Välkommen till

A

Bredd och Impact

Agenda spår A – del 1

11.00– 11.10	Inledning – Fredrik och Staffan
11.10–11.22	FormAI, Minal Suresh Patil
11.22–11.35	iSecure, Alessio Bucaioni
11.35–11.48	AORTA, Ali Balador
11.48–12.00	Robust wireless infrastructure, Aamir Mahmood
12.00–13.00	LUNCH
13.00–15.00	Fortsättning

STÄLL DIN FRÅGA

Menti: 1688 4975

Projekt

FormAI,
Minal Suresh Patil

Menti.com 1688 4975

FormAI AT SCANIA

Formally Verified AI-generated software

Minal Suresh Patil

Researcher

23rd January 2025

World of Volvo, Gothenburg

SCANIA IN BRIEF

FOUNDED
1891

100+
Countries

1,500
Service Points

56,927
employees

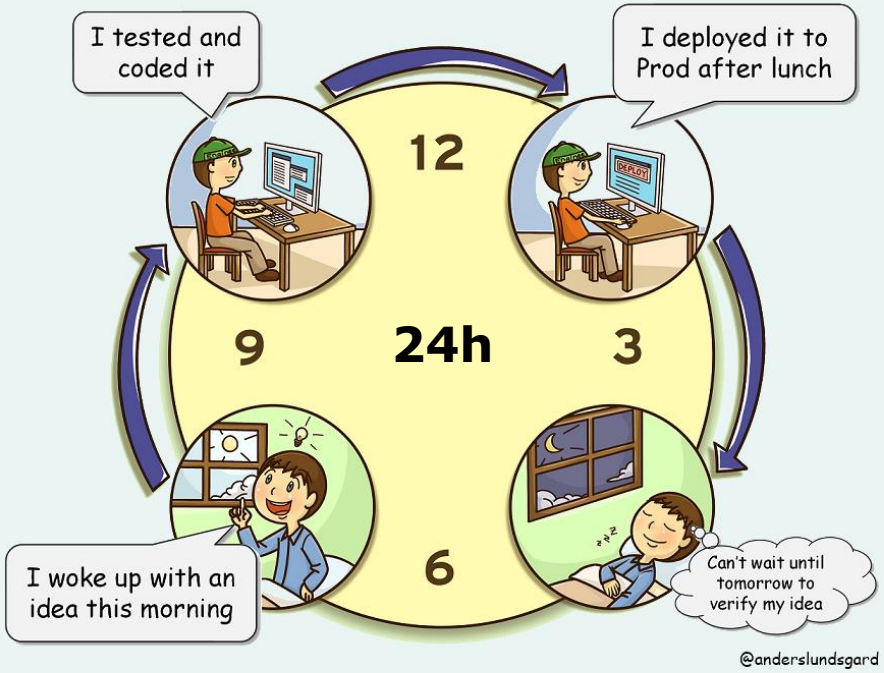
R&D spend
9,909
SEK m.

NET SALES
170,004
SEK m.

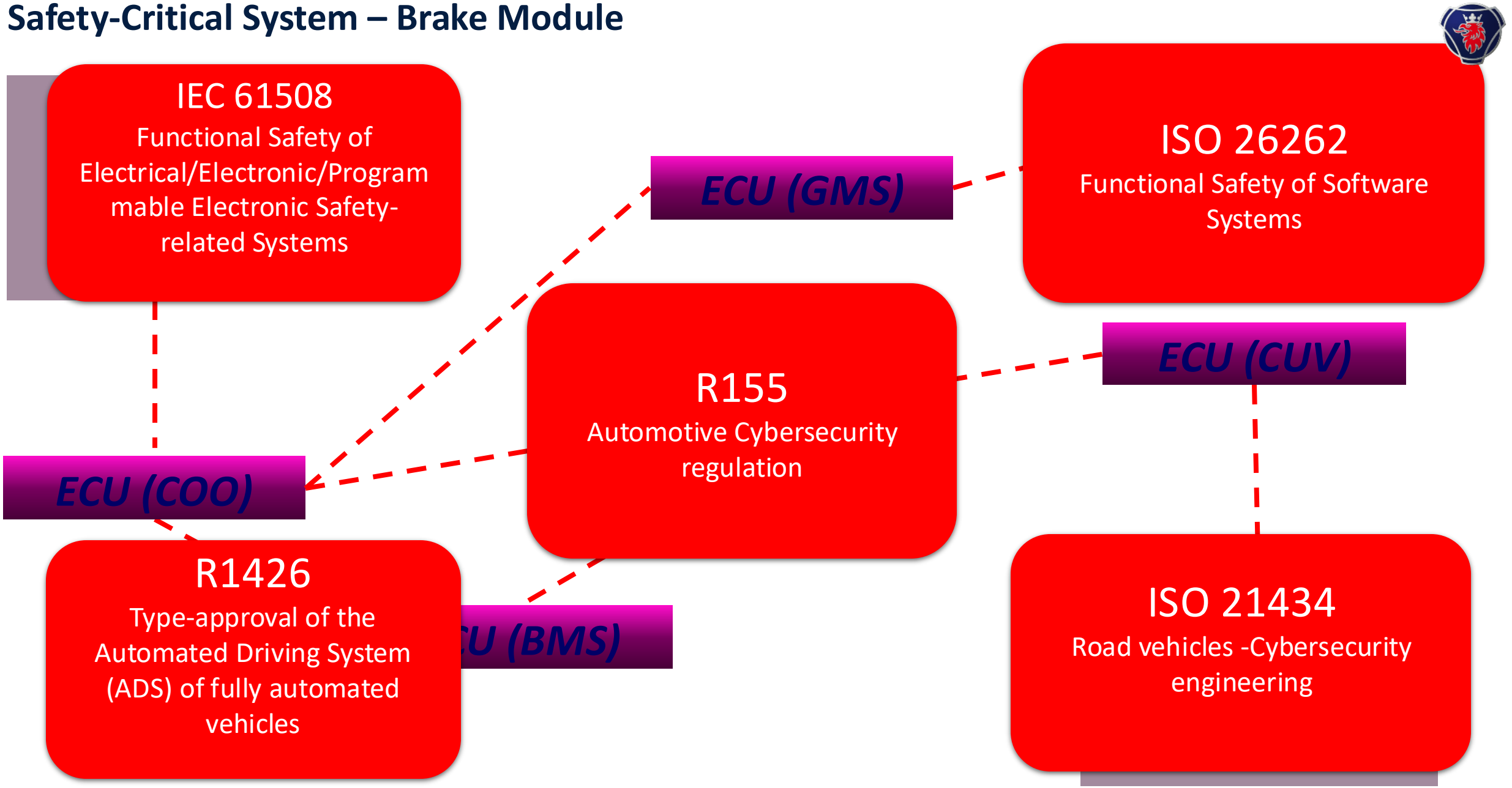
Last red figures
1934

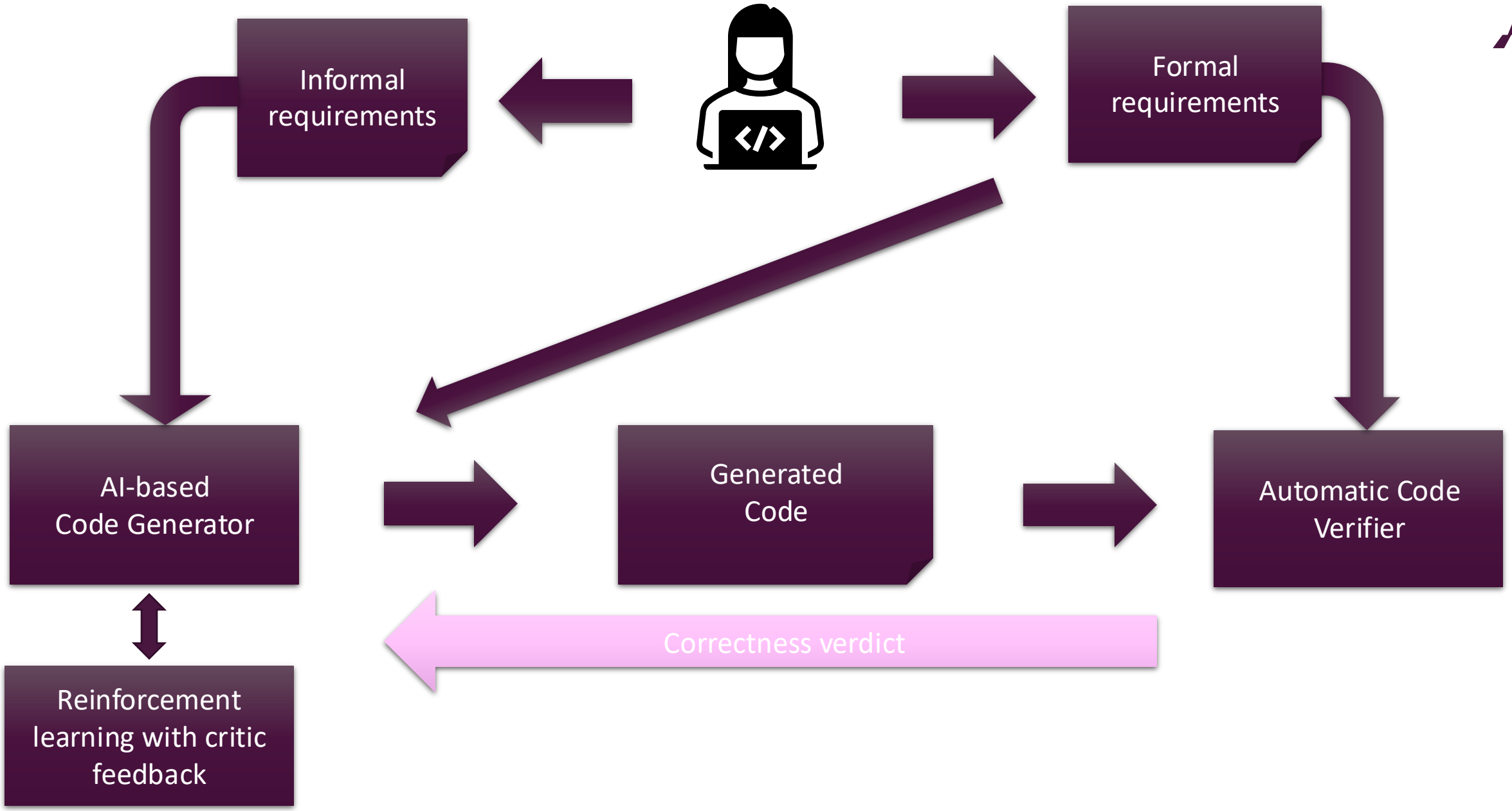
2022

Software Development from Years to Hours

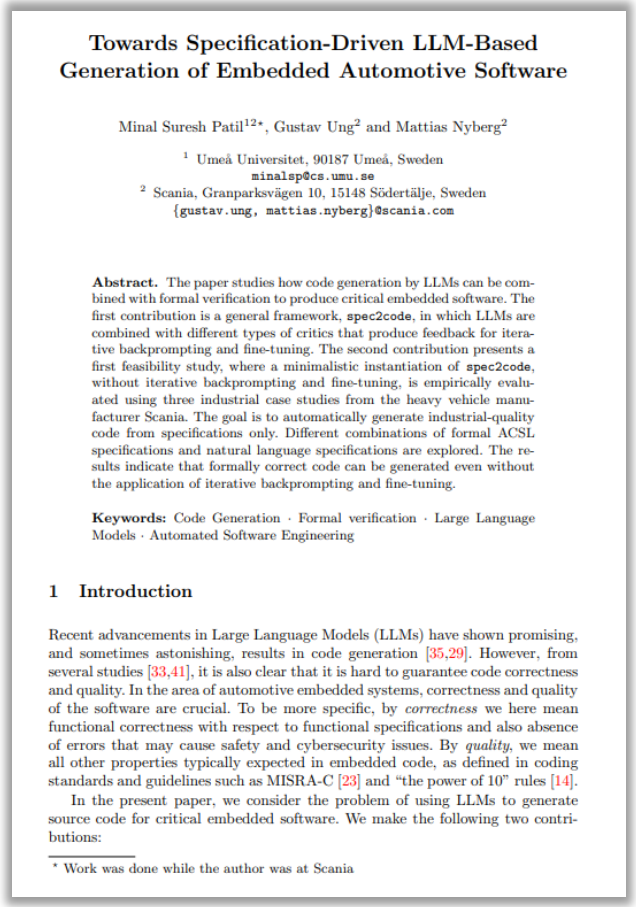


Safety-Critical System – Brake Module




Informal requirements
Formal requirements
AI-based Code Generator
Generated Code
Automatic Code Verifier
Correctness verdict
Reinforcement learning with critic feedback

Publications



Towards Specification-Driven LLM-Based Generation of Embedded Automotive Software

Minal Suresh Patil[1,2]*, Gustav Ung[2] and Mattias Nyberg[2]

[1] Umeå Universitet, 90187 Umeå, Sweden
minalsp@cs.umu.se
[2] Scania, Granparksvägen 10, 15148 Södertälje, Sweden
{gustav.ung, mattias.nyberg}@scania.com

Abstract. The paper studies how code generation by LLMs can be combined with formal verification to produce critical embedded software. The first contribution is a general framework, `spec2code`, in which LLMs are combined with different types of critics that produce feedback for iterative backprompting and fine-tuning. The second contribution presents a first feasibility study, where a minimalistic instantiation of `spec2code`, without iterative backprompting and fine-tuning, is empirically evaluated using three industrial case studies from the heavy vehicle manufacturer Scania. The goal is to automatically generate industrial-quality code from specifications only. Different combinations of formal ACSL specifications and natural language specifications are explored. The results indicate that formally correct code can be generated even without the application of iterative backprompting and fine-tuning.

Keywords: Code Generation · Formal verification · Large Language Models · Automated Software Engineering

1 Introduction

Recent advancements in Large Language Models (LLMs) have shown promising, and sometimes astonishing, results in code generation [35,29]. However, from several studies [33,41], it is also clear that it is hard to guarantee code correctness and quality. In the area of automotive embedded systems, correctness and quality of the software are crucial. To be more specific, by *correctness* we here mean functional correctness with respect to functional specifications and also absence of errors that may cause safety and cybersecurity issues. By *quality*, we mean all other properties typically expected in embedded code, as defined in coding standards and guidelines such as MISRA-C [23] and "the power of 10" rules [14].

In the present paper, we consider the problem of using LLMs to generate source code for critical embedded software. We make the following two contributions:

* Work was done while the author was at Scania

AISoLA, 2024

VeCoGen: Automating Generation of Formally Verified C Code with Large Language Models

Merlijn Sevenhuijsen
Scania & KTH Royal Institute of Technology
Södertälje, Sweden
merlijn.sevenhuijsen@scania.com

Khashayar Etemadi
KTH Royal Institute of Technology
Stockholm, Sweden
khaes@kth.se

Mattias Nyberg
Scania & KTH Royal Institute of Technology
Södertälje, Sweden
mattias.nyberg@scania.com

***Abstract*—Large Language Models (LLMs) have demonstrated impressive capabilities in generating code, yet they often produce programs with flaws or deviations from intended behavior, limiting their suitability for safety-critical applications. To address this limitation, this paper introduces VeCoGen, a novel tool that combines LLMs with formal verification to automate the generation of formally verified C programs. VeCoGen takes a formal specification in ANSI/ISO C Specification Language (ACSL), a natural language specification, and a set of test cases to attempt to generate a program. This program-generation process consists of two steps. First, VeCoGen generates an initial set of candidate programs. Secondly, the tool iteratively improves on previously generated candidates. If a candidate program meets the formal specification, then we are sure the program is correct. We evaluate VeCoGen on 15 problems presented in Codeforces competitions. On these problems, VeCoGen solves 13 problems. This work shows the potential of combining LLMs with formal verification to automate program generation.**

***Index Terms*—Code Generation, Large Language Models, Formal Verification, Iterative Code Improvement.**

I. Introduction

Large Language Models (LLMs) have demonstrated versatility, excelling in various tasks [1]–[4]. One of the tasks where LLMs perform well is the generation of programs [5]–[7]. However, despite their impressive capabilities, LLMs often produce programs with errors or inconsistencies, making them unsuitable for applications requiring high assurance of correctness [8]. This lack of trustworthiness poses a significant challenge to safety-critical domains where the correctness of programs is imperative. Even minor software defects can have severe consequences in the safety-critical domains, such as financial losses or threats to human life [9]–[11]. To address the lack of trustworthiness, the present paper introduces a new tool named VeCoGen, which combines LLMs with formal verification techniques to automatically generate C programs that are correct with respect to given specifications.

VeCoGen is based upon a novel two-step process of initial code generation and iterative code improvement through feedback from a compiler and a formal verifier. In the initial code generation step, VeCoGen generates an initial set of program candidates based on natural language in English and formal specifications in ANSI/ISO C Specification Language (ACSL) [12]. The Weakest Precondition (WP) and Runtime Error (RTE) plugins of Frama-C [13] then verify the correctness of the program candidates. If all generated program candidates fail compilation or verification, VeCoGen continues to the iterative code improvement step. Here, VeCoGen parses the feedback from the compiler and verifier to guide the LLM in generating improved candidates. VeCoGen ensures that the generated program candidate is not only syntactically valid but also formally correct with respect to the formal specification. The tool can be downloaded from https://anonymous.4open.science/r/Vecogen-3008/.

Traditional works, not utilizing LLMs, have addressed the challenge of generating programs automatically [14], but they often face scalability issues [15], [16]. LLMs offer a promising solution to the scalability issues of generating a program that meets the specifications, as explored in prior research. Mukherjee et al. [17] employ LLMs in a semi-automatic framework that uses human intervention to refine and verify generated C programs. Similarly, Patil et al. [18] propose *spec2code*, a framework that combines LLMs with critics to iteratively synthesize programs. However, these existing approaches either rely on manual feedback to the LLM or do not have a tool that implements the automatic generation. In contrast, VeCoGen is the first LLM-based tool that fully automatically generates and verifies C code.

We evaluate VeCoGen on 15 competitive programming problems to assess its effectiveness in generating formally verified C programs. VeCoGen solves 13 out of 15 problems, demonstrating its ability to generate formally verified code. This initial benchmarking showcases the potential of generating formally verified C code automatically using VeCoGen, potentially allowing for use in safety-critical software development.

The paper contains the following contributions:

1) VeCoGen, a novel LLM-based code generation tool for iteratively generating formally verified C code.
2) The evaluation of VeCoGen on VeCoSet, a collection of 15 competitive programming problems.
3) An analysis of the impact of changing the configuration of the tool, i.e. type of specification used, number of generated programs in each iteration, temperature, zero- or one-shot prompting, and LLM used.

The rest of this paper is organized as follows. Section II provides background on formal verification and LLM-based code generation. Section III describes the design and implementation of VeCoGen. Section IV outlines the experimental

FormaliSE, 2025

Lessons learned and future work

- Promising results on **real** industrial case studies
- The need and importance of writing **complete** requirements
- How do we benchmark AI-generated safety-critical code?
- Explore adapting the AI-code generator to internal documentation, coding standards, and domain-specific resources for specialized contexts


THANK YOU

Projekt

iSecure,
Alessio Bucaioni

Menti.com 1688 4975

iSecure: Developing Predictable and Secure IoT for Autonomous Systems

Alessio Bucaioni

Project name

iSecure: Developing Predictable and Secure IoT for Autonomous Systems

Presenter and project coordinator

Alessio Bucaioni, Associate Professor at Mälardalen University

Participants

MDU, Addiva, CanaryBit, Sensair, Västerås Flygplats, and Västerås Mälarhamnar

Start and end date

October 2023 – September 2026

Contact

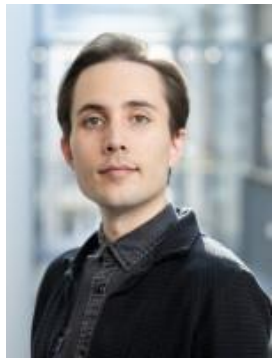
alessio.bucaioni@mdu.se

Karl Williams
Senseair

Björn Lindström
Addiva

Stefano Cucchiella
CanaryBit

Alessio Bucaioni
MDU



Sebastian Leclerc
MDU

Benoit Wastine
Senseair

Hana Oden
Västerås Mälarhamnar

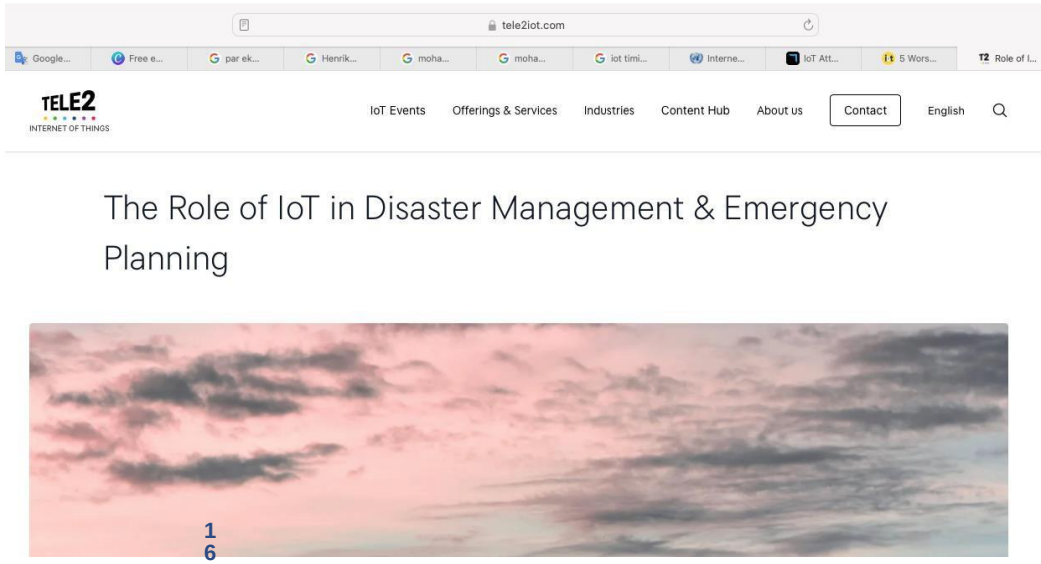
Par Ekman
Västerås Flygplats

Mohammad Ashjaei
MDU

Moris Behnam
MDU

Industrial challenges (i)

Current digital platform architectures for IoT lack of timing guarantees for short-latency data communication



Industrial challenges (ii)

Proper mechanism for data privacy and security

Relations to the call

iSecure support for heterogeneous connectivity requirements, high security of the connected devices and response time requirements

Secure digital platforms for efficient development, production and support

Edge computing, fog and cloud technologies and wireless networks and software-defined networks for industrial applications

Results (so far)

Systematic study of time critical IoT

Systematic study of over 600 publications on temporal predictability and determinism within IoT applications

Testbed

Installed tens[1] of different sensors in four different locations over Västerås municipality

Results (to achieve)

Edge-cloud architecture

A secure edge-cloud architecture with dynamic and time-predictable communication for IoT systems in industrial environments

Data sharing platform

A confidential data sharing platform for devices, systems and services

Proof-of-concept implementations

Proof-of-concept implementations in two use cases namely autonomous airport and harbor.

Thank you!
Do you have any questions?

“What we know is a drop. What we don’t know is an ocean.”

- Isaac Newton

Projekt

AORTA,
Ali Balador

Menti.com 1688 4975

AORTA: Advanced Offloading for Real-Time Applications

Welcome to the presentation on AORTA, a national project focusing on real-time robotic applications. This presentation explores the motivation, challenges, solutions, and future of this transformative initiative.

Project Duration: February 2023 - January 2026

Ali Balador, Senior Researcher, Ericsson Research, Stockholm

ali.balador@ericsson.com

January 23, 2025

VINNOVA
Sveriges innovationsmyndighet

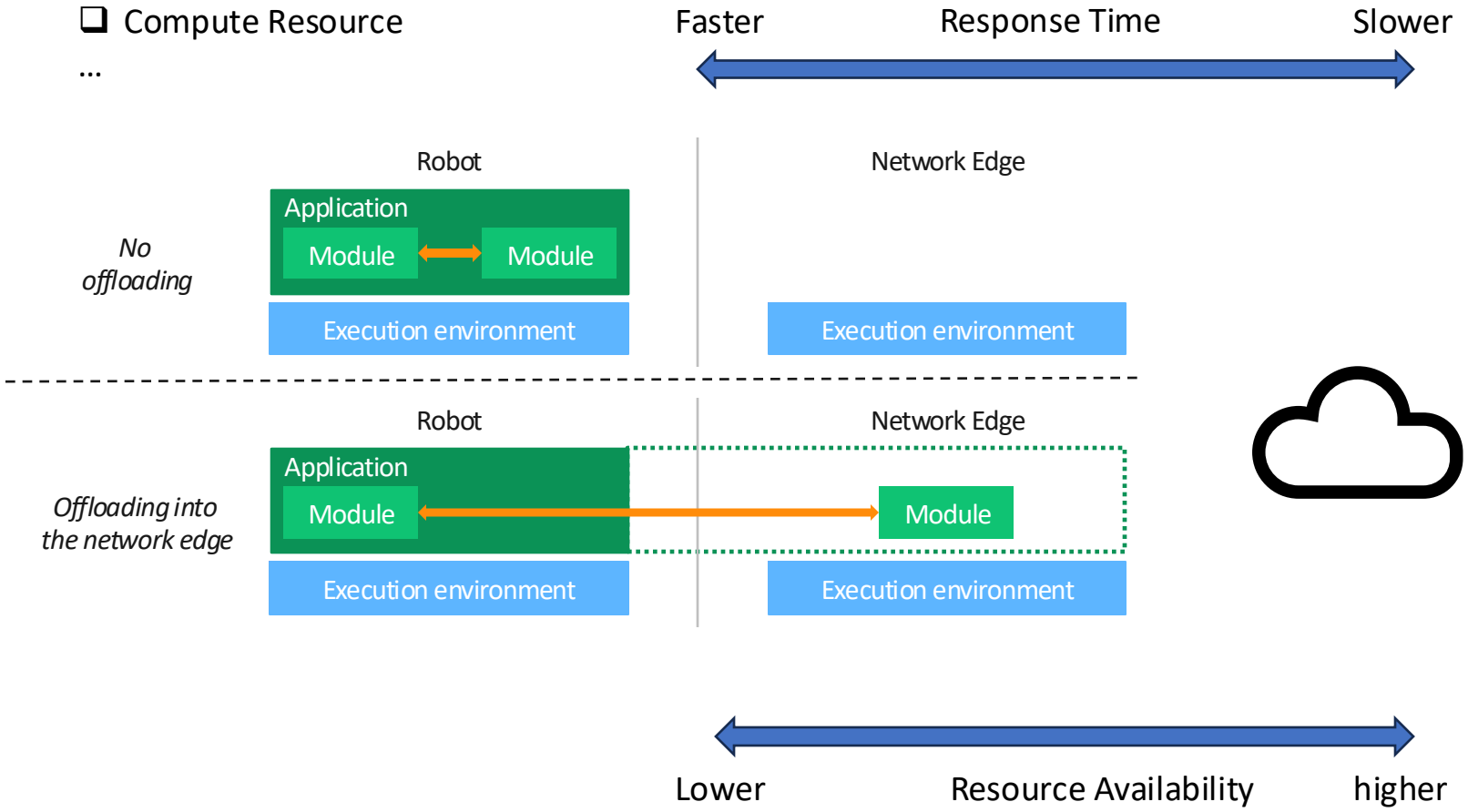
Avancerad Digitalisering

Why Offloading?

Requirements:

- ❑ Energy
- ❑ Compute Resource

…



The Need for Dynamic Offloading

Mobile Robotics

Communication quality varies in difference areas and compute needs to be adjusted accordingly to support real-time applications.

Flexibility

Provide more possibilities for software updates and maintenance maintenance remotely.

Edge Computing

Limited resources at the edge necessitate intelligent offloading. offloading.

Cost

Cost of using edge and cloud resources makes it uneconomical to always uneconomical to always offload.

AORTA: Bridging the Gap

1 **Dynamic Offloading**

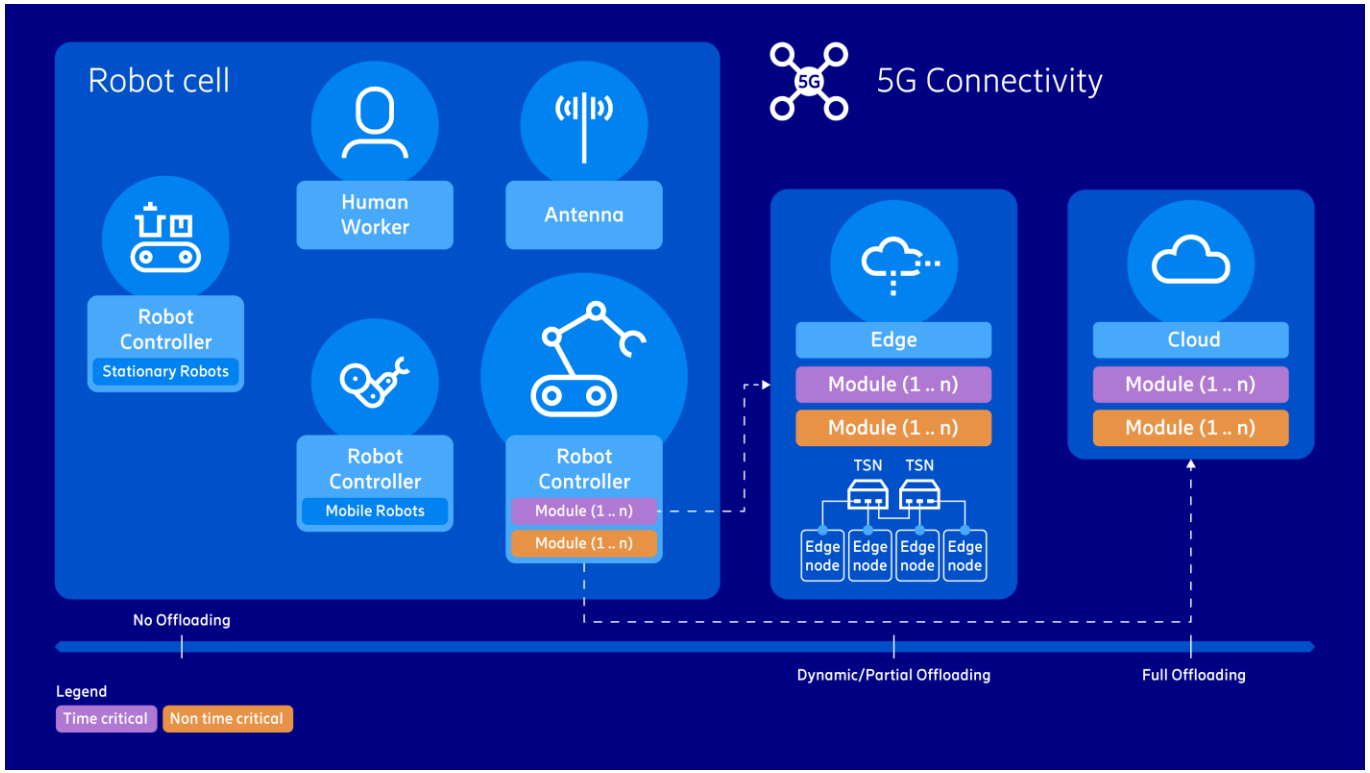
Intelligent decision-making to allocate resources efficiently. Decide where and when to offload.

2 **Enhanced Performance**

Reduced latency, improved responsiveness, and optimized utilization. utilization.

3 **Secure and Scalable**

AORTA ensures robust security and seamless scalability for diverse applications.

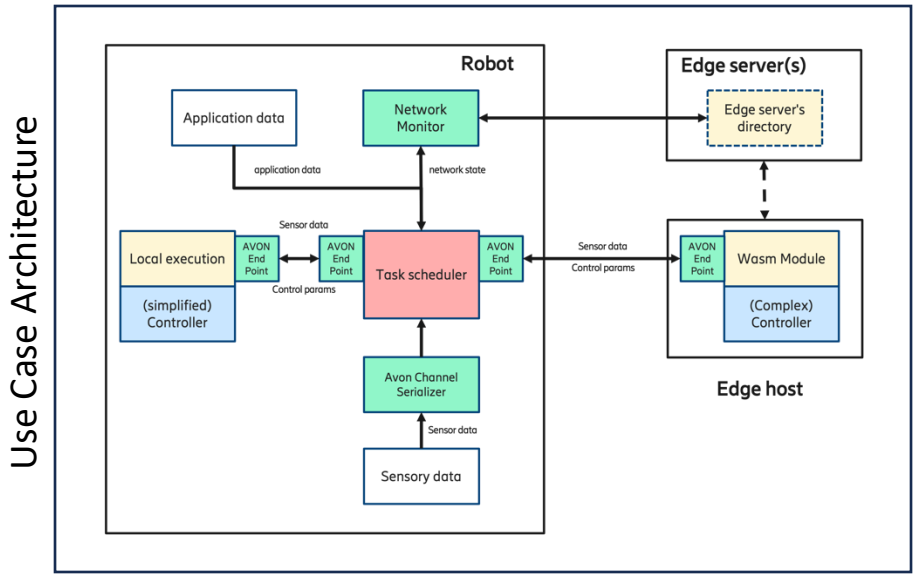


WEBASSEMBLY

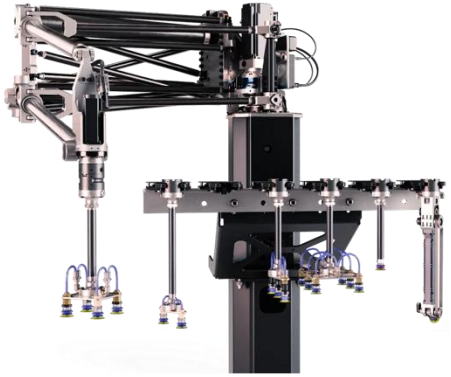
AORTA in Action: Use Case Examples

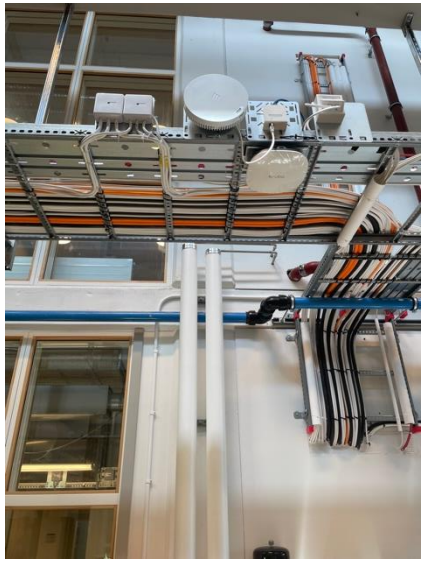
HKM Pick-and-Place Robot

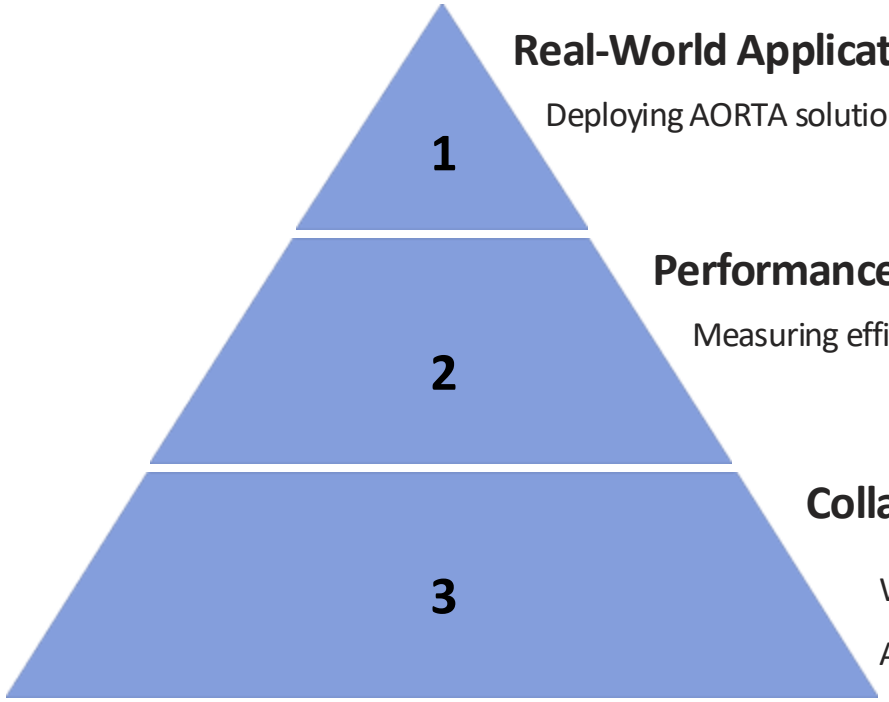
Offloads complex geometric pose planning calculations to the edge and cloud, and cloud, enabling real-time adjustments and improved precision in robotic arm robotic arm movements. This removes the need to have very power full onboard onboard compute capability and save costs.



Project Testbed







Real-World Applications

Deploying AORTA solutions in various scenarios within robotic domain.

Performance Evaluation

Measuring efficiency, latency, and scalability in realistic settings.

Collaborative Research

Working with industry partners to refine and optimize AORTA.

Future Directions and Use Cases

1

Collaborative use cases

Use cases where several robots collaborate including mobile robots.

2

WebAssembly for more complex applications

Use WASM in scenarios where we have multi-threading applications and specialized hardware (e.g., GPUs).

3

Improved decision-making algorithms

Improve decision-making algorithms considering both communication and compute resources.

COGNIBOTICS

ERICSSON

LUNDS UNIVERSITET

Mälardalens universitet

Projekt

Robust wireless infrastructure,
Aamir Mahmood

Menti.com 1688 4975

Robust trådlös infrastruktur för fjärrstyrd virkeslastning

Robust wireless infrastructure for remote timber management

Aamir Mahmood, Associate Professor

Department of Computer & Electrical Engineering

Mid Sweden University

Mittuniversitetet MID SWEDEN UNIVERSITY · Telia · ERICSSON · SCA · BIOMETRIA · VOLVO · skogforsk · VINNOVA Sveriges innovationsmyndighet

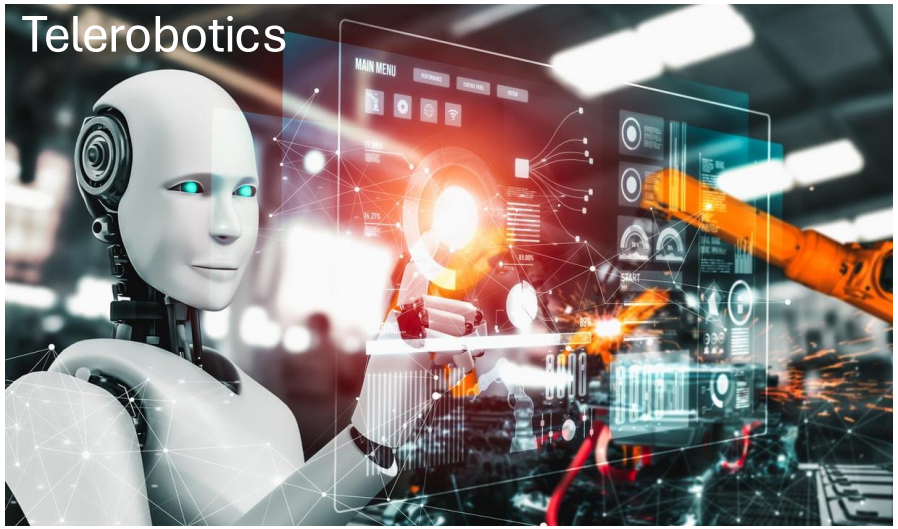
Agenda

- Teleoperation – Driving digitalization across industry verticals

- Past and Present of “Robust wireless …” project
  - Remote Timber (Vinnova, 2019-2022) – Teleoperation in forestry digital value chain
  - Robust Wireless (Vinnova, 2023-2025)

- Future of teleoperation
  - Lessons from the Past, Vision for 6G

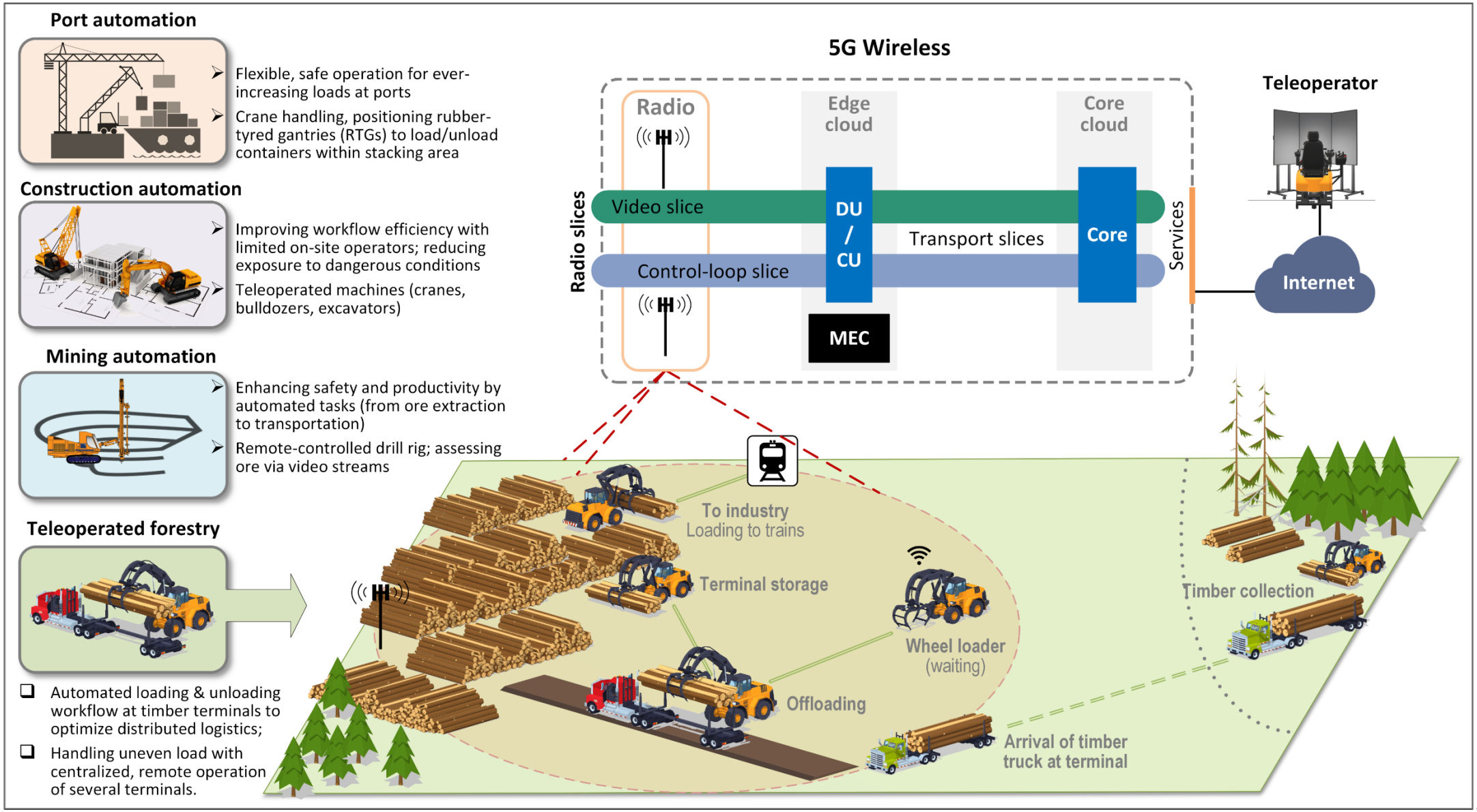
Teleoperation



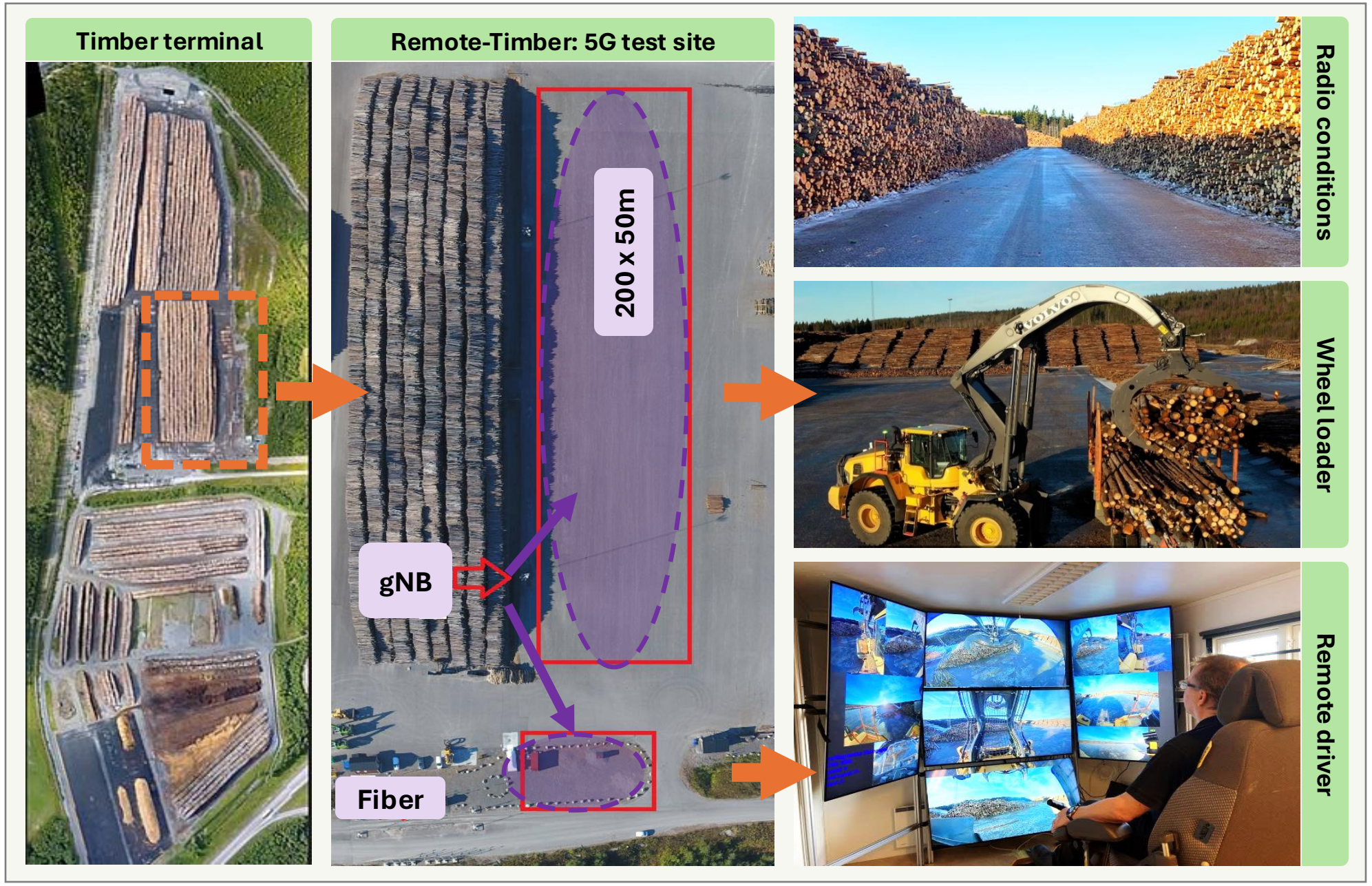
https://ukplantoperators.com/dx225lc-7x-teleoperation-showpiece-on-doosan-bauma-stand/
https://im-mining.com/tag/teleremote/

https://www.plantandequipment.news/news/product-updates/volvo-ce-tests-worlds-first-high-lift-tele-operation-over-5g/
https://www.solitonsystems.com/low-latency-video/remote-operation/telerobotics

Remote-Timber (VINNOVA, 2019—2022)

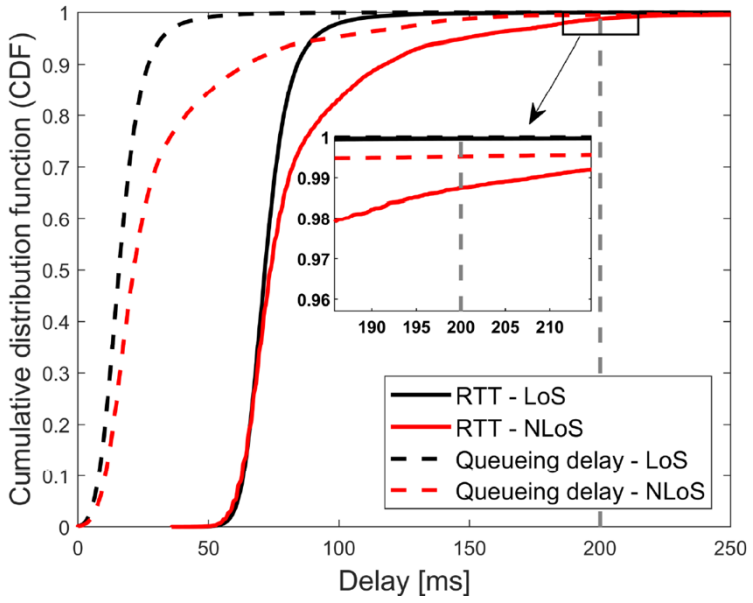


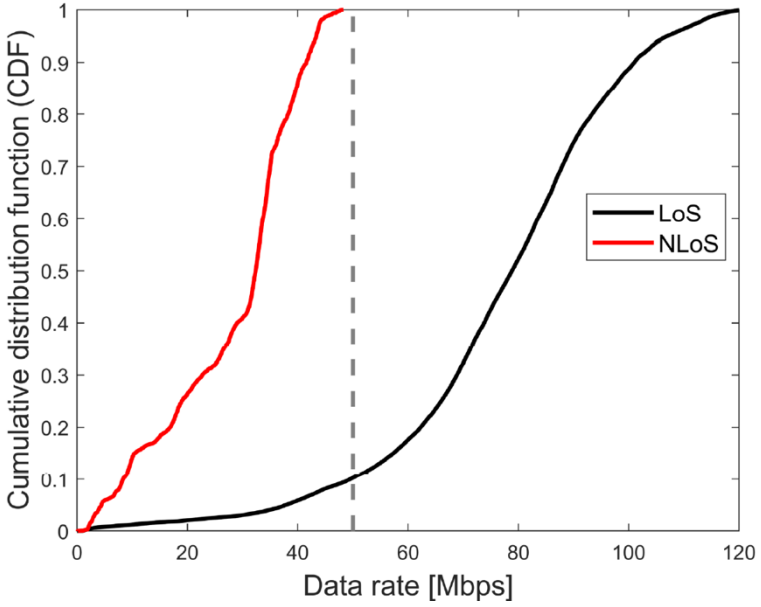
Remote Timber– Demo I



Results and Lessons Learned

- To serve one vehicle in a constrained space, over-provisioning of radio resources was needed to satisfy latency and throughput needs.
- Performance quickly degraded with distance from the base station
- **Stationary terminal; piles are not**– A flexible network infrastructure that could handle coverage needs in a dynamic timber layout
- Handling connectivity needs within a constrained set of radio resources

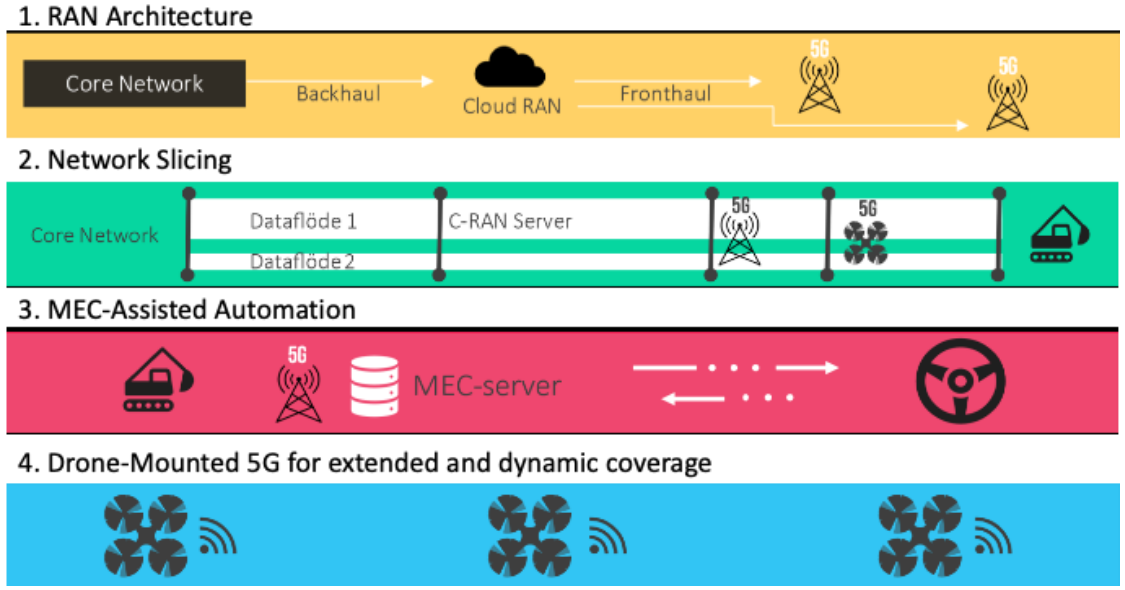




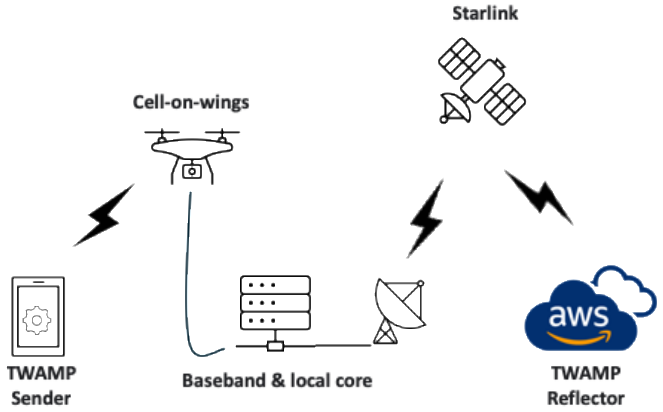
Robust wireless infrastructure for remote timber management (Vinnova, 2023 - 2025)

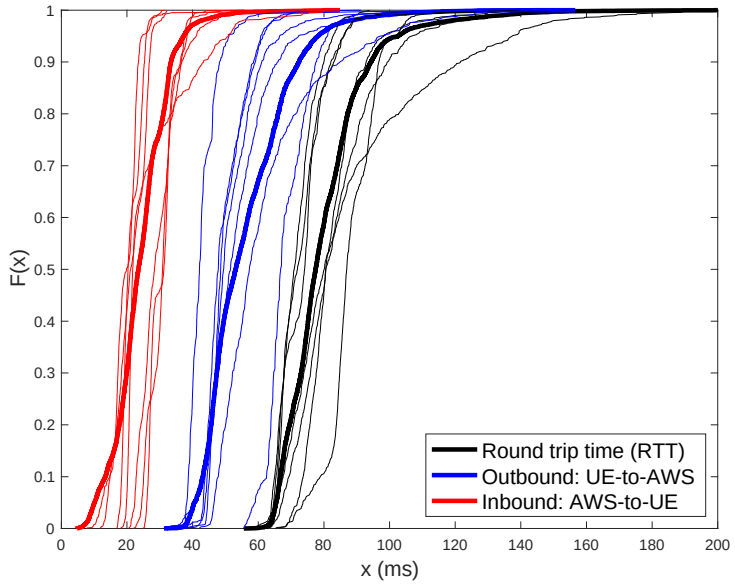
Project objectives and progress

- Build a RAN architecture in a demonstrator environment where the data flow can be guaranteed to 100% over the surface through seamless simultaneous connection.
- Create 95% coverage over a demonstrator environment modeled after a lumber terminal through drone-mounted base stations.
- Produce a demonstrator that can achieve SLA requirements with a 50% reduction in resource usage compared to Remote Timber.



Cell-on-Wings: Demo II

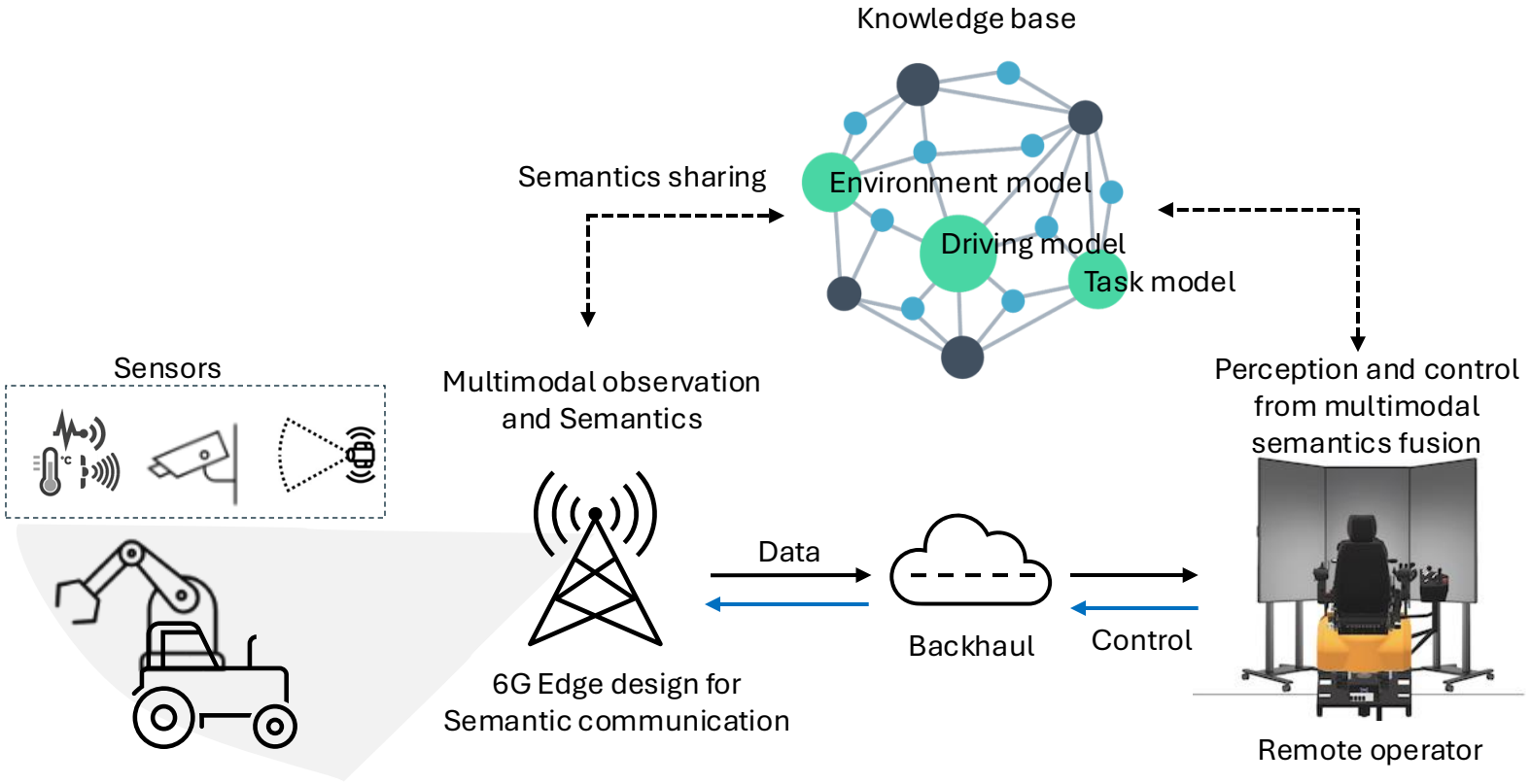


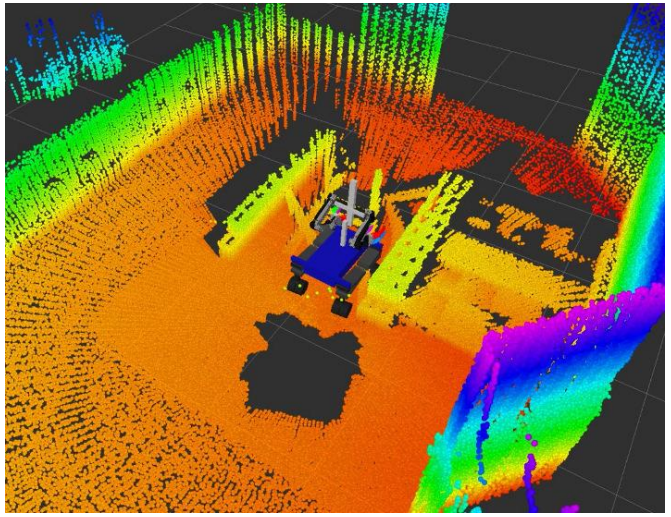


Measuring latency performance using Two-Way Active Measurement Protocol (TWAMP).


5G-utrustad drönare
fjärrstyr skogsmaskin

Future of Teleoperation





Vs

Future of Teleoperation

Yet Another Project?

- Why is 5G insufficient, and if 6G can be the savior?
  - Industry verticals need holistic approaches that address their unique requirements beyond capacity.

- Industry verticals adopt new technology with cost-benefit analysis
  - Not by chasing hype or relying solely on a research-driven focus on innovations.

- ***6G will still be insufficient***
  - If the application context and AI/ML are not correctly integrated into the co-design for communication and control for industry verticals,

Project partners

Mittuniversitetet
MID SWEDEN UNIVERSITY

BIOMETRIA

SCA

ERICSSON

Telia

skogforsk

VOLVO

VINNOVA
Sveriges innovationsmyndighet

IVA
Royal Swedish Academy of Engineering Sciences

Robust trådlös infrastruktur för fjärrstyrd virkeshantering

Vårt forskningsprojekt syftar till att lösa problemet med begränsad 5G-täckning på avlägsna virkesterminaler och andra svåråtkomliga områden. Dessa platser saknar ofta tillräcklig infrastruktur för att möjliggöra fjärrstyrning av tunga skogsmaskiner, vilket påverkar både effektivitet och säkerhet negativt.

Summering

Lunch

Entréplan, vi ses kl 13:00