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

KOMMA 2023
Kommunikation in der Automation

Tagungsband

21./22.11.2023 | 14. Jahreskolloquium

« Kommunikation in der Automation »

Ulrich Jumar, Jürgen Jasperneite (Hrsg.)

■ Eine Kooperation von: ifak inIT TH OWL

■ Unterstützt von: VDE ITG GI Gesellschaft für Informatik

Impressum

14. Jahreskolloquium

« Kommunikation in der Automation»
(KommA 2023)

21./22.11.2023 • Magdeburg

http://dx.doi.org/10.25673/111633

HERAUSGEBER

Ulrich Jumar, 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

KOMMA 2023
Kommunikation in der Automation

Vorwort

Tagungsband des 14. Jahreskolloquiums Kommunikation in der Automation – KommA 2023

Vor 13 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. Im Beitragsaufruf war damals zu lesen: „Die industrielle Kommunikation hat ihre Wurzeln in Deutschland und ist das Rückgrat jedes dezentralen Automatisierungssystems. Durch den vermehrten Einsatz von Informationstechnologien ergeben sich neue Möglichkeiten in rasanter Geschwindigkeit, gleichzeitig auch stets zu lösende Herausforderungen“. Im Grundsatz wird man diese Aussage auch heute bejahen können, der prägende äußere Einfluss der Informations- und Kommunikationstechnologien auf die Automatisierungsbranche ist allerdings zur Selbstverständlichkeit geworden. Dass damit aber weder praktisch, noch wissenschaftlich und methodisch alle Fragen beantwortet sind, offenbart das Programm des Jahreskolloquiums.

Am Forschungsstandort Magdeburg ist das Thema der industriellen Kommunikation untrennbar mit Prof. Peter Neumann verbunden. Peter Neumann, der zugleich Initiator und Gründungsvater des Instituts ifak in Magdeburg ist, verstarb im Oktober 2023 im Alter von 81 Jahren. Bereits in den 1980er Jahren widmete er sich den Feldbussen. Dies mündete in sein ausgeprägtes langjähriges Engagement in der Profibus Nutzerorganisation (PNO). Im Zeitraum 1990 – 2005 war er langjähriger Leiter des Fachausschusses „Communication Profiles“ und Mitglied des Beirates der PNO, der für die Technologieentwicklung von PPROFIBUS und PROFINET verantwortlich ist. In seinem 2018 herausgegebenen Buch zur geschichtlichen Entwicklung „Magdeburger Automatisierungstechnik im Wandel – Vom Industrie- zum Forschungsstandort“ schreibt Prof. Neumann: „Zusammenfassend kann man feststellen, dass in den siebziger Jahren der Grundstein für die verteilte Datenverarbeitung in den Automatisierungsanlagen gelegt wurde und seither die durchgreifende Anwendung der Informatik in der Automatisierungstechnik gelungen ist.“ Die methodische und angewandte Forschung zu industriellen Kommunikationstechnologien prägen das Wirken von Peter Neumann. Seine Idee des Virtual Automation Networks – VAN – wurde in einem gleichnamigen EU-Projekt mit Leben erfüllt. Die Anfänge des Jahreskolloquiums KommA hat er aktiv mitgestaltet, als hochinteressierter Fachmann hat er die KommA-Veranstaltungen bis zum Schluss verfolgt und die Diskussion mit seinen Beiträgen bereichert. Sein unermüdliches Wirken ist uns Vorbild und Ansporn!

Zum 14. Jahreskolloquium Kommunikation in der Automation, das am 21./22. November 2023 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. Ulrich Jumar

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

KommA 2023 – Kommunikation in der Automation

http://dx.doi.org/10.25673/111634

Erfahrungsbericht bei der Umsetzung der VWS Type 3 Interaktionen in einer Maintenance-Anwendung

R. Islam*, A. Wand**, Ch. Röder**, S. Stamm***, A. Dayeg ****, F. Winter ****, L. Salaj****, H. Noske*****, B. Denkena*****, Ch. Diedrich*

*Otto von Guericke University Magdeburg, Magdeburg, (christian.diedirch@ovgu.de, rafiul.islam@ovgu.de).
** SEITEC GmbH (aw@seitec.info, cr@seitec.info)
***Lauscher Präzisionstechnik GmbH, (siebo.stamm@lauscher.de)
**** FAUSER AG, (aymen.dayeg@fauser.ag, florian.winter@fauser.ag, luan.salaj@fauser.ag)
***** Universität Hannover, IFW (denkena@ifw.uni-hannover.de, noske@ifw.uni-hannover.de)

Abstract:

Intelligente Fertigungssysteme integrieren heutzutage den operativen Produktionsbetrieb mit den dispositiven Aufgaben rund um das MES (Manufacturing Execution System). In diesem Beitrag wird anhand einer instandhaltungstechnischen Aufgabe eine dezentrale Lösung für die Ableitung des Wartungsbedarfes mittels Verwaltungsschalen (VWS) und die Terminplanung durch das MES vorgestellt. Diese beruht auf den Typ 3 VWS, die mittels I4.0 Sprache die notwendigen Nachrichten austauschen. Diese Nachrichten basieren auf Kommunikationsmustern des MES Standards IEC 62264 (ISA 95). Die Architektur wird vorgestellt, mit den verwendeten Interaktionslösungen untersetzt und die daraus resultierenden Erfahrungen erläutert.

Keywords: Digital Twin, Interaktionsmuster, Maintenance, VWS, peer-to-peer

1 Einleitung

Eine effiziente Produktionsplanung und -steuerung ist prinzipiell stark von einer präzisen und vorausschauenden Instandhaltungsplanung abhängig [SEZ18]. Basierend auf einer ausreichenden Datenqualität [DEN19], [DEN20a] sollten Maschinenausfälle korrekt prognostiziert werden, um notwendige Gegenmaßnahmen wie z.B. die Umplanung bzw. Neuplanung von Aufträgen schnellstmöglich einzuleiten. Kleine und mittlere Unternehmen (KMU) können eine solche Datenbasis aufgrund begrenzter Ressourcen und der nicht wirtschaftlichen Nachrüstung von Bestandsmaschinen oftmals nicht vorhalten [KUH17]. Laut Zonta et al. sind folgende Limitierungen für eine praxisnahe Überführung ausschlaggebend [ZON20]:

- Notwendige Maschinendaten für maschinenindividuelle Instandhaltungsprognosen sind in der Praxis kaum vorhanden. Maschinenindividuelle Instandhaltungsprognosen sind damit bisher nicht realisierbar.
- Es existiert zumeist keine einheitliche Anbindung für Maschinendaten.
- Es findet zumeist keine Verknüpfung mit Produktionsplanungs- und Steuerungssystemen statt.

Dieser Beitrag berichtet von dem Verbundprojekt „BaSys4iPPS“, in dem eine Methode zur integrierten Instandhaltungs- und Produktionsplanung durch dezentrale Instandhaltungsprognose für Bestandsmaschinen von KMU entwickelt und mithilfe der BaSys-Referenzarchitektur umgesetzt wird. Die Ergebnisse werden an realen Werkzeugmaschinen der Lauscher Präzisionstechnik GmbH (Lauscher) erprobt, die als repräsentatives KMU der

Zerspanungsindustrie angesehen werden kann. Damit soll eine signifikante Reduzierung unerwarteter Produktionsstillstände und eine deutliche Erhöhung der Planungssicherheit ermöglicht werden.

Ausgangspunkt der hier dargestellten Lösung ist die Verwendung von Verwaltungsschalen [VWS22] des Typ 3 [VDI19], [BEL19] mit der I4.0 Sprache (die in das Basyx-Framework integriert werden), die vor allem die Interaktionen zwischen den Maschinen VWS und der MES VWS umsetzt. Dabei werden Interaktionsmuster angewendet, wie sie in IEC 62264 Teil 5 (auch als ISA 95 bekannt) [IEC16] definiert worden sind. In diesem Beitrag steht der Aspekt der dezentralen Agentenumsetzung zunächst im Hintergrund. Berichtet wird über die anwendungs- und kommunikationsbezogene Architektur und den Erfahrungen bei der Konzeption und Umsetzung.

2 Stand der Technik

2.1 Instandhaltungskonzepte

In der Literatur sind verschiedene Instandhaltungskonzepte beschrieben. Bei der geplanten zeitabhängigen Instandhaltung werden Instandhaltungsaktivitäten unabhängig vom tatsächlich erfassten Zustand der Komponente durchgeführt. Ziel ist es, den Abnutzungsprozess von Maschinenkomponenten durch regelmäßige Wartungen zu verlangsamen. Es wird die Annahme getroffen, dass das Ausfallverhalten bekannt ist und statistisch abgebildet werden kann. Hierfür werden Herstellerinformationen über in Massen hergestellte Maschinenkomponenten herangezogen [PEN10, SCH07]. Auf Basis dieser Informationen werden Wartungen in periodischen Abständen oder auf Basis eines Betriebsstundenzählers durchgeführt. Bei dem Einsatz von Betriebsstundenzählern werden die Betriebsstunden der jeweiligen Maschinenkomponente erfasst und mit einem oder mehreren Grenzwerten verglichen. Dieses Vorgehen wird im Folgendem als Instandhaltungsvariante 1 bezeichnet.

Die geplante zeitabhängige Instandhaltung ist bei Komponenten mit stark streuenden maximalen Laufzeiten nur bedingt für die Wartungsplanung geeignet. Die Streuung der Lebensdauer von Maschinenkomponenten erklärt sich durch verschiedene Faktoren wie Umwelteinflüsse, Zustand von Nachbaraggregaten, aber auch Vorschädigungen bei Fertigung und Montage sowie außergewöhnliche Betriebsfälle. In der Folge führen vorzeitig, bzw. unnötig durchgeführte Wartungsaktivitäten bei der zeitabhängigen Instandhaltung zu erhöhten Kosten. Maschinenzustandsüberwachungen ermöglichen die Erfassung des technischen Zustands von Maschinenkomponenten und damit eine zustandsbasierte Wartung. Die Durchführung einer Zustandsüberwachung setzt voraus, dass zustandssensitive Überwachungssignale zur Zustandsbewertung vorhanden sind [STU90, PEN10]. Die zustandsorientierte Wartung ist gegenüber der zeitorientierten Wartung bei Komponenten mit einer stark streuenden Lebensdauerverteilung im Vorteil [HAS11].

Auf Basis einer Zustandsüberwachung und Restnutzungsdauerprognose (RUL-Prognose) können Instandhaltungsmaßnahmen vorausschauend geplant werden. Zu diesem Zweck werden die folgenden Schritte vorgeschlagen. In einem ersten Schritt erfolgt die Akquise zustandssensitiver Überwachungssignale und sogenannter “run-to-failure"-Daten, die den Zustand der betrachteten Maschinenkomponente im Lebenszyklus beschreiben. Anschließend wird ein sogenannter Zustandsindex abgeleitet, der sich mit dem Zustand der Maschinenkomponente verändert. Mit dem Wissen über den Verlauf des Zustandsindex im Störungszustand einer Maschinenkomponente lässt sich dessen Verlauf in verschiedene Zustandsphasen unterteilen. Eine RUL-Prognose wird durchgeführt, wenn eine Zustandsphase erreicht wird, die einen kritischen Zustand beschreibt. Auf Basis des Zustandsindex und eines gesetzten Schwellwertes erfolgt im letzten Schritt die Prognose des Ausfallzeitpunkts. Methoden zur RUL-Prognose werden den folgenden Kategorien zugeordnet: Physikalische Ansätze, statistische Ansätze, KI-Ansätze und hybride Ansätze [LEI18]. Dieses Vorgehen wird im Folgenden als Instandhaltungsvariante 2 bezeichnet.

In der Praxis werden für das Auslösen von Wartungs- und Instandhaltungsmaßnahmen feste Intervalle der Betriebsstunden der jeweiligen Maschine herangezogen. Für Werkzeugmaschinen werden diese bereits vom Hersteller bereitgestellt. Dabei werden die jeweiligen Komponenten einer Werkzeugmaschine in Wartungspläne gleicher Betriebslaufzeit zusammengefasst. Diese Pläne enthalten jedoch nur einen reduzierten Anteil der Komponenten, eine Gesamtauflistung ist nicht vorhanden. Dies bedeutet, dass ein Teil der Komponenten lediglich Ausfall-basiert instandgehalten werden kann, was zu ungeplanten und plötzlichen Produktionsausfällen und hohen Opportunitätskosten führen kann. Durch die manuelle, wöchentliche Erfassung der Betriebsstunden kann der

Wartungsbedarf ermittelt und eingeplant werden. Dies stellt aus heutiger Sicht für viele Betreiber einen Fortschritt dar. Eine Interaktion zwischen den Maschinen und dem MES als planende Instanz für den Produktionsablauf und die Einordnung der Wartungsarbeiten ist bisher nur eingeschränkt möglich. Den Maschinen wird im MES eine Kapazität von der Fertigungsplanung zugeordnet und auch durch diese überwacht bzw. ggfs. angepasst. Hierfür werden sowohl die Verfügbarkeit von Ressourcen (wie z.B. verfügbare Belegungszustände der Maschinen, Material als auch vom Personal) in Betracht gezogen. Im Rahmen dieser Kapazität nimmt das MES die Einplanung der Produktionsaufträge entsprechend der Liefertermine vor und bildet so eine Auftragsbelegungsliste für jede Maschine. Diese Belegungsliste wird anschließend an der Maschine bereitgestellt, wodurch eine Reihenfolge der zu bearbeitenden Aufträge unter Beachtung aller relevanten Abhängigkeiten zwischen den Aufträgen auf verschiedenen Maschinen vorgegeben wird.

Im Falle einer Wartung- und Instandhaltungsmaßnahme übernehmen mehrere Personen unterschiedliche Aufgaben. Zunächst muss der Wartungsbedarf ermittelt werden, anschließend müssen die zur Wartung notwendigen Betriebsmittel vor Ort sein oder beschafft werden. Gleiches gilt für das notwendige Personal mit den dafür notwendigen Kompetenzen. Darüber hinaus muss zur Festlegung eines Wartungszeitraums, zu dem die Werkzeugmaschine weder verfügbar noch produktiv sein kann, der Produktionsplan mit in Betracht gezogen werden. Vor allem in der Auftragsfertigung mit vielen Abhängigkeiten von Aufträgen untereinander und unterschiedlichen Lieferterminen, bedarf es hoher Anstrengungen einen geeigneten Zeitraum unter der Einhaltung der vorher genannten Anforderungen zu definieren. Dies nimmt in der unternehmerischen Praxis viel Zeit und Kapazitäten in Anspruch und bietet hohes Automatisierungspotenzial.

2.2 Interaktionsmuster zwischen den Maschinen und MES

Das Anwendungsfeld in diesem Projekt betrachtet die Zusammenarbeit zwischen zerspanenden Werkzeugmaschinen und einem MES. Das MES enthält eine Vielzahl von Funktionen, die unterschiedlichen Aufgaben bei der Organisation oder der Steuerung des Produktionsablaufes zuzuordnen sind (Abbildung 1). Diese sind unter anderem im Standard IEC 62264 – auch unter ISA 95 bekannt – zusammengetragen. Dieser umfangreiche Standard definiert eine generelle Architektur, die Funktionsmodularisierung, deren Zusammenwirken, Objektmodelle und auch Interaktionsmuster. Dieser Beitrag bezieht sich dabei auf die Instandhaltungsanforderungen der Maschinen.

Bisherige Maschinenanbindungen an MES-Systeme fokussieren die Übertragung des Betriebszustands der Maschine, um so die Verfügbarkeit abzubilden und eine Steuerung der Fertigung durch Umplanungen von Aufträgen vornehmen zu können. Geplante Wartungs- und Instandhaltungsmaßnahmen werden bisher nicht übertragen, da hierzu zusätzliche Informationen notwendig sind, um diese einplanen zu können. Aus diesem Grund werden Informationen, wie die notwendige Dauer und Betriebsmittel der Wartungsmaßnahmen im übergeordneten ERP-System (Enterprise Resource Planning) hinterlegt. Ein weiteres Hindernis stellt der hohe Aufwand der Maschinenanbindung dar, die sich durch gewachsene und heterogene Maschinenparks ergeben und somit jede Maschine individuell angebunden werden muss. Für das Projekt wurde daher festgelegt, dass die Maschinen mit je einer VWS ausgerüstet werden[1]. Die VWS erhält die benötigten Daten von der Maschinensteuerung oder über andere zwischengelagerte Informationsquellen, wie z.B. Datenbanken. Auch das MES wird mit einer VWS ausgestattet, damit die standardisierte VWS-Interaktion nach VDI 2193 umgesetzt werden kann.

[1] Auf die Erläuterung der Verwaltungsschale wird hier aus Platzgründen verzichtet.

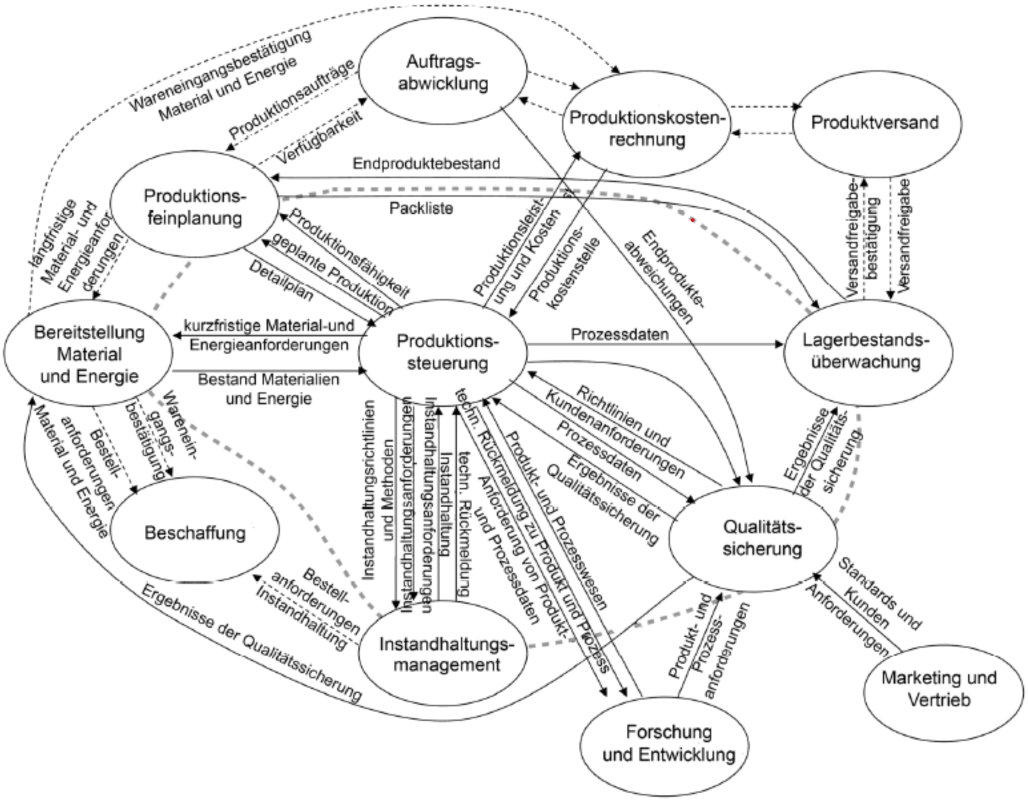
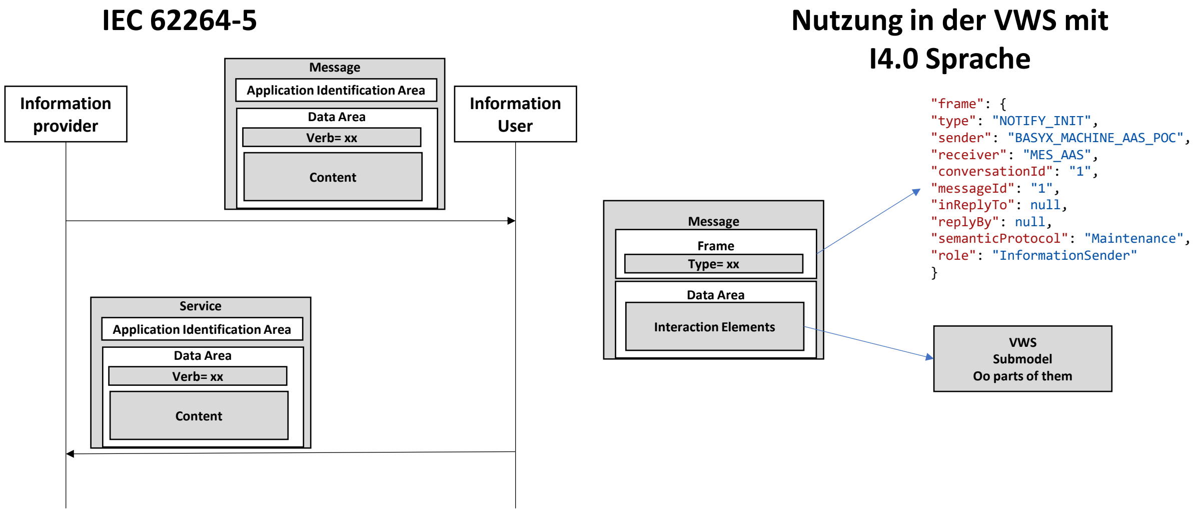

Auftrags-abwicklung
Produktionskosten-rechnung
Produktversand
Wareneingangsbestätigung Material und Energie
Produktionsaufträge
Verfügbarkeit
Produktions-feinplanung
Endproduktebestand
Packliste
langfristige Material- und Energieanforderungen
Produktionsfähigkeit
geplante Produktion
Detailplan
Produktionsleistung und Kosten
Produktions-kostenstelle
Endprodukte-abweichungen
Versandfreigabe-bestätigung
Versandfreigabe
Bereitstellung Material und Energie
kurzfristige Material-und Energieanforderungen
Bestand Materialien und Energie
Produktions-steuerung
Prozessdaten
Lagerbestands-überwachung
Richtlinien und Kundenanforderungen
Prozessdaten
Ergebnisse der Qualitätssicherung
Ergebnisse der Qualitäts-sicherung
Instandhaltungsrichtlinien und Methoden
Instandhaltungsanforderungen
Instandhaltungsanforderungen
Instandhaltung
techn. Rückmeldung
techn. Rückmeldung zu Produkt und Prozess
Anforderung von Produkt- und Prozessdaten
Produkt- und Prozesswesen
Waren-eingangs-bestätigung
Bestell-anforderungen
Material und Energie
Beschaffung
Qualitäts-sicherung
Bestell-anforderungen Instandhaltung
Ergebnisse der Qualitätssicherung
Instandhaltungs-management
Produkt- und Prozess-anforderungen
Standards und Kunden Anforderungen
Marketing und Vertrieb
Forschung und Entwicklung

Abbildung 1: Funktionsstrukturierung und Informationsaustausch nach IEC 62264 Teil 1 [IEC13]

Der Teil 5 (Business to manufacturing transactions) des IEC 62264 MES-Standards beschreibt Interaktionsmuster, die 12 Message-Typen definiert (“Verb = xx” in Abbildung 2). Diese werden jeweils für die Interaktionen zwischen den MES-Funktionstypen und der Fertigung kombiniert und mit den im Standard auch definierten Variablen der Informationsmodelle ausgestattet (“Content” in “Data Area” in Abbildung 2). Entsprechend des Aufbaus der I4.0-Sprache [VDI19] wird der Messagetyp zum „Typ“ der I4.0 Sprache und die Daten der „Data Area“ werden mit der JSON-Serialisierung der Teilmodelle der Verwaltungsschale gefüllt. Wie auch die MES-Message hat die I4.0-Nachricht einen Identifikationsbereich, der in Abbildung 2 als JSON-Text zu sehen ist. Hier ist noch einmal zu betonen, dass es sich um Interaktionsmuster auf Anwendungsebene handelt. Diese können mit verschiedenen Schicht-7-Kommunikationsprotokollen umgesetzt werden. Die Interaktionsmuster sind Kommunikationstechnologie-agnostisch.



```
"frame": {
"type": "NOTIFY_INIT",
"sender": "BASYX_MACHINE_AAS_POC",
"receiver": "MES_AAS",
"conversationId": "1",
"messageId": "1",
"inReplyTo": null,
"replyBy": null,
"semanticProtocol": "Maintenance",
"role": "InformationSender"
}
```

Abbildung 2: Abbildung IEC 62264 Teil 5 auf die I4.0 Sprache

Grundgedanke der Interaktionen sowohl von IEC 62264-5 MES-Standard und der I4.0 Sprache ist es, dass eine asynchrone Kommunikation verwendet wird. Asynchrone Kommunikation bedeutet, dass die Prozesse des Senders nicht auf eine sofortige Antwort warten, sondern eine eigene Instanz (in Abbildung 2 als “Information User” und “Information Provider” bezeichnet) der Interaktion instanziieren, in der der Ablauf des Interaktionsmusters enthalten ist. Das Interaktionsmuster ist also ein „stateful“ Protokoll. Sowohl Sender als auch Empfänger verwalten intern zu jedem Zeitpunkt den Zustand, in dem sich die Interaktion befindet. Erst wenn die Interaktion abgeschlossen ist, werden die Instanzen beendet. Es kommt also ein “peer-to-peer" Kommunikationsmodell zum Einsatz. Dieser Ansatz ist Kommunikationstechnologie-agnostisch. Das bedeutet, es ist unerheblich, ob HTTP/Rest, MQTT oder OPC UA als Kommunikationsprotokoll

3 Lösungskonzept

3.1 Die Architektur aus Anwendungssicht

Das Ziel des Vorhabens besteht in der Realisierung einer integrierten Produktions- und Instandhaltungsplanung für Bestandsmaschinen von KMU in der zerspanenden Industrie. Die zu entwickelnde Methode wird praxisnah an Werkzeugmaschinen im Bestand der Lauscher Präzisionstechnik GmbH (Lauscher) entwickelt und erprobt, die ein repräsentatives KMU in der zerspanenden Zulieferindustrie der Luft- und Raumfahrttechnik darstellt. Dabei kommt das BaSyx VWS Framework zum Einsatz. Die Daten, die über die steuerungsspezifischen Schnittstellen gelesen werden, werden in die für die Instandhaltung projektspezifisch zugeschnittene Teilmodelle der VWS in der Basyx-Software standardisiert eingelagert. Zur Integration genutzter Bestandssysteme werden mithilfe von Edge-Geräten vorhandene Maschinendaten über die jeweiligen, maschinenindividuellen Legacy-Protokolle für BaSyx und die VWS zugänglich gemacht. Diese agieren damit als Fundament für eine o. g. Planungsmethodik im MES. Eine Gesamtübersicht ist

Abbildung 3 zu entnehmen. Die Maschinen und das MES-System sind mit VWS ausgerüstet, die gesamte Kommunikation zwischen den Maschinen und dem MES wird von diesen VWS umgesetzt. Lokal werden die Daten zwischen den Maschinen und der VWS bzw. dem MES und der MES VWS über spezifischen Schnittstellen ausgetauscht. Für eine kooperative Lösung von prognostischen Planungsaufgaben werden dezentrale agentenorientierte Ansätze einbezogen.

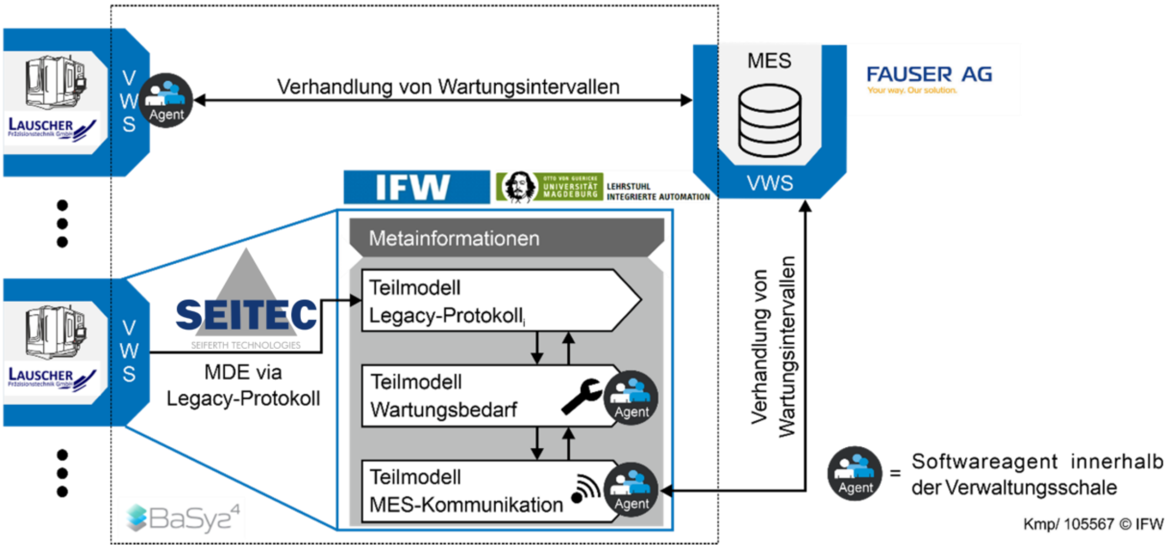

V W S
Agent
Verhandlung von Wartungsintervallen
MES
VWS
FAUSER AG
IFW
LEHRSTUHL INTEGRIERTE AUTOMATION
Metainformationen
SEITEC
Teilmodell Legacy-Protokoll
MDE via Legacy-Protokoll
Teilmodell Wartungsbedarf
Teilmodell MES-Kommunikation
Verhandlung von Wartungsintervallen
Agent = Softwareagent innerhalb der Verwaltungsschale
BaSys 4
Kmp/ 105567 © IFW

Abbildung 3: Gesamtüberblick über die integrierte Instandhaltung- und Produktionsplanung

3.2 Technologie-orientierte Architekturansatz

Die Architektur ist in Abbildung 4 dargestellt. Die Datenquellen der Werkzeugmaschinen sind die Maschinensteuerungen. In der Legacy-Variante, d.h. Steuerungen ohne frei zugänglich Steuerungsschnittstelle, werden die Daten in einer Datenbank (DB) hinterlegt. Die entsprechenden Maschinen-VWS greifen über die “IF_DB” auf diese Daten zu. Eine Edge-Komponente, die direkt auf den NC- und SPS-Teil der SINUMERIK-Maschinensteuerungen über die “IF_Edge” zugreifen kann, ist die Lösungsvariante für die Maschinensteuerungen mit frei zugänglichen Schnittstellen. Die Verwaltungsschalen lesen in beiden Varianten die benötigten Daten aus der Datenbank bzw. direkt von der Maschinensteuerung zyklisch aus. Die Verwaltungsschalen der Maschinen enthalten in der Variante 1 die Überwachung der geplanten Instandhaltungszyklen und initiieren die entsprechende Einordnung in die Produktionsplanung des MES. Das MES wird ebenfalls mit einer VWS ausgestattet, mit der entsprechenden lokalen proprietären Schnittstelle (“IF_MES”). Die Verwaltungsschalen der Maschinen und die vom MES interagieren und tauschen die Informationen mittels der I4.0-Sprache über die “IF_AAS_AAS” aus.

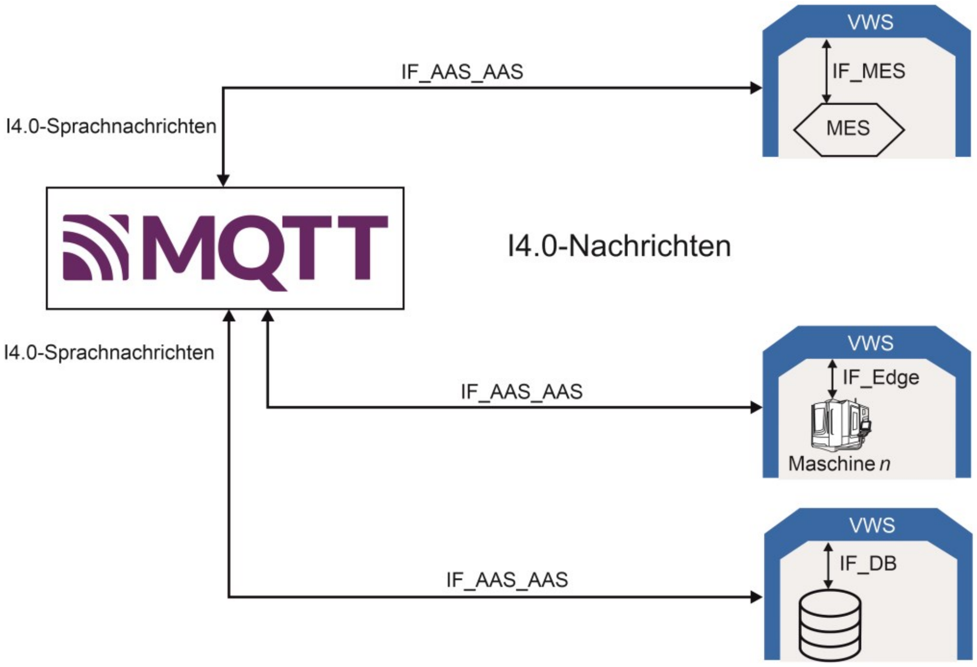


Abbildung 4: Architekturdefinition Instandhaltungsvariante 1

3.3 Die Interaktionsmuster

Abbildung 5 zeigt den Informationsfluss und die Funktionen der Instandhaltungsvariante 1, die die instandhaltungstechnischen Aufgaben, hier die Einordnung der Wartungsarbeiten in die Produktionsabläufe (Terminplanung) auf Basis der Überwachung der tatsächlich vorliegenden Betriebsstunden.

Abbildung 5 a) zeigt das zyklische pollen der Maschinendaten, insbesondere zur Ermittlung der Betriebsstunden in Instandhaltungsvariante 1.

Abbildung 5 b) zeigt mit unterschiedlichen Grautönen verschiedene, unabhängige Interaktionsprozesse, die jeweils aus der Anmeldung eines Wartungsbedarfs und deren erfolgreicher Absolvierung bestehen. Die Initiierung des Interaktionsprozesses wird von der Maschine-VWS ausgelöst und die Beendigung von der MES-VWS. Zwischen beiden Teilprozessen liegt ein nicht vorhersehbarer Zeitabschnitt, in dem physische Aktivitäten an der Maschine auszuführen sind. Das Instandhaltungspersonal gibt eine Meldung nach erfolgreicher Beendigung der Wartungsmaßnahme. Die Prozesse der verschiedenen Maschinen können ineinander verschachtelt sein.

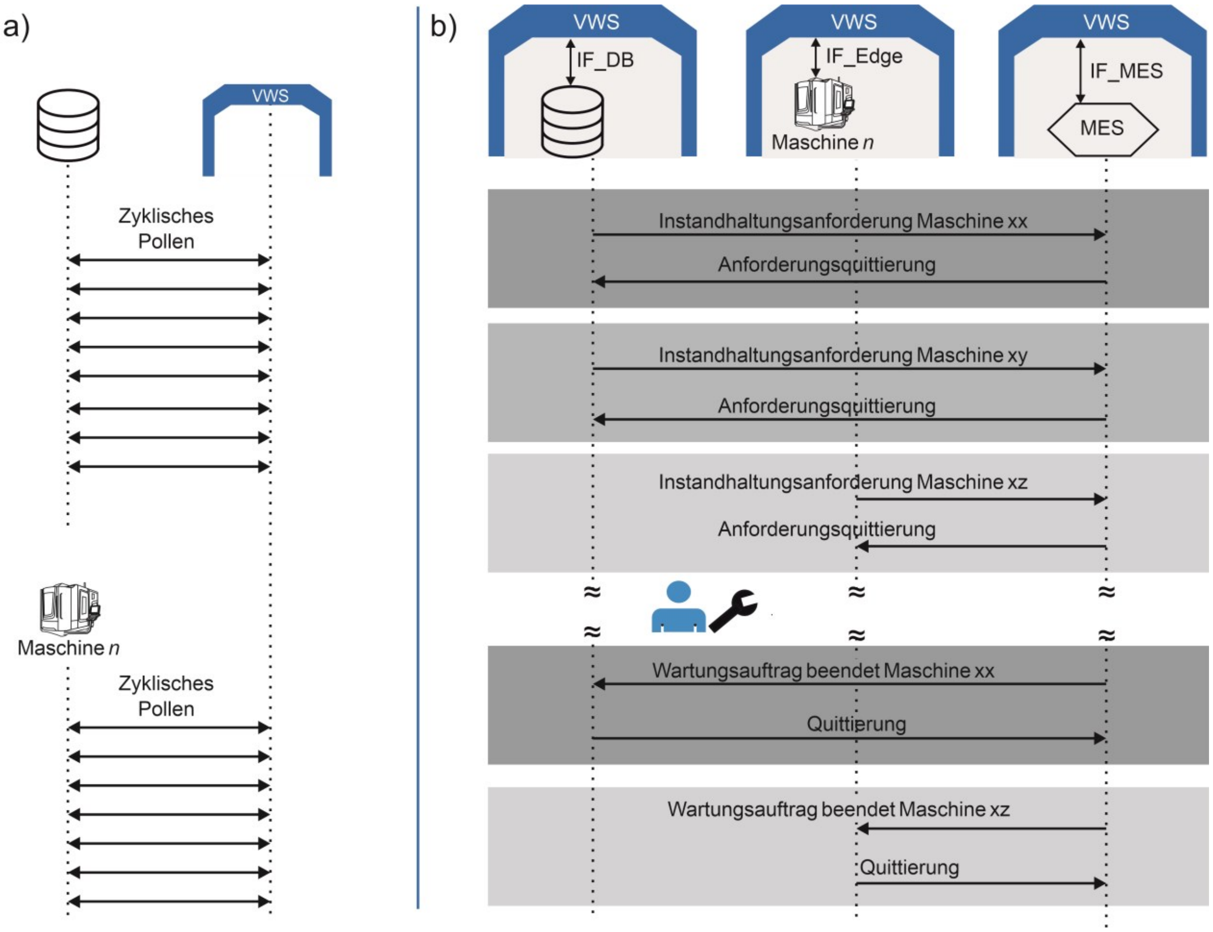

a)
VWS
Zyklisches Pollen
Maschine n
Zyklisches Pollen
b)
VWS
IF_DB
VWS
IF_Edge
Maschine n
VWS
IF_MES
MES
Instandhaltungsanforderung Maschine xx
Anforderungsquittierung
Instandhaltungsanforderung Maschine xy
Anforderungsquittierung
Instandhaltungsanforderung Maschine xz
Anforderungsquittierung
Wartungsauftrag beendet Maschine xx
Quittierung
Wartungsauftrag beendet Maschine xz
Quittierung

Abbildung 5: Betriebsstundenüberwachung für Wartung bei Maschinensteuerung mit Datenbankanschluss

Zunächst wird das Konzept des Interaktionsprinzips am Beispiel der Instandhaltungsvariante 1 erläutert. Beide Interaktionspartner (Maschinen- und MES-VWS) nehmen aktive Rollen ein, d.h. es liegt ein “peer-to-peer"-Interaktionsmuster vor. Beide Interaktionspartner agieren auf der Anwendungsseite zustandsbehaftet. Die Maschinen-VWS wartet nicht nur darauf, ob die Nachricht im MES angekommen ist, sondern ob diese den Wartungsbedarf gebucht hat. Hier ist zusätzlich zu beachten, dass zwischen der Wartungsanforderung und der Meldung, dass diese ausgeführt wurde erhebliche Zeiträume vergehen können. In einer Client-Server-Architektur müsste der Client so lange “pollen”, bis das erwartete Ergebnis verfügbar ist. Dies erfordert einen zusätzlichen aktiven Prozess im Client, hier in der Maschinen-VWS. Denn es muss nach erfolgreich durchgeführter Wartung der Stundenzähler für die jeweiligen Wartungsintervalle zurückgesetzt werden. Außerdem muss die MES-VWS den aktuellen Zustand der Wartungsintervallüberwachung jeder einzelnen Maschine verwalten. Dieses Interaktionsmuster ist durch die Informationsflussrichtung von der Datenquelle (Maschine) zur Datensenke (MES) charakterisiert, die bei einem Ereignis (hier die Überschreibung der Betriebsstunden für ein Wartungsintervall) aktiviert wird. Eine zentrale Einrichtung, die den Gesamtprozess organisiert entfällt. Zusammengefasst haben wir hier es mit einem “peer-to-peer", zustandsbasierten und ereignisgesteuerten Interaktionsmuster zu tun. Dieses Interaktionsprinzip entspricht dem einer konversations-orientierten Middleware, wie es in dem umfangreichen Artikel von Ivaki et. al. [IVA16] beschrieben ist. Die persistente Speicherung der Zustände der Interaktion für die Wiederherstellung des Interaktionszustandes bei Ausfall einer Komponente sowie das wiederholte Senden von nicht quittierten Nachrichten gehören zu den wesentlichen Eigenschaften dieses Prinzips.

Die Abbildung 6 zeigt ausschnittsweise die Interaktion zwischen den Maschinen- und MES-VWS im Detail. Wird in der Maschinen-VWS eine Grenzwertüberschreitung der Betriebsstunden für einen der Wartungsintervalle registriert, erfolgt zunächst eine Instanziierung des Interaktionsprozesses. Ist dies erfolgt, wird nach IEC 62264-5 eine “notify_init” Nachricht and die MES-VWS versendet. Darin sind die benötigen Informationen (die in einem entsprechenden Wartungs-Teilmodell in der VWS hinterlegt sind) in I4.0-Sprache (JSON-Format) enthalten. Die MES-VWS instanziiert ebenfalls einen entsprechenden Interaktionsprozess in dem die MES-VWS quittiert (“notify_accepted”), wenn diese Nachricht an das MES ausgeliefert worden ist. Dadurch ist in der Maschinen-VWS bekannt, dass eine Bearbeitung des Auftrags erfolgt. Erhält die Maschinen-VWS innerhalb einer vordefinierten Zeit keine “notify_accepted” Nachricht, so wird das “notify_init” mehrmals wiederholt. Nach 5-maliger Wiederholung wird dann abgebrochen und ein Verbindungsfehler lokal gemeldet.

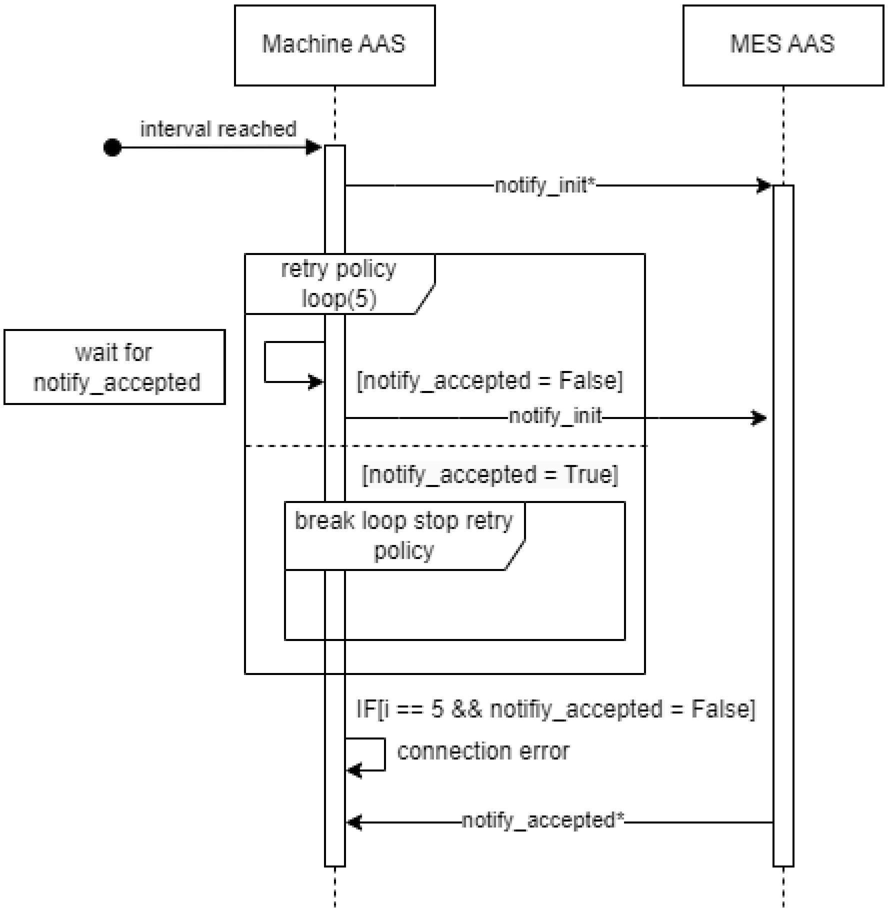


Abbildung 6: Nachricht von Maschinen-VWS and MES-VWS nach Interaktionsmuster von IEC 62264-5

Wie oben bereits erwähnt, kann die Ausführung der physischen Wartungsarbeiten Stunden oder Tage dauern. Ist der Auftrag erfüllt sendet das MES eine entsprechende Nachricht und die MES-VWS sendet eine “change”-Nachricht an die entsprechende Maschinen-VWS mit an dem zugehörigen Interaktionsprozess (Abbildung 7). Die Maschinen-VWS hat diese Nachricht ebenfalls zu quittieren (“respond”), um sicher zu gehen, dass die VWS aktiv ist. In der Maschinen-VWS ist der Betriebsstundenzähler für das Instandhaltungsüberwachungsintervall auf null zu setzen und eine lokale Kopie des Vorgangs zu erstellen (u.a. wann gemeldet, wann erfolgreich abgeschlossen). Ist die Interaktion beendet, werden beide Prozesse gelöscht. In der Maschinen-VWS erfolgt eine Speicherung der notwendigen Daten für den jeweiligen Wartungsauftrag (persistente VWS).

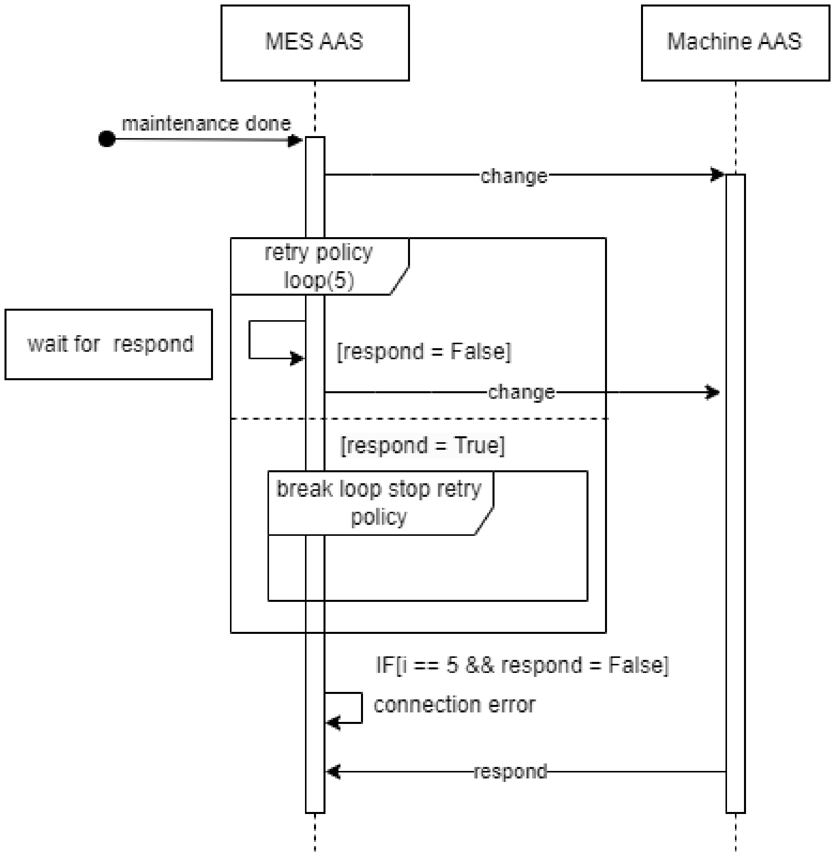


Abbildung 7: Nachricht von MES VWS an Maschinen VWS nach Interaktionsmuster von IEC 62264-5

In beiden Teilinteraktionen werden unterschiedliche Diensttypen verwendet. Der “Notify”-Dienst zeigt ein Ereignis im Absender an, er zielt nicht auf ein bestimmtes Objekt des Empfängers. Im “Change”-Dienst wird dagegen gezielt etwas beim Empfänger verändert. Beide Dienste haben also die in IEC 62264-5 definierten unterschiedlichen Bedeutungen und werden deshalb auch entsprechend in diesem Interaktionsmuster verwendet.

3.4 Agentenintegration – ein Ausblick

Der unter 3.3 geschilderte Prozess ist kommunikativ verhältnismäßig einfach umzusetzen. Trotzdem ist er für viele Unternehmen ein bedeutender Schritt, weil die Einordnung der geplanten Instandhaltung in die Produktionsplanung von der Maschine selbst automatisch eingeordnet werden kann. Der nächste Schritt, die Instandhaltungsvariante 2 (Abbildung 8), ist eine zustandsbasierte Überwachung der Maschinenkomponenten (z.B. Spindelantrieb, Hydrauliksystem). Durch eine RUL-Prognose (siehe Abschnitt 2.1) wird ein potentieller Ausfall vorhergesagt. Mit dieser Auskunft kann dann wieder im MES, automatisch eine Einordnung der Wartungsmaßnahme vorgenommen werden. Für diese Aufgabe werden dezentrale Agenten an die Maschinen-VWS angekoppelt und ein zusätzlicher Management-Agent erstellt. Auch der Management-Agent wird mit einer VWS ausgestattet, um die in der Instandhaltungsvariante 1 vorhandene Interaktionsprinzipien anwenden zu können. Von den Agenten sind folgende Aufgaben zu lösen:

- Maschinen-Agent
    - Dieser Agent überwacht die zyklisch anfallenden Maschinendaten und ermittelt daraus den Nutzungsvorrat und gegebenenfalls prognostisch einen potentiellen Komponentenausfall.
    - Dazu verwendet er den Ausfallwahrscheinlichkeitsdienst des Management-Agenten.
    - Der prognostizierte Ausfallzeitpunkt wird an den Management-Agenten gesendet.

- Management-Agent
    - Ermittlung eines potentiellen Reparaturauftrages. Ausgangspunkt ist der vorbestimmte Ausfallzeitpunkt. Dazu gehören:
        - Bestimmung der benötigten Ersatzteile und deren zeitliche Verfügbarkeit. Ersatzteile können im lokalen Lager liegen, innerhalb von 24h verfügbar sein oder erst nach mehreren Tagen oder Woche geliefert werden. Dazu existiert in unserem Anwendungsfall bei Lauscher eine entsprechende Datenbank. Diese wird vom Management-Agenten lokal abgefragt.
        - Bestimmung des Personals, die die Wartung vornehmen soll. Dazu existiert bei Lauscher eine Kompetenzmatrix, die zum Maschinentyp und zum Gewerk Auskunft gibt.
        - Bestimmung der Zuordnung des ausgewählten Personals zu den Schichten.
        - Bestimmung der Fertigungsaufträge auf anderen Maschinen, die ggf. durch die Wartung durch das zugeordnete Personal betroffen sind
    - Information aller betroffenen Maschinen-Agenten
- Maschinen-Agent
    - Alle durch die Wartung betroffenen Maschinen-Agenten verhandeln kollaborativ einen Wartungszeitraum, der möglich und kostenminimal ist
    - Dieser Wartungszeitraum wird an den Maschinen-Agent kommuniziert
- Management-Agent
    - Der Management-Agent gibt den Vorschlag an das MES weiter
    - Die Einordnung des Wartungsauftrages wird dann vom MES geprüft und vorgenommen.

Die Interaktionen zwischen allen Komponenten (auch zum Ersatzteil-Dienst, Personal-Dienst und Ausfallwahrscheinlichkeits-Dienst) verwenden die I4.0-Sprachnachrichten. Der Austausch der Nachrichten der I4.0-Sprache erfolgt über MQTT.

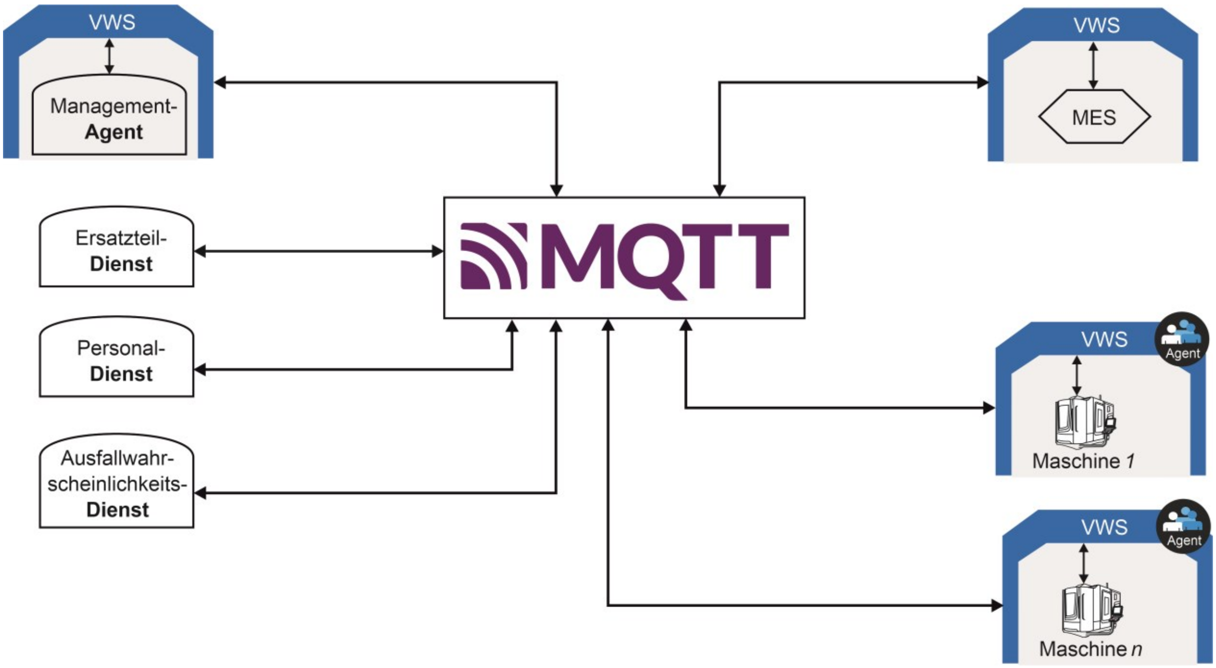


Abbildung 8: Ausblick mit Agenten (Instandhaltungsvariante 2)

4 Zusammenfassung und Erfahrungen

Im Projekt wurden folgende wesentliche Erfahrungen gemacht:

- Da alle Komponenten einen aktiven Part in der Interaktion einnehmen, ist eine asynchrone “peer-to-peer"-Interaktion zwischen den Anwendungen erforderlich.
- Eine reine Client-/Server Kommunikation auf Anwendungsseite ist nicht angebracht, da durch die Verteilung der instandhaltungstechnischen Aufgaben, zwischen Maschinen- und MES-VWS die Zustandsinformationen gehalten werden müssen.
- In der VWS muss eine persistente Datenhaltung vorhanden sein, da bei Neustart der VWS auf die bereits vorhandenen Informationsstände zurückgegriffen werden muss (z.B. wie viele Wartungszyklen wurden schon durchgeführt).
- Das “Befüllen” der VWS ist proprietär, da ein heterogener Maschinenpark verschiedene Steuerungstypen, Protokolle und Versionsstände mit sich bringt.
- Die Auswirkung der organisatorischen Änderungen in der Produktion und deren Vereinbarkeit des Qualitätsmanagementsystems mit notwendigen Zertifizierungen, müssen parallel zur Entwicklung eines hier vorgestellten Systems mitgedacht werden. Dabei stehen die Vorteile eines solchen Systems im Vordergrund, da einerseits die nicht wertschöpfenden Tätigkeiten zur Erfassung der Betriebsstunden und Abklärung der notwendigen Betriebsmittel automatisiert werden und so bei der Planung der Produktion mehrere Ressourcen gleichzeitig Berücksichtigung finden, wodurch die Fertigungsplanung optimal unterstützt wird.

Literaturverzeichnis

[BEL19] Belyaev A, Diedrich Ch (2019) Aktive Verwaltungsschale von Industrie 4.0-Komponenten - Erscheinungsformen von Verwaltungsschalen. Automation Kongress 2019. VDI Tagungsband – digital.

[DEN19] Denkena B, Dittrich M-A, Wilmsmeier S (2019) Automated production data feedback for adaptive work planning and production control. *Procedia Manufacturing* 28:18–23.

[DEN20a] Denkena B, Dittrich M-A, Keunecke L, et al. (2020) Continuous modelling of machine tool failure durations for improved production scheduling. *Prod. Eng. Res. Devel.* 14(2):207–215.

[HAS11] Hashemian HM (2011) State-of-the-Art Predictive Maintenance Techniques. IEEE Trans Instrum Meas 60:226–236. doi:10.1109/TIM.2010.2047662.

[IEC13] IEC 62264-1:2013: Enterprise-control system integration — Part 1: Models and terminology. IEC Geneva 2013.

[IEC16] IEC 62264-5:2016: Enterprise-control system integration — Part 5: Business to manufacturing transactions. IEC Geneva 2016

[IVA16] Ivaki N, Laranjeiro N, Araujo F (2018) A survey on reliable distributed communication. The Journal of Systems and Software 137:713-732. http://dx.doi.org/10.1016/jss.2017.03.028.

[KUH17] Kuhnle A, Kuttler M, Dümpelmann M, et al. (2017) Intelligente Produktionsplanung und -steuerung: Erlernen optimaler Entscheidungen. *wt Werkstatttechnik online* 107(9):625–629.

[LEI18] Lei Y, Li N, Guo L et al. (2018) Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing 104:799–834. https://doi.org/10.1016/j.ymssp.2017.11.016.

[PEN10] Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50:297–313. doi:10.1007/s00170-009-2482-0.

[SCH07] Schwabacher M, Goebel K (2007) A survey of artificial intelligence for prognostics. AAAI Fall Symposium - Technical Report.

[SEZ18] Sezer E, Romero D, Guedea F, et al. (2018) An Industry 4.0-Enabled Low Cost Predictive Maintenance Approach for SMEs. *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, S.1–8.

[STU90] Sturm A, Förster R (1990) Maschinen- und Anlagendiagnostik. Vieweg+Teubner Verlag, Wiesbaden.

[THE10] AUTOMATION, ZVEI-Fachverband; Theobald, Carolin (2010). Manufacturing execution systems (MES). Branchenspezifische Anforderungen und herstellerneutrale Beschreibung von Lösungen. ZVEI, Frankfurt.

[VDI19] VDI/VDE 2193 Blatt 1: Sprache für I4.0-Komponenten (2019) Düsseldorf: VDI, VDI/VDE 2193 Blatt 2: Sprache für I4.0-Komponenten. Interaktionsprotokoll für Ausschreibungsverfahren. Düsseldorf: VDI.

[VWS22] Platform Industrie 4.0 (2022) "Details of the Asset Administration Shell. Part 1 - The Exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC02)," Federal Ministry for Economic Affairs and Climate Action (BMWK), Berlin.

[ZON20] Zonta T, da Costa CA, da Rosa Righi R, et al. (2020) Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering* 150:106889.

http://dx.doi.org/10.25673/111635

Automation of an industrial OPC UA FX use case through the usage of proactive Asset Administration Shells

Katharina Justmann[1], Ludwig Leurs[2], Jesko Hermann[3], Martin Ruskowski[4]

Abstract: Flexible and order-driven production are essential concepts for smart factories that effectively carry out lot-size-one production. These concepts depend on manufacturer-independent information exchange throughout the production lifecycle. While Open Platform Communication Unified Architecture Field eXchange (OPC UA FX) enables standardized communication between shop floor controllers, its configuration requires additional effort. To achieve flexible communication in changing production environments the configuration needs to be automated. The Asset Administration Shell (AAS) and the associated bidding process offer a potential solution to achieve this automation. The concept of a combination of the AAS and OPC UA FX was already formulated [DWR22a]. In this paper this concept is extended with activities of the Capability, Skill, and Service model to automate the configuration and production planning process of an OPC UA FX demonstrator. This paper shows that the information models and concepts are complementary yet lack generic mappings between the defined structures and semantics.

Keywords: OPC UA Field eXchange (OPC UA FX), Asset Administration Shell (AAS), Bidding Process, Capability Skill and Service model, Industry 4.0

1 Introduction

In Industry 4.0 (I4.0), one of the main goals is to achieve flexible manufacturing by enhancing communication between the various assets that are part of the production process. Different technologies have been developed and standardized throughout the last years to achieve manufacturer independent communication. These technologies establish the foundation for flexibility in production processes and the mass production of customized lot-size-one products. Two of these standards are the Asset Administration Shell (AAS) and Open Platform Communication Unified Architecture Field eXchange (OPC UA FX). This paper shows the necessary automation of the configuration process

[1] Bosch Rexroth AG, Berliner Str. 25, 64711 Erbach, Germany, Katharina.Justmann@boschrexroth.de
[2] Bosch Rexroth AG, Bürgermeister-Dr.-Nebel-Straße 2, 97816 Lohr am Main, Germany, Ludwig.Leurs@boschrexroth.de
[3] Technologie-Initiative SmartFactory KL e.V., Trippstadter Str. 122, 67663 Kaiserslautern, Germany, Jesko.Hermann@smartfactory.de
[4] RPTU Kaiserslautern, Lehrstuhl WSKL, Gottlieb-Daimler-Str. 42, 67663 Kaiserslautern, Germany, Martin.Ruskowski@rptu.de

defined in the OPC UA FX specifications by applying these two standards to the already existing OPC UA FX demonstrator of the OPC Foundation [OP23a].

The AAS represents information and capabilities of assets in a virtual context. This information is represented by defined submodules of the AAS, allowing the standardized definition of different aspects of industrial assets throughout the entire asset lifecycle [ID23a]. On the other hand, OPC UA enables standardized vertical and horizontal communication in industrial automation. The Field Level Communication (FLC) initiative of the OPC Foundation has standardized OPC UA FX to enable a manufacturer-independent real-time controller-to-controller (C2C) communication on the shop floor. OPC UA FX enables the creation and reconfiguration of OPC UA PubSub communication between automation components during runtime [OP22a].

In the context of flexible production with the request of fast on-the-fly reconfiguration, the OPC UA FX specification does not propose an automated way for the configuration process. In Part 80 of the OPC UA FX specifications the configuration process is executed through control or system engineers that decide which process lines need to interact with each other [OP22b]. Additionally, some configuration parameters needed for establishing a connection are provided during the engineering stage. To improve these matters, defined AAS processes can help. A first conceptual approach to create a combination of the processes of the AAS and OPC UA FX was described by Ch. Diedrich et al. [DWR22a]. The combination is based on the bidding process specified in VDI/VDE 2193-2 [VD20].

This publication will further develop the concept by applying it to an industrial use case implemented in the OPC UA FX demonstrator of the OPC Foundation. A group of controller manufacturers designed and implemented the demonstrator to exhibit the applicability and interoperability of OPC UA FX in a flexible production use case. In the demonstrator users need to choose manually which controllers should be part of the production line. In this paper the demonstrator will be enhanced to showcase the automated process of creating production lines on-the-fly based on product description and the bidding process. This is achieved by enabling the product and the automation components as I4.0 components with proactive AAS. The behaviour of the different types of AAS are described in chapter 2.4. Additionally, this paper shows how the existing Capability, Skill and Service model is used to map automation functions from OPC UA FX to the AAS. The mapped automation functions and the data stored in the AAS is then used to create OPC UA FX communication connections.

2 Background and Related Work

This chapter gives a short overview of the current state of the art regarding OPC UA FX, the Capability, Skills and Service model and latest developments about the AAS. Additionally, it will present related work like the OPC UA FX demonstrator and a concept that presents first steps for enhancing OPC UA FX with the AAS.

2.1 OPC UA Field eXchange

The OPC UA FX specifications define a manufacturer-independent horizontal communication on the field level between controllers (C2C) and in the future communication to or between field devices. The current release of the OPC UA FX standard uses PublishSubscribe (PubSub) architecture with the possibility to additionally use parts of the Time-Sensitive Networking (TSN) standard. OPC UA FX defines the entities *AutomationComponent* and *ConnectionManager*. *AutomationComponents* represent assets that can perform one or more automation functions. These functions are defined as *FunctionalEntities* with input and output parameters, configuration parameters, diagnostic information, and additional identification properties in the scope of OPC UA FX. *FunctionalEntities* can represent simple subtasks like moving a robot arm to a specific position or multiple subtasks combined into larger tasks, such as entire process steps. *AutomationComponents* also include information about the assets that are used for identification purposes. The structure of *AutomationComponents* and a *ConnectionManager* are illustrated in Fig. 1. The *ConnectionManager* connects to *AutomationComponents* via OPC UA Client/Server and calls the *EstablishConnections* method to establish the communication between two *AutomationComponents*. The *ConnectionManager* creates the communication based on *ConnectionConfigurationSets* which include all necessary parameters needed for configuration. These sets are either created through an offline descriptor that is read into the *ConnectionManager* as a file or based on inputs via methods in the OPC UA address space of the *ConnectionManager*. The established communication between two *AutomationComponents* is called logical connection and is based on the OPC UA PubSub standard [OP22a].

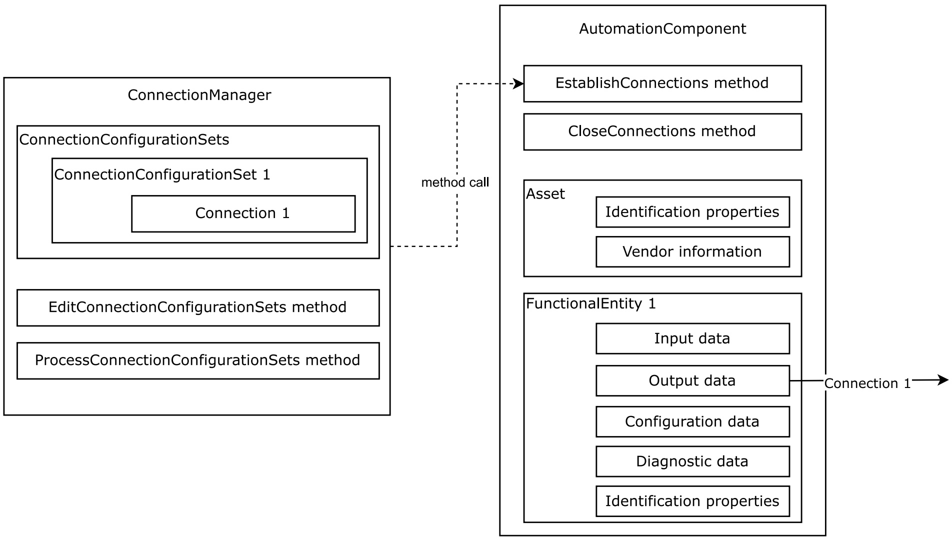

ConnectionManager
ConnectionConfigurationSets
ConnectionConfigurationSet 1
Connection 1
EditConnectionConfigurationSets method
ProcessConnectionConfigurationSets method
method call
AutomationComponent
EstablishConnections method
CloseConnections method
Asset
Identification properties
Vendor information
FunctionalEntity 1
Input data
Output data
Configuration data
Diagnostic data
Identification properties
Connection 1

Fig. 1: Structure of OPC UA FX components

2.2 OPC UA FX Demonstrator

To showcase the OPC UA FX specifications various manufacturers implemented an OPC UA FX demonstrator in the Prototyping technical working group that is part of the Unified Architecture Field eXchange working group of the OPC Foundation [OP23b]. This demonstrator presents the interoperability and on-the-fly configuration and reconfiguration during runtime of OPC UA PubSub communication connections. The controllers in the demonstrator are setup in a star network topology. The chosen user story for the demonstrator is a beer bottling plant. Each controller is an *AutomationComponent*, and each process step is modelled as a *FunctionalEntity*. The output parameter of a *FunctionalEntity* corresponds to the input parameter of the successor *FunctionalEntity*. The process steps range from selecting and washing different sizes of bottles to capping and labelling a bottle. For each process step, more than one controller represents a step to allow for more customization possibilities and to create more than one process line in parallel. For example, one controller fills a bottle with Pilsner, while another would be responsible for bottling stout beer. An overview of the demonstrator with a possible setup of communication can be seen in Fig. 2. Through a user interface of the *ConnectionManager* the user can manually select the *AutomationComponents* and the *FunctionalEntities* available in the controller that should be connected. Besides that, the user additionally decides on other parameters like publishing intervals, timeouts and the interface used for transmission. Based on this input the *ConnectionManager* creates the OPC UA PubSub configuration. The *ConnectionManager* then connects via OPC UA Client/Server to the selected *AutomationComponents* and establishes the connections between the controllers. The user needs to repeat these configuration steps for each connection that needs to be established for the entire production line.

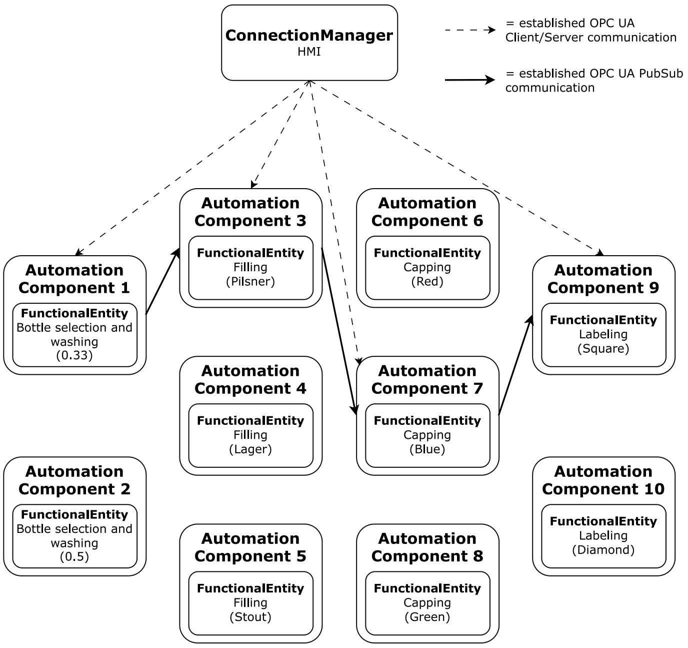


Fig. 2: Example for established communication in the OPC UA FX demonstrator

2.3 Capability, Skill, and Service Model

A working group of the “Plattform Industrie 4.0” initiative developed the Capability, Skill, and Service model (CSS model) which extends the Product, Process, and Resource model (PPR model) [Di22b]. Resources in the scope of the PPR model produce products by completing process steps, like a labelling machine for instance that puts a label onto a surface. This is one sub-process in the scope of the production process of bottling beverages for example. The implemented and executable task of the labelling machine is represented as a Skill in the CSS model. A Skill consists of a *SkillInterface* and Parameters. The *SkillInterface* is used to execute a Skill while the Parameters represent the input, output, control, and diagnostic parameters of the Skill. The abstract description of a Skill is the so-called Capabilities of a production resource. Capabilities have *CapabilityConstraints*. These constraints are conditions that need to be fulfilled either prior to, during or following the execution of the task that is linked to the Capability in question. *CapabilityConstraints* can be used to create an order of the required Capabilities if production steps are dependent on each other. Services are a representation of Capabilities in an economic sense. They extend the description of the Capability with information such as energy consumption of the production process, delivery dates, costs or even details about laws that are followed [Di22b]. Capabilities and Services also have properties. These properties describe characteristics of the entities. Capability properties can also have direct impacts on the parameters of Skills. For example, if a Capability's property describes the colour of a label, this value serves as an input for the corresponding Skill. Additionally, the CSS model defines various activities that describe how Capabilities, Skills and Services are used in an industrial use case [Di22b].

2.4 Asset Administration Shell

The AAS is a standardized way to create I4.0 components by enabling manufactures to provide the digital twin of an industrial asset. I4.0 components are industrial assets consisting of their physical and virtual representation [BVZ15]. Through various submodules the AAS can represent different aspects of assets like functionality or physical dimensions. Currently a working group of the Industrial Digital Twin Association (IDTA) is in the process of defining a submodule to represent Capabilities in the AAS as specified in the CSS-model [ID23b]. Besides the stored information also interaction patterns and semantics for an I4.0 language are defined in the scope of the AAS specifications. Three interaction patterns, passive, reactive, and proactive, are specified for the AAS. The first interaction pattern supports only passive interactions where the asset information is stored as a file format. A reactive AAS can provide the asset information via a server interface. The last type allows the AAS to proactively negotiate with another AAS. These proactive AAS use a bidding process to interact [VD20, Pl20]. When equipping an asset with a proactive AAS, the AAS consists of a reactive or passive AAS and a proactive part. The passive or reactive AAS is responsible for the storage of the asset information while the reactive part is in control of handling interactions and the bidding process [Gr22]. The bidding process defined in VDI/VDE 2193-2 describes how a customer requests different services from a contractor to fulfil the required process steps to produce a product [VD20].

2.5 Enhancing OPC UA FX with the bidding process

The concept used to enhance the OPC UA FX demonstrator was first published by Ch. Diedrich et al. [DWR22a]. The publication describes a concept for the combination of OPC UA FX and the AAS. The combination enables the configuration via OPC UA FX with the help of data from the AAS. Different submodules of the AAS contain the information needed for the *ConnectionManager* to configure the communication between the *AutomationComponents* of two adjacent process steps. Which *AutomationComponents* need to be connected is decided based on the outcome of a bidding process. One *AutomationComponent* in the scope of OPC UA FX provides one or more *FunctionalEntities*. These *FunctionalEntities* are mapped to the Services presented in the AAS. A ServiceProvider then makes an offer based on the available Service. A ServiceRequester can accept this offer based on different parameters like cost or energy consumption. If a ServiceRequester accepts the offer, the additional configuration parameters needed for the *ConnectionManager* are read out of the AAS. These parameters are then given to the *ConnectionManager*, which automatically establishes the connections between the controller of each process step [DWR22a].

3 Process of establishing connections through the bidding process

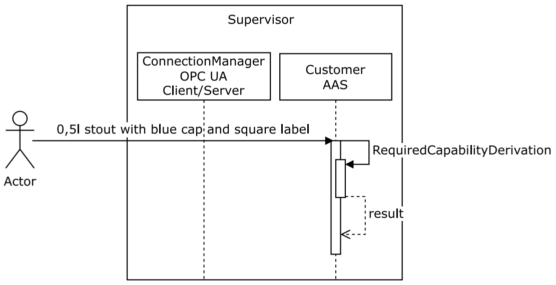
In this chapter the bidding process in the scope of the OPC UA FX demonstrator is shown. The sequence diagrams Fig. 3 to Fig. 5 illustrate the described process.

To automate the OPC UA FX demonstrator via a bidding process each *AutomationComponent* is equipped with a reactive and a proactive AAS [Gr22]. The information about each of these automation components, its Capabilities, its asset description and its OPC UA Server connection information are represented in the AAS to allow an automated configuration.

In the bidding process defined in VDI/VDE 2193-2 the user first provides either a description of the requested product or a set of processes or required Capabilities as defined in a submodule template. Such a product description could include description of the physical dimensions and features of the requested product or similar specifications. Regarding the OPC UA FX demonstrator use case the user can select a set of features for producing a bottle via an HMI. Using the following example product description, the process of handling this description, the bidding process and the following C2C configuration will be shown. The example description reads as follows: a 0.5-litre bottle of stout beer with a blue cap and square label.

This description is now used to generate a production order and create a digital representative of the product. This representative consists of the product features, requirements and later also the necessary production steps to create the product. The digital representative is an instance of an active and proactive AAS, allowing the product to be an I4.0 component and participating in the bidding process to directly create orders for its own features. The benefit of this is that information about the product is directly attached and stored during the production process. This information can later be used to comply with laws like the required digital product passport that is currently discussed in the European Union. Furthermore, a product that is represented as a I4.0 component allows more dynamic on-the-fly configurations that align with the concepts of adaptable manufacturing and self-organizing supply chains [Gr22].

An algorithm breaks the production description into features and smaller subtasks. Following the CSS model the derived subtasks or features can then be mapped either to Services or to Capabilities, depending on the additional requirements for the bidding process. When dealing with Shared Production involving multiple companies, also different laws or other non-production parameter need to be considered. In such cross-company settings, it is required to map the subtasks to Services. The OPC UA FX demonstrator showcases a use case for an adaptable factory that concentrates on internal production processes. In the scope of this use case derived features are mapped to Capabilities using the activity *RequiredCapabilityDerivation* defined in the CSS whitepaper [Di22b]. The given example results to a derivation of four Capabilities. The four Capabilities are washing a 0,5-litre bottle, bottling stout beer, closing the lid with a blue cap, and finally attaching a square label to the bottle. In the given example, the bottle must be washed before it can be filled, and the capping process step needs to be executed after filling. These dependencies are expressed through *CapabilityConstraints* [Di22b] and are used to outline a work plan. The work plan is then stored in the AAS of the product. The resulting Capabilities of the *RequiredCapabilityDerivation* with the corresponding parameters and constraints are shown in Fig. 3.


Supervisor
ConnectionManager
OPC UA
Client/Server
Customer
AAS
Actor
0,5l stout with blue cap and square label
RequiredCapabilityDerivation
result

Result			
	Capability	Parameter	Constraint
1.	select and wash bottle	size: 0.5l	
2.	fill bottle	content: stout	pre: finish select and wash bottle
3.	cap bottle	colour: blue	pre: finish fill bottle
4.	label bottle	shape: square	pre: finish cap bottle

Fig. 3: Sequence diagram and result of *RequiredCapabilityDerivation*

In the paper of Ch. Diedrich et al., the entities taking part in the bidding process are referred to as ServiceRequester and ServiceProvider [DWR22a]. However, as we are using Capabilities for mapping rather than Services in the scope of the OPC UA FX demonstrator these terms are not adopted in this paper. Instead, the terms specified in the specification of the bidding process, customers, and contractors [VD20], are used. The customer entity derives the required Capabilities from the product description and requests resources that can provide the required Capabilities to fulfil the production order. The request is made through a call for proposal which is send to the potential contractors [VD20]. To establish a more dynamic and optimized workflow in the demonstrator the customer submits requests for the next required production step outlined in the work plan instead of blocking a whole production line like in the original demonstrator. The contractor is the individual *AutomationComponent* which is responsible for the production step. It matches the required Capabilities from the proposal to its own offered Capabilities. This activity is defined as *CapabilityMatching* [Di22b]. If possible, the contractor can directly map the Capabilities. If not, the contractor can either reject the proposal as it is not able to fulfil the requirements or decompose the Capability into sub-Capabilities and attempt the matching again. *CapabilityMatching* not only compares the notations of Capabilities but also if the constraints are met and if the Capability properties match [Di22b]. A possible production step is the filling of the bottle with stout beer. Filling the bottle is a Capability of the *AutomationComponent*, filling it with stout is a property of the

Capability and also a constraint if the *AutomationComponent* is only able to fill bottles with stout and nothing else. Another constraint of this Capability is before filling, the bottle needs to be cleaned. If a set of matching Capabilities is found the contractor checks if it is feasible to produce the requested product based on the matched Capabilities, the given properties, and the capacities of the production resource using the *FeasibilityCheck* activity defined in the CCS whitepaper. Additionally, the *FeasibilityCheck* can calculate the process times and the costs of a Skill execution [Di22b]. This can be used in production settings where the aim is to optimize production with regards to changeover times. Within the scope of the OPC UA FX demonstrator the only property that is checked is whether the *AutomationComponent* is already part of a production line or not. If the *AutomationComponent* is already part of a production line, the contractor rejects the proposal; otherwise, it sends an offer to the customer. The offer includes parameters based on which the customer is able to decide on whether or not to accept the offer. In the scope of the demonstrator the offer is always accepted. Fig. 4 illustrates the sequence of steps from the customer's call for proposal to the submission of an offer.

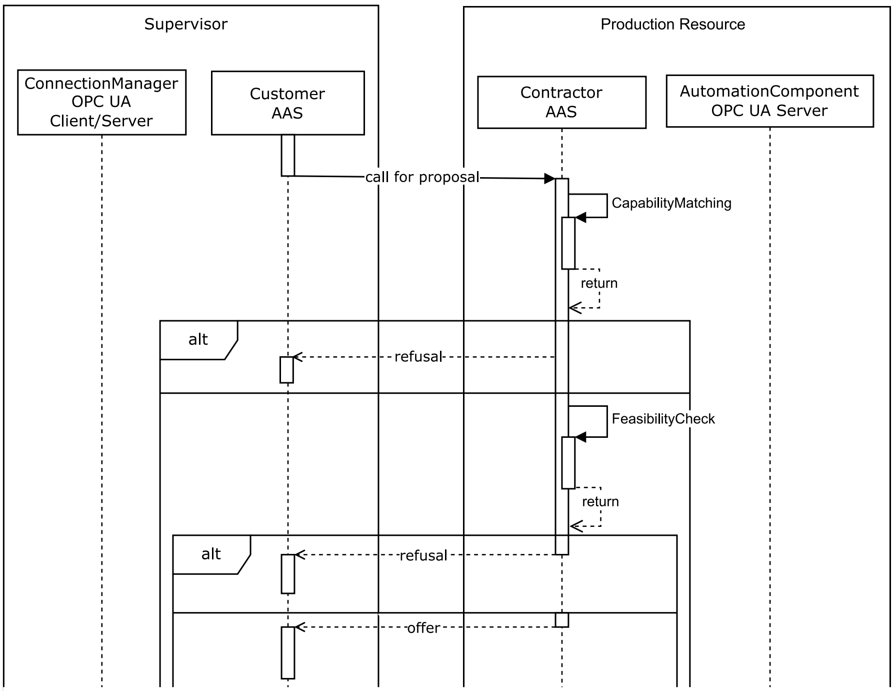


Fig. 4: Sequence diagram with *CapabilityMatching* and *FeasibilityCheck*

If the offer is accepted the configuration of the OPC UA FX communication is generated. To create such a configuration several parameters are required, which must be included in the *AutomationComponents* AAS. This enables the customer to read the configuration parameters out of the contractor's AAS. Refer to Fig. 4 for a visual representation of this steps. The contractor passes the configuration parameters to the *ConnectionManager* entity which will create the OPC UA FX configuration. The *ConnectionManager* connects then as an OPC UA Client with the OPC UA Server of the *AutomationComponent* to call the FX specific methods for establishing a OPC UA FX connection.

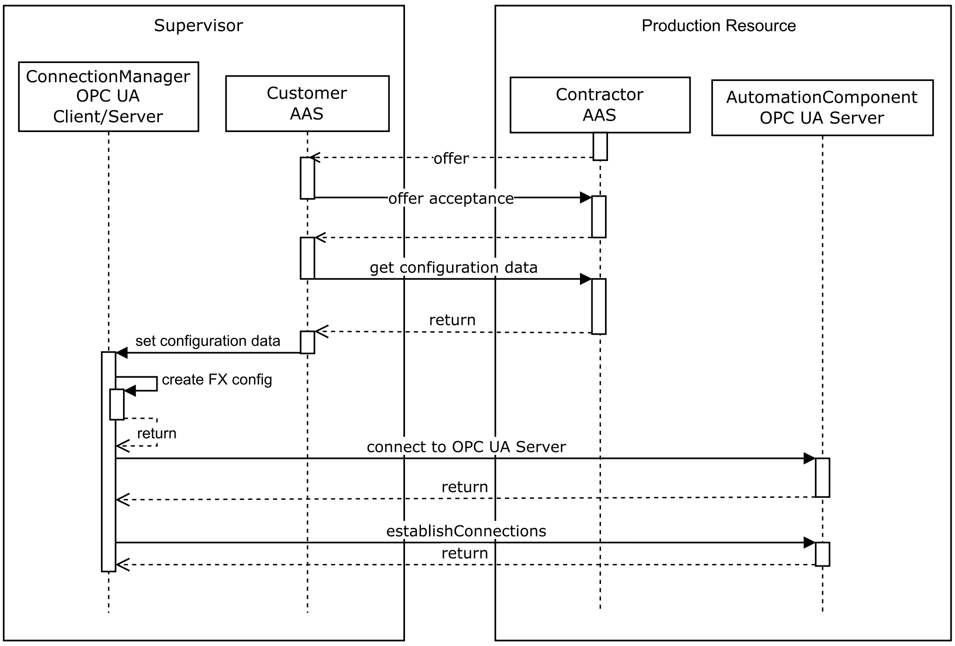


Fig. 5: Sequence diagram for getting the OPC UA FX configuration parameters

4 Conclusion and further Work

This paper utilised the concept by Ch. Diedrich et al. [DWR22a] to automate the configuration process of an OPC UA FX demonstrator during runtime in response to varying production requirements. In addition, it was shown that the CSS model offers a way to map implemented OPC UA FX *FunctionalEntities* to Capabilities and Skills within the AAS. It also illustrates how the activities defined for the CSS model enable quick on-the-fly production planning and reconfiguration of the communication between process resources even when dealing with abstract production descriptions as inputs. To fully achieve the I4.0 adaptable and optimized manufacturing concepts further work is required to implement the combination of I4.0 technologies in an actual industrial context. For instance, it is required to implement generic mappings between different domain specific semantics of various I4.0 principles.

5 References

[DWR22a] Diedrich, C.; Werner, T.; Riedl, M.: Interaktion zwischen Steuerungen auf der Basis von OPC UA FX und deren Konfiguration durch Verwaltungsschalen. In: Automation 2022, Baden-Baden, S.19-30, 2022.

[OP23a] OPC Foundation News, https://opcfoundation.org/news/press-releases/the-opc-foundation-releases-the-opc-ua-field-exchange-uafx-specifications/, Accessed 5 Oct 2023.

[ID23a] IDTA e.V.: Specification of the Asset Administration Shell Part 1: Metamodel, 2023.

[OP22a] OPC Foundation: OPC Unified Architecture Field eXchange (UAFX) Part 81: UAFX Connecting Devices and Information Model 1.00.00, 2022.

[OP22b] OPC Foundation: OPC Unified Architecture Field eXchange (UAFX) Part 80: UAFX Overview and Concepts 1.00.00, 2022.

[VD20] VDI/VDE-GMA: VDI/VDE 2193, Blatt 2 - Sprache für I4.0-Komponenten - Interaktionsprotokoll für Ausschreibungsverfahren, VDI/VDE-Richtlinien, Beuth Verlag GmbH, 2020.

[BVZ15] BITKOM e.V.; VDMA e.V.; ZVEI e.V.: Umsetzungsstrategie Industrie 4.0. Ergebnisbericht der Plattform Industrie 4.0, 2015.

[Di22b] Diedrich, C. et.al.: Information Model for Capabilities, Skills & Services. Definition of terminology and proposal for a technology-independent information model for capabilities and skills in flexible manufacturing. Plattform Industrie 4.0, Berlin, 2022.

[ID23b] IDTA e.V. Content Hub AAS Submodell Templates, https://industrialdigitaltwin.org/content-hub/teilmodelle, Accessed 29 Sep 2023.

[Pl20] Plattform Industrie 4.0: Verwaltungsschale in der Praxis. Wie definiere ich Teilmodelle, beispielhafte Teilmodelle und Interaktion zwischen Verwaltungsschalen (Version 1.0), BMWi, Berlin, 2020.

[Gr22] Grunau, S. et.al.: The Implementation of Proactive Asset Administration Shells: Evaluation of Possibilities and Realization in an Order Driven Production. In Kommunikation und Bildverarbeitung in der Automation. Technologien für die intelligente Automation, vol. 14. Springer Vieweg, Berlin u.a., 2022.

[OP23b] OPC Foundation Field Level Communications (FLC) Initiative, https://opcfoundation.org/flc/, Accessed 5 Oct 2023.

http://dx.doi.org/10.25673/111642

Edgeshark: Einblick in die virtuelle Kommunikationswelt (nicht nur) von Containern

Ein (OpenSource) Beitrag zu einer konvergierenden IT/OT

Dr. Harald Albrecht[1]

Abstract: Nicht zuletzt mit aktuellen „vPLC"-Offerten gewinnt der Einsatz von IT-Technologien wie Docker und Kubernetes zum Virtualisieren und Orchestrieren der Industrieautomation an Fahrt. Die IT verfügt über bewährte Blaupausen zum Strukturieren von Applikationen und deren virtueller Kommunikation auch in größeren Container-Systemen. In der Praxis stoßen Anwender, Entwickler, Experten und Forschende aus Automatisierung und Industrie-Kommunikation immer wieder auf überraschendes oder unerwünschtes Systemverhalten. Umfangreiche Online-Dokumentation und heutige Kommandozeilen-Werkzeuge sind dabei sowohl Hilfe als auch Hürde, um die virtuellen Systeme und Netzwerke zu durchdringen und diagnostizieren: warum kommen keine Telegramme in der Anlage an, warum gehen keine Daten die die Cloud, …? In der eigenen mehrjährigen Praxis als „containerisierter Systemarchitekt" u.a. bei der Siemens „Industrial Edge" ist deshalb das anfangs interne Projekt „Edgeshark" entstanden und stetig weiterentwickelt worden. Siemens steuert nun das Projekt (MIT-Lizenz) zur Github OpenSource Community bei. In diesem Beitrag steht jedoch nicht das Werkzeug im alleinigen Mittelpunkt, sondern es soll vielmehr wie ein guter Ausbilder Interessierten einen einfachen Zugang zu verschiedenen Elementen der virtuellen (und auch realen) Kommunikation in Containersystemen ermöglichen. Der Zugang kann explorativ über eine Weboberfläche erfolgen, genauso kann Edgeshark auch per REST API einfach in neuen Aufgaben eingebunden werden. Weiterhin existiert eine Integration samt Live-Übertragung zum OpenSource-Werkzeug Wireshark.

Keywords: Software-Container, virtuelle Kommunikation, Diagnose, Wissenstransfer

1 „Wirrtuelle" Kommunikation?

Was kommt mit den IT-Technologien zur Virtualisierung und Orchestrierung in Form von Docker und Kubernetes auf Automation und Industrie-Kommunikation zu? Müssen wir Automatisierenden und „Industrie-Kommunizierenden" uns überhaupt zumindest in Teilen mit diesen Technologien auseinandersetzen oder können wir sie nicht „einfach nur benutzen"? Und lassen sich bisherige Erfahrungen mit virtuellen Netzwerken überhaupt in die virtuelle Welt der Container so einfach übertragen?

In der Praxis stoßen Beteiligte – egal, ob industrieller Anwender, Systemadmins, Entwickler, Netzwerker, Automatisierende, … – bald auf überraschendes oder auch unerwünschtes Systemverhalten. Zwar stehen den Betroffenen vielfältige Quellen offen,

[1] Siemens AG, DI CTO, Gleiwitzer Straße 555, 90475 Nürnberg, harald.albrecht@siemens.com

beispielsweise die Docker-eigene Dokumentation [DoDo] und die bekannte Frage-und-Antwort-Plattform „Stack Overflow“ [SO]. Daneben existiert auch ein schier unüberschaubarer Markt an kommerziellen Schulungen. Diese vielfältigen Angebote können jedoch nur eingeschränkt die Fragen beantworten: wie sieht die virtuelle Kommunikation *konkret in diesem Moment* in meinem (vermutlich) liebevoll und mühsam konfigurierten Automatisierungs-System wirklich aus? “Macht” das System das, was ich ihm als Konfiguration vorgegeben habe? Erschwerend kommt hinzu, dass virtuelle Kommunikation in Container-Systemen heute weitestgehend anderen Strukturen und Richtlinien folgt, als es bislang bei virtuellen Netzwerken wie beispielsweise der IEEE 802.1 VLANs Praxis ist.

Die zum Beantworten dieser Fragen heute verfügbaren Programmier-Schnittstellen und Werkzeuge sind zwar *prinzipiell* ausreichend. Sie erfordern jedoch ein sehr hohes Maß an Verständnis der unterlagerten mehreren Ebenen der Virtualisierungstechnologien, wie beispielsweise die sog. „namespaces“ des Linux-Kernels [Namsp7] sowie deren spezielle Nutzung für Container.

Gängige und beliebte Werkzeuge zur Netzwerkdiagnose – insbesondere Wireshark [WS] – funktionieren zwar grundsätzlich auch mit Containern, sind jedoch ihrer nicht gewahr. Letztlich sollen sich Anwender auch nicht mit OS-eigenen Referenzen wie „/proc/12345/ns/net“ befassen müssen, die nur über komplizierte Kommandozeilen-Ketten in speziellen Terminalsitzungen direkt in Spezial-Container hinein ausgeführt werden können („war das Argument nun „-t“ samt Prozessnummer oder doch „--net“ und eine procfs-Referenz?“). Außerdem kann und will nicht jeder industrielle Anwender es erlauben, zur Diagnose seiner Container-unterstützten Automatisierungsanwendung sich tief in sein laufendes System auf eine Kommandozeile zu begeben.

2 OpenSource-Projekt Edgeshark

Seit Jahren mangelt es unserer Beobachtung nach an Werkzeugen, um die (virtuelle) Kommunikation in Container-Systemen auch für Einsteiger einfach begreif- und diagnostizierbar zu gestalten. Weder der kommerzielle Markt noch die OpenSource Gemeinschaft bieten derartige Werkzeuge an. Das kommerzielle Angebot „Cloudshark“ [CS] adressiert im Gegensatz zu dem hier verfolgten Ziel speziell das *Cloud-gestützte Auswerten von Paketaufzeichnungen*, ohne dass vor Ort noch ein Wireshark-Programm installiert werden muss. Wie allerdings die Paketaufzeichnung erfolgte, bleibt in diesem speziellen Fall außer Acht.

Anwender, Entwickler und Forschende müssen deshalb üblicherweise und bei entsprechendem Leidensdruck mit verschiedenen Hilfsmitteln sowie länglichen und fehlerträchtigen Kommandozeilen versuchen, mehr schlecht als Recht ihre Probleme bei der Container-Kommunikation zu diagnostizieren. In Folge bauten in der Vergangenheit Entwickler dafür vorübergehend das Wireshark-Programm mit in ihre Anwendungs-Container ein.

Nicht zuletzt für System- und Softwarearchitekten war es zudem eine ständig wiederkehrende Herausforderung, Anwendern ein glaubhaftes und nachvollziehbares Bild der zumeist vielgestaltigen Kommunikation in Docker-Systemen zu vermitteln.

In diesem mehr als unbefriedigenden Zustand entstand in der Siemens AG zunächst intern das Projekt „Edgeshark“. Dieses trägt Siemens nun als OpenSource-Projekt mit MIT-Lizenz auf Github [ES] zur OpenSource Community bei – nicht zuletzt, um damit Automatisierungsanwendern aus Industrie und Forschung beim sanften Einstieg in die „containerisierte Automatisierung“ zu unterstützen.

Der Projektname „Edgeshark“ ist der Historie verdankt, dass die meisten bisherigen Anwender das Werkzeug zuerst im Rahmen der Siemens Industrial Edge-Plattform [SIE] kennengelernt und (erfolgreich) einsetzen konnten. Edgeshark ist jedoch nicht auf die Industrial Edge-Plattform beschränkt, sondern kann auf jedem (Docker) Container-Host eingesetzt werden. Weiterhin erkennt Edgeshark die Besonderheiten lokaler Kubernetes-in-Docker-Installationen („KinD“, [KinD]), wie Pods und KinD-Knoten-Container, ohne dass dazu ein Zugriff auf das Kubernetes-API nötig ist. Aktuell umfasst das Edgeshark-Projekt rund 23.500 Zeilen Quellcode und 13.000 Zeilen Dokumentation, sowohl als Endanwender- als auch Programmdokumentation.

3 Architektur

Den Kern von Edgeshark bilden die beiden (containerisierten) Dienste einerseits für das Auffinden der virtuellen Netzwerkstruktur und deren Zustands („ghostwire“ genannt) und andererseits das Erfassen und Streamen von Netzwerkpaketen („packetflix“), siehe die folgende Abb. 1. Die Ausgestaltung als einfach konsumierbare Dienste erlaubt eine möglichst vielfältige Nutzung (und zugleich Wiederverwendung) in unterschiedlichen Anwendungen. Zudem werden Einschränkung auf ein bestimmtes Programmiersprachen-Ökosystem in der Nutzung vermieden. Nicht zuletzt folgt das Projekt hier den in der IT und den Cloud-Technologien bewährten Mustern.

Auf die Edgeshark-Dienste kann per REST- sowie Websocket-APIs auf vielfältige Weise zugegriffen werden:

- interaktiv von einem gängigen Webbrowser,
- interaktiv oder auch programmiert vom OpenSource-Werkzeug Wireshark zur Analyse des Netzwerkverkehrs von/zu Containern.
- von eigenen (neuartigen) Applikationen, gegebenenfalls unter Zuhilfenahme der “csharg” Go-Bibliothek.

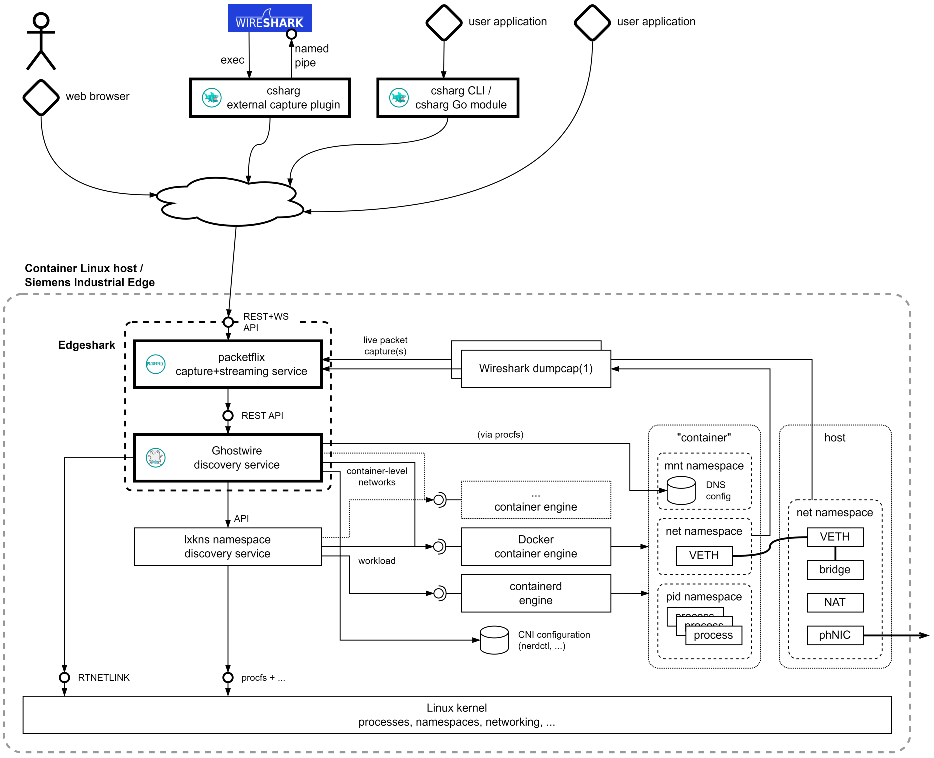


Abb. 1: Die Edgeshark-Architektur im Überblick.

Ein wichtiger Vorteil ist, dass keinerlei Änderungen an anderen Applikationen bzw. Containern vorgenommen werden müssen: die Edgeshark-Dienste „wandern“ sozusagen im Rahmen der ihnen erteilten Rechte durch das Linux-System mit seinen Containern, um die erforderlichen Informationen zu sammeln.

Die Edgeshark-Dienste kapseln hierbei die inhärente Komplexität beim Auffinden der über Container und möglicherweise verschiedene Container-Engines verteilten Netzwerk- und Konfigurationsinformationen. Anwender (und Anwendungen) können sich somit auf ihre eigentlichen Ziele wie Diagnose und ein besseres Verständnis der virtuellen Kommunikationsstrukturen konzentrieren.

Der Ghostwire-Dienst verfügt über mehrere Erweiterungspunkte („extension points“), um beispielsweise einfach APIs weiterer Container-Engines nachzurüsten oder neue Orchestrierungs-Markierungen zu nutzen. Eine experimentelle Unterstützung für podman [Podman] ist vorhanden.

Der Ghostwire-Dienst unterstützt nur Linux. Eine Unterstützung der Windows-Kommunikations-Architektur ist aufgrund der unklaren Informationslage sowie als in der Regel untypisches Einsatzgebiet von Containern nicht vorgesehen. Ebenso wird auch WSL2 bislang nicht unterstützt. Pull Requests sind herzlich willkommen und sollten freundlicherweise zuvor mit dem Edgeshark-Projekt besprochen werden.

4 Ghostwire Discovery-Dienst

Der Ghostwire-Dienst – so benannt in Anspielung auf die virtuellen „Leitungen“ zwischen Containern – ist für das Auffinden von virtuellen und realen Netzwerk-Schnittstellen zuständig, sowie deren „Verschaltungstopologie“, siehe auch Abb. 2 mit dem Datenmodell.

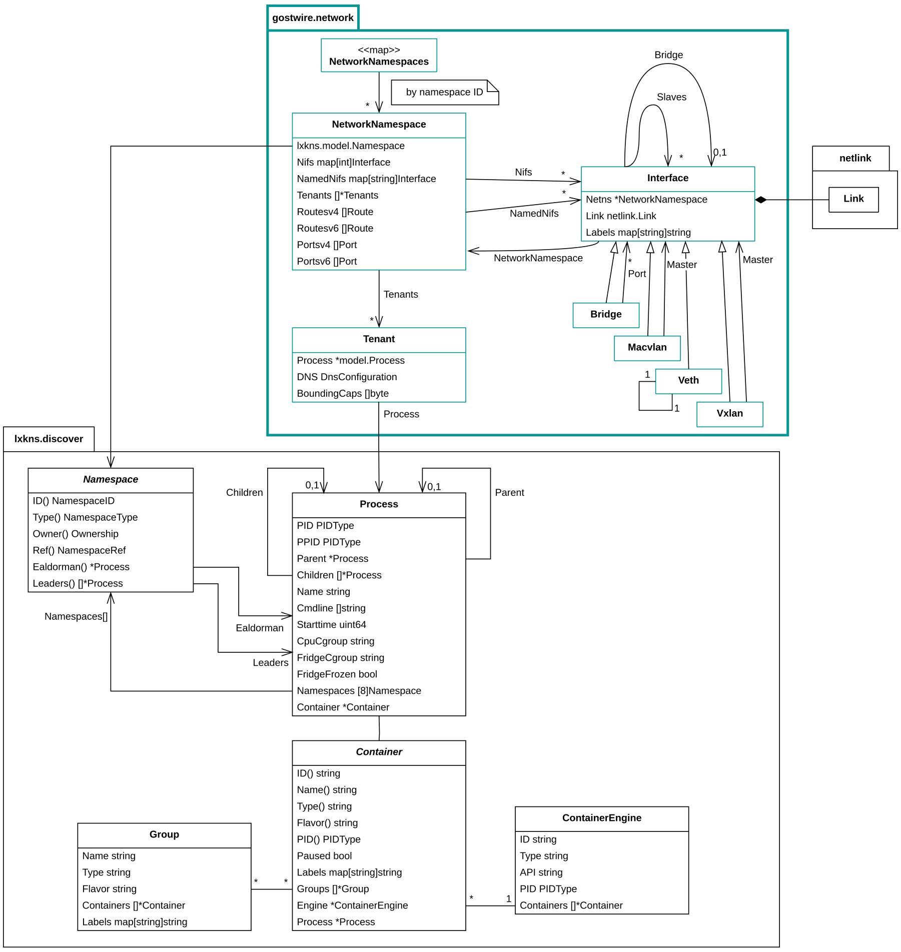


Abb. 2: Das Datenmodell.

Dabei greift der Dienst auf einen (integrierten) Erkundungsdienst für die sogenannten Linux-Kernel „namespaces“ [Namsp7] sowie deren Beziehungen zu Prozessen und Containern zurück. Auf der Basis dieser grundlegenden Informationen sammelt Ghostwire dann verschiedene Netzwerk- und Kommunikations-spezifische Daten und stellt diese als JSON-Datenmodell an seinem REST API bereit:

- reale und virtuelle Netzwerkschnittstellen und deren topologische Beziehungen (wie beispielsweise paarweise „VETH“-Netzwerkkarten).
- IP-Adress- und Routen-Konfiguration der verschiedenen (virtuellen) IP-Stacks sowie deren Zuordnung zu Containern, Pods, …
- die Konfiguration der DNS „stub resolver“ („DNS-Clients“ im umgänglichen Sprachgebrauch) innerhalb der Container.
- geöffnete TCP- und UDP-Port sowie die diese Ports bedienenden Prozesse (sowie entsprechend darauf aufsetzende „höhere“ Konstrukte wie Container, Pod, …)
- Weiterleitungen von TCP- und UDP-Ports an andere Ports.
- Container-Details zu den erteilten Berechtigungen („capabilities“).

5 Auf Erkundung

Die Edgeshark-Weboberfläche basiert rein auf der am REST API als JSON-Modell bereitgestellten Daten. Technisch werden die JSON-Daten hierbei erst im Browser selbst in einer React-Web-Anwendung [React] in ihre HTML-Darstellung überführt. Damit lassen sich einfach und bequem spezialisierte Ansichten nur auf Teile des Informationsmodells umsetzen.

So zeigt die Weboberfläche zu Beginn zunächst eine topologische Ansicht der “virtuellen Leitungen” zwischen den Containern, siehe Abb. 3. Hier sind auch die bei Containern für die Automatisierung vermehrt anzutreffenden verschiedenen Kommunikationsverbindungen erkennbar, wie in Abb. 3 rechts gesondert gezeigt. So besitzt der hier gezeigte Anwendungscontainer nicht nur die insbesondere bei IT-Anwendungen übliche Anbindung an eine virtuelle Bridge (virtueller Switch) im Container-Host (Schnittstelle „eth0“). Zusätzlich verfügt diese Anwendung noch über eine Schnittstelle „eth1“ (vom Typ „MACVLAN“), die die direkte Kommunikation auf Ebene von Ethernet-Telegrammen beispielsweise mit Feldgeräten ermöglicht. Nicht jede Automatisierungs-Anwendung benötigt diese doppelte Anbindung, sie wird in der Regel immer dann benötigt, wenn nicht-IP-basierte Feldgeräte-Protokolle direkt benutzt werden müssen.

Durch Antippen oder Anklicken von Containern und Netzwerk-Elementen können Anwender einfach in eine stärker fokussierende Kommunikationsansicht mit höherem Detailgrad wechseln. In der in Figur 4 beispielhaft gezeigten Ansicht können Anwender unter anderem rasch erkennen, ob und welche Kommunikationsendpunkte in ihren Containern konkret in diesem Moment vorhanden sind.

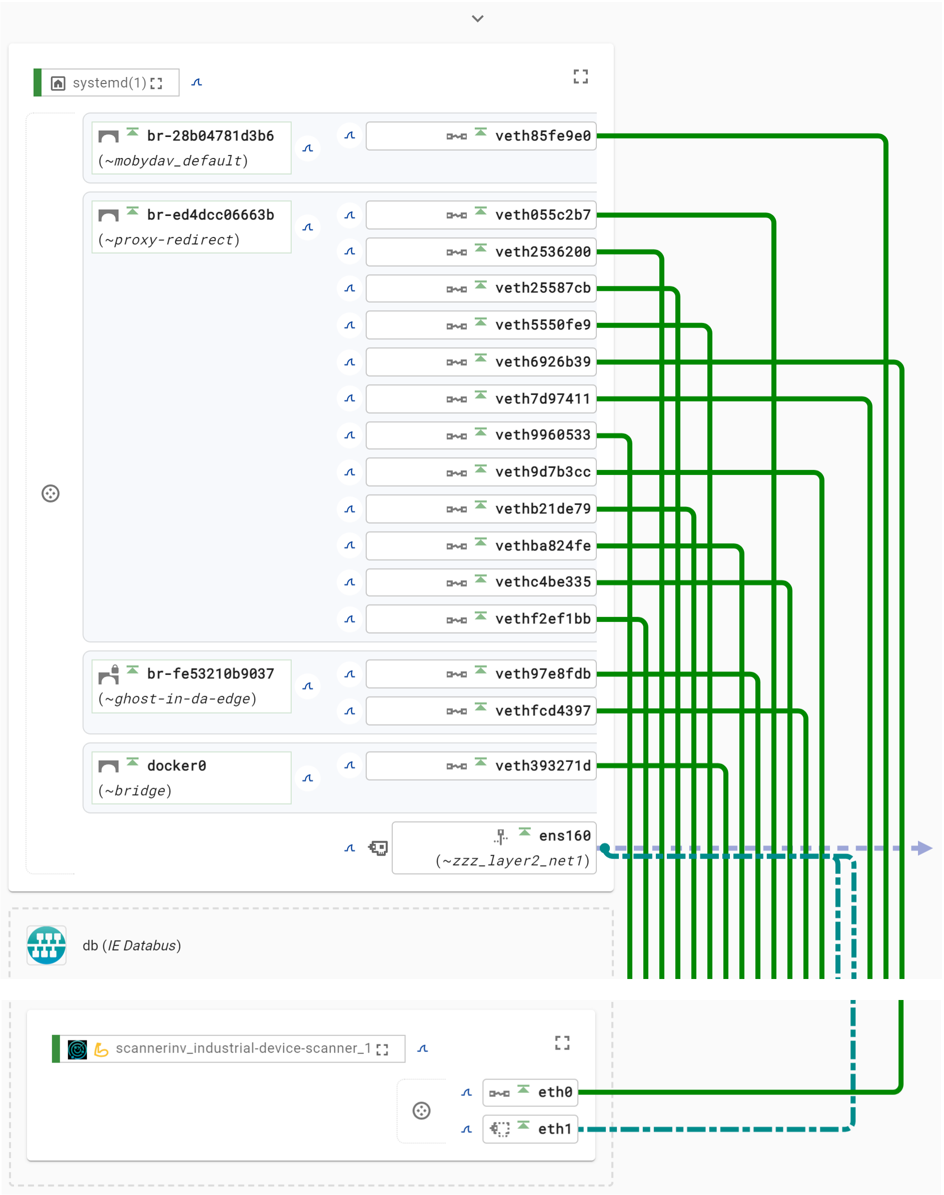
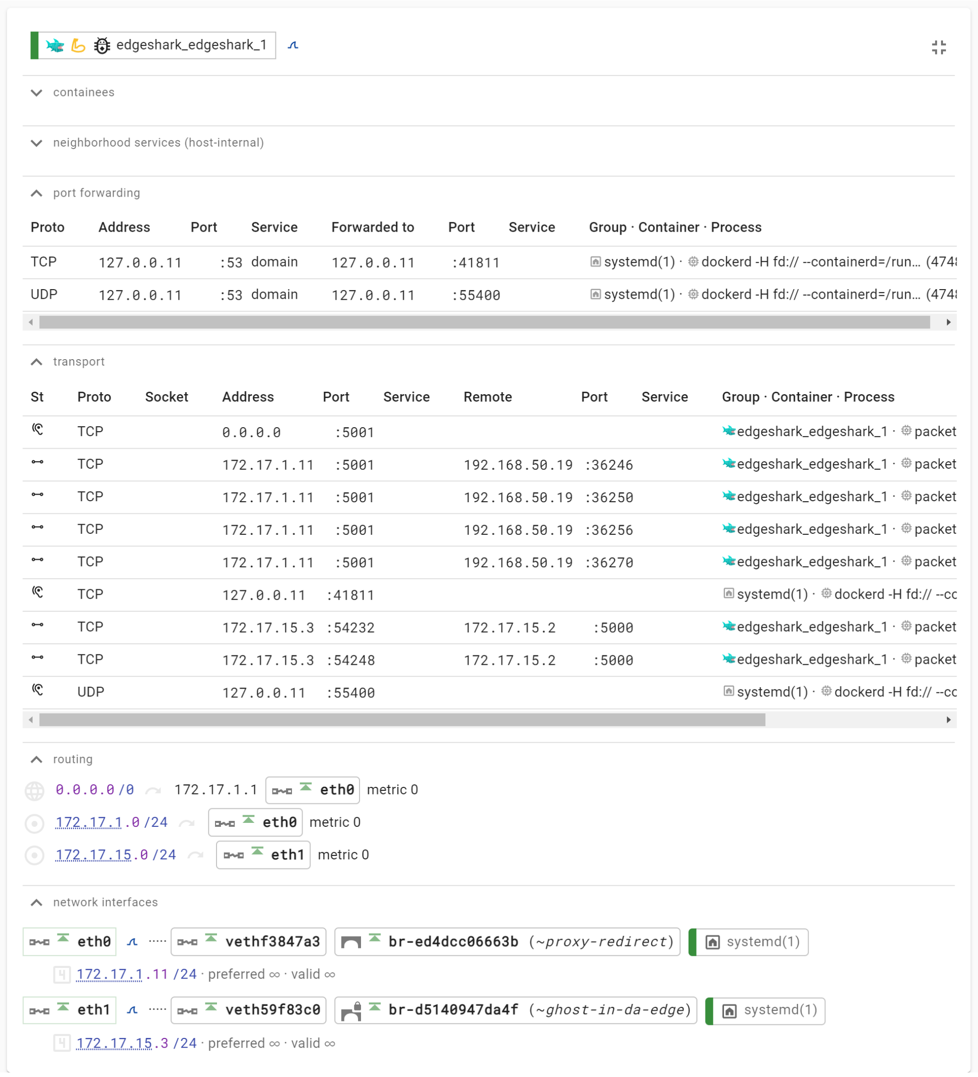


Abb. 3: Einfacher visueller Einstieg inklusive Navigationshilfen in die teilweise komplexen Kommunikationsbeziehungen in Container-Systemen.



edgeshark_edgeshark_1

containees

neighborhood services (host-internal)

port forwarding

Proto	Address	Port	Service	Forwarded to	Port	Service	Group · Container · Process
TCP	127.0.0.11	:53	domain	127.0.0.11	:41811		systemd(1) · dockerd -H fd:// --containerd=/run… (474
UDP	127.0.0.11	:53	domain	127.0.0.11	:55400		systemd(1) · dockerd -H fd:// --containerd=/run… (474

transport

St	Proto	Socket	Address	Port	Service	Remote	Port	Service	Group · Container · Process
👂	TCP		0.0.0.0	:5001					edgeshark_edgeshark_1 · packet
⊶	TCP		172.17.1.11	:5001		192.168.50.19	:36246		edgeshark_edgeshark_1 · packet
⊶	TCP		172.17.1.11	:5001		192.168.50.19	:36250		edgeshark_edgeshark_1 · packet
⊶	TCP		172.17.1.11	:5001		192.168.50.19	:36256		edgeshark_edgeshark_1 · packet
⊶	TCP		172.17.1.11	:5001		192.168.50.19	:36270		edgeshark_edgeshark_1 · packet
👂	TCP		127.0.0.11	:41811					systemd(1) · dockerd -H fd:// --cc
⊶	TCP		172.17.15.3	:54232		172.17.15.2	:5000		edgeshark_edgeshark_1 · packet
⊶	TCP		172.17.15.3	:54248		172.17.15.2	:5000		edgeshark_edgeshark_1 · packet
👂	UDP		127.0.0.11	:55400					systemd(1) · dockerd -H fd:// --cc

routing

0.0.0.0/0 172.17.1.1 eth0 metric 0

172.17.1.0/24 eth0 metric 0

172.17.15.0/24 eth1 metric 0

network interfaces

eth0 ····· vethf3847a3 br-ed4dcc06663b (~*proxy-redirect*) systemd(1)

172.17.1.11/24 · preferred ∞ · valid ∞

eth1 ····· veth59f83c0 br-d5140947da4f (~*ghost-in-da-edge*) systemd(1)

172.17.15.3/24 · preferred ∞ · valid ∞

Abb. 4: Detaillierter Einstieg in die aktuell wirksame Kommunikationskonfiguration eines Containers.

Nachdem diese Informationen dem JSON-Datenhaushalt des Discovery-API-Endpunktes entstammen, können derartige Informationen auch automatisiert ausgewertet werden – beispielsweise zur Prüfung von Containern hinsichtlich ihrer exponierten Endpunkte, zum Überwachen bestimmter Kommunikationsflüsse sowie vieles mehr.

Im Gegensatz zu den bekannten CLI-Werkzeugen wie beispielsweise “netstat” [netstat8] und “show socket” [Shows8] enthält das Ghostwire-Informationsmodell nicht nur alle Endpunkte über alle virtuellen Netzwerkstacks eines Hosts hinweg, sondern auch deren Zuordnung zu Containern und ggf. Composer-Projekten, Kubernetes-Pods, und dergleichen.

6 Packetflix Capture Streaming-Dienst

Von etlichen Heimroutern – wie beispielsweise den „Fritz!box“en – ist die Funktion bekannt, auf Wunsch einen Mitschnitt des Netzwerkverkehrs aufzeichnen und dann später auf den eigenen Rechner herunterladen zu können. Nun ist eine derartige Download-Funktion nicht grundsätzlich falsch, erscheint jedoch in ihrer konkreten Ausführung etwas aus der Zeit gefallen.

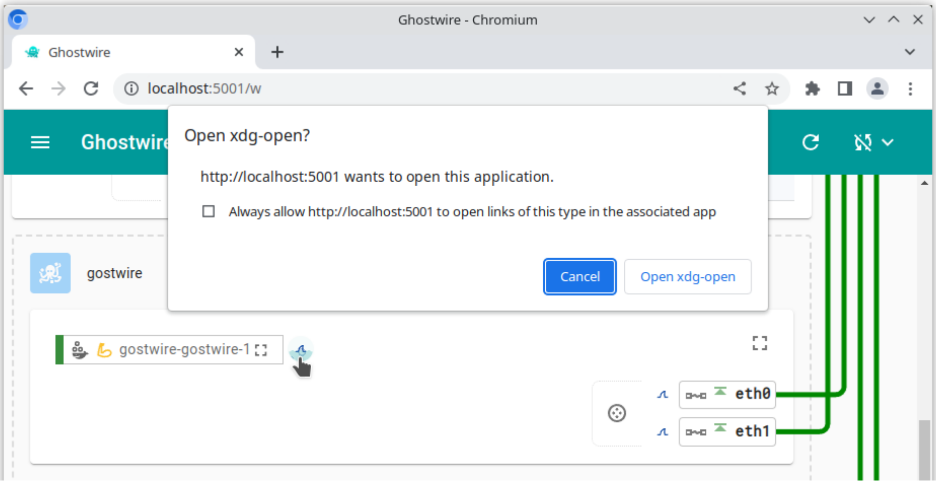
Viele Menschen fahren heute nicht mehr umständlich in eine Videothek [VT], um sich dort audiovisuelle Medien wie Blu-rays oder gar VHS-Kassetten [VHSKa] auszuleihen, zuhause anzusehen und dann (eventuell) wieder zurückzubringen. Vielmehr ist es inzwischen gängig, audiovisuelle Medien bequem zuhause oder jederzeit unterwegs per „streaming“ zu konsumieren, wie Netflix, Spotify und YouTube zeigen (um nur eine kleine Auswahl an Streaming-Diensten beispielhaft zu nennen).

Der „Packetflix“-Dienst überträgt somit konsequenterweise das heutige Streaming-Modell auf das Mitschneiden und Analysieren von Netzwerkverkehr insbesondere bei Containern. Eine einfache Websocket-API nimmt dabei beispielsweise einfach nur den Namen des „anzuzapfenden“ Containers entgegen und beginnt daraufhin, die mit geschnittenen Netzwerkpakete an die Gegenstelle zu übertragen.

Intern kümmert sich der Packetflix-Dienst darum, den Container-Namen unter Zuhilfenahme des Ghostwire-Dienstes in die benötigten Ressourcen-Referenzen auf der Ebene der Linux-API abzubilden. Damit startet er danach das Wireshark-Werkzeug „dumpcap“ [WSdc], wobei er es zudem in den Kontext des Netzwerkstacks des betreffenden Containers einblendet. Wie in Abb. 1 angedeutet, wird im Dienst das vergleichsweise schlanke dumpcap verwendet, nicht jedoch ein vollumfänglicher Wireshark selbst. Tatsächlich nutzt Wireshark dumpcap selbst in verschiedenen Anwendungsszenarien als externe Proben-Entnahmestelle.

Als besondere Funktion schreibt Packetflix zusätzliche Informationen in den Paketdatenstrom hinein, die den untersuchten Container eindeutig identifizieren (Name, ID, …). Damit wird das heute übliche Problem vermieden, dass die gängigen Werkzeuge dumpcap, tcpdump [Tcpdmp1] und Wireshark keinerlei Wissen über Container besitzen und es somit ansonsten schwerfallen würde, nachträglich die Quelle der Paketdaten zu identifizieren: ohne Nachhilfe hinterlegt dumpcap wenig aussagekräftige Informationen wie Schnittstelle „eth0“ des Hosts „52c001de42bef1“.

Der Zugriff auf den Packetflix-Dienst erfolgt im einfachsten Fall aus der Weboberfläche heraus, kann aber auch über Kommandozeilenprogramme erfolgen (Wireshark einbezogen). In der Weboberfläche stehen hierfür Schaltflächen bereit, die eine stilisierte „Haifischflosse“ in Anspielung auf Wireshark zeigen, siehe auch Abb. 5.



Open xdg-open?

http://localhost:5001 wants to open this application.

☐ Always allow http://localhost:5001 to open links of this type in the associated app

Abb. 5: Einfacher Wechsel zum Wireshark-Tool zwecks Live-Analyse von Paketen.

Antippen oder Anklicken startet nach Rückfrage beim Anwender den lokal auf dem eigenen Rechner installierten Wireshark. Dieser verbindet sich daraufhin aufgrund der ihm übermittelten „Kontaktdaten“ mit dem Packetflix-Dienst und zeigt danach ohne weiteres Zutun die live übertragenden Netzwerk-Pakete an. Der Packetflix-Dienst besitzt hierbei selbst keine Wireshark-Komponente.

Die Verbindung zwischen Wireshark und dem Packetflix-Dienst stellt ein sogenanntes „external capture plugin“ [Extcap] her; hierbei handelt es sich um ein eigenständiges kleines Binärprogramm, das von Wireshark automatisch erkannt und bei Bedarf herangezogen wird.

7 Zusammenfassung

Mit dem OpenSource-Projekt „Edgeshark“ [ES] steht ein einfach nutzbares und erweiterbares Werkzeug zur Verfügung, um die virtuelle Kommunikationsstrukturen von Containern darzustellen, zu diagnostizieren und auszuwerten. Edgeshark soll dabei aber lediglich ein Einstiegspunkt sein: weitere, neuartige Anwendungen können beispielsweise direkt auf den Datenhaushalt in JSON-Kodierung über ein REST API zugreifen. Die Codebasis steht unter der MIT-Lizenz und nutzt aktuell die weit verbreiteten Sprach- und Ablaufplattformen Go [Go] sowie React [React] in Verbindung mit Typescript.

Zukünftige Container-APIs können bei Bedarf über Erweiterungspunkte zugerüstet werden.

Literaturverzeichnis

[Caps7] Capabilities(7) man page, https://man7.org/linux/man-pages/man7/capabilities.7.html, Stand 02.06.2023.

[CS] Cloudshark, https://www.qacafe.com/analysis-tools/cloudshark/qa-cloudshark-personal-saas/, Stand 02.06.2023.

[DoDo] Docker Documentation, https://docs.docker.com/, Stand 02.06.2023.

[ES] Edgeshark Project auf Github, https://github.com/siemens/edgeshark.

[Extcap] extcap(4) manual page, https://www.wireshark.org/docs/man-pages/extcap.html, Stand 02.06.2023.

[Go] https://go.dev/, Stand 02.06.2023.

[KinD] Kubernetes-in-Docker, https://github.com/kubernetes-sigs/kind, Stand 02.06.2023.

[Namsp7] namespaces(7) man page, https://man7.org/linux/man-pages/man7/namespaces.7.html, Stand 02.06.2023.

[Netstat8] netstat(8) man page, https://man7.org/linux/man-pages/man8/netstat.8.html, Stand 02.06.2023.

[Podman] https://podman.io/, Stand 02.06.2023.

[React] https://react.dev/, Stand 02.06.2023.

[Shows8] show sockets manpage, https://man7.org/linux/man-pages/man8/ss.8.html, Stand 02.06.2023.

[SIE] Siemens Industrial Edge, https://www.siemens.com/de/de/produkte/automatisierung/themenfelder/industrial-edge.html, Stand 02.06.2023.

[SO] Stack Overflow, https://stackoverflow.com/, Stand 02.06.2023.

[Tcpdmp1] tcpdump(1) man page, https://man7.org/linux/man-pages/man1/tcpdump.1.html, Stand 02.06.2023.

[VHSKa] Video Home System Kassetten, https://de.wikipedia.org/wiki/Video_Home_System#Kassetten, Stand 02.06.2023.

[VT] Videothek (Wikipedia), https://de.wikipedia.org/wiki/Videothek, Stand 02.06.2023.

[WS] Wireshark Organisation, https://www.wireshark.org, Stand 02.06.2023.

[WSdc] Wireshark Organisation, https://www.wireshark.org/docs/man-pages/dumpcap.html, Stand 02.06.2023.

http://dx.doi.org/10.25673/111636

Compliance with Industrial Security Standards by Implementing Remote Attestation

Florian Kohnhäuser and Sören Finster

ABB AG, Corporate Research, Ladenburg, Germany
{florian.kohnhaeuser,soeren.finster}@de.abb.com

Abstract. To mitigate the risk of cyber threats on industrial systems, security standards are currently emerging and providing an important framework to ensure security. While security standards define desired security outcomes, they often lack specific implementation strategies. This leads to the application of inconsistent or inadequate security measures. In this work, we focus on a novel security measure called remote attestation, which is capable of verifying the authenticity and integrity of remote devices and systems. We analyze remote attestation and its relation to the industrial security standards IEC 62443, NERC CIP, NIST SP 800, ISO/IEC 27002, and PCI DSS. In detail, we map remote attestation to requirements of the analyzed security standards, highlighting the degree to which these requirements can be fulfilled by remote attestation. The results demonstrate that remote attestation is highly relevant to the analyzed security standards and offers both technical mitigation of cyber threats as well as compliance with well-established security standards.

1 Introduction

With the increasing connectivity and openness of industrial systems, security has become a crucial requirement. To address the need for security, regulations and standards are becoming more and more important. They serve as a framework that organizations, products, and services must satisfy to mitigate security threats. While security standards specify the desired outcome in terms of security requirements, they typically lack instructions and implementation strategies to reach that outcome. Some security requirements, such as secure communication between industrial components, are comparatively easy to implement, as all modern communication protocols have built-in security modes. Yet, other requirements, such as secure auditing, logging, and device integrity, are much harder to fulfill. This gap between the definition and implementation of security requirements can result in inconsistent or inadequate security measures, leaving systems and data vulnerable to potential threats and compromises.

Remote attestation is an emerging security technology that addresses the growing concerns on the trustworthiness and integrity of computer systems [4]. Attestation allows to verify the integrity of the hardware, software, and configuration of a remote system. During verification, tampering of the remote device is detected, which provides a strong defense against various threats, including malware and unauthorized modifications. However, remote attestation is not yet well-understood regarding its provided security capabilities and compliance with security standards, which hampers its adoption in practice.

In this work, we analyze remote attestation regarding its compliance with industrial security standards. To this end, we first provide an introduction into remote attestation and security compliance (Chapter 2). Next, the provided security properties of remote attestation are mapped to the well-established security standards IEC 62443 [6], NERC CIP [7], NIST SP 800 [8], ISOIEC 27002 [9], and PCI DSS [10] (Chapter 3). The mapping lists specific requirements of the analyzed security standards that can be fulfilled by implementing remote attestation. In specific, it is investigated to which degree the requirements can be fulfilled and how potential gaps can be addressed. It is shown that remote attestation is highly relevant to the analyzed security standards (Chapter 4).

2 Background

2.1 Remote Attestation

Remote ATtestation procedureS (RATS) [5] are a security measure to verify the integrity and trustworthiness of a remote device or system. Although RATS have been proposed two decades ago [4], they recently gained attention due to the availability of secure hardware, open source implementations, and standardization efforts [5]. RATS functioning relies on secure hardware that allows a remote system, commonly referred to as the verifier, to assess the integrity of another system, known as the prover. This assessment involves generating a unique cryptographic signature or measurement of the prover's software, hardware, and configuration. The verifier compares this measurement against a predefined reference, or known-good configuration, to determine whether the prover has been compromised or altered in any way. The primary purpose of RATS are to establish trust in remote devices or systems, ensuring that they have not been tampered with. Their security goals include detecting unauthorized modifications, protecting against malware, and providing evidence of the remote system's trustworthiness, thereby enhancing security in scenarios like remote device management, secure bootstrapping, and establishing secure communication.

2.2 Security Compliance

Security regulations and standards serve as a framework that organizations, products, and services must adhere to, in order to ensure the protection of sensitive data, maintain customer trust, and mitigate the risk of cyber threats. By complying with these standards, actors demonstrate their commitment to data security, privacy, and integrity by establishing robust security controls, implementing best practices, and undergoing regular assessments to identify and address vulnerabilities. Security regulations and standards exist at national and international level, such as the German BSI KRITIS regulation [1] and European

Cyber Resilience Act [2]. Although they exist for various industries, this work focuses on the standards regarding the electric utility industry (NERC CIP), payment card industry (PCI DSS), organizational processes (ISO 27002), government agencies (NIST SP800), and industrial control industry (ISO 62443).

3 Evaluation

In this section, we map Remote ATtestation procedureS (RATS) [5] to the requirements of the industrial security standards NERC CIP, NIST SP 800, ISO/IEC 27002, PCI DSS, and IEC 62443. In case a requirement can be fulfilled by implementing RATS, we assess whether it has a low, medium, or high relevance to RATS, and describe its relevance, including gaps, in detail.

3.1 NERC CIP

North American Electric Reliability Corporation Critical Infrastructure Protections relates to the preparedness and response to serious incidents that involve the critical infrastructure assets in the electrical power grid [7].

CIP-005 1.5 Malicious communication (low relevance)

Requirement: Have one or more methods for detecting known or suspected malicious communications for both inbound and outbound communications.

Relevance: While RATS do not monitor communication, outbound malicious communication must originate in a local process. RATS provide means for observing local processes and detecting unwanted changes in them. Thus, the root cause for outbound malicious communication can be detected using RATS.

CIP-007 2.1-4 Patch management (medium relevance)

Requirement: A patch management process includes tracking, evaluating, and installing cybersecurity patches for relevant cyber assets. This involves identifying sources for patch releases and conducting evaluations every 35 days. After evaluation, one of these actions must be taken: (i) apply the patches, (ii) create a dated mitigation plan, or (iii) revise an existing mitigation plan.

Relevance: While this patch management process does not require for checking if patches have been applied, its intention shows that all relevant security patches should be installed. With RATS, periodic checking of the software that is currently in use is done. This automatically provides means of checking that software is on the desired patch level.

CIP-007 3.1 Detect malicious code (high relevance)

Requirement: Deploy methods to deter, detect, or prevent malicious code.

Relevance: This is a direct function of RATS with the capability to even detect malicious code on already compromised systems. This is a feature that the suggested measures (e.g., antivirus) do not provide.

CIP-007 3.2 Malicious code mitigation (high relevance)

Requirement: Mitigate the threat of detected malicious code.

Relevance: RATS provide timely and automatic detection and can initiate manual processes and automatic measures to mitigate the threat.

CIP-007 3.3 Up to date measures (high relevance)

Requirement: For methods identified in part 3.1 that use signatures or patterns to detect malicious code, have a process to update of the signatures or patterns.

Relevance: This requirement is relevant in two ways. First, RATS must be provided with information about the known-good software state. This corresponds to the signatures or patterns described in this requirement. Second, RATS can measure the installed signatures or patterns on a system and thus provide checking if signatures or patterns are installed correctly.

CIP-008 1.1 and 1.4 Incident response (low relevance)

Requirement: Establish one or more processes to identify, classify, and respond to cyber security incidents.

Relevance: RATS verification services can play a supporting role in incident response for quickly identifying and analyzing security incidents.

CIP-010 2.1 Configuration change detection (high relevance)

Requirement: Monitor at least once every 35 calendar days for changes to the baseline configuration. Document and investigate detected unauthorized changes.

Relevance: As configuration can be included in RATS reports, the automatic and regular monitoring of changes is a direct result of implementing RATS.

CIP-010 3.1-4 Vulnerability assessment (high relevance)

Requirement: Conduct a paper or active vulnerability assessment every 15 months. Additionally, perform an active vulnerability assessment in a test environment every 36 months, mimicking the production environment's configuration, and document the results along with any differences from the test environment.

Relevance: RATS can cover large parts of vulnerability assessments and reduce manual effort. This is very relevant, as assessments need to be done regularly.

3.2 PCI DSS V3.1.2

Payment Card Industry (PCI) Data Security Standard (DSS) is a set of security requirements and best practices designed to protect payment card data and prevent data breaches within organizations that handle credit card transactions.

Do not use vendor-supplied defaults (2) (low relevance)

Requirement: Do not use vendor-supplied defaults for system passwords and other security parameters.

Relevance: RATS can help ensure that such defaults are not present on target systems, if the corresponding databases (e.g., password database) is checked against the whitelisted version by the verifier.

Malware and anti-virus (5) (high relevance)

Requirement: Protect all systems against malware and regularly update anti-virus software or programs.

Relevance: Detection and escalation of unwanted software (e.g., malware) is a core functionality of RATS. Additionally, RATS can protect against missed updates of anti-virus software by checking against the whitelisted current version.

Protection from known vulnerabilities (6.2) (medium relevance)

Requirement: Ensure that all system components and software are protected from known vulnerabilities by installing vendor-supplied security patches. Install critical security patches within one month of release.

Relevance: RATS provide a constant monitoring of the installed software and therefore helps ensuring that all software is free from known vulnerabilities.

Audit trails (10.2) (low relevance)

Requirement: Implement automated audit trails for all system components to reconstruct the following events: (i) use of and changes to identification and authentication mechanisms, and (ii) creation and deletion of system-level objects.

Relevance: Some of the audit trails can be fulfilled by applying RATS. Especially changes to system-level objects or authentication databases will show up in RATS reports and further actions can then be initiated by the verifier.

Deploy change-detection (11.5) (high relevance)

Requirement: Deploy a change-detection mechanism (e.g., file-integrity monitoring tools) to alert personnel to unauthorized modification of critical system files, configuration files, or content files; and configure the software to perform critical file comparisons at least weekly.

Relevance: This is a direct requirement for the service RATS provide securely.

3.3 ISO/IEC 27002

ISO/IEC 27002 is an international standard that provides guidelines and best practices for information security management, helping organizations establish and maintain effective security controls and risk management processes [9].

Asset Management (8.1.1) (low relevance)

Requirement: Assets associated with information and information processing should be identified and an inventory of these assets should be maintained.

Relevance: Especially with the further clarification in mind, that prescribes asset inventory to be “accurate, up to date, consistent and aligned with other inventories”, RATS can provide a significant portion of this requirement.

User access provisioning (9.2.2) (high relevance)

Requirement: A formal user access provisioning process should be implemented to assign or revoke access rights for all user types to all systems and services.

Relevance: User access rights can be included in RATS reports. This provides a timely checking current user access rights compared with desired access rights. Especially with the further clarification to ”periodically reviewing access rights with owners of the information systems or services”, RATS can provide at least a source for the needed information.

Use of privileged utility programs (9.4.4) (high relevance)

Requirement: The use of utility programs that might be capable of overriding system and application controls should be restricted and tightly controlled.

Relevance: RATS provide the desired tight control of the usage of all programs, especially privileged utility programs.

Controls against malware (12.2.1) (high relevance)

Requirement: Detection, prevention and recovery controls to protect against malware should be implemented, combined with appropriate user awareness.

Relevance: Detection and escalation of unwanted software (e.g., malware) is a core functionality of RATS. The implementation of RATS and their integration with SIEM systems therefore provide a large part of this requirement.

Installation of software on operational systems (12.5.1) (low relevance)

Requirement: Procedures should be implemented to control the installation of software on operational systems.

Relevance: While RATS typically do not control the installation of software, implementation guidance point f) “an audit log should be maintained of all updates to operational program libraries;”, is part of RATS functionality.

Management of technical vulnerabilities (12.6.1) (high relevance)

Requirement: Information about technical vulnerabilities of information systems being used should be obtained in a timely fashion. The organization’s exposure to such vulnerabilities should be evaluated and appropriate measures should be taken to address the associated risk.

Relevance: Since RATS can provide an inventory of the used software, the evaluation of the exposure to known technical vulnerabilities is easy to implement by simply checking the inventory of used software on the verifier.

Secure system engineering principles (14.2.5) (medium relevance)

Requirement: Principles for engineering secure systems should be established, documented, maintained and applied to any information system.

Relevance: RATS can be a crucial part of secure systems engineering.

Responsibilities and procedures (16.1.1) (medium relevance)

Requirement: Management responsibilities and procedures should be established to ensure a quick, effective and orderly response to information security incidents.

Relevance: Implementation guidance for this requirements lists “procedures for monitoring, detecting, analysing and reporting of information security events and incidents;”, which can be fulfilled with RATS functionality.

Collection of evidence (16.1.4) (medium relevance)

Requirement: The organization should identify, collect, acquire and preserve information, which can serve as evidence for security breaches.

Relevance: RATS provide identification and collection of information which can serve as evidence.

Technical compliance review (18.2.3) (high relevance)

Requirement: Information systems should be regularly reviewed for compliance with the organization’s information security policies and standards.

Relevance: RATS regularly evaluate systems for compliance. The clarification demands “Technical compliance should be reviewed preferably with the assistance of automated tools”. RATS provide such automated tools.

3.4 NIST SP800 – 171A

NIST SP 800-171 is a set of guidelines and controls by the National Institute of Standards and Technology (NIST) to enhance the security of Controlled Unclassified Information (CUI) in non-federal systems and organizations [8].

Audit and accountability (3.3.1) (medium relevance)

Requirement: Create and retain system audit logs and records to the extent needed to enable the monitoring, analysis, investigation, and reporting of unlawful or unauthorized system activity.

Relevance: RATS creates the required system audit logs that allow to monitor, analyze and investigate unlawful or unauthorized system activity.

Configuration management (3.4.1) (medium relevance)

Requirement: Establish and maintain baseline configurations and inventories of organizational systems (including hardware, software, firmware, and documentation) throughout the respective system development life cycles.

Relevance: RATS can at least support if not fulfill especially the following requirements: (3.4.7) Restrict, disable, or prevent the use of nonessential programs, functions, ports, protocols, and services; (3.4.8) Apply deny-by-exception (blacklisting) policy to prevent the use of unauthorized software or deny-all, permit-by-exception (whitelisting) policy to allow the execution of authorized software; (3.4.9) Control and monitor user-installed software.

Security Assessment (3.12.1) (medium relevance)

Requirement: Periodically assess the security controls in organizational systems to determine if the controls are effective in their application.

Relevance: RATS can be a part of security assessments and especially help to fulfill the requirement for periodic assessments since RATS assessments can be done automatically.

System and information integrity (3.14.2, 3.14.3) (high relevance)

Requirement: Provide protection from malicious code at designated locations within organizational systems. Monitor system security alerts and advisories and take action in response.

Relevance: Detection and escalation of unwanted software (e.g., malware) is a core functionality of RATS. The implementation of RATS and their integration with SIEM systems therefore provide a large part of this requirement.

3.5 IEC 62443

IEC 62443 is an international series of standards addressing cybersecurity for operational technology in automation and control systems [6]. These standards apply a risk-based approach to prevent and manage security for both entire systems (IEC 62443-3-3) and their components (IEC 62443-4-2). Five security levels (SL0-SL4) are described, with SL4 offering the highest security guarantees.

CR 1.2 – Software process and device identification and authentication (low relevance)

Requirement: Components shall provide the capability to identify itself and authenticate to any other component. If the component is running in the context of a human user the identification and authentication of the human user may be part of the component identification and authentication process.

Relevance: RATS typically provide means to unique identify devices through a hardware root of trust. E.g., trust established via a TPM chip and its associated certificates can be used to uniquely identify and authenticate devices.

CR 3.2 – Protection from malicious code (medium relevance)

Requirement: The application product supplier shall qualify and document which protection from malicious code mechanisms are compatible with the application and note any special configuration requirements.

Relevance: RATS aim at remotely detecting malicious code on devices. Thus, this requirement directly applies to RATS.

CR 3.4 – Software and information integrity (high relevance)

Requirement: Components must support integrity and authenticity checks on software, configurations, and data. The results of the integrity checks shall be recorded and reported. For SL 3 and SL 4, a configurable entity must be automatically about unauthorized changes.

Relevance: This requirement fully matches RATS, as RATS are about performing integrity checks on software, configuration, and further data, as well as reporting the results to an external party. Note that other mechanisms, such as secure boot, provide integrity checks, but are unable to report the result in a secure way to an external party. However, care must be taken to achieve SL3 and above, as RATS are typically invoked by an external party. Thus, to achieve SL3 and higher, provers must be equipped with the feature to perform self-checks.

EDR 3.12 – Provisioning product supplier roots of trust (medium relevance)

Requirement: To validate the authenticity and integrity of hardware, software, and data, a trusted source of data, known as the ”root of trust” is required. This root of trust can be cryptographic hashes of known-good software or the public part of an asymmetric cryptographic key pair used for verifying cryptographic signatures. The root of trust is crucial for verifying critical components before booting to ensure that the system starts in a known secure state.

Relevance: RATS builds upon cryptographic hashes of known-good software states as a root of trust in order to validate that software, firmware, and data are uncompromised. Therefore, this requirement applies to RATS.

EDR 3.13 – Provisioning asset owner roots of trust (medium relevance)

Requirement: To safeguard component security when extending functionality, asset owners should be able to validate and approve origins, necessitating the provision of secure ”roots of trust” by product suppliers that can differentiate between authorized and unauthorized origins.

Relevance: RATS can be implemented in a way that the trust established by the product supplier is extended to the asset owner.

EDR 3.14 – Integrity of the boot process (medium relevance)

Requirement: Embedded devices shall verify the integrity of the firmware, software, and configuration data needed for the component’s boot and runtime processes prior to their use.

Relevance: RATS verify the integrity of the firmware, software, and configuration data, but perform the verification after executing the component. Nevertheless, there are existing modifications to RATS that also enable a local verification, e.g., the IMA-appraisal feature of the Linux Integrity Measurement Architecture.

CR 6.2 – Continuous monitoring (high relevance)

Requirement: Components shall provide the capability to be continuously monitored using commonly accepted security industry practices and recommendations to detect, characterize and report security breaches in a timely manner.

Relevance: RATS goal is to continuously monitor whether malicious code and data is being executed on a remote component. Thus, this requirement fully maps to RATS. To ensure a timely response, the frequency in which the verifier quires and checks the integrity of the prover needs to be chosen appropriately.

4 Conclusion

The increasing connectivity of industrial systems made security a crucial requirement, leading to the emergence of security regulations and standards. These standards serve as a framework to mitigate security threats, but often lack specific implementation guidance, which can result in inadequately implemented security measures. Remote attestation is a promising security measures to ensure the integrity of remote devices and systems. However, its security capabilities and compliance with standards is not well-understood. This work analyzed remote attestation's alignment with industrial security standards and showed that it is highly relevant to the analyzed standards, in particular, NERC CIP and IEC 62443. Thus, remote attestation not only mitigates cyber threats on a technical level, but also provides proof for strong security by providing compliance with well-known security regulations and standards.

References

1. Manuel Atug, "Zertifizierungen im Kontext KRITIScher Infrastrukturen: Vorgaben und Möglichkeiten für KRITIS-Betreiber", Datenschutz und Datensicherheit (2020).
2. Chiara, Pier Giorgio. "The Cyber Resilience Act: the EU Commission's proposal for a horizontal regulation on cybersecurity for products with digital elements: An introduction." International Cybersecurity Law Review (2022).
3. Leander, Björn, Aida Čaušević, and Hans Hansson. "Applicability of the IEC 62443 standard in Industry 4.0/IIoT." Proceedings of the 14th International Conference on Availability, Reliability and Security. 2019.
4. Sailer, Reiner, et al. "Design and implementation of a TCG-based integrity measurement architecture." USENIX Security symposium. Vol. 13. No. 2004.
5. Birkholz, H., et al. "RFC 9334 Remote ATtestation procedureS (RATS) Architecture." (2023).
6. ISA-62443 Security for Industrial Automation and Control Systems. Standard, International Society of Automaton (2017).
7. North American Electric Reliability Corporation (NERC) Cyber Security Standards. https://nerc.com/pa/Stand/Pages/Cyber-Security-Permanent.aspx (2006).
8. NIST Special Publication 800-82: http://dx.doi.org/10.6028/NIST.SP.800-82r2 (2015)
9. Information technology - Security techniques - Code of practice for information security controls (ISO/IEC 27002) (2017).
10. Payment Card Industry Data Security Standard (PCI-DSS) Version 3.2.1 (2018).

http://dx.doi.org/10.25673/111637

Experimental Validation of a RDMA-based High Availability Concept over a deterministic IP-based network

Thomas Kampa[1], Daniel Grossmann[2]

Abstract: The edge computing paradigm continues to disrupt areas such as the automation domain where hardware-based implementations, i.e., programmable logic controllers (PLCs), have dominated for decades. The shift to virtualization and data center technologies offers both new opportunities and risks. By providing the first experimental validation of Remote Direct Memory Access (RDMA) over a deterministic IP-based network, we converge the IT and OT domains with a high availability approach for virtualized PLCs. Synchronizing the state of two PLC instances via RDMA provides fault tolerance across multiple data centers and exemplifies the potential of data center technologies for the automation industry. The experimental validation demonstrates the excellent match between the communication requirements of RDMA and the characteristics of the deterministic network. Finally, the need for deterministic acyclic communication in the pursuit of IT/OT convergence is illustrated.

Keywords: DetNet; vPLC; Real-time

1 Introduction

IT/OT convergence has been the target of various research efforts in the past and continues to attract work, often with the goal of taking advantage of the mature and highly scalable IT infrastructure in the proprietary environment of the automation domain. A prominent example is the virtualization of workloads with real-time constraints. Often referred to as edge computing, it continues to make inroads into areas where bare-metal deployments were considered the only option.

In our previous work, we presented an approach for this class of applications to meet their real-time requirements in failover scenarios through state synchronization using Remote Direct Memory Access (RDMA) [KEG23]. The concept was applied to a redundant Programmable Logic Controller (PLC) with a hot standby application running in parallel to the active PLC on a different physical infrastructure component, providing fault tolerance in failure scenarios.

The goal of this work is the subsequent experimental validation with PROFINET RT communication and a deterministic IP network with jitter and latency bounds [Ba19]. Our

[1] Technische Hochschule Ingolstadt, AImotion Bavaria, Esplanade 10, 85049 Ingolstadt, Germany;
AUDI AG, Auto-Union-Straße 1, 85057 Ingolstadt, Germany, thomas.kampa@audi.de

[2] Technische Hochschule Ingolstadt, AImotion Bavaria, Esplanade 10, 85049 Ingolstadt, Germany,
daniel.grossmann@thi.de

previous work only measured the synchronization time as a function of state size, missing the holistic view of the end-to-end process [KEG23]. Furthermore, this work includes a first evaluation for RDMA over a deterministic network with the aim to verify the capabilities of the mentioned network under the load of lower prioritized network traffic, i.e., TCP/IP based communication.

This work is structured as follows: Section 2 provides related work in the topics RDMA, DetNet, and virtualization. Section 3 introduces the concept and requirements for the deterministic network. An experimental validation is performed in Section 4 and discussed in Section 5. Finally, Section 6 concludes this work and gives an outlook on further research opportunities.

2 Related Work

This section provides an introduction to the state of research for RDMA, DetNet, and virtualization in the OT domain.

Through the usage of RDMA, memory synchronization between two hosts is possible without the involvement of the CPU or caches of the respective host, reducing latency and improving throughput [Zh17]. Dropped or out of order packets lead to retransmission, which is why RDMA requires lossless transmission of frames. A possible solution is the Priority Flow Control (PFC) defined in IEEE 802.1Qbb [Mi18]. For RDMA to work over Ethernet, the term Converged Ethernet (CE) was coined to include properties such as priorization of RDMA over lower priority traffic. Today, CE poses the basis of current RDMA over Converged Ethernet (RoCE) implementations [Mi18][Zi23].

However, the use of PFC can lead to a network-wide deadlock. To mitigate this issue, a current research trajectory are improved RoCE Network Interface Cards (ICN) [Mi18]. Previous work has already pointed out that traffic uncertainty poses the biggest challenge in simultaneously balancing congestion control on one hand and achieving high throughput and deterministic latency on the other [Zh21].

DetNet, that is currently being standardized by the IETF, could be a suited technology to ensure deterministic behavior and congestion prevention for routed environments [IE22]. In our previous work we proposed RDMA over DetNet (RoDN) as the next natural evolution step of RoCE in the IT/OT convergence area [KEG23]. DetNet is promised to provide deterministic upper and lower bounds for packets and zero congestion due to planned resource allocation. Its characteristics constitute a perfect match with the properties of CE required for RoCE. The devices used in the validation are deterministic IP routers that have been proven to have comparable characteristics and support IE protocols [Ba19].

While RDMA is extensively being used in data centers, so far it has not reached the OT domain. This might change with the introduction of edge computing and the virtualization of critical services such as PLCs. The need for high availability and resilience in the automation world has been historically immense, which is why the same properties will be required from a virtualized consolidated architecture as outlined in our previous work [KEG23].

3 Methods

This section introduces the high availability concept of a stateful real-time application with high availability requirements in the OT context, motivates the need for a deterministic IP-based network to connect data centers, and illustrates the communication requirements for this use case.

3.1 Synchronization Concept

In our previous work, we introduced a concept for high availability of vPLCs through state synchronization based on RDMA [KEG23]. The concept was applied to a redundant PLC with a hot standby application running in parallel to the active PLC on a different physical infrastructure component, providing fault tolerance in failure scenarios. The state of each application is synchronized every task cycle to keep the state of both instances consistent. Using RDMA over UDP-based synchronization reduced the average synchronization time by up to 99.39% for a modified software PLC from CODESYS. Fig. 1 shows a timing diagram of the synchronization process. It is important to note that the whole IEC task that is being synchronized must be completed before synchronization is performed. Also, the entire state is synchronized, even those variables that are not marked as persistent, to allow seamless failover.

3.2 RDMA over DetNet

In our previous work, we motivated the need for network-based scheduling of synchronization tasks in an architecture with consolidated vPLCs [KEG23]. Centralized management of the synchronization of multiple applications with their respective backups may be necessary to avoid burst scenarios and high retransmission rates, especially with UDP-based RoCE. Purely application-triggered synchronization is not desirable due to the lack of a holistic view of network and host synchronizations. A first step in this direction is to evaluate the suitability of RDMA over DetNet.

From a compatability perspective, RoCE should work without modifications for any given Ethernet-based DetNet implementation because it relies on UDP as the transport protocol. Prioritizing RDMA over lower-priority traffic is typically done by utilizing the priority bits in the Ethernet / IP header. An implementation based on holistic network-based scheduling is desirable, but out of scope for this work due to limitations in the configuration of deterministic IP routers and the timing of synchronization between vPLC instances.

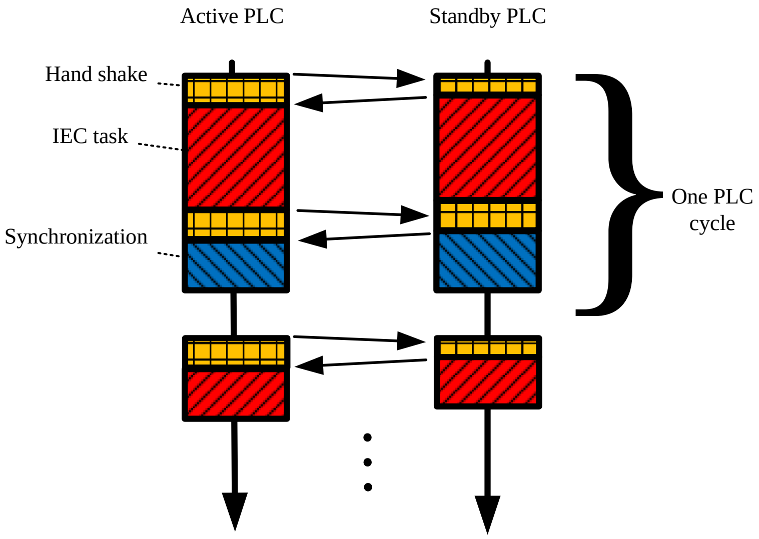


Fig. 1: Timing diagram of two PLCs, an active PLC and a standby PLC, synchronizing their states and cyclically computing a task.

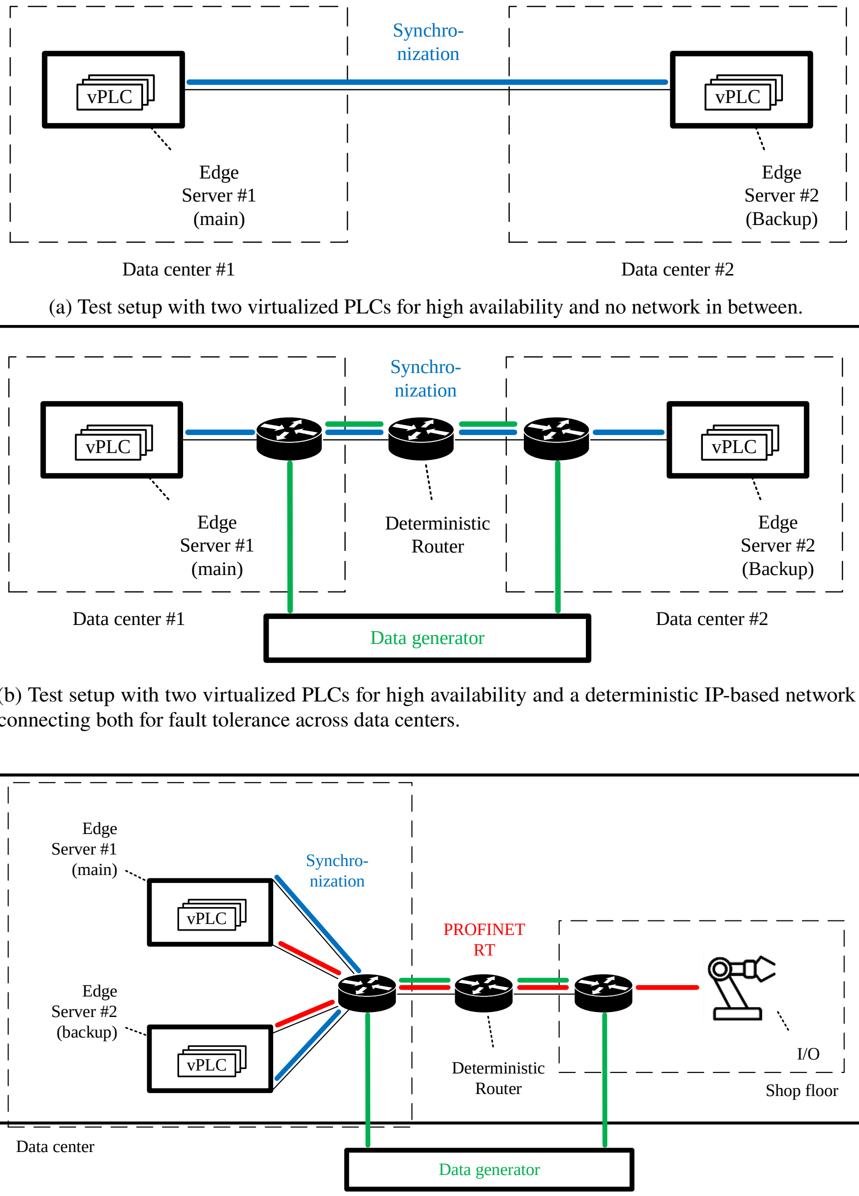
4 Experimental Validation

This section includes a description of the experimental setup, the test cases as well as the result.

4.1 Setup

Three different test setups are defined to evaluate the suitability of a deterministic IP-based network for RDMA and PROFINET RT. Fig. 2a gives a baseline for synchronization between two hosts with no network nodes in between. Fig. 2b includes three deterministic IP routers to mimic inter-data center communication. Fig. 2c evaluates vPLCs synchronizing over one hop and communicating over three hops with their respective I/Os. We consciously avoided running RDMA in parallel to IE traffic on the same link, since we believe that in real deployments both traffic classes would transverse different dedicated physical paths.

To evaluate the prioritization of PROFINET RT and RDMA-based traffic, best-effort traffic in form of UDP-packets generated by iPerf3, noted as data generator in the respective setups and generating enough traffic to fully utilize the whole bandwidth of the link between the routers. The uplinks of the edge servers have a 1 Gbit/s interface for RDMA and PROFINET, whereas the connection between the routers have a bandwidth of 10 Gbit/s.



(a) Test setup with two virtualized PLCs for high availability and no network in between.

(b) Test setup with two virtualized PLCs for high availability and a deterministic IP-based network connecting both for fault tolerance across data centers.

(c) Test setup with two virtualized PLCs for high availability and a deterministic IP-based network connecting both with PROFINET I/Os.

Fig. 2: Test cases with different network topologies and areas of interest.

Since DetNet is still in the standardization process, this work utilized deterministic IP networking routers that have been proven to provide determinism for real-time Industrial Ethernet (IE) traffic such as PROFINET [Ba19].

4.2 Results

Tab. 1 displays the values of the synchronization time as a function of state size and amount of network nodes in between, i.e., Fig. 2a equals zero network nodes and Fig. 2b equals three network nodes. It is apparent that the addition of three deterministic routers only increases the synchronization time by high two to low three-digit microseconds, whereas the Standard Deviation (SD) almost remains unchanged. It is important to note that the measurement had a millisecond accuracy limitation due to an implementation in the CODESYS vPLC.

The conducted validation after Fig. 2c resulted in no watchdog timers expiring with an update time of 1 ms and a watchdog timer of 3 ms.

Network nodes	State Size	Min	Max	Average	SD
0 network nodes	100 KB	0.00	2.00	0.957	0.205
3 network nodes	100 KB	1.00	3.00	1.019	0.136
0 network nodes	1 MB	8.00	10.00	8.978	0.146
3 network nodes	1 MB	9.00	10.00	9.056	0.231
0 network nodes	10 MB	89.00	90.00	89.224	0.417
3 network nodes	10 MB	89.00	90.00	89.274	0.446

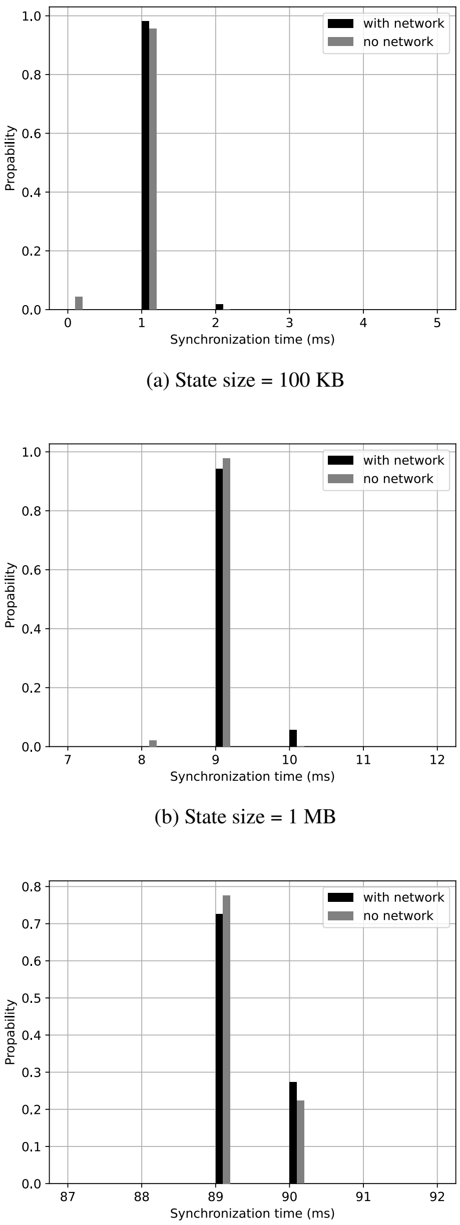
Tab. 1: Statistical analysis of the state synchronization times of the CODESYS PLC with either none or three network nodes in between. All values are in milliseconds if not described otherwise.

5 Discussion

The following section discusses the presented results and provides further observations.

5.1 Meeting Real-Time Requirements of vPLCs

The envisioned concept is suitable for applications running on virtual machines and containers and could be an enabler for the virtualization of real-time critical applications such as control functions in the automation and process industries. As displayed in Section 4.2 the routed environment that mimics a data center connection fulfills the requirements in terms of real time and utilization. Moreover, PROFINET traffic gets prioritized over best-effort traffic and meets its stringent real-time requirements.



(a) State size = 100 KB

(b) State size = 1 MB

(c) State size = 10 MB

Fig. 3: Synchronization time with and without the deterministic network after 2a) and b).

5.2 Failover Times

Failover times, i.e., the time it takes for the standby PLC to take over from the failed active one, were not part of the conducted experiments. This is mostly due to the high variety of scenarios that one might construct, depending on setup and configuration. Possible things to consider are:

- The frequency of the heartbeat and the related timeout after which a PLC is deemed to be lost.
- The implementation of the redundancy mechanism, i.e., does the heartbeat only check in between tasks or continuously.
- The protocol involved, e.g., stateful (PROFINET) and stateless (EtherCAT) protocols.
- Whether the I/Os in the field require a reconnection, e.g., PROFINET S1 vs S2 devices.

We expect the involved failover time to be in the high double digit millisecond range or there might be no failover at all and a seamless takeover happens through the standby PLC.

5.3 Resiliency Options

Depending on the requirements of the application, different deployment options can be realized to achieve varying degrees of resiliency.

The concept presented in this work allows seamless failover between two hosts with no loss of state or interruption of the process, if the application and the respective I/Os allow it. However, these benefits come with a price to pay: The state is being synchronized every single PLC task cycle, leading to high bandwidth requirements for even small state sizes if the synchronization time is desired to be kept low. Moreover, a second PLC instance is running in a hot-standby mode at all times, leading to increased resources utilization and additional licenses from the PLC vendors.

Another, less costly option, provides the saving of the state into a database or share, which in turn could also be conducted with RDMA. This way, a new vPLC would have to be spun up and synchronized with the last known state before restarting operations. While this might safe an additional PLC running in parallel, it might lead to loss of data in the split seconds following the failure of the primary PLC, leading to possible increased recovery times.

Finally, only parts of the state could be saved and stored every cycle, reducing the load on the network and the time for saving even further. This could either be only the data that has changed since the last save, a method that we called partial synchronization in our previous work [KEG23]. Another prime target could be persistent data, which is often written into a persistent storage in case of a power outage of the PLC. This albeit small change could already reduce the synchronization time mutliple times, increasing performance and reducing network load.

5.4 Network Utilization Pattern

PLC task cycles can vary in their duration time due to co-routines, varying hardware utilization and other external effects, especially in a virtualized environment. The synchronization of the state is conducted after completing a full task as described in Fig. 1. Varying task cycle times do not effect the network utilization due to PLC - I/O communication as shown in Fig. 4. However, converging the communication of the PLCs with their respective I/Os and the synchronization of PLCs in between creates an remarkable network utilization pattern which is illustrated in the lower portion of Fig. 4. This illustration shows that acyclic communication patterns need to coexist with cyclic communication patterns, both requiring deterministic real-time properties.

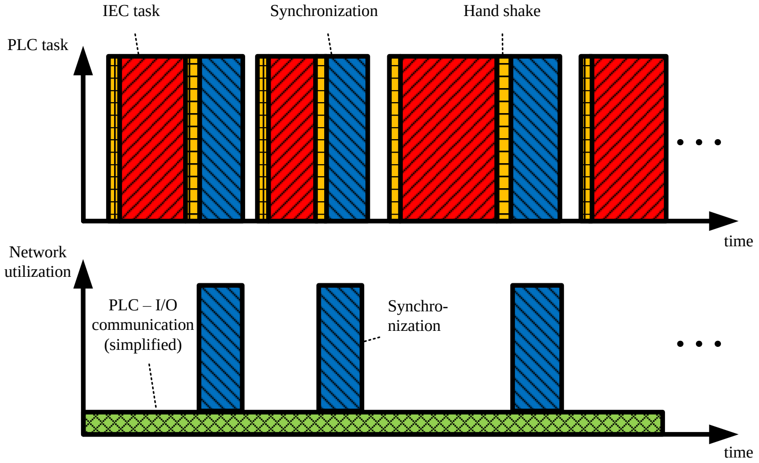


Fig. 4: How variable IEC task duration can lead to acyclic communication patterns with real-time constraints in synchronization scenarios.

6 Conclusion

In this work, we have performed an experimental validation of RDMA over a deterministic IP-based network. A high-availability concept for vPLCs is used as an evaluation example to highlight the requirements for this specific combination of technologies. The results prove that deterministic networking complements RDMA well, even under high network utilization, and poses a suitable alternative to pure RoCE to guarantee packet delivery and enable deterministic behavior.

Future research may focus on evaluating other resiliency options and improving state synchronization as well as evaluating the failover time across different deployment options. In addition, acyclic communication patterns need more attention in the automation domain as they may emerge as another class of traffic with similar deterministic requirements as cyclic real-time traffic.

Acknowledgment

The presented concept was developed within the cluster “Digital Production” of the Research Institute AImotion Bavaria of the Technische Hochschule Ingolstadt. The project is funded by AUDI AG. We thank Dr. Zhe Lou and Sushrey Sunil Sawant for providing the Deterministic IP networking setup.

References

[Ba19] Badar, A.; Lou, D. Z.; Graf, U.; Barth, C.; Stich, C.: Intelligent Edge Control with Deterministic-IP based Industrial Communication in Process Automation. In (Lutfiyya, H., Hrsg.): 15th International Conference on Network and Service Management: 1st International Workshop on Analytics for Service and Application Management (AnServApp 2019) : International Workshop on High-Precision Networks Operations and Control, Segment Routing and Service Function Chaining (HiPNet + SR/SFC 2019) : October 21-25 2019, Halifax, Canada. IEEE, [Piscataway, NJ], 2019, ISBN: 9783903176249.

[IE22] IETF: Deterministic Networking (detnet), 2022, URL: https://datatracker.ietf.org/wg/detnet/documents/, Stand: 09. 10. 2023.

[KEG23] Kampa, T.; El-Ankah, A.; Grossmann, D.: High Availability for virtualized Programmable Logic Controllers with Hard Real-Time Requirements on Cloud Infrastructures. In: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). IEEE, 2023.

[Mi18] Mittal, R.; Shpiner, A.; Panda, A.; Zahavi, E.; Krishnamurthy, A.; Ratnasamy, S.; Shenker, S.: Revisiting network support for RDMA. In: Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication. ACM, New York, NY, USA, 2018.

[Zh17] Zhang, Y.; Gu, J.; Lee, Y.; Chowdhury, M.; Shin, K. G.: Performance Isolation Anomalies in RDMA. In: Proceedings of the Workshop on Kernel-Bypass Networks. 2017.

[Zh21] Zhang, J.; Zhang, Y.; Guan, Z.; Wan, Z.; Xia, Y.; Pan, T.; Huang, T.; Tang, D.; Lin, Y.: HierCC: Hierarchical RDMA Congestion Control. In: 5th Asia-Pacific Workshop on Networking (APNet 2021). S. 29–36, 2021.

[Zi23] Zilong Wang; Layong Luo; Qingsong Ning; Chaoliang Zeng; Wenxue Li; Xinchen Wan; Peng Xie; Tao Feng; Ke Cheng; Xiongfei Geng; Tianhao Wang; Weicheng Ling; Kejia Huo; Pingbo An; Kui Ji; Shideng Zhang; Bin Xu; Ruiqing Feng; Tao Ding; Kai Chen; Chuanxiong Guo: SRNIC: A Scalable Architecture for RDMA NICs. In. S. 1–14, 2023, ISBN: 978-1-939133-33-5, URL: https://www.usenix.org/conference/nsdi23/presentation/wang-zilong.

http://dx.doi.org/10.25673/111638

6G NeXt — Joint Communication and Compute Mobile Network: Use Cases and Architecture

Sergiy Melnyk[1], Qiuheng Zhou[1], Hans D. Schotten[1,2], Wolfgang Rüther-Kindel[3], Fabian Quaeck[3], Nick Stuckert[3], Robert Vilter[3], Lisa Gebauer[3], Mandy Galkow-Schneider[4], Ingo Friese[4], Steffen Drüsedow[4], Tobias Pfandzelter[5], Mohammadreza Malekabbasi[5], David Bermbach[5], Louay Bassbouss[6], Alexander Zoubarev[6], Andy Neparidze[6], Arndt Kritzner[7], Jakob Hartbrich[8], Alexander Raake[8], Enrico Zschau[9], Klaus-Jürgen Schwahn[10]

Abstract:

The research on the new generation mobile networks is currently in the phase of defining the key technologies to make 6G successful. Hereby, the research project *6G NeXt* is aiming to provide a tight integration between the communication network, consisting of the radio access as well as backbone network, and processing facilities. By the concept of split computing, the processing facilities are distributed over the entire backbone network, from centralised cloud to the edge cloud at a base station. Based on two demanding use cases, *Smart Drones* and *Holographic Communication*, we investigate a joint communication and compute architecture that will make the application of tomorrow become reality.

Keywords: 6G, agile link adaptation, split computing, high-speed backbone, geo-distributed computing, uav, anti-collision system, wireless closed loop, holographic communication, quality of experience, split rendering, metaverse

1 Introduction

The idea of offloading demanding processing tasks to powerful machines by splitting an application into client and server parts has been known for decades. Nevertheless, the

[1] German Research Center for Artificial Intelligence (DFKI GmbH), Kaiserslautern, Germany, {sergiy.melnyk, qiuheng.zhou, hans_dieter.schotten}@dfki.de

[2] University of Kaiserslautern-Landau, Kaiserslautern, Germany, schotten@rptu.de

[3] Technical University of Applied Sciences Wildau, Wildau, Germany, {wkindel, quaeck, stuckert, vilter, lisa_josephine.gebauer}@th-wildau.de

[4] Deutsche Telekom AG, Berlin, Germany, {mandy.galkow-schneider, ingo.friese, steffen.druesedow}@telekom.de

[5] Technische Universität Berlin & ECDF, Berlin, Germany, {tp, mm, db}@mcc.tu-berlin.de

[6] Fraunhofer FOKUS, Berlin, Germany, {louay.bassbouss, alexander.zoubarev, andy.neparidze}@fokus.fraunhofer.de

[7] Logic Way GmbH, Schwerin, Germany, kritzner@logicway.de

[8] Technische Universität Ilmenau, Ilmenau, Germany, {jakob.hartbrich, alexander.raake}@tu-ilmenau.de

[9] SeeReal Technologies GmbH, Dresden, Germany, ez@seereal.com

[10] Schönhagen Airport, Trebbin, Germany, drschwahn@edaz.de

concepts such as edge computing introduced in 5G networks, enhance the clients with new capabilities making new types of server-client-splits possible. Sixth-generation networks are aiming to go even beyond this concept by flexibly distributing processing power over the whole backbone network. That is the core topic of the research project *6G NeXt – 6G Native Extensions for eXtended Reality (XR) Technologies*. Provided the tight integration with the communication infrastructure, the split computing approach would significantly enhance the 6G network experience. Based on two example use cases, *Smart Drones* anti-collision system and *Holographic Communication* (*HoloCom*), the project *6G NeXt* aims to develop and evaluate a joint connectivity and compute infrastructure, which will introduce new processing speeds in conjunction with highly dynamic geo-distributed computing capabilities within a mobile network.

Future outlooks for aviation anticipate an increasing number of remotely operated flight missions as well as an increasing number of manned flights [Rü22]. Until today, many strategies have been developed on the topic of how to implement unmanned aerial vehicles (UAVs) in the civil aviation airspace. In order to ensure a safe operation in shared airspace, a common anti-collision system will be necessary in the future. However, the key to success lies in the connectivity between different aircraft. In our work, we primarily focus on *Smart Drones*, a drone to drone anti-collision but future extensions to manned aviation are also conceivable.

The other application is supposed to bring the today's videoconferencing to the next level. By means of dedicated holographic displays, a real holographic 3D image can be created. Furthermore, the integrated eye-tracking capability allows the change of the view angle with no noticeable delay. This will provide people with a native 3D experience of their communication partners. However, the amount of holographic data is tremendous, which makes video processing on the client device impossible. Thus, the majority of processing steps such as rendering need to be performed on the side of the mobile network.

The remainder of the paper is as follows. In the next two sections, we provide a description of the targeted use cases. The appropriate architecture called to support our anti-collision system is shown in the Sect. 4. Here, we also provide a detailed description of technical solutions, required to support this architecture. Finally, Sect. 5 summarises the paper.

2 Smart Drones

The idea of an anti-collision system for drones or UAVs is derived from the research project VIGA (Virtual Instructor for General Aviation) [Rü22]. Here, the approach was based on a simulation of an aircraft that runs faster than the real life. In this way, evaluations of the possible flight trajectory can be estimated and warnings about hazardous situations can be provided. The proposed *Smart Drones* anti-collision system extends this approach to the UAVs. To do so, the application is split into two parts, which exchange data via a wireless link:

- Ground side: ground-based simulation engine with high processing power, which captures the UAVs' flight data and performs predictions of the flight paths.
- UAV side: flight controller, which keeps the communication to ground stations as well as controls the UAV accordingly

The flight path simulations are offloaded to the ground station in order to reduce the required computation power and thus the takeoff weight of the UAVs. The ground station is meant to collect all data streams of each individual UAV flying in the specific sector. The transmitted data contains intended mission profile data and UAV-specific data (flight model values) as well as altitude, position, course, and velocity information (state values), which are acquired on the drone at a high rate. The digital twin representation of the UAVs is fed by this UAV-specific information and the simulation is carried out with respect to the currently received state values. If a collision scenario is predicted by the system, evading trajectories will be calculated and simulated with the digital twins before the manoeuvre is transmitted and performed by the real UAV.

The performance of the system relies on the timeliness of the transmitted data. Thus, low latency and high reliability in the communication link are important since the overall time of data transmission, simulation, and maneuver execution has to be minimized. Moreover, the movement of the UAVs through different sectors requires dynamic data distribution. This requires a software architecture that is capable of allocating the UAV's data dynamically and handing over existing simulation data to other sectors.

To prove the feasibility of *Smart Drones* system, a real-life testbed setup is planned at Schönhagen Airfield (EDAZ). The airfield is located to the south of Berlin and Potsdam and it is one of the largest commercial airfields in Germany. Moreover, the airfield supports the research activities in the region. The long-term cooperation with universities and research institutes, such as German Aerospace Center (DLR), TU Berlin, and others, leads to newly emerging research and development projects. The Technical University of Applied Science of Wildau is a permanent cooperation partner, that maintains its own research aircraft. The most recent developmental step is the construction of an aviation safety centre with notable partners from the air traffic control and safety sectors. The centre should make a name for itself through its participation in national and international research activities, the provision of test environments, and simulations of threat scenarios for training and research, including the development and implementation of training modules.

3 Holographic Communication

Holographic communication is expected to become the next generation of video communication, with key players such as Ericsson and Google announcing their respective investments into the technology in recent years [Er23; La21]. While Ericsson aims for a device setup built around commodity hardware, Google's Starline "relies on custom-built hardware and highly specialized equipment". However, what both solutions have in common is the general processing pipeline for holographic communication, which can be applied to any concrete implementation. It consists of three steps: capturing, rendering, and display. In the capturing step, a 3D camera setup, which can consist of one or more cameras, is used to record the communication participant. In the rendering step, the resulting data is then processed into the 3D representation of the recorded participant, before being displayed on a device capable of playing back holographic video.

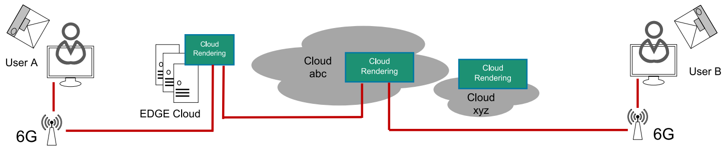


Fig. 1: *HoloCom* use case setup

The goal for *HoloCom* is the development of a real-time video communication system in a low-latency and high-bandwidth environment using specialized capturing and playback hardware to achieve an immersive and photo-realistic experience. Fig. 1 shows the exemplary setup of the use case. It details a split computing approach for the holographic communication pipeline, where capturing, rendering, and display are distributed between the communication participants and a local edge or cloud. When user A wants to call user B, image and depth information are continuously captured by a camera array on user A's side. The extremely high volume of data, potentially exceeding multiple terabytes per minute [Fr23], is transmitted via a 6G network to the nearest edge cloud, where it serves as input to a headless instance of the 3D engine Unity. In Unity, a 3D representation of user A is constructed from this data. Subsequently, the 3D representation is captured by a virtual camera, serving the data in a format required by the holographic display on user B's side. The display used mimics visual information used by the eyes to process depth in real-world objects, making the communication partner appear as they would in a face-to-face conversation. Enabled by this approach and coupled with eye tracking, the user is, within limits, free to move around in the physical space. The same pipeline is executed for user B, highlighting the necessity for high bandwidth and low latency, ensuring low delay without asynchronicity to achieve a good user experience.

4 System Architecture and Key Technologies

As described in previous sections, *Smart Drones* system as well as *HoloCom* are complex applications. On the one hand, they consist of mobile clients, which maintain wireless connection to the mobile network. On the other hand, the processing resources such as ground station or rendering facilities, are provided on the backbone network. Moreover, in order to cope with tight latency requirements, it is advantageous to split the processing tasks between edge clouds in the vicinity of the clients as well as high-performance centralised clouds.

Thus, resource orchestration requires tight coordination between communication and computing infrastructure, based on the behavior of the clients. In this paper, we provide a multilayer architecture, which is called to enhance the inter-layer coordination. Whereas the architecture shown in Fig. 2 is generic and applies to both applications, its functionalities we will exemplary explain its functionalities based on the *Smart Drones* system.

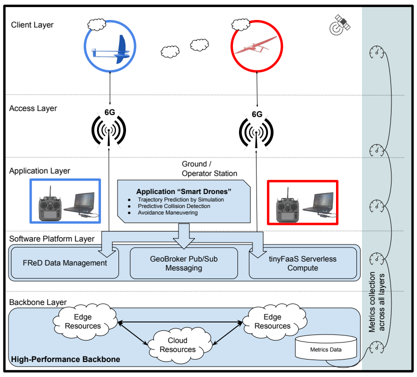


Fig. 2: Joint communication and compute architecture

This architecture will be able to provide the optimal resource allocation for all participants throughout the whole stack. Here, the UAVs are assigned to the *Client Layer*. They are logically connected to the ground control station, which is distributed on the *Application Layer* by means of mobile radio access network (RAN) at the *Access Layer*.

On the other hand, RAN maintains a physical connection to the *Backbone Layer*, where the high-performance backbone provides a convolution of communication network and computing infrastructure. In order to orchestrate the allocation of the applications to the processing resources, we introduce *Software Platform Layer*. Here, Function-as-a-Service (FaaS) based edge-cloud software platform provides an abstraction layer for the cloud-based applications running on *Application Layer*. The performance parameters of the functionalities on all layers are captured on the *Backbone Layer*. Here, optimisation of the resource allocation is performed. Taking into account the dynamically changing application environment, *Backbone Layer* anytime ensures appropriate resource allocation according to application demands.

4.1 Radio Access Network

In order to integrate the split computing infrastructure with the communication network, the latter should be able to flexibly react to the computation demands of the applications on the one side, and to the environmental influences on the quality of provided communication links on the other side. Therefore, it is necessary to address the inherent uncertainty and dynamic nature of the wireless environment and radio access network. While these requirements cannot always be guaranteed, machine learning (ML) models can be used to adapt to the spatiotemporal dynamics of wireless networks in advance. By providing predictive analytics to both end-user applications and radio access network elements, ML models can help to ensure the most efficient and effective operation of the system. Those are making it possible to make highly accurate predictions about changes in quality of service (QoS) and radio key performance indicators (KPIs), like radio environment maps (REMs), channel distribution information maps and spectral efficiency within the radio access network [Pe22; Zh22b]. By providing predicted QoS information, radio resource management schemes can offer more reliable future QoS guarantees to individual users, even when poor performance is expected. The characteristics of channel prediction methods on real communication data are also tested and verified by establishing an software-defined radio (SDR) based cellular communication and channel measurement [Zh22a].

Furthermore, the reliability of the communication channel can be increased by the adjustment of the coding schemes based on the predicted QoS values. In order to efficiently react to the changes of the channel conditions, the coding scheme should feature appropriate flexibility and agility. In our work, we investigate the utilisation of fountain codes to fulfil these requirements. A fountain encoder is able to generate an arbitrary number of codewords out of a payload packet, whereas only a certain number of them is required to decode the packet. Thus, the coding rate is not fixed, but can be adjusted “on the fly”. Since the fountain codes can withstand a drop of codeword, but a bit error, they should be utilised in conjunction with forward error correction (FEC). Even more, it was shown, that combination of error-detecting codes with fountain codes gains the performance of the communication system [Be08; Ka20]. The major advantage of fountain codes is their flexibility. Based on predicted channel quality, the coding rate can be flexibly adjusted in order to meet the QoS requirements [Me22]. Moreover, fountain coding schemes are capable of heterogeneous traffic requirements for different application types. Since a fountain encoder can be flexibly re-adjusted for any piece of payload, it will provide the agility to set up the optimal coding rate based on the traffic type as well as on the predicted channel conditions.

Final flight tests at Schönhagen Airport require an according wireless infrastructure. The modifications of wireless protocols will be realized and evaluated based on SDR platforms such as USRP. This concept is appropriate for ground-based infrastructure, whereas UAVs needs to be equipped with lightweight flight-capable terminal devices. For this purpose, the algorithms described above will be integrated with a dedicated communication module *NetMobilBox* [LW], the development of which goes back to the *5G NetMobil* project. Besides multi-connectivity capabilities, it provides the interfaces for the integration of additional sensors to capture positions and movement state information as well as communication to the UAV flight controller.

4.2 High-Performance Backbone

The performance and efficiency of a new network generation will be determined by high-performance radio interfaces with application-optimized radio protocols as well as by ultra-fast software stacks, intelligent distribution of computing tasks, and the deep integration of artificial intelligence (AI) to optimize the overall system. The concept of a High-Performance Backbone is based on the ability to deploy the computation tasks at places, where sufficient network capabilities, CPU/GPU, and memory are available. In the following, the key technologies for a joint communication and computing infrastructure are presented.

4.2.1 Split Computing

Mobility-related services and applications can be found in the automotive-, rail- or drone industries. They are typical examples of distributed applications and split computing. The *Smart Drones* application illustrates how the proposed architecture supports a smart distribution of computing tasks. In this scenario, drone steering and control, trajectory prediction, predictive collision detection, and avoidance maneuvering tasks are distributed between drone, edge, and central cloud deployments as seen in Fig. 2.

Distribution of computation tasks in the backbone layer based on KPIs enhances the concept of edge computing and provides new possibilities for future applications. The core of this concept is to match the requirements of an application such as *Smart drones* anti-collision system with the most suitable resources and connections that are available based on KPIs. *3GPP* standardized in release 18 of its 5G specification an *Architecture for enabling Edge Applications*. The goal of this architecture is to host applications in an edge cloud close to the base station and thus near to the clients in order to reduce end-to-end latency.

4.2.2 Cross-Layer Metrics Function

Before a computing task can be assigned to a certain computing infrastructure, it needs distinct knowledge about available cloud resources and connectivity characteristics. A new proposed metric function aims at gathering metrics and measurement data from *Application*, *Software Platform*, and *Access Layer*, as shown in the Fig. 2 on the right side.

4.2.3 Discovery and Broker Service Function

A *Discovery Service* [Mic22] is a typical part of a microservice architecture. Clients running in different locations or changing their locations need to find their optimal endpoint to connect to. The *Discovery Service* provides e. g. the UAVs clients with the network address of an anti-collision system service. Additionally, a service broker among others helps cloud-based services to find free cloud capacities as well as also to start a new instance based on requirements described with KPIs. When a service recognizes a decreasing service quality due to missing memory or CPU/GPU in the current cloud environment, it might ask the *Discovery* as well as *Broker Service Function* for a new deployment possibility. The *Service Broker Function* performs two tasks. First, it recommends a suitable cloud resource with sufficient connectivity for increasing service capacities. Second, it helps the service to deploy and start a new service instance using e. g. Docker and automated infrastructure tools.

In order to achieve the best suitable network connectivity, it is important to understand the nature of the traffic, e. g. in the surrounding of an aerodrome. The most crucial traffic parts must be identified and treated with a higher priority. This may be achieved by routing it to dedicated network slices with inherently higher traffic prioritisation, potentially combined with dedicated and reserved radio resources for such slices. As an additional mechanism, individual traffic flows can be elevated to a higher priority level on the fly by using QoS related application programming interfaces (APIs) as specified in the CAMARA Telco Global API Alliance.

4.3 Edge-Cloud Software Platform

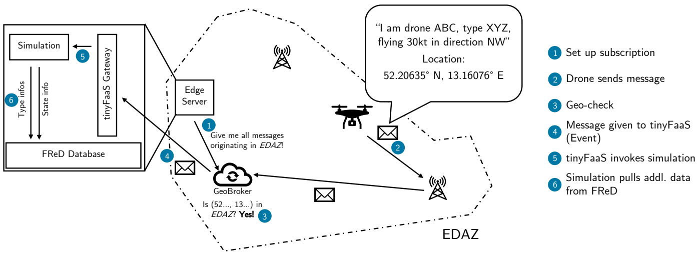

Simulation
5
tinyFaaS Gateway
6
Type infos
State info
FReD Database
Edge Server
1
Give me all messages originating in *EDAZ*!
4
GeoBroker
Is (52..., 13...) in *EDAZ*? **Yes!** 3
"I am drone ABC, type XYZ, flying 30kt in direction NW"
Location:
52.20635° N, 13.16076° E
2
EDAZ
1 Set up subscription
2 Drone sends message
3 Geo-check
4 Message given to tinyFaaS (Event)
5 tinyFaaS invokes simulation
6 Simulation pulls addl. data from FReD

Fig. 3: A workflow of the drone anti-collision system using serverless abstractions for the edge-cloud.

In order to make the high-performance communication and computation backbone suitable for applications, we need to use novel software abstractions. A serverless abstraction level can provide a unified interface to geo-distributed cloud and edge resources managing computation, messaging, and data replication.

We rely on distributed Publish/Subscribe (Pub/Sub) to provide messaging across distributed services, clients, and applications in the edge-to-cloud serverless platform. We use the distributed *GeoBroker* [HB20] that extends Pub/Sub with geographical context: *GeoBroker* can automatically filter messages not only by topic but also by geographic relevance so that subscribers receive only messages published in their proximity. This introduces a natural geographic sharing of responsibilities among replicas of an application service.

Serverless computation is implemented using the *tinyFaaS* [PB20]. It is a lightweight implementation of the FaaS paradigm, where applications are composed of interconnected stateless functions that are invoked in an event-driven manner.

Data replication is a key concern in distributed edge-to-cloud platforms: Always using a central cloud replica of a database incurs a high access latency for edge services. *FReD* is a data management middleware for the edge-to-cloud continuum that manages data replication for applications [HGB20]. Using geographically distributed nodes, applications simply specify a set of locations to replica a table of data to.

In Fig. 3, it is shown, how these abstractions can support the *Smart Drones* anti-collision system. The edge service simply specifies a geographic context that it deems relevant, e. g., the *EDAZ* airport perimeter. Any drone can send its location and status updates to a *GeoBroker* instance without specific knowledge on *which* service is interested in this data and *where* an instance of this service is running. *GeoBroker* then forwards this message to the edge service after confirming its topical and geographical relevance (with a check, whether the origin location lies in the specified geo-fence). The edge service lets *tinyFaaS* invoke the simulation component in an event-driven manner: *if* a message is received from a drone, *then* execute the simulation. A simulation instance is then automatically started by *tinyFaaS*. This instance can access a local copy of relevant data through an interface to *FReD*, e. g., to retrieve information about the maneuvering capability of a type of drone. It completes its calculations and returns instructions to the drone if necessary.

5 Conclusion

The 5th generation of mobile networks is still in the roll-out phase. However, the technologies introduced there, such as the edge computing concept, show certain limitations. Their capability will be not sufficient to serve a number of future-oriented applications, which are currently emerging. The UAVs control and anti-collision system *Smart Drones* and the *HoloCom* communication system, described in this paper, are just some of them. On the one hand, these applications set high demands on the capability of the wireless link to the clients, such as high reliability and low latency. On the other hand, a sophisticated distributed processing engine on the backbone network is required, which should be flexible to follow the mobile clients on their track. In the future 6G networks, communication and processing cannot be considered separately, but they will grow together into a joint communication-and-computing infrastructure.

Acknowledgements

Funded by the Bundesministerium für Bildung und Forschung (BMBF, German Federal Ministry of Education and Research) – 16KISK174K, 16KISK176, 16KISK177, 16KISK178, 16KISK179, 16KISK180, 16KISK182, 16KISK183.

References

[Be08] Berger, C. R. et al.: Optimizing Joint Erasure- and Error-Correction Coding for Wireless Packet Transmissions. IEEE Transactions on Wireless Communications 7/11, pp. 4586–4595, Nov. 2008.

[Er23] Ericsson Holographic Communication, 2023, URL: https://www.ericsson.com/en/ericsson-one/holographic-communication, visited on: 10/06/2023.

[Fr23] Friese, I. et al.: True 3D Holography: A Communication Service of Tomorrow and Its Requirements for a New Converged Cloud and Network Architecture on the Path to 6G. In: International Conference on 6G Networking (6GNet). IEEE, 2023.

[HB20] Hasenburg, J.; Bermbach, D.: GeoBroker: Leveraging Geo-Context for IoT Data Distribution. Elsevier Computer Communications 151/, pp. 473–484, Feb. 2020.

[HGB20] Hasenburg, J.; Grambow, M.; Bermbach, D.: Towards A Replication Service for Data-Intensive Fog Applications. In: Proceedings of the 35th ACM Symposium on Applied Computing, Posters Track (SAC '20). ACM, Brno, Czech Republic, pp. 267–270, Mar. 2020.

[Ka20] Karrenbauer, M. et al.: A Study on the Application of Rateless Coding in Non-Cellular MIMO Systems for Machine-Type Communication. IFAC 53/2, pp. 8243–8248, 2020.

[La21] Lawrence, J. et al.: Project Starline: A high-fidelity telepresence system. ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 40(6)/, 2021.

[LW] Logic Way Kommunikationsmodule, URL: https://logicway.de/pages/dimmcpu.shtml, visited on: 03/28/2023.

[Me22] Melnyk, S. et al.: Wireless Industrial Communication and Control System: AI Assisted Blind Spot Detection-and-Avoidance for AGVs. In: IN4PL. Pp. 1301–1306, Mar. 2022.

[Mic22] Service Discovery in Microservices, Nov. 2022, URL: https://www.baeldung.com/cs/service-discovery-microservices/, visited on: 04/26/2023.

[PB20] Pfandzelter, T.; Bermbach, D.: tinyFaaS: A Lightweight FaaS Platform for Edge Environments. In: Proceedings of the Second IEEE International Conference on Fog Computing (ICFC 2020). IEEE, Sydney, NSW, Australia, pp. 17–24, Apr. 2020.

[Pe22] Perdomo, J. et al.: QoS Prediction-based Radio Resource Management. In: 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). Pp. 1–6, 2022.

[Rü22] Rüther-Kindel, W. et al.: VIGA — Virtual Instructor for General Aviation. In: CEAS EuroGNC 2022. May 2022.

[Zh22a] Zhou, Q. et al.: Deep Learning-Based Signal-to-Noise Ratio Prediction for Realistic Wireless Communication. In: 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring). Pp. 1–5, 2022.

[Zh22b] Zhou, Q. et al.: Performance Evaluation over DL-Based Channel Prediction Algorithm on Realistic CSI. In: 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). Pp. 1–5, 2022.

http://dx.doi.org/10.25673/111639

Feldtest eines vernetzten autonomen Einschienenfahrzeugs MonoCab in einem 5G-Campusnetz

Timo Siekmann[1], Boris Rohde[2]

Abstract: Die Bedeutung von automatisierten und autonomem Fahren nimmt auf der Straße und Schiene sowie in der Produktion (AGV) und Logistik (Indoor und Outdoor) kontinuierlich zu.[G.] Ein Technologiebaustein dafür ist eine anforderungsgerechte Vernetzung der Fahrzeuge. WLAN (Wireless Local Area Network) und Mobilfunk sind je nach Anwendungsfeld mögliche Kandidaten für diese Vernetzung. Die Mobilfunkgeneration 5G verspricht neue Funktionen wie u.a. Network Slicing, Uplink orientierte Mobilfunknetze, Layer 2/3 Transperency und Ultra Low Latency Communication (URLLC).[RP18] In diesem Beitrag werden 5G Messungen auf einem Testfeld mit State of the Art Hardw- und Software durchgeführt. Anhand der Messergebnisse im Beitrag ist zu sehen, dass 5G Campusnetzte bereit sind die im Projekt erhobenen Anforderungen unter bestimmten Bedingungen zu erfüllen, Bis das volle Potential von 5G ausgeschöpft werden kann müssen die Hersteller jedoch weitere Funktionalitäten in ihre 5G Hardware implementieren.

Keywords: 5G Campusnetz; Vehicle to Vehicle Communication

1 Einleitung

Ein möglicher Weg zur Verkehrswende sowie der Minimierung von CO^2 ist die Automatisierung von ÖPNV Fahrzeugen wie z.B. Busse und Bahnen. Insbesondere die Automatic Train- oder die Remote Train Operation stellen hohe Anforderungen an die End-to-End Latenz sowie die Zuverlässigkeit des drahtlosen Kommunikationssystems.[3G19b] Dieser Beitrag beschreibt die Funknetzwerkplanung und Messung der Netzwerkperformance sowie Auswertung der Ergebnisse am Beispiel des 5G Feldtests im Projekt 5G-SIMONE (SIcher MObil VerNEtzt), gefördert durch den 5G.NRW Förderwettbewerb vom Ministerium für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen (MWIKE),mit einem automatisierten Einschienenfahrzeug MonoCab, welches in Abbildung 1 zu sehen ist (https://www.monocab-owl.de). Das Fahrzeug fährt auf einer Schiene mit zwei hintereinander angeordneten Rädern und es wird durch zwei elektromechanische Kreisel sowie einer Verschiebemasse stabilisiert.

[1] Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Land timo.siekmann@iosb-ina.fraunhofer

[2] Wireless.Consulting GmbH, Neulehenstraße 8a 33790 Halle (Westfalen), Deutschland br@wirelessconsulting.de

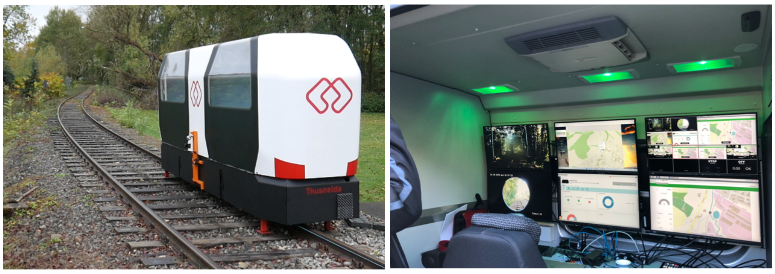


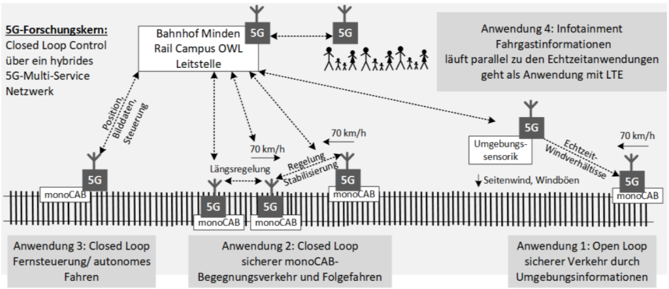
Abb. 1: MonoCab und Fraunhofer Leitstand

2 Fahrbetriebs Use-Cases

Die Folgenden Fahrbetriebs Use-Cases sollen mittels 5G Campusnetz als Übertragungsmedium relasiert werden.

→ **Sicherer Verkehr durch Umgebungsinformationen**
→ **Sicherer MonoCab-Begegnungsverkehr und Folgefahrten**
→ **Fernsteuerung/ automatisiertes Zugvertrieb**
→ **Infotainment und Nutzerinteraktion**

Die Fahrbetriebs Use-Case erfordern ein Closed loop Regler zu Regler Vernetzung von zwei oder mehr MonoCabs. Für eine optimale Regelung ist hochzuverlässige Funkverbindung mit einer geringen Latenz notwendig, besonders um Folgefahrten oder Begenungsverkehr zu realisieren. Eine Übersicht der einzelnen Use-Cases ist in der 2



5G-Forschungskern:
Closed Loop Control über ein hybrides 5G-Multi-Service Netzwerk
Bahnhof Minden Rail Campus OWL Leitstelle
5G
5G
Anwendung 4: Infotainment Fahrgastinformationen läuft parallel zu den Echtzeitanwendungen geht als Anwendung mit LTE
Position, Bilddaten, Steuerung
70 km/h
Längsregelung
Regelung Stabilisierung
70 km/h
Umgebungs-sensorik
Echtzeit-Windverhältisse
70 km/h
Seitenwind, Windböen
5G
monoCAB
Anwendung 3: Closed Loop Fernsteuerung/ autonomes Fahren
Anwendung 2: Closed Loop sicherer monoCAB-Begegnungsverkehr und Folgefahren
Anwendung 1: Open Loop sicherer Verkehr durch Umgebungsinformationen

Abb. 2: 5G-SIMONE Use-Cases

Für die Fahrzeug<-> Fahrzeug Vernutzung ist die Vehicle to Vehicle (V2V) sowie Vehicle to Infrastructure (V2I) eine hohe Relevanz.

Insbesondere die V2V Vernetzung im 5G Bereich verspricht eine besonders geringe Latenz im Vergleich zur konventionellen 5G Kommunikation zwischen 5GUEs und gNodeBs. [WLK18]

3 Erhobene Anforderungen an die MonoCab Funkkommunikation

Die Anforderungen an das Kommunikationssystem, welches das MonoCab mit anderen MonoCabs sowie der Infrastruktur verbindet ergeben sich aus physikalischen Randbedingungen und gesetzlichen Vorgaben. Aus den physikalischen Randbedingungen wie zum Beispiel den Bremsweg des MonoCabs der durch die Bremskraft und den Schienen-Radkontakt vorgegeben ist oder aber auch durch die gesetzliche Vorgabe der Kameraauflösung beim Remote fahren, lassen sich Anforderungen an ein Kommunikationssystem ableiten. Zur Anforderungsanalyse wurden alle Funktionen und Applikationen des MonoCabs betrachtet und daraus acht Datenprofile abgeleitet. Die Datenprofile beschreiben die Applikationen wie Videoübertragung oder Hotspot und Notrufdaten mittels Kommunikationstechnik Parametern. Ein Datenprofil kann mit der Framesize, der Kommunikationsrichtung (Uplink oder Downlink), der maximalen Datenrate und Latenz ein Abbild der Applikation bilden.

Use-Case	Framesize (Bytes)	Delay RTT	Uplink (kbit/s)	Downlink (kbit/s)
Management	128	<100 ms	100	100
PLC	128	<50 ms	50	50
Video Fernsteuerung	1300	150 ms	55000	100
Audio Notruf	200	<100 ms	200	200
Infotainment	640	<150 ms	1000	1000
Video Public Safety	1300	500 ms	50000	100
Hotspot	1300	Best Effort	Best Effort	Best Effort
V2X	128	<50 ms	500	500

Tab. 1: Use-Case Anforderungen

Management

Unter Network Monitoring wird der Datenverkehr, der zum Betrieb von den Netzwerkteilnehmern Industrial Computer notwendig ist, bezeichnet. Dazu gehören unter anderem Protokolle wie das SNMP (Simple Network Management Protocol) , RADIUS (Remote Authentication Dial-In User Service) oder auch NTP (Network Time Protocol). Der Datendurchsatz wurde aus Datenlogs sowie Erfahrungen aus vergangen Projekten mit 100 kBit/s im Uplink sowohl auch im Downlink 100 kBit/s angenommen.

Control Data

Das MonoCab operiert automatisiert und kann ohne externen Eingriff fahren, bis ein Hindernis oder eine Fehfunktion erkannt wird. In diesem Fall kann sich ein externer Mitarbeiter aus einem Leitstand mit dem MonoCab verbinden und sofern möglich das

MonoCab manuell per Remote Control steuern. Für die niedrig latente Steuerung hat dieses Datenprofil der höchste Latenzanforderung. Der Datenaustausch erfolgt bei diesem Datenprofil zyklisch. https://www.overleaf.com/project/644b9c6ecc9dd84f779dd579

Video Fernsteuerung

Unter dem Datenprofil Video Fernsteuerung ist die Bitrate einer Schwarz-Weiß Kamera des Partners Deutsche Bahn Systemtechnik herangezogen worden. Für die Bitrate wurde eine Kameraeinstellung von 1920x1080 Pixel (HD) unkomprimiert genutzt.

In diesem beschriebenen Szenario wird die Übertragung in der Uplink Richtung benötigt, die 5G Campusnetzte sind hingegen öffentlichen Mobilfunknetzen durch variable Uplink-Downlink Ratio Konfiguration in der Lage hohe Uplink Datenraten zur Verfügung zu stellen. Die von Consumern genutzten öffentlichen Mobilfunknetze sind hingegen Downlink orientiert.

Audio Notruf

Das Datenprofil Audi Notruf umfasst die erforderlichen Daten im Up- und Downlink für eine Voice over IP Datenübertragung. Mit Hilfe des Audio Notruf soll im Störung- oder Gefahrenfall eine Full Duplex Verbindung zwischen MonoCab und Leitstand möglich sein.

Infotainment

Das Datenprofil Infotainment beinhaltet die Übertragung von Informationen welche zum Betriebsablauf des MonoCab notwendig sind wie die Nummer des MonoCabs, die Position des MonoCabs und die Fahrtrichtung. Mittels der Informationen kann landseitig ein Scheduling der MonoCabs erfolgen und ein Betriebskonzept umgesetzt werden. Im Fahrzeug selber sollen Informationen wie die nächste Haltestelle und der Fahrplan angezeigt werden.

Video Public Safety

Das Video Public Safety steht für eine Public Safety Überwachung, der Überwachung der Fahrgäste im Innenraum der Kabine. Die Public Safety Kamera ist ebenfalls mit einem Industriecomputer verbunden, welcher als RTSP Server agiert. Dadurch kann aus einem Leitstand auf den Kamerastream zugegriffen werden. Die Innenraumüberwachung unterliegt keinen Echtzeit Anforderungen und hat damit eine maximale Video Latenz von 500 ms.

Hotspot Data

Das Hotspot Data Profil umfasst einen WLAN Hotspot innerhalb des MonoCabs, diesen können Fahrgäste nutzen um weitere Informationen zum MonoCab Fahrplan zu erfragen oder um einen WLAN Traffic offload durchzuführen. Der WLAN Hotspot wird durch die WAN (Wide Area Network) Verbindung des 5G-Campusnetztes zur Verfügung gestellt und stellt eine Best Effort Datenrate sowie Latenz zur Verfügung.

V2X Data

Das Datenprofil V2X Data beschreibt die Kommunikation zwischen MonoCab und MonoCab sowie einem Leitstand. Der Betrieb von dem MonoCab benötigt den Eingriff von einer Leitstelle in Störungsfällen oder Gefahrensituationen. Diese Leitstelle überwacht den ordnungsgemäßen Betrieb des MonoCabs und kann im Fehlerfall das Fahrzeug Remote steuern. Dazu ist eine Übertragung von Daten wie die Fahrzeugposition, Geschwindigkeit sowie MonoCab spezifische Regelungsparameter elementar. Neben der Kommunikation mit einem Leitstand berücksichtigt dieses Datenprofil eine Kommunikation zwischen zwei oder mehr MonoCabs. Die Fahrzeuge tauschen Fahrzeugstatus wie Fahrzeugposition und Regler Werte unter einander aus und bilden so eine Closed-Loop Regler Vernetzung über ein 5G-System. Zusätzlich zu der Funkkommunikation über 5G gibt es die V2X Technologie. Der V2X Standard beschreibt die Kommunikation von Fahrzeug zu Fahrzeug oder Fahrzeug zu Infrastruktur ohne Verwendung von einer Netzwerkmanagement Hardware wie einem WLAN Acess Point oder einem 5G Core. Die Technologie kann in Deutschland im Frequenzband von 5.9 GHz genutzt werden, dieses Frequenzband ist dabei von dem Standard IEEE 802.11p sowie 5G CV2-X nutzbar.

4 Planung der 5G Netzwerkabdeckung an der Teststrecke für den Feldtest

Das MonoCab Testfeld im Extertal, welches mit dem 5G Campusnetz abgedeckt werden soll ist 400 Meter lang. Damit im gesamten Gleisbereich eine hinreichende 5G Netzabdeckung verfügbar ist, müssen die Radio Units entsprechend platziert und parametrisiert werden. Die Dimensionierung des 5G-Campusnetztes erfordert im Optimalfall einen Empfangspegel am 5G UE, welcher höher ist als die Receiver Sensitivity des 5G UE Transceiver Chips ist. Zur Planung der Auslegung des 5G Campusnetztes wird die Netzabdeckung anhand einer Simulation für den Downlink ausgerechnet. Dazu wurden die Modelle aus dem Technical Report der 3GPP genutzt.[3G19a] Diese Modelle wurden aus Berechnungen und Labormessungen hergeleitet und versprechen eine hohe Annäherung an reale Umgebungen. In dem ersten Schritt werden die Pfadverluste berechnet und im zweiten Schritt mit dem Realwert verglichen.

Centerfrequency = 3.750 GHz maximale Distanz= 400 m x$_{\sigma LOS}$= 1.7 dB Receiver Sensitivity= -84,5 dBm

$PL'_{RMa-NLOS} = 161.04 - \log_{10} *(5m) + 7.5 \log_{10}(5m) - (24.37 - 3.7(5m/10m) \log_{10} *(10m) + (43.42 - 3.1 \log_{10}(10m))(\log_{10}(400.09m) - 3) + 20 \log_{10} *(3.75GHz) - (3.2(\log 10(11.75 * 1.5m))^2 - 4.97)$

Empfangspegel am Receiver= $gNodeBSendeleistung - PL'_{RMa-NLOS}$

Empfangspegel am Receiver @ 400m= $24dBm - 133.31dB = -109.31dBm$

Empfangspegel am Receiver @ 200m= $24dBm - 121.18dB = -97.189dBm$

Wie oben beschrieben, wurde eine 5G Lizenz für die Testmessung bei der BNetzA beantragt. Dabei wurden 100 MHz im Frequenzbereich von 3,7 GHz bis 3,8 GHz genehmigt. Daraus ergibt sich einer Center Frequency von 3,75 GHz.

Die Betrachtung des None Line of Side Pfadverlust mit einem Rural Macro Modell zeigt, dass die theoretische Empfangsfeldstärke nach 400 m am 5G UE -109 dBm beträgt. Hinzu kommen noch weitere Einflussfaktoren wie die Dämpfung der MonoCab Kabine welche das Signal ebenfalls abschwächt. Erste Tests mit einer Remote Radio Unit zeigen im letzten Bereich der Teststrecke Verbindungsabbrüche, bedingt dadurch das der Empfangspegel am 5G-UE zu gering ist, können keine 5G User- sowie Control Plane Daten mehr empfangen und gesendet werden.

Eine weitere Kalkulation mit zwei Remote Radio Units ergibt eine deutlich bessere Empfangsfeldstärke am 5G-UE. Wenn zwei Remote Radio Units im Abstand von 200 Meter eingesetzt werden, ergibt sich jeweils nach 200 Metern ein minimaler Empfangspegel von -97 dBm. Der Versuch auf dem Testfeld zeigt keine Verbindungsabbrüche mehr und ein Video Teststream über das 5G-System läuft ruckelfrei. Da das verwendete Nokia NDAC System ein Single Cell System ist, sind beide Radio Units in der gleichen physikalischen Zelle eingebunden und führen unter Bewegung der 5G-UEs somit ein intra cell handover aus.

Im Folgenden wurde das 5G Campus Netzwerk vor Ort aufgebaut und eine Network Coverage Survey durchgeführt. Die Survey erfolgt mittels Mess Software auf einem Tablet, welches mit dem 5G-System verbunden ist. Die Survey Software auf dem Tablet nimmt die Empfangswerte vom 5G-Chip des Tablets entgegen und kartiert diese. In der Survey wird ausschließlich der 5G Downlink betrachtet. Als Metrik für den Empfangspegel wird der RSRP Wert von dem Secondary Synchronization Signal genutzt. Die Abbildung 3 zeigt die verschiedenen Signalpegel an der Strecke mit zwei installierten ASiR-pRRH Radio Units. Dabei ist zu sehen das nahe der Remote Radio Unit der Empfang bei -68 dBm liegt, dieser Wert ist nahezu der bestmögliche Empfangspegel. Außerdem lässt sich beobachten das sich der Empfangspegel mit steigender Entfernung zur Radio Unit verringert. Die Survey Ergebnisse zeige außerdem, dass die Simulationsergebnisse sehr nah an den gemessenen Werten liegen.

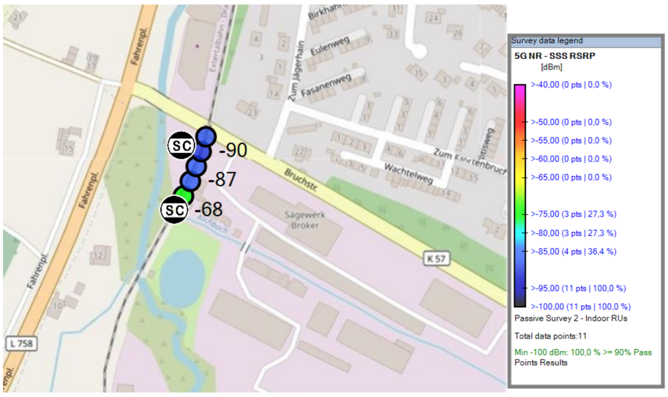


Abb. 3: Coverage Survey auf dem MonoCab Testfeld 1

5 Durchführrung und Ergebnisse der Feldtest

Nachdem die Funktionalität der Applikationen erprobt wurde, soll die Leistungsfähigkeit des 5G Campusnetzes gemessen werden. Die Leistungsbewertung wie z.B. die End-to-End Latenz kann bei einer Videoübertragung durch eine Glass to Glass Messung gemessen werden. In diesem Verfahren filmt die Kamera einen Monitor auf dem eine Atomuhrzeit angezeigt wird und überträgt den Videostream auf einen Ausgabemonitor. Auf dem Ausgabemonitor muss ebenfalls eine Atomuhrzeit dargestellt werden, dann kann durch ein Einfrieren des Bildschirms die Latenz zwischen Videoaufnahme und Empfang gebildet werden. Das Verfahren eignet sich sehr gut für die Ermittlung von Video Latenzen jedoch aber nicht von der Ermittlung der Latenz bei Datenübertragungen, daher wir hier ein anderes Mesverfahren benötigt. Eine gute Vergleichbarkeit bietet das ITU-T Standard Messverfahren ITU-T y.1564. In diesem Verfahren können die Datenprofile übertragen und ausgewertet werden. Das Messverfahren besteht aus zwei Messungen, zum einem werden alle Datenprofile sequentiell ausgeführt und unabhängig voneinander ausgewertet. Dabei werden die einzelne Datenprofile Schrittweise ausgeführt und die Datenrate iterativ erhöht, die Schritte sind 50%, 75%, 90% der Datenrate und 100 % der Datenrate. Die Zweite Messung überträgt alle Datenprofile parallel über das 5G-System und stellt die maximale Datenrate eines MonoCabs dar. Dieser Test ermöglicht die Auswertung der Quality of Service pro Datenprofil, dadurch kann eine Beeinflussung unterschiedlicher Datenprofile sicher erkannt werden.

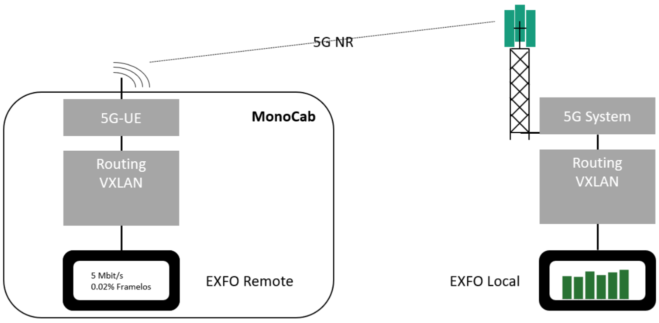


Abb. 4: Messaufbau am 03.10.2023

Funktechnisch wurde diese Strecke mit einem Nokia NDAC 5G-System SA ausgeleuchtet, welches mit einer physikalischen Radio Cell ausgestattet ist indem zwei ASiR-pRRH Radio Units platziert wurden. Die ASiR-pRRH Radio Units verfügen über eine maximale Sendeleistung von 250mW/ 24dBm. Diese Sendeleistung ist für einen Empfangspegel von > -100 dBm entlang der Strecke ausreichend. Dem Datenblatt der Modemhersteller zufolge soll am Modem jedoch ein Empfangspegel von >-84.5 dBm anliegen [Que22]. Die Messergebnisse zeigen das eine Konnektivität zwischen dem gNodeB und dem Modem auch zwischen der empfohlenen Empfangsfeldstärke und dem Verbindungsverlust besteht, diese jedoch keine hohe sowie zuverlässige Dienstgüte aufweist.

Anhand der Ergebniss Plots, welche die einzelnen Datenprofile in Abhängigkeit der Empfangsfeldstärke aufzeigen, lässt sich ableiten das die RTT Latency mit fallender Empfangsfeldstärke steigt. Dieses Verhalten zeigen die Abbildung 5, es werden die eingeführten Datenprofile auf der X-Achse und die RTT Latenz auf der Y-Achse aufgetragen. Die Abbildung zeigt die Messergebnisse des Service Configuration Test bei einem Empfangspegel am 5G UE von -50 dBm, -85 dBm und -100 dBm. Die schwarz gestrichelte Linie in der Abbildung ist bei 50 ms gezogen und dient als Orientierungshilfe sowie als maximale Latenzgrenze für das Datenprofil PLC sowie V2X. Weiterhin zeigt das Ergebnis in 5, dass das 5G-Campusnetz bei einem Empfangspegel von -100 dBm nicht in die notwendige Dienstgüte für das Datenprofil Video Fernsteuerung aus Tabelle 1 erreicht.

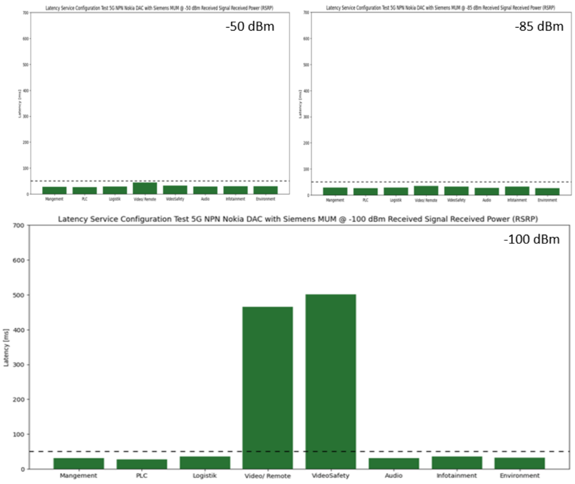


Abb. 5: Service Configuration Test -50dBm, -85dBm, -100dBm

In der Abbildung 6 werden die Ergebnisse des Service Performance Test bei einem Empfangspegel von -50dBm, -85 dBm und -100 dBm am 5G UE dargestellt. Die Betrachtung der Latenz der Datenprofile zeigt das mit einem Empfangspegel von -50 dBm sowie -85 dBm die Dienstgüte aller Datenprofile eingehalten werden kann. Bei einem Empfangspegel von -100 dBm ist jedoch zu sehen das keine Dienstgüte mehr eingehalten werden kann. Außerdem sind die Latenzen der Datenprofile alle gleich groß, was eine Überlastung des 5G UEs vermuten lässt.

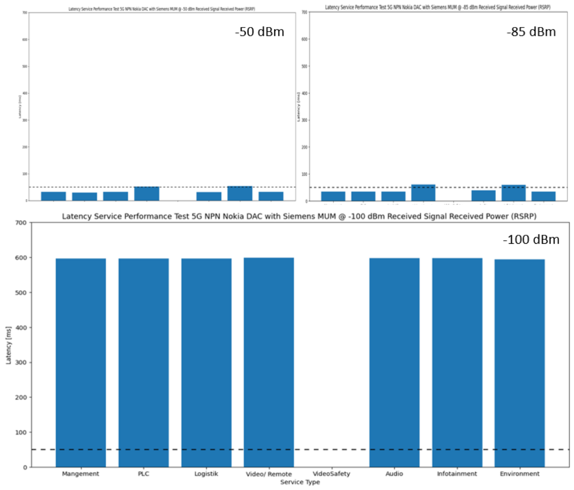


Abb. 6: Service Performace Test für -50dBm, -85dBm und -100dBm

6 Fazit

Zusammenfassend kann gesagt werden, dass sich das 5G Campusnetzwerk im Testfeld ähnlich verhalten hat wie die Simulationsergebnisse in Hinsicht auf die Netzwerkabdeckung im Verhältnis zur Reichweite. Die zuvor definierten Anforderungen konnten unter Rücksicht auf einen ausreichenden Empfangspegel erfüllt werden. Ein Ergebnis dieses Beitrags ist also das der Empfangspegel ausreichend stark sein muss um die gegebene Latenz nicht zu überschreiten. Die Ergebnisse zeigen auch das verschieden Datenprofile und damit verschiedene Protokolle nicht priorisiert werden.

Ausblickend auf weitere Funktionen aus dem 3GPP Standard wie Network Slicing und URLLC soll im weiteren Verlauf die Priorisierung auf PDU Session oder Protokoll Level untersucht werden. Außerdem sind weitere Feldtest im Projekt 5G-SIMONE mit Radio Units für eine Fernfeld Funkausleuchtung geplant.

Literaturverzeichnis

[3G19a] 3GPP: Technical Specification Group Radio Access Network; Study on channel model for frequencies from 0.5 to 100 GHz. Technical Report (TR) 38.901, 3rd Generation Partnership Project (3GPP), 09 2019. Version 15.1.0.

[3G19b] 3GPP: Technical Specification Group Services and System Aspects;Mobile Communication System for Railways. Technical Specification (TS) 22.289, 3rd Generation Partnership Project (3GPP), 12 2019. Version 17.0.0.

[G.] G.Fuchslocher: , BMW schickt Autos autonom durchs Werk. https://www.automobil-produktion.de/produktion/bmw-schickt-autos-autonom-durchs-werk-637.html. (Accessed: Feb. 17, 2023).

[Que22] Quectel Wireless Soultions Co., Ltd. RM50xQ Series Hardware Design, 2022. Status released.

[RP18] Rao, Sriganesh K; Prasad, Ramjee: Impact of 5G technologies on industry 4.0. Wireless personal communications, 100:145–159, 2018.

[WLK18] Wang, Jiadai; Liu, Jiajia; Kato, Nei: Networking and communications in autonomous driving: A survey. IEEE Communications Surveys & Tutorials, 21(2):1243–1274, 2018.

http://dx.doi.org/10.25673/111640

Virtual Automation Network Simulation (VANSIM): A tool for shared 5G campus networks in industrial working and co-working spaces

Parva Yazdani, Gustavo Cainelli, Dr.-Ing. Lisa Underberg[1]

Abstract: 5G networks are currently adapted more and more in industrial applications due to their high-speed data transfer, low latency, and increased capacity. To address growing demands and increase cost-effectiveness, the idea of shared 5G campus networks has emerged, akin to co-working spaces known from office environments. This is especially beneficial for small and medium enterprises in industrial parks looking to incorporate 5G technology into their automation needs. In order to ensure the network can meet the diverse application requirements of all companies involved, a simulation platform called Virtual Automation Network Simulation (VANSIM) is under development. VANSIM aims to allow companies to validate and test the feasibility and compatibility of shared 5G networks pre- and post-installation from the perspective of the automation applications. This paper focuses on validating this simulation platform for passive environmental influences by comparing its results to real-world measurements.

Keywords: Industrial wireless networks, Shared campus network, Private networks, Non-public networks, Co-working spaces, Network management, Network simulation, 5G network, Application requirements.

1 Introduction

In recent times, considerable attention has been directed toward 5G campus networks due to their remarkable attributes compared to preceding cellular network generations. These intrinsic capabilities have opened doors to various applications, such as industrial automation, remote healthcare, and autonomous vehicles. The surging demand for 5G connectivity, coupled with its substantial costs and environmental ramifications, has prompted the investigation regarding the concept of shared 5G campus networks. These networks enable multiple companies to share a single 5G infrastructure, particularly relevant for smaller enterprises in clusters like industrial parks (dt. "Gewerbegebiet").

Utilizing a shared network offers cost-saving benefits for individual companies, encompassing both initial installation and ongoing operational and maintenance expenses. However, the suitability of a 5G campus network within an industrial park relies on its capacity to meet the diverse application requirements of all companies concurrently. For example, one company may rely on a network of RFID tags and readers to manage logistics, another needs to monitor equipment remotely using IoT devices, such as temperature sensors. The first requires low-latency communication to ensure real-time updates, while the latter needs to transmit large volumes of data. These unique

[1] Institut für Automation und Kommunikation (ifak e.V.), ICT & Automation, Werner-Heisenberg-Straße 1, 39106 Magdeburg, parva.yazdani@ifak.eu, gustavo.cainelli@ifak.eu, lisa.underberg@ifak.eu.

communication needs must be accommodated within the shared 5G campus network infrastructure. Therefore, it is essential to assess the feasibility of implementing such a shared network before installation and thoroughly test its performance afterward. This research aims to facilitate this evaluation process by introducing a simulation platform called Virtual Automation Network Simulation (VANSIM).

In the following sections, we introduce a general model for shared 5G campus networks. This model sets the foundation for our document's primary focus, which involves the modeling and analysis of passive environmental factors affecting shared 5G communication within industrial and co-working spaces. We then offer an overview of our investigations into channel models and present an in-depth analysis of our findings.

1.1 Related work

In the previous work [YCU23], the authors introduced VANSIM and investigated its validity against passive environmental conditions. Their findings suggested that modifications to current channel models, particularly in terms of fading calculations, were necessary. Building upon this foundation, our work follows a similar structure and aims to further validate and improve VANSIM.

While previous research, such as [Or19] and [AC19], has motivated the deployment of non-public 5G networks, and web articles like [Gr18] have highlighted the benefits of shared network usage, and papers like [Dü19] have delved into signal propagation within indoor industrial environments, our primary goal sets us apart. Our ultimate objective is to develop a platform that not only evaluates propagation conditions but also focuses on the broader perspective of applications. In essence, we are providing a holistic assessment from the applications' perspective, considering their diverse needs and the specific conditions in which they operate.

1.2 Area of consideration

The findings outlined in this paper are derived from the research project "5G Industrial Working and Co-Working Space (5GIWCoW)". The project consortium comprises companies located within the Technology Park Ostfalen (TPO) close to Magdeburg, as illustrated in Figure 1. The specific project partners are enumerated in [5G23]. Presently, preparations are underway to deploy a dedicated, non-public 5G network intended for shared utilization. Within this context, the project partners are strategically planning the implementation of 5G communication for a pilot system, encompassing the following illustrative applications such as programming of welding robots and enabling remote control of robotic systems.

TPO has served as the geographical domain under consideration for the validation of the VANSIM in the previous work of the authors [YCU23]. In the present work, new measurements have been conducted at Wissenschaftshafen in Magdeburg. These measurement results have been used for further validations and modifications of the physical layer representation in VANSIM.

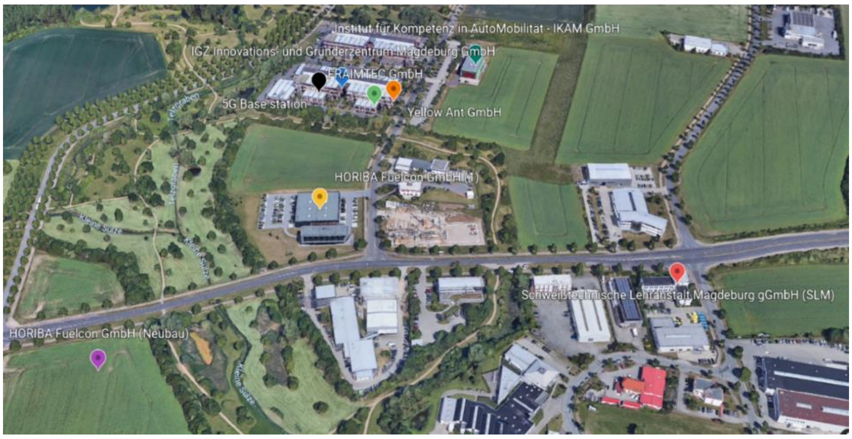

Institut für Kompetenz in AutoMobilität - IKAM GmbH
IGZ Innovations- und Gründerzentrum Magdeburg GmbH
FRAIMTEC GmbH
5G Base station
Yellow Ant GmbH
Telzgraben
HORIBA Fuelcon GmbH (1)
Kleine Sülze
Kleine Sülze
Schweißtechnische Lehranstalt Magdeburg gGmbH (SLM)
HORIBA Fuelcon GmbH (Neubau)
Kleine Sülze

Figure 1: Position of the project participants and the shared 5G base station (gNB) [Go22].

This work was funded in parts by the project “5GIWCoW” under contract 165GU041D, funded by the Federal Ministry of Transport and Digital Infrastructure, Germany.

2 Methodology

2.1 General

The approach begins with the development of a model for the shared use of a non-public 5G network inspired by a model of the wireless industrial automation system outlined in [IE22]. This model is then implemented using an open-source network simulator. Subsequently, it has been examined how environmental factors such as intervisibility, mobile or static objects, and natural environmental conditions impact the shared 5G network’s performance in an industrial park. This would help us to validate the physical layer of the channel model in VANSIM and progress toward our goal, which is a reliable platform that can consult the industrial park members on the effectiveness and feasibility of the shared network before and after installation.

2.2 Model for 5G communication in industrial working and co-working spaces

Drawing from a general model of the wireless industrial automation system outlined in [IE22], we have developed a tailored model designed for the deployment of shared non-public 5G networks within industrial parks. In contrast to the original model, which accounts for the coexistence of various wireless systems within a unified radio environment, the model, depicted in Figure 2, is designed for a single 5G network shared among different companies, with distinct applications and distinct radio environments. In this model, we view the various applications in the industrial park as a wireless industrial automation system. This system encompasses one or more distributed automation systems, multiple radio environments, and a singular 5G communication infrastructure. Each distributed automation system serves as a representation of a specific application, with the potential for multiple applications to be associated with a single company.

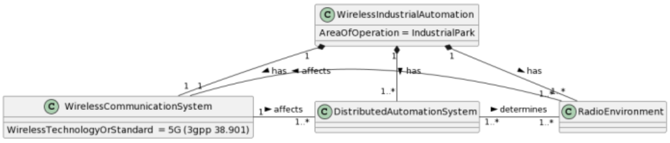

WirelessIndustrialAutomation
AreaOfOperation = IndustrialPark
1
1
1
has
affects
has
has
1
1
1..*
1..*
1..*
WirelessCommunicationSystem
WirelessTechnologyOrStandard = 5G (3gpp 38.901)
1
affects
1..*
DistributedAutomationSystem
determines
1..*
1..*
RadioEnvironment

Figure 2: The class model of the wireless industrial automation system based on [IE22].

Given the possibility of having varied environmental conditions within each distributed automation system, the need arises for one or more models of these radio environments. For instance, if one automation system involves temperature and humidity sensors in both indoor and outdoor environments when modeling the transmission medium, it's essential to represent both environments accurately, as they vary significantly. The specifications of the distributed automation system, for example, the positioning of wireless devices, primarily determine the behavior and attributes of the radio environment. Conversely, changes in the radio environment have an impact on the 5G communication system, indirectly influencing the distributed automation system.

The class models of the 5G communication system, distributed automation system, and radio environment and their interfaces are demonstrated in Figure 3. According to this model, the radio environment influences can be classified into two categories: active and passive influences. Active influences arise from disruptions caused by external wireless devices or equipment emitting electromagnetic waves within the relevant frequency range. In contrast, passive influences are a consequence of the physical environment, which includes factors like stationary and moving obstructions, reflective surfaces, and transmission distances. These passive influences result in path loss and fading effects. For this paper, our primary focus will be on exploring the passive influences.

The impact of the radio environment on the application can be assessed by examining network-related performance parameters, including the Signal to Interference plus Noise Ratio (SINR). These parameters are provided by the communication system's functions and can be quantified using specialized equipment, such as the Mobile Network Scanner (TSMA) from Rohde & Schwarz. TSMA measures signal strength independent of any device being active on the network. Furthermore, we can gauge the impact of the radio environment by considering application-related performance parameters, such as transmission time or update time. These parameter values are established in connection with logical links that interconnect locally distributed automation functions.

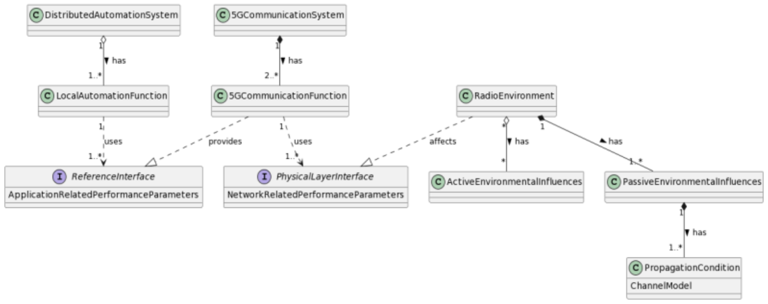


Figure 3: The class models of the WCS, DAS, and RE in accordance with [IE22].

2.3 Simulation platform for 5G communication in industrial working and co-working spaces

VANSIM is designed to be adaptable for various use cases, whether in an industrial park with multiple companies or a single company with diverse applications. The foundation is built upon the open-source 5G new radio network simulator, Simu5G, which empowers users to create new modules, algorithms, and protocols [Si23]. This framework is inherently well-suited for implementing the shared 5G communication system model and is built on top of OMNeT++.

The network simulators provide a virtual representation of the network infrastructure, protocols, and components involved in a 5G network deployment that replicates the behavior of a real 5G network. It enables users to evaluate the impact of different configurations or algorithms and test various scenarios without the need for real-world deployment. There exists plenty of 5G network simulators that could have been used to implement VANSIM such as Ns-3 [Ns23], NetSim [Bo23], and OPNET [Op23]. Among these, we have chosen OMNeT++ because of its application perspective which makes it well-suited for implementing the distributed automation system model, presented in Section 2.2.

Our objective is to evaluate the suitability of existing Simu5G channel models for the specific conditions in industrial working and co-working spaces in TPO. This section provides an overview of Simu5G and its channel models.

Simu5G supports both Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD) modes, supporting heterogeneous gNBs (macro, micro, pico, etc.). The simulator employs a customizable channel model at the physical layer, although it lacks the granularity of modeling resource elements, thus not including reference signals. Consequently, the received power calculations in Simu5G are independent of the specific resource element used for signal transmission.

The 5G device and gNB are implemented as compound modules, allowing connections to other modules for network composition. Both include the New Radio stack and all its sublayers. Additionally, Simu5G supports Carrier Aggregation (CA) and various numerologies.

Communication through Simu5G occurs via an OMNeT++ message exchange when a MAC Protocol Data Unit (PDU) is transmitted from sender to receiver. Upon receiving the message, the receiver employs various channel models to determine the received power. These channel models are capable of computing path loss, slow fading, and fast fading. The specific path loss calculation utilized is contingent upon the chosen scenario. Available scenarios in Simu5G's NR channel model are Urban Macrocell, Urban Microcell, Rural Macrocell, and Indoor [Side23].

The path loss models consider two critical parameters for each scenario: distance and Line-of-Sight (LOS) or Non-Line-of-Sight (NLOS) conditions. These are modeled according to 3gpp documentation [3G22]. The log-normal distribution model calculates slow fading, with a specified standard deviation for both LOS and NLOS scenarios. Fast fading in Simu5G primarily depends on the doppler shift induced by a moving 5G device. In the absence of a doppler shift, and with a stationary 5G device, the fading remains constant.

3 Evaluation of models for passive environmental influences

3.1 Conductions of real-world measurements

The mobile network scanner measurements provide essential values for assessing VANSIM. These measurements indicate the Reference Signal Received Power (RSRP) of the second synchronization signal within the Physical Broadcast Channel (PBCH) block. This parameter, termed SS-RSRP, reflects the power level of the synchronization signals received by the 5G device from the base station; further details can be found in [Ko19].

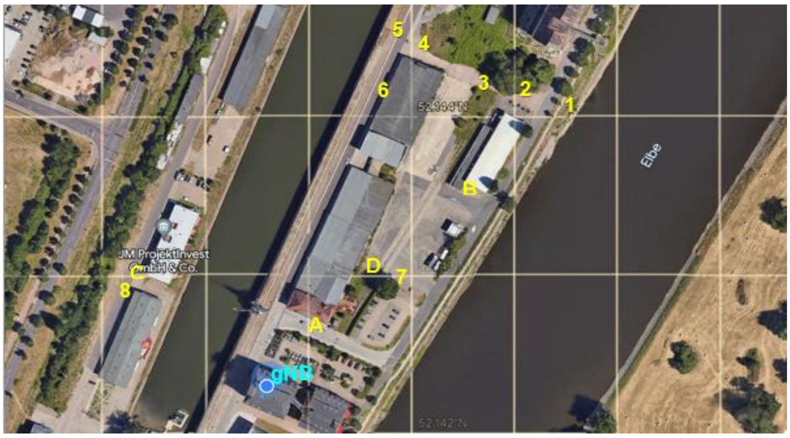


Figure 4: Different measurement points at Wissenschaftshafen [Go23].

For the measurement campaign conducted at Wissenschaftshafen, the TSMA was specifically set up to measure 5G band n78. This band, with compatibility spanning from 10 to 100 MHz and a carrier frequency range of 3.3 to 3.8 GHz, aligns with the parameters of Wissenschaftshafen‘s network. Following these configurations, the base station was positioned on one of ifak's balconies with a height of 17.76 m, and the TSMA antenna was placed in various LOS and NLOS locations to measure SS-RSRP. The distinct measurement points, depicted in Figure 4, will be utilized in the subsequent sections to validate the incorporation of passive environmental influences in VANSIM.

3.2 Channel modeling in VANSIM

In this section, the modification and configuration of the channel model implemented in Simu5G are presented. These enhancements are guided by a comparative analysis between the simulation results and empirical measurement results obtained within the Wissenschaftshafen area, aiming to closely replicate real-world conditions. As concluded in [YCU23], for the stationary devices, the simulation results were found to be time invariant. This has shown the limitation in the existing Simu5G, which does not account for the dynamic influence of the physical environment on the signal strength, i.e. it needs to be time variant. The channel model has been modified further in this work to incorporate these environmental influences.

Table 1 is a list of the main parameters in Simu5G that are set according to the 5G network’s configuration at Wissenschaftshafen.

Parameter name	Value/unit
gNB transmit power (P_{tx})	23 dBm
Bandwidth	100 MHz
Subcarrier spacing	30 KHz
TDD ratio (DL:UL)	4:1
Carrier frequency (f_{c})	3.75 GHz

Tab. 1: Main parameters of the channel model.

The Received Signal Reference Power (RSRP) is determined by equation (1). The sender’s transmit power is determined by P_{tx} and is adjustable. The computation of distance-related path loss is predicated using the eta power law (a_{pl}) as, for example, described in [UWR20] shown in equation (2). The main reason is that the a_{pl} model can easily be adapted to a specific environment by adapting eta (η) as opposed to using e.g. free space loss. Subsequently, the slow fading phenomenon (a_{sf}) is characterized by a log-normal distribution [Qu23] as shown in equation (3). Noise Figure (F) and implementation loss (I) are also defined in this calculation. These refinements collectively contribute to a more realistic representation of the wireless communication channel within the Simu5G framework.

$$\text{RSRP} \;=\; P_{\text{tx}} - a_{\text{pl}} - a_{\text{sf}} - F - I \tag{1}$$

$$a_{\text{pl}} = \eta \cdot 10\log_{10}\left({}^{d}\!/_{\text{m}}\right) + 20 \cdot \log_{10}\left({}^{f_{\text{c}}}\!/_{\text{GHz}}\right) + 32.44 \tag{2}$$

$$a_{\text{sf}} = 10 \cdot \log_{10}\left(e^{\mathcal{N}(\mu,\sigma)}\right);\;\; \forall\sigma = \sqrt{\log(1+\sigma_{\text{l}}{}^{2})}, \mu = 0 \tag{3}$$

3.3 Results and Discussion

In this section, first, the parameter η for each measurement point is estimated. This estimation is performed by excluding the a_{sf} calculation. Subsequently, we proceed to estimate the variance of the log-normal distribution (σ_{l}) for each data point, accomplished by omitting the a_{pl} calculation. The detailed findings derived from these analyses are presented in Table 2 for reference and examination.

Inter-visibility	**Measure-ment point**	**η (F+I included)**	**η (F+I = 30 dB)**	**η (F+I = 50 dB)**	**σ_{l}**
NLOS	Point 1	5.25	3.79	2.83	0.6
	Point 2	4.89	3.44	2.48	1.1
	Point 3	4.86	3.41	2.44	0.6
	Point 4	5.22	3.78	2.82	0.3
	Point 5	4.7	3.25	2.29	0.2
	Point 6	5.3	3.84	2.87	0.2
	Point 7	5.65	3.79	2.54	1.3
	Point 8	6.54	4.77	3.59	0.3
LOS	Point A	5.45	3.11	1.54	0.1
	Point B	4.62	3.05	1.99	0.3
	Point C	5.76	4.03	2.88	0.6
	Point D	5.65	3.81	2.57	0.7

Tab. 2: Estimated parameters for measurement points.

To prove the robustness and consistency of the calculations, we calculated η in three modes as shown in Table 2. In the first mode, we factored in the *F* and *I* within the computation of the a_{pl}. Subsequently, in the second and third modes, we explored scenarios where *F+I* was estimated at 30 and 50 dB, respectively. The results have shown a consistent distribution across all three modes. This confirms the reliability and uniformity of the RSRP calculations within the VANSIM. Besides, the value of η in LOS scenarios is supposed to be around 2 which results in the free path loss. With this fact, it can be concluded that the approach with *F+I* = 50 dB is more reasonable than the other two.

In Figure 5, the distribution of one NLOS and one example LOS point is illustrated. The blue distribution shows the real-world RSRP behavior, while the orange distribution visualizes the predicted RSRP behavior. It can be seen that if we use the estimated parameters extracted from the same point, the calculated RSRP distribution looks close to the measured one.

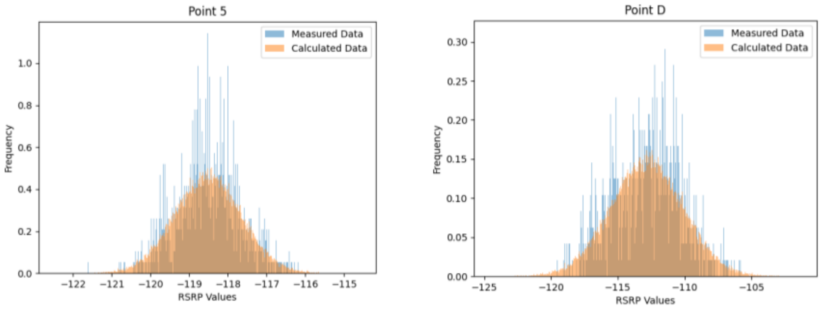


Figure 5: Points 5 and D simulation results vs TSMA measurements using corresponding estimated parameters.

On the other hand, if we take the average of estimated parameters derived from two NLOS points to predict the RSRP at a third NLOS point, the distributions are different. Figure 6 depicts the calculated RSRP values of points 1 and 5, using the estimated parameters obtained from measurements conducted at points 2 and 4. An intriguing observation emerges from the comparison: the prediction accuracy for point 1 appears to be better than that for point 5. This could be due to the passive environmental influences, such as signal reflections originating from structures located on the opposite side of the harbor, which exclusively affect point 5. However, the inaccuracy is – from a wireless system's perspective – reasonably small.

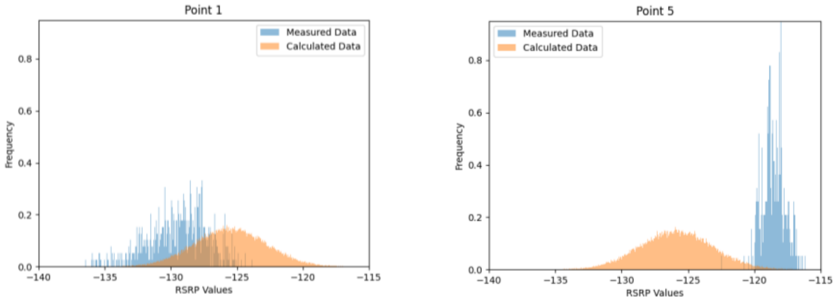


Figure 6: Points 1 and 5 simulation results vs TSMA measurements using the average of the estimated parameters for points 2 and 4.

From these results, it can be concluded that the conventional categorization of LOS and NLOS could be extended to achieve a representation of real-world conditions that are accurate enough to ensure a simulative validation adhering to industrial applications' requirements. To enhance the predictive precision, additional categories are to be introduced that account for diverse passive environmental influences including reflection, scattering, and diffraction. Another category could be the obstructed LOS. Point 3 is an example of such a propagation condition. This nuanced categorization promises a more comprehensive and realistic modeling of wireless signal propagation in complex environments.

4 Conclusions

In order to predict the behavior of the network it is important to have a stochastic estimation of the passive environmental influences. This aspect has been implemented and investigated in the present paper. Specifically, slow fading has been implemented to mimic temporal variations through a lognormal distribution. This serves as a step towards achieving a more realistic representation of wireless channel characteristics.

To validate these modifications, the channel model was assessed using the measurements obtained from the 5G network deployed at Wissenschaftshafen considering LOS and NLOS locations. The consistency of the RSRP calculations has been validated and further steps have been identified.

Our future work will focus on specifying more categories and critical locations in our measurement campaigns to enhance the adaptability of our simulation. Moreover, we will validate VANSIM for application-related performance parameters using the Funk-Transfer-Tester (FTT) results. FTT, a device developed by ifak, assesses communication solutions for industrial automation applications by emulating the communication behavior of the application and measuring application-related performance parameters. An essential component of FTT is the Multiface, which integrates real devices with various communication interfaces into the test system and generates test data traffic [If23]. FTT tests in accordance with [Vd19] and [5G19]. See [Gr08] for more details. Finally, we aim to test the shared network concept through the use cases that consider several properties with different physical environments to validate VANSIM for shared network scenarios.

References

[3G22] 3GPP TR 38.901: 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on channel model for frequencies from 0.5 to 100 GHz, V 17.0.0, 2022.

[5G19] 5G-ACIA, Performance Testing of 5G Systems for Industrial Automation (White Paper), Tech. Rep. April, 2021, https://www.5gacia.org/publications/.

[5G23] 5G Industrial Working + Co-Working für den Mittelstand, https://5g-teleport.de/. Accessed: 16 May 2023.

[AC19] 5G-ACIA. 5G non-public networks for industrial scenarios. 5G Alliance for Connected Industries and Automation white paper. 2019.

[Bo23] https://www.boson.com/netsim-cisco-network-simulator. Accessed: 16 May 2023.

[Dü19] Düngen, M. et.al.: Channel measurement campaigns for wireless industrial automation. at – Automatisierungstechnik, 2019.

[Go22] Google Earth V 9.176.0.1, Technologiepark Ostfalen, Gewerbegebiet in Barleben Sachsen-Anhalt, Available at: https://earth.google.com/, Accessed: 24 November 2022.

[Go23] Google Earth V 10.38.0.0, Wissenschaftshafen in Magdeburg Sachsen-Anhalt, Available at: https://earth.google.com/, Accessed: 14 August 2023.

[Gr08] Gnad, A.; Rauchhaupt, L.: Multi-functional interface for test of industrial wireless solutions, 2008.

[Gr18] Grijpink, F. et.al.: Network sharing and 5G: A turning point for lone riders. McKinsey&Company, 2018.

[IE22] IEC 62657-3:2022 ED1, Industrial networks - Coexistence of wireless systems - Part 3: Formal description of the automated coexistence management and application guidance.

[If23] ifak e.V., https://www.ifak.eu/en/irl_magdeburg. Accessed: 30 October 2023.

[Ko19] Kottkamp, M., et.al.: 5G New Radio: Fundamentals, procedures, testing aspects. München: Rohde & Schwarz, 2019.

[NS23] https://www.nsnam.org/about/. Accessed: 16 May 2023.

[OP23] https://opnetprojects.com/opnet-network-simulator/. Accessed: 16 May 2023.

[Or19] Ordonez-Lucena, J. et.al.: The use of 5G Non-Public Networks to support Industry 4.0 scenarios. IEEE Conference on Standards for Communications and Networking, 2019.

[Qu23] https://www.ques10.com/p/47939/log-normal-shadowing-1/. Accessed: 07 November 2023.

[Si23] http://simu5g.org/. Accessed: 14 May 2023.

[Side23] http://simu5g.org/description.html. Accessed: 15 May 2023.

[UWR20] Underberg, L; Willmann, S; Rauchhaupt, L: Traffic model integration of a 5G system into industrial communication. 26 October 2020.

[Vd19] VDI/VDE, VDI/VDE 2185-4 Radio-based communication in industrial automation - Metrological performance rating of wireless solutions for industrial automation applications, Bd. Part 4, 2019.

[YCU23] Yazdani, P.; Cainelli, G.; Underberg, L.: Shared 5G campus network in industrial working and co-working spaces. In 24. Leitkongress der Mess- und Automatisierungstechnik AUTOMATION 2023, Baden-Baden 2023.

http://dx.doi.org/10.25673/111641

Performance Evaluation and Application of Real-Time Communication with 5G IIoT

Niladri Mondal[1], Dimitri Block [2], Björn Kroll[3], Florian Klingler[4]

Abstract: In communication systems, high data rates combined with low end-to-end latencies are prime necessities for allowing a wide variety of applications, e.g., streaming of video and data or in context of IoT Systems. In contrast, applications in Industry Automation require deterministic end-to-end latencies with guaranteed deadlines. In communication systems, data rates, reliability and the achievable end-to-end latency are often a trade-off (e.g., due to buffering of data, and overall systems-design). Further, most communication systems are optimized for high data rates only, yet, deterministic end-to-end latencies are required for most Industrial Use-Cases, which are still not considered well enough in research and standardization. In this paper we focus on low-latency communication, and, outline the importance of this research aspect. Consequently, we propose a novel Mini-Slot approach for 5G and beyond communication systems to tackle the problem of minimizing uplink- and downlink communication latencies in cellular networks under load. First evaluations of our approach in context of a feasibility study show promising results. As comparison in realistic experiments with Rel-15-based 5G Commercial off-the-shelf (COTS) hardware, a baseline scenario (unoptimized) shows a maximum latency up to 49.04 ms. In contrast to that, our novel mini-slot approach allows to lower the maximum end-to-end communication latency to 15.51 ms. This way, our mini-slot approach constitutes as enabler for low-latency communication by using Rel-15-based 5G COTS and User Equipment (UE) hardware for industrial use-cases, without the need to wait for further releases of 5G systems.

1 Motivation

5G New Radio (NR) is the new radio access technology developed and standardized by Third Generation Partnership Project (3GPP). It is a convergent wireless technology enabling diverse use cases in industrial applications. It can support real-time communication, which is essential for many industrial scenarios that require low latency and high reliability. One of the potential applications of 5G-based real-time communication is fieldbus communication, which are network protocols for connecting sensors, actuators, controllers and other devices in industrial automation systems. By using 5G instead of wired fieldbus, industrial applications can benefit from the wireless advantage for flexibility, mobility or motion in plants and machines in e.g. discrete manufacturing, intralogistic applications and process industry.

[1] TU Ilmenau, IoT Engineering, Ehrenbergstraße 29, 98693 Ilmenau, Germany. mondal@wnc-labs.org

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

[3] Fraunhofer IOSB-INA, Campusallee 1 , 32657 Lemgo, Germany. bjoern.kroll@iosb-ina.fraunhofer.de

[4] TU Ilmenau, IoT Engineering, Ehrenbergstraße 29, 98693 Ilmenau, Germany. klingler@wnc-labs.org

5G is a novel technology that defines various features to address different use cases. However, early Commercial off-the-shelf (COTS) 5G networks and devices utilize different subsets of 5G features and therefore act differently. Thus, it is essential to conduct experiments in COTS 5G networks and devices to evaluate their real-world performance and limitations.

2 Problem Description

A fieldbus is a communication system that connects various devices and sensors in an industrial network. There are different types of fieldbus protocols, such as PROFINET IRT, EtherCAT, Sercos III, and Ethernet Powerlink [Wo17]. These protocols have some minor differences in their data formats, addressing schemes, and error handling mechanisms, but they all require real-time performance from the physical layer of the network. This means that the data transmission and reception must meet strict timing constraints to ensure the reliability and safety of the industrial processes. Therefore, when considering the use of 5G for fieldbus applications, it is important to understand the limitations and challenges of achieving real-time communication over a wireless medium.

In order to quantify real-time communication limitations, we focus on two aspects which outline prime metrics to tackle real-time communication in an industrial context: The timing added for packet transmission and their reliability of successful transmission. The timing aspect can be measured by the One Way Latency (OWL), while the reliability aspect can be measured by the packet error rate. Therefore, by quantify and optimizing the transmission latency and the packet error rate, the real-time capability for fieldbus protocols can be validated.

A pre-configured 5G wireless communication technology is implemented in hardware setups cloning a smart warehouse scenario described in Fig. 1. The test scenario consists of a controller such as a Programmable Logic Controller (PLC) and a PROFINET RT device (CC A&B). The device is attached to one of the moving columns of the smart warehouse. Communication between the PLC and the device takes place with the help of an applied industrial 5G IIoT solution. The end-to-end latency observed in the system is to be investigated, attempted to be reduced from it's default initial value, suffering least or no loss in channel throughput.

3 State of the art in 5G Communication

5G was distributed under 3GPP release 15. It is a key component of the 5G standard and is designed to support a wide range of use cases and deployment scenarios. It operates on a range of frequency bands from low-band frequencies to millimeter-wave frequencies. The low-band frequencies are below 6 GHz carrier range and are also known as sub-6 GHz band or Frequency Range (FR)1. The bandwidth for FR1 ranges from 5 MHz to 100 MHz. The

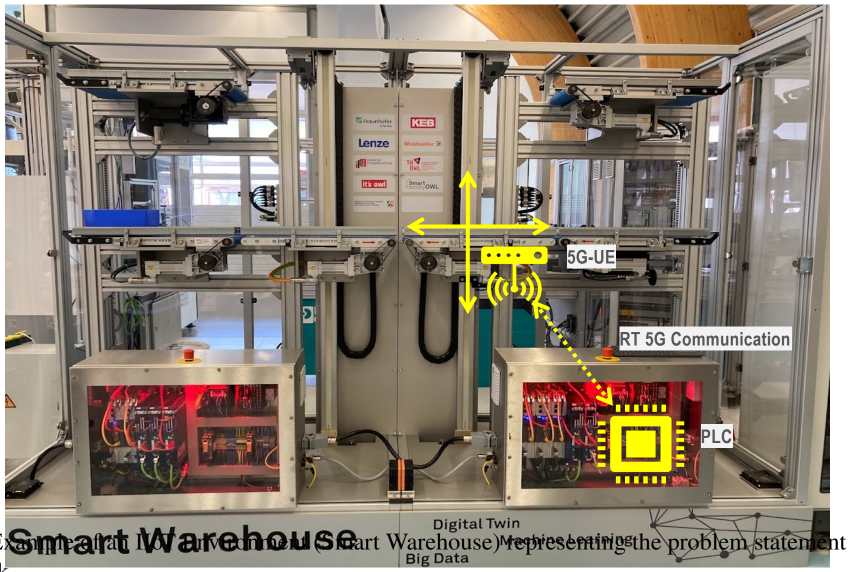


Fig. 1: Example of an IIoT environment (Smart Warehouse) representing the problem statement of this work.

millimeter-wave band operates above 6 GHz carrier frequency range. It is also called FR2 and includes the bandwidths of 50 MHz, 100 MHz, 200 MHz, and 400 MHz.

The millimeter-waves have high frequency range and because of this, the signals can easily get degraded in quality. 5G is designed to provide higher data rates, lower latency, and more reliable connectivity than previous generations of cellular networks which enables it to deliver better performance and efficiency in wireless communication.

A 5G frame is of duration 10 ms in time which is further divided into 10 sub-frames. Each of these sub-frames contains one or more number of slots. These slots accommodate the symbols which are meant to be transmitted over the 5G wireless channel. Each of the slots in a 5G NR frame can be allotted with Downlink symbols, Uplink symbols or Flexible symbols. A symbol assigned as Flexible symbol, can be used for Downlink and Uplink traffic as per the requirement. Each slot in 5G NR can accommodate a maximum of 14 OFDM symbols.

3.1 Related Work

There are a large number of factory automation IIoT use cases that leverage the benefit of cellular wireless communication technologies. With respect to industrial cellular 5G technology, there have been previously researched work, where the 5G wireless network is modified to suit the latency requirements. It is based on specific use cases whether end-to-end latency is considered, or latency is closely observed only between the immediate radio communication systems, i.e within the PHY and MAC layers of the 5G base station and UE indicated by L1+L2 latency in [Wi16]. In order to reduce the observed end-to-end latency in factory automation scenarios, different communication protocols and layers have also been modified. In the work of Natale Patriciello et al. [Pa18], the IP packet size was modified. After modifying the packet size and maintaining the packet generation rate, the packet arrival rate tends to increase in the Radio Link Control buffer. Kernel RLC buffer size was modified in the work of [Ir22] to investigate the effects on end-to-end latency with an increased receiving and sending buffer sizes in the end node machines in the communication system. The solutions proposed were appropriately suited for the specific use cases. An attempt to reduce the wireless communication latency by modifying the radio frame scheduling was performed in [Lä14]. One of the main focuses of the work has been on battery powered devices where it is crucial to conserve more energy. Performance improvement at the cost of battery life can considerably affect the operational time of such devices. The work in [Lä14] was also based on consumption of low energy with the purpose of increasing the battery life and performance. The performance was compared to that of a 4G technology-based implementation. In the work of Jens Pilz et al. [Pi16], 5G network was evaluated for tactile use cases. Such application scenarios require low latency communication, generally below 1 ms with high reliability. The research was carried out using a Software Defined Radio (SDR).

4 Research Project 5G4Automation

The overall objective of the 5G4Automation project is to design a methodology for the development of 5G products and services in the Industry 4.0 context. This will be implemented in the form of a kit equipped with methods, guidelines and concrete implementations. This kit is intended to enable small and medium-sized enterprises in particular to develop 5G products and services on a company- and application-specific basis. In this way, a contribution is also to be made to the technical sovereignty of companies and to Germany as a technology location overall. A general overview could be found in [5G423]

5 Our Minislot approach in the 5G Context

In the following we outline the basic 5G System as well as our novel mini slot approach to achieve real-time communication in an industrial context.

5.1 5G Mini Slot Approach

5G NR allows scheduling of the 5G frame in mini-slots. A mini-slot is described as the minimum scheduling unit in 5G NR. It is an enabler of a key feature in 5G called

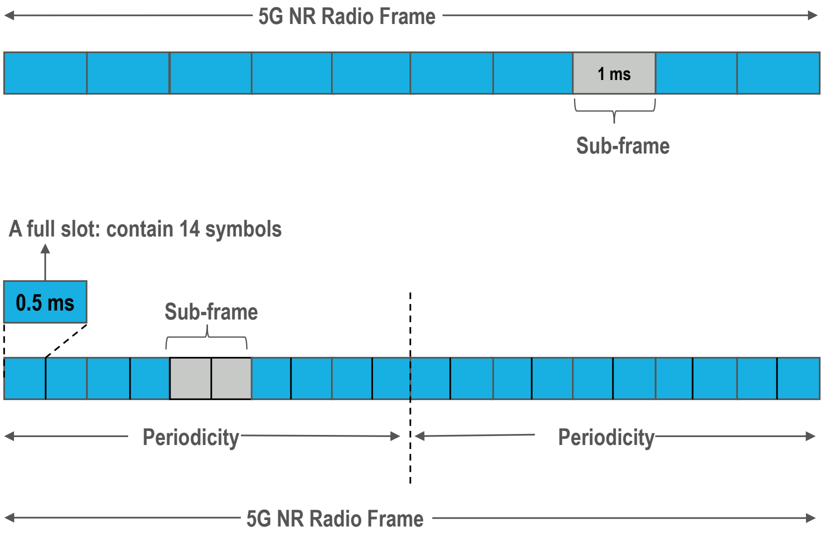
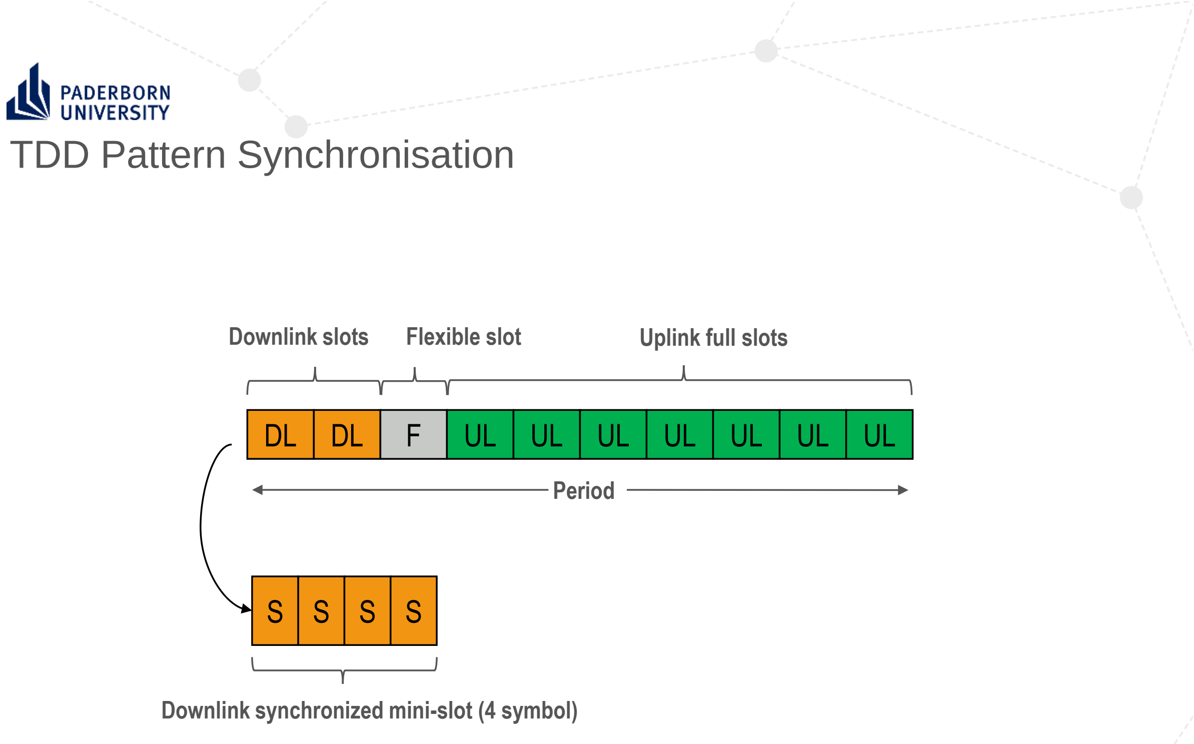


Fig. 2: The figure shows a TDD Synchronised 5G NR Frame and slot division per sub-frame at sub-carrier spacing 30 kHz used in this work.

Ultra-Reliable Low-Latency Communications (URLLC). This The mini-slot can be started at any OFDM symbol in a 5G NR slot and it can be inserted into an ongoing transmission. This enables services like URLLC along with, for example, enhanced Mobile Broadband (eMBB) traffic. Hence, it is implemented in an asynchronous mode with respect to a standard scheduled slot, which is synchronous. A mini-slot can contain either 2, 4 or 7 OFDM symbols. It can be used to transmit and receive user data in PUSCH or PDSCH channels. Hence, shorter and crucial data can be communicated using mini-slots with lower latency than that of a standard scheduled slot of 5G NR. A low latency communication provides faster and richer services because, then the 5G network can process large amounts of data in shorter time period.

This study tackles a real-time communication requiring scenario, where it is assumed that the control data for an Sensor Actuator Interface (SAI) is transmitted from a PLC over a 5G wireless channel. This is considered a Downlink (DL) network traffic. The crucial sensor data is transmitted from SAI to the PLC over the 5G wireless channel and is known as the Uplink (UL) network traffic in the problem scenario. In this regard, it is necessary that the control data from the PLC is transmitted as soon as possible to the SAI to avoid receiving any invalid or inaccurate data in the PLC. Also, the PLC is required to analyze and respond in real-time with control signals over 5G wireless network to the SAI, if there is a consequent action that needs to be performed by an actuator. In event-based systems such as PLCs and SAI, synchronous arrival of data is crucial. The operations in these systems are cyclic in nature. Therefore, if data arrives at a time that has missed the start of a cycle



TDD Pattern Synchronisation

Page 4

Fig. 3: TDD Synchronised 5G NR Frame containing two 4 symbol Downlink Mini-Slot

In order to solve this problem in favor of a reduced end-to-end latency and deterministic communication, mini-slots are designed in downlink direction of the network traffic in Time Division Duplex (TDD) mode. A synchronous mini-slot will introduce more deterministic traffic flow with reserved slots for transmitting control signals from the PLC to the SAI. Two downlink directed mini-slots are designed to be implemented in the 5G NR radio frame as indicated in the Fig. 3. The sensor will generate frequent data to be analyzed corresponding to the problem scenario. Hence, it is crucial that the sensor data is transmitted almost continuously to the PLC. Each of the seven uplink slots as seen in figure Fig. 3, transmit data at it's maximum capacity of 14 OFDM symbols per uplink slot. The TDD synchronized pattern uses one flexible slot. The symbols in the flexible slot can be assigned as downlink or uplink symbols based on traffic needs of 5G UE.

6 Real-Time Limitations of Current Industrial 5G Systems

Of the shelf 5G systems are optimized for private cellular usage patterns, like Streaming Data for example. Those system do not have a good configuration for the automation industry. As there are complete different usage patterns. Here the upload of small data frames (e.g. 128 Byte) is a good fraction of the overall traffic. Also the download speed has no value from this perspective.

In our measurements we do not focus on the achieved datarate, as for our main use case (profinet) the achieved end-to-end latency constitutes the prime metric for evaluation. To tackle this, we orchestrated sockperf and ITU-T Y.1564 for measuring the one-way latency under load.

In difference to sockperf the ITU measurements inducting additional load to the system which acts as general system load. Thus both measurements stress testing the COTS, but with different goals. Sockperf focusing on the sole system performance, ITU-T Y.1564 have the focus on shared networks.

Both measurements also outlining the limitations in current (release 15) 5G systems. As the Round Trip Time (RTT) in the worst case scenario could be more than 45 ms a lot of possible automation network types won't operate properly and need new network configurations and thus new safety evaluations.

6.1 Load-based Measurements with sockperf

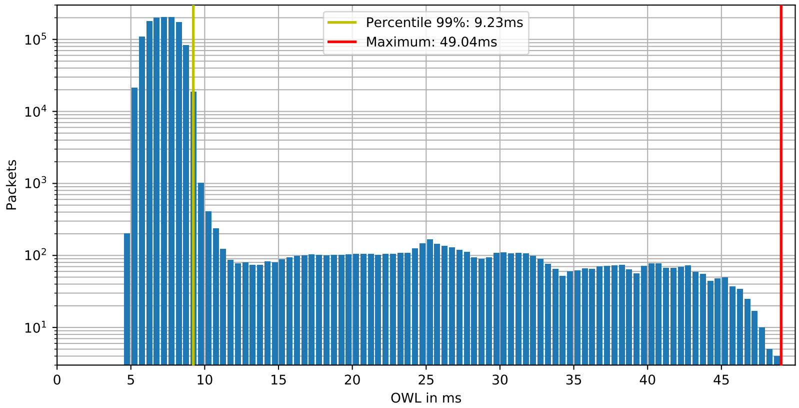


Fig. 4: Histogram of unidirectional latency measurements under load of COTS Rel-15-based 5G User Equipment (UE) and Stand Alone (SA) Non-Public Networks (NPN) (14 Byte payload with 100 µs transmission interval)

The sockperf test was performed in which for every fifth transmitted packet the server replies with a packet to emulate a utilized communication channel. This way, the measured RTT is being divided by to (as being done by sockperf) to gather the one-way latency.

The testbench consists of an Amarisoft Callbox acting as COTS and a Simcom 8200EA connected to a linux system as UE. The total runtime was 600 s which resulted in 5999964 Messages sent. The 99% percentile latency is 9.23 ms as shown in figure 4.

6.2 ITU-T Y.1564 Measurements

The main purpose of ITU-T Y.1564 is to provide network operators with a standardized method for measuring the end-to-end performance of Ethernet-based networks. This includes the performance of the network itself, as well as the performance of any devices or applications that are running on the network.

The test bench consists of two EXFO Ethersam ITU-T Y.1564 devices and two Mikrotik routers. The routers are connected to the 5G system as UE and the N6 interface. Both devices Internet Protocol (IP) address are static and run on the same RouterOS software (Version 7.11.2). Both devices are running as Virtual Tunnel Endpoint (VTEP) and the Ethernet ports are used for the EXFO connection as well as the connection to the 5G system.

In total three tests where performed with 128 Byte payload and additionally a 3000 Byte payload system load test (not shown). Each measurements creates 230.000 Ethernet frames to give an overview of the jitter and latency. Each Test was repeated at different signal levels to ensure a realistic view on the systems overall latency. The results are shown as an overview in Table 1. As the focus here is the mean performance each of the three measurements is averaged and summarized.

signal strength(dBm)	OWL(ms)	jitter(ms)
-55	18.300	10.417
-60	16.583	9.183
-65	17.067	10.633
-70	18.300	11.733
-75	16.867	9.900
-80	17.683	10.317
-85	17.317	10.033

Tab. 1: Table of ITU-T Y.1564 measurements of COTS Rel-15-based 5G UE and SA NPN (128 Byte payload)

This measurements could taken as additional reference values for the average system performance of an of the shelf 5G system. They show, that the average OWL is 17.45 ms with a jitter of 10.32 ms.

7 Performance Evaluation

After making initial modifications to the 5G network as intended in section 5, we have observed an overall improvement in the performance of the 5G system in terms of OWL. The maximum OWL achieved over a test duration of 30 minutes is shown to have improved by 68.78%. For 99 percentile of the 5G network traffic, the OWL has been depicted to have reduced by 2.4%. This also points to an enhancement of reliability in 5G system used for communication scenarios requiring critical low latency.

8 Conclusion

We presented a novel Mini-Slot approach for scheduling downlink and uplink data transmission in a 5G and beyond cellular networking context. By taking the specific requirements of

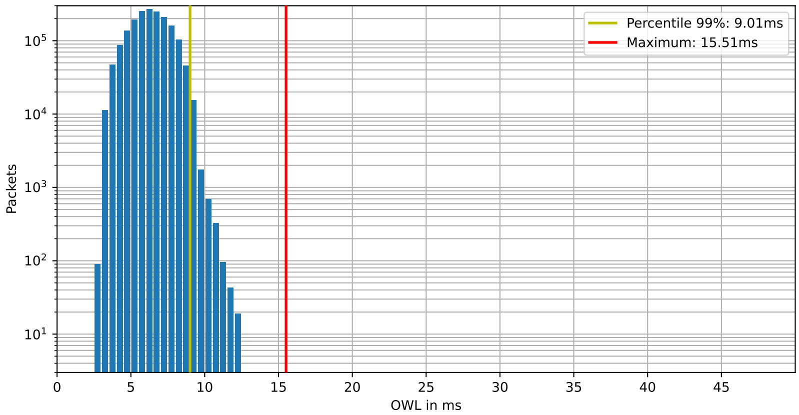


Fig. 5: Histogram of unidirectional latency measurements under load of COTS Rel-15-based 5G UE and SA NPN (14 Byte payload with 100 µs transmission interval) after modifying the 5G NR radio frame.

field bus communication for applications in the IIoT domain into account, our approach allows to lower the end-to-end communication latency on idle and under load communication links by up to 13.8% for typical industrial use-cases. Based on these results of a first feasibility study by using a Rel-15-based 5G Commercial off-the-shelf (COTS) and User Equipment (UE), our Mini-Slot approach constitutes as enabler for current and future application domains in the context of Industry Automation, where the prime requirement is low latency wireless communication.

References

[5G423] 5G4Automation Research Project, 2023, URL: `https://5g.nrw/best-practice/5g4automation/`, visited on: 09/25/2023.

[Ir22] Irazabal, M.; Lopez-Aguilera, E.; Demirkol, I.; Nikaein, N.: Dynamic Buffer Sizing and Pacing as Enablers of 5G Low-Latency Services. IEEE Transactions on Mobile Computing 21/3, pp. 926–939, 2022.

[Lä14] Lähetkangas, E.; Pajukoski, K.; Vihriälä, J.; Berardinelli, G.; Lauridsen, M.; Tiirola, E.; Mogensen, P.: Achieving low latency and energy consumption by 5G TDD mode optimization. In: 2014 IEEE International Conference on Communications Workshops (ICC). Pp. 1–6, 2014.

[Pa18] Patriciello, N.; Lagen, S.; Giupponi, L.; Bojovic, B.: 5G New Radio Numerologies and their Impact on the End-To-End Latency. In: 2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). Pp. 1–6, 2018.

[Pi16] Pilz, J.; Mehlhose, M.; Wirth, T.; Wieruch, D.; Holfeld, B.; Haustein, T.: A Tactile Internet demonstration: 1ms ultra low delay for wireless communications towards 5G. In: 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Pp. 862–863, 2016.

[Wi16] Wirth, T.; Mehlhose, M.; Pilz, J.; Holfeld, B.; Wieruch, D.: 5G new radio and ultra low latency applications: A PHY implementation perspective. In: 2016 50th Asilomar Conference on Signals, Systems and Computers. Pp. 1409–1413, 2016.

[Wo17] Wollschläger, Martin and Sauter, Thilo and Jasperneite, Jürgen: The Future of Industrial Communication. In: IEEE Industrial Electronics magazine. IEEE, 2017.

http://dx.doi.org/10.25673/111643

Simulations of the 5G-TSN bridge delay: towards a joint QoS model

Niklas Ambrosy[1], Azarakhsh Abazari[1] and Lisa Underberg[2]

Abstract: To integrate 5G mobile radio into Time-Sensitive Networking (TSN), the 3rd Generation Partnership Project (3GPP) specified the model of a virtual 5G-TSN bridge. This contains TSN translators which map principles such as time synchronization and Quality of Service (QoS) mechanisms from TSN to 5G. However, practical implementations with fine-granular QoS differentiation are not available yet. Therefore, in this paper, we examine the transmission delays of frames of the eight TSN traffic classes by simulating different scenarios while varying the QoS parameters priority, periodicity, and frame length. Our research contribution includes indications for the 5G bridge delay and Packet Delay Budget (PDB) depending on the traffic characteristics in a converged wired and wireless 5G-TSN network. This serves as a basis for the development of TSN translators and finally of a joint QoS model.

Keywords: 5G, Time-Sensitive Networking, Quality of Service, simulations, delay

1 Introduction

Flexible production processes and applications for Industry 4.0 require mobility and effortless reconfigurability and therefore a combination of deterministic wired and wireless communication technologies. This is specifically essential for time-critical machine communication. Possible use cases include wireless human-machine interfaces with emergency stop or automated guided vehicles to ensure personal safety in mobile applications. TSN and 5G are considered key technologies to meet the communication requirements of these use cases in converged wired and wireless networks.

TSN is the umbrella term for multiple IEEE 802.1 sub-standards that enable real-time capabilities and determinism for Ethernet. TSN includes mechanisms for time synchronization, bounded latency, high reliability, and dedicated resource management [IE23a]. The IEC/IEEE 60802 TSN Industrial Automation Profile intends to explicitly standardize the use of TSN in industrial automation, but is currently still in the draft stage [IE23b]. According to IEEE 802.1Q – Strict Priority, the traffic types in industrial communication are assigned to a total of eight traffic classes [In19], as listed in Tab. 1. A traffic class is identified using the Priority Code Point (PCP) as part of the Virtual Local Area Network (VLAN) tag. Priorities range from 0 to 7, where 7 represents the highest and 0 the lowest priority.

[1] Volkswagen AG, Berliner Ring 2, 38440 Wolfsburg, {niklas.ambrosy, azarakhsh.abazari}@volkswagen.de
[2] Institut für Automation und Kommunikation e. V., Werner-Heisenberg-Str. 1, 39106 Magdeburg, lisa.underberg@ifak.eu

Traffic Type	Periodicity [ms] (periodic/sporadic)	Data Delivery Guarantee	Data Size [bytes] (fixed/variable)	Criticality
Isochronous	0.1-2 (p)	Deadline	30-100 (f)	High
Cyclic synchronous or asynchronous	0.5-20 (p)	Latency	50-1000 (f)	High
Network control	50-1000 (p)	Bandwidth/data rate	50-500 (v)	High
Events	10-50 (s)	Latency	100-200 (v)	High
Alarms	2000 (s)	Latency	100-1500 (v)	Medium
Configuration & Diagnostics	N/A (s)	Bandwidth/data rate	500-1500 (v)	Medium
Audio/Video	A: 40 / V: 10 (p)	Bandwidth/data rate, latency	1000-1500 (v)	Low
Best effort	N/A (s)	None	30-1500 (v)	Low

Tab. 1: Traffic types and properties according to Industrial Internet Consortium [In19]

Tab. 2 shows the assignment of traffic types to traffic classes and priorities according to IEC/IEEE 60802, which differs slightly from that of the Industrial Internet Consortium. The combination of both serves as the traffic model for the simulations and will be discussed later.

Traffic Class	Priority (PCP)	Traffic Type
7	6	Isochronous
6	5	Cyclic synchronous
5	4	Cyclic asynchronous
4	7	Network control
3	3	Alarms & Events
2	2	Configuration & Diagnostics
1	1	Best effort high
0	0	Best effort low

Tab. 2: Traffic classes and priorities according to IEC/IEEE 60802 [IE23b]

5G as the fifth mobile radio generation is expected to meet industrial performance requirements, e.g., with the Ultra-Reliable Low Latency Communication (URLLC) feature to support time-critical machine communication. QoS in 5G is specified in 3GPP TS 23.501. Each QoS flow between the User Plane Function (UPF) and the User Equipment (UE) contains a certain QoS profile with multiple QoS parameters [5G21]:

- *Resource Type* defines how strictly other parameters should be handled. Guaranteed Bit Rate (GBR), Delay-Critical GBR, or Non-GBR can be distinguished.

- *Priority Level* indicates a flow's priority in relation to other flows for scheduling resources. Unlike TSN, the lowest value corresponds to the highest priority.
- *Packet Delay Budget (PDB)* sets an upper time limit for the delay between the UPF and the UE, before the packet is counted as lost.
- *Packet Error Rate (PER)* defines the reliability level by providing an upper bound on the number of incorrectly received or lost packets divided by the total number of received packets. The larger the packet and the lower the PDB, the higher the PER.
- *Maximum Data Burst Volume (MDBV)* indicates the data amount that can be sent without exceeding the PDB.

To integrate 5G into TSN, 3GPP specified the model of a virtual 5G-TSN bridge, as shown in Fig. 1 [3G22]. This model contains TSN translators which map information and parameters, e.g., for time synchronization and QoS, between TSN and 5G [RCK20]. 5G and TSN parameters can be translated as follows:

- TSN PCP ≙ 5G Priority Level
- TSN periodicity ≙ 5G transfer interval
- TSN frame length ≙ 5G MDBV.

One important aspect of QoS is the time delay that frames experience when traversing the 5G-TSN bridge, expressed by the PDB or bridge delay (BD). TSN AF determines and reports the minimum and maximum BD per port pair and traffic class to the CNC to check whether the delay requirements of the TSN stream to be added can be met.

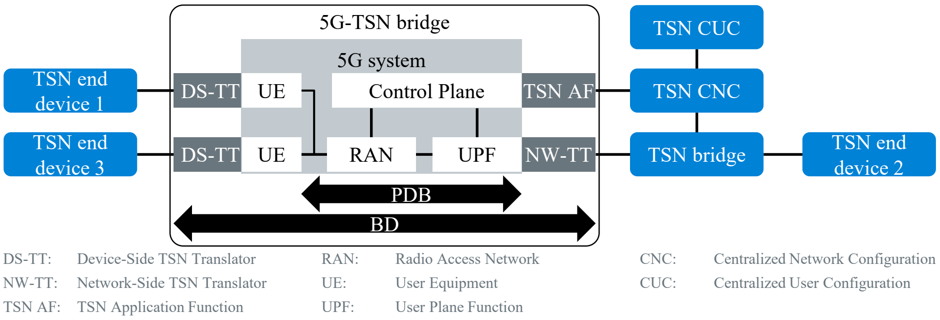

5G-TSN bridge
5G system
TSN end device 1
TSN end device 3
DS-TT
DS-TT
UE
UE
Control Plane
RAN
UPF
TSN AF
NW-TT
TSN CUC
TSN CNC
TSN bridge
TSN end device 2
PDB
BD
DS-TT: Device-Side TSN Translator
NW-TT: Network-Side TSN Translator
TSN AF: TSN Application Function
RAN: Radio Access Network
UE: User Equipment
UPF: User Plane Function
CNC: Centralized Network Configuration
CUC: Centralized User Configuration

Fig. 1: 5G system as a virtual TSN bridge according to 3GPP TS 23.501 [3G22]

TSN traffic types and classes have different characteristics and requirements in terms of priority, periodicity, and frame length. They affect the delays within the 5G system and need to be considered to determine realistic BD and PDB values for frames of different TSN traffic types and classes. Using simulations of TSN traffic over 5G in OMNeT++, this paper examines the delays to provide indications for the parameters BD and PDB as a basis for a pre-configured 5G-TSN QoS mapping table, complementing previous theoretical considerations and analytical calculations. Thus, our paper contributes to the concretization of a joint 5G-TSN QoS model.

The remainder of this paper is organized as follows: Section 2 presents relevant related work. Section 3 explains the simulation framework. Section 4 presents the simulation results, which are discussed in Section 5. Section 6 concludes the paper with a summary and an outlook. Note that the paper contains an appendix with boxplot diagrams.

2 Related work

This section provides an overview of simulations of 5G delays in OMNeT++, which can be identified as relevant related work. Prototype implementations that reflect the scope of the simulations are not known.

Martenvormfelde et al. investigate only one TSN traffic class with 1 ms periodicity and a frame length of 256 bytes in downlink (DL) or 64 bytes in uplink (UL) according to the 5G-ACIA traffic model [5G19], for which they vary the UL/DL slot size [Ma20]. Magnusson and Pantzar and Satka et al. use their independently developed TSN translators and 5G link model and validate it in a use case with two different examples of TSN traffic classes, i.e., with two different PCP values, but without addressing all parameters of the TSN traffic classes [MP21], [Sa22]. Rost and Kolding use the commercial version OMNEST and also the 5G-ACIA traffic model. They achieve a bridge delay of 1 ms and less by replicating the 5G radio access network (RAN) through randomly selected signal-to-interference-to-noise-ratio (SINR) values [RK22].

Our simulations differ from the previous ones as follows: Based on the value ranges discussed in [AU22] and [AU23] as part of previous work, we use fixed worst-case and variable values for the parameters priority, periodicity, and frame length for each of the eight TSN traffic classes as input for the simulations.

3 Simulation framework

This section explains the simulation model in OMNeT++ with INET and Simu5G. INET Framework is a model library for the OMNeT++ simulation environment. It provides protocols, agents, and other models for communication networks, such as models for the Internet stack or wired and wireless link layer protocols. Several other simulation frameworks take INET as a base and extend it into specific directions, e.g., Simu5G [Bo23]. Simu5G simulates the data plane of the 5G RAN (according to 3GPP Release 16) and core network. It allows the simulation of 5G communications with multiple features and provides 3GPP-compliant protocol layers [VN20]. In this paper, INET 4.4.1 and Simu5G 1.2.1 are used.

To evaluate suitable values for the BD and PDB per traffic class, we simulate the 5G transmission times with Simu5G in the following four scenarios with different numbers of UEs and according to the traffic parameters shown in Tab. 3:

1. DL (UPF to UE)
2. UL (UE to UPF)
3. UE to UE (UL+DL)
4. Mixed (UL+DL between UPF and UE)

TSN traffic class	**TSN traffic type**	**TSN priority/ 5G priority level**	**TSN periodicity/ 5G transfer interval [ms]**	**TSN frame length/ 5G MDBV [bytes]**
7	Isochronous	6 / 2	1	68
6	Cyclic synchronous	5 / 3	10	500
5	Cyclic asynchronous	4 / 4	30	500
4	Network control	7 / 1	500	250
3	Alarms & Events	3 / 5	2000	800
2	Configuration & Diagnostics	2 / 6	1000	1000
1	Best effort high	1 / 7	25	1250
0	Best effort low	0 / 8	1000	1500

Tab. 3: Configuration of TSN traffic parameters for 5G transmission time simulations

In factories, the most likely scenario is an overarching 5G network for the entire plant site and several separate TSN networks per production cell, group of production cells, or, at its largest, an entire production hall. To evaluate scalability, we increase the number of UEs, starting from one UE. For better comparability, all of them are located in the same position at a distance of 10 m from the gNB. In total, the simulation contains 22 iterations (four scenarios, each with different numbers of UEs and without and with prioritization).

A timer is configured within the network to timestamp the packets at the sender and receiver (depending on the scenario). Fig. 2 shows a schematic drawing of the components and modules. To implement prioritization, modules need to be adapted by proprietary developments (depicted in green boxes): The server and gNB are modified by a *Ppp* compound module, including sub-modules called *DropTailQueue*, *Classifier* and *PriorityScheduler* to support eight priority queues. The modules *CbrSender* and *CbrReceiver*, which are genuinely used to transmit constant bit-rate (CBR) packets over the network, are modified to enable labeling packets in order to later being prioritized in the Ppp modules.

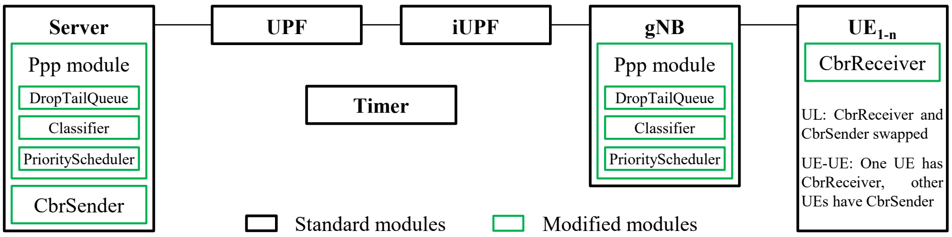


Fig. 2: Schematic drawing of the components and modules in OMNeT++

4 Simulation results

This section presents the simulation results of 32 iterations, including four scenarios with four different numbers of UEs each, both with and without prioritization.

Tab. 4 shows the maximum transmission times per iteration and traffic class. In all scenarios, a similar pattern can be observed. Large packets and packets sent with a low periodicity take longer. Prioritization affects the transmission times. They are higher in UL than in DL since the 5G time-division duplex (TDD) pattern typically provides fewer time slots and capacity for UL. The UL/DL slot ratio cannot be configured or changed in OMNeT++. Mixed traffic increases the network load and consequently the transmission times compared to pure UL or DL traffic. UE-UE communication always includes UL+DL and represents the most complex scenario with the highest transmission times.

Traffic classes 4, 6, and 7 with the highest priorities and traffic class 1 benefit from prioritization and rather in UL. In general, delays of small packets up to 500 bytes and low periodicities up to 25 ms improve. Two exceptions are traffic classes 1 (with 1250 bytes) and 4 (with 500 ms), where the prioritization also has a positive effect. Other traffic classes experience additional delays due to the packet queues introduced with the prioritization.

The more UEs are used, the higher are the mean values, maxima and outliers of the transmission time. In DL, UEs perform differently, although they are located in the same distance to the gNB. There is only one absolute maximum, but the mean values are similar. Due to separate time measurements per UE, multiple values can be obtained for DL. The UL provides only one value since the time is measured only at the server. The minimum values of the transmission time are 4-7 ms in DL, 4-15 ms in UL, and 9-20 ms for UE-UE (where three UEs perform better than two). For mixed traffic, minimum values are 5-7 ms in DL and 4-15 ms in UL. They seem to be independent of the packet size and periodicity.

Iteration	Scenario	Prioritization	Traffic class 7 1 ms, 68 bytes	Traffic class 6 10 ms, 500 bytes	Traffic class 5 30 ms, 500 bytes	Traffic class 4 500 ms, 250 bytes	Traffic class 3 2000 ms, 800 bytes	Traffic class 2 1000 ms, 1000 bytes	Traffic class 1 25 ms, 1250 bytes	Traffic class 0 1000 ms, 1500 bytes
1	DL 1 UE	N	13	13	12	14	12	11	11	8
2		Y	12	12	12	12	12	12	12	12
3	DL 2 UEs	N	21	21	19	22	19	18	20	15
4		Y	21	21	21	21	21	21	21	21
5	DL 3 UEs	N	46	43	40	46	29	36	42	28
6		Y	43	43	43	43	43	43	43	43
7	UL 1 UE	N	28	28	27	24	20	23	25	22
8		Y	28	28	27	28	26	26	26	23
9	UL 2 UEs	N	35	35	30	32	29	28	31	27
10		Y	31	31	30	31	30	30	30	27
11	UL 3 UEs	N	53	53	47	46	43	42	53	37
12		Y	51	51	49	51	49	49	49	39
13	UE-UE 2 UEs	N	35	35	33	32	27	31	34	30
14		Y	34	34	34	34	34	34	34	30
15	UE-UE 3 UEs	N	43	43	39	40	38	37	43	35
16		Y	44	44	40	44	40	40	40	35
17	Mixed 1 UE	N	DL: 13 UL: 28	DL: 13 UL: 28	DL: 12 UL: 27	DL: 14 UL: 24	DL: 12 UL: 20	DL: 11 UL: 23	DL: 11 UL: 25	DL: 8 UL: 22
18		Y	DL: 12 UL: 28	DL: 12 UL: 28	DL: 12 UL: 27	DL: 12 UL: 28	DL: 12 UL: 26	DL: 12 UL: 26	DL: 12 UL: 26	DL: 12 UL: 23
19	Mixed 2 UEs	N	DL: 23 UL: 35	DL: 23 UL: 35	DL: 22 UL: 30	DL: 23 UL: 31	DL: 21 UL: 29	DL: 20 UL: 28	DL: 21 UL: 31	DL: 16 UL: 27
20		Y	DL: 21 UL: 33	DL: 21 UL: 33	DL: 21 UL: 30	DL: 21 UL: 33	DL: 21 UL: 30	DL: 21 UL: 30	DL: 21 UL: 30	DL: 21 UL: 26
21	Mixed 3 UEs	N	DL: 42 UL: 54	DL: 40 UL: 54	DL: 33 UL: 46	DL: 37 UL: 49	DL: 30 UL: 45	DL: 35 UL: 42	DL: 39 UL: 51	DL: 29 UL: 35
22		Y	DL: 43 UL: 57	DL: 43 UL: 57	DL: 43 UL: 56	DL: 43 UL: 57	DL: 43 UL: 56	DL: 43 UL: 56	DL: 43 UL: 56	DL: 43 UL: 43

Tab. 4: Maximum transmission times per iteration and traffic class

Fig. 3-10 depict boxplot diagrams of those iterations where differences and effects are visible and comparable. The mean values and box sizes (lower and upper quartile) tend to be highest at traffic class 0 and 2, which is due to the combination of large packets and long time intervals. The whiskers of the boxplot diagrams end at the 5th and 95th percentile. The arithmetic mean, minimum and maximum values are given as numbers in the diagrams. The median is depicted as an orange line within each boxplot.

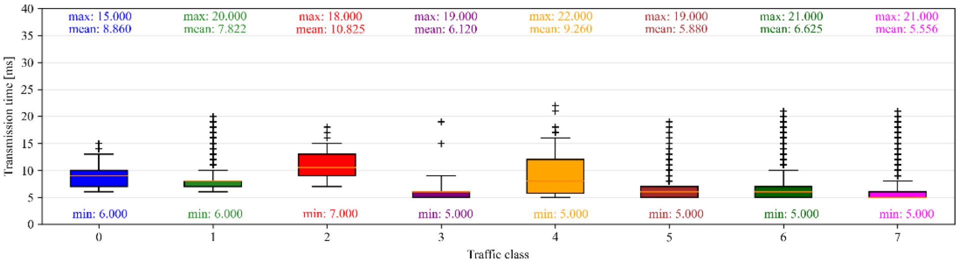


Fig. 3: Two UEs in DL without prioritization (iteration 3)

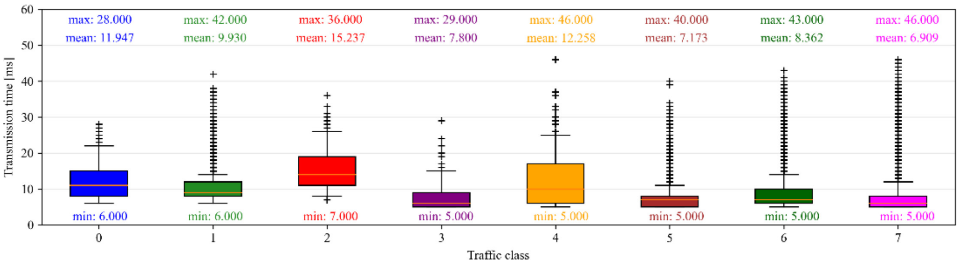


Fig. 4: Three UEs in DL without prioritization (iteration 5)

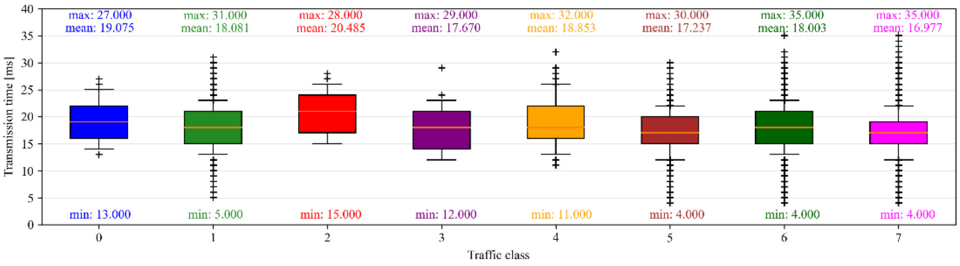


Fig. 5: Two UEs in UL without prioritization (iteration 9)

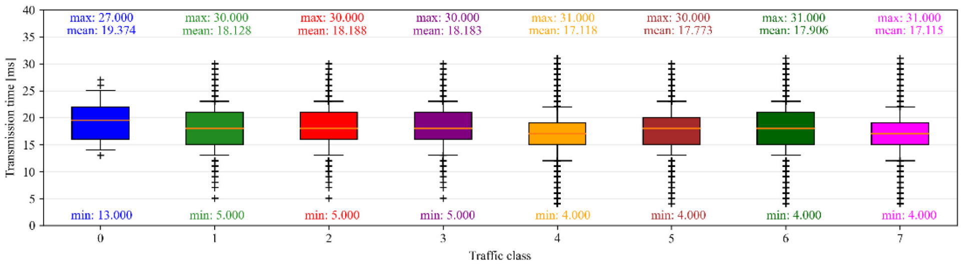


Fig. 6: Two UEs in UL with prioritization (iteration 10)

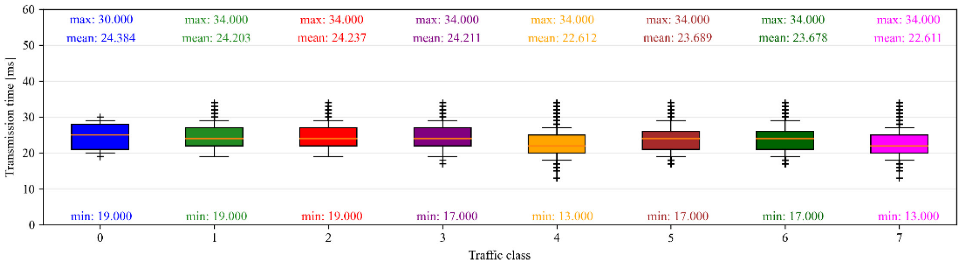


Fig. 7: UE to UE with prioritization (iteration 14)

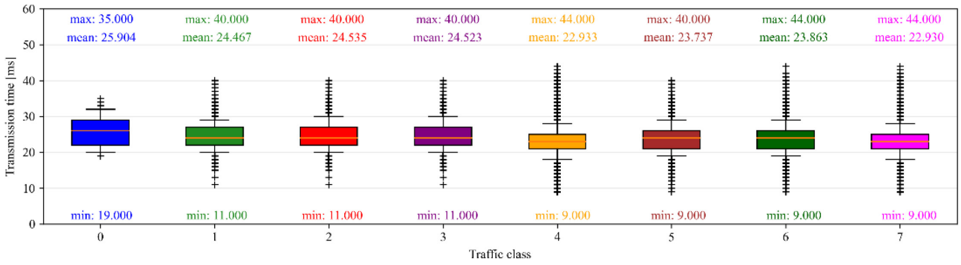


Fig. 8: Three UEs (two senders and one receiver) with prioritization (iteration 16)

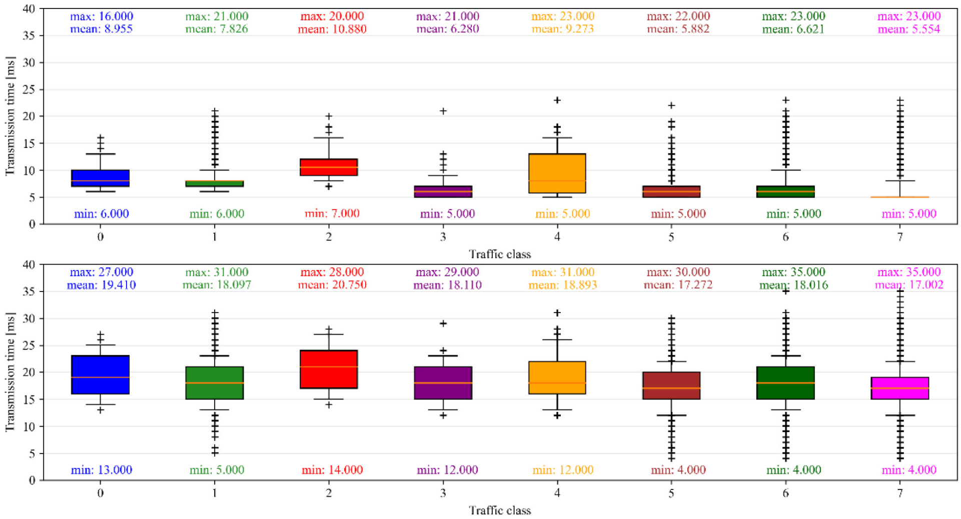


Fig. 9: Mixed traffic with two UEs and without prioritization; a) DL and b) UL (iteration 19)

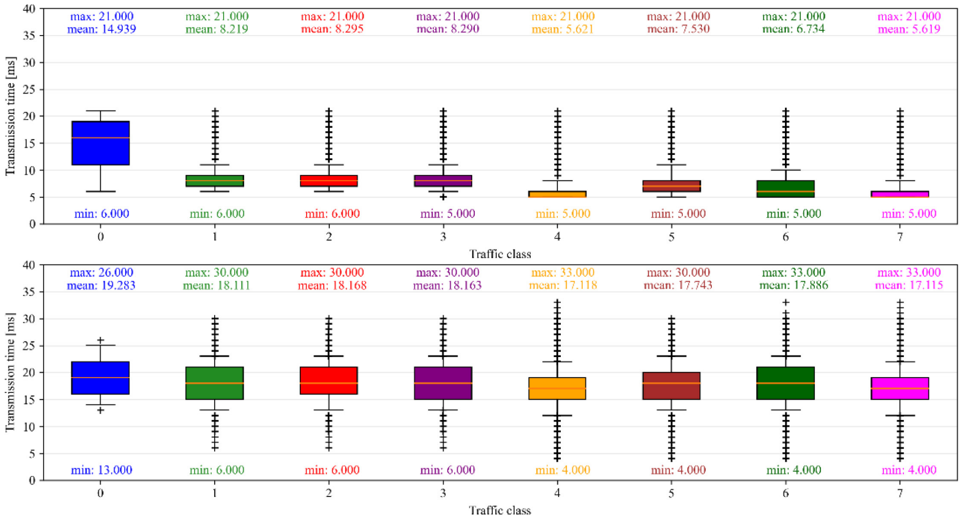


Fig. 10: Mixed traffic with two UEs and prioritization; a) DL and b) UL (iteration 20)

5 Discussion

In this section, the simulation results are discussed and conclusions for the 5G BD and PDB in the context of TSN are drawn.

The results are not reliable for more than three UEs since the transmission times are in the range of seconds. With four and more UEs, anomalies with packet loss and extremely high delays occur in each scenario and traffic class, e.g., in DL up to 16.9 s with five UEs or up to 46.7 s with ten UEs. Note that some packet loss may have occurred because the simulation had already finished when the packets were received. Therefore, the scalability of the simulation model is limited and the original plan with 50 and 100 UEs needs to be discarded. However, it can be estimated that the difference between the delays for prioritized and non-prioritized packets increases proportionally when the number of UEs increases. This issue can be further solved in more advanced simulations. Mobility of the UEs was not considered, but can be implemented in a future refinement of the simulation model. Although all UEs at the same location do not represent reality, this simplification was chosen for better comparability of the results.

Moreover, it is relevant to interpret the meaning of the simulation results for the BD and PDB. Transmission time can be equated with the BD. The transmission times are measured at the application level, since they are always about 1 ms between UPF and UE, regardless of the packet characteristics. Consequently, the applications at the sender and receiver cause the delays, which is assumed to be comparable to the delays introduced by NW-TT and DS-TT in the 3GPP bridge model, i.e., due to the conversion between wired and wireless transmissions. However, the simulated BDs significantly exceed expected values for time-critical TSN transmissions over 5G.

QoS in general and specifically the PDB cannot be simply set in Simu5G. The maximum transmission time values could be defined as PDBs of TSN traffic classes, but they still exceed the expectations for URLLC, which is obviously not supported by Simu5G. Prioritization could be implemented with a positive effect on high priority traffic classes. The more traffic occurs in the network due to larger packets, shorter periodicities and/or more UEs, the longer transmission times (≙ BD) result and the more a possible PDB is exhausted. It is questionable whether PDBs can simply be scaled down to values in the µs range and whether future 5GS will be able to comply with them under all circumstances.

6 Conclusion

In this paper, we analyzed relevant QoS parameters of the eight TSN traffic classes and simulated the transmission delays of typical TSN traffic flows in a 5G system. The results show that the delays depend on the parameters priority, periodicity, and frame length as well as on the examined scenarios and the number of UEs. The 5G bridge delays are too high for time-critical TSN traffic and the PDB cannot simply be limited to a certain value.

The QoS mechanisms need to be further investigated on the way to a joint 5G-TSN QoS model for future factory networks. Therefore, future work includes enhancements of the simulation model, e.g., a mobility model and improved scalability for more UEs, and experiments with QoS in real 5G-TSN implementations.

7 References

[3G22] 3GPP: TS 23.501: System architecture for the 5G System (V18.0.0), 2022.

[5G19] A 5G Traffic Model for Industrial Use Cases, 2019.

[5G21] 5G-ACIA: 5G QoS for Industrial Automation, 2021.

[AU22] Ambrosy, N.; Underberg, L.: Traffic priority mapping for a joint 5G-TSN QoS model. In (Jasperneite, J.; Jumar, U. Eds.): Kommunikation in der Automation. Beiträge des Jahreskolloquiums KommA 2022, Lemgo. Institut für industrielle Informationstechnik - inIT der Technischen Hochschule Ostwestfalen-Lippe, Lemgo, pp. 28–38, 2022.

[AU23] Ambrosy, N.; Underberg, L.: 5G packet delay considerations for different 5G-TSN communication scenarios: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN), 2023.

[Bo23] Bojthe, Z. et al.: What Is INET Framework? https://inet.omnetpp.org/Introduction.html, accessed 24 Apr 2023.

[IE23a] IEEE: Time-Sensitive Networking (TSN) Task Group. 1.ieee802.org/tsn, accessed 10 Aug 2023.

[IE23b] IEC/IEEE 60802: TSN Profile for Industrial Automation - D2.0, 2023.

[In19] Industrial Internet Consortium: Time Sensitive Networks for Flexible Manufacturing Testbed - Characterization and Mapping of Converged Traffic Types, 2019.

[Ma20] Martenvormfelde, L. et al.: A Simulation Model for Integrating 5G into Time Sensitive Networking as a Transparent Bridge: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, pp. 1103–1106, 2020.

[Ma21] Martenvormfelde, L. et al.: Co-configuration of 5G and TSN enabling end-to-end quality of service in industrial communications: Kommunikation in der Automation (KommA 2021) 12. Jahreskolloquium, 18.11.2021 in Verbindung mit dem Industrial Radio Day, 17.11.2021 Tagungsband, Magdeburg, 2021.

[MP21] Magnusson, A.; Pantzar, D.: Integrating 5G Components into a TSN Discrete Event Simulation Framework. Masterarbeit, Västerås, Sweden, 2021.

[RCK20] Rost, P. M.; Chandramouli, D.; Kolding, T.: 5G plug-and-produce - How the 3GPP 5G System facilitates Industrial Ethernet, 2020.

[RK22] Rost, P. M.; Kolding, T.: Performance of Integrated 3GPP 5G and IEEE TSN Networks. IEEE Communications Standards Magazine 2/6, pp. 51–56, 2022.

[Sa22] Satka, Z. et al.: Developing a Translation Technique for Converged TSN-5G Communication: 2022 IEEE 18th International Conference on Factory Communication Systems (WFCS). IEEE, pp. 1–8, 2022.

[VN20] Virdis, A.; Nardini, G.: 5G New Radio User Plane Simulation Model for INET & OMNeT++. Description. http://simu5g.org/description.html, accessed 24 Apr 2023.

The results, opinions, and conclusions expressed in this publication are not necessarily those of Volkswagen AG.

http://dx.doi.org/10.25673/111646

LoRaWAN Range Extension for Environments with High Attenuation

Martin Böhm, Olaf Gebauer, Diederich Wermser

Research Group Communication Systems
Ostfalia University
Salzdahlumer Str. 46/48
38302 Wolfenbüttel, Germany
{ma.boehm | ola.gebauer | d.wermser}@ostfalia.de

Abstract: The deployment of widespread LPWAN sensor networks, such as LoRaWAN, in high attenuation environments like forests and mines, faces a significant challenge due to the considerable reduction in communication range. LoRaWAN gateways are typically deployed to expand network coverage. These gateways require internet connectivity, often realized using cellular networks. However, in such deployment areas, cellular coverage may be limited or entirely unavailable. The use of LoRaWAN range extenders to extend coverage is inevitable. However, current LoRaWAN range extenders provide only limited coverage extension as they lack multi-hop support, necessitating the presence of online gateways. In this work, a new approach for LoRaWAN range extenders, named the LoRaWAN GW2ED Range Extender, is presented. Briefly explained, the new solution combines a LoRaWAN gateway, a local middleware, and an LoRaWAN end-device without the need for internet access by the range extender. The new approach is compared to other range extender solutions. The implementation of the concept is illustrated by architecture and sequence diagrams. In field tests within a dense forest, each PoC LoRaWAN GW2ED Range Extender extended the LoRaWAN range by an additional kilometer. This new approach is fully compatible with the LoRaWAN specifications as well as with LoRaWAN Commercial Off-The-Shelf (COTS) end-devices.

1 Introduction

LoRaWAN (Long-Range Wide Area Network) is a well-established LPWAN (Low-Power Wide-Area Network) technology for (I)IoT (Industrial Internet of Things) applications, providing considerable coverage within an unlicensed radio spectrum. However, in areas with high attenuation, the range of LoRaWAN is significantly reduced, as opposed to line-of-sight conditions where distances of over 10 km can be achieved. Environments such as forests, mines, or industrial production halls exemplify scenarios with notable attenuation challenges. For environments with sufficient cellular network coverage, LoRaWAN coverage can be increased by deploying more LoRaWAN gateways with cellular network backend connectivity, i.e., reducing the distances to LoRaWAN sensors. However, there are environments where sufficient cellular network coverage is not given, and the extension of LoRaWAN range is inevitable for the operation of widespread sensor networks, as illustrated in Figure 1.

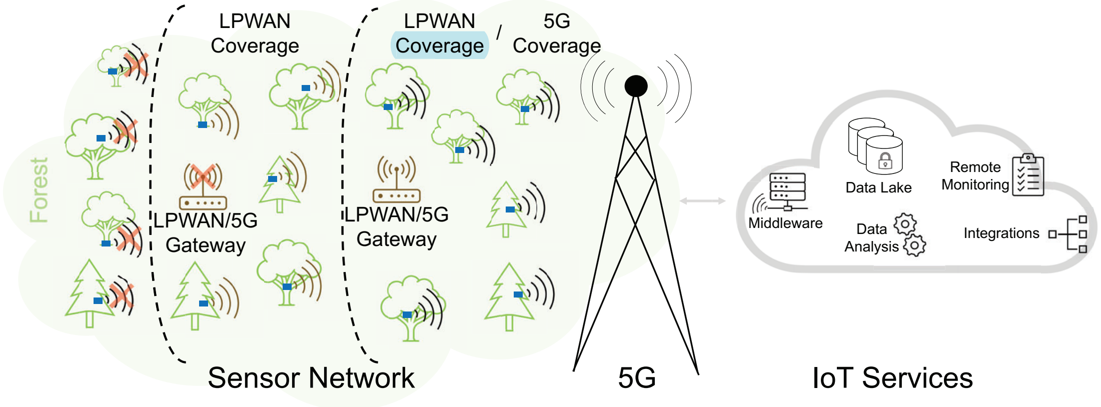


Figure 1: LPWAN forest sensor network for e.g., forest health monitoring challenged by high attenuation and off-cellular-network-coverage.

Within this work, known alternatives for extending the LoRaWAN range are investigated, including LoRaWAN Relays [2] und LoRa Relays [3]. Moreover, a new solution, named LoRaWAN GW2ED (Gateway to End-Device) range extender is presented, which overcomes significant disadvantages of the solutions known so far. Briefly explained, the LoRaWAN GW2ED solution combines a gateway and an end-device, attached back-to-back. This paper provides an exploration of the architecture and mechanisms of the new LoRaWAN GW2ED solution. Mechanisms implementing a LoRaWAN GW2ED solution, such as sensor data relaying, downlink message forwarding, and remote management functionalities, are illustrated in comprehensive diagrams. Additionally, the new solution is compared to alternative range extender solutions. The paper also presents the evaluation results obtained from a PoC (Proof-of-Concept) implementation of a LoRaWAN GW2ED Range Extender. An important aspect of this work is, that the new solution is compatible with existing COTS (Commercial Off-The-Shelf) LoRaWAN end-devices.

This work is structured as follows: Section 2 addresses challenges associated with operating sensor networks in high-attenuation environments and discusses solutions for extending the LoRaWAN range, including an overview of related work. Next, in Section 3, the new solution for extending the LoRaWAN range, called LoRaWAN GW2ED, is introduced, along with architecture and sequence diagrams. The implementation of a LoRaWAN GW2ED range extender is presented in Section 4. Section 5 focuses on the evaluation of the PoC implementation, which was tested within a forest environment. This evaluation includes a comparison of the new solution with alternative range extenders. Finally, Section 6 concludes this paper and presents future work.

2 LoraWAN Range Extender Concepts

Problem Statement

The deployment of sensor networks in environments with high attenuation, such as forests, mines, and industrial production halls, faces a critical challenge due to the significantly reduction in communication range. For instance, in a study conducted by Villarim et al. [VL19], the communication range of LoRa technology was evaluated in both urban and forest settings. In urban environments, they achieved a maximum range of 2.1 km, while in forest areas, the range dramatically decreased to 800 m and 232 m. Expanding LoRaWAN coverage typically involves the deployment of additional LoRaWAN gateways, which rely on active internet connectivity. However, in high attenuation environments, sufficient cellular coverage may not be given. This underscores the necessity for a solution that operates independently of cellular internet connectivity. Therefore, this paper explores solutions for extending LoRaWAN coverage through the use of range extenders.

LoRaWAN Basics

Further, relevant information about LoRaWAN important for paper is given. LoRaWAN is a well-established LPWAN technology that operates within an unlicensed radio spectrum, such as the 868Mhz band in the EU. LoRaWAN's coverage can be extended by deploying additional LoRaWAN gateways, making it particularly advantageous for environments with high attenuation, where the signal strength is reduced. In contrast, NB-IoT, another LPWAN technology, relies on cellular licensed communications bands, and its coverage extension is dependent on the services provided by mobile network operators. LoRaWAN benefits from a robust market, offering a diverse range of COTS smart sensors, which are sensor device equipped with built-in LPWAN functionality, as well as a wide range of COTS LoRaWAN gateways.

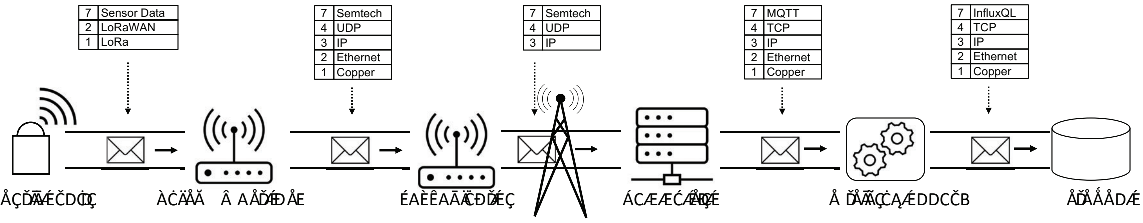


Figure 2: Data flow from smart sensor to a database exemplified with LoRaWAN, the Semtech UDP Packet Forwarder, MQTT and InfluxDB.

Furthermore, some key technical aspects of LoRaWAN are described. Figure 2 shows the data flow from a smart sensor to a database, including each entity involved in the process as well as network protocols utilized in the process. Smart sensors utilize the LoRa communication protocol to transmit their data. LoRaWAN gateways receive these LoRa messages and typically relay them over the internet to a middleware, using protocols like the Semtech UDP Packet Forwarder (P.F.). The middleware, often hosted within a cloud-based environment, comprises multiple software modules. The LoRaWAN network server manages the LoRaWAN layer for end-devices, performing functions such as deduplication handling of uplink message (detect when multiple gateways forward the same uplink message), downlink queueing, interaction with the application server, and join server. The application server handles the payload of the end-device, including payload decryption. The decrypted payload is then made accessible for further processing, often by publishing it on an MQTT broker. The join server manages the OTAA (Over-The-Air Activation) process for end-devices, including the exchange of session keys. Widely used middleware solutions include The Things Stack and the open-source Chirpstack [CS23]. Subsequently, from the MQTT broker, a service can store the sensor data within a database, such as the time-series database InfluxDB, for further utilization. Figure 3 provides an uplink message from an end-device, such as a sensor's temperature value, in the LoRaWAN architecture. Within the radio range of ED1, two LoRaWAN gateways are present, forwarding the message over the internet to the middleware, where the deduplicated message is processed. In LoRaWAN, all devices belong to the same middleware.

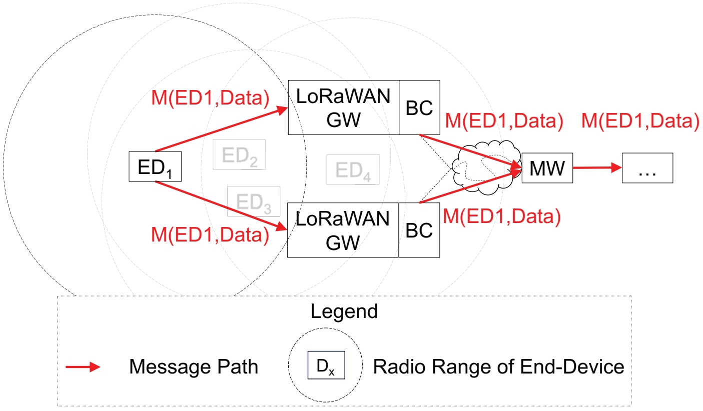


Figure 3: Uplink consideration for the LoRaWAN architecture (GW = gateway; MW = middleware such as Chirpstack or The Things Network; ED = end-device; BC = backend connectivity; M(x,y) = message with sender device ID x and payload y).

The LoRaWAN specification describes three distinct classes of devices. In Class A, end-devices transmit data by sending uplink messages at any time. Following the uplink transmission, there are two short downlink windows available for receiving data. Class A end-devices are commonly used for environmental monitoring and are typically configured to transmit their measured data at predefined intervals or in response to alarm triggers, such as reaching a sensor threshold. These end-devices are often powered by batteries with lifespans up to 10 years. Class B involves gateways transmitting time-synchronization beacons to schedule downlink messages to end-devices. This class introduces a structured downlink communication schedule for devices. Class C devices are typically mains-powered and maintain a continuous receive windows, allowing them to be in a constant listening mode for incoming data.

As previously mentioned, extending the coverage of LoRaWAN is typically accomplished by deploying additional gateways. However, in regions characterized by limited internet accessibility, such as areas with inadequate cellular network coverage, the expansion of the LoRaWAN range becomes essential for the successful operation of an extensive sensor network. Further, two existing solutions for range extenders are presented.

LoRaWAN Relays

In September 2022, the LoRa Alliance introduced a LoRaWAN Relay specification [TS22]. This specification describes the deployment of relay devices positioned between LoRaWAN end-devices and a LoRaWAN gateway, as illustrated in Figure 4.

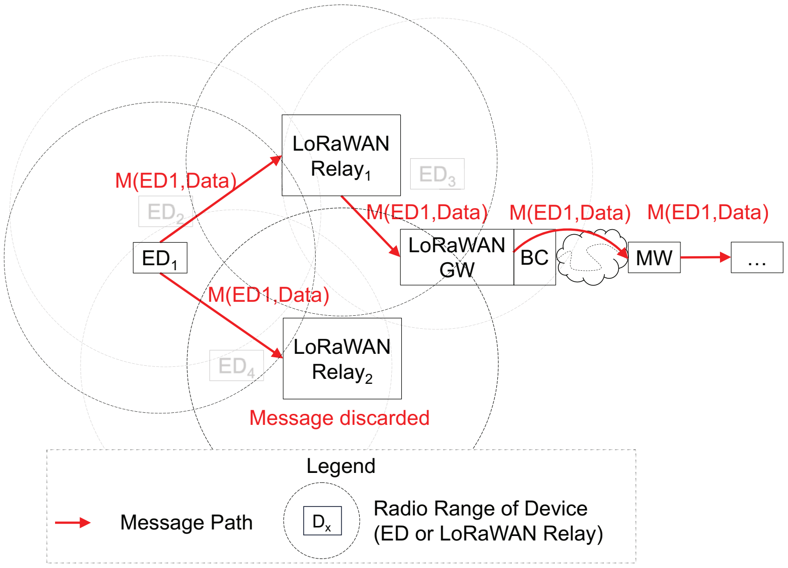


Figure 4: Uplink consideration for LoRaWAN Relay scenario (GW = gateway; MW = middleware; ED = end-device; BC = backend connectivity, M(x,y) = message with sender device ID x and payload y).

The LoRaWAN relay specification employs Wake On Radio (WOR) technology. When an end-station initiates an uplink message, a WOR frame is transmitted, activating the relaying station to await the impending uplink transmission. Afterwards, the received uplink data from the end-station is forwarded to the gateway. As both end-stations and relay devices typically operate in sleep mode, waking up only for data transfer, relay devices can also be powered by batteries. Additionally, to accommodate the increased delay introduced by the relaying mechanism, a third receive window for end-stations has been introduced. The visualization of this relaying mechanism is depicted in Figure 5.

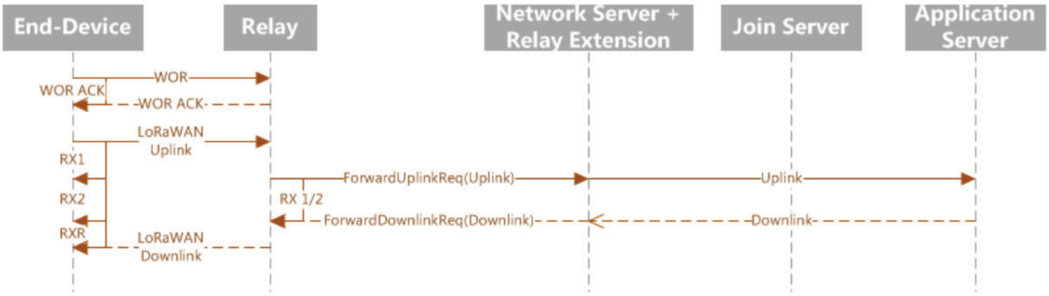


Figure 5: Sequence diagram of a LoRaWAN end-device join request using a LoRaWAN relay [TS22].

Currently, LoRaWAN relays have been designed to support a maximum of 16 different end-devices [SE23]. An end-device must be registered at a LoRaWAN relay while also remaining registered at the same middleware. Messages received at a relay from an end-device that does not belong to that LoRaWAN relay are discarded, as visualized in Figure 4. However, uplink message can only received when a relay is ready (awake) for data reception. It's important to note that relays cannot be cascaded, meaning that the relay devices cannot be daisy-chained. End-devices must implement the necessary functionality, implying that existing devices from manufacturers require firmware updates to enable this functionality. Additionally, commercially available LoRaWAN relays are not yet accessible in the market.

LoRa Relays

On the other hand, LoRa Relays offer the capability to relay LoRa frames, as illustrated in Figure 6. Mamour et al. proposed a LoRa relay that can be integrated into an existing LoRa network [MC19]. Their approach involved storing downlink messages at the relay and forwarding them after the next uplink transmission, thus enabling downlink functionality. Typically, a relay requires some form of configuration to determine which devices' data it should relay, as indiscriminate relaying of all received LoRa data could potentially violate duty cycle regulations. Their solution sidesteps the need for a management protocol by introducing an observation phase, eliminating the necessity for control messages between end-devices, gateways, and relay devices. During this phase, triggered by a device restart, the system logs all devices that transmit within a specific timeframe. Only messages from the learned devices are subsequently relayed. Furthermore, during this phase, transmission intervals of the smart sensors are observed to put the range extender to sleep to lower its power consumption when no messages are expected to be received. However, this mechanism limits the operation of the range extender to interval-based transmissions, making it unsuitable for applications like fire detection that require alarm-based operation. Changing the transmission interval of a smart sensor also necessitates a new observation phase. For a large-scale deployment, a management protocol is essential; otherwise, all relay stations would need to be manually restarted for the observation phase.

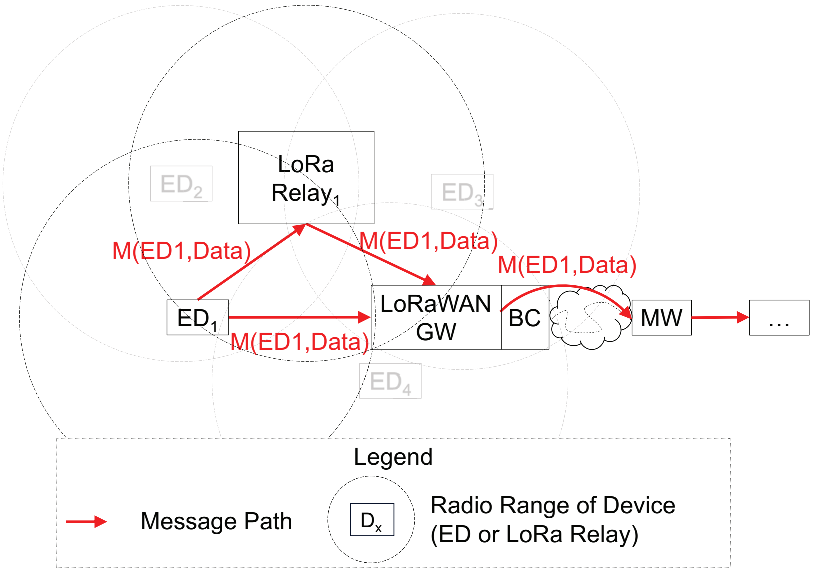


Figure 6: Uplink consideration for LoRa Relay scenario (GW = gateway; MW = middleware; ED = end-device; BC = backend connectivity; M(x,y) = message with sender device ID x and payload y).

An end-device's message can be received by both a LoRaWAN gateway and a LoRa relay simultaneously, with the relay forwarding the same message to the same LoRaWAN gateway, as illustrated in Figure 6. End-devices remain registered with the middleware. Unlike end-device, LoRa relays are not registered with a middleware. Additionally, LoRa relays cannot be cascaded, which imposes limitations on their operational range.

Satellite-based internet access, as exemplified with Starlink, offers the potential for deploying sensor networks independent of cellular coverage. Depending on the use case, the utilization of satellite-based internet can be advantageous. However, for environments like mines, such a system does not work. In forestry applications, the use of such technology is feasible, but several considerations must be taken into account. Firstly, the power consumption of such a system is significantly higher compared to a mobile network router, particularly complicating autonomous operation using e.g., a photovoltaics system. Secondly, a predominantly unobstructed view of the sky is required, which is hindered, in particular, by forest canopy. In environments where conventional mobile network coverage is absent for extended distances, this solution can be employed for the gateway with internet connectivity, thereby increasing LPWAN coverage through the use of range extenders.

The previously introduced range extenders are primarily engineered for low-power operation, limiting their capacity to support only a small number of smart sensors per extender. These solutions are not well-suited for network scenarios involving a higher number of end-devices. Additionally, these extender solutions lack the capability to be cascaded, necessitating the deployment of additional LoRaWAN gateways with internet access. Nevertheless, it's worth noting that cellular coverage is frequently unavailable in high attenuation environments. Consequently, there is a need for a range extender solution that can address these limitations.

In the upcoming section, a novel approach called LoRaWAN GW2ED range extender is introduced. The range extender solutions presented in this section will later be compared to this innovative approach.

3 LoRaWAN GW2ED Range Extender Solution

This section presents a novel approach for extending LoRaWAN coverage, called LoRaWAN GW2ED Range Extender. Compared to known range extenders so far, this solution is capable of operating in network scenarios with a higher number of end-devices. It is fully compatible with all existing LoRaWAN COTS smart sensors. In contrast to LoRaWAN relays, which require a firmware update that may not be offered from vendors, the new range extender solution offers seamless compatibility. Furthermore, these range extenders can be cascaded, resulting in extensive coverage, which can be achieved with just one LoRaWAN gateway with internet access.

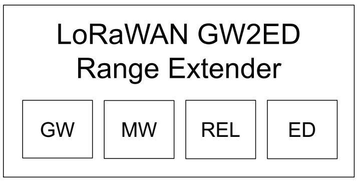


Figure 7: Internal architecture of a LoRaWAN GW2ED Range Extender (GW = gateway; MW = middleware such as Chirpstack; REL = range extender logic; ED = end-device).

The new approach, LoRaWAN GW2ED, consists of four primary components, as depicted in Figure 7. The first component is a LoRaWAN gateway, providing both uplink and downlink capabilities for connected devices. The second component is an offline middleware connected to the gateway. The third component is the range extender (RE) logic, and the last component is a LoRaWAN end-device responsible for relaying data received from the RE logic to the next LoRaWAN GW2ED range extender or LoRaWAN gateway with backend connectivity. Additionally, the end-device can also receive downlink messages, which are then forwarded to the RE logic. The RE logic acts as an intermediary between the middleware and the range extender's end-devices, managing uplink and downlink messages from end-devices and handling remote management tasks, including device registration and error message forwarding. It encapsulates other end-devices' message payloads and device information for later mapping of payload data to the respective end-device.

Device Registration and Encrypted Communication

In this solution, end-devices are registered within the middleware of the nearby range extender. This registration can be performed remotely with the help of a management protocol. The remote registration process includes the exchange of a device-specific Application Keys, which are necessary for OTAA. As a result, communication between end-devices and the middleware is encrypted. The device registration process, including the encrypted segments, is illustrated in Figure 8.

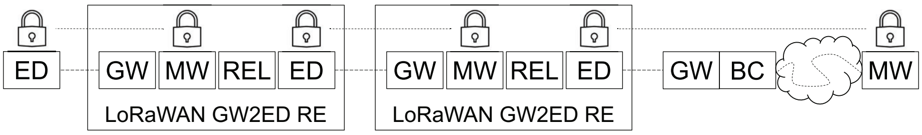


Figure 8: Device registrations within LoRaWAN GW2ED Range Extender scenario (GW = gateway; MW = middleware; REL = range extender logic; ED = end-device; BC = backend connectivity).

In LoRaWAN, each end-device can only be registered with one middleware. Messages originating from end-devices, which might be received by multiple range extenders, are exclusively processed by the specific range extender where the end-device is registered. An uplink message scenario is visualized in Figure 9. In this scenario, the end-device ED_1 sends an uplink message, like a temperature reading. Both GW2ED range extenders, RE_2 and RE_3, receive the message as they are within the radio range of ED_1. However, ED_1 is registered in the middleware of RE_3. Therefore, RE_2 discards the message, and the message is encapsulated in RE3 before being relayed further. In this case, the message is received by RE_1 and RE_2. Once again, RE_2 discards the message, as the end-device registered in RE_3 is unknown to RE_2. RE_1 then forwards the same encapsulated message to the LoRaWAN gateway. After reaching the (cloud-based) middleware, a decapsulation entity extracts the message to retrieve the source end-device information and the payload of ED_1.

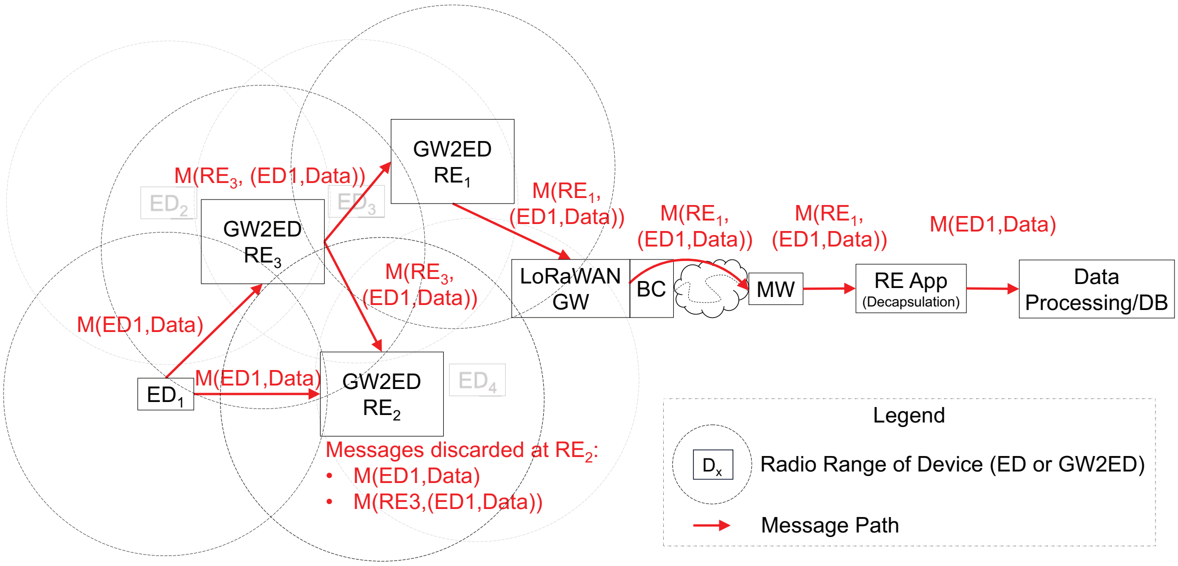


Figure 9: Uplink consideration for LoRaWAN GW2ED Range Extender scenario. (GW = gateway; MW = middleware; ED = end-device; RE = range extender; BC = backend connectivity; M(x,y) = message with sender device ID x and payload y; M(x,(y_1,y_2)) = message with sender device ID x, source device id y_1 and payload y_2).

Uplink Messages

Next, consider the example sequence diagram, which elucidates the sequence of events for an uplink message from a LoRaWAN end-device, as depicted in Figure 10. In this scenario, an end-device is already registered with the middleware of the adjacent LoRaWAN GW2ED range extender. It is assumed that a successful Join procedure, encompassing key exchange, has been previously completed.

1. The end-device transmits its uplink data using LoRa.
2. The LoRaWAN gateway of the range extender receives the message and forwards it to the middleware, possibly employing the Semtech UDP Packet Forwarder.
3. The middleware receives the message and passes it to a data integration service, which may publish the message on an MQTT broker.
4. The range extender logic component receives the message, encapsulates it by adding information related to the original end-device to the original payload, and forwards the message to the range extender's end-device. The end-device sends the encapsulated message using LoRa.
5. If there are additional range extenders, it's recognized that the message has already been relayed. Therefore, no further information regarding the uplink message is added.
6. Through the last gateway in the series, which likely has internet connectivity, the received data is sent to the cloud-based middleware.
7. The message is decapsulated to extract the source end-device information and the payload of ED$_1$.

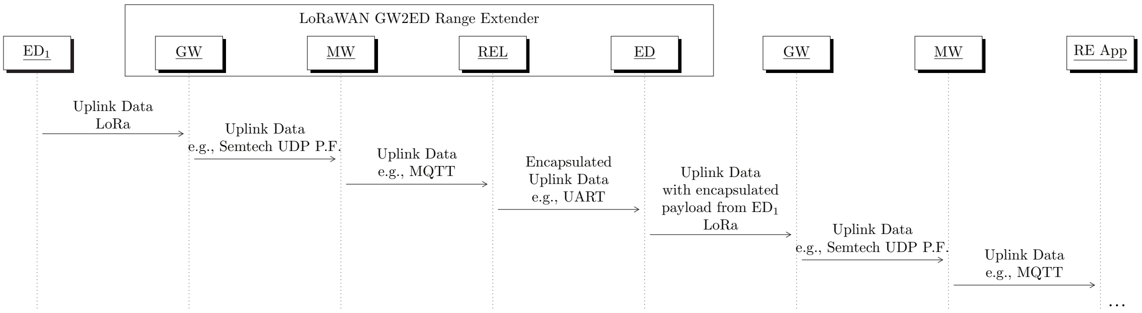


Figure 10: Sequence diagram of an end-device's uplink message when utilizing a LoRaWAN GW2ED Range Extender. (GW = gateway; MW = middleware; ED = end-device; RE = range extender; REL = range extender logic).

Downlink Messages

Downlink messages are also supported by the LoRaWAN GW2ED Range Extender. In this process, downlink messages from the cloud-based middleware for a specific end-device are relayed to the end-device through one or more range extenders. The sequence diagram in Figure 11 illustrates the flow of a downlink message. In this example, the end-device of the range extender is a LoRaWAN Class C device, ensuring constant accessibility.

1. A range extender application creates a downlink for the end-device of the range extender with an encapsulated downlink payload for ED1 and passes it to the middleware.
2. The cloud-based middleware (on the right) forwards the downlink message to the gateway, which then relays the message to the end-device of the range extender via LoRa.
3. The range extender logic component decapsulates the payload and queues a new downlink message for ED1 at the range extender's middleware.
4. In this example, the target end-device is a LoRaWAN Class A device, which means it must wait for an uplink message to open a downlink window. Once this occurs, the middleware of the range extender forwards the downlink message through the gateway to the target end-device.

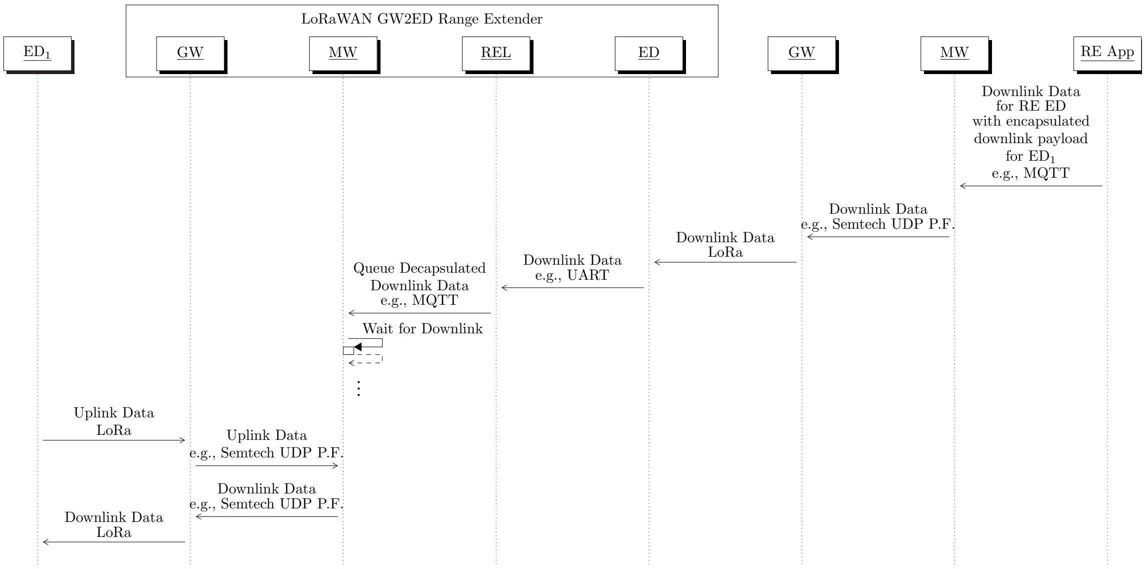

LoRaWAN GW2ED Range Extender
ED_1
GW
MW
REL
ED
GW
MW
RE App
Downlink Data for RE ED with encapsulated downlink payload for ED_1 e.g., MQTT
Downlink Data e.g., Semtech UDP P.F.
Downlink Data LoRa
Downlink Data e.g., UART
Queue Decapsulated Downlink Data e.g., MQTT
Wait for Downlink
Uplink Data LoRa
Uplink Data e.g., Semtech UDP P.F.
Downlink Data e.g., Semtech UDP P.F.
Downlink Data LoRa

Figure 11: Sequence diagram of a downlink message to an end-device in a LoRaWAN GW2ED Range Extender scenario. (GW = gateway; MW = middleware; ED = end-device; RE = range extender; REL = range extender logic).

Remote Management

The range extender can only be accessed remotely through LoRa messages, requiring a remote management protocol for configuration changes. This protocol encompasses tasks like adding new end-devices to a range extender and modifying the network's topology, which determines the connections between different range extenders. To establish these connections, the end-device of one range extender must be registered with the middleware of the next range extender, as previously illustrated in Figure 8.

The process of adding a new end-device to a range extender is visualized in the sequence diagram in Figure 12. Here, the range extender application generates a downlink message to initiate the addition of a new device. This message includes information specifying the target range extender, as well as details about the new end-device, including the DevEUI (64-bit) and the AppKey (128-bit).

1. The range extender application generates the downlink message, which is then sent through the middleware via the gateway to the end-device of the range extender. If the message is intended for another range extender, it is relayed through a series of downlink messages until it reaches the relevant range extender.
2. Within the range extender, the range extender logic component registers the new device using the middleware's API.
3. At a later point in time, the target end-device can initiate an OTAA Join procedure at the middleware of the respective range extender.

4. Upon the successful completion of the standardized Join Accept response, the end-device becomes operational and ready to communicate.

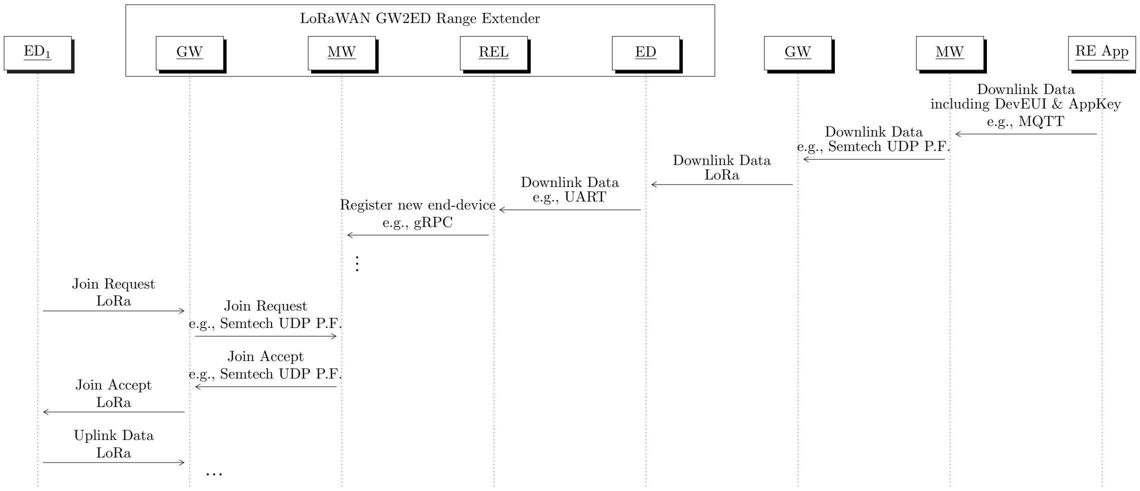

LoRaWAN GW2ED Range Extender
ED$_1$
GW
MW
REL
ED
GW
MW
RE App
Downlink Data including DevEUI & AppKey e.g., MQTT
Downlink Data e.g., Semtech UDP P.F.
Downlink Data LoRa
Downlink Data e.g., UART
Register new end-device e.g., gRPC
Join Request LoRa
Join Request e.g., Semtech UDP P.F.
Join Accept e.g., Semtech UDP P.F.
Join Accept LoRa
Uplink Data LoRa
...

Figure 12: Sequence diagram of the end-device registration process for a LoRaWAN GW2ED Range Extender scenario as part of the remote management procedure. (GW = gateway; MW = middleware; ED = end-device; RE = range extender; REL = range extender logic).

4 Implementation of LoRaWAN GW2ED Range Extender

This section describes the PoC implementation of the LoRaWAN GW2ED Range Extender with the following components and configurations:

- **Gateway:** The Dragino DLOS8N served as the gateway in this implementation. It utilized the Semtech UDP Packet Forwarder for data forwarding, directing data to the local middleware.
- **Middleware**: Chirpstack, an open-source middleware, was chosen for this setup. It was run in Docker containers on a Raspberry PI 3B operating Raspberry PI OS. Additionally, an MQTT Broker, specifically Mosquitto, was deployed as a Docker container.
- **End-Device:** The LoRaWAN USB Adapter LA66 by Dragino was employed as the end-device in this implementation. It was connected to the Raspberry PI and configured as part of the LoRaWAN network. This end-device was registered with the cloud-based middleware, which was also Chirpstack. A second end-device from a second range extender with an identical setup was registered with the middleware of the first range extender.
- **Range Extender Logic:** The range extender logic, responsible for managing communication and data relay, was developed using the Python programming language. This Python application included an MQTT subscription to receive data from the middleware. The received data was then forwarded to the LoRaWAN USB Adapter end-device using a serial connection. This allowed the end-device to send uplink messages to the next LoRaWAN gateway or range extender. Furthermore, the range extender logic processed downlink data received at the range extender's end-device. It analyzed the payload to determine the target of the message. The downlink message could be intended for the same range extender, another range extender, an end-device registered with the local middleware, or an end-device registered with another range extender. For tasks such as new device registration, the script utilized a gRPC API interface to register the device with its local middleware. To transmit downlink messages to another range extender, the range extender logic scheduled the downlink message by publishing it at the MQTT Broker, addressing the end-device of the next range extender.

Regrettably, it was found that the LoRaWAN USB Adapter could operate solely as a LoRaWAN Class A device. As a result, downlink messages were only accessible to the range extender when a downlink window was activated following an uplink message. An adapted sequence diagram for this situation is depicted in Figure 13.

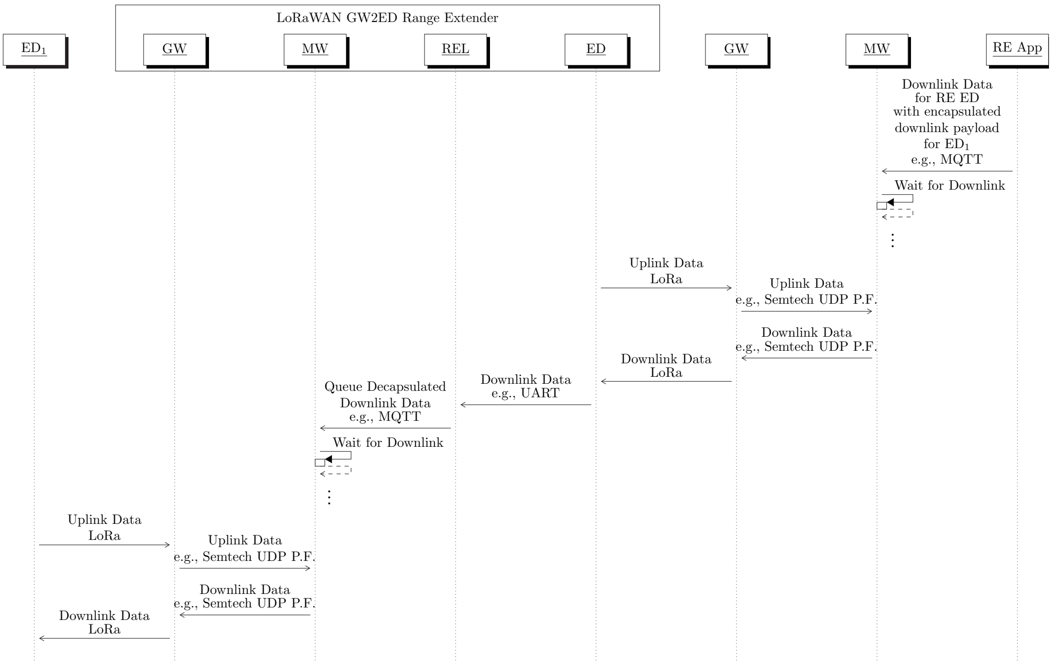

LoRaWAN GW2ED Range Extender
ED$_1$
GW
MW
REL
ED
GW
MW
RE App
Downlink Data for RE ED with encapsulated downlink payload for ED$_1$ e.g., MQTT
Wait for Downlink
Uplink Data LoRa
Uplink Data e.g., Semtech UDP P.F.
Downlink Data e.g., Semtech UDP P.F.
Downlink Data LoRa
Downlink Data e.g., UART
Queue Decapsulated Downlink Data e.g., MQTT
Wait for Downlink
Uplink Data LoRa
Uplink Data e.g., Semtech UDP P.F.
Downlink Data e.g., Semtech UDP P.F.
Downlink Data LoRa

Figure 13: Sequence diagram of a downlink message in the PoC implementation of the LoRaWAN GW2ED Range Extender scenario (GW = gateway; MW = middleware; ED = end-device; RE = range extender; REL = range extender logic).

5 Evaluation

This section evaluates the approach of the LoRaWAN GW2ED range extender. The PoC implementation is tested for functionality and range in a high attenuation environment. Following this, the new solution is compared with other range extender solutions.

Evaluation of PoC Implementation

Two PoC range extenders were tested within a forested area, which represents a high attenuation environment. Figure 14 shows an outdoor, solar-powered LoRaWAN gateway with internet access, while the other images portray the range extender enclosed in a case and affixed to a tree. All antennas and gateways were positioned at approximately 4 meters in height. The LoRaWAN field tester from Adeunis has been used as an end-device, which was registered with the second range extender's middleware. To access the performance, debugging messages transmitted from the range extender to the cloud-based middleware were used to determine the RSSI and spreading factor of each range extender.

Without the range extender, the field tester, registered at the cloud-based middleware, achieved a range of approximately 700 meters using Spreading Factor 7 and around 1000 meters when using Spreading Factor 12. Remarkably, similar results were obtained with the range extender, signifying that each range extender effectively extended the range by approximately 1 kilometer.

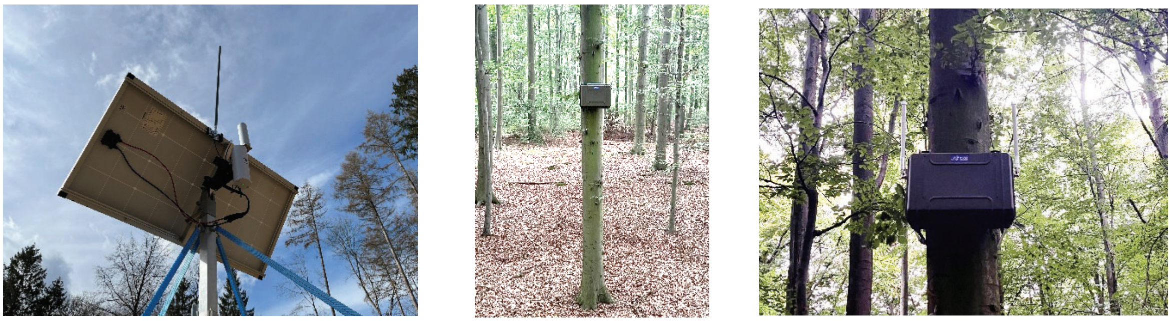


Figure 14: Solar-powered LoRaWAN gateway with internet access and the LoRaWAN GW2ED Range Extender located in a forest in Wolfenbüttel, Germany.

Comparison of the LoRaWAN GW2ED Range Extender with other LoRaWAN range extender solutions

Table 1 offers a comprehensive comparison between the developed solution presented in this work and other range extender solutions introduced in Section 2. The comparison covers various aspects, each of which will further be elaborated while discussing several of these aspects.

	LoRaWAN Relays	**LoRa Relays**	**LoRaWAN GW2ED**
Extendibility by multi-hop	No	No	Yes
Commercial availability	No	No	No
Remote management	Yes	No	Yes
Maximum number of supported end-devices	16	Low	High
Downlink capabilities	Yes	Yes	Yes
Power consumption	Low	Low	Higher
Support of COTS end-devices	(Yes)	Yes	Yes

Table 1: Comparison of LoRaWAN GW2ED Range Extender solution with other range extender solutions.

Following the discussion of the aspects of the comparison:

Extendibility by multi-hop: This aspect is about the support for cascading multiple range extenders. Only the new solution (LoRaWAN GW2ED Range Extender) provides this feature. Cascading range extenders is inevitable for operating sensor networks in high attenuation environments with inadequate cellular network coverage.

Commercial availability: Currently, none of the presented range extender concepts are currently commercially available, even though LoRaWAN Relays are already standardized in Version 1.0.

Remote management: Remote management includes features like remote device registrations or modifying the range extender topology. It is supported by both LoRaWAN Relays and LoRaWAN GW2ED Range Extender. However, LoRa Relays, as presented, do not currently support remote management, which limits their suitability for larger deployments.

Maximum number of supported end-devices: The LoRaWAN Relays specification limits the number of end-devices per relay to 16. The number of supported end-devices for the LoRa Relay is not explicitly defined, but it's estimated to be low due to the relay having to wake up for each uplink message. In contrast, the LoRaWAN GW2ED solution supports scenarios with a higher number of end-devices, limited only by the compliance to the duty cycle regulations.

Downlink capabilities: All solutions support the use of downlink mechanisms.

Power consumption: LoRaWAN Relays and LoRa Relays are designed for a battery-powered operation and have low power consumption. The current LoRaWAN GW2ED concept is not optimized for low power consumption, resulting in a higher power consumption compared to the other range extender solutions. Additionally, the PoC implementation of the LoRaWAN GW2ED solution is also not optimized for low power consumption. To reduce energy consumption of the LoRaWAN GW2ED solution, an interval-based operation is possible, as described in Reference [BW23]. In this approach, the sending interval of end-devices is synchronized with the on-time of gateways using downlink messages, allowing gateways (or in this case, the range extender) to be turned off during off-time intervals, significantly reducing power consumption. It is important to note that all range extender solution are currently available as PoC implementations, making it challenging to compare their power consumption accurately.

Support for Commercial Off-The-Shelf (COTS) end-devices: Support of COTS end-devices is available for all solutions. However, COTS end-devices using LoRaWAN Relays require firmware updates in order to support the new relaying specification.

6 Conclusion & Future Work

The deployment of LoRaWAN in high attenuation environments presents significant challenges for establishing efficient sensor networks due to the considerable reduction in the LoRaWAN range. Recognizing this need for LoRaWAN range extension, as evidenced by the recent LoRaWAN relay specification, this work has proposed a new approach for a LoRaWAN range extender, the LoRaWAN GW2ED Range Extender, which operates independently of internet connectivity. Existing specified range extender solutions have limitations, such as the absence of cascading mechanisms. The LoRaWAN GW2ED Range Extender is unique in providng extendibility through multi-hop capability.

The results demonstrate that each range extender effectively extends the LoRaWAN range by a distance comparable to that of a standard LoRaWAN gateway. In the tested scenario within a forest environment, an initial LoRaWAN range of approximately 1 km was extended by an additional 1 km for each added range extender. Importantly, this solution is fully compatible with existing LoRaWAN COTS end-devices and does not require any firmware modifications on these devices. Furthermore, the solution supports both uplink and downlink capabilities and can be used in network scenarios with a higher number of end-devices. It adheres to existing LoRaWAN procedures and protocols without any alterations.

In the future, the integration of all components into a single device to reduce power consumption will be considered. Further refinement of the management protocol is essential to support additional features, such as error message handling within range extenders, including frame counter violations of end-devices. Features like multi-path routing are feasible but require changes to existing mechanisms as presented in this work. In multi-path routing scenarios, ensuring that downlink messages are not sent redundantly to end-devices is crucial.

Additionally, for areas lacking internet connectivity and aiming to extend LoRaWAN coverage, exploring a LoRaWAN gateway that utilizes satellite-based LoRa for uplink communication could be considered as an alternative approach [AB22].

7 Acknowledgment

This work was funded by the Federal Ministry of Transport and Digital Infrastructure as a part of the research project 5G Smart Country (45FGU117).

8 References

[AB22] Afhamisis, Mohammad, Sebastian Barillaro, and Maria Rita Palattella, "A Testbed for LoRaWAN Satellite Backhaul: Design Principles and Validation," in 2022 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2022.

[BW23] M. Böhm and D. Wermser, “Sensor Networks for Forestry Applications operating with Limited Power Supply using LPWAN COTS Equipment,” in 2023 Mobile Communication-Technologies and Applications, 27. ITG-Symposium. VDE, 2023

[CS23] ChirpStack, open-source LoRaWAN® Network Server, https://www.chirpstack.io/

[MC19] D. Mamour and P. Congduc, “Increased flexibility in long-range IoT deployments with transparent and light-weight 2-hop LoRa approach,” in 2019 Wireless Days (WD). IEEE, 2019, pp. 1–6.

[SE23] The New LoRaWAN® Relay Feature: A Powerful Tool for LoRa® and LoRaWAN Networks, https://blog.semtech.com/the-new-lorawan-relay-feature

[TS22] LoRaWAN® Relay Specification TS011-1.0.0, https://resources.lora-alliance.org/technical-specifications/ts011-1-0-0-relay

[VL19] M. R. Villarim, J. V. H. de Luna, D. de Farias Medeiros, R. I. S. Pereira, C. P. de Souza, O. Baiocchi, and F. C. da Cunha Martins, “An evaluation of LoRa communication range in urban and forest areas: A case study in brazil and portugal,” in 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2019, pp. 0827–0832.

http://dx.doi.org/10.25673/111735

Positionierung und Vermessung von Komponenten für Indoor-Lokalisierungs- und drahtlose Kommunikationssysteme in Industrieumgebungen

Florian Hufen, Timo Siekmann, Harry Fast, Holger Flatt, Sebastian Schriegel
Fraunhofer IOSB, Institutsteil für industrielle Automation (IOSB-INA, Campusallee 1, 32657 Lemgo
{florian.hufen, timo.siekmann, harry.fast, holger.flatt, sebastian.schriegel}@iosb-ina.fraunhofer.de

Abstract: Der Beitrag beschäftigt sich mit der Positionierung und Vermessung von Komponenten für Indoor-Lokalisierungs- und drahtlose Kommunikationssysteme in Industrieumgebungen. Bisher wurden häufig Stativmontagen oder Flugdrohnen zur Positionsveränderung verwendet, die jedoch nicht immer flexibel einsetzbar sind und reproduzierbare Messungen erlauben. Daher wird ein neuartiges Konzept vorgestellt, welches ein mobiles Liftsystem zur automatisierten Positionierung der Komponenten im 3D-Raum nutzt. Das 3D-Liftsystem verfügt über einen horizontal sowie vertikal verfahrbaren Teleskoparm und bietet die Möglichkeit Langzeittests durchzuführen sowie Umgebungsveränderungen aufzuzeichnen. Das Konzept wurde als Fallstudie umgesetzt und im Kontext einer realen Industrieumgebung zur Vermessung von 5G-Komponenten angewendet. Ein Vergleich zu herkömmlichen Positionierungsmethoden anhand verschiedener Eigenschaften zeigt, dass das vorgeschlagene Konzept Eignungskriterien für umfangreiche, langwierige und unterbrechungsfreien Messkampagnen mit 79 % am höchsten erfüllt, während eine Umsetzung mit Stativen und Flugdrohnen nur 64 % bzw. 54 % erreichen.

1 Einleitung

Die Vermessung und Evaluation von Komponenten der Indoor-Lokalisierung und drahtlosen Kommunikation in Industrieumgebungen erfordert häufig eine flexible und reproduzierbare Positionierung des Testequipments in allen drei Dimensionen. So wird beispielsweise für den Betrieb von bestimmten 5G-Funknetzen eine maximale Feldstärke an der Grenze zu anderen benachbarten Funknetzen vorgeschrieben, welche in einer Höhe von 3 Metern über dem Boden gemessen wird [Euro20]. Bei der Entwicklung, Evaluierung und Zertifizierung von Indoor-Lokalisierungstechnologien wie z. B. Ultra Wide Band (UWB) oder Bluetooth Low Energy (BLE) werden ähnliche Messungen an verschiedenen Positionen im 3D-Raum durchgeführt [BBTM21]. Diese häufig wechselnden Messpositionen werden heute über eine Montage an Stativen [BBTM21] oder mithilfe ferngesteuerter Flugdrohnen [PRKS19] realisiert. Letztere sind jedoch häufig nur begrenzt positionsstabil und beschränkt in ihrer Einsatzdauer sowie der Kapazität, Testequipment mit Strom zu versorgen. Zudem sind Flugdrohnen in Bezug auf ihre Traglast und Einsatzgebiete stark reglementiert. Beim Einsatz von Stativen erfolgt der Positionswechsel häufig unter erhöhtem manuellem Aufwand durch die Montage oder einem erneuten Einmessen der Geräteposition, um eine Reproduzierbarkeit der Messungen zu gewährleisten.

In diesem Beitrag wird daher ein Konzept für ein mobiles 3D-Liftsystem mit ausfahrbarem Teleskoparm vorgestellt, welche die Nachteile der zuvor erwähnten „herkömmlichen“ Positionierungsmethoden adressiert. Dieses 3D-Liftsystem ermöglicht es, Geräte automatisiert an verschiedene Positionen im 3D-Raum zu fahren sowie die Vermessung von Funk- und Kommunikationstechnologien zu vereinfachen. Neben der Vorstellung des Konzeptes beschreibt der Beitrag ebenfalls dessen Umsetzung sowie Anwendung anhand verschiedener Beispiele. Im Anschluss werden die Eigenschaften des Liftsystems mit den herkömmlichen Positionierungsmethoden verglichen und ein Fazit gezogen.

2 Vorstellung des Konzeptes

Das Kernstück des neuartigen Messkonzeptes bzw. der Positionierungsmethode bildet ein mobiles 3D-Liftsystem mit ausfahrbarem Teleskoparm. Dieses 3D-Liftsystem ermöglicht es für Funk- und Kommunikationssysteme, an verschiedene Positionen im 3D-Raum zu fahren, und eröffnet dadurch vielfältige Möglichkeiten für automatisierte Untersuchungen im Bereich der drahtlosen Kommunikationstechnologie.

Ein Hauptvorteil dieses Liftsystems besteht in der effizienten und leicht reproduzierbaren Durchführung von Messungen. Durch die automatische Positionierung der Geräte im 3D-Raum können Messungen an verschiedenen Standorten durchgeführt werden, ohne manuelle Eingriffe oder Neupositionierungen. Dadurch wird Zeit gespart und die Genauigkeit und Zuverlässigkeit der Messungen verbessert.

Zusätzlich zu den Positioniermöglichkeiten, die die ausfahrbaren Teleskoparme bieten, wird die Mobilität des Systems bzw. die Größe des Messraumes durch die Kompatibilität zu fahrerlosen Transportsystemen (FTS) erhöht. Diese Mobilität ermöglicht einen flexiblen Einsatz des Liftsystems für verschiedene Anwendungsbereiche. So kann das mobile 3D-Liftsystem beispielsweise für die Vermessung von Funkabdeckungen in großen Gebäuden, wie z. B. Industriehallen, oder auf Freiflächen eingesetzt werden. Auch die Untersuchung der Signalqualität in verschiedenen Umgebungen oder von komplexen Forschungsfragen wird dadurch möglich.

Durch die Fähigkeit des 3D-Liftsystem, eine ausdauernde, Akku-basierte Stromversorgung für die drahtlosen Kommunikationssysteme zu Verfügung zu stellen, sind mit dem hier beschriebenen Konzept auch ganz- oder mehrtägige Messungen bzw. Messreihen möglich. Die Durchführung solcher Langzeittest ermöglicht eine detaillierte Untersuchung der Systeme unter realen Bedingungen und berücksichtigt Umgebungseinflüsse, die sich im Laufe der Zeit verändern können. Die dabei ermittelten (Mess-) Daten können über die ebenfalls von dem 3D-Liftsystem bereitgestellte WLAN-Schnittstelle direkt zur Auswertung übertragen werden. Ein Anwendungsbeispiel für solche Langzeittests aus dem Industrieumfeld ist die Untersuchung von lagerndem Material oder flexiblen Fertigungszellen. Durch die kontinuierliche Überwachung der Umgebung können Signalstärke, Interferenzen und andere relevante Parameter umfassend analysiert werden, um die Effizienz und Zuverlässigkeit der beobachteten Kommunikationstechnologie zu verbessern.

3 Umsetzung des Konzeptes durch Aufbau des Systems

Zur Umsetzung des im vorherigen Abschnitt erläuterten Konzeptes wurde das ebenfalls dort eingeführte 3D-Liftsystem konzipiert und schließlich konstruiert. Um die präzise und flexible Bewegung von Funkkomponenten in drei Dimensionen zu ermöglichen, besteht das 3D-Liftsystem aus verschiedenen Komponenten, die im Folgenden näher beschrieben werden. Abb. 1 gewährt dabei zunächst einen grafischen Überblick und benennt die einzelnen Bestandteile:

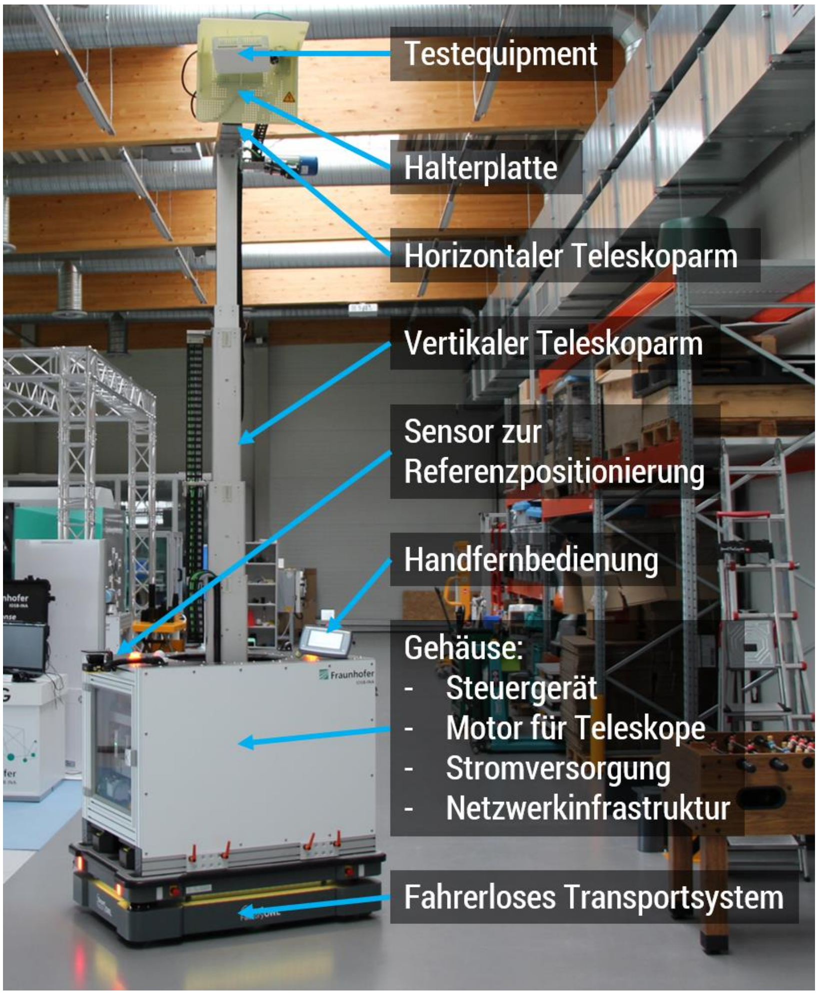


Abb. 1: Bestandteile des 3D-Positioniersystems

Das aus Aluminiumprofilen bestehende Gehäuse des 3D-Liftsystems beherbergt verschiedene Komponenten: Zunächst enthält es ein zentrales Steuergerät, welches die gesamte Funktionalität des Systems steuert. Außerdem befindet sich darin ein Motor, welcher für das Verfahren der beiden Teleskoparme verantwortlich ist. Des Weiteren beinhaltet das Gehäuse einen 3,2 kWh starken LiFePo-Akku sowie Spannungswandler bzw. Netzteile für 24V und 230V, welche sowohl die einzelnen Komponenten des Systems, als auch an dem Teleskoparm montierte Kommunikationssysteme für mehrtätige Messungen mit Strom versorgen. Eine weitere Möglichkeit zur Versorgung des Testequipments bietet die Gigabit-Netzwerkinfrastruktur durch einen verbauten Power over Ethernet (PoE) Switch. Die kabelgebundene Netzwerkinfrastruktur ermöglicht ebenfalls eine Datenübertragung von der Halterplatte bis zu den im Gehäuse verbauten WLAN-Gateway. Das 3D-Liftsystem wurde als Modul designt und hat die Grundmaße einer DIN Euro-Palette. Dadurch kann das 3D-Liftsystem auf vielfältige Weise, wie z. B. mit einem Hubwagen oder auf einem fahrerlosen Transportsystems, bewegt oder abgestellt werden. An der Außenseite des Gehäuses befindet sich außerdem eine abnehmbare Handfernbedienung, mit der die Teleskoparme durch manuelle Bedienung präzise in den drei Dimensionen gesteuert werden kann. Eine Regelung der Verfahrgeschwindigkeit ist ebenfalls möglich. Zur Absicherung des Betriebs ist an der Fernbedienung ein Totmannschalter vorhanden.

Des Weiteren ist an der oberen Gehäuseabdeckung ein Sensor zur Referenzpositionierung montiert. Dieser dient dazu, mithilfe von laserbasierten Messverfahren eine Referenzposition für das Liftsystems zu bestimmen. Dadurch können die Teleskoparme auf der Basis zum Referenzpunkt relativer Koordinaten bewegt werden, welches eine hohe Genauigkeit und Wiederholbarkeit der Bewegungen gewährleistet.

In der Mitte der oberen Gehäuseabdeckung befindet sich eine Aussparung, aus der der vertikale Teleskoparm herausragt. Dieser besteht aus vier einzelnen Segmenten und kann bis zu einer Höhe von 5 Metern in einer Abstufung von einem Millimeter ausgefahren werden. Zusätzlich ist eine Rotation um die eigene Achse mit bis zu 358° möglich. An der Spitze des vertikalen Teleskoparms ist ein weiterer, kleinerer Arm befestigt, welcher über 1,5 Metern horizontal ausgefahren werden kann. Beide Teleskoparme werden beim Verfahren von jeweils einem Motor angetrieben. Ist die gewünschte Position erreicht, wird aufgrund der selbsthemmenden Eigenschaft der Teleskoparme der Motor abgeschaltet. Dies führt nicht nur zu einer hohen Positionsstabilität, sondern wirkt sich auch positiv auf den Energiebedarf des Systems, besonders bei Langzeittests, aus.

Eine Halterplatte zur Montage von Testequipment befindet sich am Ende des horizontalen Teleskoparms. Dort können drahtlose Kommunikationssysteme oder andere Komponenten mit einem Gewicht von bis zu 15 kg sicher auf einer Lochplatte sowie durch zusätzliche Halterungen befestigt werden. Des Weiteren befinden sich dort eine Schutzkontaktsteckdose mit 230V Spannungsversorgung aus dem Gehäuse sowie ein Ethernet-Anschluss zur Anbindung an die Netzwerkinfrastruktur und PoE. Die dafür notwendigen Kabel werden mithilfe von flexiblen Kabelschleppen über die Teleskoparme ins Gehäuse geführt.

4 Anwendung des Konzeptes

Das konstruierte 3D-Liftsystem wurde in der SmartFactoryOWL, einem gemeinsamen Reallabor des Fraunhofer IOSB-INA und der Technischen Hochschule OWL für Industrie 4.0 sowie einem 5G-Anwendungszentrum, in Betrieb genommen und befindet sich dort für Fallstudien im Einsatz. Als Anwendungsbeispiel für das vorgestellte Positionierungs- und Vermessungskonzept werden im Folgenden bereits mithilfe des 3D-Liftsystems durchgeführte Messungen von 5G-Komponenten im Kontext von Industrieumgebungen beschrieben. Darauf folgt eine weitere Anwendungsmöglichkeit als Ausblick für eine zukünftig angestrebte Automatisierung eines Prüfprozesses für Indoor-Lokalisierungsgeräte.

4.1 Wireless Messungen mit dem 3D-Liftsystem in industriellen Umgebungen

In der Industrie 4.0 spielt drahtlose Kommunikationstechnik eine immer größere Rolle. So werden breitbandige drahtlose Kommunikationssysteme für die Vernetzung von Fahrzeugen, Menschen und Maschinen im Produktions- und Logistikumfeld eingesetzt [LSRP18]. Damit die verschiedenen Kommunikationssysteme wie Mobilfunk oder Wi-Fi ordnungsgemäß funktionieren, müssen diese professionell geplant und parametrisiert werden. Bereits in der Planungsphase kann eine Kartierung der Funkabdeckung sehr wertvoll sein. Neben der Funkabdeckung auf Personenhöhe fordern diverse Use-Cases immer öfter eine Funkabdeckung im 3D-Raum. Dazu zählen z. B. Kommunikationsmodule an Gabeln von Flurförderfahrzeugen oder mehrstöckige Anlagen und Maschinen wie Lackierstraßen. Eine Möglichkeit, theoretische Simulationen mit Messungen zu validieren, bietet hier das 3D-Liftsystem. An der Halterplatte am Ende des Teleskoparms kann ein Mobilfunkmodem zur Messung der Funkabdeckung angebracht werden. Die Spannungsversorgung erfolgt dabei über das 3D-Liftsystem selbst. Mit dem 3D-Liftsystem kann entweder an vielen verschiedenen Punkten gemessen oder eine statische Langzeitmessung durchgeführt werden.

5G-Campusnetze dürfen laut Gesetzgeber eine maximale Feldstärke an der Grundstücksgrenze der Campusnetzzuteilung nicht überschreiten. Diese Feldstärke wird jedoch auf einer Höhe von 3 m über dem Boden gemessen [Euro20]. Das 3D-Liftsystem kann während der Installation der aktiven Netzwerkkomponenten, wie z. B. von Radio Units, in kritischen Bereichen platziert werden und als Messwertaufnehmer unterstützen. Speziell im 5G-Bereich können relevante Parameter der Basisstation iterativ unter Berücksichtigung der Feldstärke am Messwertaufnehmer angepasst werden. Relevante Parameter für die Funknetzabdeckung und somit auch der Feldstärke am Rand von Grundstücken sind der horizontale und vertikale Öffnungswinkel der Remote Radio Unit, Tilt der Radio Unit, Antenna Gain sowie die Sendeleistung. Flugdrohnen bzw. Unmanned Area Vehicles (UAV) eignen sich nicht für diese Art der Messung, da diese häufig nicht innerhalb von Gebäuden fliegen können und oft eine aktive Steuerung bzw. Kontrolle benötigen, wodurch sich der Arbeitsaufwand für die Person erhöht, welche die 5G-Einstellungen parametrisiert. Aufgrund der Möglichkeit nach Erreichen der gewünschten Position das 3D-Liftsystem zu parken, ist kein aktiver Eingriff während der Messung mehr notwendig. Außerdem sind die Teleskoparme durch die Verwendung von Gewindespindeln eigensicher und werden selbst bei Spannungsverlust des Systems nicht unbeabsichtigt eingefahren.

Eine weitere Anwendung im 5G-Bereich ist die Verwendung des 3D-Liftsystems als mobiler Träger einer Remote Radio Unit, wie in der Abb. 2 zu sehen ist. Die Funkeinheit kann an der Halteplattform angebracht und mit Spannung versorgt werden, dadurch ist ein 5G-Netzwerk sofort und ohne baulichen Eingriff in Industriehallen einsetzbar. Mittels dieser Technik können mögliche Positionen von Mobilfunkinfrastruktur im Rahmen einer Messkampagne evaluiert werden.

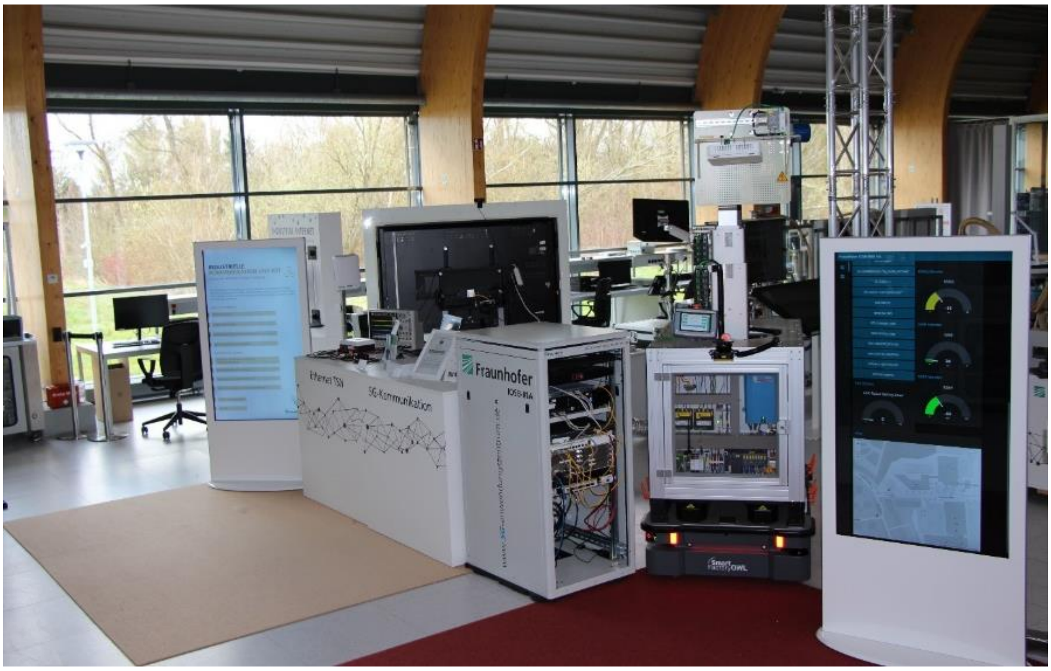


Abb. 2: 5G-Anwendungszentrum mit 3D-Liftsystem

Neben der Verwendung zur Positionierung von Mobilfunkinfrastruktur und -endgeräten eignet sich das 3D-Liftsystem auch für Wi-Fi Clients und Wi-Fi Access Points. Wi-Fi Clients werden in Hand-Scannern und Gabelstaplern in der Industrie- und Lagerlogistik eingesetzt, auch hier ist eine lückenlose Funknetzabdeckung möglich. Diese wird ebenfalls mit dem 3D-Lifsystem oberhalb der Personenhöhe ermöglicht. Dazu kann das 3D-Liftsystem auch Wi-Fi Roaming-Tests in Problem- oder Grenzzonen ermöglichen. Das Wi-Fi Roaming wird dann initiiert, wenn ein Wi-Fi Access Point in der näheren Umgebung bessere Kommunikationseigenschaften liefert als der aktuell verwendete.

4.2 Automatisierung von Prüfprozessen für Komponenten von Indoor-Lokalisierungssystemen

Omlox ist ein offener Standard für Ultra Wide Band (UWB) basierte Echtzeit- bzw. Indoor-Lokalisierungssysteme in Industrieumgebungen, welcher von der Profibusnutzerorganisation (PNO) zertifiziert wird. Neben Softwarekomponenten besteht ein omlox-System aus einer Satelliteninfrastruktur und so genannten „Trackables" bzw. Tags. Mithilfe verschiedener, z. B. Laufzeit-basierter Verfahren kann die Position eines Trackables innerhalb einer von mehreren Satelliten aufgespannten Lokalisierungszone ermittelt werden. Um ein Zertifikat für ein omlox-System erhalten zu können, müssen Hersteller ihre Geräte von einem unabhängigen, akkreditierten Labor prüfen lassen, welches diverse Test durchführt. Diese Tests erfordern unter anderem die Durchführung von mehreren Messungen mit omlox-Tags an verschiedenen Positionen in zwei oder drei Dimensionen innerhalb einer Lokalisierungszone von 3 x 6 x 3 Metern. Bisher wird die Vielzahl Umpositionierungen der Tags von Labormitarbeiter manuell durchgeführt, was viel Zeit in Anspruch nimmt und die Kosten einer Prüfung erhöht.

Die Umsetzung des in diesem Beitrag vorgeschlagenen Konzeptes könnte in Zukunft den Prüfprozess durch eine automatisierte Positionierung von Tags in Bezug auf Zeit und Kosten deutlich optimieren. Unter der Voraussetzung, dass im Prüflabor ausreichen Platz vorhanden ist und kein Tag niedriger als die Höhe des komplett eingefahrenen vertikalen Teleskoparms positioniert werden muss, kann die Anwendung des zuvor beschriebenen 3D-Liftsystems sinnvoll sein. Diese Einschränkung einer notwendige Minimalhöhe für die Positionierung, könnte durch eine zukünftige Verkleinerung des 3D-Liftsystems oder durch eine Anhebung des Bodenlevels der Lokalisierungszone auf Softwareebene vermieden werden. Alternativ könnte das 3D-Liftsystem lediglich zur Positionierung bei weiterführenden Tests in realen Produktionsumgebungen, außerhalb eines reglementierten Prüflaborsetups eingesetzt werden, z. B. in großen Lagerhallen.

5 Auswertung

Die Umsetzung des Konzeptes lässt einen qualitativen Vergleich mit herkömmlichen Positionierungsmethoden zu. In Tab. 1 werden die verschiedenen Methoden hinsichtlich ihrer Eignung für die Umsetzung von Messungen verglichen. Die Bewertung wird dabei anhand der Eignung für verschiedene Aspekte, welche bei einer Messung von Bedeutung sind, vorgenommen. Aus der Summe der Bewertungspunkte (siehe Aufschlüsslung in der Bewertungsskala der Tabelle) wird eine Gesamteignung berechnet.

Für die Umsetzung von umfangreichen, langwierigen und unterbrechungsfreien Messreihen an einer Position sind insbesondere die dauerhafte Verfügbarkeit des Systems von Bedeutung. Diese ist durch die Eigenschaften des Systems in vollem Umfang gegeben, da das 3D-Liftsystem nicht nur über eine Stromversorgung für den Eigenbetrieb verfügt, sondern auch für die zu vermessenden Geräte Energie bereitstellt. Zudem arbeitet das 3D-Liftsystem besonders energieeffizient, da die Teleskoparme selbsthemmend sind und keinen Strom verbrauchen, wenn sie nicht gerade verfahren werden. Herkömmliche Stative sind ebenfalls für langwierige Messreihen geeignet, bieten aber nicht den Komfort einer integrierten Energieversorgung für Messequipment. Flugdrohnen bzw. UAVs sind für lange, umfangreiche und unterbrechungsfreie Messreihen aufgrund der begrenzten Akkukapazitäten nicht geeignet.

Die Genauigkeit der Positionierung von Geräten mithilfe von Stativen und auch dem 3D-Liftsystem liegt im Zentimeterbereich. Bei Stativen ist die Positioniergenauigkeit von der verwendeten Messmethode abhängig (nicht Gegenstand dieser Betrachtung). Dabei erfordert jeder Positionswechsel ein erneutes Einmessen und erzeugt somit einen hohen Aufwand bei der Einrichtung weiterer Messpositionen. Gegebenenfalls sind mit dem Stativ auch genauere Positionierungen (im Millimeterbereich) möglich. Dahingegen erfordert ein mit dem 3D-Liftsystem durchgeführter Positionswechsel kein erneutes Einmessen, stattdessen wird die Position auf der Basis des zurückgelegten Verfahrweges zentimetergenau ermittelt. Bei Flugdrohnen ist ebenfalls eine einfache Neupositionierung umsetzbar, allerdings ist eine dauerhafte und genaue Positionierung nicht immer möglich, z. B. bei Luftzügen oder ähnlichen Umgebungseinflüssen [PRKS19]. Dies wirkt sich bei UAVs auch auf die Reproduzierbarkeit von Tests aus, wenn z. B. eine Position erneut angefahren werden muss. Zudem führt die häufig fehlende GPS-Verbindung im Indoor-Bereich bei Flugdrohnen zu grundsätzlich Problemen bei der automatisierten Positionierung. Auch bei der Verwendung von Stativen ist eine exakte Reproduktion einer vorherigen Messposition nicht ohne einen erhöhten Aufwand möglich, da diese erst durch wiederholtes Einmessen und ggf. iterative Anpassungen gefunden werden muss. Hier bietet das 3D-Liftsystem ebenfalls den Vorteil, dass die exakte Position gespeichert und einfach erneut angefahren werden kann.

Durch die Bauweise des 3D-Liftsystems besteht die Möglichkeit, Komponenten bis zu einem Gewicht von 15 kg aufzunehmen. Zudem kann das Equipment sicher montiert werden. Stative hingegen müssen für eine größere Traglast entsprechend ausgelegt sein und können somit schnell unhandlich werden. Flugdrohnen können aufgrund ihrer Bauweise häufig nur geringe Lasten tragen und bergen außerdem die Gefahr eines Absturzes durch Kollision.

Durch seinen robusten Aufbau steht das 3D-Liftsystem jedoch der Flugdrohne und dem Stativ hinsichtlich Kompaktheit und Einsatzfähigkeit nach. Während Stative auch aufgrund ihrer Kompaktheit nahezu an jedem Ort eingesetzt werden können, bedarf es für den Einsatz des 3D-Liftsystems eines ebenen und festen Untergrunds sowie einer geeigneten Transportmöglichkeit (z. B. ein fahrerloses Transportsystem) für eine Verlegung des Systems. Flugdrohnen sind ebenfalls sehr kompakt aufgebaut, haben aber auch Einschränkungen, was die Einsatzumgebung betrifft. So gelten hier besondere Regeln, wie beispielsweise in der Nähe von Windrädern oder Flughäfen.

Bei gleicher Gewichtung aller Kriterien zeigt die Auswertung, dass das 3D-Liftsystem die meisten Eignungskriterien für umfangreiche, langwierige und unterbrechungsfreie Messkampagnen erfüllt.

Kriterium	**Stativ**	**Flugdrohne**	**3D-Liftsystem**
Einsatzdauer am Stück	++	--	++
Positionsgenauigkeit	++	0	+
Einrichtungsaufwand neuer Messpositionen	--	++	++
Reproduzierbarkeit von Messpositionen	--	-	++
Traglast & Sicherheit	+	-	++
Einsatzumgebungen	+	+	0
Kompaktheit	++	++	-
Gesamtbewertung	18/28 (64 %)	15/28 (54 %)	22/28 (79 %)
Bewertungsskala - Sehr Gut (++, 4 Punkte) - Gut (+, 3 Punkte) - Ausreichend (0, 2 Punkte) - Größtenteils ungeeignet/schlecht (-, 1 Punkt) - Ungeeignet (--, 0 Punkte)			

Tab. 1: Qualitativer Vergleich verschiedener Positionierungsmethoden hinsichtlich verschiedener Kriterien

6 Zusammenfassung & Ausblick

In diesem Beitrag wurde ein neuartiges Konzept zur Positionierung und Vermessung von Komponenten für Indoor-Lokalisierungs- und drahtlose Kommunikationssysteme in Industrieumgebungen vorgestellt. Herkömmliche Positionierungsmethoden und -konzepte haben teilweise signifikante Nachteile, wie einen erhöhten Zeitaufwand, begrenzte Traglasten, sowie mögliche Ungenauigkeiten und damit eine geringe Reproduzierbarkeit von Positionen. Als Alternative zu diesen herkömmlichen Methoden wurde das Konzept einer automatisierten Positionierung mithilfe eines 3D-Liftsystems eingeführt und dessen technische Umsetzung und die dadurch entstehenden Vorteile beschrieben. Darauf folgte eine Validierung des Liftsystems durch eine Fallstudie zur Vermessung von 5G-Komponenten sowie ein Ausblick auf ein weiteres mögliches Anwendungsfeld bei der Prüfung von Indoor-Lokalisierungssystemen wie z. B. dem neuen, herstellerübergreifenden Indoor-Lokalisierungsstandard omlox. Abschließend wurden die verschiedenen Positionierungsansätze qualitativ miteinander verglichen und ihre Vor- und Nachteile aufgezeigt. Dabei konnte gezeigt werden, dass das vorgeschlagene Konzept die Kriterien für eine umfangreiche, langwierige und unterbrechungsfreie Messkampagne mit 79 % erfüllt, während eine Umsetzung mithilfe von Stativen oder Flugdrohnen dies nur zu 64 % bzw. 54 % tun. Demnach ist das Positionierungskonzept des 3D-Liftsystems deutlich besser geeignet als die verglichenen Ansätze.

7 Literaturverzeichnis

[BBTM21] Barbieri, Luca ; Brambilla, Mattia ; Trabattoni, Andrea ; Mervic, Stefano ; Nicoli, Monica: UWB Localization in a Smart Factory: Augmentation Methods and Experimental Assessment. In: *IEEE Transactions on Instrumentation and Measurement* Bd. 70 (2021), S. 1–18

[Euro20] The European Conference of Postal and Telecommunications Administrations (CEPT) Electronic Communications Committee (ECC): ECC Recommendation (15)01, Cross-border coordination for Mobile/Fixed Communications Networks (MFCN) in the frequency bands: 694-790 MHz, 1427-1518 MHz and 3400-3800 MHz (2020)

[LSRP18] Lucas-Estañ, María del Carmen ; Sepulcre, Miguel ; Raptis, Theofanis ; Passarella, Andrea ; Conti, Mario: Emerging Trends in Hybrid Wireless Communication and Data Management for the Industry 4.0. In: *Electronics* Bd. 7 (2018), S. 400

[PRKS19] Platzgummer, Valentin ; Raida, Vaclav ; Krainz, Gerfried ; Svoboda, Philipp ; Lerch, Martin ; Rupp, Markus: UAV-Based Coverage Measurement Method for 5G. In: *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, S. 1–6

http://dx.doi.org/10.25673/111736

From Specification to Test Cases: A State-Machine-Based Approach using Image Recognition

Björn Otto,[1] Robin Gröpler,[2] Karsten Meinecke,[3] Tobias Kleinert[4]

Abstract: Software testing enables the assessment and assurance of software quality. However, writing test cases manually is time-consuming and error-prone. To avoid this, test cases can be generated automatically using Model-based Testing (MBT). MBT derives tests from a test model like a finite state machine (FSM). Such FSMs are often part of specifications, but unfortunately, often only provided as image and not in machine-readable form. Therefore, in this work we present an approach to extract machine-readable representations of FSMs from images automatically using Neural Networks. Additionally, we evaluate the applicability of our approach using a real-world specification to generate test cases from it.

The number of software-intensive systems has drastically increased over the last decades. With them, the complexity of developing and maintaining software-intense systems has also increased. However, software quality measures have to keep up with these changes, which led to a huge number of software testing methods and approaches.

One key approach is to reduce the effort of defining and implementing tests manually. This can be achieved by generating tests in an automated or at least semi-automated way. Among various other methods, Model-based testing (MBT) has been shown to be suitable for test case generation [UPL12].

In MBT, tests are derived from a model. Here, one common type of model are finite state machines (FSM). From FSMs, tests can be derived by following paths through it. This approach has many advantages. First, manual effort is drastically reduced. Second, tests are parameterizable by defining a coverage criterion (like covering all nodes or transitions in the FSM). And last, the tests are efficient, in such that multiple tests avoid testing the same aspect of the application.

The crucial part of MBT is defining the model used for test generation. This can be done manually, which comes with some effort. Fortunately, manually defining the test model is often not needed, since many specifications already include state machines as part of the

[1] Institute for Automation and Communication, Werner-Heisenberg-Straße 1, 39106 Magdeburg, Germany bjoern.otto@ifak.eu

[2] Institute for Automation and Communication, Werner-Heisenberg-Straße 1, 39106 Magdeburg, Germany robin.groepler@ifak.eu

[3] Institute for Automation and Communication, Werner-Heisenberg-Straße 1, 39106 Magdeburg, Germany karsten.meinecke@ifak.eu

[4] RWTH Aachen University, Chair of Information and Automation Systems for Process and Material Technology, Turmstraße 46, 52064 Aachen, Germany kleinert@plt.rwth-aachen.de

system or behavioral description. However, the state machines are often rendered as images and are therefore not machine-readable.

To address this issue, this work will provide a two-fold contribution: First, we will provide an algorithm to automatically extract FSMs from images in human-readable specifications. Second, we will demonstrate the usefulness of our approach by automatically generating test cases from the extracted FSMs. Connecting these steps will result in the setup shown in Figure 1.

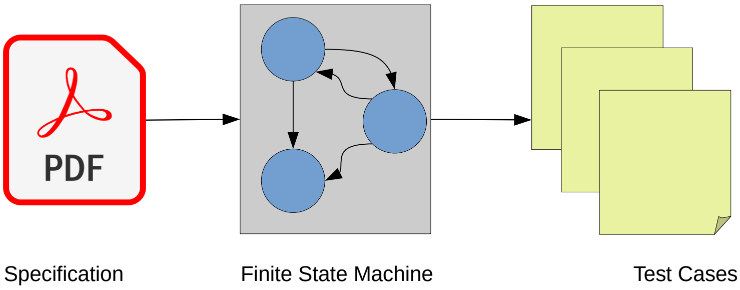


Fig. 1: Our test setup

This work is structured as follows: Section 1 lists related work, in Section 2, we describe our approach, which we evaluate in Section 3 using a fieldbus-related case study. Section 4 gives a conclusion.

1 Related Work

Our approach focuses on an end-to-end solution which yields test cases directly from a given specification by using image recognition. To best of our knowledge, this has not been studied in literature. Therefore, we focus on work related to individual phases of our approach.

1.1 Diagram Extraction

Diagram extraction has been studied extensively in literature. Most work focuses on extracting UML diagrams from images. The authors of ReSECDI[Ch22] present a method to extract class information from a rendered UML class diagram. They combine shape detection techniques (lines, rectangles) to extract the required information. They evaluate the applicability on 80 images in total.

Other work focuses on piping and instrument diagrams (P&ID) like [Yu20]. In [GZS20], the authors leverage Faster Regional Convolutional Neural Networks (Faster RCNN) to analyze P&IDs. They evaluate their approach on a commercial nuclear power plant.

The authors of [KC13] present an approach with a similar architecture to ours. Their tool Img2UML extracts information from UML class diagrams in three phases: class detection, text recognition and relationship detection. The authors validate their tool using 10 different images.

Block diagrams are analyzed in [BL22]. Here, the authors extract series of triples from rendered block diagrams. A large language model then uses these triplets to summarize the contents of the diagram.

1.2 Model-Based Testing

Model-based testing is a well-studied field in research[Di07]. Especially FSMs have been shown to be a useful type of model for test case generation[LY96]. Model-based testing is applicable in a variety of fields. However, as our case study focuses on testing a network protocol, we consider work in this field in the following, only. In [TAB17], the authors present a model-based method to test MQTT brokers. Other approaches like [PA09] test TCP/IP implementations using model-based testing.

2 Approach

Our approach is divided into two stages: First, FSMs are extracted from the specification using image processing. Second, these FSMs are leveraged to generate test cases using model-based testing.

2.1 FSM Extraction

To goal of this step is to extract all FSMs out of the specification. For this, we first extract all images out of the specification. We then apply the pipeline shown in Figure 2 to each of the extracted images.

First, we segment the image to obtain a segmentation mask as shown in Figure 3. This mask reveals where nodes and edges can be found in the input image. Given this mask, we then focus on the patches masked as nodes. We consider each continuous region a node. Additionally, we can extract the node's text using OCR. Finally, we need to check, which nodes are connected by edges. We consider two nodes as connected, if there is a continuous path of pixels labeled as edge between them.

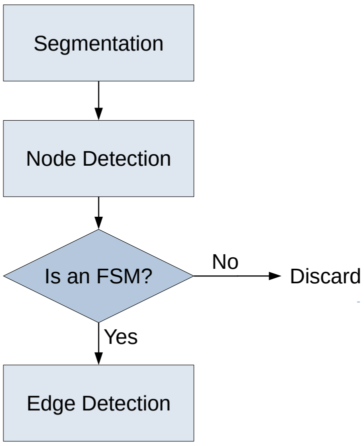
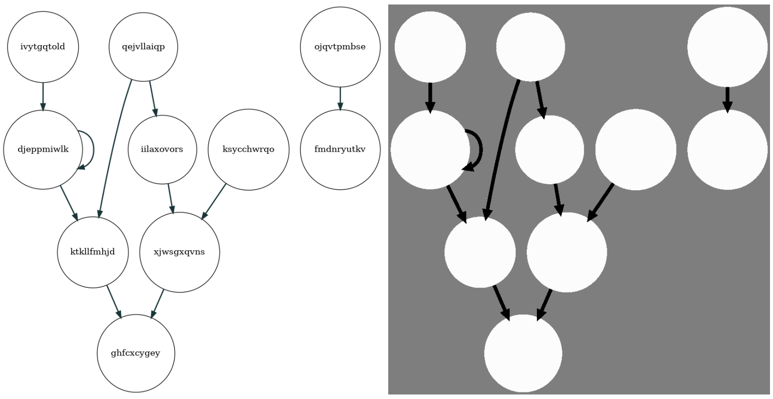


Fig. 2: FSM extraction

2.1.1 Segmentation

To goal of the segmentation step is to generate a mask image as shown in figure 3. The mask image is of the same size as the input image. For each input pixel the corresponding mask pixel has a value of 0, 1 or 2, where 0 indicates background, 1 a node and 2 an edge.



(a) Input Image

(b) Segmentation Mask

Fig. 3: Image segmentation

For the image segmentation, we use a Neural Network[Gu18]. Our network is a modified U-Net[Si21], which means it consists of an encoder and a decoder. The encoder part feeds the input image through 4 layers, each one halving the size of the image from 128x128 to 4x4 pixels. This way, the image's features are compressed to 16 values. For the encoder part, we use a pre-trained net, namely the MobileNetV2[Sa18], which is a state-of-the-art model suited for image segmentation. The decoder part does the opposite then. It consists of 4 layers doubling the size of the image, again up to a size of 64x64 pixels. The final convolution layer then yields the final segmentation mask of size 128x128 pixels. As the trained model can only be applied to image of exactly 128x128 pixels, we need to split the input image into overlapping patches of this size and merge the segmentation results afterwards.

Training our neuronal network required a reasonable amount of training data. We generated this data synthetically as follows: First we generated a random adjacency matrix. We then serialized the matrix using the dot language. Here, we also selected random colors, shapes and line widths for nodes and edges. Finally, we rendered the dot file using the graph layout engine Graphviz[El04]. Additionally, we rendered the segmentation mask. By repeating this process, we were able to generate 10.000 images for training with least effort.

For training, we used the TensorFlow[PNW20] library. As the edges naturally occupy less pixels than nodes, we weighted them 5 times more, so that they were not simply ignored during optimization. We have chosen Sparse Categorical Crossentropy as loss function and trained the model over 20 epochs. This way, we were able to achieve a (pixel-wise) accuracy of 95% on our validation set.

2.1.2 Node Detection

The goal of the node detection step is to generate a list of node masks and their enclosed text (see Figure 4). For this, we first iterate over the pixels of the segmentation mask. For each pixel masked as node, we check if it is connected directly to another node pixel. If this is not the case, the pixel is assigned a new node number. Otherwise it is assigned the node number of the connected pixel. Using a disjoint-set data structure, this can be implemented efficiently.

Along with the pixels of each node we also obtain its enclosed text using Optical Character Recognition (OCR). As specifications often contain lots of other images besides actual FSMs, we also decide which images to discard at this point. We reject all images which have less than 2 nodes. Furthermore, we reject images where the node's pixels occupy less than 20% or more than 80% of the image.

2.1.3 Edge Detection

In the last step, we decide which nodes we consider being connected by edges in the image. For this, we iterate over the pixels of the segmentation mask, again. This time, however, we number the *edge* pixels in the same way we did for the node pixels in the step before. Next, we iterate the border pixels of each node to find out, to which edges each node is connected

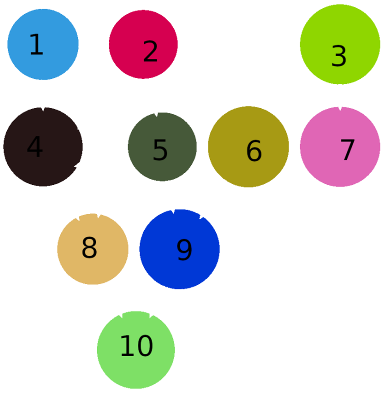


Fig. 4: Nodes mask

to. Finally, we can mark two nodes as being connected, if they are connected to at least one common edge.

To detect the direction of the edge, we use the following simple heuristic: We count the number of edge pixels adjacent to the node's pixels at each end of the edge. We then consider the side with more pixels to be the end, as it is most probably the arrow head.

2.2 Test Case Generation

For test case generation, we rely on a coverage-based approach[OK22]. For each extracted FSM, we generate a set of paths, which covers all nodes of the FSM[5]. Each path corresponds to one test case. For this, we follow the path and collect all nodes with their associated OCR-text in an array. We then pass this array to the final test execution. It interprets the node texts and executes operations accordingly. As this highly depends on the use-case, we cannot give a generalized approach for execution at this point.

For example, given the FSM in figure 5, we could obtain two test cases: `[A, B, D]` and `[A, C]`. It is then up to the text execution to map the labels `A, B, C, D` to actual test inputs and checks.

3 Case Study

To assess the feasibility of our approach, we applied it to the IEC 61158-6-10 standard. This standard describes application layer protocols for fieldbusses. For this case study, we focused on the Profinet fieldbus. As input for our approach, we used the document given in

[5] Some nodes may be covered multiple times

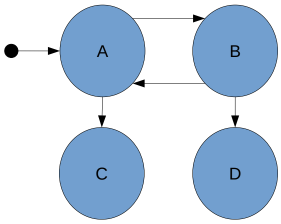


Fig. 5: Example FSM

pdf format. It consists of 1037 pages containing 274 figures in total. A few of these figures are actual state machines as the ones shown in Figure 6.

At first we converted each page to an image yielding 1037 images in total. Obviously, it would be better to consider only *figures*. However, we found these difficult to detect, because they are embedded as vector graphics within the text.

We then applied our image processing approach as described in Section 2.1. This resulted in 94 FSMs. For the test case generation and execution, we focused on the two FSMs shown in Figure 6. However, we still had to adjust the FSMs generated by our algorithm manually for two reasons: first, the extraction missed some nodes and edges, especially the intersecting ones. Second, we had to remove the loop from state K to the beginning as it would lead to only one test case iterating all the states multiple times. Finally, we used the Fences library[6] to select concrete paths through the FSMs, so that all transitions are covered. The results are shown in Table 1.

Tab. 1: Generated Test Cases

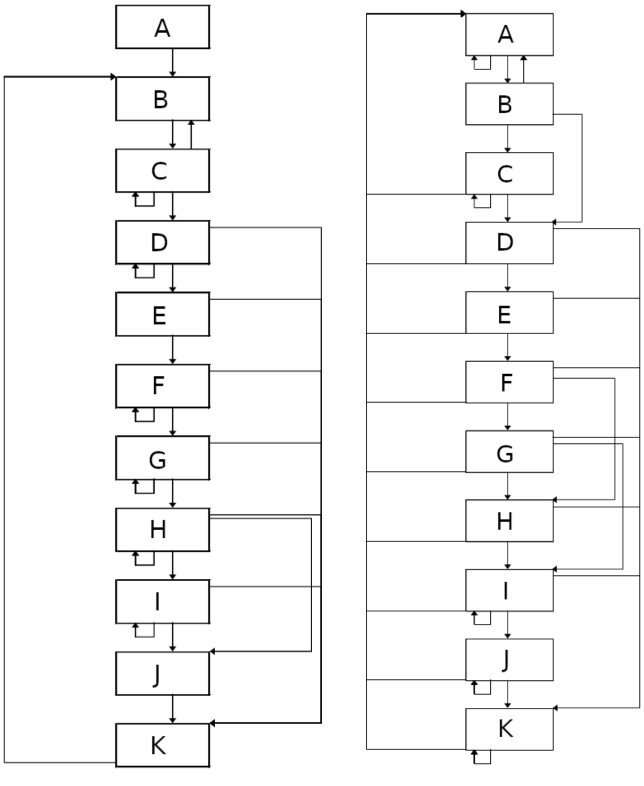
Target	States	Test Cases
Device	11	8
Controller	11	10

For the actual test execution, IEC 61158-6-10 already gives hints, how the states shall be implemented. Following this guides would result in complete test cases, which is out of the scope of this work.

4 Conclusion

In this work, we presented an end-to-end method for automated test case generation directly based on a specification. For this, we presented an approach to extract FSMs out of a specification and then use these to derive test cases. We evaluated the feasibility of our approach using the IEC 61158-6-10 standard. Evaluation showed, that our approach still

[6] https://github.com/ifak/fences



(a) Device (b) Controller

Fig. 6: State machines describing the interaction between a Profinet device and controller (from IEC 61158-6-10, node labels are simplified)

needs some manual effort. This has several minor reasons. First, the FSM extraction finds much more FSMs than we can use for actual testing. Also, the generated FSMs need some manual adjustments. And finally, the test execution still requires the states to be implemented manually.

Despite that, with the contributions made by this work, the effort of MBT can be drastically reduced since FSMs can be directly extracted out of corresponding specifications. Besides testing, our FSM extraction can be used to acquire models for design or development as well.

Future work should aim at improving the accuracy, especially for intersecting edges, i.e., FSMs with a non-planar representation.

Bibliography

[BL22] Bhushan, Shreyanshu; Lee, Minho: Block Diagram-to-Text: Understanding Block Diagram Images by Generating Natural Language Descriptors. In: Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022. pp. 153–168, 2022.

[Ch22] Chen, Fangwei; Zhang, Li; Lian, Xiaoli; Niu, Nan: Automatically recognizing the semantic elements from UML class diagram images. Journal of Systems and Software, 193:111431, 2022.

[Di07] Dias Neto, Arilo C; Subramanyan, Rajesh; Vieira, Marlon; Travassos, Guilherme H: A survey on model-based testing approaches: a systematic review. In: Proceedings of the 1st ACM international workshop on Empirical assessment of software engineering languages and technologies: held in conjunction with the 22nd IEEE/ACM International Conference on Automated Software Engineering (ASE) 2007. pp. 31–36, 2007.

[El04] Ellson, John; Gansner, Emden R; Koutsofios, Eleftherios; North, Stephen C; Woodhull, Gordon: Graphviz and dynagraph—static and dynamic graph drawing tools. Graph drawing software, pp. 127–148, 2004.

[Gu18] Gu, Jiuxiang; Wang, Zhenhua; Kuen, Jason; Ma, Lianyang; Shahroudy, Amir; Shuai, Bing; Liu, Ting; Wang, Xingxing; Wang, Gang; Cai, Jianfei et al.: Recent advances in convolutional neural networks. Pattern recognition, 77:354–377, 2018.

[GZS20] Gao, Wei; Zhao, Yunfei; Smidts, Carol: Component detection in piping and instrumentation diagrams of nuclear power plants based on neural networks. Progress in Nuclear Energy, 128:103491, 2020.

[HC20] Hagberg, Aric; Conway, Drew: Networkx: Network analysis with python. URL: https://networkx. github. io, 2020.

[KC13] Karasneh, Bilal; Chaudron, Michel RV: Extracting UML models from images. In: 2013 5th International Conference on Computer Science and Information Technology. IEEE, pp. 169–178, 2013.

[LY96] Lee, David; Yannakakis, Mihalis: Principles and methods of testing finite state machines-a survey. Proceedings of the IEEE, 84(8):1090–1123, 1996.

[OK22] Otto, Björn; Kleinert, Tobias: A Flow Graph based Approach for controlled Generation of AAS Digital Twin Instances for the Verification of Compliance Check Tools. In: IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society. IEEE, pp. 1–6, 2022.

[PA09] Paris, Javier; Arts, Thomas: Automatic testing of tcp/ip implementations using quickcheck. In: Proceedings of the 8th ACM SIGPLAN Workshop on Erlang. pp. 83–92, 2009.

[PNW20] Pang, Bo; Nijkamp, Erik; Wu, Ying Nian: Deep learning with tensorflow: A review. Journal of Educational and Behavioral Statistics, 45(2):227–248, 2020.

[Sa18] Sandler, Mark; Howard, Andrew; Zhu, Menglong; Zhmoginov, Andrey; Chen, Liang-Chieh: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520, 2018.

[Si21] Siddique, Nahian; Paheding, Sidike; Elkin, Colin P; Devabhaktuni, Vijay: U-net and its variants for medical image segmentation: A review of theory and applications. Ieee Access, 9:82031–82057, 2021.

[TAB17] Tappler, Martin; Aichernig, Bernhard K; Bloem, Roderick: Model-based testing IoT communication via active automata learning. In: 2017 IEEE International conference on software testing, verification and validation (ICST). IEEE, pp. 276–287, 2017.

[UPL12] Utting, Mark; Pretschner, Alexander; Legeard, Bruno: A taxonomy of model-based testing approaches. Software testing, verification and reliability, 22(5):297–312, 2012.

[Yu20] Yun, Dong-Yeol; Seo, Seung-Kwon; Zahid, Umer; Lee, Chul-Jin: Deep neural network for automatic image recognition of engineering diagrams. Applied sciences, 10(11):4005, 2020.

http://dx.doi.org/10.25673/111737

Investigation in Automatic Fault Detection for Scheduled Traffic and Frame Preemption in Time-Sensitive Networks

Tobias Ferfers[1], Sebastian Schriegel[1] Jürgen Jasperneite[1]

Abstract: A thorough network diagnosis is essential to cutting down the cost of network downtime in heterogeneous, time-sensitive Ethernet networks. It appears that many Time-Sensitive Networking mechanisms do not provide sufficient information about possible error sources, error recognition, or error causes. This paper examines possible symptoms and error sources of Frame Preemption and how to detect them automatically. Moreover, it examines the limitations and functionality of the Scheduled Traffic Anomaly Detection algorithm (STADA) by utilizing a test network. This research provides assistance to manufacturers of industrial automation devices, experts, and network administrators in performing FDD and root-cause analysis for Scheduled Traffic and Frame Preemption faults in Time-Sensitive networks.

Keywords: Scheduled Traffic, Time-Sensitive Networking, Fault Detection and Diagnosis, Frame Preemption

1 Motivation

Time-Sensitive Networking (TSN) for Ethernet networks introduces the possibility of Quality of Service (QoS) in Ethernet networks like, deterministic and low-latency real-time communication for control application e.g., PROFINET over TSN [Pr23]. The key mechanisms of TSN in industrial communication networks are: Time Synchronization (IEEE 802.1AS), Enhancements for Scheduled Traffic (IEEE 802.1Q) and Frame Preemption (IEEE 802.1Q). During the lifetime of a TSN device, faults may occur. Possible faults of devices, products or production plants are physical, hardware, software, aging, design fault, operating error, configuration error or production error, but also faulty network cable or rough industrial environments, temperature, humidity and many more. The additional challenge in TSN networks is the consideration of the time behavior in the network, especially in case of a fault. The previously mentioned faults can lead to (network) downtime. The cost of the downtime heavily depends on the industry branch as well as the company's size and has a large variance, according to the Ponemon Institute the average cost of network downtime in data center is about $9000 per minute [Co16].

IEEE 61158-2017 "IEEE Standard for Industrial Real-Time Communication" considers possible error sources / error symptoms, error recognition and the error handling for some

[1] Fraunhofer IOSB-INA, Campusallee 1, 32657, Lemgo {tobias.ferfers, sebastian.schriegel, jürgen.jasperneite}@iosb-ina.fraunhofer.de

components of this communication technology e.g., data link layer and physical layer [Ie17]. In comparison, most TSN standards do not provide this kind of error recognition and error handling, hence expert knowledge and experience is necessary for troubleshooting. In case of a fault, the fault detection and diagnosis (FDD) [GDC15a, GDC15b] and troubleshooting can therefore take more time in TSN networks, extend the production downtime (planned or unplanned) and increase the revenue lost.

The primary objective of FDD and root-cause analysis is to facilitate the troubleshooting process for users in the event of a fault or failure. [FSJ23] et.al. presented a concept for the automatic root-cause analysis in time-sensitive networks based on fault models. Fault models connect the symptoms of faults. to their root causes. The current state of the physical network (netload, protocol alarms, runtime measurements) is compared to network models that contain the nominal state of the network (protocols, topology, netload, schedules, configuration). An anomaly detector uses FDD technologies to detect symptoms in the physical network, then a reasoner uses fault models to find possible root causes for troubleshooting and presents the possible root causes and their probability to the network operator. To create such a system, it is necessary to investigate TSN mechanism regarding possible faults, their symptoms and root-causes. What faults in TSN key technologies can occur? How to detect faults of TSN key technologies automatically during runtime? The aim of this paper is to investigate possible faults, their symptoms and automatic detection of Frame Preemption mechanism and to evaluate the functionality and limitations of the Scheduled Traffic Anomaly Detection Algorithm (STADA) [FSJ23].

The first section describes State of the Art Diagnosis in industrial communication. The second section describes the functionality of Frame Preemption and Scheduled Traffic. The third section investigates Frame Preemption mechanism and describes possible faults and how to detect them. The fourth section handles the evaluation of STADA including a description of the algorithm, the test setup and method as well as the results. The final section is conclusion and future work. This work will support vendors of industrial automation devices, experts and administrators of TSN networks during FDD and root-cause analysis for Scheduled Traffic and Frame Preemption faults.

2 State of the Art Diagnosis

The first section of the chapter, highlights the most important terms as well as general FDD techniques. In the second part of the chapter, three examples of diagnosis in industrial communication are explained in greater detail. In the 1990s, Isermann et al. defined terms in the field of Fault Detection and diagnosis (FDD), e.g., faults, fault diagnosis, fault management, and more [IB96]. *Fault detection* is the determination of the fault's presence in a system and the time of detection. *Fault isolation* is the determination of the kind, location, and time of the detection of a fault. *Fault identification* describes the determination of the size and time-variant behavior of a fault. After a fault diagnosis, the location, size, and type of fault are known, but the root cause and the actions to be taken for rectification are unknown. Sometimes FDD and anomaly detection are used as synonyms; in principle, it is about being able to detect deviations from the normal state. The field of fault detection and diagnosis (FDD) is divided into four categories: signal-based, model-based, data driven and hybrid methods [GCD15a, GCD15b]. Signal-based methods rest upon signals that are somehow connected to the fault in time, frequency, or time-frequency domain and utilize, for example statistical information, e.g., mean value, variance, or kurtosis, for example [EM13]. Model-based approaches consider an exact (mathematical) representation of the system or process, commonly applied to physical processes, for example applied on LAN in [Fo02]. The data-driven methods are divided into two methods: statistical analysis and artificial intelligence, e.g. [An18]. Data-driven techniques use available information about the devices or network, and are often considered as an alternative to model-based approaches because there is no detailed modeling necessary.

The "IEEE61158 Standard for Industrial Hard Real-Time Communication" provides additional information and even recommendations regarding error management at both the data link layer and physical layer [Ie17]. The standard covers possible error sources, error recognition, error handling, and error registration. Loss of link, buffer overflows or underruns, timing violations for received frames, transmission errors, collisions, frame loss, incorrect physical Ethernet operating mode, and numerous other issues that are addressed in the standard. In the following, the “Loss of PollResponses” will be explained in more detail. This error indicates that no PollRepsonse frame was received in the current time slot. The standard describes multiple categories of possible error sources from physical errors e.g., loss of link Rx buffer overflow. Other possible sources of errors are, for example, defective components in the network or the use of devices whose latency does not meet the requirements. This fault is supposed to noticed in the management node cycle state machine and can be recognized if the sot timer expires and no frame was received in the slot. When a frame loss is detected, several actions are taken: Notification of other devices or components, exclusion from isochronous communication, or error logging.

PROFIENT IRT is a communication profile of PROFINET that outlines a highly synchronized (isochronous) and stringent communication protocol, meticulously engineered from the topology to the cable delays [Pr22]. Since, PROFINET IRT is highly engineered, the protocol implements mechanisms to detect changes in the topology or cable length during runtime. One part of error handling is done at the application level with the “SignOfLife” application, which has a counter that increases every cycle, and the mechanism checks through this counter if frames were lost in a cycle. A threshold is set by the user for the number of frames that are acceptable to lose. Normal errors are handled as in the non-isochronous PROFINET protocol: if a module or submodule detects a fault, an alarm is sent to the upper layer, e.g., if data processing is not finished when the next cycle starts. Since, PROFINET IRT depends very much on synchronization, details about sync errors are also given in the specification. More precisely, it describes how sync errors should be handled on the provider and consumer sides. For example, if an out-of-sync error occurs, access to data is refused, or it describes error codes, e.g., jitter out of bounds or no sync telegram within rules received. PROFINET IRT diagnoses the communication at the application level and has an alarm system for notification, but in-depth diagnosis and providing the user with the root cause are not included.

“IEC/IEEE 60802 TSN Profile for Industrial Automation” describes a set of rules for time-sensitive networking in the industrial automation field [Ie23]. For diagnosis, the profile suggests observing the YANG data model representation in the local database of the component and observing the available objects. Furthermore, the profile defines a subscriber-based notification mechanism and corresponding events, e.g., loss of link, loss of sync or periodic statistics. As with other protocols, there are mechanisms to detect certain errors, but possible causes or actions to clarify them do not exist.

3 TSN-Mechanisms Scheduled Traffic and Frame Preemption

3.1 Scheduled Traffic

The Enhancements for Scheduled Traffic (IEEE 802.1Qbv) allows the transmission of each transmission queue to be scheduled to a relative time. Transmission gates are associated with each queue [Ie16a]. The state of the gate determines whether frames can be selected for transmission (open or closed). Every port has a gate control list with ordered gate operations, and for each entry in the gate control list, there is a traffic class assigned. Depending on whether frame preemption is used or not, the gate operation of each entry allows preemption of frames. Scheduled traffic leads to a slot-based communication where one or more traffic classes are assigned to the slots (see Fig. 1). The most important parameters for scheduled traffic are: base-time (start time of the schedule), cycle time, ControlList, ControlListLength and the CycleTimeExtension.

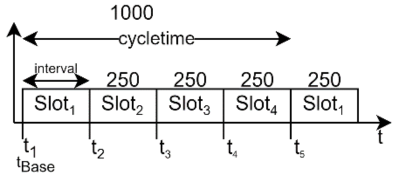
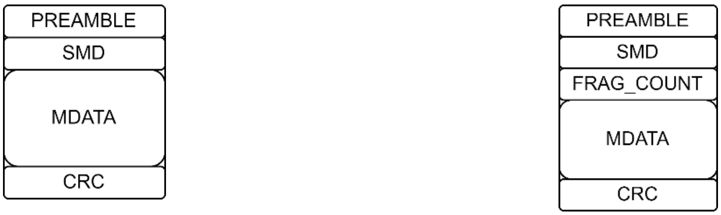


Fig. 1: Scheduled Traffic 802.1Qbv

3.2 Frame Preemption

IEEE 802.1Qbu – Enhancements for frame Preemption and IEEE 802.1br are a set of features that allows higher priority frames to interrupt the transmission of lower priority frames and resume it later [Ie16b]. Frame preemption is implemented at data link layer according to the ISO/OSI model. The MAC layer provides two services: the preemptable

MAC (pMAC) and the express MAC (eMAC). A MAC merge layer merges these two MAC services back together and preempts preemptable traffic currently being transmitted or prevents the start of the transmission of preemptable traffic. When the preemption capability is inactive, the MAC Merge sublayer does not allow express traffic to interrupt a frame provided by the pMAC service interface. In the MAC Merge sublayer, a special packet format is used called mPacket. (see Fig. 2).



a) format of express packet, complete preemptable packet or an initial fragment of a packet

b) format of a fragment of a packet

Fig. 2 : mPacket format MAC merge sublayer

The Preamble of an mPacket is identical to a MAC Preamble, the Start Frame Delimiter (SFD) is replaced by the Start mPacket Delimiter (SMD) value and identifies the type of mPacket frame e.g., verify, respond, express packet, preemptable packet start (SMD-S0 to SMD-S3), or a continuation fragment (SMD-C0 to SMD-C3). The frag_count in a fragment is a modulo-4 counter that increments each continuous fragment of a preempted mPacket. The frag_count is used to detect mPacket reassembly errors by enabling detection of the loss of up to three packets. As seen in Fig. 2 the frag_count is only included at a continuous fragment of a mPacket. The CRC field contains a cyclic redundancy check (CRC) and an indication of whether this is the final mPacket of a frame. In the final mPacket of a frame the CRC field contains the CRC of the MAC frame. For other mPackets the CRC field contains an mCRC (mPacket CRC) value for that specific mPacket. Generally, the preemption capability is enabled on the transmit direction only if it’s ensured that the link partner also supports the frame preemption capability. The process of discovering the support on the link partner relies on the exchange of additional Ethernet capability TLV in the LLDP frame. The mechanism is only enabled if the support was announced before and the preemption mechanism is disabled in case of a link failure. Only if the frame preemption functionality has been made known beforehand the verification process will be triggered. In this process a verify mPacket is sent and a respond packet is expected from the link partner. If the frame preemption capability is enabled but has not been verified yet, the MAC merge sublayer indicates verification process. Verification can be disabled, this is useful for engineered networks.

4 Possible Faults and Symptoms Frame Preemption

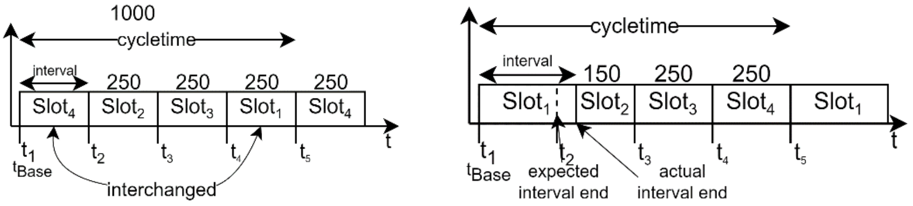
In order to diagnose the previously outlined Frame Preemption mechanism, it is imperative to distinguish between two distinct stages: initialization during the ongoing verification process (static) and diagnosis subsequent to successful verification at runtime (dynamic). Static errors describe the fact that the verification process was not successful, which can have several causes. This phenomenon may manifest itself in the absence of LLDP frames or the absence of the additional Ethernet availability for frame preemption in the frame. Possible root causes are an incorrect implementation or configuration of the device or an increased network load could that leads to frame loss of LLD frames or verification mPackets. Therefore, devices or additional measuring equipment could check if the verification was successful, e.g., during startup or after a link failure. The second possible error category pertains to dynamic errors, which may arise in the event that the frame preemption verification was successful and the mechanism is operational and functioning. The standard already provides capabilities to indicate faults, like various counters and status variables to check if the mechanism is working correctly e.g., aMACMergeAssErrorCount, count of MAC frame reassembly errors on receiver side or the MACMergeFrameSmdErrorCount is a counter of the received MAC frames / frame fragments rejected due to unknown SMD value or arriving of SMD-C when no frame in progress. For detailed diagnosis access to these counters and the current device configuration is necessary. A faulty implementation could also be the reason for incorrectly sent or re-assembled fragments, which could be noticed by the link partner (missing frames etc.). Further on, it is possible to observe the jitter of the real-time network traffic, if frame preemption should be configured but is not and, enough best effort traffic is going through the network this faulty configuration could be noticed.

Some of the evaluation of the functionality can be done by observing counters or variables already defined in the frame preemption standard, especially for the runtime errors. Additional measuring equipment or an extension of the devices is necessary to detect faults in the verification process or the jitter of real-time traffic. Furthermore, access through an API must be granted for network administrators or central diagnosis systems to evaluate the quality of the network. This must be integrated into the driver or firmware of the devices to provide this information, if it is not already the case through Management Information Base (MIB).

5 Evaluation of STADA

5.1 Description STADA

The Scheduled Traffic Anomaly Detection Algorithm (STADA), presented in [Fe23], aims to validate the correct scheduled traffic configuration based on the transmit timestamp (tx_timestamp) of a frame at runtime, the desired scheduled traffic configuration (base time of the schedule, traffic class for each slot). Based on the transmit timestamp the exceeded time in the current cycle can be calculated: *time_elapsed = (tx_timestamp – base_time) % cycle_time*. With the desired configuration and the elapsed time in the current cycle the current slot of the schedule can be determind and which frames are allowed. Then the algorithm compares whether the allowed and actual frame type match. If the frame types do not match, further comparison is done to determine if the frame type is even configured or if the interval is too short or too long or a wrong order of time slots is given. But how does this algorithm performs in a test setup and network?



(a) faulty slot order

(b) faulty interval length

Figure 1: Possible faults scheduled traffic

5.2 Test setup

The test setup for the evaluation of STADA consists of a TSN controller, TSN switch, TSN device and additional measuring equipment. To analyze the traffic on the wire a network Test access point (TAP) is set between the TSN controller and the TSN switch. The TSN controller, TSN device and the measuring device is Linux based with Intel I225 network cards. Traffic on the controller is generated by a dummy application. Generally, it is possible to implement the STADA in the device driver or firmware of the network component, except of the API to the (Linux) kernel this is highly vendor specific and closed source. Therefore, STADA was implemented on additional measuring equipment and the transmit timestamp is calculated as *transmit_timestamp = receive_timestamp – delay*. The delay was determined by the delay of the TAP and previous measurements. The measuring equipment is connected to the TSN switch for synchronization and to the network TAP for diagnosis. The schedule for evaluation has three slots: network management (PTP and LLDP), second slot real-time traffic (PROFINET) and the third slot best effort (IPv4) as seen in Tab. *1*.

Slot number	**Length [µs]**	**Traffic type**
1	200	PTP, LLDP
2	250	PROFINET
3	550	Best effort (IPv4)

Tab. 1: nominal scheduled traffic configuration

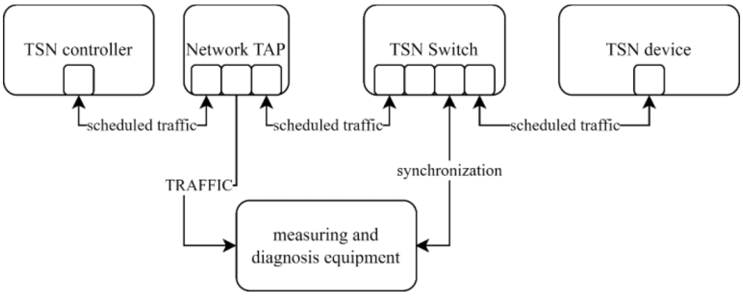


Fig. 3: Test setup for STADA

5.3 Method

As described in previous work [Fe23], possible faults for scheduled traffic are a wrong order of slots and an incorrect interval length (too short or too long). For the determination of the possible combinations for testing, it was specified that no repetitions may occur in the order. With the test schedule from the test setup this results in the following possible combination shown in Tab. 2. A short interval length is defined as half of the normal interval length, the long interval length is defined as double of the normal slot length.

Slot order	Slot 1 interval	Slot 2 interval	Slot interval 3
123	Short	Short	short
132	Normal	Normal	Normal
213	Long	Long	Long
231			
312			
321			

Tab. 2: scheduling configuration

The possible combinations based on Tab. 2 are 6x3x3x3 = 162 combinations. In the literature there are methods about reducing the test cases for example [Ho13]. In order to reduce the number of test cases, the approach of pairwise testing was selected. The idea behind pairwise testing is that some combinations of the parameters are responsible for faults, and most of the time the combination of two parameters. The construction of the possible test combinations is done via orthogonal arrays, as described in [Ho13]. After the reduction of test cases, the matrices show don’t care entries, it was defined that every time an entry consists a don’t care the normal slot interval length was chosen. With pairwise testing the combinations could be reduced to 33 test cases. The test sequence for all 33 test cases was as follows: Start synchronization, set the schedule, wait until synchronization is complete, start STADA for five minutes, procced with next test case. For all frames, the transmit time was determined based on the cycle time and logged.

5.4 Result

In all 33 executed test cases, the algorithm detected a faulty scheduled traffic configuration. However, the detailed statement (wrong order, interval length too short or too long) was not clear in most cases. The algorithm principle is responsible for this imprecise recognition. The algorithm uses the transmission time and only knows about the traffic class of the previous frame and the desired traffic class. Tab. 3 you can see some summarized measured values, e.g., mean, minimal, maximum value, from the two test cases. The first test case is the desired configuration, the second test case is the correct order but all intervals are too short. This shows that the mean value in the misconfigured case deviates significantly from the normal case, as does the standard deviation of the network management. These can be also parameters, which one puts into an extension of the algorithm. In the future, it may be possible to obtain a more precise diagnosis by closely scrutinizing multiple cycles, for instance.

Test case	**traffic type**	**Count [µs]**	**Mean [µs]**	**Std [µs]**	**Min [µs]**	**Max [µs]**
TC 1 desired	NM	6494	29	45	0,5	199
	Real-Time	326491	424	0,48	419	438
TC 4 short,123	NM	6646	269	251	0,63	599
	Real-Time	334101	601	0,11	600	604

Tab. 3 : Example Measurements results STADA evaluation

6 Conclusion

In this work possible errors and methods to detect frame preemption faults were investigated. Furthermore, the Scheduled Traffic Anomaly Detection Algorithm (STADA) was evaluated in a test setup. An example schedule with three time slots was chosen and the possible combinations determined. With pairwise testing the test cases could be reduced from 162 to 33. These 33 test cases were executed in the test setup. It could be shown that STADA generally detects a faulty configuration of the scheduled traffic. As of today, the algorithm cannot give a detailed analysis of what exactly is wrong with the schedule. Future work should concentrate on improvement of STADA with the mentioned ideas and the evaluation of the Frame Preemption fault detection ideas.

Literature

[Pr23] PROFINET over TSN, https://www.profibus.com/technology/industrie-40/profinet-over-tsn, 20.10.2023

[Co16] Cost of Data Center Outages, https://www.vertiv.com/globalassets/documents/reports/2016-cost-of-data-center-outages-11-11_51190_1.pdf, 2016

[Ie17] IEEE61158: IEEE Std 61158-2017 (Adoption of EPSG DS 301). IEEE Standard for Industrial Hard Real-Time Communication. IEEE, S.l., 2017.

[GCD15a] Gao, Z.; Cecati, C.; Ding, S. X.: A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. IEEE Transactions on Industrial Electronics 6/62, pp. 3757–3767, 2015.

[GCD15b] Gao, Z.; Cecati, C.; Ding, S.: A Survey of Fault Diagnosis and Fault-Tolerant Techniques Part II: Fault Diagnosis with Knowledge-Based and Hybrid/Active Approaches. IEEE Transactions on Industrial Electronics, 2015.

[FSJ23] Ferfers, Tobias; S Schriegel, Sebastian; Jasperneite, Juergen: Automated Root Cause Analysis in Time-Sensitive Networks based on Fault Models, International IEEE Symposium on Precision Clock Synchronization for Measurement, Control and Communication ISPCS 2023, London, United Kingdom, 2023

[IB96] Isermann, Rolf; Ballé, Peter: Trends in the Application of Model Based Fault Detection and Diagnosis of Technical Processes, IFAC Proceedings Volumes, Volume 29, Issue 1, Pages 6325-6336, 1996

[EM13] Estima, J. O.; Marques Cardoso, A.J.; A New Algorithm for Real-Time Multiple Open-Circuit Fault Diagnosis in Voltage-Fed PWM Motor Drives by the Reference Current Errors, In IEEE Transactions on Industrial Electronics, vol. 60, no. 8, pp. 3496-3505, Aug. 2013

[Fo02] Fontanini, S. T.; Wainer J.; Bernal V.; Maragon S.: Model based diagnosis in LANs, IEEE Workshop on IP Operations and Management, Dallas, TX, USA, 2002, pp. 121-125

[An18] Anusasamornkul, Tanapat: A Network Root Cause Analysis and Repair System, 2018 6th International Symposium on Computational and Business Intelligence (ISCBI), 2018

[Pr22] Profibus Nutzerorganisation e.V.: Isochronous Mode – Guideline for PROFNET IO, Version 1.3, 2022

[Ie23] IEC/IEEE 60802 TSN Profile for Industrial Automation, https://1.ieee802.org/tsn/iec-ieee-60802/, draft 2.1, 20.10.2023

[Ie16a] IEEE Standard for Local and metropolitan area networks -- Bridges and Bridged Networks - Amendment 25: Enhancements for Scheduled Traffic, in IEEE Std 802.1Qbv-2015 , no., pp.1-57, 18 March 2016

[Ie16b] IEEE Standard for Local and metropolitan area networks -- Bridges and Bridged Networks -- Amendment 26: Frame Preemption, in IEEE Std 802.1Qbu-2016 (Amendment to IEEE Std 802.1Q-2014) , vol., no., pp.1-52, 30 Aug. 2016

[Ho13] Hoffmann, W. Dirk: Software-Qualität, 2. Auflage, Springer Vieweg, 2013

http://dx.doi.org/10.25673/111741

5G-Based Localization in Industrial Environments

A survey on challenges of localization via 5G in industrial scenarios

Bjarne Frischkorn [1], Michael Knitter[2], Wolfgang Endemann[3], Rüdiger Kays[4]

Abstract: This paper focuses on challenges occurring when using the 5G NR standard as a real-world application for precise localization in industrial environments. The different aspects of a mobile network based localization approach are discussed. First an overview on mobile network setup is given. Based on a mobile network emulation a first localization is conducted in an indoor laboratory. Afterwards the influence of indoor channel properties and the arising problems are discussed. With the results from this discussion, a new system model is introduced to improve the localization accuracy down to one meter.

Keywords: 5G, Survey, Localization, Indoor, Rising Edge

1 Introduction

The use of autonomous industrial vehicles in companies not only offers opportunities for increasing productivity, but also the risk of accidents. For this reason, strict regulations are in place, e.g. by specifying low speeds or special right-of-way rules. For an increased productivity and safety, all internal road users must be able to be informed of the current position of other participants in order to adjust their trajectory and thus avoid accidents.

Goal of the 5G SAIFE project is to establish a 5G based real time positioning solution for factory traffic participants. At the same time, the capabilities of the 5G standard shall be used for a low latency, high precision positioning. Key enhancements for a more precise positioning are defined in 3GPP 5G standard release 16 and further enhanced in release 18, whereas practical guidelines and implementation as missing at the current point in time.

[1] Communications Technology Institute (CTI), TU Dortmund, Otto-Hahn-Straße 4, 44227 Dortmund, bjarne.frischkorn@tu-dortmund.de

[2] Communications Technology Institute (CTI), TU Dortmund, Otto-Hahn-Straße 4, 44227 Dortmund, michael.knitter@tu-dortmund.de

[3] Communications Technology Institute (CTI), TU Dortmund, Otto-Hahn-Straße 4, 44227 Dortmund, wolfgang.endemann@tu-dortmund.de

[4] Communications Technology Institute (CTI), TU Dortmund, Otto-Hahn-Straße 4, 44227 Dortmund, ruediger.kays@tu-dortmund.de

The roadmap of 5G provided by the 3GPP states that release 18 will be finished by end of 2023 respective start of 2024 [3G23].

This paper mainly presents evaluation and research gathered during the 5G SAIFE project.

Availability of feasible hardware solutions for test and measurement showed up as a key challenge. For FR2, with higher bandwidths, there does not exist “off the shelf” hardware for campus networks. In FR1 some proprietary hardware exists. The proprietary hardware only supports software releases up to release 15. As localization is introduced in release 16, the existing equipment does not support localization. In this paper a solution to emulate 5G localization is presented. Afterwards, localization measurements are conduced and evaluated. These results are refined by theoretical considerations based on a new system model.

The following section gives a brief overview on related work.

2 Related Work

Wireless localization is a widespread topic of discussion. [Tr16] shows the different approaches of wireless positioning in mobile communications. These approaches are classified in [Ga20] into signal strength based, time based and angular based procedures.

The release 16 standard introduces improved localization to 5G NR mobile communication networks. In this release, the positioning reference signal (PRS) is introduced for usage of the three classes of localization approaches [Dw21].

As shown in [Hu22], bandwidth is the most important factor for high accuracy localization. The 5G standard defines bandwidths of up to 100 MHz in frequency range 1 (FR1) and 2 GHz in frequency range 2 (FR2), enabling improved localization compared to previous mobile communication standards. With increased bandwidths, time based localization approaches have advantages, compared to power based approaches which do not depend on the bandwidth but suffer from low accuracy due to randomness of small-scale fading.

[Tr21] shows angular based approaches for FR2. A lack of available hardware does not allow for an implementation in practice. Resulting from the lack of hardware for angular based approaches and insufficient localization from power based approaches, time based approaches are the most used and promising techniques for 5G indoor localization [Pa22].

For time based approaches there are three main techniques: Round-Trip-Time (RTT), Time of Arrival (ToA) and Time Difference of Arrival (TDoA).

Localization accuracies vary depending on the used algorithms and scenarios. In [Ha23] the authors simulate a 5G FR2 scenario where an accuracy of 50 cm for 90 % of all cases is achieved. A more than generous bandwidth of 1200 MHz was used. At time of writing and to the author's knowledge, no systems for use in industrial environments exist that make use of FR2 of 5G NR.

Measurements with equipment for FR1 localization has been discussed in literature. The authors of [Pa23] use a TDOA approach. To increase the accuracy of the measurements the results are oversampled with a factor of 16 resulting in an accuracy of 4 m without averaging and calibration. Taking averaging into account and by properly calibrating the devices, 90 % of all measurements give an accuracy of better than 3 m. The main source of errors is a lack of synchronization.

In [Ru22a] a power based approach was used to determine the accuracy of localization in FR1. To increase precision, a neural network was used. This method uses a technique called fingerprinting where the receiver stores channel state information (CSI) and received signal strength indicators (RSSI) before the active localization in a database. During active localization the so called iPos-5G algorithm uses an AI to compare actual CSI and RSSI information to the database. The resulting accuracy dependents on the scenario. In an office environment the authors achieve an accuracy of 2.39 m, and 3.26 m in a corridor. This technique requires extensive measurements before active localization. It is questionable, how long the fingerprinting will stay valid in a non-static environment. In a follow-up publication, the authors improved the accuracy to 2.35 m in 90 % of all cases [Ru22b]. Combining fingerprinting and angular based approaches in a neural network is shown in [Zh21] with a mean accuracy of 50 cm in simulations. A sensor fusion approach combining LIDAR data and signal strength data in a simulation by [Mu21] results in a high error of 6.55 m.

In [Ga17] the authors present an excellent discussion on using super-resolution algorithms like MUSIC or ESPRIT for TOA and TDOA localization. While these approaches are more tolerant regarding noise, the increased needs for computational power makes these algorithms undesirable compared to the approach shown later in this paper.

Our previous work evaluated channel impulse responses (CIR) in industrial environments [Kn22] with respect to ToA measurement. The paper shows that the CIR does not consist of a single line-of-sight (LOS) impulse but rather of a LOS impulse and multiple echoes which arrive shortly after the initial LOS component. Echoes result from multiple reflections in indoor propagation scenarios. Using high bandwidth systems, it is possible to identify each echo and remove it afterwards. For low bandwidth systems the superimposed echoes deform the LOS path impulse. For a 5G system this will be discussed in section 5 and following sections.

The following section covers the lack of available commercial equipment or software for setting up a 5G NR localization network and test environment.

3 Getting Started

The first approach to 5G localization in industrial environments is to use already existing networks from public mobile communication providers. In 5G exists the option that parts of a network can be dedicated to a special service. This method is called slicing. However, current mobile operators do not offer network slicing. Moreover, mobile operators do not offer localization even as a service on their own. Thus a campus network needs to be could be implemented at industrial sites, though still lacking any location service.

As mentioned in section 1, there is no hardware available to conduct localization in self operated campus networks. Based on software defined radios (SDR), Open Source software solutions such as O-RAN and SRS-RAN offer an alternative to proprietary hardware. While it is possible to set up a running 5G base station (gNB) using SRS-RAN, the software is also lacking the location management function (LMF), which is needed for localization.

Since off the shelf solutions do not exist, gNBs and UE have to be emulated to be able to conduct experimental research. The emulation has to match the requirements provided by the standard. Therefore, vector signal generators (VSG) are used as gNBs. Each VSG is fitted with an individual 5G test signal. Table 1 shows parameters for the 5G test signals.

Parameter	**Value**
Transmission Frequency	3.75 GHz
Bandwidth	100 MHz
Subcarrier Spacing	30 kHz
Number of Subcarriers	3276
Frame Length	10 ms
Number of Slots	20 Slots
OFDM Symbols per Slot	14 Symbols per Slot

Tab. 1: Parameters of the 5G test signal

A 10 ms frame consisting of 20 slots is used for transmissions. In the first transmitted slot the synchronization signal block (SSB) is transmitted to synchronize UE with gNBs. On the four following slots, each of the four different gNBs transmit their channel state informations (CSI) and PRS. The first OFDM symbol is used for the CSI, the second is left out and the remaining 12 symbols are used for the PRS. Each gNB has its own distinctive PRS signal depending on the assigned cell ID. After the five initial slots, the whole process gets repeated three times to match the frame length of 20 ms.

Meanwhile the UE is modeled by a portable handheld spectrum analyzer. The UE samples the whole received signal which is later post processed in MATLAB.

As a localization technique, a time based approach is used. The chosen methods are ToA and TDoA to determine 2D coordinates. With this approach, only three gNBs are needed for localization. However, using more gNBs gives the opportunity to use different base station combinations for the case that one measurement is off due to interference. For every gNB the time of flight needs to be determined. This is conducted by correlating the PRS which is unique for each gNB, to the received overall signal. From the obtained time delays, the slot offset needs to be subtracted to normalize all results.

In the following section the results and challenges of the experiment are discussed.

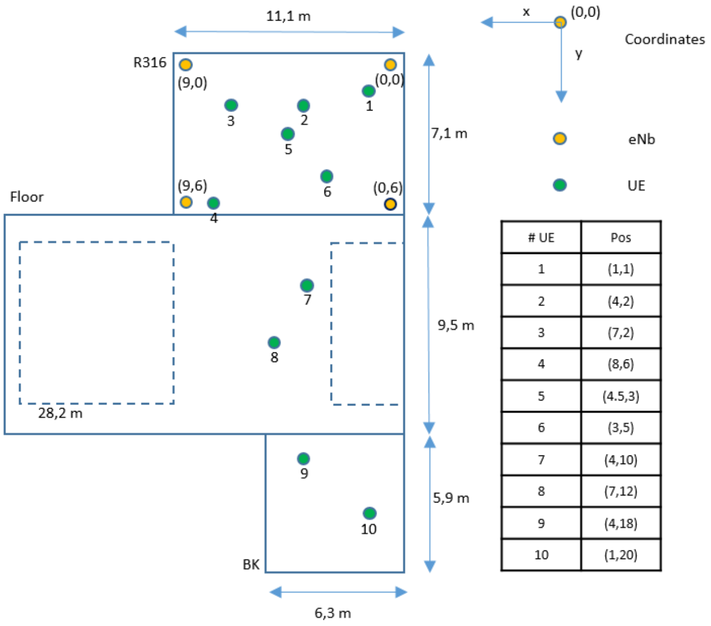
4 Emulation of 5G Localization using 5G Test Signals

For the emulation of 5G localization, four base stations are used. The site consists of the main room, in which the gNBs are placed and two adjacent rooms, to emulate shadowed LOS scenarios. The UE is moved across the site to the positions 1-10. Locations of the gNBs and the UE can be taken from fig 2.

As shown in section 2, the synchronization is the most important source of errors. To reduce the impact of synchronization errors, the VSGs are manually synchronized.

The measurement starts by an impulse issued by a signal generator, which also starts the transmission by the gNBs. Furthermore, to evaluate drift of the generators, position 5 is placed at the exact center of the four gNBs. Every occurring synchronization mismatch can be corrected during the post processing by correcting the time of flights, as these have to be equal.

The post processing in MATLAB correlates the received signal with each of the original PRS signals. To improve accuracy, the literature shown in section 2 recommends to use oversampling. In this experiment, an oversampling factor of 10 is used. Then the maximum of the oversampled correlation result is selected and converted into a time of flight value.



# UE	Pos
1	(1,1)
2	(4,2)
3	(7,2)
4	(8,6)
5	(4.5,3)
6	(3,5)
7	(4,10)
8	(7,12)
9	(4,18)
10	(1,20)

Fig 1: Layout of the scenario with marked positions of base stations and UE

Figure 3 shows the exemplarily correlation result for position 6.

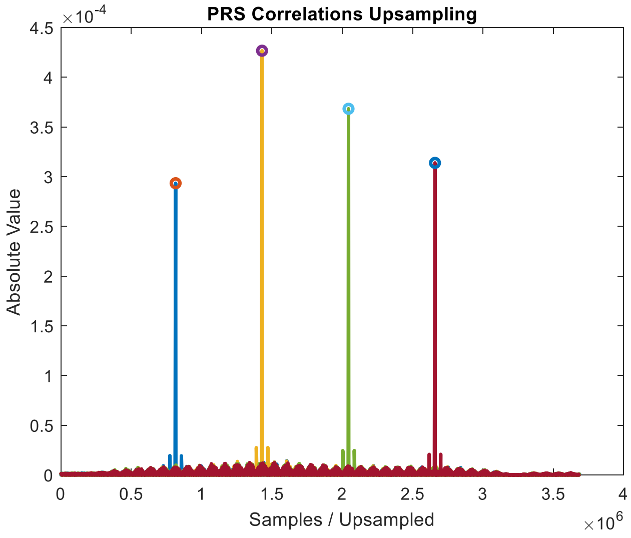


Fig 2: Correlation result of position 6

When looking at the peaks of the correlation, the influence of echoes can be examined. For individual UE and gNB position combinations, there are multiple peaks as shown in fig 4. A maximum search would lock onto the higher peak, even if it logically does not make sense. The LOS component of the signal will always arrive first at the receiver. Therefore, the first peak with a significant amplitude can be interpreted as the time of flight.

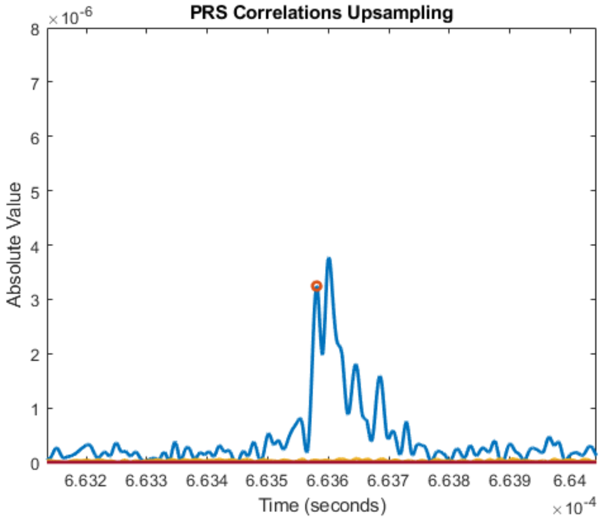


Fig 3: The improved maximum algorithm locks onto the first significant maximum and not on the global maximum

The derived positions for all nine positions, including time of flight correction and improved peak detection, are depicted in table 2

Real Position (X,Y)	**Calculated Position (X,Y)**	**Deviation In m**
POS1 (1,1)	(1.4, -1.65)	0.76 m
POS2 (4,2)	(4.37,1.44)	0.67 m
POS3 (7,2)	(6.98, 0.76)	1.24 m
POS4 (8,6)	(15.78, 14.43)	11.47 m
POS5 (4.5,3)	(4.5,3)	0 m
POS6 (3,5)	(1.46, 4.83)	1.55 m
POS7 (4,10)	(2.49, 5.65)	4.60 m
POS8 (7,12)	(4.87, 5.21)	7.11 m
POS9 (4,18)	Not measured	Not measured
POS10 (2,20)	(0.96, 4.97)	15.07 m

Tab. 2: Calculated positions and the deviation to the real positions

Despite the made corrections, an exact localization is not possible for every combination of gNBs. When observing the correlation results, the influence of echoes on the main LOS peak is obvious. The following section focuses on modeling the influence caused by echoes.

5 System Model for a new Approach

An industrial environment suffers from multiple echoes which arrive close to the LOS path. To model the LOS and echoes mathematically, dirac impulses are used. Through the band limitations, impulses get deformed into SINC impulses. Understanding the influence bandlimited impulses have on each other is key in improving the accuracy of localization in wireless communication networks.

In a first investigation, one SINC impulse with an amplitude of 1 is superimposed with a second SINC impulse of same amplitude. When there is no time difference between these two SINC impulses, the sum results in one SINC impulse with doubled amplitude. Now, a time offset is applied to the second impulse. For small time offsets the resulting signal corresponds to a slightly time shifted SINC impulse with an amplitude greater than1. The original SINC impulse could not be extracted from the superimposed signal. For large time offsets, the superimposed signal has two distinctive peaks at exact the positions of the peaks of the single SINC impulses. The scenario shows that for sufficient big time delays a maximum peak detection can detect the correct time of flight of the LOS. If the time difference between LOS and one echo is smaller, the maximum of the signal is shifted to a later point of time. Thus, distorting the calculated time of flight.

The model, consisting of a LOS path and one echo, is extended to fit a more real scenario by adding multiple echoes. The channel impulse response becomes:

$$h_d(t) = \sum_{k=0}^{E} a_k SINC\big(B(t - E_k T)\big)$$

Where $a_k$is a complex scaling factor, B is the bandwidth of the signal, $T$is the sampling Time and $E_k$is the time delay for an echo k. In the following section, an improved localization approach is presented based on this system model.

6 Approach to improve 5G Localization Accuracy

The system model presented in section 5 consists only of scaled SINC impulses. As already shown, the peak of the LOS path is deformed by following echo impulses. The signal is now divided into a main peak area, rising edge area and a pre-oscillation area. If the following echo impulses are close to the LOS impulse, the maximum of the function (the main peak) is shifted to a later point of time. For a real system, the maximum cannot be shifted to an earlier point of time. However, for some LOS-echo combinations, the maximum shows up before the time of the LOS impulse peak. These special cases are based on seldom phase combinations of LOS and echo paths.

The question arrives which area of the superimposed signal should be evaluated to have the best match to the LOS path only?

Observing the pre-oscillation area makes no sense as this area is mostly influenced by noise due to very low amplitude of the SINC impulse in that area. The rising edge is left as an interesting area to observe. The rising edge of the whole signal is dominated by the first received impulse. As stated before, the first received impulse in a LOS scenario is always the LOS impulse. Only the maximum is shifted by subsequent impulses.

7 Simulation of Approach

When simulating a channel model with different numbers of echoes, it gets clear that the resulting superimposed signal will either have a distinct maximum, when the echoes arrive significantly later than the LOS path, or a distorted maximum due to short time offsets between echoes and LOS. As seen in previous work and section 4, the latter case is mostly present in industrial environments. This scenario is now simulated by using the system model provided in section 5. In 200 different realizations, 25 superposed impulses are used to create channel impulse responses. The LOS impulse has a normalized amplitude of one. Fig. 5 presents a selection of 10 out of the 200 realizations.

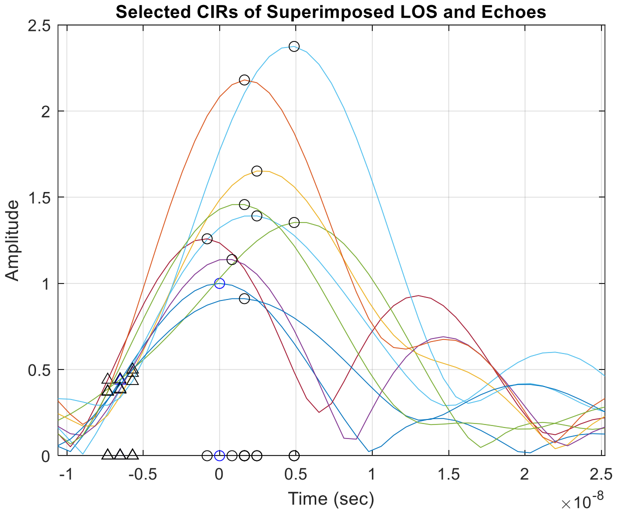


Fig 4: Multiple CIR of different realizations. The spread of the rising edge is significantly smaller than the spread of the maxima.

The figure shows that a pure maximum detection algorithm has a wider spread compared to an algorithm which locks onto the rising edge. A smaller spread of time values reduces the resulting error of a ToA or TDoA algorithm. Of course, the underlying time offset has to be compensated.

The following section makes use in practical measurements with the improved simulation approach.

8 Emulation of 5G Localization using VNA

The newly introduced approach is now applied to a new set of measurements. As the approach focuses on reducing the time error of one link, the experiment will consist of emulating only one base station UE combination. A whole TDoA or ToA procedure is not needed. When the accuracy of one distance measurement is increased, the resulting localization accuracy is also increased.

The measurements are conducted in a 7 m by 7 m room. From a Vector Network Analyzer (VNA) one port is used as a receiver (UE) and one port is used as a transmitter (base station). The UE is placed at 25 different positions in the room. To detect unpredicted changes in the channel, a third antenna is placed at a reference position. For every position the measurement is conducted with greater bandwidth (7.5 GHz) and a restricted bandwidth of 100 MHz. The restricted bandwidth is used as a substitute for a 5G FR1 signal while the full bandwidth allows a better understanding of the channel.

The received CIRs are post processed with the proposed rising edge algorithm and the conservative maximum approach. For the full bandwidth, the two approaches determine the distance between transmitter and receiver correctly. The biggest deviation is about 10 cm. When using the limited bandwidth, the errors naturally increase. Here, the rising edge algorithm allows for a better distance measurement accuracy with a variance of 0.252 m and a standard derivation of 0.502 m compared to the maximum peak algorithm which has a variance of 1.403 m and standard deviation of 1.184 m. Obviously wrong detections are excluded. The measured distances for a 100 MHz system with both approaches are depicted in Fig. 6.

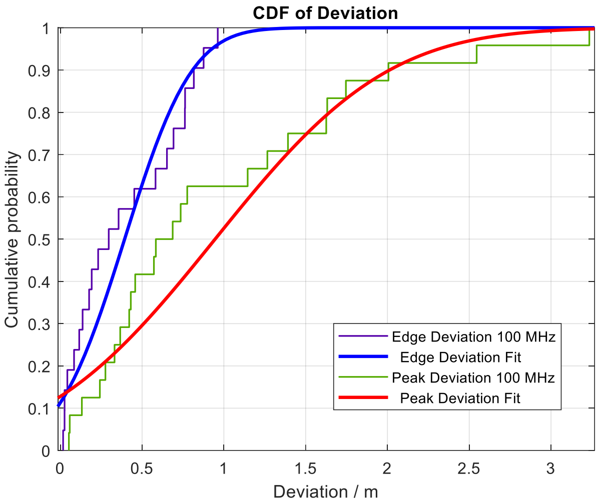


Fig 5: Comparison of deviations of rising edge and maximum approach

As shown, the usage of a rising edge detection algorithm allows an increase in accuracy. The standard derivation of the accuracy error can be halved by using the rising edge detection, compared to a maximum search detection.

9 Conclusion and Future Work

This paper shows the practical use case of 5G localization in industrial environments and the challenges occurring during measurement implementation. Existing 5G public communication networks do not provide the service of localization. The alternative is a campus network which depends on availability of off the shelf components. As localization using signal generators, transmitting 5G test signals, suffers from critical synchronization, a VNA based approach helps to overcome discussed issues. This paper discussed a different methodology to identify the LOS path. A basic maximum search is proposed and afterwards improved by using the rising edge of the CIR. With this approach, an accuracy of sub 1 m is possible for indoor scenarios with 5G NR FR1 bandwidth limitations.

Future work investigates trajectories of moving targets in close indoor environments to further improve localization accuracy.

10 References

[3G23] 3GPP Releases, https://www.3gpp.org/specifications-technologies/releases, as of 06.09.2023

[Tr16] TAHAT, Ashraf, et al. A look at the recent wireless positioning techniques with a focus on algorithms for moving receivers. IEEE Access, 2016, 4. Jg., S. 6652-6680.

[Ga20] GARCÍA, Adrián Cardalda; MAIER, Stefan; PHILLIPS, Abhay. Location-Based Services in Cellular Networks: from GSM to 5G NR. Artech House, 2020.

[Dw21] DWIVEDI, Satyam, et al. Positioning in 5G networks. IEEE Communications Magazine, 2021, 59. Jg., Nr. 11, S. 38-44.

[Hu22] HUANG, Siyu, et al. Positioning Performance Evaluation for 5G Positioning Reference Signal. In: 2022 2nd International Conference on Frontiers of Electronics, Information and Computation Technologies (ICFEICT). IEEE, 2022. S. 497-504.

[Tr21] TRIVEDI, Meet Ameet; et al.. Localization and Tracking of High-speed Trains Using Compressed Sensing Based 5G Localization Algorithms. In: 2021 IEEE 24th International Conference on Information Fusion (FUSION). IEEE, 2021. S.1-8.

[Pa22] PAPP, Zsófia, et al. TDoA based indoor positioning over small cell 5G networks. In: NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2022. S. 1-6.

[Ha23] HÄGER, Simon; GRATZA, Niklas; WIETFELD, Christian. Characterization of 5G mmWave high-accuracy positioning services for urban road traffic. In: Proc. IEEE VTC-Spring. 2023.

[Pa23] PALAMÀ, Ivan, et al. From Experiments to Insights: A Journey in 5G New Radio Localization. In: 2023 21st Mediterranean Communication and Computer Networking Conference (MedComNet). IEEE, 2023. S. 74-82.

[Ru22a] Y. Ruan, L. Chen, X. Zhou, G. Guo and R. Chen, Hi-Loc: Hybrid Indoor Localization via Enhanced 5G NR CSI," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp.

[Ru22b] RUAN, Yanlin, et al. iPos-5G: Indoor positioning via commercial 5G NR CSI. IEEE Internet of Things Journal, 2022, 10. Jg., Nr. 10, S. 8718-87331-15, 2022, Art no. 5502415, doi: 10.1109/TIM.2022.3196748

[Zh21] ZHANG, Zhaohan, et al. AoA-and-amplitude fingerprint based indoor intelligent localization scheme for 5G wireless communications. In: 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2021. S. 1-5.

[Mu21] MUKHTAR, Hind; EROL-KANTARCI, Melike. Machine learning-enabled localization in 5g using lidar and rss data. In: 2021 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2021. S. 1-6.

[Ga17] GAO, Caicai; WANG, Guohua; RAZUL, Sirajudeen Gulam. Comparisons of the super-resolution TOA/TDOA estimation algorithms. In: 2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL). IEEE, 2017. S. 2752-2758.

[Kn22] KNITTER, M.; KAYS, R. Channel Sounding Measurements for 5G Campus Networks in Industrial Environments. In: 2022 32nd International Telecommunication Networks and Applications Conference (ITNAC). IEEE, 2022. S. 1-6.

http://dx.doi.org/10.25673/111749

Die Bedeutung des Digital Twins auf Basis einer aktiven Verwaltungsschale für Kommunikations-Komponenten

Betrachtung am Beispiel von Industriesteckverbindern

Andreas Huhmann

Abstract:

Kommunikations-Komponenten sind in der industriellen Applikation eine Grundlage der Digitalisierung. Dabei wurde lange außer Acht gelassen, dass sie als reale Assets ebenfalls der Digitalisierung unterliegen. Damit rückt in dieser Betrachtung der Digital Twin von Kommunikations-Komponenten in den Mittelpunkt. Das Schlüsselelement der Digitalisierung ist die Verwaltungsschale (AAS Asset Administration Shell).

Der Ursprung dieser Betrachtung ist in der vierten industriellen Revolution (Industrie 4.0) zu finden. Im Referenzarchitekturmodell ist verankert, dass auch jede industrielle Komponente einen Digital Twin besitzt. Das verbindende Element zwischen realer Komponente und Digital Twin ist die Verwaltungsschale. Steckverbinder als passive Komponenten werden zukünftig durch eine passive Verwaltungsschale repräsentiert. Die Nutzung der Verwaltungsschale vereinfacht die Integration der Komponente im Rahmen des Engineerings. In Fällen, dass der Steckverbinder smarte Zusatzfunktionen besitzt, bildet eine aktive Verwaltungsschale das Mittel der Wahl. Dabei bleibt der Charakter der Komponente erhalten und es wird verzichtet, den Steckverbinder in eine komplexe Netzwerkkomponente zu transformieren. Diese Beschränkung ist entscheidend, damit der Steckverbinder universell einsetzbar bleibt, zum Beispiel auch in einer Infrastrukturkomponente. Die aktive Verwaltungsschale kann genutzt werden, um die Zustände des Steckverbinders im Betrieb zu erfassen und Zustandsübergänge einzuleiten.

Durch diese neue Betrachtung ist der smarte elektrische Connector (SmEC) in einen ganzheitlichen Ansatz für alle Steckverbinder eingebettet. Das führt zu einem skalierbaren Konzept, das applikationsabhängig den passenden Funktionsumfang zur Verfügung stellt.

Allgemein wird über den Steckverbinder sehr gut transparent, welche weitreichenden Konsequenzen Industrie4.0 auf alle Assets hat. Die Erkenntnisse sind daher sehr gut auf weitere Komponenten, auch weitere Kommunikations-Komponenten, übertragbar.

Keywords: Asset Administration Shell, Digital Twin, Connectivity, Smart electrical Connector,

HARTING Technology Group, Strategy Consultant, Marienwerder Straße 3, 32339 Espelkamp andreas.huhmann@HARTING.com

1.1 Ableitung des digital Twin aus dem RAMI Model

Bereits das Referenzmodel RAMI4.0 beschreibt das Asset als Typ und Instanz. Hierin ist die Grundlage für den digital Twin mit einer Verwaltungsschale gelegt. Die Verwaltungsschale wurde zu Beginn des Referenzmodells noch nicht eindeutig spezifiziert. Diese Definition mit all ihren Sub-Modellen findet durch die internationale Standardisierung (IEC) erst heute statt. Die AAS (Asset Administration Shell) kann also als die konsequente Umsetzung und Grundstein der Industrie 4.0 gedeutet werden.

Die standardisierte AAS bildet eine Grundlage für die Zusammenarbeit innerhalb von Ökosystemen. Unternehmensübergreifende Prozesse und der Datenaustausch in der Industrie werden damit einheitlich möglich (Manufacturing-X und Catena-X.)

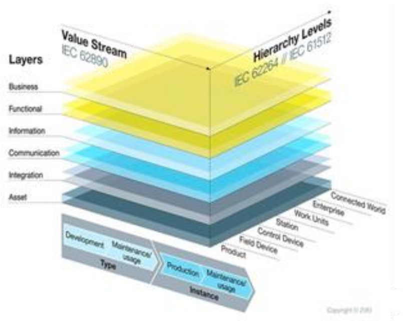


Abb. 1: Referenzarchitektur-Model RAMI4.0

1.2 Der digital Tin eines passiven Steckverbinders

Im RAMI4.0 sind Typ und Instanz wichtige Begriffe, die sich auf Assets beziehen. Ein Typ ist eine abstrakte Beschreibung eines Assets, während eine Instanz eine konkrete Realisierung dieses Typs ist.

Der Steckverbinder als elektromechanische Komponente ist durch spezielle Eigenschaften gekennzeichnet, die für den Typ des Steckverbinders relevant sind. Eine weitergehende Individualisierung des Steckverbinders in Rahmen einer Instanz ist demgegenüber zumeist für einfache elektromechanische Komponenten nicht notwendig.

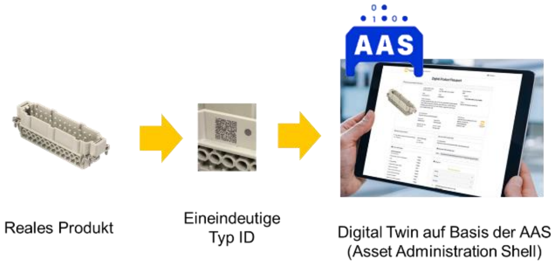


Abb. 2: Die Verwaltungsschale des Steckverbinders

Wird die konsequente Ableitung aus dem RAMI4.0 betrachtet, handelt es sich um eine konsequente Ableitung. Sich dieses vor Augen zu führen, halte ich für äußerts wichtig, da wir uns immer noch in der Umsetzungsphase von Industrie 4.0 befinden:

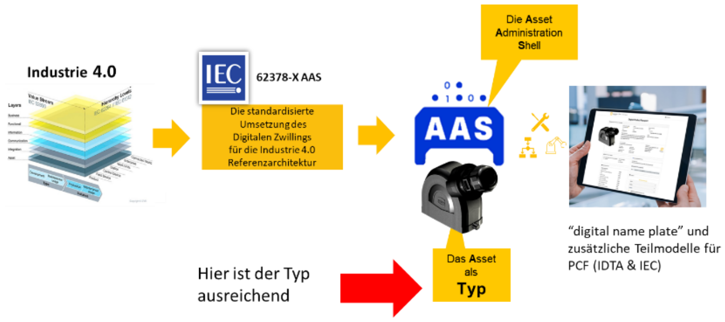

Industrie 4.0
IEC
62378-X AAS
Die standardisierte Umsetzung des Digitalen Zwillings für die Industrie 4.0 Referenzarchitektur
Die Asset Administration Shell
AAS
"digital name plate" und zusätzliche Teilmodelle für PCF (IDTA & IEC)
Hier ist der Typ ausreichend
Das Asset als Typ

Abb. 3: Ableitung der Verwaltungsschale aus RAMI4.0

Bereits durch diese Form der Typ Verwaltungsschale ergeben sich weitreichende Vorteile bei der Integration von Komponenten wie Steckverbindern in die kundenseitigen Prozesse.

Beim Engineering des Steckverbinders entsteht die Verwaltungsschale des Steckertyps mit den folgenden Teilmodellen:

- Identification
- Nameplate Technical Data
- Documentation
- MCAD, ECAD, BOM, Service
- Capabilities, Operational Data

- Digital Product Passport

Wird der Steckverbinder isoliert aus dem Aspekt Nachhaltigkeit betrachtet, so soll:

- der PCF (Product Carbon Footprint) minimal sein (Cradle to Gate)
- das eingesetzte Material nicht schädlich für Mensch und Umwelt sein und in eine Kreislaufwirtschaft eingebettet werden
- die Produktion sozial- und umweltgerecht sein
- das Produkt langlebig sein

Bezogen auf die ökologische Nachhaltigkeit stellt die Typ AAS ein geeignetes Mittel dar, den Product Carbon PCF transparent zu machen, der über das Produkt ohne AAS so nicht identifizierbar wäre.

Ziele für die Nachhaltigkeit eines Steckverbinders sind:

- direkt reduzierter CO_2-Fußabdruck des Steckverbinders durch Konstruktion und Produktion
- Verwendung von umweltfreundlichen Materialien,...
- Reduzierung des CO_2-Fußabdrucks im gesamten Lebenszyklusdurch technische Eigenschaften und Integrations-Schnittstellen (Engineering, Installation, Betrieb: technische Merkmale, Gewicht,...)

Diese Angaben können ideal in der AAS dokumentiert werden. Es entsteht daher Transparenz ohne die Nachhaltigkeit nicht nachhaltig umsetzbar ist.

1.3 Der digital Twin eines intelligenten Steckverbinders

In Fällen, dass der Steckverbinder smarte Zusatzfunktionen besitzt, so bildet eine aktive Verwaltungsschale das Mittel der Wahl, um zustandsabhängige Funktionen abzubilden.

Der smarte Steckverbinder wird in der Normierung als SmEC (Smart Electrical Connector) bezeichnet.

Die aktive Verwaltungsschale kann genutzt werden, um die Zustände des Steckverbinders im Betrieb zu erfassen und Zustandsübergänge einzuleiten.

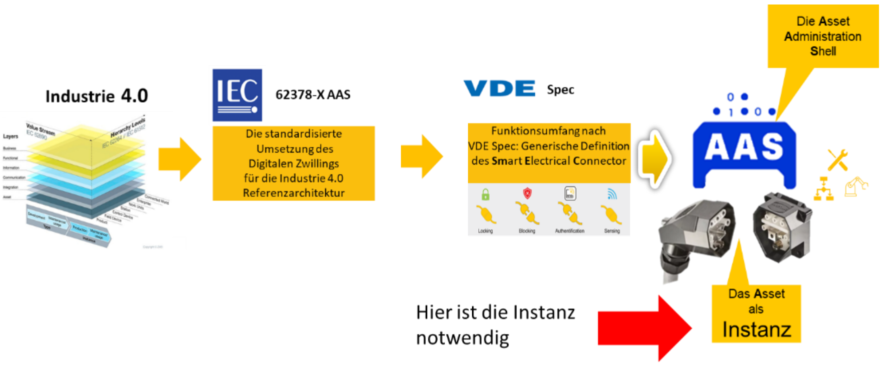

Industrie 4.0
Layers
Value Stream
Hierarchy Levels
IEC 62378-X AAS
Die standardisierte Umsetzung des Digitalen Zwillings für die Industrie 4.0 Referenzarchitektur
VDE Spec
Funktionsumfang nach VDE Spec: Generische Definition des Smart Electrical Connector
Locking
Blocking
Authentification
Sensing
0 1 0
AAS
Die Asset Administration Shell
Das Asset als Instanz
Hier ist die Instanz notwendig

Abb. 4: Ableitung der Verwaltungsschale eines intelligenten Steckverbinders

Der im Engineering entstandene Digitale Zwilling wird als Instanz im Betrieb verwendet. Der Steckverbinder kommuniziert über OPC-UA mit den überlagerten Systemen. Der Steckverbinder kann zustandsabhängige Funktionen besitzen, wie eine zustandsabhängige Verriegelung.

Eine neue Dimension der effizienten Interoperabilität zwischen Geschäftsanwendungen schafft eine ganzheitliche Sicht auf Produkte wie dem SmEC über den gesamten Produktlebenszyklus mit einer standardisierten AAS (z. B. SAP S/4 HANA Materialstamm / digitales Typenschild).

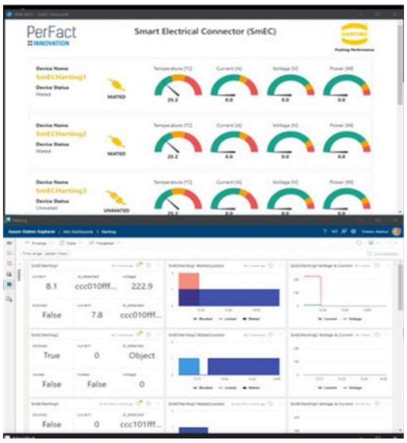


Abb. 5: Zustände eines SmEC im produktiven System

1.4 Beispiel Use Case für einen intelligenten Steckverbinder

Neben den klassischen Business Prozessen der Information Technology (IT) ist die Einbindung eines intelligenten Steckverbinder wie beschrieben auch im Kontext der Operational Technologie (OT) zu betrachten.

Im Lebenszyklus eines Steckverbinders sind mit höchster Relevanz die Use-Cases zu betrachten, die eine autarke Funktion als Installationskomponente in hochmodularen Produktionsanlagen darstellen und im Lebenszyklus oftmals gesteckt werden.

Die Kenntnis der einwandfreien Funktion, des „Fitnesszustandes" kann daher äußerst wichtig sein. Da dieser Zustand vom individuellen Steckverbinder abhängig ist, wird dieser zur eineindeutigen Instanz und ist kommunikationstechnisch mit seinem Digitalen Zwilling in Form der Verwaltungsschale (AAS) verbunden.

Bei einem flexiblen Anlagendesign, dass eine Rekonfiguration durch den Anwender erfordert, nimmt der SmEC als Schnittstelle eine Schlüsselfunktion ein. Es geht dabei um das sichere und richtige Verbinden und Stecken. Dabei wird die Identifikation des SmEC durch NF RFID Technologie dazu genutzt, zu erkennen, ob der richtige Steckverbinder am Einsatzort gesteckt wird. Die zustandsabhängige Verriegelung sichert den Steckvorgang dadurch ab, dass beispielsweise ein Ziehen unter Last verhindert wird. Hierzu kann der SmEC autonome Funktionen enthalten.

Abb. 5: Der Use-Case des SmEC in der Smart Factory KL

Der Zustand des SmEC wird standardisiert übertragen, ein präferierter Übertragungskanal ist dabei OPC UA. Die Zustände werden zur Steuerung der Gesamtinfrastruktur einer modularen Produktionsanlage an ein Superior System, wie in Abb. 6 zu sehen, übermittelt.

Das Superior System hat dabei auch die Möglichkeit Zustandsübergänge des SmEC auszulösen, beispielsweise um das Ziehen eines Steckverbinders und damit die Um-Konfiguration zu ermöglichen. Dieses Management einer Produktionsanlage wird als Production Level 4 Use-Case in der Smart Factory KL bearbeitet.

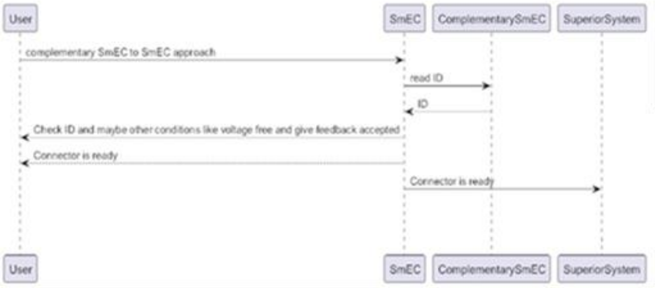


Abb. 6: Diagramm der zustandsabhängigen Verriegelung eines SmEC

1.5 Resümee und Ausblick

Durch diese neue Betrachtung ist der smarte elektrische Connector (SmEC) in einen ganzheitlichen Ansatz für alle Steckverbinder eingebettet. Das führt zu einem skalierbaren Konzept, das applikationsabhängig den passenden Funktionsumfang zur Verfügung stellt. Es werden so maßgeschneiderte Steckverbinder möglich, die im gesamten Lebenszyklus Nutzen generieren. Die AAS wird so zum universellen Standard, auch für Steckverbinder.

Bezogen auf die industrielle Kommunikation sehe ich durch die AAS die Chance, Installationskomponenten an Life Cycle Services zu koppeln, ohne diese zu einem Automation-Device zu transformieren. Der Vorteil liegt darin, dass dadurch der Aufwand beherrschbar bleibt, da die Automatisierungsarchitektur unverändert bleibt. So können ehemals passive Komponenten auf der Ebene der Komponenten verbleiben. Die vor vielen

Jahren propagierte Aufwärtsintegration wird durch die AAS in eine Digitalisierung ohne Hierarchiewechseln abgebildet (abb.7). Damit wandert die industrielle Kommunikations-Komponente ohne Systembruch in die virtuelle Welt.

Ein Steckverbinder bleibt Steckverbinder.

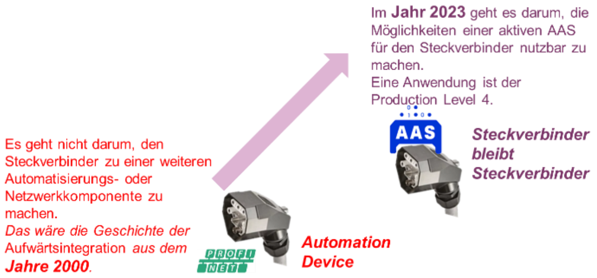

Im Jahr 2023 geht es darum, die Möglichkeiten einer aktiven AAS für den Steckverbinder nutzbar zu machen.
Eine Anwendung ist der Production Level 4.
AAS
Steckverbinder bleibt Steckverbinder
Es geht nicht darum, den Steckverbinder zu einer weiteren Automatisierungs- oder Netzwerkkomponente zu machen.
Das wäre die Geschichte der Aufwärtsintegration aus dem Jahre 2000.
PROFI NET
Automation Device

Abb. 7: Das Resümee „Steckverbinder bleibt Steckverbinder“

Literaturverzeichnis

[1] Smart Factory KL; Arbeitsdokument 2022/2023, „Anforderungen einer modularen und re-konfigurierbaren Produktion nach Production Level 4 an eine smarte Steckverbinder-Schnittstelle“ sowie weitere Arbeitsdokumente der Arbeitsgruppe Production Level 4 Infrastructure

[2] Andreas Huhmann; Vortrag Electronica 2022, SPS 2022, „Mit dem intelligenten Steckverbinder zu Production Level 4“

[4] IDTA (Industrial Digital Twin Association e.V.), Submodel_Template_VVS, 2022

[5] DKE/AK 651.0.3; „Steckverbinder mit Zusatzfunktion“ (Arbeitsgruppeninternes Draft)

[6] Andreas Huhmann, Vortrag Steckverbinder Kongress 2023, Würzburg

[8] DIN SPEC 91345:2016-04 - Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)

http://dx.doi.org/10.25673/111742

Standardisierte Maschinenanbindung an ein Produktionsleitsystem über die Asset Administration Shell

Gewinnung und Auswertung von Maschinendaten für die Prozessoptimierung

Alexander Schließmann[1], Melanie Stolze[2], Matthias Riedl[3], und Tizian Schröder[4]

Abstract: Die Verwaltung unterschiedlicher Maschinentypen von verschiedenen Herstellern in einer heterogenen Produktionsanlage ist aufgrund proprietärer Schnittstellen und unterschiedlicher Informationsmodelle umständlich. Dies erschwert eine ganzheitliche Planung und macht die Prozessoptimierung komplex. Der Beitrag zeigt die Herausforderungen bei der Integration verschiedener Maschinen in ein Produktionsleitsystem auf und schlägt eine generische Lösung unter Verwendung von Asset Administration Shells (AAS) als digitale Zwillinge und etablierter Kommunikationsprotokolle wie OPC UA vor. Zudem werden Vorgehen und Tools für die praktische Umsetzung des Konzepts exemplarisch beschrieben.

Keywords: Asset Administration Shell, Digitaler Zwilling, Maschinenanbindung, OPC UA

1 Einleitung

Produktionsanlagen für eine spezifische Bauteilefertigung bestehen in der Regel aus einem heterogenen Maschinenpark. Dabei werden unterschiedliche Maschinentypen eingesetzt, die unter anderem von verschiedenen Herstellern stammen. Diese Typenvielfalt erschwert eine übergreifende Planung und Auswertung der Maschinendaten, da herstellerspezifische Schnittstellen und unterschiedliche Informationsmodelle Stand der Technik sind. Ein entscheidender Punkt für den Anlagenbetreiber stellt das Fehlen bzw. der außerordentlich hohe Aufwand einer durchgängigen Prozessoptimierung über den gesamten Maschinenpark dar. So ist es für ihn sehr komplex, fundierte Aussagen über Investitionen in neue Maschinen und/oder Änderungen im Maschinenpark zu treffen, da diese nur isoliert pro Maschine getroffen werden können. Eine Prognose zu Auswirkungen auf den Gesamtprozess ist in der Regel nicht möglich.

[1] FORCAM ENISCO, Strategisches Produktmanagement, Herrenberge Str. 56, 71034 Böblingen, alexander.schliessmann@forcam.com

[2] Institut für Automation und Kommunikation e.V., IKT & Automation, Werner-Heisenberg-Str. 1, 39106 Magdeburg, melanie.stolze@ifak.eu

[3] Institut für Automation und Kommunikation e.V., IKT & Automation, Werner-Heisenberg-Str. 1, 39106 Magdeburg, matthias.riedl@ifak.eu

[4] Institut für Automation und Kommunikation e.V., IKT & Automation, Werner-Heisenberg-Str. 1, 39106 Magdeburg, tizian.schroeder@ifak.eu

Eine erfolgversprechende Möglichkeit, die Planungs- und Optimierungsaufgaben zu lösen, liegt in der Nutzung digitaler Zwillinge der Maschinen bzw. der darin verbauten Komponenten. Zur übergreifenden Nutzung müssen die Informationen, sowie der Zugriff auf digitale Zwillinge und deren Verwendung vereinheitlicht werden. Konkret lassen sich somit vorab Aussagen treffen, welches Bauteil auf welcher Maschine mit welcher Technologie zu bestimmten Kosten produziert werden kann. Die dazu notwendigen informationstechnischen Technologien zur interoperablen Bereitstellung der Informationen durch beispielsweise OPC UA oder der Asset Administration Shell (deutsch Verwaltungsschale) aus der Plattform Industrie 4.0 sind mittlerweile vorhanden bzw. reifen sehr schnell.

Der Artikel möchte ein Integrationskonzept aufzeigen, das die breite Facette an standardisierten Schnittstellentechnologien nutzt und eine Anbindung an die Asset Administration Shell ermöglicht. Kapitel 2 identifiziert zu Beginn die Schwierigkeiten in der aktuellen Vorgehensweise zur Anbindung einer Maschine an ein Produktionsleitsystem. Zudem wird gezeigt, welche der bisher entwickelten Teilmodelle der Asset Administration Shell zur Problemlösung beitragen und welche notwendigen Informationen zur Maschinenanbindung noch zu modellieren sind. Zusätzlich wird auf den Zusammenhang der Asset Administration Shell mit dem Equipment Behaviour Catalogue eingegangen. In Kapitel 3 folgt die Vorstellung des entwickelten Konzepts, das als Basis die gewonnenen Ergebnisse aus dem vorhergehenden Kapitel nutzt. Hierbei wird vor allem auf den anwendungsfallspezifischen Inhalt der Asset Administration Shell und der zu entwickelnden Teilmodelle eingegangen. Zusätzlich erfolgt die Erläuterung einer Möglichkeit zur praktischen Integration der Maschinenanbindung an den digitalen Zwilling. Abgeschlossen wird der Artikel mit einer Zusammenfassung, die die Integrationsreife des Konzeptes analysiert und mögliche weitere Forschungsthemen aufzeigt. Die hier vorgestellten Ergebnisse und Ideen stammen aus dem vom Bundesministerium für Wirtschaft und Klimaschutz geförderten Forschungsprojekt TwinMaP [Tw23].

2 Stand der Technik

2.1 Schwierigkeiten bei der Maschinenanbindung an ein Produktionsleitsystem

Bei der Maschinenanbindung an ein Produktionsleitsystem existieren zwei Herausforderungen, und zwar die Gewinnung der Maschinendaten und deren Interpretation. Die Datengewinnung gestaltete sich vor einigen Jahren noch schwierig, da Standards wie OPC UA und dessen Companion Specifications für den interoperablen Datenaustausch wenig Anwendung fanden. Die Integration von OPC UA in Maschinen steigt jedoch von Jahr zu Jahr, und trägt somit zur Verbesserung der Datengewinnung bei. Jedoch gibt es immer noch einen sehr großen Bestand an Maschinen, die nicht über den OPC UA Standard verfügen. Von serieller Kommunikation, über firmenspezifische

Protokolle einschließlich einer nicht vorhandenen Kommunikationsmöglichkeit ist die Bandbreite an Herausforderungen für die Datengewinnung groß. Zur Schließung dieser Informationslücken in der Datengewinnung gibt es mittlerweile gute Überbrückungslösungen wie z. B. FORCAM FORCE EDGE.

Die zweite Herausforderung ist die produktionssemantische Interpretation der Daten. Aus den gewonnenen Rohsignalen einer Maschine können nicht direkt betriebswirtschaftlich nutzbare Kennzahlen abgeleitet werden. So erlaubt das Signal „Kühlmittel ein“ keinen Rückschluss auf den Ressourcenverbrauch der Maschine, hier sind weitere Informationen wie der Verbrauch an Energie und Kühlmittel relevant. Das Gleiche gilt für die Ermittlung des Maschinenbetriebszustands „Produktion“. Zur Bestimmung des Betriebszustands werden kontinuierliche Werte wie Vorschub, Spindelvorschub, Drehzahl, Temperatur, etc. benötigt, deren Zusammenhang über eine Schaltfunktion beschrieben ist. Abhängig vom Maschinentyp sind zudem besondere Betrachtungen für mehrkanalige Maschinen, Multistation oder Mehrspindler notwendig. Eine beispielhafte Definition von „Produktion“ kann sich dann aus dem gleichzeitigen Anliegen folgender Rohsignale ergeben:

> Produktion = Automatik-Betrieb UND Programm läuft UND (Vorschubschalter = 100%) UND (Eilgang = 100%) UND (Spindeldrehzahl = 100%) UND NC Betriebsbereit

Signale bzw. Signalkombinationen, die nicht in dieser Definition enthalten sind, führen zum Maschinenstillstand bzw. verallgemeinert ausgedrückt zur Maschinenstörung.

Zusätzlich zu diesen Maschinensignalen gibt es Maschinenmeldungen. Diese führen aber nicht zu einem Statuswechsel der Maschine und sind deswegen für die Ermittlung der Kennzahlen nicht relevant.

2.2 Asset Administration Shell

Die Asset Administration Shell (AAS) ist ein Konzept, das von der Plattform Industrie 4.0 entwickelt wurde, um die horizontale Interoperabilität von Geräten und Systemen zu verbessern und auch Geschäftsprozesse automatisiert abzuwickeln. Eine AAS kann Daten über ihr verwaltetes Asset aus unterschiedlichen Quellen aufnehmen, beispielsweise durch das Mithören auf dem Feldbus, das Lesen von Diagnoseinformationen aus einem Gerät oder die Vorverarbeitung in einem Edge Device. Die aufgenommenen Daten werden über standardisierte Teilmodelle [ID23a] in der AAS gespeichert und anderen Teilnehmern zur Verfügung gestellt.

Für die generische Anbindung einer AAS an ein Asset kann das Asset Interface Description (AID) Teilmodell der IDTA genutzt werden (IDTA-Nr.: 02017). In diesem Teilmodell werden sowohl die Metainformationen zu den einzelnen Kommunikationstechnologien als auch die Endpunkte der Assetsignale erfasst. Der Entwurf des Teilmodells basiert auf der Web of Things Thing Description [KMK23]. In

[OKD23] steht beschrieben, dass mit dem Teilmodell die Einbindung von Kommunikationstechnologien wie HTTP, MQTT, OPC UA, Modbus-TCP, PROFIBUS und PROFINET möglich ist. Der Aufbau und die Spezifikation des Teilmodells ist derzeitig noch nicht veröffentlicht, jedoch ist in [OKD23] der Aufbau des Teilmodells teilweise beschrieben.

Ein weiteres in Entwicklung befindliches Teilmodell der IDTA ist das Asset Interface Mapping Configuration (AIMC) Teilmodell (IDTA-Nr.: 02027). Dieses Teilmodell wird in Verbindung mit dem AID-Teilmodell genutzt, um anderen Teilmodellen den Zugriff über die Endpunkte auf die Assetdaten zu beschreiben. Somit bildet das AID-Teilmodell die zentrale Beschreibung aller Endpunkte für die Assetdaten, während das AIMC-Teilmodell die Datenverteilung nach Anwendungsbereich vornimmt.

Andere bereits veröffentlichte Teilmodelle wie das digitale Typenschild [ID22b], die technischen Daten [ID22a] oder die Übergabedokumentation [ID23b] zu einem Asset, können auch für die AAS eines Assets implementiert werden. Sie spielen in diesem Artikel eine untergeordnete Rolle und werden deshalb nicht näher beschrieben.

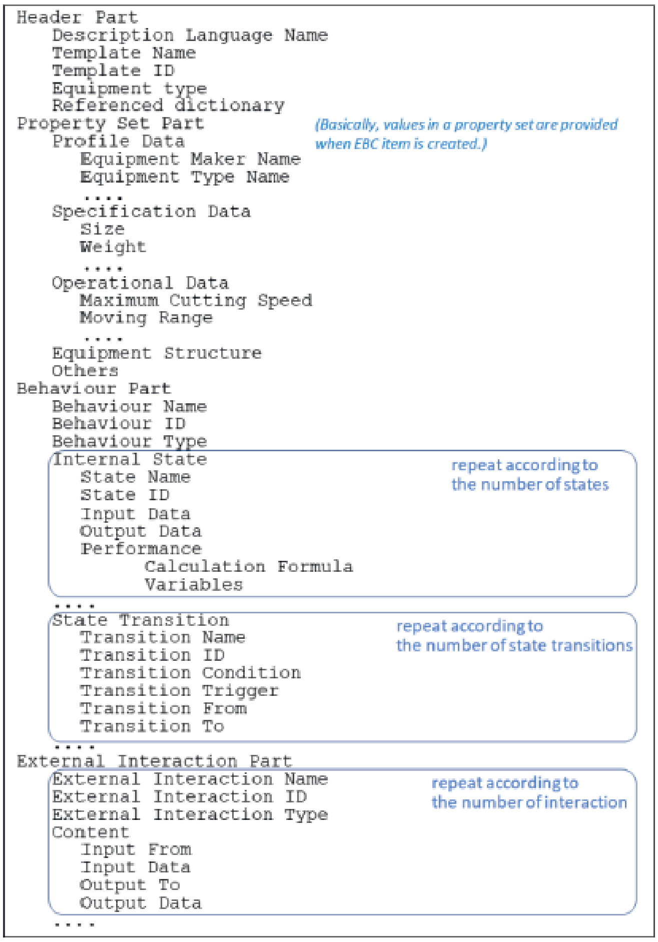
2.3 Equipment Behaviour Catalogue

Zusätzlich zu der von Deutschland getriebenen AAS-Initiative gibt es noch den Ansatz des Equipment Behaviour Catalogue (EBC) der ISO/AWI 16400-5 [ISOc]. Die Gemeinsamkeit beider Ansätze spiegelt sich darin wider, dass sie Systeme beschreiben. Die AAS dient vorrangig der einheitlichen Beschreibung von Assets/Objekten/Entitäten, um einen interoperablen Datenaustausch zu ermöglichen. Der EBC wird motiviert durch die formale Beschreibung von Automatisierungssystemen und deren Interaktionen, um darauf aufbauend Simulationssysteme zu erstellen. Die grundlegende Struktur eines Automatisierungssystems setzt sich mit dem EBC aus der Beschreibung

- der fertigungstechnischen Kompetenzen (z. B. Größe, Schnittgeschwindigkeit, maximale Bearbeitungsgröße),
- den möglichen Zuständen (z. B. Standby, Einfahren, Einrichten, Produktion),
- den Bedingungen für die Transitionen (z. B. Tür geschlossen UND NC-Programm läuft UND Override = 100%) und
- der produktionstechnischen Parameter zusammen.

Abb. 1 zeigt, wie so eine Beschreibungsstruktur aufgebaut ist.

Im EBC ist jedoch nicht festgelegt, wie die genannten Informationen erfasst bzw. dargestellt werden. Eine Verknüpfung von EBC und AAS kann insofern erfolgen, dass der EBC zur Erstellung von Teilmodellen mit in der AAS genutzt werden kann, um einen Mehrwert für interoperable Verhaltensbeschreibungen zu erzeugen [ISOa].



```
Header Part
   Description Language Name
   Template Name
   Template ID
   Equipment type
   Referenced dictionary
Property Set Part                    (Basically, values in a property set are provided
   Profile Data                      when EBC item is created.)
     Equipment Maker Name
     Equipment Type Name
     ....
   Specification Data
     Size
     Weight
     ....
   Operational Data
     Maximum Cutting Speed
     Moving Range
     ....
   Equipment Structure
   Others
Behaviour Part
   Behaviour Name
   Behaviour ID
   Behaviour Type
   Internal State                         repeat according to
     State Name                           the number of states
     State ID
     Input Data
     Output Data
     Performance
          Calculation Formula
          Variables
   ....
   State Transition                   repeat according to
     Transition Name                  the number of state transitions
     Transition ID
     Transition Condition
     Transition Trigger
     Transition From
     Transition To
   ....
External Interaction Part
   External Interaction Name            repeat according to
   External Interaction ID              the number of interaction
   External Interaction Type
   Content
     Input From
     Input Data
     Output To
     Output Data
   ....
```

Abb. 1: Beispiel für die formale Beschreibungsstruktur eines EBC-Templates [ISOb]

3 Konzept zur Maschinenanbindung via Asset Administration Shell

Das in diesem Abschnitt vorgestellte Konzept zur Maschinenanbindung an ein Produktionsleitsystem basiert auf den zwei, in [SS14] beschriebenen, Transformationsschritten. Diese werden von dem Produktionsleitsystem ausgeführt und nutzen als Basis Daten aus der Maschinen-AAS. Damit werden aus den Maschinenrohsignalen und anderen maschinenzugehörigen Daten das Logbuch für das Produktionsleitsystem maschinell erstellt.

Zu Beginn der Anbindung der Maschine an das Produktionsleitsystem muss die Anbindung der Maschine an die AAS erfolgen. Dies geschieht in diesem Konzept mit Hilfe des in Kapitel 2.2 beschriebenen Asset Interface Description Teilmodells und der Nutzung von OPC UA. Nach der Verbindung von Maschine und AAS folgt der erste Transformationsschritt, der als Eingangsdaten die in der AAS berechneten Betriebssignale und Mengenmeldungen der Maschine erhält. Das setzt eine Basistransformation in der AAS voraus, die die Rohsignale durch logische Verknüpfungen in für das Produktionsleitsystem interpretierbare Betriebssignale wandelt, wie es in Kapitel 2.3 beschrieben wurde. Aus der ersten Transformation resultieren die Maschinenkennzahlen, die in dem VDMA-Einheitsblatt 66412-1 [VDM09] mit insgesamt 26 definierten Kennzahlen beschrieben sind und einen Einblick in den Maschinenstatus geben. Diese Kennzahlen werden in der Werks-AAS gespeichert, da sie sowohl Ist-Daten aus der Maschine als auch Planungsdaten von Aufträgen und anderen Ressourcen benötigen. Mit Hilfe dieser Kennzahlen ist die Ableitung von Stillstandszeiten, Störungen oder der Produktionszeit von Maschinen möglich. Diese Informationen sind in Teilen aber noch nicht aussagekräftig genug, um den Betriebszustand definieren zu können. Beispielsweise können Stillstandszeiten erfasst werden aber nicht der Grund für den Stillstand, da die zeitlichen Zusammenhänge mit dem Stillstand (z. B. Rüsten, Pause etc.) nicht in den Maschinensignalen enthalten sind. Deshalb folgt eine zweite Transformation, die die vorher berechneten Maschinenkennzahlen und den aktuellen Schichtplan der Maschinen-AAS nutzt und daraus den Betriebszustand der Maschine berechnet. Dieser Betriebszustand kann für betriebswirtschaftliche Entscheidungen in der Prozessoptimierung genutzt werden.

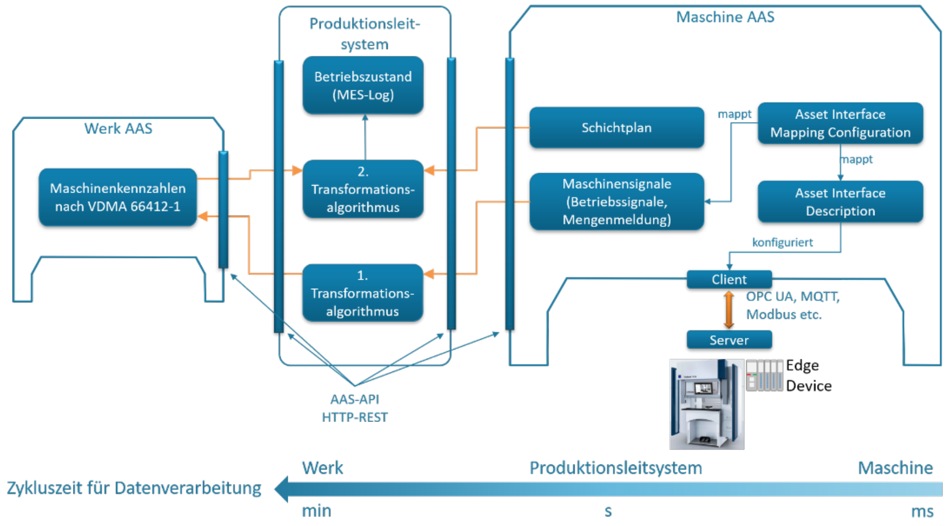


Abbildung 2: Konzept zur Maschinenanbindung an ein Produktionsleitsystem mit Hilfe der AAS

In Abb. 2 ist der soeben beschriebene Prozess mit den Teilmodellen, die die erwähnten Informationen beinhalten, dargestellt. Zusätzlich sind an einem Zeitstrahl die zur Verfügung stehenden Zykluszeiten der einzelnen Systeme zu sehen. Daraus geht eindeutig hervor, dass von der Maschine keine harten Echtzeitdaten für die Transformationsberechnungen benötigt werden und die AAS neben den verschiedenen Kommunikationstechnologien wie OPC UA, MQTT, Modbus etc. eine interoperable Alternative bildet.

In den folgenden Kapiteln wird die Umsetzung des beschriebenen Konzeptes erläutert. Hierbei wird die Implementierung der AAS und Verknüpfung dieser mit der Maschine beschrieben. Zudem wird ein erster Entwurf der Teilmodelle „Maschinensignale“ und „Maschinenkennzahlen“ beschrieben.

3.1 Implementierung des digitalen Zwillings

Zur praktischen Erprobung der vorgestellten Konzepte dient eine Multivendoranlage mit mehreren Assets. Jedes dieser Assets ist über eine AAS vom Typ 2 repräsentiert und ist so für übergeordnete Systeme über eine standardisierte HTTP-REST-Schnittstelle [PI21] zugänglich. Hierüber können Daten von den Assets gelesen und bei Bedarf zurückgeschrieben werden. Das Informationsmodell dieser AAS wird mit dem AASX Package Explorer [ID23c] erstellt und folgt einem einheitlichen Metamodell [PI22].

Für die Entwicklung der lauffähigen Softwareinstanzen der Typ 2-AAS dient das BaSyx .NET SDK [Ba23] des BaSys-Projekts als Ausgangspunkt. Das SDK wurde dahingehend erweitert, dass die mit dem AASX Package Explorer generierten Pakete im aasx-Dateiformat importiert werden können und somit eine aufwändige manuelle Erstellung des VWS-Informationsmodells im Quellcode erspart bleibt. In dem hier gewählten Szenario erfolgt das Hosting der jeweiligen AAS zu den Assets in einem gemeinsamen AAS-Repository. Alternativ ist eine Instanziierung mehrerer eigenständiger, verteilter AAS-Instanzen möglich.

Um Aktualdaten eines Assets in der AAS zu speichern, liest ein Client in der AAS die anlagenseitig, über einen OPC UA-Server, zur Verfügung gestellten Daten. Abb. 3 zeigt einen vereinfachten Ausschnitt des VWS-Informationsmodells im AASX Package Explorer und des OPC UA-Informationsmodells in einem grafischen OPC UA-Client zu den für die Kommunikationsintegration relevanten Assets. Die Modelle enthalten Maschinenrohsignale, die in einem nachgelagerten Schritt zu aggregierten Prozessgrößen zusammengeführt werden können. Die mit einem grünen Icon gekennzeichneten Variablen-Knoten sind mit den AAS zu verknüpfen, um Aktualdaten für das Produktionsleitsystem bereitzustellen. In der Abbildung sind die zuzuordnenden Elemente anhand ihrer Namensgleichheit zu erkennen, was für das vorzunehmende Mapping aber keine Voraussetzung ist. Die hier beschriebene Anbindung einer Maschine an die AAS ist eine Variante. Eine zweite Variante ist aber auch über die Verwendung der AAS als Beschreibungsmittel möglich. Hierbei beschreibt die AAS zum einen den Zugriff auf die

Maschinendaten und welche Maschinendaten für das Produktionsleitsystem miteinander verrechnet werden müssen, um die gewünschte Kennzahl zu erhalten. Somit liest das Produktionsleitsystem erst die Beschreibung in der AAS und kommuniziert dann direkt mit der Maschine, um die Anzahl an Datentransfers und somit die Zykluszeit für die Datenverarbeitung zu verringern.

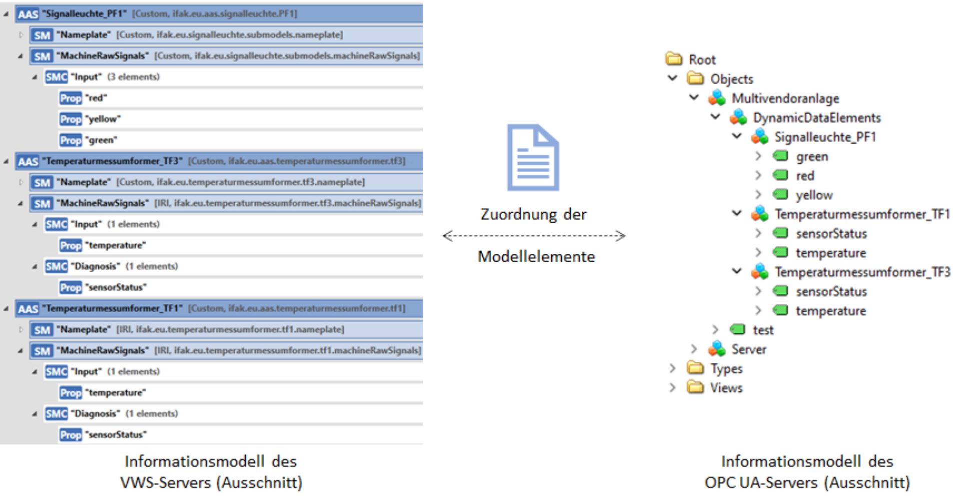

AAS "Signalleuchte_PF1" [Custom, ifak.eu.aas.signalleuchte.PF1]
SM "Nameplate" [Custom, ifak.eu.signalleuchte.submodels.nameplate]
SM "MachineRawSignals" [Custom, ifak.eu.signalleuchte.submodels.machineRawSignals]
SMC "Input" (3 elements)
Prop "red"
Prop "yellow"
Prop "green"
AAS "Temperaturmessumformer_TF3" [Custom, ifak.eu.aas.temperaturmessumformer.tf3]
SM "Nameplate" [Custom, ifak.eu.temperaturmessumformer.tf3.nameplate]
SM "MachineRawSignals" [IRI, ifak.eu.temperaturmessumformer.tf3.machineRawSignals]
SMC "Input" (1 elements)
Prop "temperature"
SMC "Diagnosis" (1 elements)
Prop "sensorStatus"
AAS "Temperaturmessumformer_TF1" [Custom, ifak.eu.aas.temperaturmessumformer.tf1]
SM "Nameplate" [IRI, ifak.eu.temperaturmessumformer.tf1.nameplate]
SM "MachineRawSignals" [IRI, ifak.eu.temperaturmessumformer.tf1.machineRawSignals]
SMC "Input" (1 elements)
Prop "temperature"
SMC "Diagnosis" (1 elements)
Prop "sensorStatus"
Informationsmodell des VWS-Servers (Ausschnitt)
Zuordnung der Modellelemente
Root
Objects
Multivendoranlage
DynamicDataElements
Signalleuchte_PF1
green
red
yellow
Temperaturmessumformer_TF1
sensorStatus
temperature
Temperaturmessumformer_TF3
sensorStatus
temperature
test
Server
Types
Views
Informationsmodell des OPC UA-Servers (Ausschnitt)

Abb. 3: Ausschnitte der Informationsmodelle von AAS und Asset

Für die kommunikationstechnische Anbindung der Assets an die AAS wurde das BaSyx-SDK um eine durch Teilmodelle konfigurierbare Asset-Interface-Klasse erweitert. Instanzen dieser Klasse implementieren einen OPC UA-Client, der die Anbindung des Assets über OPC UA an die AAS realisiert. Auch eine Anbindung über MQTT wird durch diese Klasse unterstützt. Weitere Kommunikationsprotokolle können leicht ergänzt werden. Die durch den OPC UA-Client vom Asset gelesenen Variablenwerte werden in AAS-internen Variablen abgelegt und sind somit über die durch das SDK implementierte REST-API auslesbar. Analog erfolgt dies für das Schreiben von Variablenwerten.

Die Konfiguration der AAS-Instanzen erfordert allgemeine Angaben wie den HTTP-Endpunkt unter dem die API der AAS erreichbar ist oder der Pfad zu der zu importierenden aasx-Datei, um das Informationsmodell aufzuspannen. Zudem sind die spezifischen Angaben für die Integration von AAS und Asset nötig. Dazu dienen künftig die in Kapitel 2.2 erwähnten Asset Interface Description (AID)- und Asset Interface Mapping Configuration (AIMC)-Teilmodelle. Diese enthalten unter anderem den Endpunkt, unter dem der Asset-seitige OPC UA-Server erreichbar ist, aber auch das konkrete Mapping zwischen dem Informationsmodell der AAS und dem OPC UA-Adressraum. Informationsmodellelemente der AAS tragen als Identifikatoren u. a. eine IdShort. Das Mapping ist als Zuordnung von IdShort-Pfaden der AAS zu NodeIds der OPC UA-Knoten realisiert und enthält auch die Festlegung der Zugriffsrechte. Bis zur

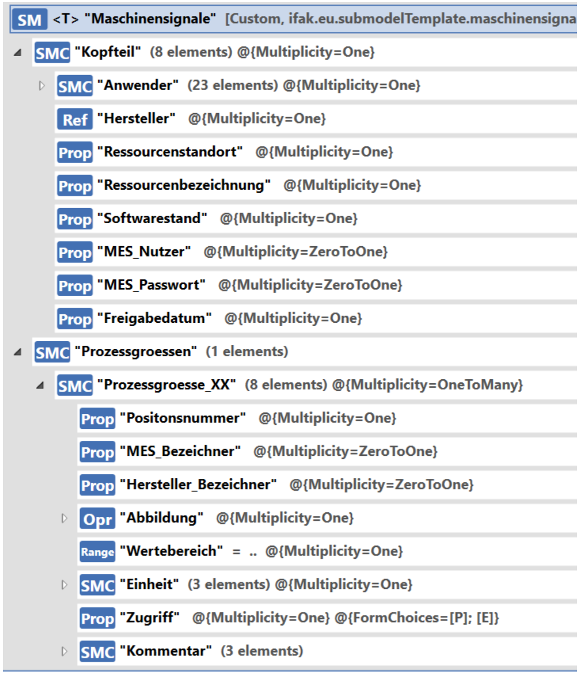
Veröffentlichung der AID- und AIMC-Teilmodelle wird im hier beschriebenen Demonstrator eine JSON-Konfigurationsdatei genutzt, die diese Angaben enthält.

3.2 Teilmodelle im Detail

Die Beschreibung der aus den Maschinenrohsignalen (s. Kapitel 3.1) berechneten Maschinensignale (Betriebssignale und Mengenmeldungen) und der daraus berechneten Maschinenkennzahlen nach VMDA 66412 erfolgt in Teilmodellen, wie sie in Abb. 2 gezeigt wurden. Das Teilmodell des Schichtplans wird in diesem Artikel nicht näher beschrieben. Neben dem AID und AIMC-Teilmodell (s. Kapitel 2.2) der IDTA sind bisher keine weiteren Entwicklungsarbeiten hinsichtlich der eben genannten Teilmodelle bekannt, sodass diese neu entwickelt wurden.

Maschinensignale

Das Teilmodell-Template sieht entsprechend der VDMA 66412-10 [VDM15] eine Gliederung der Informationen in zwei Bereiche vor. Der erste Bereich wird als Kopfteil bezeichnet und enthält alle für das Produktionsleitsystem relevanten Metadaten, wie den Anwender der Maschine, deren Standort und Bezeichnung.



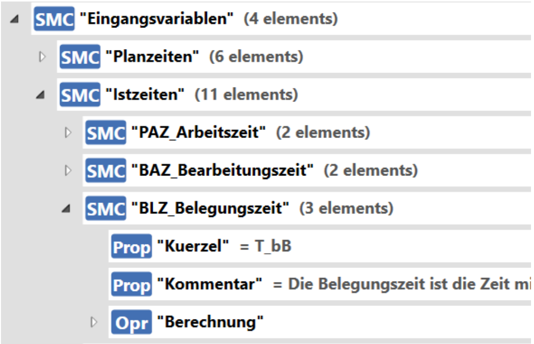
```
SM <T> "Maschinensignale"  [Custom, ifak.eu.submodelTemplate.maschinensigna
◢ SMC "Kopfteil"  (8 elements) @{Multiplicity=One}
   ▷ SMC "Anwender"  (23 elements) @{Multiplicity=One}
     Ref "Hersteller"  @{Multiplicity=One}
     Prop "Ressourcenstandort"  @{Multiplicity=One}
     Prop "Ressourcenbezeichnung"  @{Multiplicity=One}
     Prop "Softwarestand"  @{Multiplicity=One}
     Prop "MES_Nutzer"  @{Multiplicity=ZeroToOne}
     Prop "MES_Passwort"  @{Multiplicity=ZeroToOne}
     Prop "Freigabedatum"  @{Multiplicity=One}
◢ SMC "Prozessgroessen"  (1 elements)
   ◢ SMC "Prozessgroesse_XX"  (8 elements) @{Multiplicity=OneToMany}
        Prop "Positonsnummer"  @{Multiplicity=One}
        Prop "MES_Bezeichner"  @{Multiplicity=ZeroToOne}
        Prop "Hersteller_Bezeichner"  @{Multiplicity=ZeroToOne}
      ▷ Opr "Abbildung"  @{Multiplicity=One}
        Range "Wertebereich"  =  ..  @{Multiplicity=One}
      ▷ SMC "Einheit"  (3 elements) @{Multiplicity=One}
        Prop "Zugriff"  @{Multiplicity=One} @{FormChoices=[P]; [E]}
      ▷ SMC "Kommentar"  (3 elements)
```

Abb. 4: Teilmodell-Template "Maschinensignale"

Der zweite Bereich enthält alle wichtigen Betriebssignale und Mengenmeldungen, auch Prozessgrößen genannt, die für die Berechnung der Maschinenkennzahlen notwendig sind. Dabei setzt sich eine Prozessgröße aus zwei Bezeichnern zusammen. Einen den der Maschinenhersteller und einen den der Produktionsleitsystemhersteller vergibt. Weiterhin wird die Abbildung mehrerer Rohsignale zu der beschreibenden Prozessgröße vorgenommen. Im Teilmodell wird die Abbildung durch eine Operation realisiert, die als Eingangsvariablen die Rohsignale erhält und als Ausgangsvariable das Ergebnis für die Prozessgröße berechnet. Die Operation kann dabei von dem Produktionsleitsystem selbst durchgeführt oder an eine andere Applikation delegiert werden. Zusätzlich zu der Abbildung muss der Wertebereich und die Einheit der Prozessgröße sowie der Zugriff auf diese Prozessgröße beschrieben werden. Mit dem “Zugriff” ist das zyklische Abfragen (Polling) der Prozessgröße oder das eventbasierte Abfragen bei Wertänderungen durch eine Applikation gemeint. Mit Hilfe der Multiplizitäten neben jedem Teilmodellelement ist definiert, wie oft das entsprechende Element im Teilmodell vorhanden sein darf. Der grobe Aufbau des Teilmodell-Templates ist in Abb. 4 zu sehen.

Maschinenkennzahlen nach VMDA 66412

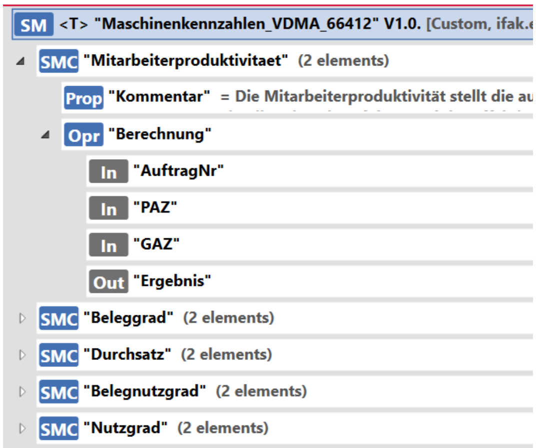
Das Teilmodell-Template gliedert sich nach der VDMA 66412-1 [VDM09] in zwei Bereiche auf. Der erste Bereich beschreibt die Eingangsvariablen, die für die Berechnung der Kennzahlen notwendig sind und aus vielen verschiedenen AAS extrahiert werden müssen. Dazu wird wiederum eine Untergliederung der Eingangsvariablen in Planzeiten, Istzeiten, logistische Mengen und Qualitätsdaten vorgenommen. In jeder Rubrik sind Variablen mit ihrem Kürzel (soweit vorhanden), ihrer Berechnung (wenn es zusammengesetzte Signale sind) und einem Kommentar versehen. Ein Beispiel für eine Variable aus der Rubrik Istzeit ist in Abbildung Abb. 5 zu sehen. Genau wie bei den Maschinensignalen, wird die “Eingangsvariable” mit Hilfe einer Operation berechnet. Die Eingangsvariablen werden dabei als Referenzen auf Elemente anderer AAS abgebildet. Das führt zu einer Entkopplung zwischen Maschinenkennzahl und der betrachteten Ressource.



```
◢ SMC "Eingangsvariablen"  (4 elements)
    ▷ SMC "Planzeiten"  (6 elements)
    ◢ SMC "Istzeiten"  (11 elements)
        ▷ SMC "PAZ_Arbeitszeit"  (2 elements)
        ▷ SMC "BAZ_Bearbeitungszeit"  (2 elements)
        ◢ SMC "BLZ_Belegungszeit"  (3 elements)
              Prop "Kuerzel"  = T_bB
              Prop "Kommentar"  = Die Belegungszeit ist die Zeit mi
            ▷ Opr "Berechnung"
```

Abb. 5: Teilmodell-Template "Maschinenkennzahlen" - Eingangsvariablen im Detail

Den zweiten Teil des Teilmodells bilden die 26 Kennzahlen des VDMA-Einheitsblattes. Zu diesen Kennzahlen gibt es immer einen Kommentar, der den Zweck der Kennzahl beschreibt, und eine Operation. Die Operation hat als Eingang Referenzen auf die vorher erwähnten Eingangsvariablen. Somit erfolgt eine Entkopplung der Kennzahlen von den Eingangsvariablen, die sich je nach betrachteter Ressource ändern können. Ein Beispiel dafür ist in Abbildung Abb. 6 dargestellt.



```
SM <T> "Maschinenkennzahlen_VDMA_66412" V1.0. [Custom, ifak.e
◢ SMC "Mitarbeiterproduktivitaet"  (2 elements)
    Prop "Kommentar"  = Die Mitarbeiterproduktivität stellt die au
  ◢ Opr "Berechnung"
      In "AuftragNr"
      In "PAZ"
      In "GAZ"
      Out "Ergebnis"
▷ SMC "Beleggrad"  (2 elements)
▷ SMC "Durchsatz"  (2 elements)
▷ SMC "Belegnutzgrad"  (2 elements)
▷ SMC "Nutzgrad"  (2 elements)
```

Abb. 6: Teilmodell-Template "Maschinenkennzahlen" - Maschinenkennzahl im Detail

Das Teilmodell der Maschinenkennzahlen ist zu 100% an das VDMA-Einheitsblatt angelehnt, weshalb einzig die benutzerdefinierte Anpassung der Referenzen für die Eingangsvariablen notwendig ist.

4 Zusammenfassung

Mit Hilfe der AAS und den Teilmodellen aus dem beschriebenen Konzept ist es möglich, Maschinendaten an ein Produktionsleitsystem unabhängig vom Maschinentypen bereitzustellen. Durch das AID-Teilmodell erfolgt die Beschreibung der protokollspezifischen Anbindung an ein Asset und über das AIMC-Teilmodell die Verknüpfung der Maschinenrohsignale in andere Teilmodelle. Dies dient, vor allem in Verbindung mit dem Teilmodell „Maschinensignale“, der schnelleren Integration von Maschinen in ein Produktionsleitsystem. Dazu müssen die Verbindungen zwischen den Eingangsvariablen der im Teilmodell „Maschinensignale“ aufgeführten Prozessgrößen zu den Signalen aus dem AIMC-Teilmodell gesetzt werden. Die Entkopplung der Variablen hat den Vorteil, dass z. B. eine Änderung an einem Namen einer Variablen in einem der Teilmodelle keine Auswirkungen auf das andere Teilmodell hat. Das gleiche gilt für das Teilmodell „Maschinenkennzahlen“, dass in sich geschlossen anwendbar ist und in den Referenzen zu den Werten für die Eingangsvariablen gelegt werden.

Bezogen auf die vielen Informationsquellen zur Berechnung der Kennzahlen ist dies von Vorteil. Wendet man konsequent die definierten Teilmodell-Templates an, können sogar die Referenzen aus den Teilmodellen teilweise statisch festgelegt werden, sodass bei einem Wechsel der angehängten Informationsquelle (Maschine) keine Änderungen im Produktionsleitsystem notwendig sind. Das ist aber nur dann möglich, wenn die Teilmodelle in ihrer vorgegebenen Struktur genutzt werden, da die Referenzen auch relative Pfade beinhalten können. Am Beispiel des Teilmodells „Maschinensignale" ist der Pfad zu einem Wert in einer fest vorgegebenen Prozessgröße immer gleich aufgebaut, egal von welcher Maschine Daten gelesen werden sollen.
Da sich das Konzept derzeit in Entwicklung befindet und die Implementierung und Validierung der Teilmodelle erst noch im Anfangsstadium ist, kann sich der Aufbau der selbst entworfenen Teilmodelle noch verändern. Trotzdem soll diese Arbeit anderen Lesern dienen, die im Begriff sind, selbst solche Teilmodelle zu erstellen. Dadurch entsteht die Möglichkeit, projektübergreifend einen gemeinsamen Konsens zu finden, der für eine spätere Spezifizierung der Teilmodelle bei der IDTA von Vorteil ist.

Literaturverzeichnis

[Ba23] BaSyx .NET SDK-Repository, https://github.com/eclipse-basyx/basyx-dotnet, 04.09.2023.

[ID22a] Industrial Digital Twin Association (IDTA): IDTA 02003-1-2 Generic Frame for Technical Data for Industrial Equipment in Manufacturing. Submodel Template of the Asset Administration Shell, Version 1.2, 2022.

[ID22b] Industrial Digital Twin Association (IDTA): IDTA 02006-2-0 Digital Nameplate for Industrial Equipment. Submodel Template of the Asset Administration Shell, Version 2.0, 2022.

[ID23a] AAS Submodel Templates, https://industrialdigitaltwin.org/content-hub/teilmodelle, 30.08.2023

[ID23b] Industrial Digital Twin Association (IDTA): IDTA 02004-1-2 Handover Documentation. Submodel Template of the Asset Administration Shell, Version 1.2, 2023.

[ID23c] IDTA: AASX Package Explorer-Repository. https://github.com/admin-shell-io/aasx-package-explorer, abgerufen am 04.09.2023.

[ISOa] ISO/TC 184/SC 5: ISO/DIS 16400-1 - Automation systems and integration - Equipment behaviour catalogues for virtual production system. Part 1: Overview, Edition 1.

[ISOb] ISO/TC 184/SC 5: ISO/DIS 16400-2 - Automation systems and integration - Equipment behaviour catalogues for virtual production system. Part 2: Formal description of a catalogue template, Edition 1.

[ISOc] ISO/TC 184/SC 5: ISO/AWI 16400-5 - Automation systems and integration - Equipment behaviour catalogues for virtual production system. Part 5: Integration of EBC templates in production system design and operation, Edition 1.

[KMK23] Kaebisch, S.; McCool M.; Korkan E.: Web of Things (WoT) Thing Description 1.1. World Wide Web Consortium, Camebridge, 2023.

[OKD23] K. O.; Kaebisch, S.; Diedrich, C.: Integration Concept of PROFIBUS and PROFINET devices into Asset Administration Shell using Asset Interface Description submodel. In (VDI Wissensforum GmbH): AUTOMATION 2023. VDI Verlag GmbH, Düsseldorf, S. 259-278, 2023.

[PI21] Plattform Industrie 4.0: Details of the Asset Administration Shell - Part 2 - Interoperability at Runtime – Exchanging Information via Application Programming Interfaces (Version 1.0RC02). Federal Ministry for Economic Affairs and Climate Action (BMWK), 23.11.2021.

[PI22] Plattform Industrie 4.0: Details of the Asset Administration Shell - Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC02). Federal Ministry for Economic Affairs and Climate Action (BMWK), 30.05.2022.

[SS14] Schließmann, A.; Strahlberger, M.: Kennzahlen und Prozessgrößen wohl definiert: MES-Kennzahlen basierend auf Maschinenzuständen. Zeitschrift für wirtschaftlichen Fabrikbetrieb, Band 109 Heft 7-8, S. 549-551, 2014.

[Tw23] TwinMaP, www.twinmap.de, 23.08.2023.

[VDM09] Verband Deutscher Maschinen- und Anlagenbau e.V. (VDMA): VDMA Einheitsblatt 66412-1 - Manufacturing Execution Systems (MES). Kennzahlen, 2009.

[VDM15] Verband Deutscher Maschinen- und Anlagenbau e.V. (VDMA): VDMA Einheitsblatt 66412-10 - Manufacturing Execution Systems (MES). Daten für Fertigungskennzahlen, 2015.

(Stand der Autorinnen- und Autorenrichtlinien: Mai 2023)

http://dx.doi.org/10.25673/111743

Automatic generation and orchestration of active Asset Administration Shells with IO-Link

Benjamin Evans[1], Victor Chavez[2], Jörg Wollert[3]

Abstract: This paper presents a proof of concept for automatically generating and orchestrating active asset administration shells (AAS) with IO-Link. AAS are software-based representations of physical assets that enable interoperability and standardised communication across different industrial systems. IO-Link is a widely adopted communication protocol for sensors and actuators in industrial automation. Our method uses an approach to generate AASs based on the IO-Link device description files. The generated AASs can then be orchestrated to form a distributed system that provides dynamic information about the status and performance of the connected assets. We demonstrate the effectiveness of our method through a proof of concept that involves the automatic generation and orchestration of AASs for a fluid processing unit equipped with pressure and flow sensors and a pump. The results show that our approach reduces the time and effort required to create and maintain active AASs.

Keywords: Asset Administration Shells; IO-Link; Industry 4.0

1 Introduction

The onset of the fourth industrial revolution has led to the emergence of smart factories, where intelligent field devices generate vast amounts of data that can be used for tasks such as monitoring and predictive maintenance. However, this technological advancement has also brought about a surge in interfaces, data streams, and documentation, posing navigation challenges for companies and their engineers [Gr20],[Re17]. Questions about which devices are installed where, which manuals correspond to specific devices, and other similar concerns have become daunting challenges in the modern industrial landscape.

In response to these challenges, the concept of digital twins has emerged as a promising solution[Kr18]. These digital replicas of physical assets act as repositories for a wealth of data and documents that might otherwise be scattered and elusive. While digital twins have shown great potential, there is a growing recognition that a standardised approach is needed to harness their full benefits. This is where the Asset Administration Shell (AAS) comes into play, representing not only a shift towards standardisation but a transformative leap for modern industrial devices in their entirety.

[1] FH Aachen, Institut für angewandte Automation, Goethe Straße 1, 52064 Aachen, Deutschland benjamin.evans@alumni.fh-aachen.de

[2] FH Aachen, Institut für angewandte Automation, Goethe Straße 1, 52064 Aachen, Deutschland chavez-bermudez@fh-aachen.de

[3] FH Aachen, Institut für angewandte Automation, Goethe Straße 1, 52064 Aachen, Deutschland wollert@.fh-aachen.de

AASs offer a framework that allow companies to employ consistent file standards and communication platforms across various field devices, fostering seamless integration and management. However, the challenge lies in the creation and configuration of AASs to efficiently gather and encapsulate this critical information.

This paper is dedicated to addressing this fundamental challenge and presents a proof of concept that demonstrates a solution based on IO-Link. We delve into the intricacies of AAS creation and setup, showcasing a practical approach to streamline the management of digital twins for contemporary industrial devices. In doing so, we aim to pave the way for more efficient, standardised, and interoperable industrial systems in the era of Industry 4.0.

2 Background

AASs have gained prominence in the German industrial landscape, promoted as the standard for digital twins by Plattform Industrie 4.0. This initiative, backed by Germany's Ministry of Economic Affairs and Climate Action and Ministry of Education and Research, has set the stage for AAS adoption [PI23].

The fundamental concept behind AASs is to function as digital repositories, encapsulating all the data associated with the physical assets they represent. This is accomplished through a hierarchical structure known as "submodels", which serve as the internal organisational system for AASs. Each submodel comprises various submodel elements, such as the asset's product name, manufacturer details, or dynamic operational parameters. The composition of elements within a submodel is specific to the submodel in question. While there are no rigid rules governing the structure of a submodel or the submodels an AASs must encompass, best practices often involve utilising the templates provided by the "Details of the Asset Administration Shell Part 1"[Ba22] or those offered by the Industrial Digital Twin Association [IT23] on their GitHub page.

This constitutes the core essence of an AAS; however, discussions surrounding AASs often imply the existence of an AAS runtime environment or server.

2.1 Different types of AASs

AASs can be split into three different groups, as described in [Sc22]:

- Type 1 (Passive AAS): AAS and submodels are serialised files
- Type 2 (Reactive AAS): AAS is a runtime instance accessible via a standardised API
- Type 3 (Proactive AAS): AAS has the ability to actively negotiate with other AASs

It is important to note that each successive type encompasses the functionalities of the preceding types. This research project primarily concerns itself with the generation of basic reactive AASs, necessitating serialised AASs and supplementary runtime instances to facilitate this process.

It's worth noting that there is still limited research in the domain of fully automated generation and deployment of reactive AASs. This area warrants further exploration and investigation.

3 Related Work

Asset Administration Shells (AASs) have witnessed a surge in research interest in recent years. These research efforts span various facets of AAS functionalities, with a substantial number of researchers directing their focus towards Proactive AASs and the intricate digital infrastructures essential for enabling such use-cases [SLl22], [Ye21].

A noteworthy trend is the widespread adoption of the AASX Package Explorer [AP23] as a tool for creating AASs in research projects. This tool has found application in multiple publications such as those by Pribiš et al., Sakurada et al. and Ocker et al. [PBD21], [SLl22], [Oc21]. It serves as a practical solution when the scope of a research project involves the creation of a limited number of AASs. However, this approach proves less suitable for industrial applications, where the demand for AAS generation is substantial. An ideal solution lies in the automated generation of the requisite AASs. Nonetheless, this task presents significant complexity beyond initial expectations.

Initial proposals have surfaced, such as the concept suggested by Miny et al., which explores automated AAS creation through domain-specific language elements [Mi20]. Additionally, Xia et al. have developed a method for generating AASs using neural language models and semantics [XJW22]. This approach, leveraging neural language models, demonstrates promising initial results, although it can encounter challenges related to data mappings, it exhibits versatility beyond specific communication protocols. Moreover, research has been conducted on the conversion of AutomationML to AAS serialisation, as outlined by Lüder et al. [Lü20]. Nevertheless, this approach necessitates the presence of asset information in an AutomationML format before serialisation, rendering it less suitable for a "plug and produce" scenario.

4 Methods

Our research highlighted a practical challenge regarding AASs in industrial adoption. Manual AAS creation using tools like the AASX Package Explorer, while valuable for beginners and understanding the AAS structure, is not feasible for manufacturers or engineers integrating AASs into production lines. The required technical knowledge is substantial.

To address this challenge, we initiated the development of an automated AAS creation tool designed specifically for IO-Link sensors. This tool simplifies the AAS creation process and ensures continuous updates of dynamic sensor information within the AAS. The following sections outline the methods employed to achieve these objectives.

5 Architecture Overview

An overview of the components and their connections within this architecture is depicted in Fig. 1. The architecture employed comprises distinct modules, each being essential to facilitate the automatic AAS generation. The process begins with the generation of a passive AAS, followed by the transmission of operational data to the newly deployed AAS runtime instance.

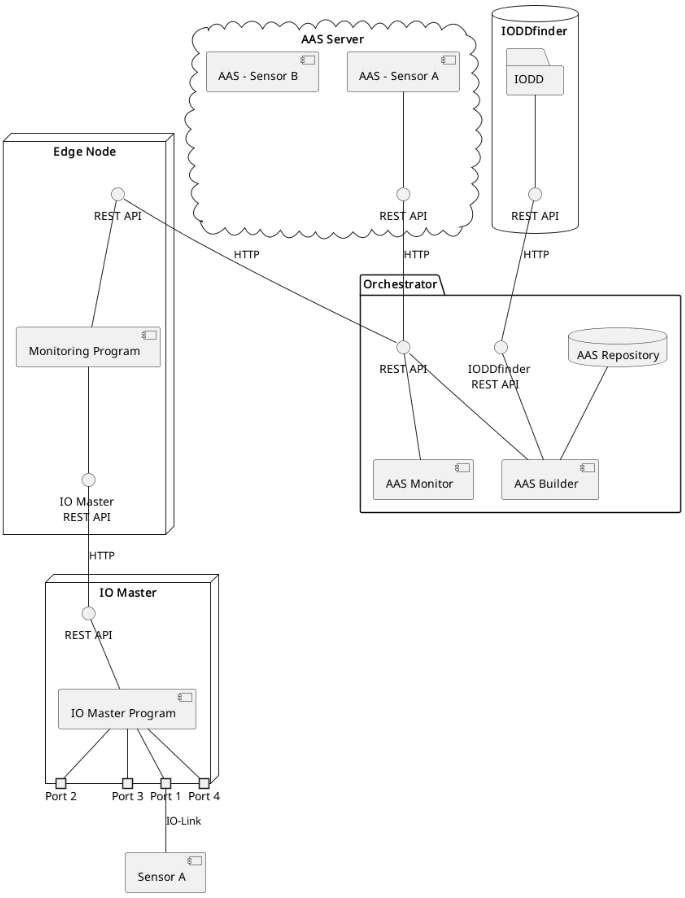


Fig. 1: Component Diagram of presented architecture

5.1 Generation of the passive AAS

The cornerstone of AAS generation in this architecture is the Edge Node, alternatively referred to as an Edge Adapter. Its primary role is to establish the communication protocol between the asset and the broader AAS infrastructure. The specific Edge Node used is determined by the asset's communication protocol. In this project, we designed the Edge Node to work with the IO-Link protocol.

It's important to note that while passive AASs, represented as Type 1, do not require direct communication with the asset, an Edge Node becomes necessary from Type 2 onwards. This holds true unless the sensor itself offers the requisite interface directly to the AAS runtime instance. In our architecture, the Edge Node's significance extends to the generation of passive AASs. This is accomplished by harnessing the information provided by IO-Link's I/O Device Descriptions (IODDs). In the initial phase of the program, the first task is to acquire information about the status of the IO-Master's ports. Once the Edge Node determines which ports are in use by responsive devices, it proceeds to conduct individual polls on these devices. The polled information includes details such as the device manufacturer, product ID, serial number, hardware revision, and firmware revision. This data, in conjunction with the IODD, gathered by the orchestrator, is essential for the creation of an AAS Nameplate submodel. This communication can also be seen in Fig. 2.

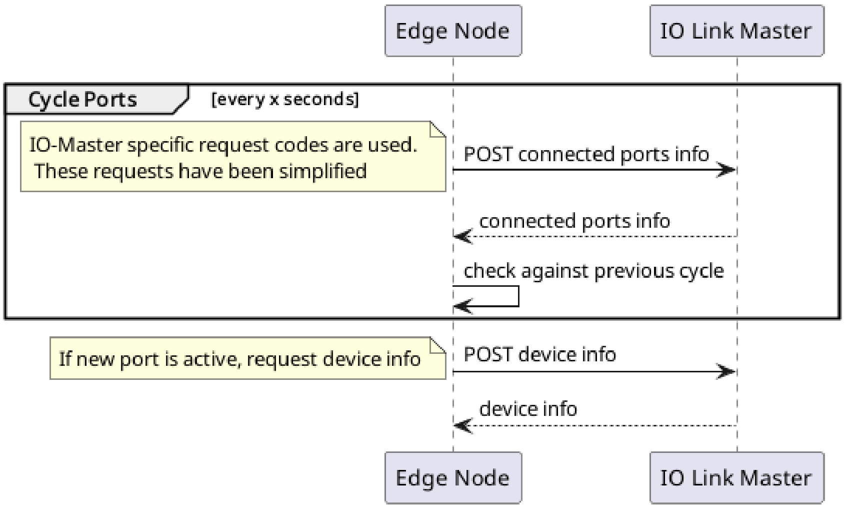

Edge Node
IO Link Master
Cycle Ports
[every x seconds]
IO-Master specific request codes are used.
These requests have been simplified
POST connected ports info
connected ports info
check against previous cycle
If new port is active, request device info
POST device info
device info
Edge Node
IO Link Master

Fig. 2: Communication with IO-Link master

Following the collection of device data, this information is transmitted to the orchestrator via its REST API.

While this approach is tailored for IO-Link, it's essential to highlight its adaptability. The Edge Node used here, specifically designed for IO-Link data transmission, can be substituted with an Edge Node suited to alternative communication protocols for transmitting the asset's operational data. However, it's crucial to note that the creation of a passive AAS for the asset requires the provision of a device profile. Proceeding to the next step, the orchestrator initiates a request for the IODD files from IODDfinder, an online IODD database run by the IO-Link community. Subsequently, the orchestrator cross-references the information obtained from the Edge Node with its locally stored lookup table of known assets. Fig. 3 shows this communication.

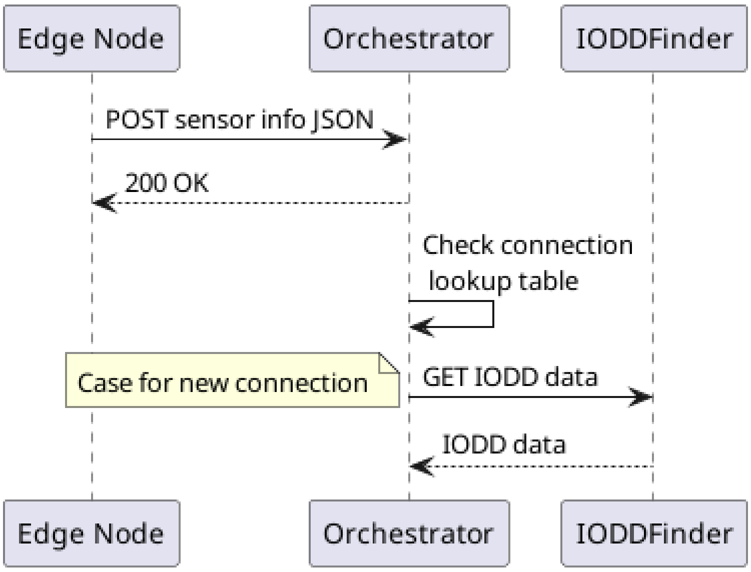


Fig. 3: Retrieval of IODD

If the device has previously undergone AAS generation, and its nameplate information is already available in the orchestrator's records, there is no need to retrieve this information from the IODD. In such cases, the orchestrator launches an AAS runtime instance (AASX Server docker image [AS23]) using the corresponding AASX file stored in its repository. However, if the orchestrator does not find an entry for the asset in question, it proceeds to request the necessary IODD files via the IO-Link IODDfinder API. Subsequently, the orchestrator parses these IODD files to extract the required information. This step can be seen in Fig. 4.

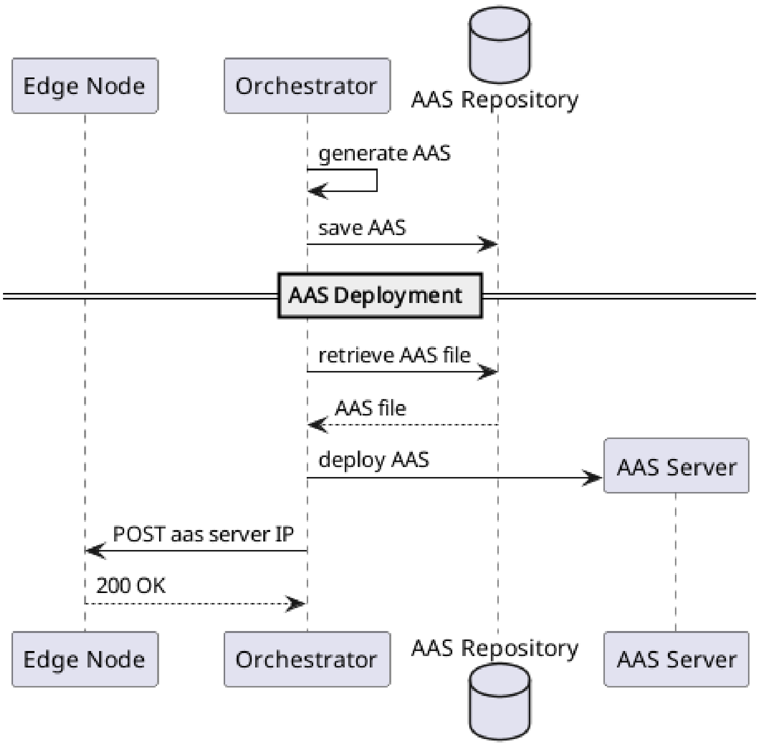


Fig. 4: Generation and deployment of AAS Server

For a comprehensive list of submodel elements used in the Nameplate submodel, along with their corresponding IODD variables or sensor parameters, please refer to Tab. 1.

AAS Nameplate submodel element	IO-Link[a]	
idShort	IODD attribute	Index Service Data Unit[b]
URIOfTheProduct	-	Index 21, Subindex 0 (Serial No.) & Index 19, Subindex 0 (Product Id)
ManufacturerName	DeviceIdentity •vendorName	Index 16, Subindex 0
ManufacturerProductDescription	DeviceVariantCollection •DeviceVariant •Description	-
ContactInformation	-	-
ManufacturerProductRoot	DeviceVariantCollection •DeviceVariant •Description	-
ManufacturerProductFamily	DeviceIdentity •DeviceFamily	-
ManufacturerProductType	DeviceIdentity •DeviceName	Index 19, Subindex 0
OrderCodeOfManufacturer	-	-
ProductArticleNumberOfManufacturer	DeviceIdentity •DeviceName	Index 19, Subindex 0
SerialNumber	-	Index 21, Subindex 0
YearOfConstruction	-	-
DateOfManufacture	-	-
HardwareRevision	-	Index 22, Subindex 0
FirmwareRevison	-	Index 23, Subindex 0
SoftwareVersion	-	-
CountryOfOrigin	-	-
CompanyLogo	DeviceIdentity •VendorLogo	-
Markings	-	-
AssetSpecificProperties	DeviceFunction	-

[a] For IODD specification v1.1, other IODD specifications are also viable

[b] Can be dynamically retrieved from the IO-Link device itself

Tab. 1: IO-Link to AAS mapping ("-" stands for no information available)

5.2 Transmission of operational data

Once the orchestrator has assembled the requisite information, generated the AAS, and deployed a runtime instance, it proceeds to notify the asset's Edge Node about the IP address of the AAS runtime. This information enables the asset's Edge Node to establish direct communication with the AAS runtime's API.

This step is pivotal as it facilitates seamless communication between the Edge Node and the AAS, bypassing the orchestrator. In larger systems, this approach prevents the orchestrator from becoming a potential bottleneck. Subsequently, the Edge Node periodically updates the operational data from the asset via the IO-Master and directly transmits this data to the AAS runtime instance using its REST API. This communication cycle can be seen in Fig. 5.

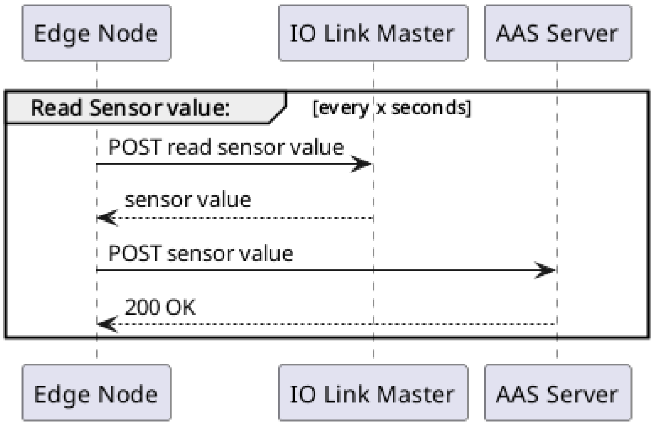


Fig. 5: Dynamic asset data update cycle

6 Validation

To evaluate the effectiveness of the proof of concept, we conducted tests on a sensor testbed equipped with multiple IO-Link sensors. Fig. 6 shows the testbed used. The testbed served as a comprehensive platform for assessing the capabilities of the presented architecture with a diverse array of sensors. Our proposed architecture demonstrated its ability to seamlessly communicate with all sensors within the testbed. Moreover, it successfully generated AASs with Nameplate submodels for these sensors. For a visual representation of one such AAS, please refer to Fig. 7

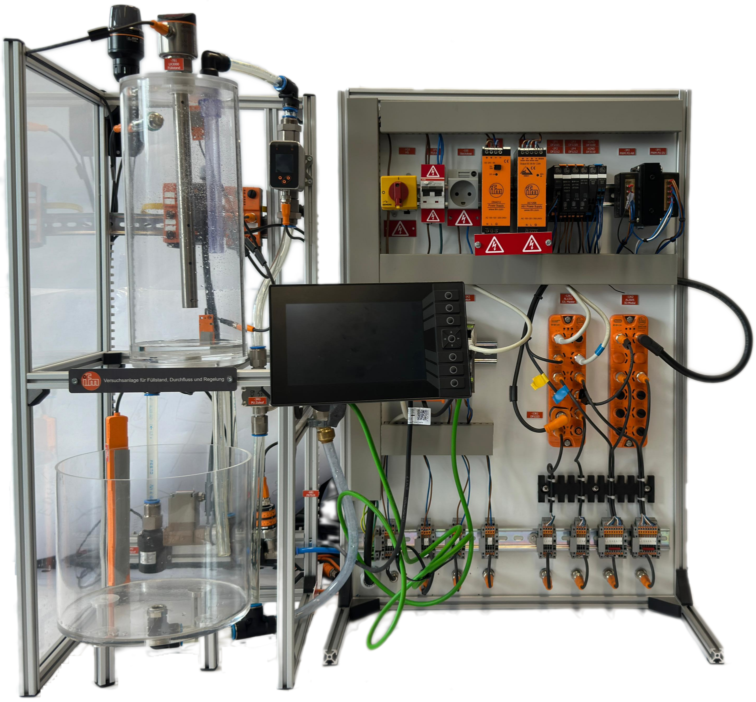
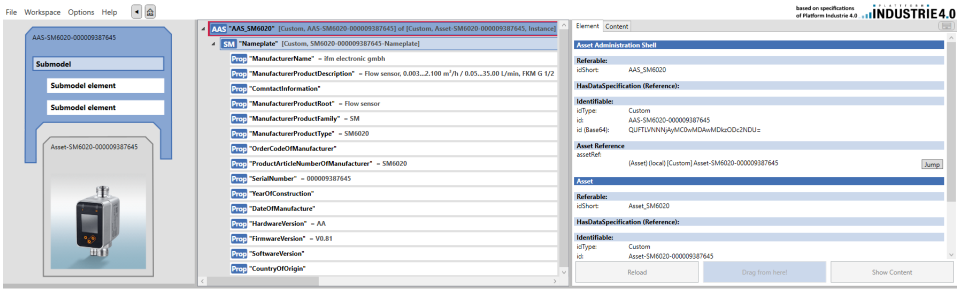


Fig. 6: Sensor Testbed



AAS-SM6020-000009387645

Submodel

Submodel element

Submodel element

Asset-SM6020-000009387645

AAS "AAS_SM6020" [Custom, AAS-SM6020-000009387645] of [Custom, Asset-SM6020-000009387645, Instance]

SM "Nameplate" [Custom, SM6020-000009387645-Nameplate]

- Prop "ManufacturerName" = ifm electronic gmbh
- Prop "ManufacturerProductDescription" = Flow sensor, 0.003...2.100 m³/h / 0.05...35.00 L/min, FKM G 1/2
- Prop "ComntactInformation"
- Prop "ManufacturerProductRoot" = Flow sensor
- Prop "ManufacturerProductFamily" = SM
- Prop "ManufacturerProductType" = SM6020
- Prop "OrderCodeOfManufacturer"
- Prop "ProductArticleNumberOfManufacturer" = SM6020
- Prop "SerialNumber" = 000009387645
- Prop "YearOfConstruction"
- Prop "DateOfManufacture"
- Prop "HardwareVersion" = AA
- Prop "FirmwareVersion" = V0.81
- Prop "SoftwareVersion"
- Prop "CountryOfOrigin"

Asset Administration Shell

Referable:
idShort: AAS_SM6020

HasDataSpecification (Reference):

Identifiable:
idType: Custom
id: AAS-SM6020-000009387645
id (Base64): QUFTLVNNNjAyMC0wMDAwMDkzODc2NDU=

Asset Reference
assetRef:
(Asset) (local) [Custom] Asset-SM6020-000009387645

Asset

Referable:
idShort: Asset_SM6020

HasDataSpecification (Reference):

Identifiable:
idType: Custom
id: Asset-SM6020-000009387645

Fig. 7: Generated AAS from the Testbed opened in the AASX Package Explorer

7 Discussion & Outlook

The architecture presented in this study exhibits promise for automating the creation of digital nameplates within AASs. This capability could prove increasingly valuable in the context of potential future requirements for digital product passports (DPPs) [EU23], which may become mandatory for manufacturers in the future. Initiatives are already underway to establish AASs as the de-facto standard for DPPs [DP23], and manual AAS generation is unlikely to remain a feasible option. However, it's crucial to acknowledge that this architecture has certain limitations. One notable limitation is the reliance of edge nodes on existing device descriptions for assets, which are subsequently reformatted into a compatible AAS structure. As detailed in Tab. 1, there are instances where certain AAS nameplate submodel elements lack equivalents within the IODD or ISDU. This limitation is likely to extend to other communication protocols not specifically designed with AASs in mind. Nonetheless, this architecture excels in delivering a true "plug and produce" functionality, demonstrated by its capacity to seamlessly generate basic reactive AASs.

Bibliography

[AP23] Industrial Digital Twin Association e. V., 2023, URL: https://github.com/admin-shell-io/aasx-package-explorer.

[AS23] Industrial Digital Twin Association e. V., 2023, URL: https://github.com/admin-shell-io/aasx-server.

[Ba22] Bader, S.; Barnstedt, E.; Bedenbender, H.; Berres, B.; Billmann, M.; Boss, B.; Braunisch, N.; Braunmandl, A.; Clauer, E.; Diedrich, C.; Flubacher, B.; Fritsche, W.; Garrels, K.; Gatterburg, A.; Hankel, M.; Heppner, S.; Hoffmeister, M.; Jänicke, L.; Jochem, M.; Ziesche, C.: Details of the Asset Administration Shell. Part 1 -The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC02), Mai 2022.

[DP23] Industrial Digital Twin Association e. V., ZVEI e. V., 2023, URL: https://dpp40.eu/.

[EU23] European Health and Digital Executive Agency (HaDEA), 2023, URL: https://hadea.ec.europa.eu/calls-proposals/digital-product-passport_en.

[Gr20] Grabowska, S.: Smart Factories in the Age of Industry 4.0. Management Systems in Production Engineering 28/2, S. 90–96, 2020, URL: https://doi.org/10.2478/mspe-2020-0014.

[IT23] Industrial Digital Twin Association e. V., 2023, URL: https://industrialdigitaltwin.org/.

[Kr18] Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W.: Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine 51/11, 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018, S. 1016–1022, 2018, ISSN: 2405-8963, URL: https://www.sciencedirect.com/science/article/pii/S2405896318316021.

[Lü20] Lüder, A.; Behnert, A.-K.; Rinker, F.; Biffl, S.: Generating Industry 4.0 Asset Administration Shells with Data from Engineering Data Logistics. In: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). Bd. 1, S. 867–874, 2020.

[Mi20] Miny, T.; Thies, M.; Epple, U.; Diedrich, C.: Model Transformation for Asset Administration Shells. In. S. 2207–2212, Okt. 2020.

[Oc21] Ocker, F.; Urban, C.; Vogel-Heuser, B.; Diedrich, C.: Leveraging the Asset Administration Shell for Agent-Based Production Systems. IFAC-PapersOnLine 54/1, 17th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2021, S. 837–844, 2021, ISSN: 2405-8963, URL: https://www.sciencedirect.com/science/article/pii/S2405896321009563.

[PBD21] Pribiš, R.; Beňo, L.; Drahoš, P.: Asset Administration Shell Design Methodology Using Embedded OPC Unified Architecture Server. Electronics 10/20, 2021, ISSN: 2079-9292, URL: https://www.mdpi.com/2079-9292/10/20/2520.

[PI23] Plattform Industrie 4.0, 2023, URL: https://www.plattform-i40.de.

[Re17] Ren, G.; Hua, Q.; Deng, P.; Yang, C.; Zhang, J.: A Multi-Perspective Method for Analysis of Cooperative Behaviors Among Industrial Devices of Smart Factory. IEEE Access 5/, S. 10882–10891, 2017.

[Sc22] Schnicke, F.; Kuhn, T.; Klausmann, T.; Grüner, S.; Porta, D.: Architecture Blueprints for the Application of the Industry 4.0 Asset Administration Shell. In. S. 1–8, 2022.

[SLl22] Sakurada, L.; Leitao, P.; la Prieta, F. D.: Agent-Based Asset Administration Shell Approach for Digitizing Industrial Assets. IFAC-PapersOnLine 55/2, 14th IFAC Workshop on Intelligent Manufacturing Systems IMS 2022, S. 193–198, 2022, ISSN: 2405-8963, URL: https://www.sciencedirect.com/science/article/pii/S2405896322001938.

[XJW22] Xia, Y.; Jazdi, N.; Weyrich, M.: Automated generation of Asset Administration Shell: a transfer learning approach with neural language model and semantic fingerprints. In: 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA). S. 1–4, 2022.

[Ye21] Ye, X.; Hong, S. H.; Song, W. S.; Kim, Y. C.; Zhang, X.: An Industry 4.0 Asset Administration Shell-Enabled Digital Solution for Robot-Based Manufacturing Systems. IEEE Access 9/, S. 154448–154459, 2021.

http://dx.doi.org/10.25673/111744

Intralogistics application with a fleet of robots on a private 5G campus network

Marc Kalter[1], Dennis Karbach[1], Christian Schellenberger[1], Eric Schöneberg[2], Axel Vierling[3], Hans D. Schotten[1], Daniel Görges[2], Karsten Berns[3]

Abstract: This paper presents the concept and the current state of implementation of a semi-autonomous robot fleet for logistics applications in a campus environment. The communication is realized via a private 5G SA network. The robot fleet performs its logistics tasks semi-autonomously on campus and can deliver mail or parcels, for example. Sensor data (GPS, camera images, 2D and 3D laser scanners, ...) is sent to a central computing unit in the control center via the 5G interface to analyze and store live data and influence the robot's actions at real time to save costs of the robot and conserve energy to increase operating time. The operator in the control center can intervene in unusual situations at any time and remotely control the robots via 5G. The described system is being tested with a fixed private 5G SA network and a nomadic 5G SA network as public cellular networks are not performant enough in regards to low latency and upload bandwidth. The nomadic network approach opens up further application scenarios such as company premises or events. The system has so far been built and tested with one robot. The expansion of the robot fleet with different platforms is currently in progress.

Keywords: 5G, controlcenter, intralogistics, private network, robot fleet, teleoperator, ROS2

1 Introduction

Logistics tasks arise in many areas of industry and also in the private sector. Size, weight and geometry of the transported goods as well as environment, time schedule and safety makes those tasks highly heterogeneous. The tasks can be very varied in size, weight and geometry of the object to be transported. The completion of these tasks is time-consuming and expensive. If chosen carefully, some tasks can be structured and reduced in complexity to be automated. Due to their clear structure and low complexity of the actual task, they can be automated. In a closed system, such as a fully automated warehouse, this is already state of the art. Structurally more complex environments - e.g. including interaction with people and in heterogeneous environments - automation becomes more complex. In more complex environments and in environments with interaction with people the implementation is much more complicated. The presented concept of a semi-automated logistics system in combination with a private 5G network tries to close this gap. The diversity and complexity created by such an environment are reflected in the concept software requirements, as well

[1] RPTU Kaiserslautern-Landau, Institute for Wireless Communication and Navigation

[2] RPTU Kaiserslautern-Landau, Division of Electromobility

[3] RPTU Kaiserslautern-Landau, Robotics Research Lab

as the radio communication requirements. Both aspects are explained and illustrated by measurements below.

2 5G and Intralogistics

A semi-autonomous robot fleet that can be centrally managed by one operator places new demands on communication technology. Common non-cellular wireless communication technologies like WiFi or LoRaWAN are not able to fulfill some of the rising demands. These non-cellular wireless communication technologies are still predominant especially in the industrial environment [3]. However, none of them can cover all the aspects that are required to be able to map a multidimensional use case such as this one due to coverage, throughput and latency deficiencies. The deployment of such a fleet of robots raises questions about communications technology.

Can 5G cover the requirements for:

- Remote Control of the Robot?
- Semi-Autonomous Driving?
- Outsourcing Sensor Evaluation?

As an example, uploading 3D point clouds can require a data rate of up to 1 Gbps per robot [5]. These and other questions regarding 5G communications will be addressed and investigated with the use case.

3 System Description

Figure 1 shows the schematic overview of the intralogistics concept where the robot generally drives in autonomous mode and can directly communicate with infrastructure like doors and elevators. Logistics orders can be created and managed by the operator from the control station. The robot performs its logistics tasks semi-autonomously, i.e. it is able to independently plan and travel its route to the destination on campus. In its typical operation

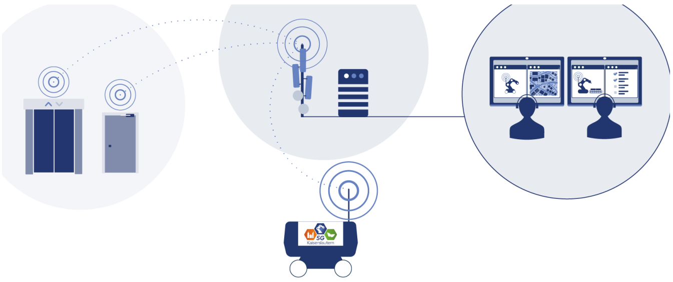


Fig. 1: Schematic System Overview

mode the AGV completes its tasks autonomously and awaits the next order without the operator having to intervene. In the event of an incident (unknown obstacles, roadblock, critical sensor values, etc.), the robot reports to the control station and passes the decision on how to proceed to the operator. If the situation has been bypassed or is not critical, the operator can switch back to autonomous operation. This concept allows the operator to manage and monitor the complete robot fleet. The robot is also able to communicate with the campus infrastructure via 5G and, for example, call the elevator and send it to the desired floor.

3.1 System Architecture

Components of the demonstrated use case are the robot fleet, the control center and the campus infrastructure. All components communicate via the 5G network and make the data available in a centralized data base which is used for a central dashboard in the control center. The individual components are presented in this section.

3.2 Private 5G Network

The campus of RPTU and the area of the adjacent research institutions in Kaiserslautern are covered by five outdoor radio sites. Each site consists of one radio head split into two 2x2 sectors for optimal coverage. The radio heads are connected to the central base band unit (BBU) centrally located in the data center of the university. The BBU is connected to the core whose local breakout is directly connected to the control center hardware and edge cloud for the sensor data offloading. Compared to public cellular networks private 5G networks can be adapted to fit a lot of different applications. Cellular networks have the advantage of predictable network access times through centralized medium access control instead of best effort medium control in Wi-Fi and can therefore achieve lower and more reliable latencies [7]. Using dedicated spectrum, cellular networks do not suffer from interference from surrounding networks and have greater coverage area with one base station. The drawback of cellular networks is higher capex and opex costs compared to enterprise Wi-Fi. Private cellular networks provide high configurability compared to public cellular networks, but require highly trained staff to operate. Table 1 shows the performance of different cellular networks.

	private 5G SA	public 4G	public 5G NSA
Avg. throughput Downlink	~ 700 Mbit/s	~ 90 Mbit/s	~ 240 Mbit/s
Avg. throughput Uplink	~ 300 Mbit/s	~ 18 Mbit/s	~ 110 Mbit/s
Latency	~ 10 ms	~ 25 ms	~ 20 ms

Tab. 1: Performance Data of Cellular Networks [7]

3.3 Intralogistics Robot Hardware

The general idea is a heterogeneous fleet of robots that is adapted to the logistics task at hand. A common carrier platform and different body types for the individual tasks are also conceivable. The robot already built consists of the base (drive unit), the outer webs with sensors and a rugged, industrial computer, and the exchangeable transport box. This transport box is independent of the robot and has its own 5G interface, display and GPS module. The robot is equipped with a comprehensive sensor package, which enables it to operate on a busy site such as a university campus. The sensor package includes safety-relevant sensor technology for the robot and its environment such as the two 2D laser scanners as virtual bumpers on the corners of the robot. A 3D laser scanner to create maps of its environment, and two stereo cameras (front and rear) for remote or autonomous operation. Two easily accessible emergency stop switches on the body can switch off the drive locally, in addition to software emergency stop routines.

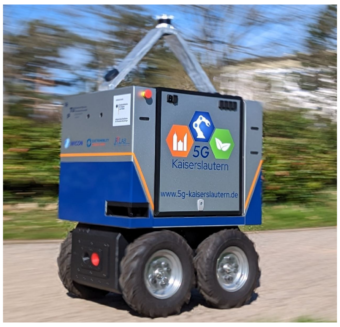


Fig. 2: Intralogistic Robot at the RPTU

Figure 3 gives a schematic overview of the internal hardware connections. The Industrial Computer serves as the main computational unit, running open source robotics software as described in details in Section 3.4. Consequently, all sensor, communication and actuation interfaces must finally converge to it on this higher software level. 5G communication with the private network as described in Section 2 is enabled by a Quectel RM500Q module. Low level software in actuation and sensors is handled by a Pixhawk 4, which again connects to Sabertooth motor drivers, a built-in 9-degrees-of-freedom IMU, encoders and a GPS module. The high level sensors are connected to the industrial computer according to Figure 3. The Ouster OS0-128 is a high-resolution 3D LiDAR on top of the robot, which enables navigation in complex environments. Both Sick Tim561 2D LiDARs are mounted on diagonal corners of the robot. They provide safety by minimizing blind angles. ZED2 stereo cameras with a broad field of view at the front and rear of the robot enable video streaming for a remote human operator and can be included in autonomous navigation.

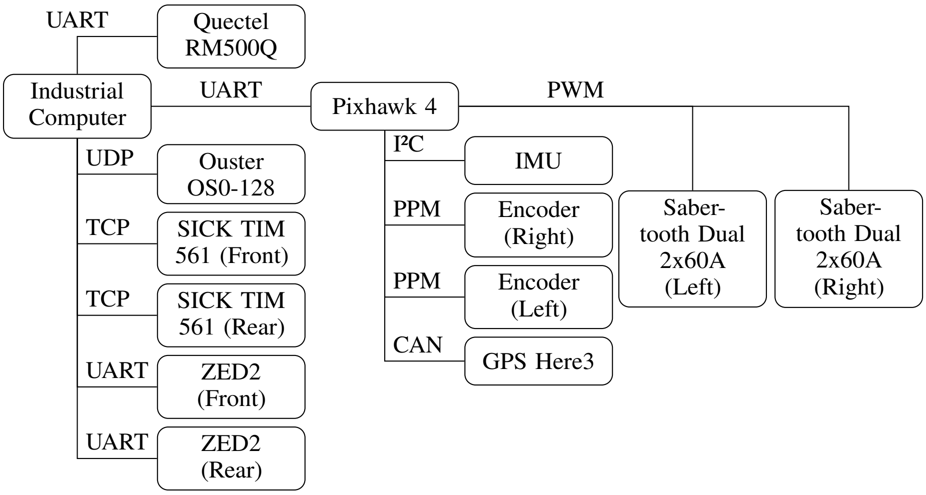


Fig. 3: Intralogistics Robot - Hardware Connection Graph

3.4 Intralogistics Robot Software

The main computational unit runs Robot Operating System 2 (ROS2) as middle-ware software. Robot Operating System is an open-source robotics middle-ware software framework. It provides a flexible distributed software architecture, utilizing Data Distribution Service (DDS) [4] real-time communication. [2]. Consequently, all sensors, communication and actuation must finally converge to the computer on this higher software level. Once data is available within the ROS network, it can be broadcasted to any other network client within the same network - i.e. via 5G.

The entire software stack can be separated into ROS nodes - executables - each communicating via DDS with other nodes forming a graph. Any node within the graph can run on any of the aforementioned network clients - provided they do not have explicit hardware connection. See Figure 3, hardware modules Pixhawk 4 and ZED2 each connect to the ROS graph with their own ROS nodes, yet use UART (USB) communication and therefore must run on the industrial computer. The Quectel RM500Q is not part of the ROS network, but provides operating system level network access. This leaves Ouster OS0-128 and Sick Tim 561 as viable choices to provide access solely via 5G - even if the sensors are physically on the same robot. Still, such architecture lacks justification. Benefits in smaller required local computational power using cloud servers, smaller power consumption or cost in hardware on the robot are insignificant.

Nonetheless, a robot powered by 5G and running ROS can be closely monitored and subject to manual intervention, allowing for adjustments to its behavior or parameters as needed. This creates opportunities for human operators to oversee and manage multiple robots simultaneously, including human emergency backup operation, remote diagnostics and error handling, remote robot operation and decision making in non-accessible or dangerous environment. As an example and simplified:

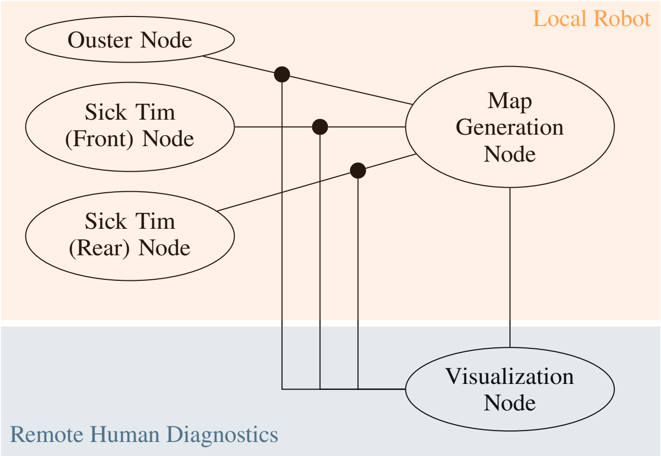


Fig. 4: ROS2 Distributed Software Architecture - Remote Human Diagnostics

Locally some sensor data (e.g. Ouster and two Sick Tim) is transferred to another ROS node which builds a map. An operator can now observe both raw sensor data and the resulting map - where all data is transferred in real-time via 5G. The resulting remote visualization can be seen in Figure 5, where grey and black color represent a global map, white dots represent a single line of Ouster's 3D pointcloud, and both light and dark orange nuances represent the front and rear Sick Tim respectively. Any node-to-node communication can be observed by an operator, where currently the software opens approximately 50 active nodes

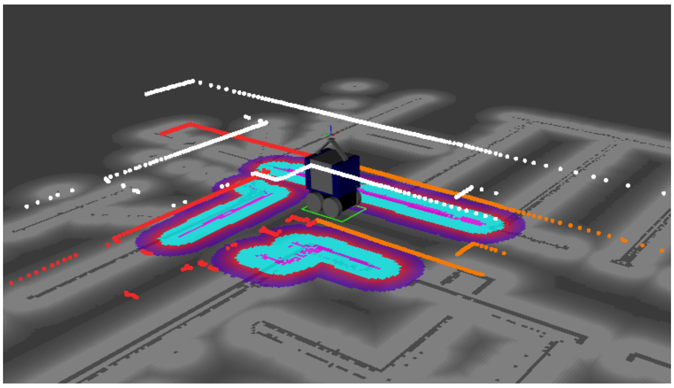


Fig. 5: Remote Diagnostics Visualization

and 100 vertices. The requirements in communication for such an architecture is discussed in details in Section 4 and Section 5.

3.5 Control Center

All data generated by and about the 5G network and the robots is systematically gathered and stored within a database, facilitating real-time analysis as well as future utilization. This data repository paves the way for a multitude of applications, such as enhancing artificial intelligence capabilities or the creation of digital twins. The collected data can be harnessed to train and improve artificial intelligence algorithms, enabling more efficient decision-making and automation. By leveraging this data, it becomes possible to create digital replicas or twins of physical systems, aiding in simulations, monitoring, and predictive maintenance. Utilizing historical and real-time data, predictive analytics can be employed to foresee potential issues, optimize operations, and streamline resource allocation. Accurate data can inform resource allocation, helping organizations allocate resources effectively and minimize costs. Real-Time data analysis allows the user to control and monitor the whole process conveniently from a central point.

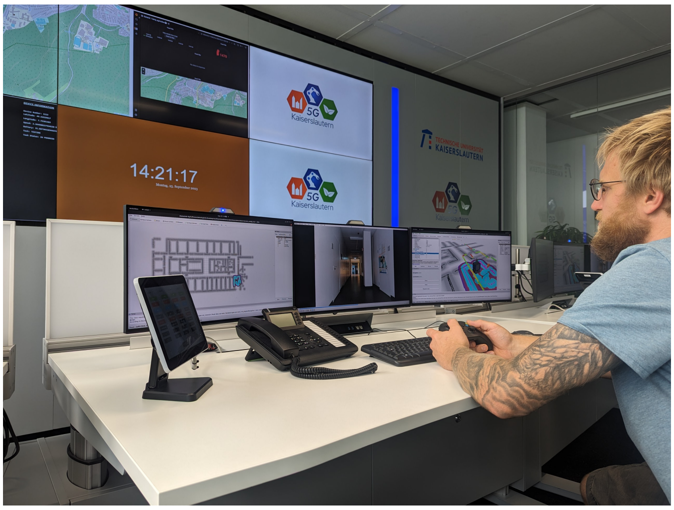


Fig. 6: Control Center with Operator at the RPTU Kaiserslautern

The operator can access any robot from the fleet at any time from his workstation and display its data separately. An overview map of the complete campus operational area including all current robot locations is provided. In the event of a detected abnormal situation, the operator can display the video streams (front and back) and all sensor data at his workstation to assess the situation and make a decision. This semi-automated process with human-in-the-loop allows constant improvement towards full automation.

4 Differentiation of 5G from other Wireless Technologies

To weigh which of the sensor data can potentially be offloaded to an edge cloud server, knowledge of the required bandwidths is crucial. Safety-relevant data, such as the Sick safety lidar, are not considered further here, since the evaluation of these must always happen on the platform. Table 2 gives an overview of the required bandwidths of the very data-intensive applications such as video and point cloud stream of the ZED and the 3D ouster lidar, where the specified data for the camera applies only to the front camera.

Pointcloud				Video-Stream			
Dots	Lines	Frequency	Data Rate	Resolution		Freuqency	Data Rate
512	128	10 Hz	62 Mbit/s	WQHD	1440p	15 fps	8.5 Mbit/s
1024	128	10 Hz	123 Mbit/s	FullHD	1080p	30 fps	12.5 Mbit/s
2048	128	10 Hz	247 Mbit/s	HD	720p	60 fps	7 Mbit/s

Tab. 2: Required bandwidth of sensor data

An overview of average throughputs and latencies in cellular networks has already been given in Table 1 in Section 2. The following table shows the average throughput and the maximal range in the various WiFi standards.

#	Standard (IEEE)	Frequency	Theoretical Data Rate	Practical Data Rate	Max. Range
4	802.11n	2,4 / 5 GHz	600 Mbit/s	300 Mbit/s	100 m
5	802.11ac	5 GHz	6936 Mbit/s	870 Mbit/s	50 m
6 / 6E	802.11ax	2,4 / 5 / 6 GHz	9608 Mbit/s	1200 Mbit/s	50 m

Tab. 3: Data throughput of WiFi [6]

The specified theoretical data rate corresponds to the calculated maximum of the data rate, taking into account all performance features provided for in the respective standard. In practical implementation, however, there are limitations due to which this data rate cannot be realized. A practical data rate that is closer to the WiFi equipment in practice is therefore more suitable for comparing WiFi standards. The specified practical data rate corresponds to the data rate that is usually possible with purchasable devices. Two antennas and a channel width of 80 MHz in the frequency range of 5 GHz are taken into account. Depending on the equipment, the value of this practical transmission rate can also be higher or lower.

Figure 7 qualitatively differentiates actual WiFi from public mobile radio standards such as 5G and LTE. Especially in use cases of autonomous driving in complex large public environments like the university campus in Kaiserslautern with adjacent research institutes show the necessity of using 5G. While LTE simply offers too little bandwidth, the coverage of WiFi is insufficient and justified by only small ranges of the access points. In addition, the bandwidth of WiFi networks decreases significantly and latencies even more, especially in busy environments with many end devices. 5G offers a variety of methods to minimize these disadvantages and to adapt the network more optimally to the needs of the respective network subscribers. Examples include carrier aggregation, the possibility of using small cells or multi-antenna systems (MIMO), variable alignment to the end device (beamforming) and virtually shared networks (networkslicing). Another key benefit of using 5G is significantly lower latency. This is particularly crucial for applications in which robots in busy environments have to be controlled remotely from the control center.

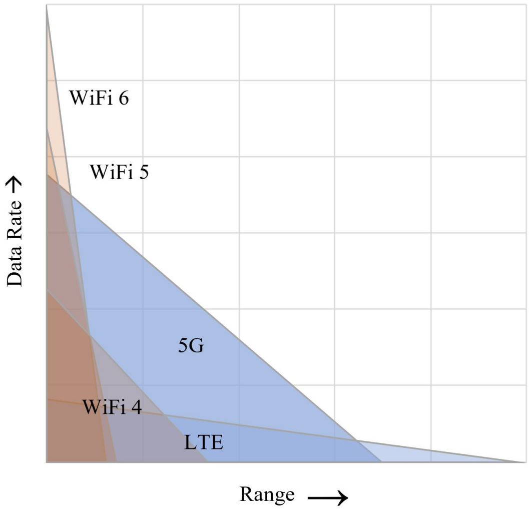


Fig. 7: Qualitative delimitation of WiFi, LTE and 5G

5 Conclusion and Future Work

The presented concept and implementation shows a possible handling of robotics applications in a busy environment like a university campus. A fully autonomous operation would be difficult both legally and technically. Nevertheless, the semi-autonomous approach with an operator in the control center allows an economical use of the robot fleet [1]. In the future, the existing robot will be joined by robots of other designs and their operation tested in terms of manageability by the operator and utilization of the 5G network.

Literature

[1] Stephan Ludwig u. a. „Reference Network and Localization Architecture for Smart Manufacturing Based on 5G“. In: *Advances in System-Integrated Intelligence: Proceedings of the 6th International Conference on System-Integrated Intelligence (SysInt 2022), September 7-9, 2022, Genova, Italy*. Springer. 2022, S. 470–479.

[2] Steven Macenski u. a. „Robot Operating System 2: Design, architecture, and uses in the wild“. In: *Science Robotics* 7.66 (2022), eabm6074. DOI: `10.1126/scirobotics.abm6074`. URL: `https://www.science.org/doi/abs/10.1126/scirobotics.abm6074`.

[3] Lara Nehrke u. a. „Survey on Usage of 5G Campus Networks in Intralogistics“. In: *Advances in System-Integrated Intelligence: Proceedings of the 6th International Conference on System-Integrated Intelligence (SysInt 2022), September 7-9, 2022, Genova, Italy*. Springer. 2022, S. 480–488.

[4] Gerardo Pardo-Castellote. „Omg data-distribution service: Architectural overview“. In: *23rd International Conference on Distributed Computing Systems Workshops, 2003. Proceedings.* IEEE. 2003, S. 200–206.

[5] Christopher Vincent Poulton u. a. „Long-Range LiDAR and Free-Space Data Communication With High-Performance Optical Phased Arrays“. In: *IEEE Journal of Selected Topics in Quantum Electronics* 25.5 (2019), S. 1–8. DOI: `10.1109/JSTQE.2019.2908555`.

[6] Adian Fatchur Rochim u. a. „Performance comparison of wireless protocol IEEE 802.11ax vs 802.11ac“. In: *2020 International Conference on Smart Technology and Applications (ICoSTA)*. 2020, S. 1–5. DOI: `10.1109/ICoSTA48221.2020.1570609404`.

[7] Christian Schellenberger u. a. *Leveraging 5G private networks, UAVs and robots to detect and combat broad-leaved dock (Rumex obtusifolius) in feed production*. 2023. arXiv: `2305.00430 [cs.RO]`.

http://dx.doi.org/10.25673/111745

A Communication Concept Using 5G for the Automated Driving Monorail Vehicle MONOCAB

Andre Bröring[1] Arne Neumann[2] Andreas Schmelter[3] Jürgen Jasperneite[4]

Abstract:

The MONOCAB is an innovative monorail vehicle designed to operate in two directions simultaneously on a single rail track. To ensure smooth operations and efficient fleet management, various communication needs arise. This paper outlines four common use cases and identifies nine communication requirements for the MONOCAB. Based on this, it presents a communication concept utilizing 5G technology, covering Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication, as well as time critical communication to an edge application in a central control centre and non-time critical communication for fleet management and provision of information for the MONOCAB users.

Keywords: 5G Communication; Railway; V2X; Remote Control; Security

1 Introduction

Increasing the attractiveness of rural areas depends to a large extent on accessible mobility services and the associated connections to surrounding regions. In addition to usability and thus acceptance, a requirement for these services is cost-effectiveness. This essentially depends on the market introduction costs and the subsequent operating and maintenance costs, especially for personnel and expensive infrastructure. These factors can be kept low by reusing existing infrastructures and reducing operating costs [Fl20].

The MONOCAB, shown in Figure 1, is an automated driving monorail vehicle that can drive on a single rail track with several vehicles simultaneously in two directions, enabling a bidirectional on-demand mobility service [GMS21]. The prerequisite for such an automated driving vehicle are suitable technological measures, which require reliable internal and external communication mechanisms. In order to avoid heavy investments in rebuilding the traditional rail communication systems used by trains with signal lights, railway crossing gates and sensors, and sensors for rail clearance signals, a proprietary communication system should be used. This should enable the communication between the MONOCABs and a central control centre, as well as between individual MONOCABs and between the MONOCABs and the infrastructure, for example at railroad crossings. Such a communication system must comply with time barriers in transmission time, allow sufficient bandwidth,

[1] Institute Industrial IT - inIT, Campusallee 6, 32657 Lemgo, Germany, andre.broering@th-owl.de
[2] Institute Industrial IT - inIT, Campusallee 6, 32657 Lemgo, Germany, arne.neumann@th-owl.de
[3] Institute Industrial IT - inIT, Campusallee 6, 32657 Lemgo, Germany, andreas.schmelter@th-owl.de
[4] Fraunhofer IOSB-INA, Campusallee 1, 32657 Lemgo, Germany, juergen.jasperneite@iosb-ina.fraunhofer.de

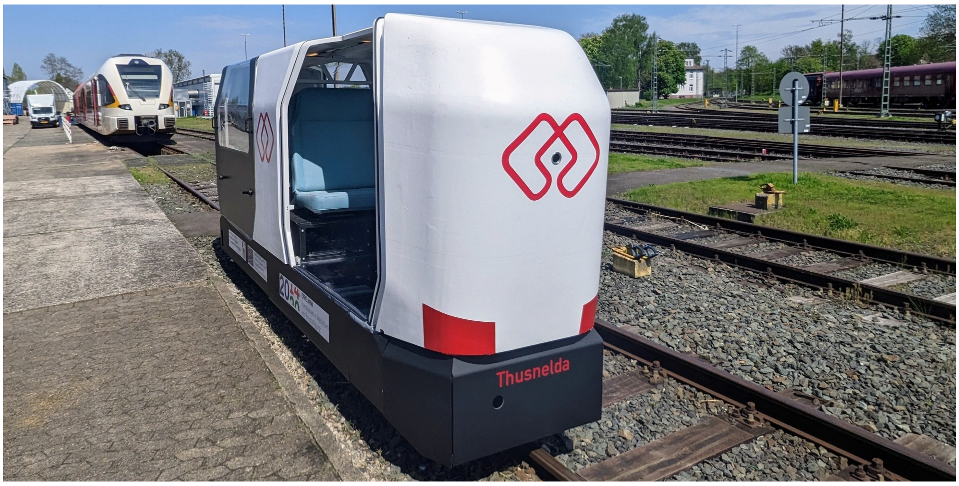


Fig. 1: Photo of the MONOCAB Demonstrator named ’Thusnelda’ in 2023

e.g. for video streams of the MONOCAB surroundings, and be robust against (wilful) disturbances.

This paper presents a corresponding communication concept based on the 5G communication standard. Section 2 describes the typical use cases and requirements for the MONOCAB communication. Based on that, Section 3 gives an overview of available technologies for the required communication system, leading to the communication concept in Section 4.

2 Use Cases and Requirements

2.1 Use Cases

In driving operation, situations, such as preceding, following, and oncoming MONOCABs, typical infrastructure along the track, as well as malfunctions of the vehicle itself are relevant. This minimum selection of driving situations results in the external communication interfaces, which are briefly described by four Use Cases for the communication system:

Use Case 1: As the MONOCABs pass each other close on a single track, a communication to oncoming MONOCABs in a **Vehicle-to-Vehicle (V2V)** communication can be used to exchange status information, such as position, velocity, acceleration, and vertical stabilization status. In addition, the V2V communication can support the object identification by sharing sensor data for environmental perception of objects along the rail track. For following MONOCABs, the communication can enable a closer distance by the synchronization of the velocity and acceleration, similar to the platooning of road trucks during collaborative cruise control [TJS16]. The data exchanged can complement the local sensor data of each MONOCAB for an enhanced vertical stabilization control and driving operation. This requires very low latencies to enable an appropriate reaction to new data.

Use Case 2: The **Vehicle-to-Infrastructure (V2I)** communication, for the connection to the infrastructure, such as railroad crossings of MONOCAB tracks and roads, can optimize the coordination of MONOCABs and road traffic. A communication to connected cars can even enhance the coordination. Sensor data from infrastructure systems in occupied sections can be exchanged for the environmental perception of objects and provide environmental data, such as wind, rainfall, and temperature, to adapt the MONOCAB driving operation, for example by a reduced driving speed.

Use Case 3: In case of emergencies or obstacles on the rail track that can not be identified by the system enabling the automated driving, a human operator can take over the control and move the MONOCAB out of the unclear situation via **remote control from an external control centre**. Therefore, the transmission of remote control commands as well as video streams of the MONOCABs surroundings and inside are required. This necessitates a low latency and high uplink data rate from the MONOCAB in a one-to-one communication between MONOCAB and control centre. As an assumption, the speed during the remote control is limited to 6 km/h to enable a safe breaking distance.

Use Case 4: In addition to the remote control from the control centre, less time-critical data for a **MONOCAB fleet management**, such as the battery status, has to be transferred continuously from all MONOCABs to the control centre. In the other direction, the control centre distributes information with the next stop and driving job for each MONOCAB. This communication requires a continuous communication between many MONOCABs and a single endpoint in a bidirectional many-to-one communication.

According to the four described use cases, the communication demands and capabilities of a MONOCAB can be contextualized close to the automotive domain, which puts many efforts into similar functionalities, such as teleoperated driving and V2X communication. On the other hand, technologies from the railway domain could be relevant, as the MONOCAB is operated on a rail track and has less possible driving manoeuvres compared to a road vehicle, as it is bound to the rail track. Nevertheless, the railway domain has different requirements for the typical much larger and faster trains. In addition, the railway technologies have to be compatible with the legacy communication systems on the existing rail tracks simultaneously used by accompanying trains. For the MONOCAB, a concept with an independent proprietary communication system will be proposed here.

2.2 Requirements

As mentioned in the Use Cases, a low latency for the direct control via remote control is needed. In addition to the communication system latency, video encoding and decoding, application latencies on MONOCAB and control centre side, as well as the reaction time of the operator have an impact on the total service latency for the remote control operation. In a study, the latency of the communication system is assumed with only one third of the total end-to-end service latency [5G21].

In the railway domain, a study assumes a total roundtrip latency for the communication system including the transmission of the video streams and control signals of 20 ms for a train remote control below 40 km/h [Ce21]. In the automotive domain, different projects identified different maximum latency values for teleoperated driving. In one project, the total latency of the communication system is assumed to be 60 ms for driving slower than 50 km/h [5G21]. As the MONOCAB is much smaller and lighter than a train and driving only 6 km/h in a remote control situation, we adapt the values from the automotive domain. This results in a latency of 60 ms roundtrip time for the communication system during remote control operation (**REQ.01**).

In addition to the low latency during the remote control, there is high data throughput needed. The uplink data throughput from a MONOCAB is expected with about 33 MB/s (**REQ.02**). This consists of four high definition video streams for the MONOCAB surroundings and inside, each using 8 MB/s. The control commands, status information and potential audio signals can be expected with less than 1 MB/s [5G21].

For the V2V communication in a platooning scenario, a service level latency of 50 ms is required in the automotive domain [5G23] and adapted for the MONOCAB (**REQ.03**). The V2V communication should be possible for a MONOCAB speed up to 80 km/h, resulting in a relative speed of 160 km/h (approx. 44 m/s) for oncoming MONOCABs (**REQ.04**). The resulting minimum range depends on the breaking distance of the MONOCAB as well as the time to establish a V2V communication. As both values are currently unknown, testes with the MONOCAB and a V2V communication system are pending. For now, a V2V communication distance of 440 meters (**REQ.05**) is assumed to enable a communication between two oncoming MONOCABs within 10 seconds at maximum speed.

Another important topic are different demands on service quality in terms of latency and data rate while communicating to multiple endpoints simultaneously. As a result, there is a prioritization of certain communication relations needed, for example with a higher priority for the remote control situations and safety relevant communication (**REQ.06**).

During all the use cases and scenarios, interference with private communication devices along the MONOCAB track and of MONOCAB passengers should be avoided (**REQ.07**). The MONOCAB is only used in a very limited area of reactivated disused tracks (**REQ.08**).

With regard to ICT (Information and Communication Technology) security of the MONOCAB, it is necessary to create security that is independent of other communication participants, or to establish a base-line security that works in every network without further latencies due to additional data processing for encryption and signatures (**REQ.09**). In the context of the objectives of the MONOCAB, the authenticity and integrity of the data and, depending on the data type, the confidentiality must be in the foreground of such a solution, depending on the type of data transmitted and the resulting need for protection. On the one hand, driving commands are transmitted from the control centre to the vehicle during the remote control. Secrecy is unnecessary here, but the protection of authenticity and integrity is. The same applies to the transmitted data for the MONOCAB fleet management. The situation is different with regard to any audio and image transmissions from the interior of the vehicle. These must additionally be protected with regard to their confidentiality.

3 State of the Art

5G is the first generation of mobile networks enabling the realization of stand-alone, standard-based private wireless networks in addition to public land mobile networks [Ro19] in order to support a broad range of vertical industries including mobility. These private networks can be implemented in a separate network infrastructure as well as an isolated virtual 5G network slice of a public infrastructure [5G19].

In [He22] the future 5G for railways (5G-R) was introduced, that in comparison to 5G shows some advantages, such as a higher reliability and handover success rate. Nevertheless, V2V communication is not part of the development. The end-to-end latency tends to be higher and the data rate tends to be smaller for 5G-R compared to 5G [He22]. The focus of 5G-R is the integration into the existing complex and diverse railway applications that have different requirements compared to the MONOCAB. Since traditional train control systems, such as the European Train Control System (ETCS), are not relevant for the MONOCAB operation, a proprietary system should be used that enables the usage of a rail track only with MONOCABs. As a result, normative standards and solutions targeting the high demands for railway systems with high-speed railways and large wagon trains are out of scope here.

In the automotive domain, scenarios and use cases similar to the use cases described for the MONOCAB are currently under development. According to the roadmap of the 5GAA (5G Automotive Association) for C-V2X (Cellular-V2X), automated driving, including automated parking and teleoperated driving similar to remote control for the MONOCAB, are possible in local areas and campuses using 4G. A transition to 5G-based communication is started. The sharing of dynamic objects, sensor signals, complex interactions, and cooperative manoeuvres are in development [5G22]. As mentioned before, the technologies used for the automotive domain are taken into account for the MONOCAB communication system due to similarities in the use cases and requirements.

For the V2X communication, there are two main solutions available and under development. C-V2X is part of the 3rd Generation Partnership Project (3GPP) standards. A direct communication via LTE-V2X was introduced in 3GPP Release 14 using the PC5 interface. The following 3GPP releases up to Release 17 bring up enhancements for the C-V2X communication, such as a latency of about 1 ms and a transition from LTE-V2X to 5G-V2X [MVH21]. The other standard for the V2X communication is based on IEEE

802.11p, using a WLAN-based direct communication. The enhanced standard 802.11bd is in development [MVH21, NCP19]. A comparison of the two standards from the 3GPP and IEEE leads to advantages and greater support of the C-V2X communication. It has benefits, such as longer communication distances, an integration into existing cellular infrastructure, and the support of the 5GAA [5G16] presumably leading to a wide adoption in the automotive domain.

With regard to ICT security, 5G has been significantly further developed compared to previous mobile radio standards. On the part of standardisation, security is specified, starting with security management through network monitoring systems to the classic protection goals of confidentiality, integrity, and authenticity. A challenge for ICT security in 5G networks lies in the inhomogeneity between network operators and the multitude of user equipment (UEs) involved. A study commissioned by the European Union [NI20a] shows potential vulnerabilities in the 5G network, which are located both in possible misconfigurations, but also in the confluence of the most diverse device manufacturers and network providers with different implementations. Overall, it can be seen that the security features of 5G standardisation are fine, but their implementation and thus use are often optional and thus dependent on the device manufacturer and mobile network provider [NI20a, para. 3.3]. A more general overview on IT-Security issues, based on the basic technologies used in 5G, is given by [Ah17].

Similar to the developments of the 3GPP C-V2X standard, the 3GPP Release 16 and Release 17 introduce new general features for 5G useful for the presented MONOCAB Use Cases. This includes the support of ultra reliable low latency communication (uRLLC) [Fo21]. Up to now, most hardware that is currently available only supports 3GPP Release 15 [Ha23]. Hence, a realization of the presented features is not possible at the time of writing this paper.

4 Communication Concept

Based on the state of the art, the use cases, and requirements, a concept for the MONOCAB communication is developed. The overall communication is shown in Figure 2. Here, network transitions to the Controller Area Network (CAN) protocol integrate the MONOCABs control units while all other end devices utilize IP based protocols.

4.1 Communication to Control Centre

In order to be independent of network operators and to limit the interference of the MONOCAB communication with the communication of passengers as well as other users of the public 5G network along the track (**REQ.07**), the MONOCAB communication system is based on a standalone NPN. Therefore, the MONOCAB communication can operate in a 5G network using a licensed private frequency for the limited area around the MONOCAB

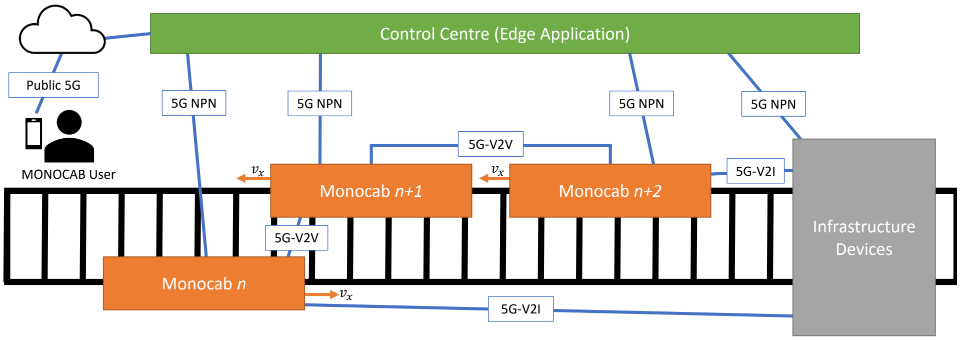


Fig. 2: Communication of three MONOCABs, the Control Centre, Infrastructure, and User

rail track (**REQ.08**). To make the system more resilient, e.g. in case of a network failure of the standalone NPN, the public network can act as a backup solution.

The solution based on an NPN also enables to run the MONOCAB control centre as an edge application. This can reduce the latency and increase the Quality of Service (QoS) [HYW19], especially for the remote control (**Use Case 3**). In a 5G outdoor field test with one MONOCAB, the added latency when using a 5G communication compared to a wired communication was about 26 ms from 119 ms (wired) to 145 ms (with 5G) glass-to-glass video stream service latency [Gu23]. However, more extensive practical tests to measure the latency (**REQ.01**) and data throughput (**REQ.02**) of a remote control application when operating a whole fleet of MONOCABs in a 5G network are pending.

The monitoring and management of the MONOCABs and the infrastructure can be realized as an edge application as well. The control centre can collect data, e.g. about the status of railroad crossings, from the infrastructure facilities. This non-time critical data can be transferred to and managed in a cloud service. From the cloud service, selected data can be made available for the MONOCAB users. To separate the different time-critical and non-time-critical communication, different slices or QoS profiles can be used for remote control (**Use Case 3**) and fleet management (**Use Case 4**) to fulfil **REQ.06**.

4.2 V2X Communication

The 5G-V2X or IEEE 802.11p based communication can be used to realise **Use Case 1** and **Use Case 2**. The advantage of the integration into the planned 5G communication infrastructure for the MONOCAB leads to a recommendation to adopt the 5G-V2X standard. An additional benefit is the support of the 5GAA [5G16], enabling a compatibility with the road traffic without multiple hardware and enabling the integration of road traffic and road safety messages, as well as pedestrian and bicycle protection as a part of 5G-V2X [5G22] into the MONOCAB communication.

A 4G-based V2V latency test using the PC5 interface resulted in a latency around 30 ms, hence lower than the 100 ms defined in the 3GPP Release 14 [MVH21] and fulfilling **REQ.03**. Another advantage of the C-V2X standards in comparison to IEEE 802.11p is a low end-to-end delay for longer distances. According to a simulation, a communication of safety broadcast messages could be realized with the C-V2X solution between two vehicles over a distance greater than 525 m with a delay of 4.2 ms, outperforming the simulation results when using IEEE 802.11p [Th18]. The distance is also fulfilling **REQ.05**. Newer systems following newer releases aim to further decrease the latencies. In another simulation, the V2X sidelink communication in the mmWave-band was independent of the vehicle velocity up to 300 km/h [Ki21], hence fulfilling **REQ.04**. However, the range in the higher frequency bands (mmWave-band) is limited and was out of scope in the simulations [Ki21]. Tests and simulations for the V2X communication with different velocities in the Sub-6-GHz frequency bands with a higher range are pending.

In addition to the infrastructure monitoring via the proposed edge application in the control centre, data can be exchanged directly between a MONOCAB and close infrastructure devices, for example at railroad crossings, using C-V2I communication (**Use Case 2**) similar to the C-V2V communication.

4.3 ICT Security

Each of the identified protection goals in **REQ.09** requires cryptographic procedures, such as encryption and signatures, which use different algorithms. The New Radio Integrity Algorithm (NIA) and New Radio Encryption Algorithm (NEA) procedures, which are AES-based and offer good protection, are primarily to be used [Co22]. A particular challenge arises from the combination of the required low latencies and cryptographic protocols and their partly non-deterministic computing time, for example for key generation, data encryption and decryption. These must not be violated under any circumstances. Established procedures, such as IPsec, come into consideration here [NI20b]. What is promising about these protocols is that they create end-to-end security, which is also demanded or recommended in other scientific works [Ku18, Zh22, Ah17].

5 Conclusions and Outlook

The concept presented in this publication shows to what extent the current mobile radio standard 5G can be used in the example of the automated driving monorail vehicle MONOCAB in theory. The criteria for this were developed on the basis of selected use cases and requirements. A concept covering the use cases and requirements is presented based on the state of the art in the 5G technology. Due to missing hardware supporting the needed functionalities of the recent 5G standards, practical implementations and tests for the whole concept are pending in order to prove the fulfilment of all presented requirements.

First experiments for a 5G-based MONOCAB communication were already executed and documented in [Gu23]. Nevertheless, more measurements with future hardware supporting the newer 3GPP Releases should be executed, especially in the environment of a rail tracks in rural areas that can be used by MONOCABs. The resulting characteristics of the 5G communication and the influence on the mentioned use cases can be analysed to further improve the concept.

Acknowledgement

This work was funded by the Ministry of Economic Affairs, Innovation, Digitalization, and Energy of the State of North Rhine-Westphalia in the research project “5G SIMONE - sicher.mobil.vernetzt.”

Bibliography

[5G16] 5GAA Automotive Association: The Case for Cellular V2X for Safety and Cooperative Driving. White Paper, 2016.

[5G19] 5G-ACIA: 5G for Automation in Industry. White Paper, 2019.

[5G21] 5GAA Automotive Association: Tele-operated Driving (ToD): System Requirements Analysis and Architectures. Technical Report, 2021.

[5G22] 5GAA Automotive Association: A visionary roadmap for advanced driving use cases, connectivity technologies, and radio spectrum needs. White Paper, 2022.

[5G23] 5GAA Automotive Association: C-V2X Use Cases and Service Level Requirements Volume II. Technical Report, 2023.

[Ah17] Ahmad, Ijaz; Kumar, Tanesh; Liyanage, Madhusanka; Okwuibe, Jude; Ylianttila, Mika; Gurtov, Andrei: 5G security: Analysis of threats and solutions. In: 2017 IEEE Conference on Standards for Communications and Networking (CSCN). pp. 193–199, 2017.

[Ce21] Cellarius, Bastian; Fritzsche, Richard; Lohmar, Thorsten; Kuo, Fang-Chun: Design of an FRMCS 5G E2E System for Future Rail Operation. Study Report, 2021.

[Co22] Cowperthwaite, Alex: A 5G Security Overview: Features, Rewards, and Risks of 5G Technology. KROLL, 2022.

[Fl20] Flasskamp, Martin et. al.: Vorstudie vernetzte Mobilität OWL. Ministerirum für Verkehr des Landes NRW, 2020.

[Fo21] Fodor, Gábor; Vinogradova, Julia; Hammarberg, Peter; Nagalapur, Keerthi Kumar; Qi, Zhiqiang Tyler; Do, Hieu; Blasco, Ricardo; Baig, Mirza Uzair: 5G new radio for automotive, rail, and air transport. IEEE Communications Magazine, 59(7):22–28, 2021.

[GMS21] Griese, Martin; Mousavi, Seyed Davood; Schulte, Thomas: Modeling the vertical dynamics of a self-stabilizing monorail vehicle. In: 2021 9th International Conference on Control, Mechatronics and Automation (ICCMA). IEEE, pp. 205–210, 2021.

[Gu23] Gustin, Denis; Siekmann, Timo; Kroll, Björn; Kleen, Philip; Schriegel, Sebastian; Jasperneite, Jürgen: Outdoor Field Test of 5G-based V2X Communication for Real-Time Monitoring and Remote Control of a Monorail Vehicle. In: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). IEEE, pp. 1–6, 2023.

[Ha23] Hardesty, Linda: Chips are holding back 5G private wireless. `https://www.fiercewireless.com/private-wireless/private-network-use-cases-expand-verizon`, 2023. Accessed: 2023-09-05.

[He22] He, Ruisi; Ai, Bo; Zhong, Zhangdui; Yang, Mi; Chen, Ruifeng; Ding, Jianwen; Ma, Zhangfeng; Sun, Guiqi; Liu, Changzhu: 5G for Railways: Next Generation Railway Dedicated Communications. IEEE Communications Magazine, 60(12):130–136, 2022.

[HYW19] Hassan, Najmul; Yau, Kok-Lim Alvin; Wu, Celimuge: Edge computing in 5G: A review. IEEE Access, 7:127276–127289, 2019.

[Ki21] Kim, Junhyeong; Noh, Gosan; Kim, Taehyoung; Chung, Heesang; Kim, Ilgyu: Link-level performance evaluation of mmwave 5g nr sidelink communications. In: 2021 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, pp. 1482–1485, 2021.

[Ku18] Kumari, K. Anitha; Sadasivam, G. Sudha; Gowri, S. Shymala; Akash, Sebastin Arockia; Radhika, E.G.: An Approach for End-to-End (E2E) Security of 5G Applications. In: 2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS). pp. 133–138, 2018.

[MVH21] Miao, Lili; Virtusio, John Jethro; Hua, Kai-Lung: PC5-based cellular-V2X evolution and deployment. Sensors, 21(3):843, 2021.

[NCP19] Naik, Gaurang; Choudhury, Biplav; Park, Jung-Min: IEEE 802.11 bd & 5G NR V2X: Evolution of radio access technologies for V2X communications. IEEE access, 7:70169–70184, 2019.

[NI20a] NIS Cooperation Group: Cybersecurity of 5G networks EU Toolbox of risk mitigating measures. `https://digital-strategy.ec.europa.eu/en/library/cybersecurity-5g-networks-eu-toolbox-risk-mitigating-measures`, 2020.

[NI20b] NIS Cooperation Group: A Guide to 5G network security 2.0. `https://www.ericsson.com/en/security/a-guide-to-5g-network-security`, 2020.

[Ro19] Rostami, Ahmad: Private 5G Networks for Vertical Industries: Deployment and Operation Models. In: 2019 IEEE 2nd 5G World Forum (5GWF). pp. 433–439, 2019.

[Th18] Thota, Jayashree; Abdullah, Nor Fadzilah; Doufexi, Angela; Armour, Simon: Performance of car to car safety broadcast using cellular V2V and IEEE 802.11 P. In: 2018 IEEE 87th Vehicular Technology Conference (VTC Spring). IEEE, pp. 1–5, 2018.

[TJS16] Tsugawa, Sadayuki; Jeschke, Sabina; Shladover, Steven E: A review of truck platooning projects for energy savings. IEEE Transactions on Intelligent Vehicles, 1(1):68–77, 2016.

[Zh22] Zhao, Jinghao; Li, Qianru; Yuan, Zengwen; Zhang, Zhehui; Lu, Songwu: 5G Messaging: System Insecurity and Defenses. In: 2022 IEEE Conference on Communications and Network Security (CNS). pp. 37–45, 2022.

http://dx.doi.org/10.25673/111746

Systemanalyse und Entwurf einer intelligenten Ladeinfrastruktur der Elektromobilität in Smart-Grid-Umgebungen

Niklas Dreyer[1], Michael Peter Siemon[1], Andreas Pretschner[1]

Stichwörter

Elektromobilität, intelligente Ladeinfrastrukturen, Smart Grid, Vehicle-to-Grid, OCPP 2.0.1, OSCP 2.0, IEC15118-20, Protokoll Verifikation und Konformität, Tests

Zusammenfassung

Im Paper werden die drei wichtigsten technischen Kommunikationsstandards einer modernen Ladearchitektur, der ISO 15118-20, das *Open Charge Point Protocol* (OCPP) in der Version 2.0.1 und das *Open Smart Charging Protocol* (OSCP) 2.0 betrachtet. Welche Softwareanforderungen werden gestellt und in welchem Umfang müssen diese implementiert werden, um eine Ladeinfrastruktur als intelligenten Verbraucher oder Energiespeicher nutzen zu können? Wie kann man den Systementwurf testen, wie skaliert dieser Entwurf? Reicht es, wie bisher die Kommunikation zwischen einzelnen Akteuren zu standardisieren oder sind weitreichendere Vorgaben notwendig, um die E-Mobilität und den Energiesektor nachhaltig zu koppeln? Die im Paper benutzte Methodik der schrittweisen partiellen Implementierung der notwendigen Softwarekomponenten ermöglicht den Aufbau einer realen *smart charging architecture* auf deren Grundlage, ausgehend von einer „Minimalvariante“ bis zur „Vollversion“, das elektrische Fahrzeug aktives Speicherelement im geregelten elektrischen Netzwerk sein kann.

1 Einleitung

Die Nationale Plattform Zukunft der Mobilität (NPM)[2] wurde im Jahr 2018 einberufen und ist in sechs Arbeitsgruppen weiterhin aktiv. Die Umsetzung der Beschlüsse des NPM gestaltet sich schwierig, die Entscheidungswege durch alle Instanzen der Politik und Industrie sind lang. Es stellt sich die Frage, ob die Ziele der NPM dank neuer Normen und Standards umsetzbar ist und welchen Beitrag dazu Forschungseinrichtungen leisten können. Das folgende Paper untersucht konkrete Fragestellungen der 6 (Normung)[3] der NPM und möchte die Frage beantworten, welche konkreten, technischen und logischen Schritte nötig sind, um eine intelligente und bidirektionale Ladeinfrastruktur zu etablieren. *Die DIN EN IEC 63110 / VDE 0122-110-1 Protokoll zum Management von Lade- und Entladeinfrastruktur für Elektrofahrzeuge definiert bereits die abstrakte Kommunikation einer Ladestation mit einem lokalen oder zentralen Energiemanagementsystem, nimmt aber noch keine Vereinheitlichung der konkreten Kommunikation vor.* ***Aus diesem Grund besteht zusätzlicher Normungs-***

[1] Hochschule für Technik, Wirtschaft und Kultur (HTWK) Leipzig - Fakultät Ingenieurwissenschaften
[2] https://www.plattform-zukunft-mobilitaet.de/
[3] https://www.plattform-zukunft-mobilitaet.de/schwerpunkte/ag-6/

und Standardisierungsbedarf. [1, S. 9] *By 2025, the automotive market demand for electric vehicles is expected to exceed 6.5 million units per year worldwide.* [2] Ohne den Ausbau entsprechender Infrastrukturen könnte die vermehrte Nutzung elektrischer Energie für den Kraft- und Personenverkehr zu einer Überlastung führen. Im Ansatz, das Elektrofahrzeug nicht als statischen Verbraucher, sondern vielmehr als dynamisch regelbare Last und darüber hinaus als Erzeuger (*Vehicle-to-Grid*) anzusehen, besteht die Chance, das Versorgungsnetz mit einem erhöhten Regelvolumen und einer möglichen Rückspeisung zu entlasten. *Smart charging means connecting charging points between users and operators using an smart EV charge management method to avoid overloading and/or destabilizing the grid.* [3] Die Nutzung batterieelektrischer Kraftfahrzeuge als steuerbare Last und elektrischer Energiespeicher wurde bereits in diversen Forschungsarbeiten und Studien theoretisch betrachtet. Konsens dieser Arbeiten war, dass eine Nutzung als steuerbare Last oder Speicher die Effizienz von volatilen Energieträgern wie Wind oder Photovoltaik erhöhen kann. *From the power grid's perspective, EV charging through EVSC could help maintain/improve the power grid operating condition while providing additional services to the operators, such as frequency regulation* [4] Der Ausbau und die Automatisierung von Ladeinfrastruktur werden zur effektiveren Nutzung von volatilen Energieträgern und somit zur Stabilisierung des Energienetzes beitragen, wenn diese in einem gewissen Grad plan- und steuerbar wird. Hieraus lassen sich zentrale Fragen für den Bereich der angewandten Prozessinformatik ableiten: Welche konkreten Umsetzungen müssen getroffen werden, um die einzelnen Teilnehmer mit den notwendigen informationstechnischen Normen und Standards auszustatten? Welche Aufgaben stellen sich für den einzelnen Teilnehmer einer Ladeinfrastruktur bei einem hohen Automatisierungsgrad der Ladeplanung?

2 Ziel der Arbeit

Wie einleitend beschrieben, besteht der Bedarf einer ausgebauten Ladeinfrastruktur mit Möglichkeit von geplanten, bidirektionalen Ladezyklen. Die technischen Grundsteine für eine Umsetzung sind gelegt und werden seitens der NPM unterstützt. [1] Aus dem Blickpunkt der angewandten Ingenieurwissenschaft, bzw. der Prozessinformatik wird sich dieses Paper mit der Umsetzung von technischer Kommunikation und Datenverarbeitung aller notwendigen Teilnehmer einer intelligenten Ladearchitektur befassen. Zunächst werden die einzelnen Beteiligten anhand der aktuell genutzten, bzw. veröffentlichten Standards und Normen erläutert und benannt. Anschließend werden verschiedene Ausbaustufen definiert und hinsichtlich der Aufgaben für die einzelnen Teilnehmer beschrieben. Mithilfe dieser Informationen wird gezeigt, mit welchen konkreten technischen Softwareimplementierungen eine intelligente, bidirektionale Ladearchitektur umsetzbar ist. Das Ziel ist es, eine Bewertung des Aufwands für die Hauptteilnehmer getrennt nach den Ausbaustufen zu schaffen. Dabei liegt der Fokus auf den technischen Kommunikationsstandards / Normen und nicht auf Finanzdienstleistungen oder dem elektrischen Netz. Die Anforderung an den Umfang einer Software ist von der Programmiersprache oder dem zugrunde liegendem System abgehoben.

3 Topologie - Anforderungsanalyse

In dieser Arbeit werden die Kommunikationsstandards und Normen für die technische Umsetzung einer intelligenten Ladeinfrastruktur für batterieelektrische Fahrzeuge im Kontext der Einbindung in elektrische Verteilnetze (*smart grid*) betrachtet. Dabei sind insbesondere die Akteure und Teilnehmer der Ladeinfrastruktur mittels vier aktueller Normen und Standards zu berücksichtigen (Abbildung 1). Die farbig markierte Benennung der fünf Teilnehmer wurde aus den Normen und Standards übernommen.

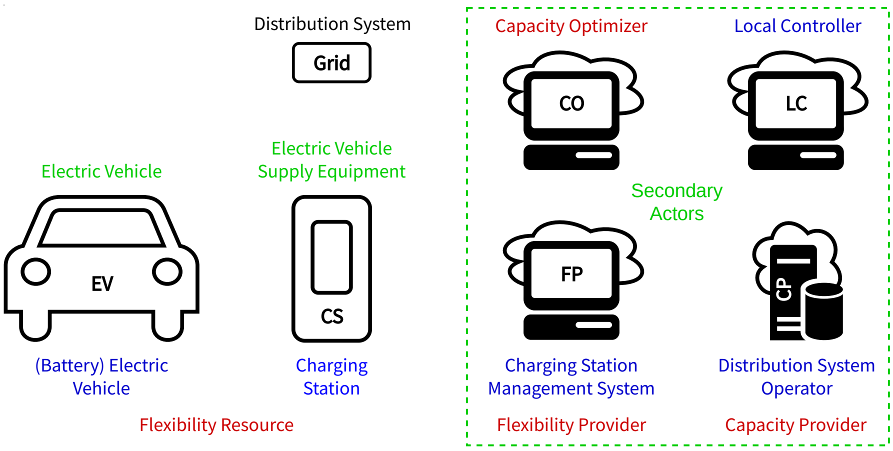


Bild 1: Technische Akteure einer öffentlichen Ladeinfrastruktur

Grün stellt die Namensgebung in der ISO 15118 (Abschnitt 3.2) dar. Es ist deutlich erkennbar, dass der Fokus dieser Norm auf der Kommunikation zwischen Fahrzeug und Ladestation (*primary actors* [5, S. 26]) liegt, da weitere Teilnehmer unter *secondary actors* zusammengefasst sind [5, S. 26]. Die in blauer Farbe dargestellten Bezeichnungen entstammen dem OCPP 2.0.1 (Abschnitt 3.3), es werden mit Ausnahme des *Capacity Optimizer* alle Teilnehmer einzeln definiert. Das OCPP standardisiert die Schnittstelle zwischen *Charging Station* und *Charging Station Management System* [6, S. 2], da in einigen Use-Cases (UCs) des Protokolls explizit die Akteure *Electrical Vehicle* [7, S. 269] und *Distrbution System Operator* [7, S. 231] inkludiert werden, sind diese auch namentlich definiert. Beim *Local Controller* [6, S. 15] handelt es sich laut Definition um eine *Charging Station* ohne EVSE, bzw. ohne Ladeanschluss für ein EV. Da der LC kein eigenständiger Teilnehmer sein muss, sondern auch nur ein Dienst innerhalb des CSMS, wird er in diesem Paper in das CSMS inkludiert, sodass die Gesamtarchitektur simplifiziert wird. Die roten Definitionen sind die des OSCP 2.0 (Abschnitt 3.4), dieses Protokoll dient der standardisierten Kommunikation zwischen *Flexibility Provider* [8, S. 4] / *Charging Station Management System* und *Capacity Provider* [8, S. 4] / *Distrbution System Operator*. Zusätzlich wird der Teilnehmer *Capacity Optimizer* [8, S. 4] in die Topologie gebracht, dieser dient mit einer Verbindung zum *Flexibility Provider* als Informationsquelle für die Infrastruktur. Da dieser nicht zwingend eine eigene Instanz sein muss und logisch auch im *Flexibility Provider* angesiedelt sein kann, wird er mit diesem zusammengefasst. Im weiteren Verlauf werden folgende Begriffe, bzw. Abkürzungen verwendet:

- *(Battery) Electric Vehicle / Flexibility Resource (Car Part))* → **EV**
- *Electric Vehicle Supply Equipment / Charging Station / Flexibility Resource)* → **CS**
- *Charging Station Management System / Flexibility Provider / Local Controller* → **FP**
- *Distribution System Operator / Capacity Provider / Capacity Optimizer* → **CP**

3.1 Ladesysteme von Elektrofahrzeugen (DIN 61851)

Die IEC 61851-1 [9] ist eine internationale Norm für das Laden von Elektrofahrzeugen. Sie enthält Richtlinien für den Entwurf, die Prüfung und die Installation von Ladesystemen für Elektrofahrzeuge. Die Norm deckt sowohl Wechselstrom- als auch Gleichstrom-Ladesysteme ab und legt Anforderungen für die in diesen Systemen verwendeten Steckverbinder, Kabel und Ladegeräte fest. Sie enthält auch Sicherheits- und Leistungskriterien, die erfüllt sein müssen, damit ein Ladesystem als konform mit der Norm gilt. [10] Die IEC 61851-1 soll gewährleisten, dass Ladesysteme für Elektrofahrzeuge sicher, zuverlässig und interoperabel sind, und das Wachstum des Marktes für Elektrofahrzeuge unterstützen.

Funktionsanalyse - Softwareanforderungen / Kommunikationsumfang

Die Kommunikation zwischen **EV** und **CS** ist in der IEC 61851-1 beschrieben, seitens des Fahrzeugs können Widerstandswerte zwischen dem *Control Pilot* und dem Schutzleiter (PE) angepasst werden, sodass die angelegte Spannung sich verändert [9, S. 29]. Dies dient zur Übermittlung der Ladebereitschaft seitens des **EV**. Parallel dazu wird von der **CS** eine Pulsweitenmodulation genutzt, um dem **EV** die zur Verfügung stehende elektrische Leistung in Form von Strom pro Phase mitzuteilen [9, S. 24].

3.2 Kommunikation Elektrofahrzeug - Ladeinfrastruktur (ISO 15118)

Die ISO 15118 [5] [11] [12] ist eine Norm für die Kommunikation zwischen **EVs** und **CSs**. Sie definiert ein Protokoll für den sicheren und effizienten Austausch von Informationen und Energie zwischen diesen beiden Systemen. Diese Norm ermöglicht es **EV**, automatisch einen Ladevorgang mit einer kompatiblen **CS** einzuleiten und durchzuführen, ohne dass ein direktes menschliches Eingreifen erforderlich ist. Sie soll dazu beitragen, die breite Einführung von **EV** zu erleichtern und die Entwicklung eines nachhaltigeren Verkehrssystems zu unterstützen. [10]

Softwareanforderungen

Die von der ISO 15118 Normfamilie definierte Kommunikationsarchitektur ist in Analogie zu dem OSI-Referenzmodell aufgebaut [12, S. 1,16f]. und basiert zu einem Teil auf existierenden, teils weit verbreiteten Protokollen und Technologien, wie HPGP[4] (PLC), Wi-Fi[5], IPv6[6], UDP[7], TCP[8] und TLS[9]. [11, S. 32,73] [12, S. 51ff,54ff] Für den Austausch von Nachrichten über das von der ISO 15118 beschriebene *Vehicle-to-Grid*-Kommunikationsprotokoll ist der Aufbau einer TCP/IPv6-Verbindung über das Ladekabel (PLC) oder per Wi-Fi notwendig [12, S. 128ff]. Zudem beschreibt die Norm ein eigenes Protokoll (V2G Transfer Protocol) für die Steuerung des Datenverkehrs [12, S. 82ff], Bei den V2G-Nachrichten handelt es sich um, aus XML-Schemen abgeleiteten, Datenstrukturen und -typen. Die Repräsentation, d.h. (De-)Kodierung, dieser Nachrichten wird auf von einem sogenannten EXI-Coder[10] übernommen [12, S. 87ff], welcher frei erhältlich ist, jedoch eigens in die Ladesoftware eingebunden werden muss. Der Informationsaustausch zwischen **EV** und **CS** ist zustandsbehaftet und hängt von der gewählten Ladetechnologie ab. Für die Implementierung des V2GP bedeutet dies, dass eine oder mehrere Zustandsmaschinen aus den in der Norm definierten Kommunikationsanforderungen abgeleitet werden müssen. Die Transitionsbedingungen und Operationen werden von der Norm nur implizit vorgegeben, wodurch sich die Rekonstruktion der Zustandsmaschinen als sehr komplexer Prozess herausstellt.

3.3 Kommunikationsprotokoll für das Laden (OCPP 2.0.1)

Das „Open Charge Point Protocol“ ist ein Kommunikationsprotokoll für **CS** und **FP**. Es definiert einen Standardsatz von Nachrichten [7, S. 310-340] und Operationen [7, S. 15-310], die für die Fernverwaltung und -steuerung von **CS** verwendet werden können. OCPP 2.0.1 [13] [6] [7] [14] [15] ist die neueste Version dieses Protokolls und bietet Unterstützung für neue Funktionen wie Lastausgleich [7, S. 203-244], Authentifizierung [14, S. 62-99] und Preisinformationen [7, S. 184-192]. Sie enthält Verbesserungen bestehender Funktionen,

[4] HomePlug Green PHY Specification
[5] IEEE 802.11-2020, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications
[6] IETF RFC 8200, Internet Protocol, Version 6 (IPv6) Specification (July 2017)
[7] IETF RFC 768, User Datagram Protocol (August 1980)
[8] IETF RFC 793, Transmission Control Protocol [...] Protocol Specification (September 1981)
[9] IETF RFC 8446, The Transport Layer Security (TLS) Protocol Version 1.3 (August 2018)
[10] W3C EXI 1.0, Efficient XML Interchange (EXI) Format 1.0, W3C Recommendation (March 2011)

wie die Unterstützung mehrerer Ladepunktanschlüsse und eine verbesserte Diagnose [7, S. 276-292].

Softwareanforderungen

Aus der Protokolldokumentation des OCPP 2.0.1 [15] lassen sich folgende Mindestanforderungen an eine Schnittstelle extrahieren. *„For the connection between a Charging Station and a Charging Station Management System (CSMS) using OCPP-J, the CSMS acts as a WebSocket server and the Charging Station acts as a WebSocket*“ [15, S. 4] Als Basis dient eine WebSocket[11] Verbindung, welcher eine HTTP[12] Anfrage vorhergeht [15, S. 5]. „*The whole message consisting of wrapper and payload MUST be valid JSON encoded with the UTF-8 character encoding.*“ [15, S. 8] Die übertragenen Daten beim OCPP sind im JSON[13] [14] Format zu strukturieren und in UTF-8[15] zu codieren. Neben den Anforderungen an die Übertragung, gibt es die Grundimplementierung der internen Abläufe (UCs) [13, S. 13], welche nach der Protokollvorgabe umgesetzt werden müssen. **B**01-**B**04 (*Booting a Charging Station*); **B**05-**B**07 (*Configuring a Charging Station*); **B**11-**B**12 (*Resetting a Charging Station*); entweder **C**01, **C**02 oder **C**04 (*Authorization options*); **E**01 (eine aus S01-S06), **E**02-**E**03, **E**06 (eine aus S01-S06), **E**07-**E**08, entweder **E**09 oder **E**10 und **E**11-**E**13 (*Transaction mechanism*); **G**01, **G**03-**G**04 (*Availability*); **G**05 und **N**07 (*MonitoringEvents*); **J**02 (*Sending transaction related Meter values*); **P**01-**P**02 (*DataTransfer*); Nur wenn ein Akteur diese ca. 30 UCs implementiert hat, kann er von OCPP 2.0.1 Unterstützung sprechen.

3.4 Erweiterte Anwendungsfälle für das Laden (OSCP 2.0)

Im OSCP 2.0[16] [8] werden Anwendungsfälle beschrieben, in denen die Nachrichten allgemeiner angewendet werden als in OSCP 1.0[17], das speziell auf das intelligente Laden von Elektrofahrzeugen durch einen Verteilernetzbetreiber (DSO) ausgerichtet war. Der Grund für die Verwendung allgemeinerer Begriffe ist, dass diese Spezifikation die Möglichkeiten des Protokolls nicht auf das intelligente Laden von **EVs** beschränken will. Dies wird durch die Integration von **EVs** in größere Energie-Ökosysteme, einschließlich PV, stationäre Batterien, Wärmepumpen und andere Geräte, vorangetrieben. Weitere Änderungen sind die Umstellung auf JSON[18] / REST[19] [8, S. 23], zusätzliche Prognosetypen (Erzeugung, Verbrauch, Fallback) [8, S. 27-28] und eine Nachricht zur Meldung von Fehlern.

Softwareanforderungen

Für den Datenaustausch zwischen Teilnehmern, die das OSCP in der Version 2.0 implementieren, werden zwei HTTP[20] Verbindungen benötigt [8, S. 23] um eine bidirektionale Kommunikation zu ermöglichen. Das bedeutet für jeden Teilnehmer die Bereitstellung eines HTTP Servers und die Nutzung eines HTTP Clients pro Verbindung. *The protocol is based on HTTP combined with JSON formatting (mimetype application/json). It fits within a RESTful architecture.* [8, S. 23] Daten werden im JSON13 14 Format strukturiert. Für die HTTP Server Implementierung sind Zertifikate für die verschlüsselte Verbindung notwendig. *[..] only server side certificates in order to set up a secure SSL connection.* [8, S. 23] Es gibt keine Grundimplementierung der internen Abläufe (UCs) wie bei OCPP (Abschnitt 3.3). OSCP 2.0 beinhaltet insgesamt sieben UCs [8, S. 10-21], diese werden als zu implementieren angenommen.

[11] IETF RFC 6455, The WebSocket Protocoll (December 2011)
[12] IETF RFC 2616, Hypertext Transfer Protocol — HTTP/1.1(June 1999)
[13] IETF RFC 7515, JSON Web Signatures (JWS)(May 2015)
[14] IETF RFC 7518, JSON Web Algorithms (JWA) (May 2015)
[15] IETF RFC 3629, UTF-8, a transformation format of ISO 10646 (November 2003)
[16] https://www.openchargealliance.org/protocols/oscp-20/
[17] https://www.openchargealliance.org/protocols/oscp-10/
[18] 13
[19] IETF TFC 6690, Constrained RESTful Environments (CoRE) Link Format (August 2012)
[20] IETF RFC 2616, Hypertext Transfer Protocol — HTTP/1.1(June 1999)

4 Varianten der Systemarchitektur

Ausgehend von den beschriebenen Standards sollen im Folgenden vier Varianten einer *smart charging architecture* definiert werden. Die Methodik der folgenden Spezifikation richtet sich nach dem Grad und Aufwand der Umsetzung/Implementierung unter realen Bedingungen. Der Ansatz soll zeigen, dass man die Integration der elektrischen Batteriespeicher in das elektrische Netz unter bestimmten Bedingungen erreichen kann - dies aber nicht im Sinne einer vollständigen Implementierung, sondern nur durch eine schrittweise (partielle) Umsetzung.

	Ausspeisung	Einspeisung
dynamisch	„passiver“ Verbraucher	„passiver“ Erzeuger
geplant	„aktiver“ Verbraucher	„aktiver“ Erzeuger

In der Tabelle sind die vier Betrachtungen eines **EV**, bzw. der gesamten *Flexibility Resource* aus Sicht des elektrischen Energienetzes dargestellt. Grundlage sind die in Abschnitt 3 beschriebenen technischen Möglichkeiten der Summe aller Standards. *Dynamisch* bedeutet in diesem Fall, dass der **CS** kein Zeitraum für die Energieübertragung bekannt ist, sodass allein ein Faktor wie der aktuelle Preis die Richtung und Stärke der Übertragungsleistung bestimmen kann. Bei *geplant* liegen Informationen wie Abfahrtzeitpunkt und Ladestand zu diesem Zeitpunkt vor. Aus dieser Betrachtung lassen sich verschiedene Architekturen der Ladeinfrastruktur ableiten. Als Basis dient eine Mindestimplementierung, welche nicht primär die Funktionen einer intelligenten Infrastruktur innehat. In diesem Sinne verstehen sich die folgend beschriebenen Architekturversionen nicht als unabhängige Realisierungen, sondern als funktionale Erweiterungen der Version 0 (Minimalimplementierung).

Variante 0: Minimalimplementierung

In der niedrigsten Ausbauversion einer möglichen *smart grid architecture* gibt es drei Akteure, diese sind in Abbildung 2 dargestellt.

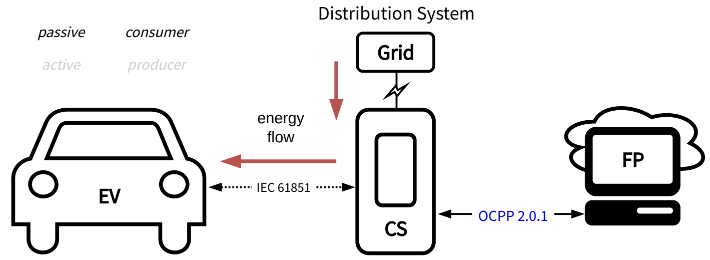


Bild 2: Minimalausbau einer öffentlichen Ladeinfrastruktur

Für diese Minimalimplementierung ist die IEC 61851 als Kommunikation zwischen **EV** und **CS** vorgesehen. Wie in Abschnitt 3.1 beschrieben, können damit Ladeleistungswerte von der **CS** an das **EV** übertragen werden. Diese müssen in der **CS** Software vorgesehen sein, als Basis dienen physikalische Grenzen wie die Anschlussleistung der Anlage oder thermische Belastung. Als Grundlage des technischen Datenaustausches zwischen **CS** und **FP** ist das OCPP 2.0.1 vorgesehen. Die in Abschnitt 3.3 beschriebenen UCs die ein Teilnehmer mindestens implementieren muss, um OCPP-konform arbeiten zu können, sind in diesem Szenario umzusetzen. Intelligentes Laden ist in diesem Fall nur im Sinne einer Übertragung von z.B. *charging limits* durch den UC **K**09 [15, S. 250f] möglich.

Variante 1: „SCready“

In der funktional erweiterten Variante wird die Infrastruktur um den **CP** und damit zwingend auch um das OSCP 2.0 (Abschnitt 3.4) erweitert.

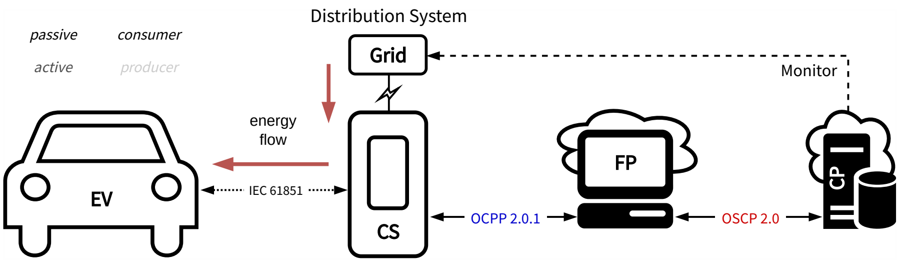


Bild 3: Erweiterte Ausbauversion einer öffentlichen Ladeinfrastruktur

Es bedarf bei einer Implementierung von OSCP 2.0 [8] mindestens die UCs für die *Connectivity* (2) [8, S. 10-12] und die Definition von **group_id**’s [8, S. 27-28]. Dabei handelt es sich um einen Bereich, in dem eine oder mehrere *Flexibility Resource(s)* an das elektrische Netz angeschlossen sind. („*The id of the area in which the Flexibility Provider has Flexibility Resources connected to the grid.*“ [8, S. 27]). Eine **group_id** setzt voraus, dass ein **CP** die von ihm betriebenen und verwalteten Teile des Stromnetzes unterteilt hat und diese Informationen über die Bereiche mit dem **FP** teilt. Der **CP** überträgt eine Leistung für diese Bereiche zu bestimmten Zeitpunkten in der nahen Zukunft (max. 24h) an den **FP**. Seitens des OCPP müssen diese Informationen der zur Verfügung stehenden Ladeleistung vom **FP** an die in den jeweiligen **CS** weitergeleitet werden. Vorab ist diese Ladeleistung zu parametrisieren, so diese nicht der vollen Anschlussleistung entspricht. Konkret wird z.B. der UC **K**01 [7, S. 233f] implementiert um der **CS** oder einem einzelnen Anschluss eine Ladeleitung vorzugeben. Die **CS** setzt diese Vorgaben um und passt durch Mittel der IEC 61851 die Ladeleistung an. Der Energiefluss ist unidirektional, nicht planbar aber in der Leistung regulierbar.

Variante 2: „FullSC“

Die nächste Erweiterung in der Architektur ist die ISO 15118-20 (Abschnitt 3.2). Durch diese wird der passive Teilnehmer **EV** zu einem potenziellen Akteur.

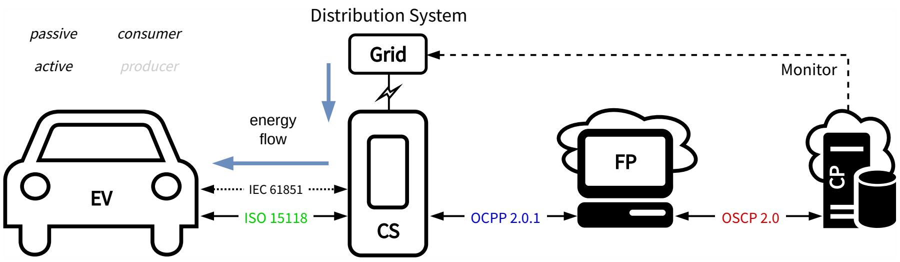


Bild 4: Dritte Ausbaustufe einer öffentlichen Ladeinfrastruktur

Im OSCP gibt es in Version 2 keine zwingende Erweiterung gegenüber der Version 1. Es ist jedoch angemessen, den UC zur Messwertübertragung (*Distribute measurements* [8, S. 18-19]) zu implementieren, da der **CP** hierdurch aktuell genutzte Leistungen in seinen Bereichen kennt und auf Basis dieser Informationen handeln kann. Durch die Verfügbarkeit von Informationen wie *departure time* oder *charging schedule* vom **EV** können OCPP-UCs wie **K**08 [15, S. 248ff] genutzt werden um eine Ladeplanung zu implementieren. Des Weiteren verfügt OCPP über drei UCs für die Nutzung der ISO 15118 an einer **CS**. Diese sind auf Abläufe der ISO Norm angepasst und in dieser Architekturvariante vollständig umsetzbar (**K**15-**K**17 [15, S. 259-267]). Da der **CS** die oben genannten

Informationen bekannt sind, führt das zu einem weiteren Bedarf an Entscheidungslogik. Diese muss nicht nur mit aktuellen Werten arbeiten, sondern auch in die Zukunft planen und Energiemengen zuweisen. Dabei muss der **CP** für gesamte Gebiete die Leistung für einzelnen Bereiche einteilen, diese dem **FP** übertragen und dieser die Entscheidungen für einzelne **CS** treffen. Die wesentliche Erweiterung findet hier durch die Einführung einer *High-Level*-Kommunikation zwischen **EV** und **CS** auf Basis der Norm ISO 15118 statt. Für die Konfiguration und Steuerung von Ladevorgängen umfasst die Norm ein zweistufiges Verfahren. Zunächst findet eine Einigung auf einen Energie-Transfer-Service und dessen Parametrisierung statt [12, S. 154ff,361ff]. Die **CS** stellt dem **EV** auf Anfrage alle von ihr unterstützen Lade-Services dar (*ServiceDiscovery*). und liefert ggf. weitere Informationen über einen oder mehrere Services in Form von Serviceparameter-Listen (*ServiceDetail*). Beispiele hierfür sind die unterstützen Steckerbelegungen, Steuer-Modi (Scheduled, Dynamic), Mobilitätsbedarfsmodi und Preisgestaltungsoptionen. Das **EV** übersendet abschließend den gewählten Lade-Service und seine Auswahl zugehöriger Parameter (*ServiceSelection*). In der zweiten Stufe findet, in Abhängigkeit des gewählten Service (und dessen Parametern), ein Austausch von Ladeparametern statt [12, S. 134ff,159ff,209ff]. EV und **CS** kommunizieren ihre (physischen) Grenzwerten für die Leistungsübertragung, die zu jedem Zeitpunkt eingehalten werden müssen (*ChargeParameterDiscovery*). In Abhängigkeit des gewählten Steuer-Modus wird daraufhin der Ablauf des bevorstehenden Energieaustauschs verhandelt (*ScheduleExchange*). Im dynamischen Modus informiert das **EV** die **CS** über die Energiemenge, die es plant während des Ladevorgangs bis zu einem bestimmten Abfahrtszeitpunkt aufzunehmen. Die **CS** kann diese Sollwerte akzeptieren oder verweigern. Im geplanten Modus bietet die **CS** dem **EV** mögliche Ladepläne (Schedules) an, welche detaillierte, in mehrere Zeitabschnitte unterteilte, Leistungs- und Preisinformationen für den gesamten Ladezeitraum enthalten. Bei der Berechnung dieser Ladepläne werden alle relevanten Soll- und Grenzwerte sowie Vorgaben von SAs berücksichtigt. Der Energietransferprozess kann durch das **EV** (und indirekt durch die **CS**) gestartet, ausgesetzt oder beendet werden. Zusätzlich kann das **EV** der **CS** sein voraussichtliches Ladeleistungsprofil bereitstellen (*PowerDelivery*). Während des Ladevorgangs findet ein kontinuierlicher Austausch von leistungs- und energiebezogenen Soll- und Ist-Werten von **EV** und **CS** statt (*ChargeLoop*). Das **EV** kann außerdem die aktuellen Zählerstände der **CS** abfragen. Der Energiefluss ist unidirektional, planbar und in der aktuellen Leistung regulierbar.

Variante 3: „UltraSC“

Die Änderung von „FullSC“ zum maximal implementieren Funktionsumfang der Architekturversion „UltraSC“ bestehen in einer Erweiterung der Infrastruktur um den **CO** und des *bidirectional power transfer* (BPT).

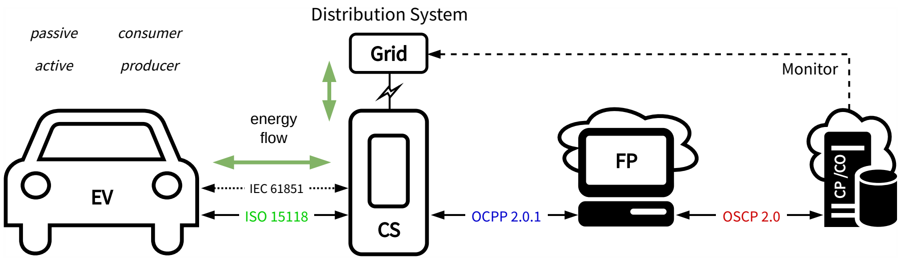


Bild 5: Vollausbau einer öffentlichen Ladeinfrastruktur

Zwischen **CP** und **FP** besteht spätestens hier eine vollständige Implementierung aller UCs des OSCP. In diesem Szenario gibt es zusätzlich einen **CO**, dessen Rolle laut Protokoll eine additive zu der des **CP** ist. Es sollen anhand zusätzlicher Informationen wie Wetterdaten ein optimierter *capacity forecast* [8, S. 7] gebildet werden. Dies bedeutet in jedem Fall eine zusätzliche Entscheidungslogik für den **FP**. Die ISO 15118-20 umfasst einen Satz an Services für das bidirektionale Laden [12, S. 361ff], bei dem neben der Speisung aus dem Netz auch eine Einspeisung erfolgen kann. Die Fähigkeit des bidirektionalen Ladens können sich **EV** und **CS** gegenseitig während der Service-Aushandlung bestätigen. Die Nachrichtenparameter-Sätze für das bidirektionale Laden ersetzen und/oder erweitern die ursprünglichen Parameter-Sätze um die für eine inverse Energieübertragung relevanten Grenz-, Soll- und Ist-Werte. Hier sind beispielsweise die maximale inverse Ladeleistung, der gewählte Generator-Modus (netzbildend/netzfolgend) und die Kanal-Architektur zu nennen [12, S. 313ff]. Wie beim unidirektionalen Laden wird zwischen den Lademodi *dynamic* und *scheduled* unterschieden. Der Verlauf der Einspeisung kann demnach ebenso in Form von Ladeplänen ausgehandelt werden. Für die Umsetzung des bidirektionalen Ladens ist softwareseitig die Implementierung der Service- und Nachrichten-Parameter vorzunehmen. Der Energiefluss ist bidirektional, voll planbar und in der aktuellen Leistung regulierbar.

5 Validierung

Eine grafische Repräsentation der vorhergehenden Betrachtungen (Abbildung 6) verdeutlicht, dass die Normen und Standards für die verschiedenen Ausbaustufen (Versionen der Systemarchitektur) einer intelligenten Ladeinfrastruktur zu unterschiedlich großen Anteilen abgedeckt sind.

Die Implementierung einer Ladearchitektur nach Version 3 „UltraSC“, welche den bidirektionalen und geplanten Energietransfer unterstützt, erfordert einen erheblichen Entwicklungsaufwand. Zusätzlich müssen sich Teilnehmer wie **FP** und **CP** bereit erklären, Informationen zu Netzbeschaffenheit und Messwerten auszutauschen. Die Bereiche außerhalb der mit Strichlinien hervorgehobenen Kreise für die Standards, stehen qualitativ für die nicht durch Normung abgedeckten Softwarebestandteile, welche zusätzlich definiert und implementiert werden müssen. Dazu gehören Entscheidungsprozesse zur lokalen Lastverteilung und die Umsetzung von Kapazitätsvorhersagen zu konkreten Ladeprofilen. Diese sind bislang durch keine der betrachteten Normen ausreichend beschrieben, da die Standardisierung an dieser Stelle auf einer übergeordneten Ebene stattfinden muss. Weiterhin sind diese Standards nicht zwingend gekoppelt und auch einzeln werden die UCs zu *smart charging* nicht für die Mindestimplementierung vorgesehen. Es ist legitim eine **CS** oder ein **FP** OCPP 2.0.1 anzubieten, ohne einen der 17 *smart charging* UCs abzudecken. Solange es keine Verpflichtung zur Verbindung von **FP** und **CP** mit OSCP 2.0 oder einem anderen Protokoll gibt, OCPP 2.0.1 in der Mindestimplementierung keine *smart charging* UCs vorsieht und die ISO 15118-20 keine Pflicht bei **EV** und **CS** sind, ist für den öffentlichen Raum eine solche Ladearchitektur flächendeckend nur schwer zu realisieren.

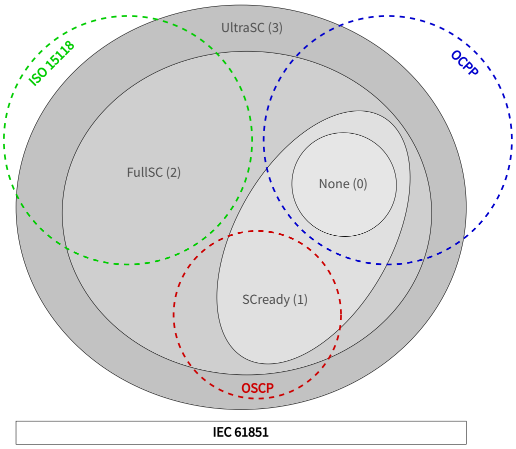


Bild 6: Qualitative Betrachtung der Schnittmengen zwischen Normen und Ausbaustufen

6 Work in Progress - Ausblick

Die vorgestellten Lösungen können als Umsetzungsrichtlinie und „Leistungsklassen“ verstanden werden. In diesem Fall ist es möglich für alle Teilnehmer im *smart grid* verbindliche Umsetzungsrichtlinien vorzugeben und im weiteren Ausbau kompatibel zu verschiedenen Anbietern der elektrischen Ladeinfrastruktur, den Fahrzeugherstellern und den Energieversorgern zu bleiben. Eine weitere Hoffnung ist die im Juli 2022 vorgestellte IEC 63110-1:2022[21]. Diese soll ein Teil der oben genannten Probleme angehen und für vollumfängliche Standards in der Ladeinfrastruktur sorgen.

Im Laufe der Entwicklung wurde durch die Autoren eine Testumgebung entwickelt, die einerseits als Hardware-In-The-Loop System, mit konkreter Charger-Hardware (AC) und Ladecontroller und simulierten Software-Back-End als auch als Software-In-The-Loop System mit simulierter Hardware und konkretem Software-Back-End System betrieben werden kann. Mit diesem Stand der Testumgebung kann die Variante 0 (Minimalimplementierung) umgesetzt werden. Die dafür notwendigen **UCs** des Kommunikationsprotokolls OCPP Version 2.0.1 sind adressiert und wurden schon teilweise realisiert. Die ausstehende Realisierung der darauf folgenden Varianten erfordert einerseits die Implementierung des Kommunikationsprotokolls OSCP Version 2.0 als auch die Anbindung eines simulierten elektrischen Teilnetzes (verfügbar im Institut Elektrische Energietechnik der HTWK).

Literatur

[1] Z. u. T. Nationale Plattform Zukunft der Mobilitaet Arbeitsgruppe 6 „Normung, Standardisierung, “Schwerpunkt-roadmap intelligentes lastmanagement,” April 2020.

[2] A. Belkaaloul and B. A. Bensaber, “Anonymous authentication protocol for efficient communications in vehicle to grid networks,” in *2021 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–5, 2021.

[3] K. Hajar, B. Guo, A. Hably, and S. Bacha, “Smart charging impact on electric vehicles in presence of photovoltaics,” in *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, vol. 1, pp. 643–648, 2021.

[4] O. Sadeghian, A. Oshnoei, B. Mohammadi-Ivatloo, V. Vahidinasab, and A. Anvari-Moghaddam, “A comprehensive review on electric vehicles smart charging: Solutions, strategies, technologies, and challenges,” *Journal of Energy Storage*, vol. 54, pp. 1–24, October 2022.

[5] “Road vehicles – vehicle to grid communication interface – part 1: General information and use-case definition,” March 2018.

[6] OpenChargeAlliance, “Ocpp 2.0.1 - part 1 - architecture & topology,” techreport, Open Charge Alliance, Utrechtseweg 310, 6812 AR Arnhem, Niederlande, December 2019.

[7] OpenChargeAlliance, “Ocpp 2.0.1 - part 2 - specification,” techreport, Open Charge Alliance, Utrechtseweg 310, 6812 AR Arnhem, Niederlande, December 2019.

[8] OpenChargeAlliance, “Oscp 2.0 - specification,” techreport, Open Charge Alliance, Utrechtseweg 310, 6812 AR Arnhem, Niederlande, October 2020.

[9] “Konduktive ladesysteme fuer elektrofahrzeuge - teil 1: Allgemeine anforderungen,” November 2001.

[10] N. P. E. (NPE), “Die deutsche normungs-roadmap elektromobilitaet 2020,” April 2017.

[11] “Road vehicles – vehicle to grid communication interface – part 3: Physical and data link layer requirements,” August 2016.

[12] “Road vehicles — vehicle to grid communication interface -part 20: 2nd generation network layer and application layer requirements,” April 2022.

[13] OpenChargeAlliance, “Ocpp 2.0.1 - part 0 - introduction,” techreport, Open Charge Alliance, Utrechtseweg 310, 6812 AR Arnhem, Niederlande, December 2019.

[14] OpenChargeAlliance, “Ocpp 2.0.1 - part 2 - appendices,” techreport, Open Charge Alliance, Utrechtseweg 310, 6812 AR Arnhem, Niederlande, December 2019.

[15] OpenChargeAlliance, “Ocpp 2.0.1 - part 2 - json over websockets implementation guide,” techreport, Open Charge Alliance, Utrechtseweg 310, 6812 AR Arnhem, Niederlande, December 2019.

[21] https://webstore.iec.ch/publication/60000

Impressum

14. Jahreskolloquium

« Kommunikation in der Automation »
(KommA 2023)

21./22.11.2023 • Magdeburg

DOI :http://dx.doi.org/10.25673/111633

HERAUSGEBER

Ulrich Jumar, 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