept vollständig verworfen, da sie für die Applikation nicht nutzbar ist. Bei TCP hat der verwendete Algorithmus für die Stausteuerung (congestion control) einen großen Einfluss auf die Übertragungszeiten. Diese Algorithmen regeln das Sendefenster (congestion window), um dynamisch auf unterschiedliche Qualität der Verbindung reagieren zu können. Die Verwendung eines dedizierten TCP Stacks ermöglicht es, den Einfluss

unterschiedlicher Algorithmen für die Stausteuerung sowie unterschiedlicher Parametrierungen dieser Algorithmen auf die applikativen Kenngrößen analysieren zu können. Somit ist eine allgemeinere Aussage über die Eignung eines Kommunikationssystems für eine bestimmte Anwendung als bei der Verwendung einer einzigen Implementierung möglich.

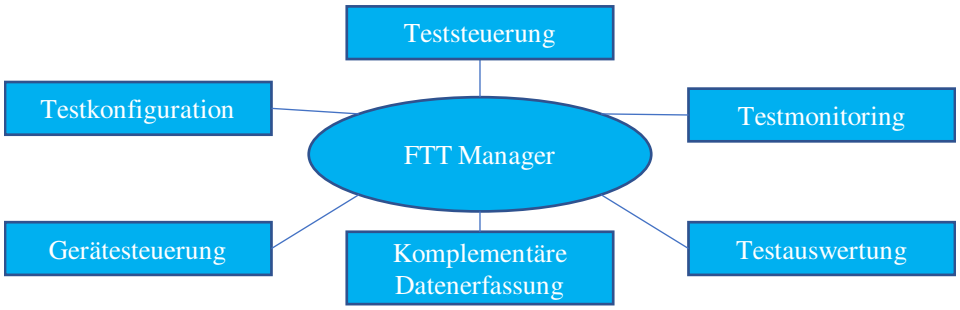


Abb. 1: Funktionalitäten FTT Manager

Die aktuellen Anwendungen von Funkkommunikationssystemen, beispielsweise fahrerlose Transportsysteme (FTS), erzeugen eine zusätzliche Herausforderung bei reproduzierbaren Tests: sie operieren in einer dynamischen Umgebung. Die Position von Geräten des System Under Test (SUT) lässt sich nicht immer a priori festlegen. Es ist daher mindestens notwendig, bestimmte Umgebungsbedingungen parallel zu den eigentlichen Performancetests automatisiert zu dokumentieren. Dies sind bspw. Positionen von Komponenten, aber auch Funkkanalparameter von Funkgeräten. Das Testkonzept wurde um die Möglichkeit erweitert, testspezifische Routinen zur Ermittlung dieser Werte parallel zum Performancetest auszuführen, und diese Werte parallel zu den Werten der Kenngrößen zu dokumentieren. Abb. 1 zeigt eine Übersicht der Funktionalitäten, die der FTT Manager nach dieser Erweiterung enthält. Eine besondere Herausforderung sind hierbei die deutlich unterschiedliche Abtastraten, sowie die Variabilität dieser. Bei logischen Verbindungen im Testsystem ist das Transfer Interval sehr genau definiert und wird auch – sofern von Kommunikationssystem her möglich – sehr deterministisch eingehalten. Andere Protokolle, die zum Auslesen der Ergebnisse verwendet werden, bieten diese Echtzeiteigenschaft jedoch nicht. Die Messwerte werden daher unter Verwendung von hinreichend genauen Zeitstempeln mit den Kommunikationszyklen der Performancemessung synchronisiert. Durch diese Synchronisation sind a posteriori Informationen über die Umgebung zum Zeitpunkt der Messung bzw. des Tests rekonstruierbar.

4 Einsatzmöglichkeiten des Testsystems

4.1 Aussagekraft von anwendungsbezogenen und physikalischen Kenngrößen

Im Projekt REBAKO [Re24] wurden mit Hilfe des FTT, sowohl Werte von anwendungsbezogenen als auch von physikalische Kenngrößen (phyKG) aufgenommen. Ziel der Betrachtung war eine Zustandsanalyse und -bewertung mit Hilfe der Kenngrößen vorzunehmen, um ein Koexistenzmanagement mit den Daten zu versorgen (Abb. 2). Die anwendungsbezogenen Kenngrößen (anwKG) wurden mit dem Multiface gemessen, die phyKG wurden aus den verwendeten Kommunikationsgeräten ausgelesen und im FTT Manager verarbeitet. Zu Zustandsbewertung wurden die Werte nicht immer direkt verwendet, sondern auch deren statistische Parameter oder eine Regression von mehreren Werten. Exemplarisch wurde der Einfluss von Störungen, aus zuvor zusammengetragenen phyKG, untersucht (Abb. 3). Daraus ergaben sich Kandidaten die in einem Validierungsszenario erprobt wurden.

Im Validierungsszenario wurde basierend auf der Zustandsanalyse und -bewertung der Koexistenzzustand beeinflusst. Die Beeinflussung des Koexistenzzustands kann durch eine Änderung der Kommunikation in Frequenz, Zeit und Raum erfolgen oder durch einen Eingriff in die Anwendung. Im Validierungsszenario war die Anwendung ein fahrerloses Transportfahrzeug (FTF). Hier konnte ein Eingriff über die Geschwindigkeit des FTF oder eine räumliche Umfahrung des Störers vorgenommen werden.

Abb. 2: Komponentenübersicht

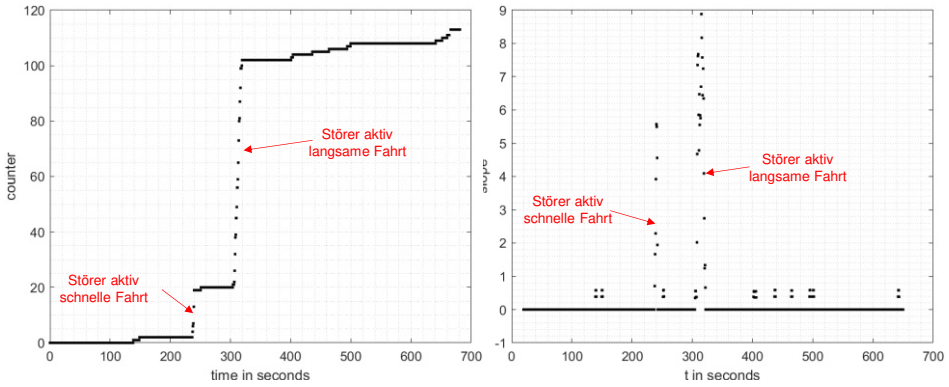


Abb. 3: links: Absolutwert ifOutErrors, rechts: Lineare Regression über die letzten fünf Werte [Wi23]

Zusammenfassend sind zwei Punkte hervorzuheben. PhyKG können für die Anwendung als Indikator für „nicht funktionsfähig“ genutzt werden. Eine praxisrelevante Aussagekraft für die Funktionsfähigkeit der Kommunikation einer Anwendung haben anwKG. Für die Zustandsbewertung der Kommunikationsverbindung ergänzen sich die phyKG und die anwKG und liefern ein aussagekräftiges Ergebnis.

4.2 Erkenntnisgewinn durch erhöhte Stichprobengröße

Wie eingangs erwähnt erschließt die Unterstützung neuer Bezugsschnittstellen vollkommen neue Anwendungsfelder für das Testsystem. Aktuell liegt noch ein starker Fokus auf der Analyse der Einsatzmöglichkeiten von Mobilfunk (5G) für die industrielle Kommunikation. Diese Kommunikationssysteme sind deutlich komplexer als bisherige, in dieser Domäne verwendete. Bei diesen herkömmlichen Funkkommunikationssystemen wie bspw. WiFi [Ie22a], Bluetooth [Bl24], Zigbee [Zi23], IO-Link Wireless [Io23], gibt es stets einen Funkkanal zwischen Sender und Empfänger einer Nachricht. Bei einem 5G System kommt eine gänzlich andere Architektur zum Einsatz. Der Kommunikationsweg von einem Funkgerät (User Equipment, UE) zum anderen ist vereinfacht: UE → 5G-RAN (Radio Access Network) → 5G Core → 5G-RAN → UE (vgl. Abb. 4). Es sind signifikant mehr Komponenten beteiligt, die alle von unterschiedlichen Herstellern kommen können. Die dadurch massiv erhöhte Komplexität zusammen mit der üblicherweise Unkenntnis über die konkreten Implementierungen von UE und Core resultieren in einem von außen nicht deterministisch wirkendem Verhalten. Damit auch für diese Kommunikationssysteme eine Aussage über die Zuverlässigkeit getroffen werden kann, muss mit einer sehr großen Stichprobe und einer langen Messdauer gearbeitet werden.

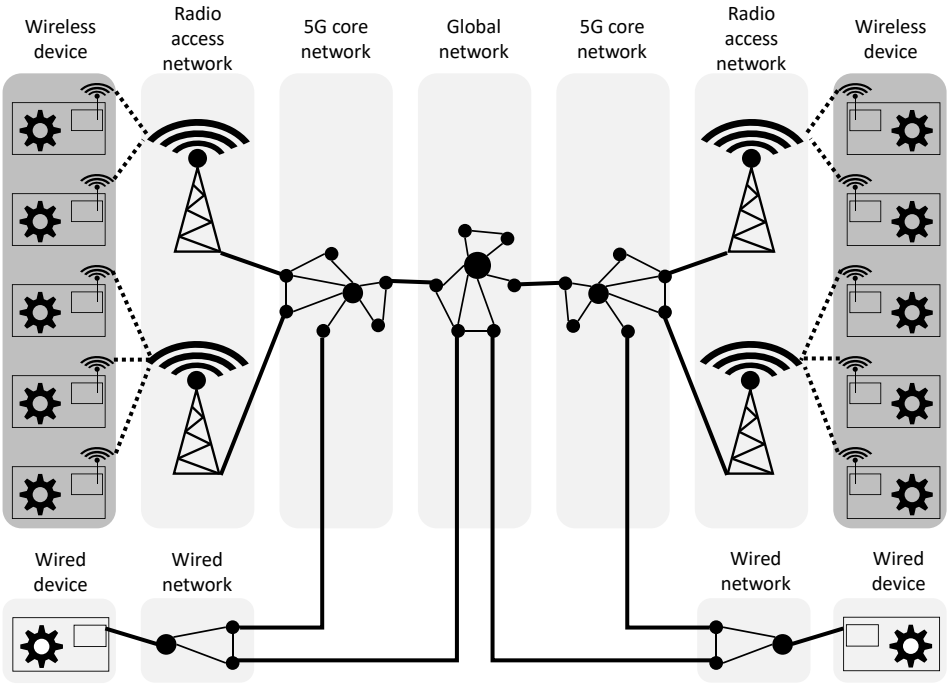


Abb. 4: Kommunikationsweg bei 5G

Nachfolgend sind Beispiele für Erkenntnisse aus realen Messungen aufgeführt, die nur durch diese Langzeittests zu erhalten waren.

- Kurzfristige Erhöhung der Übertragungszeit (TT)

  Es hat sich ein Verhalten gezeigt, dass sich nach einigen Stunden die Übertragungszeit für einen begrenzten Zeitraum um fast eine Größenordnung erhöht hat. Dieses wurde für einige Sekunden beobachtet, danach verhielt sich die Kommunikation ohne manuellen Eingriff wieder wie zuvor. Dieses Verhalten sorgt in einer Anlage zumindest für einen temporären Abbruch der Kommunikation.

- Drift der Übertragungszeit

  Ebenfalls nach einigen Stunden konnte ein Drift der Werte der Übertragungszeit beobachtet werden. Der Durchschnittswert stieg auf ca. das doppelte des ursprünglichen Wertes.

Die Komplexität der Funkkommunikation wurde durch weitere Messungen unterstrichen. Es wurden unter gleichen Umgebungsbedingungen wie Ort, Eigenschaften des Funkkanals und Temperatur Messungen mit verschiedenen UEs durchgeführt. Um Einflüsse der unterschiedlichen Bauformen der UEs auszuschließen, wurde dieselbe Antennenkonfiguration für die Messungen verwendet. Es zeigt sich, dass die

Implementierung der UEs selbst einen entscheidenden Einfluss auf die Übertragungszeit, aber auch auf die erreichbare Datenrate hat.

Diese Messergebnisse unterstreichen die Notwendigkeit, bei der Planung, aber auch beim Aufbau einer Anlage nicht nur Simulationsergebnisse für die Signalstärke zu verwenden, sondern jeweils auch Tests mit den tatsächlichen Komponenten durchzuführen, um die Einhaltung der Anforderungen nach [Me17] und [Me20] abzusichern.

5 Zusammenfassung

Im Paper wird zunächst das Testkonzept erläutert. Mit Hilfe von realen Messungen des FTT, der u.a. aus dem Multiface und dem FTT Manager besteht, wurde zum einen aufgezeigt welche erhöhte Aussagekraft das Zusammenbringen von anwendungsbezogenen Kenngrößen und physikalischen Kenngrößen erzielt werden kann. Des Weiteren wurde am Beispiel von Messungen in einem 5G Campusnetz der Erkenntnisgewinn durch eine Hohe Stichprobengröße herausgestellt. Ein Funktionstest ersetzt keine solide Validierung des installierten Kommunikationssystems in realer Umgebung.

Danksagungen

Die Arbeiten wurden zum Teil unterstützt vom Projekt „Industrial Radio Lab Germany“ (Förderkennzeichen 16KIS1013), gefördert vom Bundesministerium für Bildung und Forschung, dem Projekt „5GIWCoW“ (Förderkennzeichen 165GU041D) gefördert vom Bundesministerium für Digitales und Verkehr, sowie vom Projekt „REBAKO“ (Förderkennzeichen: 21529 BG), gefördert vom Bundesministerium für Wirtschaft und Klimaschutz.

Literaturverzeichnis

[Bl24] Core Specification Working Group: Bluetooth Core Specification v6.0, August 2024

[EX25] FTB 5GPro - Complete, all-in-one 4G and 5G test solution, https://www.exfo.com/en/products/field-network-testing/ethernet-testing/ftb-5gpro/, 15.10.2025

[Gn17] Gnad, A.; Rauchhaupt, L.; Willmann, S.; Oeffler, T: Verteiltes Testsystem zur Bewertung der Zuverlässigkeit von Kommunikationssystemen für Industrie 4.0 Anwendungen, KommA 2017, Magdeburg, 2017

[Me17] VDI/VDE 2185 Blatt 4: Funkgestützte Kommunikation in der Automatisierungstechnik – Messtechnische Performancebewertung von Funklösungen für industrielle Automatisierungsanwendungen, November 2017

[Me20] VDI/VDE 2185 Blatt 1: Funkgestützte Kommunikation in der Automatisierungstechnik – Anforderungen und Grundlagen, August 2020

[Ie22a] IEEE: 802.11-2020/Cor 1-2022 – IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks--Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Corrigendum 1 – Correct IEEE 802.11ay Assignment of Protected Announce Support bit, Dezember 2022

[Ie22b] IEEE: 802.3-2022 - IEEE Standard for Ethernet, 2022

[Io23] IO-Link Community: IO-Link Wireless – System Extensions, June 2023

[Re24] DFAM, Deutsche Forschungsgesellschaft für Automatisierung und Mikroelektronik e.V., REBAKO Abschlussbericht, Stand: 30.09.2024

[RS25] R&S®TSMA6 Autonomous Mobile Network Scanner Bedienhandbuch, https://www.rohde-schwarz.com/de/handbuch/r-s-tsma6-autonomous-mobile-network-scanner-bedienhandbuch-handbuecher_78701-572421.html, 15.10.2025

[VI25] OneAdvisor 800 Wireless Plattform, https://www.viavisolutions.com/de-de/produkte/oneadvisor-800-wireless-plattform#overview, 15.10.2025

[Wi19] Willmann, S.; Meier, M.; Rauchhaupt, L.; Wiebusch, N.; Meier, U.: "Towards Validation of Wireless Coexistence Management," IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 2019, pp. 2855-2862, doi: 10.1109/IECON.2019.8926957.

[Wi23] Willmann, S.: Zuverlässigkeitsbewertung von industrieller Funkkommunikation; Manuskript Dissertation, 2023

[WKR17] Willmann, S.; Krätzig, M.; Rauchhaupt, M.: "Methodology for holistic assessment of dependability in wireless automation," 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 2017, pp. 1-4, doi: 10.1109/ETFA.2017.8247724.

[Zi23] Connectivity Standards Alliance: Zigbee Specification, Revision 23, March 2023

http://dx.doi.org/10.25673/121097

Studie zur Wirtschaftlichkeit von Ethernet-APL

Kostenvergleich einer Referenzanlage: 4-20mA Technologie versus Ethernet-APL im CapEx Bereich

Sedat 1 Yildirim 1[1], Alen 2 Mahendrarajah 2[2] und Norbert 3 Große 3[3]

Abstract: Trotz der weit verbreiteten Dominanz der 4-20 mA-Technologie gewinnt Ethernet-APL zunehmend an Bedeutung und verspricht signifikante Vorteile über den gesamten Anlagenlebenszyklus. Doch wie wirkt sich die neue Technologie tatsächlich auf die Investitionskosten aus? Um diese Frage zu beantworten, hat der NAMUR APL Taskforce eine Referenzanlage definiert, auf deren Basis drei unterschiedliche Anlagenkonzepte entwickelt und verglichen wurden: 4-20 mA mit Stammkabel (Variante A), 4-20 mA mit Remote I/Os (Variante B) und Ethernet-APL (Variante C). Der Fokus der Untersuchung lag auf den Hardwarekosten sowie den zentralen Arbeitsschritten bei Planung, Installation, Integration und Inbetriebnahme der Anlage. Während eine direkte monetäre Bewertung der Betriebsvorteile aufgrund der fiktiven Anlage nicht möglich ist, zeigt die Analyse dennoch an einigen Stellen deutliche Unterschiede zwischen den Technologien, die interessante Erkenntnisse für die zukünftige Anlagenplanung liefern

Keywords: Ethernet-APL, Wirtschaftlichkeit, Systemvergleich, Engineering, Installation, Integration, Inbetriebnahme

1 Einleitung

Traditionell sind Feldgeräte einzeln über Punkt-zu Punkt-Verbindungen mit einem Leitsystem verbunden. Dies erfordert einen hohen Verkabelungsaufwand mit Stammkabeln und den Einsatz von Rangierverteilern im Schaltraum und Feldverteilern im Feld. Die Einführung von Remote-I/O-Systemen brachte hier Verbesserungen, da sie eine dezentrale Erfassung der Prozesssignale im Feld ermöglicht. Hierdurch können Platz und Installationskosten eingespart werden. Die 4-20 mA/HART-Technologie hat sich bis heute aufgrund ihrer hohen Zuverlässigkeit und geringen Komplexität als Standard etabliert. Im Gegensatz dazu haben sich komplexere Feldbussysteme wie z. B. PROFIBUS PA und Foundation Fieldbus (FF) nur begrenzt durchsetzen können. Ein vielversprechender Ansatz ist der Einsatz von Ethernet als Schlüsseltechnologie in der Prozessindustrie. Ethernet hat sich derzeit vor allem in der Informationstechnik und in der Fertigungsautomatisierung etabliert. Hier bietet der Einsatz bereits heute eine Vielzahl von Vorteilen, die Installation, Integration, Inbetriebnahme und Betrieb effizienter machen.

[1] Technische Hochschule Köln, PLT-Labor, Betzdorferstraße 2, 50679 Köln, yildirim@plt-labor.de,
[2] Technische Hochschule Köln, PLT-Labor, Betzdorferstraße 2, 50679 Köln, mahendrarajah@plt-labor.de,
[3] Technische Hochschule Köln, PLT-Labor, Betzdorferstraße 2, 50679 Köln, grosse@plt-labor.de,
https://www.linkedin.com/company/plt-labor/?originalSubdomain=de

Nun soll Ethernet-APL (Advanced Physical Layer) zunehmend auch in der Prozessindustrie eingesetzt werden. Als innovative Lösung ermöglicht Ethernet-APL eine direkte digitale Datenübertragung von der Feldebene zur Prozessleitebene mit 10 Mbit/s. Dies erlaubt im Vergleich zu HART schnellere Parametrierung, effizientere Updates und einfache Fernwartung. Zudem bietet Ethernet-APL eine robuste, ex-geschützte Technik, die speziell für die Prozessindustrie ausgelegt ist und Geräte mit Energie versorgt.

Die Nutzung der Vorteile setzt die Investitionsbereitschaft der Anlagenbetreiber und ein breites Angebot der Hersteller voraus. Bisher fehlen praktische Erfahrungen und belastbare Informationen, was bei Anwendern zu Unsicherheit über Einsparpotenziale und bei Herstellern zu Zurückhaltung bei der Entwicklung führt. Diese Unsicherheit hemmt Investitionen. Daher ist es wichtig, die Einsparpotenziale aufzuzeigen. Die vier Chemieunternehmen Bayer, Covestro, Currenta und Lanxess haben sich zusammengeschlossen und das PLT-Labor beauftragt, die Wirtschaftlichkeit dieser Technologie zu untersuchen.

1.1 ZIEL DER STUDIE

Das Ziel dieser Studie ist ein objektiver und umfassender Vergleich dreier alternativer und repräsentativer Technologiekonzepte zur Anbindung der Feldebene: konventionelle 4-20mA-Technologie mit Stammkabeln (A), 4-20mA mit Remote-I/Os im Feld (B) und Ethernet-APL (C). Variante A und B stellen die am weitesten verbreiteten Lösungen in der Prozessindustrie dar, wohingegen Ethernet-APL eine neue Technologie ist. Damit Ethernet-APL sein volles Potenzial entfalten kann, sind sowohl Investitionen der Anlagenbetreiber als auch ein breites Geräteangebot der Hersteller erforderlich. Da bislang belastbare Informationen und insbesondere Vergleiche zu den Investitionskosten fehlen, untersucht diese Arbeit die Chancen und Herausforderungen von Ethernet-APL im Lebenszyklus einer Anlage und beleuchtet insbesondere die Investitionskosten.

1.2 METHODIK

Um eine vergleichbare Basis für die Analyse zu schaffen, wurden von den erfahrenen Ingenieuren der NAMUR APL Task Force spezifische Anforderungen an das Mengengerüst einer Referenzanlage im industriellen Maßstab definiert, um hieran die Technologien mit ihren verschiedenen Kostenaufwänden zu vergleichen. Die Bewertung des PLT-Labors erfolgt insbesondere unter dem Gesichtspunkt der Investitionskosten (Capital Expenditures, CapEx), um Unterschiede zwischen den drei Technologiekonzepten aufzuzeigen. Die Engineering-, Installations-, Integrations- und Inbetriebnahmekosten werden in Anlehnung an die Arbeitsschritte der NAMUR-Arbeitsblätter NA35 [2] und NA186 [3] definiert und evaluiert. Zur Validierung der Methodik und Ergebnisse wurden Gespräche mit Anwendern und Systemintegratoren geführt, um den Aufwand genau zu quantifizieren.

1.3 RAHMENBEDINGUNGEN

Die drei konzipierten Lösungen sind als Greenfield-Anlagen klassifiziert. Hinsichtlich der Anforderungen an Kapazitätsreserven und Verfügbarkeit sollen alle Lösungen möglichst gleichwertig sein, um eine vergleichbare Basis zu schaffen.

Für die Hardware-Kosten wurden Listenpreise der Hersteller verwendet, während die Kosten für die Dienstleistungen auf Basis von Stundensätzen berechnet und mit Daten aus Gesprächen mit Anwendern und Systemintegratoren validiert wurden. Beispielsweise soll der Stundensatz eines Ingenieurs, der das Basic Engineering durchführt, in der Studie mit 180 €/h angenommen werden. Den Preisen liegen europäische Standards zugrunde.

Die Betriebskosten (Operational Expenditures, OpEx) werden nicht monetär bewertet, sondern anhand von Anwendungsfällen im Betrieb analysiert, um die Auswirkungen auf die Anlagenverfügbarkeit zu veranschaulichen. Die Studie konzentriert sich auf technologische Unterschiede, bewertet nur relevante Aspekte und schließt Kosten für Sicherheit und Schulung aus, da diese noch außerhalb des Untersuchungsrahmens liegen.

2 UMSETZUNG DER REFERENZANLAGE

2.1 ANFORDERUNGEN DER NAMUR APL TASK FORCE

In Tabelle 1 ist das von der NAMUR APL Task Force vorgegebene, repräsentative Mengengerüst für die verschiedenen Signaltypen und Ex-Zonen dargestellt.

Signaltyp	Keine Zone	Zone 2	Zone 1
Analoger Eingang	0	500	200
Diskreter Eingang (inkl. Ein-/ Aus-Ventilendschalter)	0	300	100
Analogausgang (zur Steuerung von Ventilen)	0	150	50
Diskreter Ausgang (inkl. Magnetventile für Ein/Aus-Ventile)	0	150	0
Frequenzumrichtern	20	0	0
Motorschutzmodule	100		
Gesamt	120	1100	350

Tab. 1: Anforderungen der NAMUR APL Taskforce, *NAMUR*

Zu erkennen ist, dass es in den Anlagen zwei Explosionsschutzzonen (Zone 1 und Zone 2) gibt mit insgesamt 1.570 Ein- und Ausgangssignalen. Außerhalb der Explosionsschutzzonen befinden sich lediglich die Frequenzumrichter, die als Steuerungsschnittstelle zu den Motoren dienen.

2.2 VARIANTE A: 4-20 mA MIT STAMMKABELN

Variante A basiert auf der konventionellen 4- 20 mA-Technologie mit Stammkabeln, die analoge und binäre Signale über Punkt-zu-Punkt-Verbindungen von der Feldebene in die Prozessleitebene übertragen. Die Feldgeräte sind über HART parametrierbar. Wie in Abbildung 1 dargestellt, werden zwischen Prozess- und Leitebene mehradrige 4-20 mA Stammkabel zur Weiterverarbeitung der Signale eingesetzt. In der Leitebene befinden sich die Kommunikationskarten, an die ein Leitsystembus angeschlossen ist. Die Anbindung der Frequenzumrichter erfolgt über einen redundanten PROFIBUS DP um die Ausfallsicherheit zu erhöhen.
Für die Signalübertragung aus der Ex-Zone 1 werden eigensichere Ex i Leitungen verwendet. Die Stammkabel zwischen Unterverteilung und Rangierverteiler müssen ebenfalls eigensicher sein. Zusätzlich sind im Schaltraum Speisetrenner erforderlich, um die Versorgung der eigensicheren Feldgeräte in Zone 1 und Zone 2 sicherzustellen.

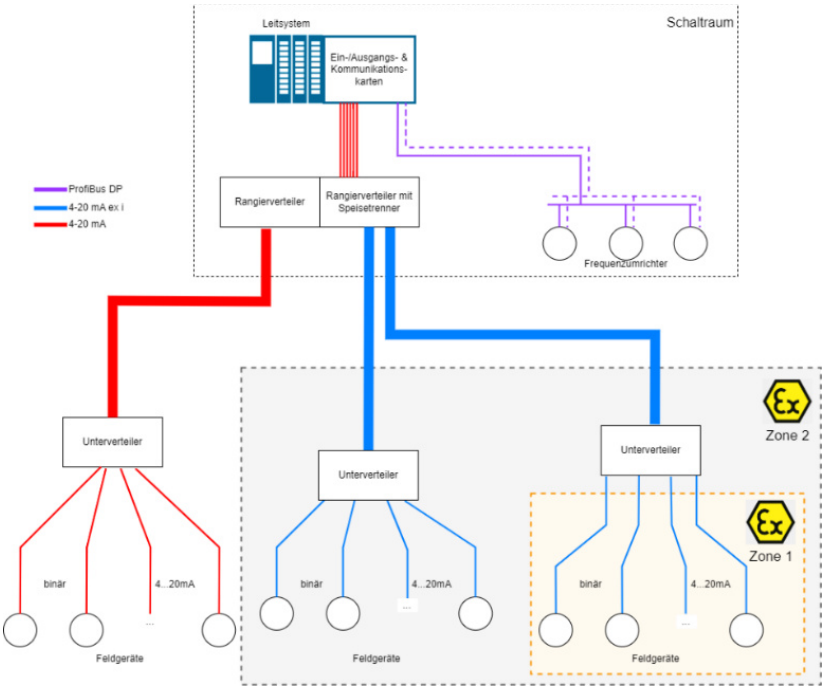


Abb. 1: Netzwerkstruktur der Variante A - 4-20 mA mit Stammkabeln

2.3 VARIANTE B: 4-20mA MIT REMOTEI/Os

Abbildung 2 zeigt die Topologie der Variante B, welche ebenfalls die 4-20 mA Technologie in der Feldebene verwendet. Jedoch werden für diese Variante die Signale über Remote-I/Os im Feld erfasst. Die Remote I/Os sind in einer Linientopologie über einen redundanten PROFIBUS DP-Backbone verbunden.
Für den Einsatz in explosionsgefährdeten Bereichen werden Ex i-geschützte PROFIBUS DP-Leitungen und DP-DP Ex i Koppler verwendet. Die Frequenzumrichter sind wie bei der Variante A über den PROFIBUS DP Backbone mit dem Leitsystem verbunden. Die Feldgeräte können auch hier über HART parametriert werden.
Obwohl diese Topologie auch mit einem PROFINET Backbone realisiert werden könnte, wurde bewusst eine Realisierung mittels PROFIBUS DP gewählt, da dieser in bestehenden Anlagen am weitesten verbreitet ist.

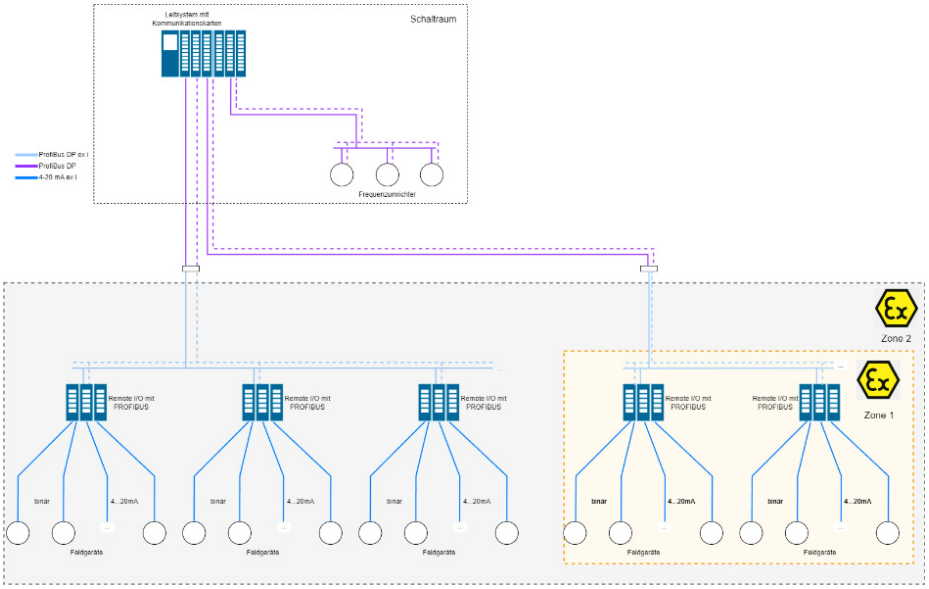


Abb. 2: Netzwerkstruktur der Variante B - 4-20 mA mit Remote I/Os.

2.4 VARIANTE C: ETHERNET-APL

Die in Abbildung 3 dargestellte Topologie der Variante C basiert auf einer PROFINET-Netzwerkstruktur, die eine einfache Ringstruktur mit dem Redundanzprotokoll MRP (Media Redundancy Protocol) verwendet. Dabei wird eine PROFINET S2-Verbindung zwischen Leitsystem und MRP-Ring, die gemäß NAMUR NE 168 [3] als empfohlener Standard für Neuanlagen gilt genutzt. Hierdurch ist eine hohe Verfügbarkeit des Gesamtsystems durch Medien- und Systemredundanz gewährleistet.

Ein Ring besteht aus zwei Managed Switches, sowie APL-Field-Switches und Remote-I/Os im Feld zur Verarbeitung von Binärsignalen. Diese Komponenten sind in dezentralen Unterverteilungen im Feld untergebracht. Durch die Nutzung der PROFINET S2-Verbindung wird die redundante Anbindung an das Leitsystem über zwei Kommunikationskarten realisiert.

Im Schaltraum werden PROFINET-Patchkabel zur Verbindung der Managed Switches und der Kommunikationskarten eingesetzt. Im Feld werden PROFINET-Verlegekabel eingesetzt. Zusätzlich bieten die Remote-I/Os und APL-Field Switches die Möglichkeit Lichtwellenleiter (LWL) über SFP-Transreceiver direkt anzuschließen.

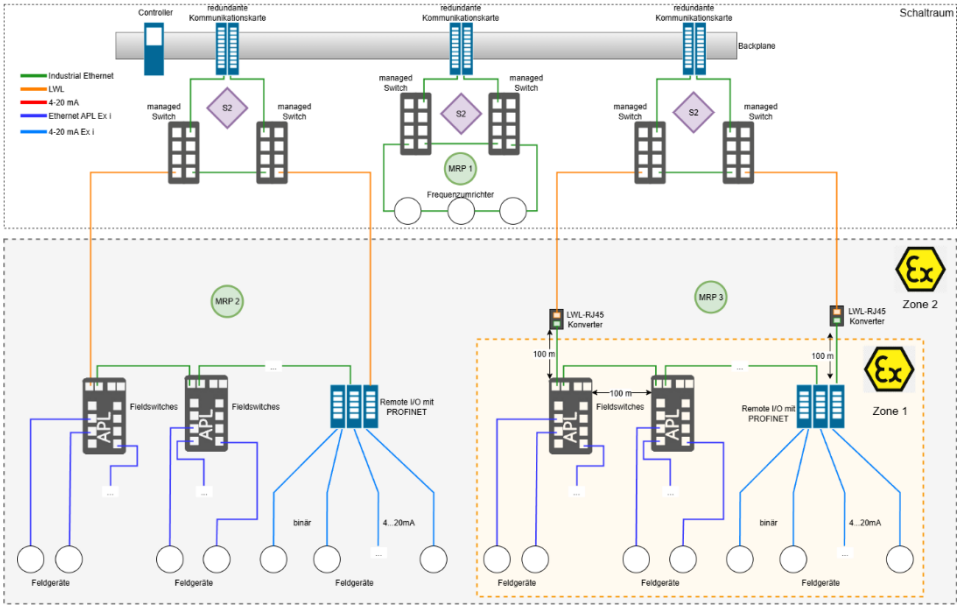


Abb. 3: Netzwerkstruktur der Variante C – Ethernet-APL.

3 VERGLEICH DER INVESTITIONSKOSTEN

Auf Grundlage des Mengengerüsts der drei Lösungen können die Investitionskosten der drei Technologiekonzepte für die Referenzanlage unter Berücksichtigung der Lebenszyklusphasen bis zum Betrieb der Anlage analysiert werden. Zu diesem Zweck werden die Unterschiede zwischen den Varianten untersucht, um die Auswirkungen auf die Investitionskosten detailliert darzustellen.

3.1 ANSCHAFFUNGSKOSTEN DER HARDWARE

Die Spezifikation der Feldgeräte ist bei allen drei Varianten gleich. Jedoch können bei den Varianten A und B die gleichen 4-20 mA Feldgeräte genutzt werden, während bei Ethernet-APL PROFINET over APL-Feldgeräte verwendet werden müssen. Die Kosten der Binärsignalgeber wurden nicht gesondert betrachtet, da ihre Funktionalität bei allen Technologien gleich ist. Eine Ausnahme bilden die Grenzwertschalter, die bei A und B zur Endlagensignalisierung der Ventile verwendet werden. Hier sind als Annahme 80% der Ventile mit jeweils zwei Grenzwertschaltern ausgestattet. Bei Ethernet-APL wird die Funktionalität der Grenzwertschalter bereits im Stellungsregler integriert, so dass Grenzwertschalter nicht mehr benötigt werden. Es hängt von der Philosophie des Unternehmens ab, ob Grenzwertschalter an Ventilen eingesetzt werden oder ob man dies über das Protokoll realisiert. Die Netzwerkstruktur umfasst das Leitsystem (Controller und E/A-Baugruppen), die Feldverteiler mit allen Hardware-Komponenten sowie die Verkabelung (Spur- und Backbone-Leitungen). Variante A erfordert zusätzlich Rangierverteiler und Speisetrenner im Schaltraum, die einen erheblichen Kostenfaktor darstellen. Bei den Varianten B und C entfällt dieser Aufwand durch den Einsatz von Feldbus, bzw. Ethernet-Netzwerken. Bei Variante B entstehen Kosten für Koppler, Feldbusabschlusswiderstände, Repeater und universelle Ein-/Ausgabekarten. Bei Variante C entstehen Kosten auf APL-Field-Switches und Remote-I/Os – für verbleibende Binärsignale werden Stand heute weiterhin Remote-I/Os benötigt. Diese sind hierbei mit PROFINET-Schnittstelle ausgeführt. Für die direkte Bereitstellung von Binärsignalen per Ethernet-APL arbeiten Hersteller bereits an einer Lösung um Remote-I/Os zu ersetzen. Die Bewertung der Beschaffungskosten beschränkt sich auf die technischen Komponenten und das Gehäuse der Verteilerkästen. Tragschienen und Kabelkanäle wurden nicht berücksichtigt. Somit sind folgende Kosten für die Hardware bei den drei Möglichkeiten zu erwarten:

Kostenpunkt	**Variante A**	**Variante B**	**Variante C**
Netzwerk-struktur	924.773 €	593.550 €	593.693 €
Feldgeräte	4.587.872 €	4.587.872 €	4.847.340 €
Gesamt	**5.512.645 €**	**5.181.422 €**	**5.441.033 €**

Tab. 2: Zu erwartende Beschaffungskosten für die drei Lösungen.

3.2 PLANUNG UND ENGINEERING

In diesem Abschnitt werden die Unterschiede im Basic- und Detail-Engineering der drei Technologien analysiert. Die betrachteten Arbeitsschritte und die zu erstellende Dokumentation orientieren sich am NAMUR-Arbeitsblatt 35 (NA 35 [1]), das als Standard für die Strukturierung von Projekten in der Prozessindustrie dient. Es werden hierbei die Unterschiede zwischen den Technologien betrachtet, während die Punkte, bei denen keine Unterschiede zu erwarten sind, nicht bewertet werden. Im Basic Engineering gibt es wenig Unterschiede zwischen den drei Technologien, da grundlegende Aufgaben wie die Definition der Anforderungen, die Erstellung von R&I Fließbildern und PLT-Stellenblättern identisch sind. Im Detail Engineering gibt es keine Unterschiede bei der Spezifikation von Feldgeräten, der Erstellung von Messstellenlisten, Stromlaufplänen und Funktionsplänen. Die in dieser Studie betrachteten Tätigkeiten im Engineering sind:

- Gerätenamen- und Adressvergabe
- Planung des Schaltraum und der Installationen
- Planung der Feldverteiler
- Planung der Kabel und Kabelwege
- Nachweis der Eigensicherheit
- Netzwerkplanung
- PLT-Stellenpläne

Hieraus haben sich für jede Aufgabe unterschiedliche Einsparmöglichkeiten ergeben. Das Basic-Engineering der Schalträume umfasst die Festlegung der Komponenten, deren Abmessungen, die Raumgröße sowie die Wärmeberechnung und eventuelle Brandschutzmaßnahmen. Aufgrund der platzintensiven Rangierverteiler und Feldbarrieren für 4-20 mA mit Stammkabeln (A) ist der Schaltraum hier deutlich größer, während bei Remote-I/Os (B) und Ethernet-APL (C) durch den Wegfall dieser Komponenten bis zu zwei Drittel weniger Platz benötigt wird, was Bau- und Klimatisierungskosten reduziert. Die Planung der Verteilerkästen unterscheidet sich bei den drei Varianten nur geringfügig. Während bei Variante A und Ethernet-APL (C) ca. 23 Minuten pro Verteiler eingeplant werden, erfordert Variante B aufgrund der Remote I/Os mit 25 Minuten etwas mehr Aufwand, der sich jedoch durch die geringere Anzahl an Verteilern relativiert. Die Planung der Kabel und Kabelwege wird bei Ethernet-APL etwas vereinfacht, da anstelle der dickeren Stammkabel für Variante A dünnere PROFINET-Verlegekabel oder Lichtwellenleiter verwendet werden können. In Variante C wird der PROFINET Backbone in einer Ringstruktur aufgebaut, was die Anzahl der zu planende Kabel reduziert und den Planungsaufwand verringert. Der Aufwand für das Zeichnen der Kabelwege im CAD-Tool bleibt jedoch unverändert, da die Kabelwege unabhängig von der Anzahl und Art der Kabel im 3D-Modell hinterlegt werden müssen. Der Nachweis der Eigensicherheit erfolgt bei Ethernet-APL durch das 2-WISE-Konzept, der im Vergleich zu den klassischen Methoden weniger Schritte wie Parametersuche und Berechnung der

zulässigen Werte erfordert. Dadurch reduziert sich der Zeitaufwand von 15 Minuten auf 5:30 Minuten pro Nachweis, der jedoch für identische Betriebsmittel wiederholt werden kann. Das bedeutet, dass die Einsparung pro Typical angenommen werden kann, jedoch nicht pro Feldgerät. Die Netzwerkarchitektur ist bei der Variante A einfacher zu planen, da sich die Topologie auf den Systembus in der Leitwarte beschränkt, während bei den Varianten B und C die Netzwerkteilnehmer über die gesamte Anlage verteilt sind, was eine aufwändigere Planung erfordert. Insbesondere bei Ethernet-APL müssen zusätzlich alle Feldgeräte als Netzwerkteilnehmer mit IP-Adressen im Netzwerkplan hinterlegt werden. Der größte Unterschied im Engineering zwischen den drei Referenzanlagen liegt in den PLT-Stellenplänen, die die Klemmstellen des Signalpfades darstellen. Variante A erfordert mit 30 Minuten pro PLT-Stelle den höchsten Aufwand, da in der angenommenen Lösung insgesamt 18 Klemmstellen dokumentiert werden müssen, während Variante B durch den Einsatz von Remote-I/Os nur 15 Minuten benötigt, da weniger Klemmstellen eingezeichnet werden müssen. Ethernet-APL vereinfacht die Erstellung weiter, da bei verpolungssicheren Anschlüssen keine Adern einzeln dokumentiert werden müssen. Somit ist es möglich die Feldinstallation mittels einer PLT-Stellenliste zu definieren, die pro Gerät die wichtigsten Informationen wie z. B. APL-Switch mit dem zugehörigen Port spezifiziert. Hierfür können 5 Minuten je PLT-Stelle angenommen werden.

3.3 INSTALLATION

Die Arbeitsschritte für Installation, Integration und Inbetriebnahme wurden in Anlehnung an das NAMUR-Arbeitsblatt NA 186 [2] definiert, wobei nur die Schritte berücksichtigt wurden, die sich zwischen den Technologien unterscheiden. In dieser Betrachtung werden folgende Tätigkeiten betrachtet und bewertet:

- Installation der Spurleitung am Feldgerät und Feldverteiler
- Verlegen und konfektionieren der Trunk-Leitung
- Installation im Schaltraum

Als Grundannahme wurde ein Zeitaufwand von 2 Minuten pro Klemmstelle angenommen, mit Ausnahme der Trunk-Leitung, hierbei konnten die Kosten pro Meter für das Verlegen abgeschätzt werden. Die Montage der Feldgeräte unterscheidet sich bei den Technologien kaum, da bei allen Varianten eine 2-Draht-Leitung zum Feldverteiler geführt und verdrahtet wird. Ein Unterschied besteht in der Schirmung. Bei 4-20 mA wird der Schirm einseitig aufgelegt, während bei Ethernet-APL eine beidseitige Schirmauflage erforderlich ist, um Störeinflüsse und Einkopplungen zu minimieren. Bei Ethernet-APL reduziert sich der Verkabelungsaufwand durch den Wegfall von 320 Stichleitungen für Grenzwertschalter. Die Installation des Backbones ist bei 4-20 mA mit Stammkabeln (A) am aufwendigsten, da 110 Stammkabel mit Gesamtkosten von 107.360 € verlegt werden müssen, während bei 4-20 mA mit Remote I/Os im Feld (B) die Kosten durch leichtere PROFIBUS-Kabel und weniger Leitungen auf 6.467 € sinken. Ethernet-APL verwendet eine Kombination aus PROFINET-Kabeln und Lichtwellenleitern, wobei die aufwendigere Installation der empfindlichen Lichtwellenleiter zu etwas höheren Kosten

als bei Variante B führt. Die Installation im Schaltraum ist bei 4-20 mA mit Stammkabeln am aufwändigsten, da ca. 14.100 Klemmstellen angeschlossen werden müssen, was einen Aufwand von 470 Stunden erfordert. Bei 4-20 mA mit Remote I/Os und Ethernet APL reduziert sich der Aufwand erheblich, da nur noch wenige Komponenten in der Leitwarte installiert werden müssen. Dies sind die Controller, die Kommunikationskarten und, bei Ethernet-APL, die Managed Switches. Dies kann mit einem Aufwand von ca. 2 Stunden bewertet werden, was einer Einsparung von 468 Stunden entspricht.

3.4 INTEGRATION

Die in der Integration durchgeführten Tätigkeiten umfassen: Geräteeinbindung in das Prozessleitsystem, Konfiguration der Feldgeräte, Programmierung der Feldgeräte. Bei den Varianten A und B werden die HART-Feldgeräte über ein Remote-I/O eingebunden, wobei klar definiert werden muss, welches Signal an welchem Kanal anliegt, was ca. 20 Minuten pro Remote-I/O erfordert. Bei Ethernet-APL entsteht durch die direkte Einbindung der APL-Feldgeräte ein höherer Aufwand, der sich aber durch die vereinfachte Handhabung und die schnellere Datenrate auf 5 Minuten pro Gerät und 3 Minuten pro Field-Switch reduziert. Die Konfiguration ist mit Ethernet-APL wesentlich schneller, da die erhöhte Datenrate bis zu zehnmal schnelleres auslesen und laden der Daten ermöglicht. Während die Konfiguration eines 4-20 mA Feldgerätes ca. 30 Minuten pro Gerät dauert, benötigt man bei Ethernet-APL nur 5 Minuten pro Gerät. Bei der Programmierung der PLT-Stellen ist bei 4-20 mA eine Skalierung der Eingänge erforderlich, um den Stromwert in den entsprechenden Messwert umzurechnen, während dies bei Ethernet-APL aufgrund der digitalen Erfassung nicht erforderlich ist. Da die Eingangsmodule jedoch auch für Diagnosezwecke verwendet werden, ist es bei Ethernet-APL dennoch sinnvoll, Grenzwerte zu definieren, was 5 Minuten pro Feldgerät erfordert.

3.5 INBETRIEBNAHME

Die Inbetriebnahme umfasst die Funktionsprüfung der PLT-Stelle sowie die Korrekturmaßnahmen bei fehlerhafter elektrischer Installation. Bei den 4-20-mA-Varianten sind ein Leitsystemtechniker im Schaltraum und ein Stellenprüfer im Feld erforderlich, was zweimal 30 Minuten pro PLT-Stelle erfordert. Bei Ethernet-APL genügt eine Person am Asset Management System (AMS), da sich die Geräte selbst melden und der Aufwand auf 10 Minuten pro PLT-Stelle reduziert wird, wobei durch den Wegfall der Grenzwertschalter auch weniger Loops zu prüfen sind. Hinzu kommt, dass der Einsatz von Ethernet-APL Verpolungsfehler verhindert, da die Kabel am Feldgerät polaritätsunabhängig angeschlossen werden können. Im Gegensatz dazu treten bei der konventionellen Technik erfahrungsgemäß Fehler bei etwa 10 % der Messstellen auf, was zu erhöhtem Aufwand bei der Inbetriebnahme führt, während sich die Fehlerquote bei Remote-I/Os durch weniger Rangierungen halbiert und bei Ethernet-APL die Fehlerquote null sein wird.

4 BETRIEBLICHE FAKTOREN

Der Einsatz von Ethernet APL zeigt bereits heute positive Auswirkungen auf den Betrieb. Ethernet-APL ist eine Enabling-Technologie für NOA (NAMUR Open Architecture), da es die parallele Ausschleusung großer Datenmengen direkt aus dem Feld über PROFINET ermöglicht – und dies direkt herstellerneutral im einheitlichen PA-DIM-Format. Bei Einsatz konventioneller Technik ist dies nur mit hohem Hardware-Aufwand über HART-Multiplexer und ein Asset Management System (AMS) mit NOA-Schnittstelle realisierbar wäre. Hinzu kommt der Aufwand für das Engineering der Gateways mit den Datenpunkten, was zusätzliche Stunden in der Planung bedeutet. Außerdem gibt es erweiterter Gerätediagnosemöglichkeiten. Auf der einen Seite können mehr und schneller Diagnosedaten zur Verfügung gestellt werden. Auf der anderen Seite kann die Diagnose über das AMS, den Controller oder auch über den Web-Server der Geräte gefahren werden. Ein weiterer Vorteil ist das herstellerunabhängige Plug-and-Produce, bei dem Feldgeräte dank generischer Gerätetreiber (PA-Profil [4]) ohne Neukonfiguration ausgetauscht werden können. Eine monetäre Bezifferung der betrieblichen Faktoren wurde im Rahmen dieses Vergleichs nicht vorgenommen.

5 BEWERTUNG DES VERGLEICHS

Eine tabellarische Übersicht aller Kostenpositionen ermöglicht die Ermittlung der Gesamtkosten für die betrachteten Phasen. Die detaillierte Darstellung der Kostenpositionen mit den zusammengefassten Planungsschritten sowie der Gesamtkosten für die drei Lösungen ist in Tabelle 1 dargestellt.

Phase	**4-20mA mit Stammkabel (A)**	**4-20mA mit Remote-I/Os (B)**	**Ethernet-APL (C)**
Beschaffung der Hardware	5.512.645 €	5.181.422 €	5.441.033 €
Schaltraum	93.000 €	37.000 €	37.000 €
Engineering	98.490 €	55.465 €	31.270 €
Installation	158.577 €	25.914 €	28.531 €
Integration	85.600 €	85.600 €	36.640 €
Inbetriebnahme	233.280 €	222.480 €	33.280 €
Gesamt	**6.181.592 €**	**5.607.881 €**	**5.607.754 €**

Tabelle 1: Gesamtkosten aller Phasen im Investitionsbereich des Anlagenlebenszyklus

Die Analyse zeigt, dass Variante B (4-20 mA mit Remote-I/Os) und die Variante C (Ethernet-APL) gleich teuer in der Beschaffung sind. Variante A (4-20 mA mit Stammkabeln) stellt derzeit die teuerste Möglichkeit dar. Trotz anfänglich hoher Hardware-Kosten aufgrund der Entwicklungskosten und der Marktneuheit von Ethernet-APL bietet die Technologie bereits heute Vorteile in personalintensiven Phasen, insbesondere angesichts des Fachkräftemangels in Installation, Integration und Inbetriebnahme. Dadurch verkürzt sich auch die Zeit bis zur Inbetriebnahme einer Anlage, so dass früher Erträge erwirtschaftet werden können. Die in dieser Betrachtung erzielten Ergebnisse sind als Tendenz für zukünftige Projekte zu verstehen. Die tatsächlichen Kosten variieren je nach Projekt und Anwender, da individuelle Preise und Bedingungen das Ergebnis beeinflussen.

6 ZUSAMMENFASSUNG

Die Ergebnisse dieses Vergleichs sind sowohl für die Hersteller als auch für die Anwender der Technologie von großer Bedeutung. Für Hersteller stellt Ethernet-APL eine große Marktchance dar, da die Technologie bereits zahlreiche Anwendungsfälle ermöglicht und als zukunftsweisend gilt. Um die Marktdurchdringung zu fördern, sind jedoch Strategien wie eine günstigere Preisgestaltung und die Entwicklung von Lösungen erforderlich, die den Einsatz von Remote-I/Os für binäre Signale überflüssig machen. Anwender profitieren trotz höherer Hardware-Kosten von erheblichen Einsparpotenzialen im Engineering, in der Installation, Integration und Inbetriebnahme. Die Vorteile im Betrieb beim laufenden Auswerten der Daten können effizientere Produktionen, weniger Ausfälle und geringeren Instandhaltungsaufwand erwarten lassen. Dies muss bei ausgeführten Anlagen beobachtet werden. Die hier vorgestellten Ergebnisse begründen sich auf der intensiven Unterstützung der beauftragenden Chemiefirmen Bayer, Covestro, Currenta und Lanxess. Weiter gab es eine enge Zusammenarbeit mit der NAMUR APL Task Force, Bilfinger, Evonik, den Systemintegratoren Actemium, Bilfinger, Etteplan, Griesemann, IOTECH, S.M.A.R.T.S Engineering, TUEG Schillings und YNCORIS, sowie den Herstellern ABB, Beckhoff, Dehn, Emerson, Endress+Hauser, HIMA, Honeywell, Indu-Sol, Krohne, LAPP, Moxa, Pepperl+Fuchs, Phoenix Contact, SAMSON, Siemens, Softing, R.Stahl, VEGA und Yokogawa.

[1] NA 35, *PLT-Planung und -Abwicklung*, NAMUR, 2019.

[2] NA 186, *Interaktionen und Verantwortlichkeiten des PLT-Engineerings*, NAMUR, 2024

[3] NE 168, *Anforderungen an ein Ethernet-Kommunikationssystem für die Feldebene*, NAMUR, 2018

[4] PROFIBUS Nutzerorganisation e.V. (PNO), „Profile for Process Control Devices", 2022

http://dx.doi.org/10.25673/121097

Fahrzeug-zu- Infrastruktur-Kommunikation (V2X) für den autonomen innerbetrieblichen Transport

C-ITS Anwendungen für fahrerlose Routenzüge im intelligenten Werksgelände

Olaf Czogalla[1], René Schönrock[2] und Mansour Abboud[3]

Abstract: Die Unternehmenslogistik befindet sich im Zuge der digitalen Transformation in einem tiefgreifenden Wandel. Mit dem Einzug autonomer Systeme und der zunehmenden Vernetzung entstehen neue Anforderungen an die Kommunikationsinfrastruktur, insbesondere im Kontext von Industrie 4.0. Gegenwärtig werden am Institut der Autoren Forschungsarbeiten zur Ertüchtigung eines Betriebsgeländes für den automatisierten innerbetrieblichen Güter- und Personentransport durchgeführt. Autonome Routenzüge erfordern eine ständige Kommunikationsverbindung mit einer Betriebsleitzentrale, über welche die Transportaufträge erteilt werden, der Transportablauf überwacht und ein gegebenenfalls erforderlicher Eingriff vorgenommen werden kann. Zur Realisierung dieser Kommunikationsverbindung wird die im allgemeinen Verkehrswesen etablierte *Vehicle-to-Infrastructure*-Kommunikation (V2X) nach dem ITS-G5 Standard in dem genannten Fall eingesetzt. Dieser Nachrichtenaustausch gestattet den Aufbau von kooperativen Diensten unter Nutzung dieser intelligenten Transportsysteme. Beispiele für diese Dienste sind *Probe Vehicle Data (PVD), Collision warning* oder *Obstacle on the Road*, welche über die Verbindung zur Betriebsleitzentrale wichtige Zustandsinformationen des autonomen Fahrzeugs bereitstellen. Im vorliegenden Beitrag werden weitere physische Übertragungswege für V2X-Nachrichten aufgezeigt, die im Rahmen des genannten Anwendungsfalls untersucht wurden. So können die standardisierten Nachrichten über ein bestehendes 4G-Mobilfunk oder das installierte betriebliche WLAN ausgetauscht werden und der Betriebsleitzentrale über allgemeine Protokolle wie *Advanced Message Queuing Protocol (AMQP)* bereitgestellt werden, um auf diese Weise Zentralen auch untereinander und überregional zu vernetzen.

Keywords: Fahrzeug-zu-Infrastruktur-Kommunikation, Fahrerloser Routenzug, Kooperative Intelligente Transport Systeme

1 Einleitung

Für die Kommunikation von autonomen Fahrzeugen, sowohl untereinander als auch zwischen Fahrzeug und zentraler Steuerungsinfrastruktur beim Transport von Gütern und Materialien, werden in der Intralogistik nach dem gegenwärtigen Stand der Technik sowohl industrielle WLAN-Netze mit WiFi 5/6 als auch private Mobilfunknetze mit 4G/LTE und

[1] Institut für Automation und Kommunikation, Werner-Heisenberg-Str. 1, 39116 Magdeburg, olaf.czogalla@ifak.eu
[2] Institut für Automation und Kommunikation, Werner-Heisenberg-Str. 1, 39116 Magdeburg, rene.schoenrock@ifak.eu
[3] Institut für Automation und Kommunikation, Werner-Heisenberg-Str. 1, 39116 Magdeburg, mansour.abboud@ifak.eu

5G Campus-Netze eingesetzt. So sind WLAN-Netze weit verbreitet in Indoor-Bereichen wie Hallen für *Automated Guided Vehicle*-Flotten (AGV) oder *Autonomous Mobile Robot*-Flotten (AMR). Hierbei kommen IP-basierte Protokolle wie MQTT oder REST/WebSockets zur Anwendung, teils auch PROFINET inkl. PROFIsafe für sicherheitsrelevante Signale.

Für Bedarfe mit hoher Flächenabdeckung in Outdoor-Bereichen und bei vorhersehbaren Latenzen kommen hingegen zunehmend 5G-Netze zum Einsatz, weil hiermit die Latenzanforderungen für AGVs im einstelligen Millisekundenbereich adressiert werden können. Als standardisierte Schnittstelle zur Kopplung von Leitsystemen mit heterogenen AGR/AMR-Flotten besteht im DACH-Raum die VDA 5050 (VDA/VDMA) als De-Facto-Standard [Vv25]. Hierüber werden Auftrags- und Statusdaten mittels MQTT-Topics (*Message Queueing Telemetry Transport*) ausgetauscht. Weiterhin wird MQTT als Mittler (*Middleware*) für geringere Datenraten zwischen Leitstand und Fahrzeugen mittels *subscribe/publish*-Flottenkommunikation genutzt. Auf der Basis von ROS 2 (*Robot Operating System*) [Ma22] kommt ebenfalls der *Data Distribution Service DDS* zum Einsatz, häufig innerhalb der Roboter und in Multi-Robot-Szenarien, sowie auch für Leitstands-Integration, wenn eine sehr feine Steuerung gewünscht ist.

Über einen solchen spezifischen Kommunikations-Stack würde demzufolge das AGV mittels WLAN oder 5G über IP mittels MQTT-Broker mit dem Leitstand über die VDV 5050-Schnittstelle [Vv25] verknüpft sein. Mit 5G würde eine bessere Abdeckung bei besserem Roaming und in einem Campusnetz erwartbar geringe Latenzzeiten für bewegte und dichte ADV/AMR-Flotten erzielbar sein, jedoch bleibt industrielles WLAN die wirtschaftlichste und erprobte Lösung für diesen Anwendungsfall.

Für den Verkehr im öffentlichen Raum, im individuellen Straßenverkehr oder im öffentlichen Personennahverkehr wird für die Kommunikation zwischen Fahrzeugen und der Verkehrssteuerungsinfrastruktur hingegen die V2X-Kommunikation nach der Technologie ITS-G5 (*Dedicated Short Range Communication- DSRC*) basierend auf IEEE 802.11p oder C-V2X (Cellular V2X) auf der Basis von 4G/5G angewendet.

Diese versetzt das Fahrzeug in die Lage, Nachrichten über Position, Geschwindigkeit, Fahrtrichtung und Linie mit stationären *Roadside units* derselben Technologie bzw. Geräteplattform auszutauschen, um auf diese Weise die Ankunft an signalisierten Knotenpunkten zu melden und eine Priorisierung der Fahrtfreigabe zu erlangen.

Mit Hilfe der V2X-Kommunikation wird die Etablierung von Diensten ermöglicht, welche die Verkehrsteilnehmer auch unterschiedlicher Verkehrsarten zur Kooperation untereinander befähigen. Diese als *Cooperative Intelligent Transport Systems C-ITS* bezeichnete Systeme dienen dem Informationsaustausch von Fahrzeugen, Infrastruktur (z. B. Ampeln), Verkehrszentralen und andere Verkehrsteilnehmer in Echtzeit, um die Verkehrssicherheit und Verkehrseffizienz zu erhöhen. Hierbei werden über Broadcast standardisierte Nachrichten wie zum Beispiel die *Cooperative Awareness Message (CAM)* zur Weitergabe der Bewegungsdaten oder die *Decentralized Environmental Notification Message (DENM)* zur Gefahrenwarnung an alle Teilnehmer im Empfangsbereich ausgesendet.

Die Anwendung und Übertragung von Kommunikationstechnologien wie V2X aus dem allgemeinen Verkehrswesen in den spezialisierten Bereich des innerbetrieblichen Transports hat einige nennenswerte Vorteile. So kann eine bestehende Kommunikationsinfrastruktur, die innerhalb des abzudeckenden Bereiches zur Verfügung steht für den Datenverkehr der Intralogistik genutzt werden, ohne dass zusätzliche ortsfeste Empfangsgeräte (z.B.: *Roadside units*) benötigt werden. Im öffentlichen Verkehrsräumen werden diese RSUs in unmittelbarer Nähe zum Verkehrsknoten installiert, vielfach um C-ITS Dienste wie ÖV-Bevorrechtigung bereitstellen zu können. Im innerbetrieblichen Transport kommt es jedoch eher auf eine kontinuierliche Datenübertragung von Zustandsdaten des Fahrzeugs an den Leitstand oder die AGV-Leitstelle an. Ein weiterer Vorteil der Nutzung von V2X in der Intralogistik besteht in der Nachnutzung von typischen sicherheitsrelevanten C-ITS Diensten wie Kollisionswarnung (DENM), Schutz schwacher Verkehrsteilnehmer (*Vulnerable road user - VRU*) und Übermittlung von Fahrzeugdaten (*Probe data*). Im Folgenden wird mit dem Dienst *Probe data* ein ausgewähltes Beispiel mit Anwendungspotential in der Intralogistik aufgezeigt.

2 Erfassung von Übertragungsparametern

Auf einem realen Werksgelände wurden Feldmessungen durchgeführt, um die zur IP-basierten Übertragung von V2X-Nachrichten erforderliche Netzabdeckung zu ermitteln. Das für autonome Fahrzeuge zu nutzende Werksgelände umfasst hierbei Abmessungen in einem Rechteck von annähernd 250 m Breite und 750 m Länge in welchem verschiedene Straßenabschnitte für den Werksverkehr zwischen den Werkshallen und weiteren Gebäuden angeordnet sind, siehe dazu Abb. 1.

Abb. 1: Werksgelände mit innerbetrieblichen Straßenabschnitten im Luftbild,
Bild: Google maps.

Für die drahtlose Kommunikation über IEEE 802.11n steht im überwiegenden Teil des Werksgeländes ein WLAN-Netz zu Verfügung. Mit Hilfe von mehrtägigen Messfahrten wurden Übertragungsparameter, wie die im WLAN auf dem Gelände örtlich erzielbare

Bandbreite erfasst. Zu diesem Zweck wurde ein Fahrzeug mit einer *Onboard unit* und geeigneten Antennenaufbauten ausgestattet, welche in einer ähnlichen Weise im autonomen Routenzug zum Einsatz kommen soll, siehe Abb. 2.

Abb. 2: Antennenaufbau auf dem Versuchsfahrzeug
Bild: Eigene Darstellung

Entsprechend den vorgesehenen Anwendungsfällen für Transportrouten von Teilen, Vorrichtungen sowie von Personen wurden alle innerbetrieblichen Straßenabschnitte der geplanten Routen in die Erfassung der Empfangsfeldstärke (*Received Signal Strength Indicator – RSSI*) und der tatsächlich erreichten Datenrate in Mbps einbezogen und so gespeichert, dass eine nachträgliche Ortsreferenzierung über den Zeitstempel der sekündlich durchgeführten Einzelmessungen möglich ist. In Abb. 3 ist die erreichte WLAN-Bandbreite für alle durchfahrenen Abschnitte dargestellt.

Abb. 3: Werksgelände mit WLAN-Bandbreite auf Straßenabschnitten,
Bild: Eigene Darstellung auf Leaflet-Kartenbasis OpenStreetMap & Contributors.

In räumlicher Nähe von WLAN-Accesspoints (AP) werden erwartungsgemäß Datenübertragungsraten von 100 Mbps und mehr erreicht, während in Abschnitten mit größerer Entfernung zum AP die erreichbaren Raten in den einstelligen Bereich absinken. Die durchgeführten Messungen hatten insbesondere das Ziel, die Bereiche des Betriebsgeländes zu identifizieren, für die eine Ergänzung des bestehenden WLAN-Netzes geboten ist, um eine lückenlose Konnektivität mit hinreichenden Übertragungsparametern für den vorgesehenen Einsatzzweck gewährleisten zu können.

3 V2X-Kommunikation zur Realisierung des Dienstes *Probe Data*

3.1 Übertragungsweg nach ETSI ITS-G5

Der standardgemäße Übertragungsweg der V2X-Kommunikation über ITS-G5/ 802.11p sieht vor, dass eine Nachricht (z.B. *Cooperate awareness message - CAM*) von der *Onboard unit (OBU)* als V2X-Sendestation unter Verwendung von bestimmten Protokollen, Kodierungen und Schichten zur V2X-Empfangsstation gelangt, welche in der Regel eine *Roadside unit (RSU)* ist, bzw. im Fall der Fahrzeug-zu-Fahrzeug-Kommunikation, wiederum eine OBU ist. Nach dem ETSI ITS-G6 Stack wird hierfür der folgende, vereinfacht skizzierte Weg durchlaufen:

1. Die OBU erstellt die CAM- oder DENM-Nachricht nach dem ETSI-Standard (EN 302 637-2 für CAM, EN 302 637-3 für DENM) mit den jeweils aktuellen variablen Werten: Fahrzeugposition, Geschwindigkeit, Richtung oder detektierte Ereignisse wie Unfall, Glatteis, Stauende als ASN.1-Datenpaket und kodiert es als UPER (*Unaligned Packet Encoding Rules*)

2. Die Nachricht wird in ein *GeoNetworking*-Paket nach ETSI EN 302 636-4-1 eingebettet. Hierbei wird eine geobasierte Adressierung angewendet, das heißt die Nachricht wird an andere V2X-Stationen im geografischen Umkreis per Broadcast mit dem Transportprotokoll BTP (*Basic Transport Protocol*) verbindungslos und portbasiert gesendet.

3. Das *Access Layer* beinhaltet das *Data Link Layer (DLL)* und das *Physical Layer*. Das DLL verwendet IEEE 802.2 LLC (Logical Link Control) mit EtherType 0x8947 (für *GeoNetworking*), DCC (*Decentralized Congestion Control*) kontrolliert Sendeintervalle und Sendeleistung, um Überlastung zu vermeiden. Die MAC-Schicht basiert auf IEEE 802.11p (ETSI ITS-G5) oder bei Cellular-V2X auf der PC5-Schnittstelle nach 3GPP Rel. 14/16 [3G17],[3G20].

4. Das *Physical Layer* arbeitet im Frequenzbereich: 5.875–5.905 GHz (ITS-G5 in Europa) mit der Modulation OFDM (bei ITS-G5) und SC-FDMA (bei C-V2X). Die Datenrate beträgt hierbei je nach Modulation and Coding Scheme 3, 6, oder 12 Mbps.

5. Die Nachricht wird über die Luftschnittstelle an alle Empfängerstationen im erreichbaren Umkreis gesendet.

6. Die Empfängerstation empfängt die Nachricht auf der V2X-Frequenz und entschlüsselt sie gemäß derselben Protokollkette in umgekehrter Richtung.

7. Eine zusätzliche Sicherheitsschicht (ITS Security) ist optional. Hierin werden V2X-Nachrichten mit digitalen Zertifikaten signiert, um die Authentizität und Integrität des Nachrichteninhalts zu gewährleisten. Weiterhin wird mittels pseudonymer Zertifikate die *Privacy*-Anforderung im öffentlichen Straßenraum erfüllt, den Aufenthaltsort einer V2X-Station nicht rückverfolgen zu können.

In Abb. 4 ist der Aufbau einer Beispielnachricht vereinfacht schematisch dargestellt.

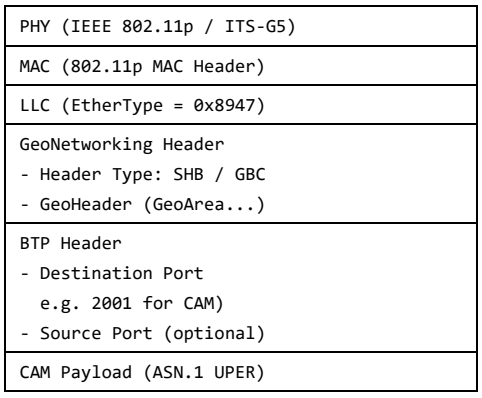


Abb. 4: CAM Nachricht – Paketaufbau ETSI Stack

Der BTP Header besteht aus 4 Bytes, die dieser Reihenfolge zuerst den Zielport (2 Bytes) und danach folgend den Quellport (2 Bytes) enthalten. Das BTP verwendet für die CAM üblicherweise den Zielport 2001 (0x07D1). Der Geonetworking-Header ist variabler Länge und enthält neben der Single Hop Broadcast für CAM den Source Position Vector mit der GPS-Position des Senders, der GeoArea (leer bei Broadcast) und der Lifetime mit typischerweise dem Wert „1“ als Hop Limit.

Die CAM-Payload beinhaltet die UPER-kodierte Nachricht mit der Fahrzeugposition, Heading, Geschwindigkeit, dem Fahrzeugtyp und optional weitere Fahrzeugdaten. Die typische Größe einer CAM-Nachricht beträgt 60-120 Bytes. Die Grundstruktur der CAM-Nachricht ist als ASN.1 Code im folgenden Listing der Abb. 5 angegeben. Die darin verwendeten Container sind im Standard ETSI TS 102894 beschrieben und aus Platzgründen hier nicht dargestellt.

```
CAM ::= SEQUENCE {
         header ItsPduHeader,
         cam CoopAwareness
}
CoopAwareness ::= SEQUENCE {
         generationDeltaTime GenerationDeltaTime,
         camParameters CamParameters
}
CamParameters ::= SEQUENCE {
         basicContainer BasicContainer,
         highFrequencyContainer HighFrequencyContainer,
         lowFrequencyContainer LowFrequencyContainer OPTIONAL,
         specialVehicleContainer SpecialVehicleContainer OPTIONAL,
         ...
}
```

Abb. 5: CAM Nachricht – ASN.1 Definition des Root Data Frames

3.2 Alternativer Übertragungsweg über IP/UDP

In IP-basierten Architekturen, wie sie im vorgesehenen Einsatzfall mit den innerbetrieblichen WLAN-Netz bestehen, können V2X-Nachrichten alternativ übertragen werden, in dem anstelle des BTP – *Basic Transport Protocol* das UDP des IP-Stacks eingesetzt wird, und entsprechende Anpassungen in der Generierung der Nachricht und des Protokollstacks vorgenommen werden.

Auch in der Anwendung von Cellular-V2X, in Verbindung mit 4G/5G und IP-Kommunikation werden CAM und DENM-Nachrichten über IP/UDP versandt. Anstelle des BTP/GeoNetworking wird auch hierzu UDP verwendet. Die ASN.1-kodierte CAM bleibt gleich, nur Transport- und Netzwerk-Schicht ändern sich, siehe Abb. 6.

Die Nutzung des IP/UDP-Übertragungsweges bietet den Vorteil, eine bestehende WLAN-Infrastruktur verwenden zu können, ohne zusätzliche RSU-Technik für den Empfang über ITS-G5 installieren zu müssen. Diese Voraussetzung ist in vielen innerbetrieblichen Werksbereichen bereits gegeben und stellt somit eine günstige Basis dar, um autonomen Verkehr in abgegrenzten Arealen zu etablieren, die – im Gegensatz zum öffentlichen Straßenraum – nicht die Anwesenheit eines Sicherheitsfahrers im Fahrzeug erfordern.

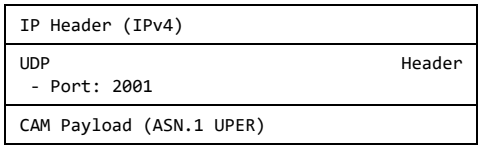


Abb. 6: CAM Nachricht – Paketaufbau für IP/UDP

In der folgenden Tabelle sind die Unterschiede zwischen beiden Übertragungswegen zusammenfassend dargestellt.

Merkmal	**ETSI ITS-G5 (GeoNet/BTP)**	**IP-basiertes V2X (5G, UDP)**
Netzwerk Layer	GeoNetworking (non-IP)	IPv6
Transport Layer	BTP (4 Byte Header)	UDP (8 Byte Header)
Typische Nutzung	Direkt V2X (Fahrzeug–Fahrzeug)	V2N, V2C (Fahrzeug–Cloud, Edge)
Adressierung	Geografiebasiert (GeoArea)	IP-Adressen
Infrastruktur	RSU, OBU	5G-GNB, Edge/MEC

Tab. 1: Vergleich ITS-G5 vs. IP/UDP-V2X

3.3 Generierung von CAM mittels Board Support Package auf Onboard unit

Die Nutzungsprofile von Nachrichtentypen in der V2X-Kommunikation nach ITS-G5 sehen vor, dass CAM-Nachrichten ausschließlich auf ITS-Stationen des Typs *Onboard unit* erzeugt und im Rahmen von dedizierten ITS-Diensten ausgesendet werden. Im Rahmen von langjährigen Forschungsarbeiten wurde am Institut der Autoren eine Hardwareplattform für die V2X-Kommunikation bestehend aus *Onboard unit* und *Roadside unit* entwickelt. Diese Plattform beruht auf einer geeigneten Microcontroller-Architektur mit dem Apalis iMX 6x8, einem ublox Vera-P174 V2X-Radiomodul sowie einer Vielzahl von

drahtgebundenen (Ethernet, USB, GPIO) und drahtlosen Schnittstellen (ITS-G5, 4G/LTE, WiFi, BLE, GNSS), siehe hierzu Abb. 7. Für die OBU wurden die für den Einsatz in Schienenfahrzeugen und Busfahrzeugen des öffentlichen Personenverkehrs erforderlichen Zertifizierungen für Brandschutz und elektromagnetische Verträglichkeit durch benannte Prüfinstitute erbracht. Die OBU wurde bereits in einer Vielzahl von Forschungsprojekten (siehe [Sc20], [Ha21], [Cz22], [Cz24]) zur V2X-Kommunikation erfolgreich weiterentwickelt und im ÖV-Regelbetrieb eingesetzt.

Abb. 7: ifak C-ITS Onboard unit (links) und Roadside unit (rechts)

Als Betriebssystem für diese Geräteplattform wird ein *Embedded Linux* eingesetzt, das auf einem *OpenEmbedded-BSP* (*Board Support Package*) basiert und mit einem angepassten Kernel der Version 6.112 operiert. Als Build-Umgebung für das OBU-BSP kommt *Yocto* zum Einsatz. Unter Nutzung des BSP kann eine Tool-Chain aufgebaut werden mit welcher die automatische Generierung von V2X-Nachrichten und deren Versand über IP/UDP ermöglicht wird.

Ein wichtiges Element der Tool-Chain ist ein ASN.1-Compiler welcher mit Hilfe essentieller Build-Tools wie der GNU Compiler Collection `gcc`, bzw. Autotools (`configure, make, make install`) aus einem frei verfügbaren Git-Repository aufgebaut wird. Der ASN.1 Compiler wird benötigt, um aus der im ETSI-Standard TS 102 900 v2.2.1 [Et25] definierten CAM-Nachricht eine C++ Quellcodedatei zu erzeugen. Die Beschreibung der CAM erfolgt im ASN.1-Format als CAM-PDU (*Packet Data Unit*). Darin verwendete Definitionen und Beschreibung der verwendeten Datenelemente und „*named values*“ sind als CDD (*Common Data Dictionary*) im Standard ETSI TS 102 894-2 V2.3.1 (2024-08) spezifiziert [Et24], die auch in weiteren V2X-Nachrichtendefinitionen verwendet werden. Beispiele hierfür sind Definitionen für die wichtigsten ITS-Datentypen wie: `StationType`, `ReferencePosition`, `Altitude`, `BasicContainer`, `Heading`, `Speed`, `VehicleWidth`, oder „*named values*“ als `CauseCodes` für Unfallereignisse. Dieser Datencontainer wurde in Vorläuferversionen des Standards auch als ITS-Container bezeichnet. Diese Datei trägt aktuell im ETSI-Repository die Bezeichnung `ETSI-ITS-CDD.asn`.

```
 * @category: Communication information.
 * @revision: Created in V2.1.1 based on StationType
 */
TrafficParticipantType ::= INTEGER {
    unknown         (0),
    pedestrian      (1),
    cyclist         (2),
    moped           (3),
    motorcycle      (4),
    passengerCar    (5),
    bus             (6),
    lightTruck      (7),
    heavyTruck      (8),
    trailer         (9),
    specialVehicle  (10),
    tram            (11),
    lightVruVehicle (12),
    animal          (13),
    agricultural    (14),
    roadSideUnit    (15)
} (0..255)
```

Abb. 8: CDD (*Common Data Dictionary*) mit *named values* als ASN.1 Definition des Typs `TrafficParticipantType` aus `StationType`

Mit der auf der *Onboard unit* somit bereitstehenden Tool-Chain kann unter Einbeziehung der nunmehr in C++ vorliegenden CAM-Definitionen und Beschreibung der ITS-Datenelemente eine Anwendung erstellt werden, die dem im Folgenden dargestellten Algorithmus beinhaltet und eine kontinuierliche Aussendung einer CAM-Message mit den aktuellen Werten der Sensoren ausführt.

Algorithm 1: Einfacher Generator für C-ITS CAM-Message

input : Vehicle position (Latitude, Longitude), Vehicle width, Vehicle length
output: ETSI conform C-ITS CAM-Message → over UDP/IP

```
Validate destination IP
if IP is valid then
    Create UDP socket
    Connect to GPS
    Enable streaming
    while 1 do
        Wait for GPS Fix
        Extract GPS data
        Initialize CAM structure
        Set CAM header
        Set generationDeltaTime
        Set BasicContainer
        Set HighFrequencyContainer
        Set mandatory fields
        Validate ranges
        Check constraints
        Encode to UPER
        Send over UDP
```

Abb. 9: Einfacher Algorithmus als Generator für C-ITS CAM-Message

Einen Ausschnitt aus der Implementation des einfachen CAM-Generators in C++ zeigt Abb. 10. Neben dem `BasicContainer` ist für die CAM-Nachricht auch die Belegung der Werte des `HighFrequencyContainer` als obligatorische Felder bindend, wenn die Werte für Geschwindigkeit, Bewegungsrichtung, Fahrzeuglänge, usw. in der CAM-Nachricht enthalten sein sollen.

```
// Set HighFrequencyContainer
cam->cam.camParameters.highFrequencyContainer.present =
HighFrequencyContainer_PR_basicVehicleContainerHighFrequency;
BasicVehicleContainerHighFrequency_t *hf =
    &cam->cam.camParameters.highFrequencyContainer.choice.
    basicVehicleContainerHighFrequency;

// Set mandatory fields
hf->heading.headingValue = its_heading;
hf->heading.headingConfidence = HeadingConfidence_unavailable;
hf->speed.speedValue = its_speed;
hf->speed.speedConfidence = SpeedConfidence_unavailable;
hf->driveDirection = DriveDirection_forward;
hf->vehicleWidth = 18; // 1.8 meters or 18 decimeters
hf->vehicleLength.vehicleLengthValue = 42; // 4.2 meters
hf->vehicleLength.vehicleLengthConfidenceIndication =
VehicleLengthConfidenceIndication_noTrailerPresent;
hf->longitudinalAcceleration.value = 0; // 0 m/s²
hf->longitudinalAcceleration.confidence = AccelerationConfidence_unavailable;
hf->yawRate.yawRateValue = 0; // 0 deg/s
hf->yawRate.yawRateConfidence = YawRateConfidence_unavailable;
```

Abb. 10: Codefragment aus C++-Implementierung des C-ITS CAM-Generators

3.4 Test der CAM-Übertragung im lokalen IP-Netz

Für den Labortest des Übertragungsweges im lokalen IP-Netz werden jeweils ein CAM-Generator (*Encoder*) als Clientapplikation und ein CAM-Decoder als Serverapplikation auf den entsprechenden Zielplattformen implementiert. Diese besteht aus der *Onboard unit* als Clientrechner aus der beschriebenen Tool-Chain, die in analoger Weise für die Zielplattform des Servers am Ort der der Betriebsleitzentrale aufgebaut wird. Beide Instanzen sind über ein lokales Netzwerk (WLAN oder LAN) verbunden. Hierzu wird dem CAM-Encoder-Client die IP-Adresse des CAM-Decoder-Servers für die Kommunikation über das UDP-Socket als Parameter übergeben.

In Abb. 11 ist eine Testimplementierung für CAM-Encoder und CAM-Decoder mit Debug-Ausgaben der übermittelten CAM-Nachrichteninhalte dargestellt. In der linken Bildhälfte ist das Ausgabefenster des CAM-Encoders auf der *Onboard unit* mit den Werten der GPS-Position (*Longitude*, *Latitude*) und den daraus abgeleiteten Werten für Geschwindigkeit und Bewegungsrichtung zu sehen. Die UPER-codierte CAM-Nachricht hat eine Länge von 40 Bytes. Im rechten Fenster sind die decodierten und identischen Werte derselben CAM-Nachricht nach Übertragung und Empfang durch den Serverrechner auf Seiten der späteren Betriebsleitzentrale zu erkennen.

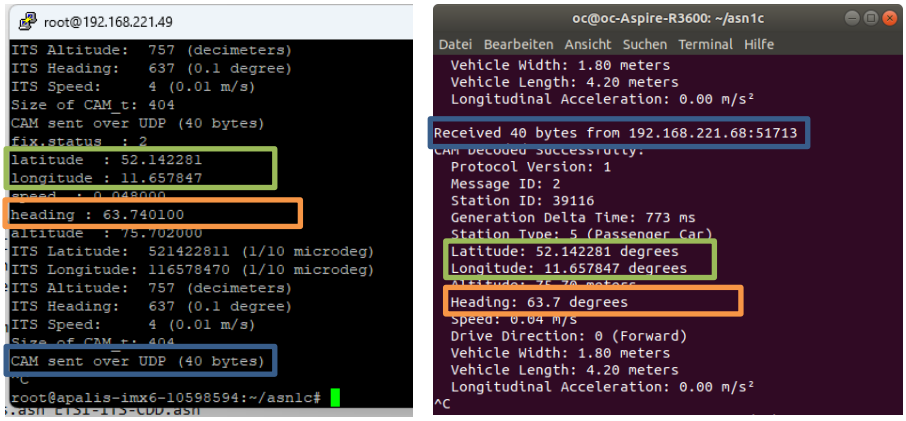


Abb. 11: Testimplementierung für CAM-Encoder (links) auf *Onboard unit* und CAM-Decoder (rechts) auf Betriebsleitserver mit Debug-Ausgaben übertragener CAM-Nachrichteninhalte

4 Ausblick zur V2X-Kommunikation mit Nutzung des *Advanced Message Queuing Protocol (AMQP)*

Um Betriebsleitzentralen auch überregional zu vernetzen, kann das offene Standardprotokoll *Advanced Message Queuing Protocol (AMQP)* für den Nachrichtenaustausch in verteilten Systemen genutzt werden. AMQP definiert, wie Nachrichten zwischen Anwendungen transportiert, formatiert und verarbeitet werden – unabhängig von der Programmiersprache oder Plattform. Hierbei wird eine Nachrichtenzustellung, aber auch Transaktionen einschließlich Verschlüsselung und Zugriffskontrolle garantiert.

Die Komponenten von AMQP sind *Producer* als Sender von Nachrichten, *Exchange* als Empfänger von Nachrichten, eine *Queue* für die Speicherung von Nachrichten sowie der *Consumer*, welcher Nachrichten aus der *Queue* liest.

AMQP bietet verschiedene Routing-Mechanismen über *Exchanges*:

- *Direct Exchange*: Routing anhand eines *Routing-Keys*
- *Fanout Exchange*: Nachricht an alle *Queues*
- *Topic Exchange*: Routing über Platzhalter und Muster
- *Headers Exchange*: Routing anhand von Header-Feldern

Anwendungen, die AMQP unterstützen, können über verschiedene Plattformen hinweg interoperabel miteinander kommunizieren. Hierfür existieren typische Implementierungen

unter Verwendung von weit verbreiteten plattformunabhängigen Message-Brokern wie *RabbitMQ*.

Obwohl AMQP kein direkter Bestandteil der V2X-Kommunikation in kooperativen intelligenten Verkehrssystemen (C-ITS) ist, kann es eine Rolle in Backend-Systemen oder Verkehrsmanagementzentralen spielen, um Echtzeitdaten zwischen Servern auszutauschen, Sensor- oder Fahrzeuginformationen weiterzuleiten bzw. Ereignisse, wie Staus oder Unfälle zu verteilen. Im vorliegenden Einsatzfall von V2X-Kommunikation für den autonomen innerbetrieblichen Transport kann das AMQP-Protokoll dazu dienen, Daten zwischen Subsystemen wie Betriebsleitzentralen, technischen Aufsichtssystemen und weiteren Leitstellen auszutauschen.

Für einen solchen Anwendungsfall soll eine Beispielarchitektur im Folgenden kurz skizziert werden (siehe Abb. 12). Es wird davon ausgegangen, dass eine V2X-Verkehrszentrale Ereignisdaten aus verschiedenen Quellen sammelt, z.B.:

- Fahrzeuge (AGV, AMR, Routenzug) über OBUs
- Verkehrssensoren (Detektoren zur Belegung und Bemessung von Verkehrsströmen)
- Intelligente Leitbaken als Navigationshilfen für autonomen Routenzug
- Verkehrssteuerungsinfrastruktur über RSUs (Lichtsignale)
- Straßenzustand (trocken, feucht, vereist) und Wetterdaten (Sichtbedingungen)

Diese Daten sollen in verschiedene Backend-Systeme verteilt werden, wie zum Beispiel die folgenden:

- Verkehrszustandsinformation für Leitstellenpersonal
- Datenanalyse und Prognosen
- Dashboards für Operatoren
- Eventgesteuerte Warnmeldungen

Ein Dienst mit der Bezeichnung *Message Ingestor* empfängt von verschiedenen Instanzen über OBUs und RSUs zum Beispiel Warnmeldungen und Statusmeldungen über eingehende Ereignisse (z. B. DENM, CAM, Stauwarnungen, Baustellenstatus) und publiziert sie an einen AMQP-Exchange-Dienst. Für diesen Übertragungsweg wird die beschriebene V2X Kommunikation über IP/UDP genutzt.

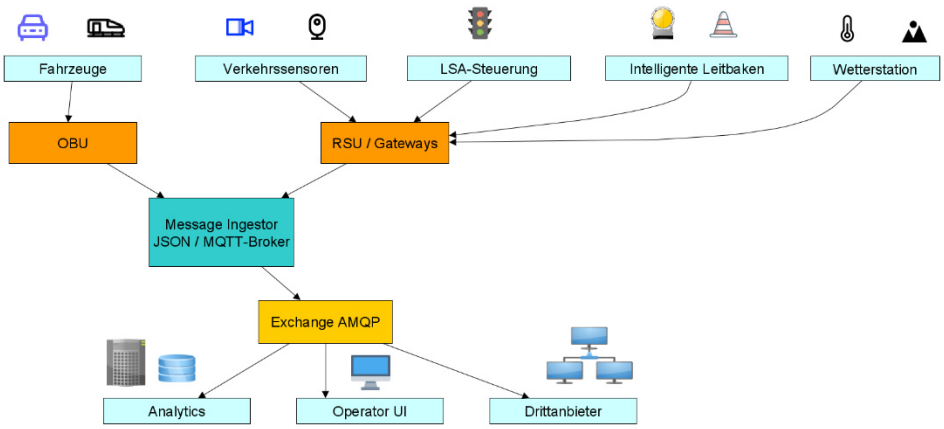


Abb. 12: Beispielarchitektur für AMQP-Integration in ein V2X-Kommunikationsnetzwerk

Der *Message Ingestor* übernimmt hierbei die Aufnahme von Nachrichten jeglicher Informationsquellen, unabhängig von deren späterer Nutzung durch verschiedene datenkonsumierende Dienste. Letztere können für ihre spezialisierten Zwecke nur bestimmte Nachrichtentopics abonnieren und somit gezielte Auswertungen und Analysen vornehmen, im Fall von *Analytics* z.B: nur auf Ereignisse mit der Topic-Gruppe `event.traffic.*`. So erhielte beispielhaft ein Web-Dashboard sämtliche Ereignisse, wie `weather.alerts` und `road.works.start`, während ein weiterer AMQP-Consumer bestimmte Ereignisse an Dispatcher, Verkehrs-APIs, z. B. nur `event.road.works.*` weiterleiteten würde.

Der Vorteil eines solchen Verarbeitungsansatzes besteht darin, dass eine Entkopplung vom Produzenten einer Nachricht mit dessen Konsumenten stattfindet und beide Instanzen nicht direkt miteinander verbunden sein müssen. Dieser Ansatz ist beliebig skalierbar. Neue Empfänger können beliebig hinzugefügt werden. Die Nachrichten können gepuffert werden und nach Erreichbarkeit später zugestellt werden, was die Fehlertoleranz erhöht. Schließlich wird die Flexibilität des topic-basierten Routings durch diese gezielte Auswahl des Nachrichtentyps gesteigert.

Danksagung

Die in diesem Beitrag behandelten Forschungsarbeiten sind zum Zeitpunkt der Erstellung im Fortgang und beruhen auf der Teilnahme des ifak am Verbundprojekt „IBATOUR“ welches von den Autoren in Projektpartnerschaft bearbeitet wird. Das Projekt wird vom Bundesministerium für Verkehr und digitale Infrastruktur (BMVI), ab 6. Mai 2025 vom Bundesministerium für Verkehr unter dem Förderkennzeichen 19FS2070G gefördert.

Literaturverzeichnis

[3G17] 3GPP, 2017. *TS 23.682, V14.5.0, Release 14: Service Capability Exposure Function (SCEF)*. ETSI / 3GPP.

[3G20] 3GPP, 2020. *TS 29.500, V16.4.0, Release 16: Service Based Interfaces – 5G Core (SBA)*. ETSI / 3GPP.

[Cz22] Czogalla, O.; Naumann, S.: *C-ITS Anwendungen für ÖPNV und Straßenverkehr im Projekt LOGIN Hannover*, 17. VIMOS-Symposium, TU Dresden, 30.11-01.12.2022

[Cz24] Czogalla, O.: *Vehicle-to-Anything*. In: Mitteldeutsche Mitteilungen, Informationen aus Wirtschaft | Wissenschaft | Gesellschaft, 33. Jahrgang, 3/2024

[Et24] ETSI TS 102 894-2 V2.3.1 (2024-08): *Intelligent Transport Systems (ITS)*; Users and applications requirements; Part 2: Applications and facilities layer common data dictionary; Release 2

[Et25] ETSI TS 103 900 V2.2.1): *Intelligent Transport Systems (ITS)*; Facilities Layer; Cooperative Awareness Service; Release 2, 2025-02

[Fa11] Fakler, Oliver; Bogenberger, Klaus; Mincheva, Eleonora: *Forschungsprojekt Aktiv-VM: Adaptive und kooperative Technologien für den intelligenten Verkehr - Verkehrsmanagement*: Schlussbericht, Transver Verlag, München, 2011.

[Ha21] Hartmann, K., Kayser, M., Harders, T., Koch, O., Finger, T., Maier, P., Schön, T., Gorisch, S., Vomend, P., Zimmermann, A., Czogalla, O., Schönrock, R., Schade, J., Mustafa, L., & Wartmann, R. (2021). *BiDiMoVe – bidirektionale multimodale Vernetzung*: Abschlussbericht (Projektlaufzeit: 10/2018–09/2021). Freie und Hansestadt Hamburg. https://www.tib.eu/de/suchen/id/TIBKAT%3A181099442X

[Ma22] Macenski, S., Foote, T., Gerkey, B., Lalancette, C. & Woodall, W., 2022. Robot Operating System 2: Design, architecture, and uses in the wild. *Science Robotics*, 7(66), eabm6074.

[Ot24] Otto, T; Bieland, D.; Czogalla, O. et al.: *FGSV 321/4 Hinweise für den Technologiewechsel C-ITS an Lichtsignalanlagen H TwLSA*. Forschungsgesellschaft für Straßen- und Verkehrswesen FGSV, Arbeitsgruppe Verkehrsmanagement, FGSV Verlag Köln, Ausgabe 2024.

[Ru16] T. Ruß, R. Schönrock, J. Krause: *Automated testing of V2X-based applications for cooperative traffic systems*, 23rd World Congress on Intelligent Transport Systems, Melbourne, 10.-14.10.2016 - 2016

[Ru16a] T. Ruß, R. Schönrock, J. Krause: *V2X-based cooperative protection system for vulnerable road users, impact on traffic*, 23rd World Congress on Intelligent Transport Systems, Melbourne, 10.-14.10.2016.

[Sc20] Schäfer, M., Hoyer, R., Hepner, E., Wang, M., Zhao, C., Schneegans, L., Schmitt, V., Hartung, K., Miltner, T., Kugler, M., Weisheit, T., & Mahler, M. (2020). *VERONIKA – Vernetztes Fahren des öffentlichen Nahverkehrs in Kassel*: Abschlussbericht, Teil 1 (öffentlich): gemeinsamer Arbeitsbericht der Projektpartner. Universität Kassel. https://www.tib.eu/de/suchen/id/TIBKAT%3A1741226813 https://doi.org/10.2314/KXP:1741226813

[Vv25] VDA/VDMA (2025): *VDA 5050 – Schnittstelle zur Kommunikation zwischen Leitsteuerung und fahrerlosen Transportsystemen (FTS) / autonomen mobilen Robotern (AMR)*, Version 2.1.0, Frankfurt am Main.

[Wr24] Wrubel, Th.; Czogalla, O.; Naumann, S.; Schönrock R.; Hoyer, R; Hepner, E. et al.: *LOGIN - Lichtsignalanlagen optimal gesteuert im Nahverkehr*- Abschlussbericht, Laufzeit:12/2020 – 05/2024, TIB Hannover

http://dx.doi.org/10.25673/121097

Industrielle 5G-Netze der Deutsche Bahn - Einsatzgebiete * Potenziale * Herausforderungen

Alexander Bentkus[1]

Abstract: Dieser Vortrag gibt einen Einblick in die Praxis, wie eine neue Technologie schrittweise einsatzfähig gemacht wurde und welche Herausforderungen weiterhin bestehen, um das volle Potenzial nutzen zu können. Die industriellen 5G Netze werden in der Instandhaltungseinrichtungen bei der Deutschen Bahn eingesetzt und nach und nach ausgerollt. Das neue DB-Werk in Cottbus ist für die schwere Instandhaltung der ICE4 Flotte zuständig und wird das erste produktive 5G-Netz ab Januar 2026 erhalten.

[1] Deutsche Bahn
Alexander.Bentkus@deutschebahn.com

Impressum

16. Jahreskolloquium

« KOMMUNIKATION IN DER AUTOMATION »
(KOMMA 2025)

11./12.11.2025 • Magdeburg

ISBN 978-3-948749-60-6

DOI: http://dx.doi.org/10.25673/121097

HERAUSGEBER

Lisa Underberg, Magdeburg
Jürgen Jasperneite, Lemgo

Institut für Automation und
Kommunikation e.V.
An-Institut der Otto-von-Guericke-Universität
Magdeburg
Werner-Heisenberg-Straße 1
39106 Magdeburg

Telefon: +49 (0)391 990140
Fax: +49 (0)391 9901590
Internet: www.ifak.eu