Age-Based Metrics for Joint Control and Communication in Cyber-Physical Industrial Systems

Maia de Sant Ana, Pedro

DOI (link to publication from Publisher):
10.54337/aau548863844

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

AGE-BASED METRICS FOR JOINT CONTROL AND COMMUNICATION IN CYBER-PHYSICAL INDUSTRIAL SYSTEMS

BY
PEDRO MAIA DE SANT ANA

DISSERTATION SUBMITTED 2023

AALBORG UNIVERSITY
DENMARK

Age-Based Metrics for Joint Control and Communication in Cyber-Physical Industrial Systems

Ph.D. Dissertation

Pedro Maia de Sant Ana

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7B
DK-9220 Aalborg

Dissertation submitted: July 2023

PhD supervisor: Professor Petar Popovski
Aalborg University

PhD committee: Associate Professor Gilberto Berardinelli (chairman)
Aalborg University, Denmark

Senior Principal Scientist Zhibo Pang
ABB Corporate Research, Sweden

Professor Pablo Ameigeiras Gutiérrez
University of Granada, Spain

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-674-4

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2023

Abstract

This thesis delves into the emerging paradigm of goal-oriented communication, particularly in the context of wireless networked control systems. Moving away from traditional communication metrics, we argue for a shift in focus towards the ultimate purpose or 'goal' of the communication process. This paradigm envisages the network and control applications as interconnected parts of a singular system, orchestrated towards a definitive purpose. We explore strategies for the network to learn about the control system's behavior, optimizing resource allocation in the process, and for the control system to adapt its parameters using network information to enhance stability and performance. To support this paradigm shift, we introduce the Age of Loop (AoL), a novel concept designed to offer a more holistic view of the control loop communication process. The AoL metric extends the application of the Age of Information (AoI) by integrating both downlink and uplink effects and their interplay, setting the stage for more effective resource allocation strategies.

Our investigation includes theoretical analyses and practical experimentation. We embark on a measurement campaign utilizing a standalone 5G network in a Bosch factory setting to measure and analyze AoL behavior, demonstrating its practical advantages over AoI. Furthermore, we illustrate the practical application of our proposed methodologies through an Automated Guided Vehicle (AGV) use-case, showing that theoretical solutions can be successfully adapted to real-world scenarios. In summary, our work propels the goal-oriented communication paradigm forward, offering innovative strategies and tools, and providing a compelling case for rethinking traditional network design principles. This study sets a robust foundation for future research and practical implementations, promising a significant impact on the future design and operation of wireless networked control systems.

Resumé

Denne afhandling dykker ned i det nye paradigme kendt som målrettet kommunikation, specifikt inden for trådløse og netværksstyrede kontrolsystemer. Vi argumenter for at skift i fokus fra traditionelle kriterier og mod at fokus direkte for målene for kommunikationsprocessen. I dette paradigme anses både netværket og kontrolsystemerne som forbundne dele af et enkelt system, der så styres mod et bestemt formål. Vi udforsker strategier for at netværket kan lære om kontrolsystemets opførsel, og derved optimere allokering af resurser, samt for at kontrolsystemet kan tilpasse sine parametre ved hjælp af netværksinformation of derved forbedre stabilitet og ydeevne. For at understøtte dette paradigmeskift, introducerer vi Age of Loop (AoL), et nyt begreb der giver et holistisk syn på kontrolløkken for kommunikationsprocessen for. AoL er en udvidelse af Age of Information (AoI), og begrebet integrerer både downlink og uplink effekter samt deres indbyrdes påvirkninger, hvilket baner vejen for mere effektive strategier indenfor allokering af resurser. Vores afhandling omfatter både teoretiske analyser samt praktisk eksperimentering. Vi har gennemført en målekampagne ved hjælp af et selvstændigt 5G-netværk i et fabriksmiljø ved Bosch for at måle og analysere AoL, samt for at demonstrere dets praktiske fordele over AoI. Desuden illustrerer vi praktisk anvendelse af vores udviklede metoder gennem eksperimentering med et Automated Guided Vehicle (AGV), hvilket viser, at vores teoretiske løsninger kan tilpasses virkelige scenarier. Vores arbejde rykker paradigmet indenfor for målrettet kommunikation betydeligt, foreslår nye og innovative strategier og værktøjer, og giver et stærkt grundlag for at genoverveje traditionelle netværksdesignprincipper. Afhandlingen lægger et solid grundlag for fremtidig forskning og praktiske implementeringer, der vil få en betydelig indflydelse på fremtidige designs og drift af trådløse og netværksstyrede kontrolsystemer

Contents

Thesis Details

Thesis Title:	Age-Based Metrics for Joint Control and Communication in Cyber-Physical Industrial Systems
Ph.D. Student:	Pedro Maia de Sant Ana
Supervisors:	Prof. Petar Popovski, Aalborg University Assoc. Prof. Beatriz Soret, Aalborg University Dr. Nikolaj Marchenko, Bosch Corporate Research

The main body of this thesis consist of the following papers.

[A] de Sant Ana, Pedro M., Nikolaj Marchenko, Petar Popovski, and Beatriz Soret. "Wireless control of autonomous guided vehicle using reinforcement learning." In *IEEE Global Communications Conference*, pp. 1-7. IEEE, 2020.

[B] de Sant Ana, Pedro M., and Nikolaj Marchenko. "Radio Access Scheduling using CMA-ES for Optimized QoS in Wireless Networks." In *IEEE Globecom Workshops*, pp. 1-6. IEEE, 2020.

[C] de Sant Ana, Pedro M., Nikolaj Marchenko, Petar Popovski, and Beatriz Soret. "Age of loop for wireless networked control systems optimization." In *IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-7. IEEE, 2021.

[D] de Sant Ana, Pedro M., Nikolaj Marchenko, Petar Popovski, and Beatriz Soret. "Control-Aware Scheduling Optimization of Industrial IoT." In *IEEE 95th Vehicular Technology Conference*, pp. 1-6. IEEE, 2022.

[E] de Sant Ana, Pedro M., Nikolaj Marchenko, Beatriz Soret, and Petar Popovski. "Goal-Oriented Wireless Communication for a Remotely Controlled Autonomous Guided Vehicle." *IEEE Wireless Communications Letters* , 2023.

[F] de Sant Ana, Pedro M., Nikolaj Marchenko, Beatriz Soret, and Petar Popovski, "Experimental Study of Information Freshness for Goal-Oriented Wireless Communication in a Factory." *IEEE Internet of Things Journal*, 2023 - *submitted*.

[G] de Sant Ana, Pedro M., Nikolaj Marchenko, Beatriz Soret, and Petar Popovski, "Goal-Oriented Wireless Communication and Control using Age of Loop" *IEEE Communication Magazine*, 2023 - *submitted*.

In addition to the main papers, the following patents, which are not included in the thesis, have been submitted during the Ph.D. studies:

[1] de Sant Ana, Pedro M., and Nikolaj Marchenko, "A Method for Efficient Radio Resource Allocation based on Application QoS Requirements," European patent, submitted 2020.

[2] de Sant Ana, Pedro M., and Nikolaj Marchenko, "A Method for Improved RF-Prediction and Planning based on Ray Tracing Enhanced with Deep Learning Neural Networks," European patent, submitted 2020.

[3] de Sant Ana, Pedro M., and Nikolaj Marchenko, "Control-Aware Scheduling Method and Apparatus," European patent, submitted 2022.

[4] de Sant Ana, Pedro M., and Nikolaj Marchenko, "Adaptive Network using Pilot-Assisted Image Transmission," European patent, submitted 2022.

[5] de Sant Ana, Pedro M., and Nikolaj Marchenko, "RF-based gesture recognition for COVID-19 monitoring and detection," European patent, submitted 2022.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Acknowledgement

I am deeply grateful for the steadfast support and guidance I received during the rigorous journey of my PhD studies. First and foremost, I wish to express my profound gratitude to Petar Popovski, Beatriz Soret, and Nikolaj Marchenko. Their unerring confidence in me and the freedom they extended shaped my intellectual growth, driving me to explore the realms of creativity in multiple projects. Their invaluable mentorship over these challenging three years has been instrumental in the successful completion of my studies.

I would also like to express my appreciation to the members of my assessment committee for taking their time to evaluate my work.

Finally, I wish to extend my sincerest thanks to Prof. Vicente A. de Sousa Jr. It was his tutelage and guidance that initiated my journey in the field of communication engineering. His influence has been pivotal in shaping my career trajectory, for which I am eternally grateful.

To each of you, I extend my sincere and heartfelt thanks for making this journey not just possible, but also intellectually rewarding and personally enriching.

Pedro Maia de Sant Ana
Aalborg University, July 3, 2023

Part I

Introduction

Introduction

1 Motivation

Communication systems, as we traditionally understand them, have been dominantly focused on the reliable exchange of data. The design and optimization of these networks is guided by specific Key Performance Indicators (KPIs), such as data throughput and latency. While these KPIs are undoubtedly important, they do not fully encapsulate the nuanced demands of more complex environments [6, 19]. This becomes particularly evident in industrial settings, where machines, sensors, and control systems must continuously interact in real-time to execute single or collaborative tasks, some of which may be critical. In such environments, ensuring smooth operation of these systems may necessitate more than just adhering to predefined link-level KPIs [11]. Particularly when considering the finite network resources at disposal, the establishment of a communication paradigm capable of fulfilling these requirements may warrant a shift in perspective, favoring more context-aware metrics and strategies that can account for the real-time and collaborative nature of tasks within these complex settings.

This is where a new paradigm known as goal-oriented communication is coming to the fore. The premise of goal-oriented communication is quite a shift from the traditional outlook [16]. Instead of concentrating solely on data transmission, the focus moves towards understanding the ultimate purpose or 'goal' of that communication. This paradigm goes beyond the "how" of communication - how to transmit data as efficiently as possible. Instead, the main idea is to take into account the "why" - why the data is being communicated and what desired outcome of that communication is.

The paradigm of goal-oriented communication compels us to rethink our foundational approach to network design [11]. Rather than treating control applications and the network as distinct entities, they are perceived as constituents of a singular, interconnected system, orchestrated with a definitive purpose. Take, for instance, a factory environment where the primary goal is to sustain fluid and efficient operations. In such a goal-oriented communication scenario, the network could discern a data stream correlated with a vital control system state and prioritize its communication. Should a complication emerge within the control system, the network possesses the agility to

adapt instantaneously, e.g. by rerouting traffic or adjusting parameters to avert disruptions and uphold operations. Concurrently, the control application could deliver constant feedback to the network, thus empowering the network to *learn*, adapt, and progressively enhance its performance.

The transition towards goal-oriented communication pledges two pivotal advantages. Firstly, it changes the way we currently envision network resource allocation. This means that rather than uniformly striving for high throughput and low latency based on preset requirements, the network can intelligently allocate resources attuned to the unique state of the control system application. Secondly, goal-oriented communication could effectively diminish the distance between the network and the applications it underpins, thereby fostering the development of more complex, intelligent systems. Strategies such as machine learning could hold a central role in this context, empowering the network to comprehend application behavior, generate forecasts, and make real-time adjustments.

While the promise of this goal-oriented communication paradigm is undeniable, it is important to acknowledge that it embodies a considerable departure from established communication design principles. It calls for novel strategies and technologies, alongside a re-imagined perspective on network functionality and structure. Foremost, achieving a granular understanding of application effectiveness is pivotal [7]. This necessitates the network's ability to distinguish the specific performance attributes and resource demands of diverse applications. Acquiring such profound comprehension is a challenging task, underlining the potential indispensability of sophisticated machine learning algorithms and models. The objective is to architect a system that not only accumulates and processes pertinent data but also effectively *learns* from and predicts the dynamic requirements of varying applications. Moreover, the transition from traditional network metrics towards application-centric goals introduces a significant dilemma: How do we delineate and measure communication effectiveness in this new context? Conventional network performance indicators like latency, throughput, and packet loss are well-understood, yet the metrics pertinent to goal-oriented communication may fluctuate extensively based on the application. Quantifying these metrics necessitates comprehension of their impact on the application's success. This also demands the creation of innovative measurement tools and methodologies. Machine learning has potential here too, aiding in discerning which metrics hold the most relevance for specific applications and dynamically fine-tuning network parameters to optimize these metrics [11].

In this context, the inclusion of the Age of Information (AoI) and the proposed Age of Loop (AoL) concepts extends the design space and potential strategies for managing the complexities of goal-oriented communication. AoI, which quantifies the freshness of information, offers a promising metric for assessing the relevance and timeliness of information [11]. This attribute could be particularly useful in systems where information freshness is critical, such as real-time monitoring systems, autonomous vehicles, and IoT applications. In fact, recent literature sees numerous researchers exploring AoI as a viable metric for analyzing and enhancing wireless networked control sys-

tems [3, 5, 8, 10, 14]. This trend underscores the emerging shift towards innovative perspectives in network design and assessment, particularly in the realm of control systems research.

Nevertheless, an intrinsic challenge with the application of Age of Information lies in its inclination to restrict optimizations to a single communication link. Consequently, it might inadvertently overlook the complex interplay between downlink and uplink communications, thereby neglecting a holistic overview of the communication process. This becomes particularly impactful in closed-loop control systems, where the importance of feedback-based control signals to system stability and performance is paramount. For instance, in our work [4], we illustrate how the application of AoI may fall short in encapsulating the complexities of these systems, resulting in sub-optimal behavior for closed-loop control. To overcome these limitations, we introduce the Age of Loop concept. AoL aims to integrate both downlink and uplink effects, as well as their interrelation. By doing so, it provides a more holistic understanding of the control loop communication process, laying the groundwork for an optimized resource allocation strategy.

In this thesis, we delve into the goal-oriented communication paradigm, paying particular attention to its application in wireless networked control systems. We design strategies that enable the network to learn about the control system's behavior, thereby optimizing resource allocation. Simultaneously, we explore scenarios where the control system can adapt its parameters by leveraging network information to enhance stability and performance. Both strategies underline an important paradigm initially put forth by Witsenhausen [18]: in the context of distributed control problems, actions taken at the control system level can directly impact the communication system, and vice versa. To accomplish this, we fundamentally rely on age-based metrics, introducing a new concept, the Age of Loop, which propels our results beyond the current state-of-the-art. Moving beyond theoretical and analytical considerations, we also initiate a measurement campaign utilizing a standalone 5G network. This allows us to measure and analyze AoL behavior within a Bosch factory setting, demonstrating its practical superiority over AoI. Moreover, we examine an Automated Guided Vehicle (AGV) use-case to illustrate that solutions conceived within an analytical framework can be readily adapted to the experimental model. This further underscores the applicability and adaptability of our proposed methodologies and concepts.

2 Thesis Objectives and Methodology

The objective of this thesis is to explore and tackle a set of key issues highlighted in the preceding section. Specifically, it intends to shed light on the following three research questions, providing preliminary responses to each:

Q1: How can we construct a comprehensive model that enables the concurrent op-

timization of control and network aspects, facilitating mutual learning between these components within the framework of goal-oriented communication?

Q2: How can we proficiently employ the Age of Information and the newly proposed Age of Loop metrics to critically assess and enhance control system performance, while concurrently optimizing network resource efficiency?

Q3: How can we harness the empirical insights derived from AoL behavior in a real-world 5G network setting to bridge the gap between theoretical constructs and practical applications, and subsequently employ these insights to enhance the operational efficiency of real-time applications, such as AGVs?

The responses to these research questions will be articulated through a series of research articles. Figure 1 offers a visual representation of the interconnected relationship between each research paper and the corresponding research question. This diagram also serves as a roadmap of the methodology deployed to fulfill the objectives of this thesis. Each research question is envisioned as part of a broader topic, with every topic being interrelated as they align toward a specific goal.

The first research question resides within the broader research realm of "Goal-Oriented Communication in Wireless Networked Control Systems." Here, we delve into the interplay between communication and control, pinpoint gaps in existing research, and propose hypotheses accompanied by potential solutions. Specifically, we endeavor to construct comprehensive models that encapsulate both control and network aspects, enabling their joint optimization. Consequently, we design learning-based control strategies that intelligently adapt to network behavior and performance, as well as fine-tune network resource allocation based on the dynamic states of the control system.

Our second research question explores solution methodologies using age-based metrics. We identify the limitations of AoI when modeling WNCS scenarios, which propels us to propose a new metric: the Age of Loop. This metric enables the formulation of learning-based solution strategies for a variety of control problems, showcasing its advantages over AoI and pushing performance results beyond current state-of-the-art standards.

The third research question is centered around the validation of solution strategies derived from the first two questions. We examine the empirical behavior of AoL data in a real-world 5G network setting, bridging the chasm between theoretical constructs and practical applications. The insights gleaned from this process are then employed to bolster the operational efficiency of real-time Automated Guided Vehicle (AGV) applications.

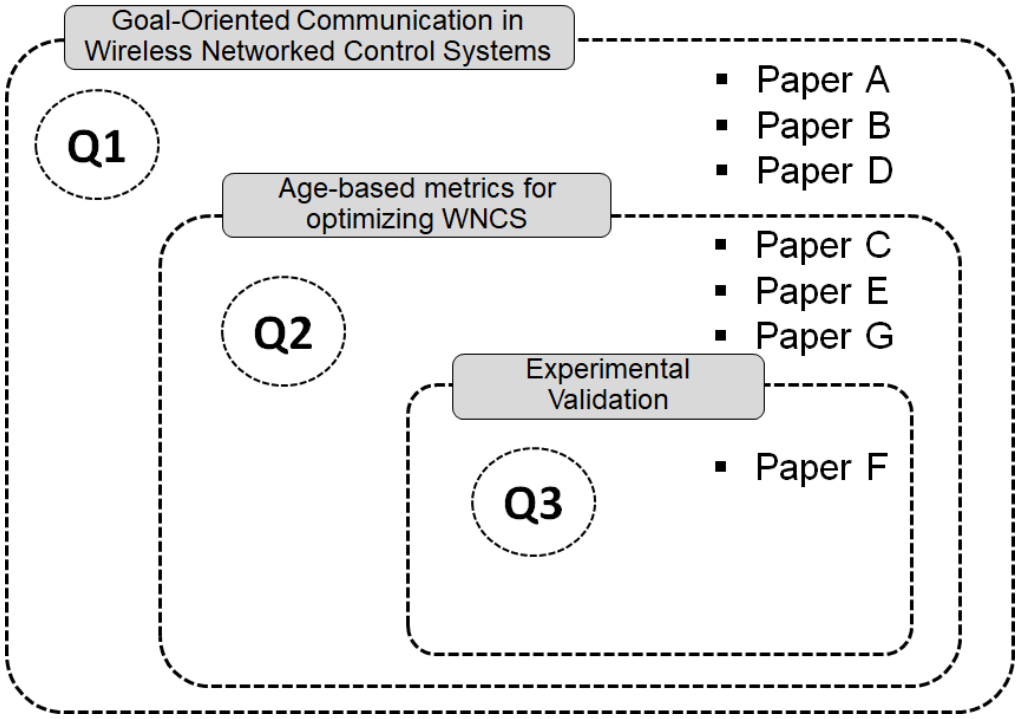


Fig. 1: Relationship between the topic, the research questions and the corresponding papers included in the thesis.

3 Thesis Outline

This thesis, structured as a compendium of papers, is divided into two main sections where significant contributions and findings are highlighted. Part I continues with Chapters 4 and 5. Chapter 4 elucidates the theoretical background pertaining to Wireless Networked Control Systems, including the Age of Information and Age of Loop. Chapter 5 features comprehensive overviews of the individual papers incorporated into the thesis, while also enumerating patent applications and open source code contributions developed during the course of the PhD research. Finally, conclusions and discussions concerning prospective research directions are detailed in Chapter 9.

Part II showcases the contributions made in the form of submitted and published scientific papers, forming the crux of the thesis. The papers are organized chronologically, methodically addressing the research questions initially outlined in this section.

4 Exploring the Core Theories: Control, Communication, and Age-based Metrics

In this chapter, we offer a concise review of the theoretical background central to our contributions. We commence by laying out the foundational elements of wireless networked control systems, along with an exploration of existing solutions proposed by current literature. Subsequently, we introduce age-based metrics and expound upon their potential utility as a method for goal-oriented communication.

4.1 Wireless Networked Control Systems

Wireless Networked Control Systems are a subset of networked control systems where the communication among various control elements, including sensors, controllers, and actuators, is wireless. There are many advantages of WNCS deployment in comparison with traditional control systems in terms of flexibility, maintainability, and lower costs. In this context, WNCS is considered an essential enabler for future industrial, logistics, and transport applications, where a high level of flexibility, data fusion, resource sharing, and cost reduction are desired.

The development of WNCS, however, presents unique challenges due to the unreliability and limited resources of wireless networks, as represented in Figure 2. For these reasons, the design and analysis of such systems require unique consideration of the wireless networking aspects along with the control system design.

The Dynamics in Control System Architecture

A conventional control system architecture encompasses a structured network of sensors, actuators, and controllers, operating in tandem to manage the behavior of the system. The workflow within this network initiates with sensors, which are responsible for capturing and transmitting the real-time state of the system to the controller. Once the controller receives these state measurements, it utilizes them to compute the control inputs. These inputs are then forwarded to the actuators, effectively influencing the system's behavior based on the inputs.

To provide a mathematical perspective, let us consider $\mathbf{X}(t) \in \mathbb{R}^n$ as the state-space representation obtained from sensor data, and $\mathbf{U}(t) \in \mathbb{R}^m$ as the control input generated by the controller and transmitted to the plant. For all integers n and m, and at any given time instance t, the system dynamics can be encapsulated by a differential equation $\dot{\mathbf{X}}(t) = f(\mathbf{X}(t), \mathbf{U}(t))$. In the canonical case of a linear time-invariant (LTI) plant, the function f represents a linear combination of the system state and control input, as described by the following equation:

$$\dot{\mathbf{X}}(t) = f(\mathbf{X}(t), \mathbf{U}(t)) = \mathbf{A}\mathbf{X}(t) + \mathbf{B}\mathbf{U}(t) \tag{1}$$

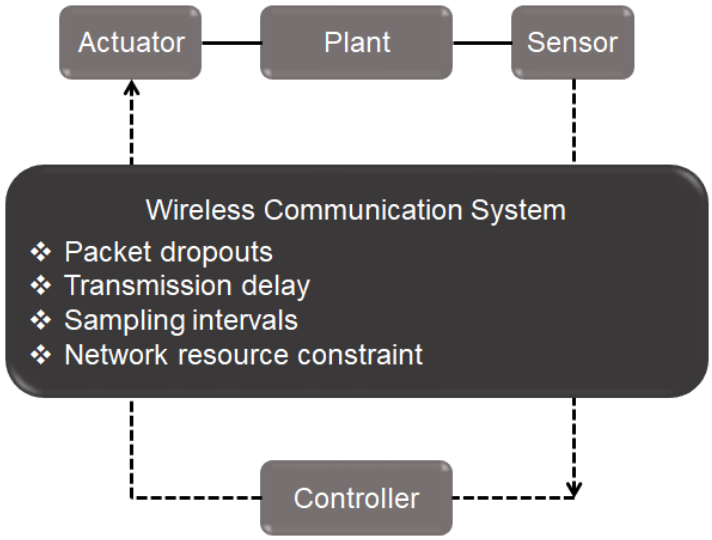


Fig. 2: Components of the control system are connected via wireless communication, thus leading to imperfections and constraints.

Here, $\mathbf{A} \in \mathbb{R}^{nxn}$ and $\mathbf{B} \in \mathbb{R}^{nxm}$ are the state transition and control input matrices, respectively, which essentially characterize the control system behavior.

The controller, leveraging the state-space representation $\mathbf{X}(t)$, computes the control input $\mathbf{U}(t)$. Conventionally, this computation is carried out by applying a state-feedback control law, which can be represented mathematically as follows:

$$\mathbf{U}(t) = \mathbf{K}\mathbf{X}(t) \tag{2}$$

In this equation, $\mathbf{K} \in \mathbb{R}^{mxn}$ symbolizes the controller gain matrix. It is through these set of equations and control laws that the entire operation of a control system is steered and monitored.

Optimal Control Design

For many control applications, it is desirable to design the controller so as to optimize a certain performance index. One common approach is to minimize the quadratic cost function, leading to what is known as the Linear Quadratic Regulator (LQR) [12]. The LQR problem is defined as follows:

$$\begin{aligned} \underset{\mathbf{U}(t)}{\text{minimize}} \quad & \int_0^\infty \left[\mathbf{X}(t)^T \mathbf{Q}\mathbf{X}(t) + \mathbf{U}(t)^T \mathbf{R}\mathbf{U}(t)\right] dt \\ \text{subject to} \quad & \dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{B}\mathbf{U}(t) \end{aligned} \tag{3}$$

Here, $\mathbf{Q} \geq 0$ and $\mathbf{R} \geq 0$ are arbitrary positive defined matrices that are chosen by the designer to specify weights between state regulation and control effort [12].

The solution to the LQR problem is a state-feedback controller of the form $\mathbf{U}(t) = \mathbf{K}\mathbf{X}(t)$, where the optimization to find the $\mathbf{U}(t)$ is done in an online fashion each time a new control command is generated, by solving the Algebraic Riccati Equation [15]:

$$\begin{aligned} &\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{Q} = \mathbf{0}, \\ &\mathbf{K} = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}, \\ &\mathbf{U}(t) = \mathbf{K}\mathbf{X}(t). \end{aligned} \tag{4}$$

For $\mathbf{A}$ and $\mathbf{B}$ controllable, the infinite time horizon LQR with $\mathbf{Q}, \mathbf{R} > \mathbf{0}$ gives a convergent closed-loop system [15], where the stability can be guaranteed.

Control system under imperfect communication

Wireless networks, distinguished by their dynamic behavior, present a set of unique challenges compared to wired control loops. As illustrated in Figure 2, constraints intrinsic to wireless communication can result in variable delays and packet dropouts, often contributing to deteriorated performance or instability in the control system.

Existing literature typically partitions solutions to these challenges into two primary categories:

- *Control over Network* [9, 17, 22]: This category concentrates on designing control systems under the premise of predictable network constraints. The primary objective is to optimize the Quality of Control (QoC) within the volatile environment induced by the network's behavior. The proposed solutions frequently depend on robust assumptions about the network's behavior, such as packet dropouts following a Bernoulli distribution or constant delay models.

- *Control of Network* [13, 21, 23]: This category primarily targets the communication network itself. It encompasses pivotal aspects such as radio resource management, routing or congestion protocols, and network topology. The principal aim is to achieve network configurations that satisfy Quality of Service (QoS) constraints.

Both categories, however, display a clear delineation between the control and the network entities, each optimized independently. This leads to proposed solutions that lean heavily on assumptions about the behavior of either the network or the control.

In response to these realities, this thesis approaches a third category: *Control-Oriented Network*. Here, the network's decisions are influenced not solely by network QoS Key Performance Indicators (KPIs), but also by the overarching goals of the application. Age-based metrics play a pivotal role in achieving this integration, and will be introduced in the subsequent section.

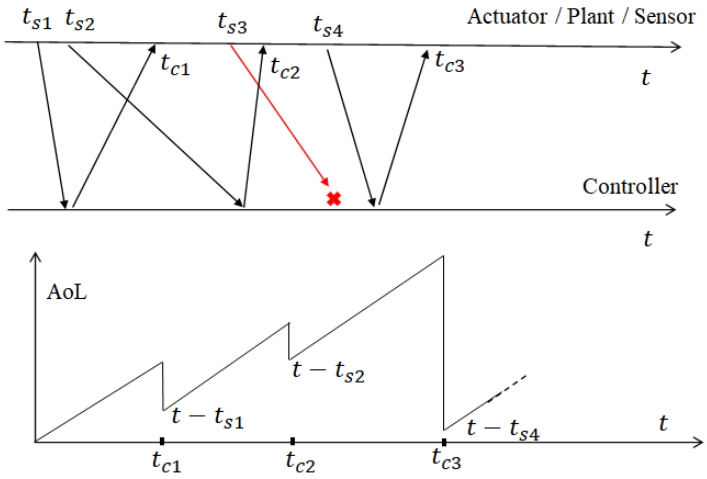


Fig. 3: Workflow of AoL behavior in WNCS.

4.2 Age of information and Age of Loop

The concept of Age of Information forms a crucial framework in understanding the freshness of knowledge related to the status of remote systems. As a quantifiable measure, AoI marks the time span between the generation of the latest received data and the present moment, represented mathematically as $\Delta(t) = t - U(t)$ where $U(t)$ indicates the time when the newest data was generated, as highlighted in [1, 20].

However, it is important to note that the AoI's formal definition is inherently applicable to a singular communication link. Studies exploring Wireless Network Control Systems (WNCS) through the lens of AoI have so far been confined to separate analyses of either uplink (UL) [3, 5, 10] or downlink (DL) [8, 14] transmissions. Nonetheless, the intrinsic structure of WNCS involves a closed-loop incorporating both UL and DL, wherein each can potentially impact the other. This creates implications on system performance and network resource utilization. To illustrate, a high UL AoI denotes less current knowledge for the controller about the plant, necessitating quicker delivery of the control signal and consequentially, greater network resource usage by the DL link.

In order to address these considerations, we introduce the Age-of-Loop metric, aimed at assessing the overall age of a WNCS closed-loop. This is clarified through the WNCS example on Fig. 3. Consider the closed-loop control system where an AGV sends its states update and expect a feedback control command from an edge controller. In this context, denote $\mathcal{S} = \{t_{s_1}, t_{s_2}, \ldots, t_{s_i}, \ldots\}, t_{s_{i+1}} \geq t_{s_i}$, a sequence of time instances where data packets at the AGV are generated and subsequently sent to the edge controller, while $\mathcal{C} = \{t_{c_1}, t_{c_2}, \ldots, t_{c_i}, \ldots\}, t_{c_{i+1}} \geq t_{c_i}$, the time instances where the corresponding

feedback control command is received by the AGV from the controller, satisfying an arbitrary arrival process where $t_{c_i} > t_{s_i}$. Additionally, we can define $\delta_i = t_{c_i} - t_{s_i}$ the total time encompassing the packet generation until the control loop is closed, which can include the time spent for transmissions (DL and UL), queuing and control computation. For a given timestamp t, the latest control loop cycle was initiated at the timestamp:

$$s(t) = \sup\{t_{s_i} \in \mathcal{S} : t_{s_i} + \delta_i \leq t\}, \tag{5}$$

such that we can define the Age of Loop as:

$$\Delta_{\mathrm{L}}(t) := t - s(t). \tag{6}$$

Remark: The AoI is defined for two independent links, requiring instantaneous and perfect feedback channel for the sender to know the age at the receiver. This is a practical limitation often neglected in the literature. In contrast, AoL can be easily implemented and exploited in a real network, as it captures the behavior of both UL/SL directions into a single metric that can be easily measured at either of the end points.

5 Summary of Contributions

This chapter presents succinct synopses for each paper included in Part II. Each summary outlines the problem addressed, the paper's primary goal, and the key findings. Thus, the thesis encompasses an assortment of theories, models, ideas, and results derived from these papers, highlighting essential discoveries and contributions in the forthcoming publications. Moreover, we enumerate patent applications and open source code contributions developed during the course of the PhD research.

5.1 Research Papers

Paper A: "Wireless control of autonomous guided vehicle using reinforcement learning"

Description

This paper identifies a significant challenge in the real-time control and coordination of mobile robots from an edge cloud infrastructure over a wireless communication network. This method is viewed as crucial for future industrial, logistics, and transport applications due to its potential for flexibility, data fusion, resource sharing, and cost reduction. The primary problem lies in achieving low latency, reliability, and stability due to uncertainties such as network-induced delays, data packet dropouts, network topology, channel fading, and network throughput. Existing models and approaches fail to adequately consider these factors, leading to non-optimal reliability and issues with control stability.

Objective

The paper's objective is to propose and examine a more comprehensive model for the control communication protocol to read the AGV state and send control commands. The aim is to enhance the performance of the Autonomous Guided Vehicle (AGV) by employing a learning-based approach to optimize vehicle speed based on channel information, thus improving both the application's stability and communication reliability. An essential part of the study also involves using a reinforcement learning approach to address the trade-off between increased speed and system stability.

Main Findings

The research demonstrates that system performance and stability are significantly affected by the reliability of the wireless link with fading. There is a notable trade-off between AGV speed, control stability, and channel quality, which can be managed effectively using a reinforcement learning approach. This approach can find the optimal speed of AGV to complete a mission path in the shortest time. As a result, the proposed solution maintains system stability on par with widely used baseline state-of-the-art controllers while reducing AGV mission time by more than 30%. The study also validates the impact of vehicle velocity on system stability from both a channel and control perspective. It shows how reinforcement learning algorithms can offer an effective solution to improve AGV stability and mission time.

Paper B: "Radio Access Scheduling using CMA-ES for Optimized QoS in Wireless Networks"

Problem

The paper addresses the challenge of resource allocation of limited radio resources for applications with various Quality of Service (QoS) requirements, a well-known issue in wireless communications. Traditional approaches to this problem often fail to balance complexity, fairness, latency, throughput, and other network requirements effectively. Moreover, the issue of reproducibility and comparison of results in machine learning (ML) techniques for radio resource management, especially Reinforcement Learning (RL), remains unresolved due to a lack of standard problem definition and open implementations.

Objective

The paper aims to propose a solution for the TimeFreqResourceAllocation-v0 problem using a learning-based black-box methodology, specifically the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). This problem, as introduced by Nokia Bell Labs in its open-source framework Wireless Suite, entails an OFDM resource allocation task

where a limited number of frequency resources need to be allocated to a large number of User Equipments (UEs) over time. The goal is to provide an approach that improves overall QoS performance and scales effectively with the growing number of UEs in terms of computational resources.

Main Findings

The study reveals that the proposed black-box optimization technique, CMA-ES, for the TimeFreqResourceAllocation-v0 problem shows superior overall performance in QoS provision compared to standard classical and RL approaches. It demonstrates that CMA-ES not only improves performance but also scales more effectively with the increasing number of UEs in terms of computational resources, a feat that is not matched by Reinforcement Learning approaches. The results of the paper confirm that the proposed approach can outperform traditional baselines provided by Nokia and even surpass Deep Reinforcement Learning techniques.

Paper C: "Age of loop for wireless networked control systems optimization"

Problem

The paper addresses the challenge in Wireless Networked Control Systems (WNCS), specifically concerning the joint design of control and communication for future wireless industrial applications. It focuses on the Age of Information (AoI) metric, which is usually defined for a single communication direction and fails to capture the closed-loop dynamics. Furthermore, conventional design approaches typically decouple control and communication systems, which leads to over-provisioning of network resources and does not effectively reflect the interplay between these systems in real-world applications.

Objective

The primary objective of this paper is to extend the concept of AoI by introducing a new metric, the Age of Loop (AoL). The AoL considers both uplink and downlink of the control loop in WNCS and thus provides a more precise system state estimation. Using this newly proposed metric, the paper aims to learn the WNCS latency and freshness bounds and to minimize the long-term WNCS cost with the least amount of bandwidth.

Main Findings

We find that the proposed AoL metric, as a measure of the age of a WNCS closed-loop, offers a more efficient and accurate evaluation of the WNCS performance than the standard AoI metric. When applied to a case study involving a remotely controlled

inverted pendulum system, the AoL metric significantly outperforms policies based on fixed latency requirements. Moreover, by proposing a bandwidth allocation policy based on the AoL and channel quality information, the authors demonstrate that it's possible to learn the system robustness and thus prevent over-provisioning of network resources in a networked control system.

Paper D: "Control-Aware Scheduling Optimization of Industrial IoT"

Problem

The research focuses on the frequency resource allocation challenge within Wireless Networked Control Systems (WNCS). In many industrial settings, remote Internet of Things (IIoT) devices vie for limited network resources. The allocation of these resources by a centralized network base station is crucial to ensure the overall control system's stability. Traditional scheduling approaches either decouple the control and communication entities leading to over-provisioning of network resources, or they do not adequately consider the time and frequency dynamics of a real network scheduler, which can impact the strict timing requirements of WNCS.

Objective

The paper's main objective is to design a joint network and control scheduling solution that estimates the control system's degradation based on the network state and utilizes this information to assign frequency resources to each device. The authors aim to develop a scheduler that adheres to the 3GPP MAC scheduling scheme and generalizes for any new IIoT device entering the network, thus significantly reducing system complexity, especially in highly dense scenarios.

Main Findings

The proposed control-aware scheduler mechanism, formulated as a Markov Decision Process (MDP), successfully estimates each IIoT device's control performance based on its current network conditions. This allocation scheme generalizes for multiple control dynamics and can be solved in polynomial time, which significantly reduces the overall system complexity. When compared to traditional scheduler baselines and evolutionary approaches, the solution provides better performance. It minimizes the overall control cost, increases system capacity, and assures the long-term stability of the overall system.

Paper E: "Goal-Oriented Wireless Communication for a Remotely Controlled Autonomous Guided Vehicle"

Problem

The study addresses the issue of precise remote trajectory tracking control of Autonomous Guided Vehicles (AGVs) over time-varying wireless channels in dynamic factory environments. One of the inherent problems in this setup is the dependency between the data rate and the resulting control accuracy for the system. While the Age-of-Information (AoI) metric offers a solution for measuring the freshness of transmitted data, it does not account for both downlink and uplink effects and their interplay. Moreover, it neglects the role of correlated fading, which significantly influences remote path tracking.

Objective

The research aims to propose a goal-oriented wireless solution that continuously adapts data transmission rates based on the Age-of-Loop (AoL) metric to achieve precise remote trajectory control of an AGV. The researchers intend to demonstrate how the physical AGV process varies as a function of AoL, linking control planning and radio resource allocation. The project also involves formulating a model that dynamically adjusts the transmission data rate, optimizes the AGV trajectory, and outperforms fixed-data rate policies and AoI-based solutions.

Main Findings

The study showed that the proposed model of a remote AGV control that dynamically adjusts the wireless transmission rate successfully optimizes the AGV trajectory. By employing the AoL concept, the model accounted for both downlink and uplink effects as well as the channel correlated fading effect. Numerical analysis affirmed the importance of the proposed solution in goal-oriented AGV applications, delivering superior performance than fixed-data rate policies and AoI-based solutions, achieving higher system trajectory accuracy. The findings indicate the potential for this approach to improve precision in complex factory scenarios in future works.

Paper F: "Experimental Study of Information Freshness for Goal-Oriented Wireless Communication in a Factory"

Problem

The problem at hand is the traditional approach of wireless data transmission that primarily focuses on rate and reliability but doesn't necessarily align with the ultimate application goal. Age of Information (AoI), a metric quantifying the freshness of information, has been the common attribute representing the relevance of information but falls short as it is constrained to a single communication link, limiting its effectiveness

for closed-loop control systems. The Age-of-Loop (AoL) metric, which takes into account both downlink and uplink effects, as well as their interplay, was introduced as an alternative to AoI. However, its practical application in closed-loop control systems with wireless links needs to be explored further.

Objective

The study aims to leverage the goal-oriented wireless communication paradigm to enhance the effectiveness of wireless data transmission in the context of Wireless Networked Control Systems (WNCS). This involves setting up a measurement campaign within a factory environment to collect and analyze AoL data and to experimentally validate its significance for WNCS problems. Furthermore, the study aims to propose a numerical model to characterize the AoL behavior and use it to address the problem of wirelessly controlled Autonomous Guide Vehicle (AGV), formulating this problem as a semi-Markov-Decision Process (MDP) where the AGV trajectory is optimized by controlling network model parameters.

Main Findings

The study successfully set up an experimental setup within a factory environment to model and analyze the wireless channel behavior of an AGV, utilizing a 5G standalone network. The research proposed a model for a remote AGV control system that dynamically adjusts the wireless transmission rate to optimize the AGV's trajectory, considering both the effects of downlink and uplink using the AoL concept, and the channel-dependent fading effect. The study showed that the obtained optimized policy using AoL provided better AGV performance than AoI baseline solutions, validating the proposed solution with experimental data. These findings underline the significance of the proposed solution for AGV applications, particularly in factory settings.

Paper G: "Goal-Oriented Wireless Communication and Control using Age of Loop"

Problem

The current communication paradigm optimizes network performance indicators such as throughput, latency, and packet loss, which may lead to over-provisioning of network resources, especially in the context of Wireless Networked Control Systems (WNCS). There is a need to go beyond content-agnostic wireless connectivity and consider the application's end-goal, which requires a high granularity level of understanding application effectiveness. Existing metrics such as Age of Information (AoI) quantify information freshness but are constrained to a single communication link, making them sub-optimal for closed-loop control problems.

Objective

This paper aims to present Age-of-Loop (AoL) as a suitable metric for goal-oriented communication in WNCS. By considering both downlink and uplink effects, AoL serves as a more holistic measure that can optimize the overall sense-connect-control cycle in WNCS, such as the remote control of Autonomous Guided Vehicles (AGVs). Using an experimental setup within a factory environment, the paper seeks to analyze AoL behavior and showcase its use as a key metric for AGV performance optimization through reinforcement learning.

Main Findings

The study successfully set up a measurement campaign using a 5G-SA network within a factory environment to empirically evaluate AoL behavior. A model was developed to analyze and optimize the trajectory of an AGV using AoL, resulting in two proposed solutions: an AGV speed controller and a network radio resource scheduler. These findings indicate that using AoL allows for the creation of goal-oriented solutions that can be managed by either the application or the network. This approach provides a practical advantage over AoI, as it does not require time synchronization for end-to-end evaluation.

5.2 Patent applications

In addition to the main papers, the following patents, which are not included in the thesis, have been submitted during the Ph.D. studies:

[1] de Sant Ana, Pedro M., and Nikolaj Marchenko, "A Method for Efficient Radio Resource Allocation based on Application QoS Requirements," European patent, submitted 2020.

[2] de Sant Ana, Pedro M., and Nikolaj Marchenko, "A Method for Improved RF-Prediction and Planning based on Ray Tracing Enhanced with Deep Learning Neural Networks," European patent, submitted 2020.

[3] de Sant Ana, Pedro M., and Nikolaj Marchenko, "Control-Aware Scheduling Method and Apparatus," European patent, submitted 2022.

[4] de Sant Ana, Pedro M., and Nikolaj Marchenko, "Adaptive Network using Pilot-Assisted Image Transmission," European patent, submitted 2022.

[5] de Sant Ana, Pedro M., and Nikolaj Marchenko, "RF-based gesture recognition for COVID-19 monitoring and detection," European patent, submitted 2022.

5.3 Open source code

As part of the work developed in Paper B, the following open source code contribution is available:

`https://github.com/nokia/wireless-suite/blob/master/wireless/agents/bosch_agent.py`

6 Conclusions and Future Work

The research insights shared in this thesis span a diverse spectrum of topics pertaining to goal-oriented communication within the context of wireless networked control systems. Primarily, the investigation has orbited around the three research questions delineated in Section 2, each emphasizing the intricate interaction between control and network elements, the use of age-based metrics, and the empirical validation of theoretical outcomes.

In what follows, we reexamine each of these three research questions, offering conclusions drawn from the findings in this thesis, while also envisioning potential trajectories for future research.

Q1: How can we construct a comprehensive model that enables the concurrent optimization of control and network aspects, facilitating mutual learning between these components within the framework of goal-oriented communication?

In Paper A, we articulate a model that incorporates the interplay between control and network, providing preliminary understanding about the mutual influence of control and communication facets and their impact on the overall application. This model serves as the groundwork for model design and problem formulations in Papers E and F. Intriguingly, our study in Paper A presented a counter-intuitive example where fast scale fading could actually be beneficial to the AGV control. This occurs by helping to de-correlate the transmissions and potentially averting a series of errors, thereby enhancing control system stability. A prospective direction for this work could be the employment of Radio Intelligent Surfaces (RIS) to artificially generate fading effects that could bolster control system stability.

Although Paper B isn't directly aligned with the primary focus of this PhD thesis, it allowed us to gain critical insights about the communication model proposed by Nokia, specifically an OFDM transmission. These insights were adapted for the WNCS problems posed in Paper D and G, thereby crafting a framework for analyzing and optimizing both control and communication KPIs.

While the integrated consideration of control and communication aspects within our models is promising, it currently remains a research-oriented initiative. We observe a

gap in contemporary 3GPP standards regarding the establishment of joint communication and control models. The introduction of "survival time" in the 3GPP Release 15 is a step forward, albeit a very generalized one, that remains application-agnostic and still decouples control and communication entities. This challenge paves the way for a potential research direction: how to effectively define control models within 3GPP standards. For additional insights, this problem parallels certain standardization challenges that are strikingly similar to those targeted by the 3GPP TR 22.876 in Release 19 regarding the study on AI/ML Model.

Q2: How can we proficiently employ the Age of Information (AoI) and the innovative Age-of-Loop (AoL) metrics to critically assess and enhance control system performance, while concurrently optimizing network resource?

In Paper C, we probe into the potential of age-based metrics for analyzing control system performance by taking the well-established inverted pendulum control problem as a case study. In this pursuit, we introduce the Age of Loop metric, serving as a measure of the age of a WNCS closed-loop. We show that this metric offers a more precise and efficient assessment of WNCS performance compared to the standard AoI. Specifically, by employing learning-based methodologies, we discover that we can correlate the stability of the control system to the current AoL state of the WNCS. This pivotal finding forms the bedrock for Papers E and F, where we extend this methodology to tackle a distinct control application - specifically, the AGV control, which minimizes trajectory error based on the AoL measurements. While Paper E emphasizes on analytical results, Paper F adopts a more experiment-oriented approach. In Paper G, we exploit AoL to exemplify a control-oriented network scheduler, where network resource allocation is tailored based on the current states of AoL, thereby averting unnecessary overprovisioning of network resources.

The exploration of AoL-oriented network schedulers, exemplified in Paper G, has potential to be broadened to encompass a variety of applications, each with their distinct requirements and limitations. Specifically, there's a vast opportunity to examine how network resource distribution in the Industrial Internet of Things (IIoT) domain can be significantly fine-tuned through the cooperative design of both network and control components. The study proposed in [2] serves as a promising example in this context, as a future research direction.

Q3: How can we harness the empirical insights derived from Age-of-Loop (AoL) behavior in a real-world 5G network setting to bridge the gap between theoretical constructs and practical applications, and subsequently employ these insights to enhance the operational efficiency of real-time applications, such as AGVs?

In Paper F, we expanded upon the work of Paper C, showcasing the tangible application of AoL within wireless networked control systems. In this study, we executed a comprehensive measurement campaign utilizing a 5G standalone network. The objective was to measure and analyze AoL behavior in an actual Bosch factory setting located in Stuttgart-Feuerbach, providing a tangible assessment of AoL performance in practical settings. We analyzed the collected AoL data and developed a Gilbert-Elliot communication model. This model numerically characterized the observed AoL behavior within the factory, and we went a step further to validate our proposed model utilizing empirical AoL data. This step serves to translate the empirical findings into a theoretical construct that can be used for further analysis and optimization.

The AoL model we constructed laid the foundation for a wireless controlled Automated Guided Vehicle application. We were thus able to statistically validate the network's impact on the AGV's track-error performance, providing a real-world demonstration of how AoL insights can be employed to enhance the operational efficiency of a real-time system.

To cap it all, we formulated a problem aimed at optimizing AGV trajectory by controlling network model parameters within a Semi-Markov Decision Process framework. The optimized policy, derived using AoL, resulted in superior AGV performance compared to baseline solutions utilizing AoI. These findings underscored the significance of the proposed solution for AGV applications, particularly within factory environments. This demonstrates the practical value of AoL insights in optimizing real-time applications, thereby effectively bridging theory and practice.

As a potential future endeavor, we aspire to address more complex factory environments by embedding AGV path control planning into the currently proposed semi-Markov Decision Process framework. The objective of this integration is to optimize not only the trajectory of a single AGV, but also the operational plans of multiple AGVs concurrently navigating within a factory environment. This takes into account the diversity in task assignments for AGVs, as different tasks might demand varying stability requirements and thus different resource allocations. By including this complex dimension of resource scheduling among multiple AGVs, we anticipate the development of a comprehensive model capable of managing dynamic conditions in real-world factory settings. For example, AGVs performing tasks that require higher precision might need more frequent updates, requiring more network resources, while those running less critical tasks could operate with less frequent updates. Our model would aim to dynamically allocate the resources based on these diverse needs to ensure overall system performance.

This holistic approach to AGV operations, planning, and scheduling will aid in achieving more efficient, flexible, and robust AGV operations across the factory floor. It is expected that such a strategy will significantly enhance operational productivity and efficiency in advanced manufacturing scenarios.

References

[1] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, 2019.

[2] J. Cao, X. Zhu, S. Sun, P. Popovski, S. Feng, and Y. Jiang, "Age of loop for wireless networked control system in the finite blocklength regime: Average, variance and outage probability," *IEEE Transactions on Wireless Communications*, 2023.

[3] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using AoI in a single-loop networked control system," in *IEEE Conference on Computer Communications Workshops*, 2019.

[4] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Age of loop for wireless networked control systems optimization," in *IEEE PIMRC*, 2021.

[5] K. Gatsis, H. Hassani, and G. J. Pappas, "Latency-reliability tradeoffs for state estimation," *IEEE Transactions on Automatic Control*, 2020.

[6] T. M. Getu, G. Kaddoum, and M. Bennis, "Making sense of meaning: A survey on metrics for semantic and goal-oriented communication," *IEEE Access*, 2023.

[7] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, 2022.

[8] Huang, Kang *et al.* , "Wireless feedback control with variable packet length for industrial IoT," *IEEE Wireless Communications Letters*, 2020.

[9] Y. Jiang, J. Fan, T. Chai, F. L. Lewis, and J. Li, "Tracking control for linear discrete-time networked control systems with unknown dynamics and dropout," *IEEE Transactions on Neural Networks and Learning Systems*, 2017.

[10] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "AoI-penalty minimization for networked control systems with packet loss," in *IEEE Conference on Computer Communications Workshops*, 2019.

[11] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, 2021.

[12] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[13] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward computing resource reservation scheduling in industrial internet of things," *IEEE Internet of Things Journal*, 2020.

[14] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the latency, rate and reliability tradeoff in wireless networked control systems for IIoT," *IEEE IoT Journal*, 2020.

[15] Z. Sun, R. Wang, Q. Ye, Z. Wei, and B. Yan, "Investigation of intelligent vehicle path tracking based on longitudinal and lateral coordinated control," *IEEE Access*, 2020.

[16] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, B. Soret *et al.*, "Semantic communications in networked systems: A data significance perspective," *IEEE Network*, 2022.

[17] X. Wang, J. Sun, G. Wang, and L. Dou, "A mixed switching event-triggered transmission scheme for networked control systems," *IEEE Transactions on Control of Network Systems*, 2021.

[18] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, 1968.

[19] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. S. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, 2022.

[20] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE Journal on Selected Areas in Communications*, 2021.

[21] Q. Yu and M. Gu, "Adaptive group routing and scheduling in multicast time-sensitive networks," *IEEE Access*, 2020.

[22] J. Zhang and E. Fridman, "Dynamic event-triggered control of networked stochastic systems with scheduling protocols," *IEEE Transactions on Automatic Control*, 2021.

[23] T. Zhu, Z. Cai, X. Fang, J. Luo, and M. Yang, "Correlation aware scheduling for edge-enabled industrial Internet of things," *IEEE Transactions on Industrial Informatics*, 2022.

Part II

Papers

Paper A

Wireless control of autonomous guided vehicle using reinforcement learning

Pedro M. de Sant Ana, Nikolaj Marchenko, Beatriz Soret and Petar Popovski

The paper has been published in the
IEEE Global Communications Conference (GLOBECOM) 2020.

Abstract

Real-time wireless networked control of an Autonomous Guided Vehicle (AGV) from an edge cloud controller is an attractive approach to reduce hardware costs of AGVs, e.g., for industrial applications. We specify a networked control protocol for AGV and investigate how system performance and stability are affected by the reliability of the wireless link with fading. Particularly, there is a trade-off between the AGV speed, the control stability, and the channel quality. Our model takes into account end-to-end latency, which includes control loops and communication. Considering the model complexity, we employ a Reinforcement Learning (RL) approach in order to find the optimal speed of AGV to complete a mission path in shortest time. The proposed solution achieves system stability at par with widely used baseline state-of-the-art controllers, while reducing the AGV mission time by more than 30%.

1 Introduction

Real-time control and coordination of mobile robots and vehicles performed from the edge cloud infrastructure over a wireless communication network is seen as an essential enabler for future industrial, logistics, and transport applications, where a high level of flexibility, data fusion, resource sharing, and cost reduction are highly desired [1]. However, achieving the low latency, reliability and stability for real-time control required in many industrial applications remains a critical challenge, due to often uncertain highly time-varying network-induced delays, data packet dropouts (data losses/packet losses), network topology, channel fading, and network throughput [2].

Such uncertainty often leads to non-optimal reliability of control applications when classic model-based approaches are employed. In contrast, in recent years the use of learning-based methods is proposed, which aim to learn how to act in particular local conditions and achieve desired control stability [3]. In this context, the authors in [4] analyze the path tracking problem of a wireless controlled AGV, where a Kalman filter is designed to provide end-to-end delay estimation. However, an important shortcoming of this approach is that the proposed latency model does not consider the channel time correlation, which is dependent on the AGV speed and directly impacts the AGV stability. This effect is actually studied by the authors in [5], where they provide an analytical approach on the relationship between the AGV velocity, the channel outages, and the control system stability, providing an upper bound, by exhaustive search, on the consecutive channel outages to retain the stability of an AGV control system. Given this background, we aim to extend the problem raised in [5], proposing a more elaborated communication protocol and using a learning methodology to optimize the vehicle velocity based on the channel information. Other related works, such as [6] and [7], focus on network planning and design optimization solutions, which is not the scope of

this paper, where we seek to enhance the AGV performance by using a learning-based approach for improving both the application stability and communication reliability.

This paper has three main contributions. First, we propose a model for the control communication protocol, which is then used to read the AGV state and send control commands. Second, we analyze the stability of AGV for the defined model and correlated wireless fading channel. Finally, we utilize a reinforcement learning approach to optimize the AGV speed while improving the system stability and, thus, achieving the shortest mission time with lower failure rates. One of the main insights obtained from our model is that there is an interesting tradeoff with respect to the speed: increasing speed may deteriorate stability, but at the same time help the wireless link to faster overcome a deep fade. The proposed learning approach is inherently capable to address this trade-off and adjust the speed accordingly.

The rest of the paper is organized as follows: in the next Section, we present the system model for the control protocol and AGV. In Section 3, we elaborate on the problem of AGV stability over imperfect channel. Finally, in Section 4, we present a reinforcement learning approach for optimizing the AGV velocity.

2 System Model for the AGV Control

2.1 Control Protocol

Figure A.1 shows the overall model of the communication protocol we use to communication with an AGV. The model shows the details of the interaction between the communication and application control loops. First, the readings of the AGV state $z(k)$, at step $k \in \mathbb{Z}^+$, are stored into memory and communicated to the controller over the communication channel. For simplicity of visualization and modeling, we assume that the uplink channel is error- and delay-free, and the input values are immediately delivered to the controller. The readings of input values are done strictly periodically with the input cycle time ΔT_{in}, as it is commonly done across various control systems [1].

At the controller, the input values $z(k)$ are also stored into the memory. A control application is called periodically, with interval time ΔT_{c}, which is also commonly done in real-time controllers. The control application gets the most recently stored input values, and produces an output $u(k)$ for the AGV, according to the predefined metrics and control goals (see 2.2). In our model we neglect the processing time required for the control algorithm.

The output parameters $u(k)$ are communicated back to the AGV. We assume that the downlink channel is not necessary error-free, and the control values can get lost. Details to the channel model and the impact of packet loss are discussed later in Section 3.2. To provide higher end-to-end reliability, it is common across industrial control communication protocols, to perform multiple transmissions of the same data at the application level. In the proposed system, as shown in Figure A.1 the transmissions are

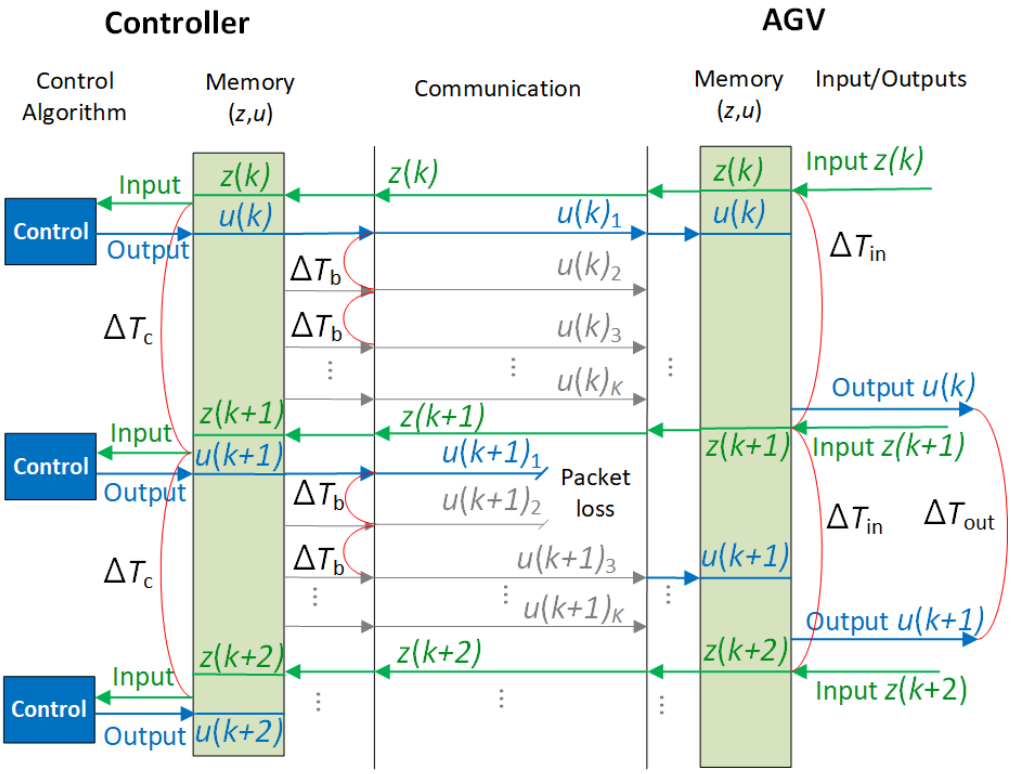


Fig. A.1: Networked control system model to operate AGV from remote controller.

initiated periodically with time interval T_b. As a result, K transmissions of the same output values are performed until new output values can be generated with the control cycle T_b. In practice, similar periodic transmission mechanism should be applied in uplink to provide reliable delivery of input values. However, in this paper we focus on studying the impact of downlink communication, assume perfect uplink transmissions, and leave the investigation of the uplink for future work.

At the AGV side, the received command values $u(k)$ are stored in memory. The output application is called periodically with the time interval ΔT_{out}, when the most recently stored command values are called from the memory and applied to the AGV drives.

In the presented model, each time cycle (ΔT_{in}, ΔT_c, ΔT_b and ΔT_{out}) can be configured independently of each other. By providing such system decoupling, a more flexible implementation of input, control, and communication systems can be achieved. In our

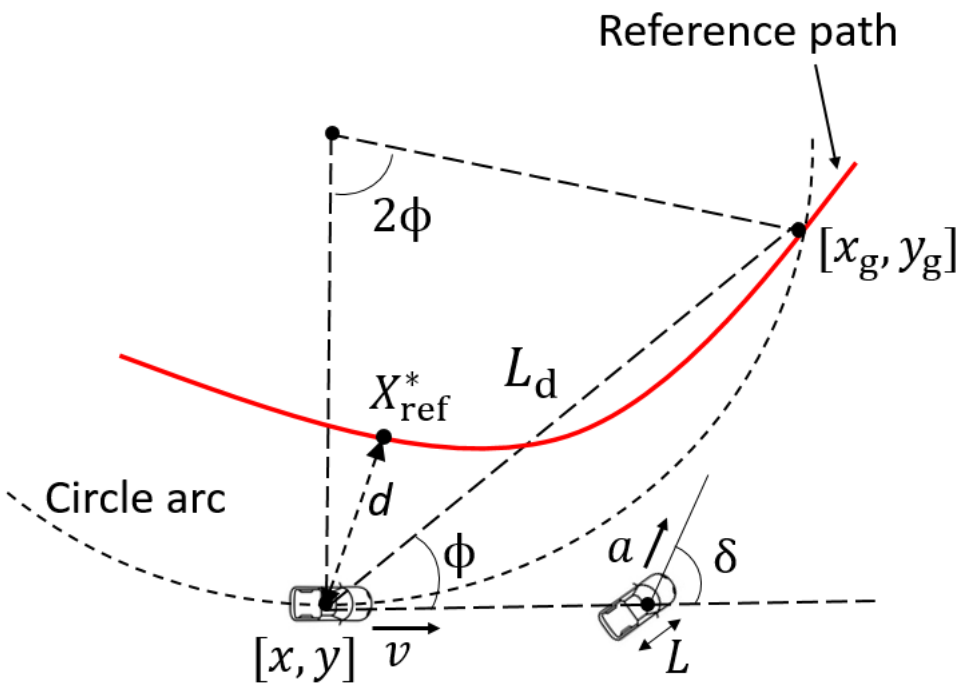


Fig. A.2: Pure Pursuit geometry model

model, we assume that the system is well synchronized, and input, output, communication and control cycles are well adjusted to each other. As a result, $\Delta T_{\text{out}} = \Delta T_{\text{in}} = \Delta T_{\text{c}} = K\Delta T_{\text{b}}$.

2.2 Pure Pursuit AGV Controller

The proposed AGV solution follows a pure pursuit control law [8], which has been proved to be a very effective path tracking strategy, including applications on two vehicles at the DARPA Grand Challenge [9] and three vehicles at the DARPA Urban challenge [10], as is commonly used as baseline comparison.

The model approach to pure pursuit control consists of geometrically calculating the curvature of a circular arc that connects the rear axle location to a goal point on the path ahead of the vehicle. The goal point is determined from a look-ahead distance L_{d} from the current rear axle position to the desired path, as exposed in Fig. A.2.

For a given AGV position, $X(k) = [x(k), y(k)]$, at time k and a predefined reference path $X_{\text{ref}} = \{(x_{\text{ref}}, y_{\text{ref}})\}$, we select the goal point, $G(k) = [x_g, y_g]$, as a value from X_{ref},

such that the condition $\|G(k) - X(k)\| = L_{\mathrm{d}}$ is satisfied. The vehicle steering angle δ then can be determined using only the goal point location and the yaw angle ϕ between the vehicle heading vector and the look ahead vector, i.e.,

$$\delta = \tan^{-1}\left(\frac{2L\sin(\phi)}{L_d}\right) \tag{A.1}$$

As defined in [11], we can set an input state vector $z(k) = [x(k), y(k), v(k), \phi(k)]$. The control output vector $u(k) = [a(k), \delta(k)]$ is defined by the AGV acceleration a and steering angle δ for a given vehicle wheelbase L.

2.3 AGV State Update

We apply linear control methods to the AGV dynamic model, such that the AGV state update is based on the linearized vehicle model [11] given by the solution of (A.2).

$$\dot{z} = \frac{\partial}{\partial z} z = f(z, u) = A'z + B'u \tag{A.2}$$

where A' and B' can be calculated as following

$$A' = \begin{bmatrix} 0 & 0 & \cos(\phi) & -v\sin(\phi) \\ 0 & 0 & \sin(\phi) & v\cos(\phi) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\tan(\delta)}{L} & 0 \end{bmatrix} \quad B' = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & \frac{v}{L\cos^2(\delta)} \end{bmatrix} \tag{A.3}$$

The system is linear, continuous and time varying. We can get discrete-time mode with forward Euler discretization with sampling time dt:

$$z_{k+1} = z_k + f(z_k, u_k)dt \tag{A.4}$$

Using first degree Taylor expansion over z and u, we obtain the discrete, time varying and linear model for the state space update as

$$\begin{aligned} z_{k+1} &= z_k + (f(z,u) + A'z_k + B'u_k - A'z - B'u)dt \\ &= (I + dtA')z_k + (dtB')u_k + (f(z,u) - A'z - B'u)dt \\ &= \bar{A}z_k + \bar{B}u_k + \bar{C} \end{aligned} \tag{A.5}$$

where $\bar{A}$, $\bar{B}$ and $\bar{C}$ are respectively given by:

$$\begin{aligned} \bar{A} &= I + dtA' \\ \bar{B} &= dtB' \\ \bar{C} &= (f(z,u) - A'z - B'u)dt \end{aligned} \tag{A.6}$$

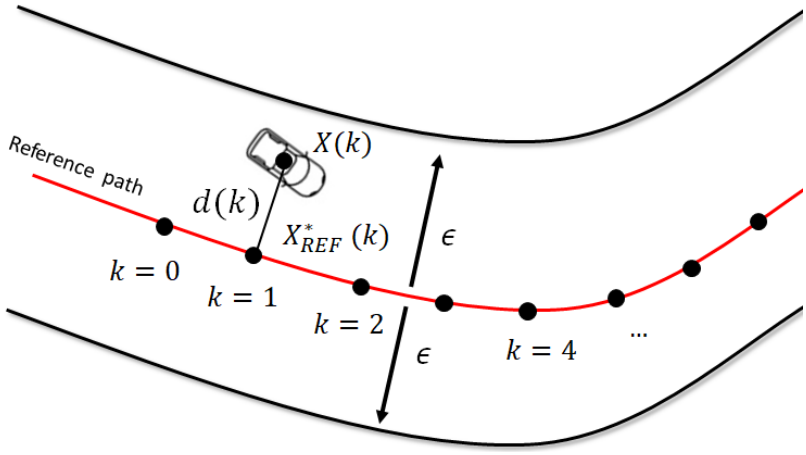


Fig. A.3: Safety stability representation.

2.4 Safety Stability Criteria

It is unarguably critical for many AGV applications to ensure driving safety, for which we determine that the AGV must operate within a feasible region constraint defined according to the road boundaries. This lateral safety corridor, as illustrated in Fig. A.3, can be expressed as a function of the cross-track error (XTE) [12]. As described in (A.7), the XTE is defined by the deviation, $d(k)$, at time k, between the actual AGV position and the closest point on the planned track. $X(k)$ denotes the AGV position and $X^*_{\text{ref}}(k)$ its nearest point at the reference track. We define that the AGV must obey the following criteria for a given maximal allowed path deviation ϵ:

$$d(k) = \|X(k) - X^*_{\text{ref}}(k)\| \leq \epsilon,\ k \in \mathbb{Z}^+ \tag{A.7}$$

As demonstrated in [11], the XTE is directly impacted by the AGV speed. Furthermore, as we will discuss in the next section, $d(k)$ might present random behavior throughout the path under imperfect channel communication. In practice, the threshold of (A.7) may be crossed out, and higher speeds v mean a higher chance of breaking the criterion. We denote p_ϵ to be the probability of exceeding the threshold, such that

$$\begin{aligned} &P[d(k) < p_\epsilon | v_2] \leq P[d(k) < p_\epsilon | v_1] \\ &\forall v_1, v_2 \in \mathbb{R} : v_2 \geq v_1 \end{aligned} \tag{A.8}$$

3 AGV Control over Imperfect Communication Channel

3.1 AGV Pure Pursuit Control over Imperfect Channel

Following the model of Fig A.1, a sequence of channel downlink outages may lead the AGV to use outdated output from the controller. For a sequence of N AGV cycles ΔT_{out} with no information update from the controller (which also is equivalent to $N \cdot K \cdot \Delta T_{\text{b}}$ consecutive channel outages) the AGV state will be updated by

$$z_{k+1} = \bar{A} z_k + \bar{B} u_{k-N} + \bar{C} \tag{A.9}$$

where u_{k-N} is the control input command outdated by N AGV cycles randomly distributed according to the channel fading effect. Therefore, as long as the constraint in (A.7) holds, we keep updating the AGV states according to (A.9).

3.2 Communication Channel Model

The communication channel is modeled as a first-order Markov process [13], also known as Gilbert-Elliot channel. Basically, the time-correlation property is represented by two states: the good state G if the packet can be successfully received; and the bad state B, otherwise. The corresponding transition probability matrix is defined as $M(x) = M(1)^x$, with:

$$M(x) = \begin{bmatrix} p(x) & q(x) \\ r(x) & s(x) \end{bmatrix}, \; M(1) = \begin{bmatrix} p & q \\ r & s \end{bmatrix} \tag{A.10}$$

Where $p(x) = 1 - q(x)$ is the probability that the transmission in the slot time i is successful, given that the transmission in slot $i - x$ was successful. Same logic applies for $r(x) = 1 - s(x)$ but for unsuccessful transmissions. Given the matrix $M(1)$, the channel properties are completely characterized. In particular, it is possible to find that the marginal probability of a packet error, ε_p, is given by:

$$\varepsilon_p = 1 - \frac{r}{1 - p + r} \tag{A.11}$$

For Rayleigh fading, ϵ is expressed as:

$$\varepsilon_p = 1 - e^{\gamma_{th}} \tag{A.12}$$

Where γ_{th} is the minimum threshold SNR required to successfully decode the received signal. The Jakes's channel correlation coefficient is given by

$$\rho = J_0(2\pi f_d T_s) \tag{A.13}$$

where $J_0()$ is the zero-order Bessel function of the first kind, f_d is the Doppler frequency shift and T_s is the sampling time. The error probability of a single back to back failure, s, is then written as [13]

$$s = 1 - \frac{Q(\theta, \rho\theta) - Q(\rho\theta, \theta)}{e^{\gamma_{th}} - 1} \tag{A.14}$$

where $Q(.,.)$ is the Marcum Q function and θ is defined as:

$$\theta = \sqrt{\frac{2\gamma_{th}}{1 - \rho^2}} \tag{A.15}$$

The main motivation for this channel model is to demonstrate the impact of time correlated channel errors at the AGV stability. We can verify that higher Doppler deviation (i.e. higher speeds) reduces the channel time correlation, thus also reducing the probability of channel constant link outages. This behavior is, in fact, well explored by the authors in [5].

3.3 The Impact of Correlated Fading on the AGV Stability

The subsections 3.2 and 2.4 demonstrated that there is a dichotomous behavior when it comes the speed impact over the AGV stability. From channel perspective, higher speeds result in a channel with a lower time correlation, thus minimizing a long burst of communication errors. Nevertheless, from control perspective, higher speeds provide higher cross-track error and, as a consequence, a higher probability of stability loss.

In order to better understand this trade-off between channel reliability and control stability, we first propose to analyze the AGV behavior under two different scenarios with Rayleigh fading as defined in the previous subsection: 1) We fix the channel SNR and vary the cross-track error criteria; 2) We establish a certain XTE criteria, and check the AGV behavior over different SNRs. In both cases, we analyze, by exhaustive searching, the stability performance of multiple vehicle velocities, thus showing if there is any optimal speed in which the AGV should achieve to minimize the number of stability failures along the path.

These scenarios will be evaluated according to the reference path showed in Fig. A.4, considering the main parameters indicated in Table E.1. The region segmentation in Fig. A.4, serves for further analysis in Section 4. For this subsection there is no difference between regions.

The Fig. A.5 illustrates the result for the first scenario, showing each AGV speed performance under different XTE criteria for a fixed 10 dB SNR. While for the second scenario, the Fig. A.6 shows the impact of multiple SNRs over the optimal velocity considering a fixed XTE criteria of 1.25 m. Basically two insightful conclusions can be highlighted from these results. First, adjusting the AGV speed according to the channel condition is not evidently intuitive. Both plots actually show the stability performance getting worse under very low speed. As explained in subsection 3.3, this

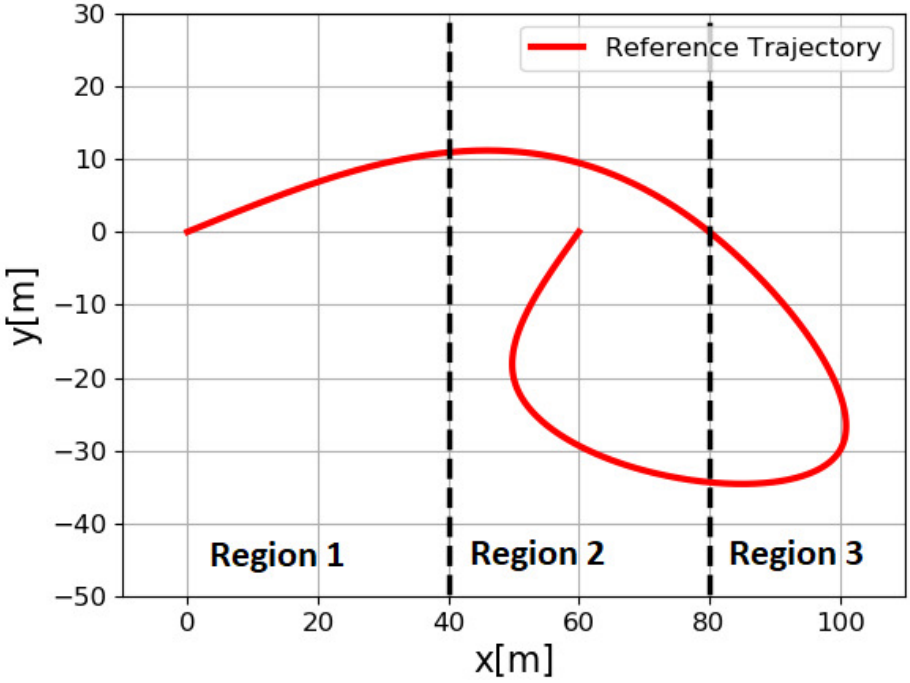


Fig. A.4: AGV reference path.

Table A.1: System Parameters

Bandwidth	200 kHz
Channel sampling time, ΔT_{b}	1 ms
Message size	78 Bytes
Carrier frequency	900 MHz
Spectral efficiency	3.12
AGV sampling time, ΔT_{in}	10 ms
Number of repeated output transmissions, K	10

happens because the channel correlation increases for lower speeds, thus boosting the number of successive link outages. Second, too high AGV velocities are not the best choice either. In these situations, besides there is a smaller channel probability of error, the system control stability is more quickly broken, as exposed by subsection 2.4.

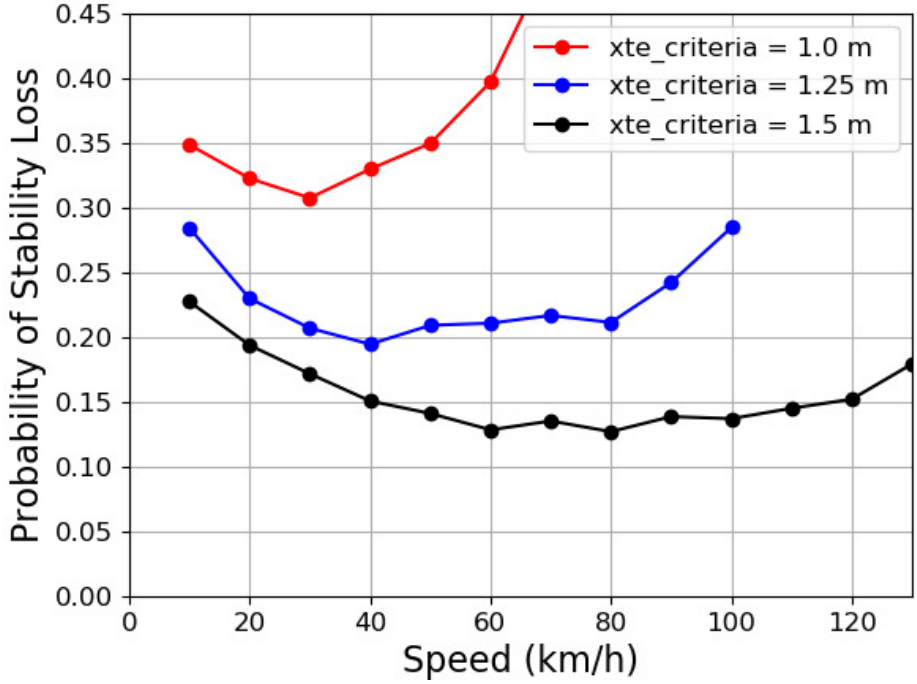


Fig. A.5: Probability of stability loss for each velocity considering different XTE criteria under fixed 10 dB SNR.

3.4 Problem Formulation

In Section 3, we demonstrated the vehicle velocity role at the trade-off between control stability and channel reliability. In other words, the channel is operating independently, and it will eventually fail, causing the AGV stability loss. However, the main question is: What the AGV can do, in turn, to minimize the stability loss?

A potential answer to this question can be exemplified in Fig. A.5 and Fig. A.6, where we showed that adjusting the AGV speed can considerably reduce the probability of stability loss under different XTE criteria and SNRs. Hence, if the AGV finds a way to optimize its speed based on the channel knowledge, it can independently increase its stability.

We can extend this problem to a more general AGV use case. Considering the path of Fig. A.4, we aim to minimize the stability loss while also minimizing the AGV mission time towards the goal point. So, the AGV must do the path as fast as possible while

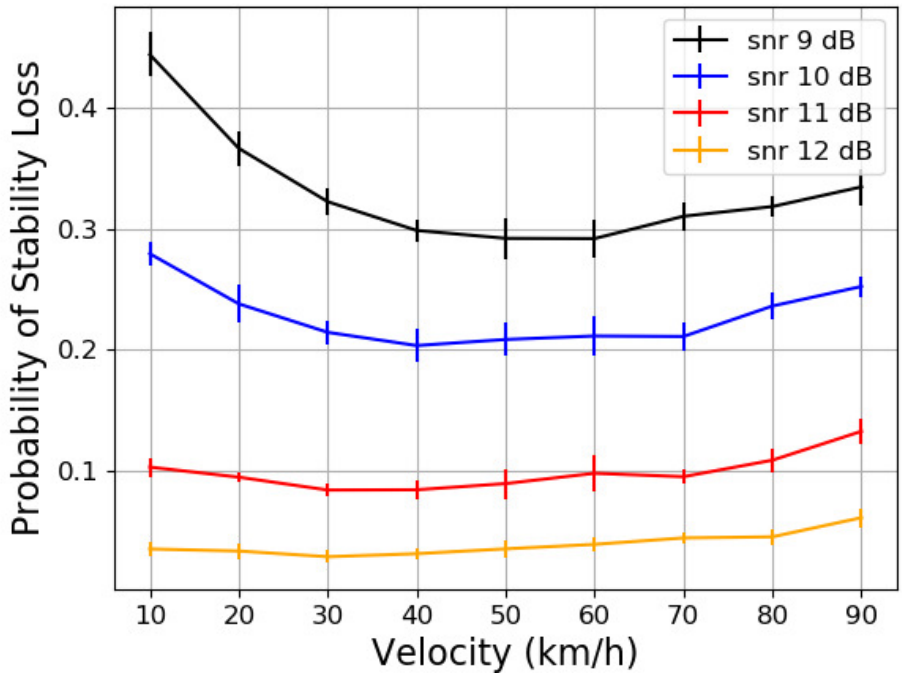


Fig. A.6: Probability of stability loss for each velocity considering 1.25 m of XTE criteria under different SNRs.

maintaining a minimal stability loss.

More formally, we can establish a control policy $\pi(u_k, z_k)$ that, for each time step k, selects the control command u_k given the current AGV state z_k throughout the AGV trajectory $\{z_1, u_1, z_2, u_2, \ldots, z_T, u_T\}$ until termination at time step T. Similar to the robot control problem introduced in [14], we can define a cost function according to (A.16), where the AGV is penalized for each time step taken until achieving the final destination, as well as highly penalized for deviating the path above a certain threshold ϵ.

$$\forall \rho_1, \rho_2 \in \mathbb{R}_*^- : \rho_1 \ll \rho_2$$
$$c(z_k) = \begin{cases} \rho_1, & \text{if } \|X(k) - X^*_{\text{ref}}(k)\| > \epsilon \\ \rho_2, & \text{otherwise} \end{cases} \tag{A.16}$$

Hence, we aim optimizing the AGV trajectory by finding an optimal control policy

$\pi^* = \{u_1^*, \dots u_T^*\}$ that minimizes the expected trajectory cost:

$$\pi^* = \underset{u_1,\dots,u_T}{\operatorname{argmin}} \sum_{k=1}^{T} \mathrm{E}[c(z_k)]$$
$$\text{s.t. } z_{k+1} = f(z_k, u_{k-N}) \tag{A.17}$$

where f is the AGV state evolution defined in (A.9).

Finding π^* means that, for every AGV state throughout the path, the controller provides an optimal control decision, more specifically accelerating or decelerating the AGV, such that (A.16) is minimized although the system is subject to random outdated control commands as a consequence of the channel conditions. This decision-making control process can be represented as a typical Markov Decision Process (MDP) framework [15], where for any given state, the AGV establishes a policy for deciding a proper action that minimizes the cost function as a result of the environment interaction. The general problem is P-complete and can be solved with polynomial complexity to the size of the state/action space [15].

4 Solution Proposal

To solve the MDP in (A.17), we advocate a reinforcement learning methodology for two main reasons. First, using a non-learning approach, such as linear matrix inequality (LMI) methods, as proposed by the authors in [16], would require the cost function to be convex [17]. Moreover, the time complexity and memory limit can exponentially grow with the number of states [17]. Second, if we take into account the variety of environments, the range of possible XTE criteria, and, especially, the need for complete channel probability knowledge, it turns out to be unfeasible to solve this MDP using traditional Dynamic Programming (DP) algorithms, such as value iteration or policy iteration.

4.1 RL Model Configuration

Observation space

We propose a discrete state space, in which both x and y coordinates of the planned path, as represented in Fig. A.4, are sampled by 10 m spacing. We also include in the observation space the instantaneous SNR. In more details, there are 11 samples for x axis (from 0 to 110 m), 6 samples in y axis (from -40 to 20 m) and 4 samples for the SNR in {9 dB, 10 dB, 11 dB, 12 dB}. This will result in a total of 264 states (11x6x4).

Action space

The proposed algorithm for each step must decide between two actions: increasing or decreasing the AGV target speed by 0.25 km/h.

Reward

As formulated in (A.16), we set the AGV to be penalized by -0.1 reward for each step taken until complete the planned path. Moreover, it will be punished by -1000 in case of stability loss.

RL algorithm

We evaluated the AGV performance applying State–action–reward–state–action (SARSA) algorithm [18], where both exploration and learning rate exponentially decay over time.

According to [19], SARSA will converge with probability one to an optimal policy and action-value function as long as all state-action pairs are visited an infinite number of times and the policy converges in the limit to the greedy policy (which is already arranged, since the exploration rate asymptotically converges to zero).

4.2 Simulation Test Scenario

We will evaluate the proposed algorithm considering the following scenario:

- The AGV will pursuit the path as established in Fig. A.4.
- The channel SNR will vary among the set $\{9, 10, 11, 12\}$ dB, as evaluated in Fig. A.6, and the SNR along the pathway will be selected according to three different path regions: $(0 < x \leq 40)$ m, $(40 < x \leq 80)$ m, and $(x > 80)$ m. So, for every simulation, we uniformly random pick the SNR for each path region.
- The AGV has a fixed cross-track error criteria of 1.25 m, as discussed in Fig A.5, with further channel parameters details as described in Table E.1.
- Monte Carlo simulation considering 100 jobs of 1000 evaluations.

4.3 Simulation Results

We compare the proposed RL framework performance with common baseline state-of-the-art controllers, such as the Stanley controller [9], the winner of the DARPA Grand Challenge in 2005, and the Linear–quadratic regulator (LQR) speed and steering controller [11], the Stanford's entry in the 2007 DARPA Urban Challenge. We leave the comparison to other more sophisticated and recent approaches, e.g., model predictive

Table A.2: Simulation results under error-free channel

		Mission Time	Stability Loss
Stanley Controller, [9]	mean	9.28 s	0.0 %
	std	0.011 s	0.0 %
LQR Controller, [11]	mean	13.46 s	0.0 %
	std	0.009 s	0.0 %
Proposed RL Solution I	mean	10.26 s	0.0 %
	std	0.0 s	0.0 %

controllers [8], for follow-up work, where more complex scenarios and system models need to be considered to employ their benefits.

We analyze their performance of stability loss and mission time under the proposed scenario defined in subsection 4.2. For comparison purposes, we also check their behavior under an error-free communication channel, where we can also explore the RL framework behavior under three different training cases:

- **Proposed RL solution I**: We first assume zero channel transmission errors over the path, so that we train the RL algorithm under an error-free scenario.
- **Proposed RL solution II**: Here, we train the RL algorithm over the proposed scenario defined in subsection 4.2, where the channel is subject to errors along the path, but we only consider the AGV position (x, y) in the state space, with unknown SNR information.
- **Proposed RL solution III**: As specified in subsection 4.1, we now train the RL algorithm considering the instantaneous SNR as part of the AGV state.

The whole implementation was carried upon the open-source library python robotics [20]. The Table A.2 provides the results for the error-free case, while the Table A.3 present the outcome of each controller under the proposed scenario.

We can verify in Table A.2 that, when the channel is error-free, it allows the controllers to achieve higher speeds (thus reducing the mission time) while fully respecting the stability criteria. In this situation, the Stanley controller is recognized to be more suitable [11].

For the results of Table A.3, where the controllers were subject to imperfect channel communication, we can highlight three relevant conclusions. First, the RL solution I (trained in the error-free channel), increases its stability loss to more than 20%. However, this approach is already enough to overcome the performance of both benchmark controllers. Second, the RL solution II provides a clear indication about the trade-off

Table A.3: Simulation Results under the test scenario

		Mission Time	Stability Loss
Stanley Controller, [9]	mean	18.77 s	23.65 %
	std	0.128 s	1.29 %
LQR Controller, [11]	mean	20.35 s	23.71 %
	std	0.056 s	2.43 %
Proposed RL Solution I (from error-free scenario)	mean	10.27 s	20.49 %
	std	0.028 s	1.03 %
Proposed RL Solution II (trained in Test Scenario without SNR)	mean	12.95 s	17.81 %
	std	0.076 s	1.71 %
Proposed RL Solution III (trained in Test Scenario with known SNR)	mean	12.95 s	15.91 %
	std	1.409 s	1.12 %

between speed and stability, such that, compared to the RL solution I, it was able to learn a new control approach capable of improving the stability loss by augmenting the mission time. Finally, when we add the SNR information to the RL algorithm (RL solution III), it was possible to achieve the best stability loss performance, while also reducing the path duration by more than 30% compared to the benchmark controllers.

5 Conclusions and Future Works

This work elaborated on a control protocol model and analyzed its use case over a cloud-controlled AGV. We could verify the vehicle velocity's impact at the system stability, considering both channel and control perspective. Based on those results, we formulated an MDP to improve the AGV stability and mission time, advocating a reinforcement learning algorithm as a candidate solution. We compared the proposed scheme with state-of-the-art baseline controllers, showing a considerable performance improvement in stability loss and mission time.

As future perspectives, we plan to analyze the impact of the uplink channel under more complex scenarios, which will probably require more advanced RL methodologies, such as deep reinforcement learning, as well as performance comparison with leading-edge controller approaches like MPC.

Acknowledgement and Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813999. Further information is available at https://windmill-itn.eu/ This document reflects the views of the author(s) and does not necessarily reflect the views or policy of the European Commission. The REA cannot be held responsible for any use that may be made of the information this document contains.

References

[1] P. Park, S. Coleri Ergen, C. Fischione, C. Lu, and K. H. Johansson, "Wireless network design for control systems: A survey," *IEEE Communications Surveys Tutorials*, 2017.

[2] X. Zhang, Q. Han, and X. Yu, "Survey on recent advances in networked control systems," *IEEE Transactions on Industrial Informatics*, 2016.

[3] M. Eisen, K. Gatsis, G. J. Pappas, and A. Ribeiro, "Learning in wireless control systems over nonstationary channels," *IEEE Transactions on Signal Processing*, 2018.

[4] C. Lozoya, P. Martí, M. Velasco, J. M. Fuertes, and E. X. Martín, "Simulation study of a remote wireless path tracking control with delay estimation for an autonomous guided vehicle," *The International Journal of Advanced Manufacturing Technology*, 2011.

[5] S. Tayade, P. Rost, A. Maeder, and H. D. Schotten, "Cloud control agv over rayleigh fading channel-the faster the better," in *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*. VDE, 2019.

[6] K. Sasaki, N. Suzuki, S. Makido, and A. Nakao, "Vehicle control system coordinated between cloud and mobile edge computing," in *Proc. Ann. Conf. the Society of Instrument and Control Engineers of Japan (SICE)*, 2016.

[7] H. Inaltekin, M. Gorlatova, and M. Chiang, "Virtualized control over fog: Interplay between reliability and latency," *IEEE Internet of Things Journal*, 2018.

[8] B. Paden, M. Cap, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Transactions on Intelligent Vehicles*, 2016.

[9] M. Buehler, K. Iagnemma, and S. Singh, *The 2005 DARPA Grand Challenge: The Great Robot Race*. Springer-Verlag Berlin Heidelberg, 2007.

[10] S. Singh, *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*. Springer-Verlag Berlin Heidelberg, 2009.

[11] S. Campbell, *Steering control of an autonomous ground vehicle with application to the DARPA Urban Challenge*. MSc thesis, MIT, 2007.

[12] K. Liu, J. Gong, S. Chen, Y. Zhang, and H. Chen, "Model predictive stabilization control of high-speed autonomous ground vehicles considering the effect of road topography," *Applied Sciences*, 2018.

[13] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Transactions on Vehicular Technology*, 1997.

[14] A. W. Moore, "Efficient memory-based learning for robot control," Ph.D. dissertation, University of Cambridge, 1990.

[15] M. L. Littman, T. L. Dean, and L. P. Kaelbling, "On the complexity of solving Markov decision problems," 1995.

[16] A. Khalil and J. Wang, "Stability and time delay tolerance analysis approach for networked control systems," *Mathematical Problems in Engineering*, 2015.

[17] J. VanAntwerp, "A tutorial on linear and bilinear matrix inequalities," *Journal of Process Control*, 2000.

[18] G. Rummery and M. Niranjan, "On-line Q-Learning using connectionist systems," *Technical Report CUED/F-INFENG/TR 166*, 1994.

[19] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[20] A. Sakai, D. Ingram, J. Dinius, K. Chawla, A. Raffin, and A. Paques, "PythonRobotics: a Python code collection of robotics algorithms," *arXiv:1808.10703*, 2018.

Paper B

Control-Aware Scheduling Optimization of Industrial IoT

Pedro M. de Sant Ana and Nikolaj Marchenko

The paper has been published in the
IEEE Globecom Workshops (GC Wkshps) 2020.

Abstract

In this paper, we explore a radio access scheduling problem, TimeFreqResourceAllocation-v0, proposed by Nokia Bell Labs in its open-source framework Wireless Suite. We elaborate on a learning-based black-box methodology using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) as a candidate solution for QoS-aware scheduling. The proposed approach achieves the best score performance compared to either baseline methodologies (e.g., Round Robin, Proportional Fair, etc.) or common Deep Reinforcement Learning approaches.

1 Introduction

Resource allocation of limited radio resources for applications with different Quality of Service (QoS) requirements is a well-known problem in wireless communications [1]. Classical approaches and their drawbacks in terms of complexity, fairness, latency, throughput, and other network requirements have been extensively studied by the research community [1–3]. In recent years, machine learning (ML) techniques for radio resource management, especially Reinforcement Learning (RL), has been increasingly gaining more attention [4]. Many authors are proposing RL based scheduling approaches for different applications, such as device-to-device (D2D) communication [5], edge computing [6], massive MIMO [7] and LTE/Wi-Fi coexistence [8, 9]. However, it remains a challenge in the research community to easily compare and reproduce those results due to a lack of standard problem definition and open implementations.

However, recently, Nokia Bell Labs has open-sourced its Wireless Suite [10] framework, a collection of well-known problem implementations that are intended to establish performance benchmarks, stimulate reproducible research and foster quantitative comparison of algorithms for telecommunication problems, especially with the use of ML techniques.

The first environment proposed by Nokia is the TimeFreqResourceAllocation-v0 [10], which simulates an OFDM resource allocation task, where a limited number of frequency resources are to be allocated to a large number of User Equipments (UEs) over time. Researchers are invited to develop a new agent that interacts with this environment and takes effective resource allocation decisions.

In this paper, we present a solution for the TimeFreqResourceAllocation problem, which we approach as learning-based black-box optimization using Covariance Matrix Adaptation Evolution Strategy (CMA-ES). The proposed approach shows better overall performance in QoS provision compared to standard classical and RL approaches. It also scales better with the growing number of UEs in terms of computational resources, which is not the case for Reinforcement Learning approaches.

The rest of this paper is organized as following: in the next Section, we provide more

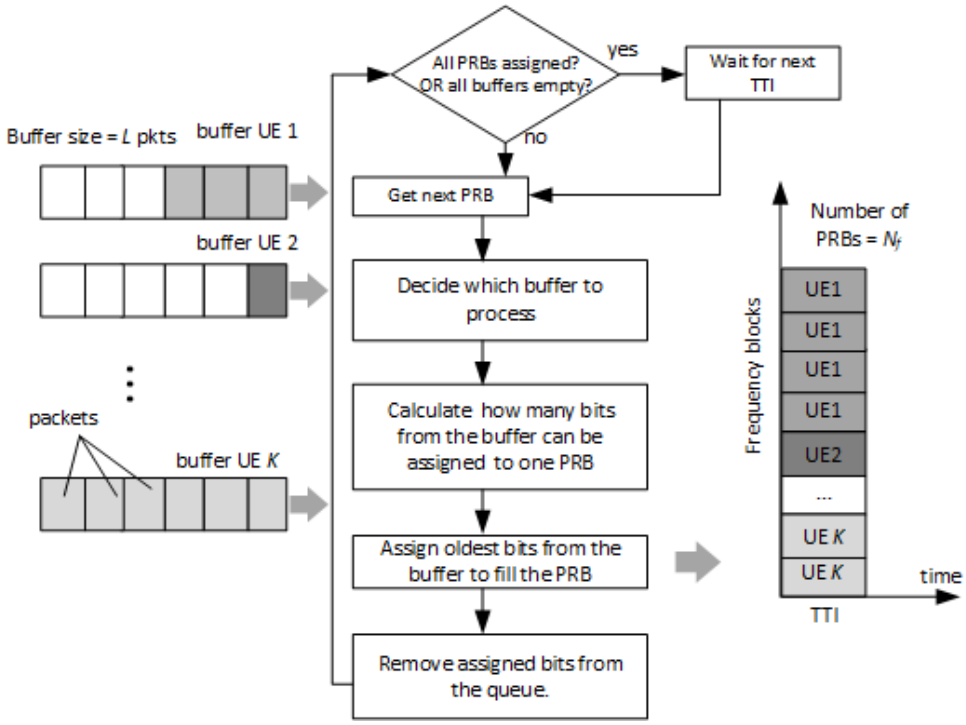


Fig. B.1: The main steps of the resource allocation process.

details about the problem modeling. In Section 3, we derive a solution proposal for the problem while presenting the obtained results in Section 4.

2 Problem Description and Modeling

As specified by the challenge description in [10], on each episode of this environment, the agent must allocate N_f downlink frequency resources to User Equipment (UEs). This takes place in a free-space environment with K UEs, where each UE has specific traffic requirements (some require high guaranteed bit rates, others low packet delivery delays, etc.). The main steps of the allocation are shown in Figure B.1. This recreates a well-known case of OFDM resource allocation, where a MAC scheduler allocates frequency resources to UEs under different radio conditions. The problem is formulated as a Markov Decision Process (MDP) and implemented using Open AI Gym [11]. The MDP properties are specified in the next subsections.

2.1 Environment Dynamics

At the beginning of an episode, a fixed number of K UEs are scattered randomly throughout an empty Euclidean space of size $1\,\mathrm{km}^2$ containing a base station (BTS) centered at coordinates $(0, 0)$. The BTS transmits with EIRP= 13 dBm. The carrier frequency is $f_c = 2655$ MHz and the system bandwidth $BW = 5$ MHz. The transmit power is distributed equally across all Physical Resource Blocks (PRBs). Free space propagation is assumed and the UEs move at random speeds in random rectilinear trajectories throughout the environment (bouncing off the edges at specular angles). The UE speeds are normally distributed as described in [12].

Each episode begins at time step $t = 0$ with $p = 0$ and TTI= 0, where p denotes the current PRB being allocated and TTI is the Transmission Time Interval. One TTI is assumed to last 1 ms exactly. The environment is then time-stepped and the TTI counter is increased by 1 when $p = t \mod N_f = 0$. The environment is run indefinitely (i.e. for a very large number of time steps).

When the environment starts, each UE gets assigned a random QoS Identifier (QI) class from a total of 4 QIs. This assignment is uniform (There are exactly K/4 UEs of the same QI and all QIs are assigned).

On the first time step of each TTI (i.e. when $t \mod N_f = 0$) new traffic packets for each UE are generated according to a random process specific for specific QoS Identifier class (see Table B.1). These packets are then added to the UE traffic buffer. A packet size in a UE buffer decreases each time step according to the UE spectral efficiency and to the number of radio resources allocated by the agent to the UE. The maximum number of packets that each buffer can store (i.e. its buffer size) is defined up to $L = 100$.

2.2 Observation Space

The state vector is a concatenation of vectors providing the following information at each time step:

- Channel Quality Inidicator (CQI) of a UE k: $q_k \in [0, 15] \forall k \in [1, K]$.
- Sizes (in bits) of all packets in each UE buffer: $S = s_{k,l} \in \mathbb{R}^{K \mathrm{x} L}$.
- Ages (in TTIs) of all packets in each UE buffer: $E = e_{k,l} \in \mathbb{R}^{K \mathrm{x} L}$, where $e_{k,l}$ is the age of the l^{th} packet of the k^{th} UE.
- QoS Identifier (QI) classes of each and all UEs as a one-hot vector: $c_k \in [0, 1, 2, 3] \forall k \in [1, K]$. The QI classes are given in Table B.1.
- Current PRB being allocated $p \in [0, ..., N_f - 1]$, which can be calculated from the current time step as $p = t \mod N_f$.

Table B.1: Traffic characteristics of each QoS Identifier class

QI classes	Resource Type	Guaranteed Bit Rate (GBR) (kbps)	Packet Delay Budget (PDB) (ms)
3 or [0,0,0,1]	GBR (Conversational Voice)	29.2	100
2 or [0,0,1,0]	GBR (Conversational Video)	1250	150
1 or [0,1,0,0]	Delay Critical GBR	10	30
0 or [1,0,0,0]	Non-GBR (web browsing)	N.A.	300

2.3 Action Space

On each time step, the agent may take one of K possible actions by assigning the current frequency resource to the k-th UE. If an action is chosen that allocates the current PRB to the k-th UE, the size of the oldest packet in the k-th UEs buffer is reduced by a number of bits equal to the number of transmitted bits. The number of bits transmitted in one PRB depends on the UE channel quality.

2.4 Reward

The agent receives a reward of 0 on all time steps except on those leading to a state wherein $p = 0$. These are called TTI transition time steps.

The reward received on the TTI transition time steps is the negative sum of non-GBR buffer sizes (to encourage the agent to empty the non-GBR queues as fast as possible), plus the negative sum of delay traffic (to encourage the agent to respect the Packet Delay Budgets). Note that this reward is calculated before new traffic is added to the UEs buffers:

$$r_t = r_t^{(GBR)} + r_t^{(nonGBR)} \tag{B.1}$$

where $r_t^{(GBR)}$ and $r_t^{(nonGBR)}$ are defined as:

$$r_t^{(GBR)} = \quad -\sum_{\substack{k=1}}^{K}\sum_{l=1}^{L} S_{k,l} \quad \begin{subarray}{l} \\ e_{k,l}>\mathrm{PDB}_k,\ c_k\in[1,2,3] \end{subarray} \tag{B.2}$$

$$r_t^{(nonGBR)} = -\sum_{\substack{k=1}}^{K}\sum_{l=1}^{L} S_{k,l} - \sum_{\substack{k=1}}^{K}\sum_{l=1}^{L} S_{k,l} \quad \begin{subarray}{l} \\ e_{k,l}>\mathrm{PDB}_k,\ c_k=0 \end{subarray} \quad \begin{subarray}{l} \\ c_k=0 \end{subarray} \tag{B.3}$$

2.5 Evaluation

After proposing a scheduler candidate, the evaluation is performed by running 16 random episodes with a maximum of 65536 time steps each, collecting the reward obtained by the agent on each time step. The result is calculated as the average reward obtained in all time steps on all episodes.

3 Solution Proposal

According to the reward definition in (B.1), we can imply that the ultimate goal is optimizing scheduling decisions in order to minimize the overall UEs buffer ($S_{k,l}$), thus maximizing the reward function.

We propose a methodology where the base station resource allocation decision is based on calculating a metric for transmission priority given the state space information. More precisely, for each time step, the current RB is assigned to the user k^* with the highest priority value among K total users according to the following equation:

$$\forall k \in [1, K]$$
$$k^* = \underset{k}{\operatorname{argmax}} \left\{ \alpha \cdot q_k + \beta \cdot \sum_{l=1}^{L} s_{k,l} + \gamma \cdot \max_{l} \{e_{k,l}\} + \mu \cdot \frac{1}{\mathrm{n}_k} \right\} \tag{B.4}$$

where α, β, γ and μ are priority weights that need to be tuned, and n_k is a fairness measure indicating how many RBs was historically allocated to the user k.

This metric is formulated to explore the observation space information and adapt its weights according to the environment requirements. More precisely, for each user, we basically consider its current CQI value, the total amount of packets in its buffer, the age of the oldest packet and a fairness metric that motivates spreading the RB assignments among all users.

In the next subsections we propose a random black-box optimization to derive tuned values for α, β, γ and μ. The biggest advantage of using such methodology instead of an RL approach is the fact that we now reduce the problem complexity from an amount that is proportional to the number of state-action values (which also grows exponentially with the number of UEs) to a continuous search in $\mathbb{R}^4$. Furthermore, it is interesting to highlight that this priority optimization does not rely on the environment conditions (number of UEs, buffer size, bandwidth, etc.), thus not requiring any re-training or transferring learning from a certain condition to another.

3.1 Random Black Box Optimization

According to the problem definition, we can imply that for any given combination of α, β, γ, and μ, the environment will allocate RBs among the users and eventually provide feedback through the reward function. So, we can formulate the environment dynamics as $f : \mathbb{R}^4 \to \mathbb{R}$ an objective function where we want to maximize the reward defined in (B.1) given the priority weights α, β, γ and μ. The goal is to find candidate solutions α, β, γ, and μ in which the values of r_t over time are as maximum as possible.

As described in [13], black-box optimization refers to the situation where function values of evaluated search points are the only accessible information on f. The restriction here is that we are not directly searching for a global optimum, but this can be

substantially beneficial in situations where this value is neither feasible nor relevant in practice. The pseudo-code 1 illustrates a general methodology of a randomized black-box search. It is interesting to highlight from this pseudo-code that variations in steps 1 and 4 mostly define different approaches for black-box optimization, such as genetic algorithms (GA).

Algorithm 1 Randomized black box search.

Input: Priority metrics: $\alpha, \beta, \gamma, \delta$
Output: Reward cost function after episode completed
Initialize distribution parameters : $\theta^{(0)}$
1: **for** generations g=0,1,2,3,... **do**
2: Sample λ independent points from distribution $P(x|\theta^{(g)}) \rightarrow x_1, \ldots, x_\lambda$
3: Evaluate the sample $x_1, \ldots, x_\lambda$ on f
4: Update parameters $\theta^{(g+1)} = F_\theta(\theta^{(g)}, (x_1, f(x_1), , \ldots, (x_\lambda), f(x_\lambda)))$
5: break, if criterion met
6: **end for**

3.2 Covariance Matrix Adaptation Evolution Strategy

The CMA methodology [13] is a randomized non-linear non-convex black-box optimization, where the search distribution, P, as indicated in step 2 of the pseudo-code 1, is a multivariate normal distribution. The main idea is sampling λ independent candidate solutions and selecting μ off-springs (e.g., the best 20% solutions). We then use these best candidate solutions for estimating the covariance matrix of the next generation. The parameters update (step 4 in pseudo-code 1) determines the overall variance of the mutation at generation g. This property plays a virtual role in the CMA strategy by providing a capacity to adaptively increase or decrease the search space for the next generation. This property is also intuitively interesting because there might be situations where we want to explore more and increase the standard deviation of the search space. However, there can also be situations where we are confident we are close to a good optimum and just want to tune the solution. In fact, this technique overcomes a limitation of traditional black-box optimization algorithms, such as GA, where the standard deviation noise parameter remains always fixed.

The pseudo-code 2 outlines the baseline procedure of CMA-ES. As described in [14], we need to define, for an arbitrary state dimension n, the following state variables:

- $m \in \mathbb{R}^n$, $\sigma > 0$, $C \in \mathbb{R}^{nxn}$, the parameters for a multi-variate normal distribution $\mathcal{N}(m, \sigma^2 C)$.
- $p_s \in \mathbb{R}^n$, $p_c \in \mathbb{R}^n$, two evolution paths, initially set to zero.

Algorithm 2 Pseudo-code for CMA-ES.

Initialize: State variables: m, σ, C, p_s, p_c
1: Set the sizes of parent and offspring population λ and μ.
2: **while** stop condition not met **do**
3: **for** $i \in \{1, \ldots, \lambda\}$ **do**
4: $x_i \leftarrow \mathcal{N}(m, \sigma^2 C)$
5: **end for**
6: sort x_i according to $f(x_i)$
7: m' = m
8: $m \leftarrow$ update-m$(x_1, \ldots, x_\lambda)$
9: $p_s \leftarrow$ update-ps$(p_s, \sigma^{-1} C^{-1/2}(m - m'))$
10: $p_c \leftarrow$ update-pc$(p_c, \sigma^{-1}(m - m', ||p_s||))$
11: $C \leftarrow$ update-C$(C, p_c, (x_1 - m')/\sigma, \ldots, (x_\lambda - m')/\sigma))$
12: $\sigma \leftarrow$ update-sigma$(\sigma, ||p_s||)$
13: **end while**

The way we model the update functions in steps 8 to 12 of the pseudo-code 2 basically define many variants of the CMA strategy. We can refer to its baseline version in [15] and some important extensions for handling fitness noise in [16] and multi-modality in [17].

3.3 Optimizing Resource Scheduling using CMA-ES

We applied CMA strategy to optimize the tuning parameters defined in (B.4), thus maximizing the reward function defined in (B.1). The CMA-ES implementation was carried using the open-source library PyBrain [18], considering a maximum of 200 evaluations with exponentially decreasing step sizes. The environment scenario was considered to be using the default parameters established by Nokia:

- Number of UEs: 32
- Number of Physical Resource Blocks (PRBs): 25
- Buffer size: 8 packets
- Effective Isotropic Radiated Power : 13 dBm
- Carrier frequency: 2655 MHz
- Maximum packet size: 41250 bits
- Mean inter-arrival time of non-GBR traffic: 10 TTIs

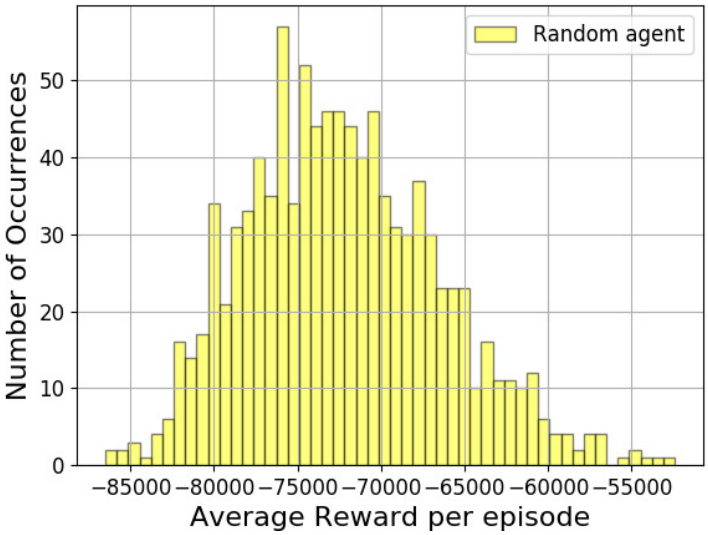


Fig. B.2: Random Scheduler.

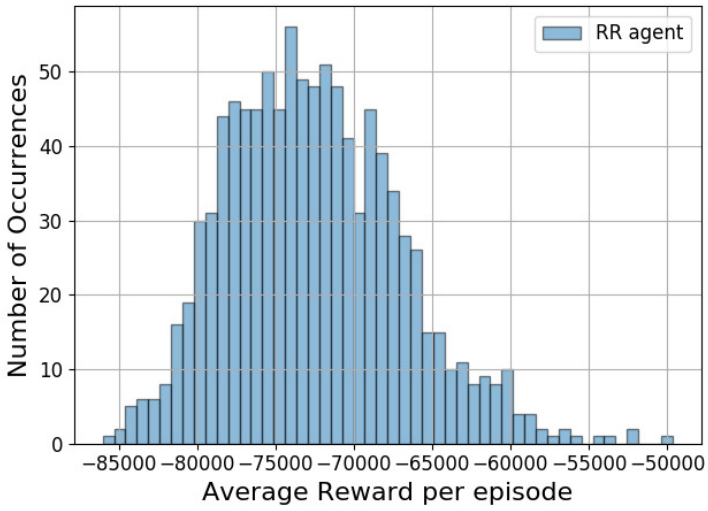


Fig. B.3: Round Robin Scheduler.

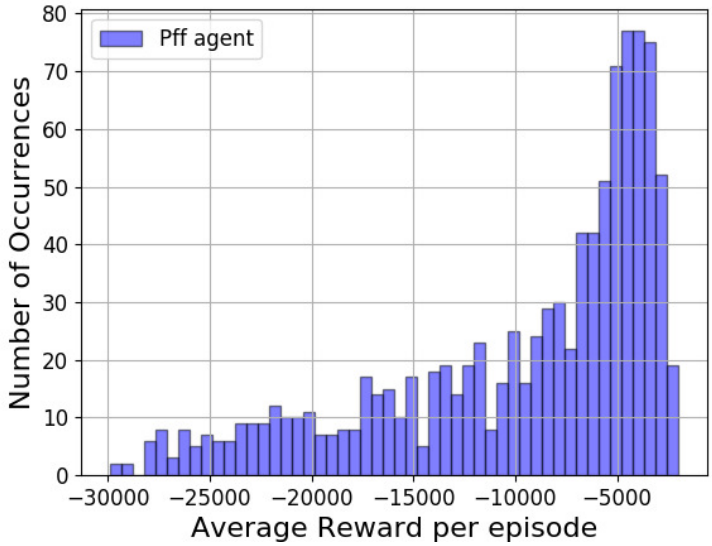


Fig. B.4: Proportional Fair Scheduler.

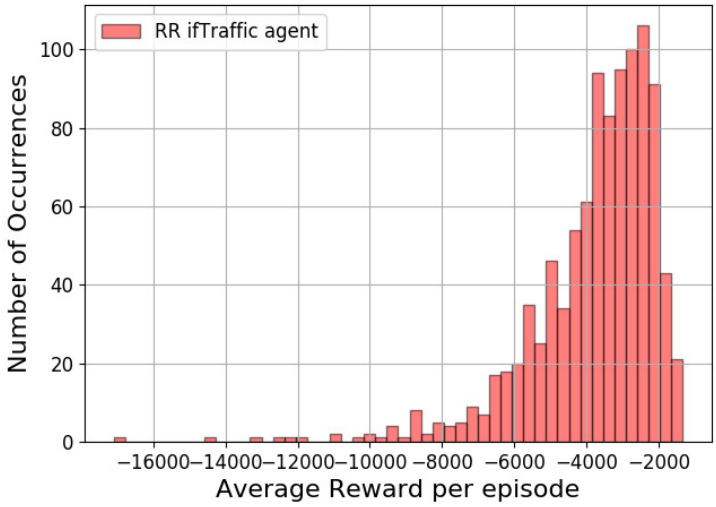


Fig. B.5: RR If Traffic scheduler.

The source code for the proposed solution after optimization can be found as *bosch_agent.py* in the Nokia repository [10]. In the next section, we compare the proposed solution result with the baseline algorithms and RL approaches.

4 Simulation Results

We compared the CMA-ES reward function performance with baseline algorithms provided by Nokia, such as Round Robin (RR) [19], Round Robin If Traffic (RR IFtraffic), which is a variant of RR that checks the need of transmission before allocating resources, Proportional Fair (Pff) [19] and a Random scheduler that decides which user to allocate RBs according to a uniform random variable. Their implementations are available in the source code in [10].

We also evaluated the performance of advanced deep reinforcement learning techniques, such as Actor-Critic (A2C) [20] and Proximal Policy Optimization (POP) [21]. Both implementations were carried using Open-AI Stable Baselines [22] using MLP policy and 400000 time steps for training over the Nokia environment.

4.1 Baseline Results

The Figures B.6, B.7, B.8 provide the simulation results for 1000 random episodes of each baseline algorithm. The mean and standard deviation of each distribution is summarized in Table B.2. As we can note, the Round Robin IfTraffic provides the best reward performance, with an average of -3781.09 over all timesteps and all episodes.

4.2 Deep RL and CMAES Results

Table B.2: Reward performance of each scheduler after 1000 epochs evaluation.

	RR	**Random**	**Pff**	**RR IF Traffic**	**A2C**	**POP**	**CMAES**
Mean:	-72547.94	-72286.80	-9717.30	-3781.09	-72240.33	-73803.17	-1531.79
Std:	5608.97	5807.31	6808.53	1769.90	6778.20	6642.11	234.63

The Figures B.6, B.7, B.8 and Table B.2 summarize the results for 1000 random episodes of the Deep RL techniques and the proposed CMA-ES. It was surprising that the Deep RL approaches, besides the long training duration, were unable to overcome even round-robin. We can presuppose some specific reasons for this poor performance, such as high variance at the value function estimation and over-fitting in some environment patterns, causing the algorithm to work under sub-optimal performance. To cope with such issues, we can think of advanced hyperparameter optimization strategies and more complex network architecture, but at the cost of higher training complexity. It is

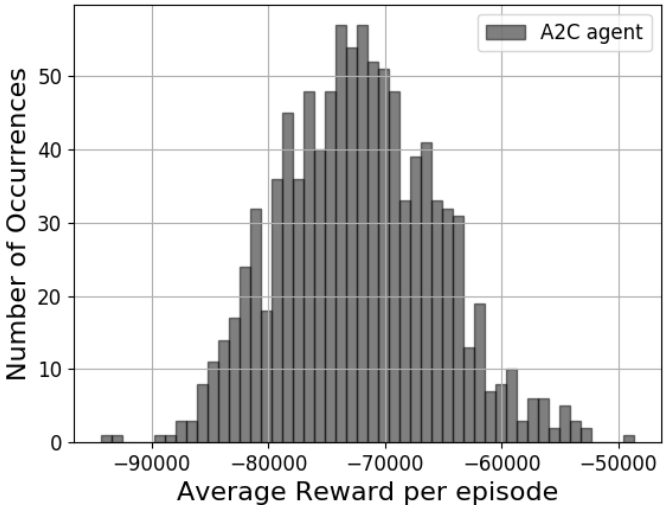


Fig. B.6: A2C RL Scheduler.

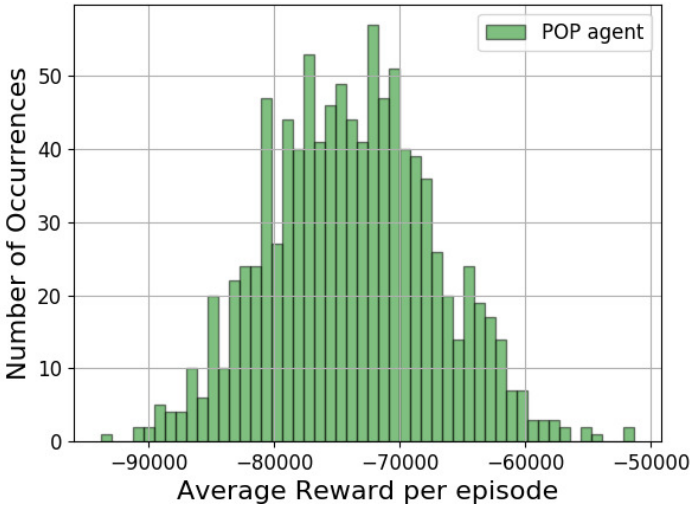


Fig. B.7: POP RL Scheduler.

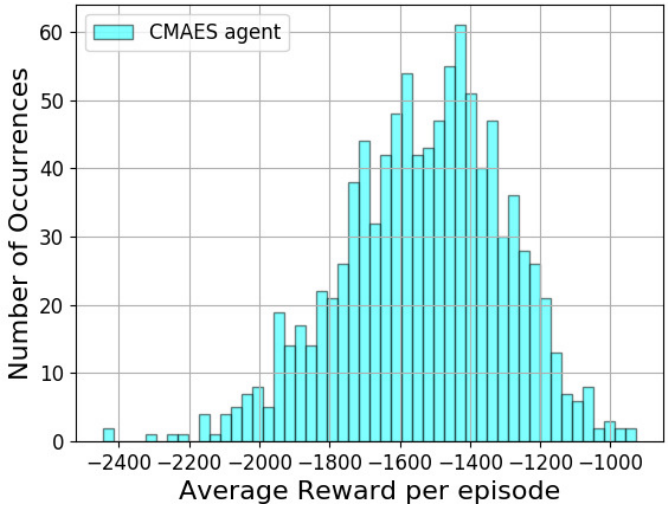


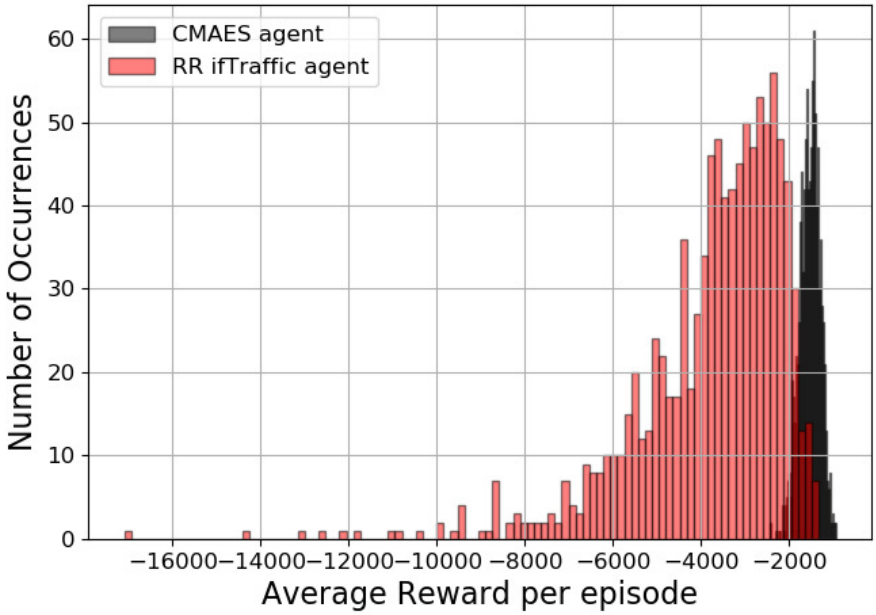
Fig. B.8: CMA-ES Scheduler.

important to emphasize, however, that, as also explored in [4], it might still be an open research question of how RL can properly be applied in such problems.

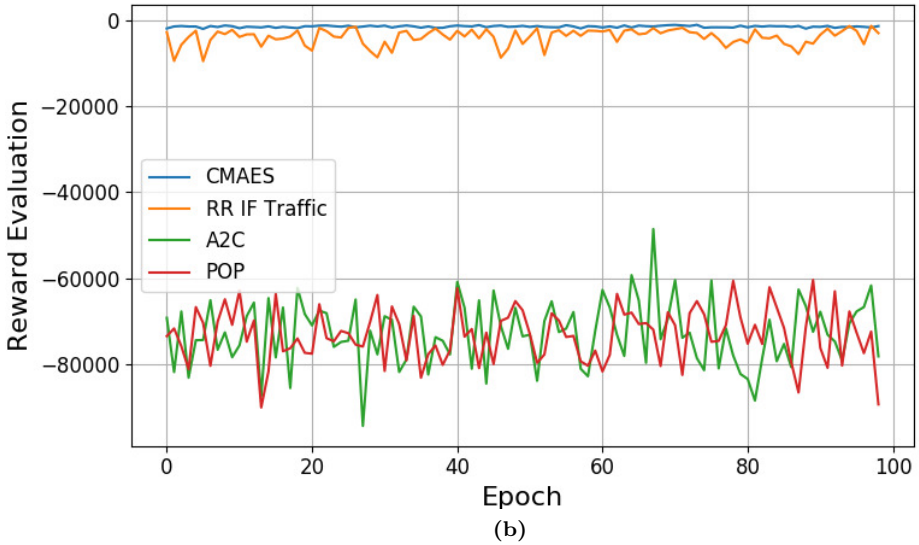
On the other hand, the CMA-ES has shown promising performance, surpassing all the baseline algorithms, including the RR IfTraffic. The Figure B.9 provides a more detailed comparison of CMA-ES performance, especially its low variance demonstrated in Figure B.9a. In Figure B.9b, we see the performance variation in the first 100 episodes.

5 Conclusions

In this paper, we elaborated on a solution for an open source problem proposed by Nokia Bell Labs regarding network resource allocation. We defined a priority metric calculation and applied a black-box optimization approach technique, named CMA-ES, to tune its parameters. The results show that the proposed approach was able to overcome traditional baselines results provided by Nokia, as well as Deep Reinforcement Learning techniques.



(a)



(b)

Fig. B.9: (a) Performance comparison between CMA-ES and RR IfTrafiic. (b) Results for the first 100 random episodes.

References

[1] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE-Advanced: A tutorial," 2015.

[2] E. Pateromichelakis, M. Shariat, A. u. Quddus, and R. Tafazolli, "On the evolution of multi-cell scheduling in 3GPP LTE / LTE-A," 2013.

[3] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," 2013.

[4] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine learning for resource management in cellular and IoT networks: Potentials, current solutions, and open challenges," *IEEE Communications Surveys Tutorials*, 2020.

[5] Y. Luo, Z. Shi, X. Zhou, Q. Liu, and Q. Yi, "Dynamic resource allocations based on Q-learning for D2D communication in cellular networks," in *Proc. International Computer Conference on Wavelet Actiev Media Technology and Information Processing (ICCWAMTIP)*, 2014.

[6] T. Yang, Y. Hu, M. C. Gursoy, A. Schmeink, and R. Mathar, "Deep reinforcement learning based resource allocation in low latency edge computing networks," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, 2018.

[7] I. Ahmed and H. Khammari, "Joint machine learning based resource allocation and hybrid beamforming design for massive mimo systems," in *Proc. IEEE Globecom Workshops*, 2018.

[8] P. M. de Santana, V. A. de Sousa, F. M. Abinader, and J. M. d. C. Neto, "DM-CSAT: A LTE-U/Wi-Fi coexistence solution based on reinforcement learning," *Telecommunication Systems*, 2019.

[9] P. M. de Santana, J. M. C. Neto, F. M. Abinader Jr, and V. A. de Sousa Jr, "GTDM-CSAT: an LTE-U self coexistence solution based on game theory and reinforcement learning," *Journal of Communication and Information Systems*, 2019.

[10] A. Valcarce, "Wireless Suite: A collection of problems in wireless telecommunications," 2020. [Online]. Available: https://github.com/nokia/wireless-suite

[11] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI Gym," 2016.

[12] "Speed distribution curves for pedestrians during walking and crossing," *Procedia - Social and Behavioral Sciences*, 2013.

[13] N. Hansen, "The CMA evolution strategy: A tutorial," *CoRR*, 2016.

[14] O. Krause and T. Glasmachers, "A CMA-ES with multiplicative covariance matrix updates," 2015.

[15] S. Kern, S. Muller, N. Hansen, D. Buche, J. Ocenasek, and P. Koumoutsakos, "Learning probability distributions in continuous evolutionary algorithms - a comparative review," *Natural Computing*, 03 2004.

[16] N. Hansen, A. S. P. Niederberger, L. Guzzella, and P. Koumoutsakos, "A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 1, 2009.

[17] A. Auger and N. Hansen, "A restart CMA evolution strategy with increasing population size," in *Proc. IEEE Congress on Evolutionary Computation*, 2005.

[18] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber, "PyBrain," *Journal of Machine Learning Research*, 2010.

[19] NS3. LTE resource allocation model. [Online]. Available: https://www.nsnam.org/docs/models/html/lte-design.html

[20] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," 2016.

[21] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[22] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, and Y. Wu, "Stable baselines," https://github.com/hill-a/stable-baselines, 2018.

Paper C

Age of loop for wireless networked control systems optimization

Pedro M. de Sant Ana, Nikolaj Marchenko, Beatriz Soret and Petar Popovski

The paper has been published in the
IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2021.

The layout has been revised.

Abstract

Joint design of control and communication in Wireless Networked Control Systems (WNCS) is a promising approach for future wireless industrial applications. In this context, Age of Information (AoI) recently has been proposed as a metric that is more representative than communication latency in conduct of systems with a sense-compute-actuate cycle. Nevertheless, AoI is commonly defined for a single communication direction, Downlink or Uplink, which does not capture the closed-loop dynamics. In this paper, we extend the concept of AoI by defining a new metric, Age of Loop (AoL)*, relevant for closed-loop WNCS problems. The AoL is defined as the time elapsed since the piece of information causing the latest action or state (depending on the selected time origin) was generated. We use the proposed metric to learn the WNCS latency and freshness bounds, and apply such learning methodology to minimize the long-term WNCS cost with the least amount of bandwidth. We show that, using the AoL, we can learn the control system requirement and use this information to optimize network resources.*

1 Introduction

Networked control systems (NCS) are an essential part of many industrial domains such as factory automation, logistics, or transportation. Wireless NCSs (WNCS) enable mobile control applications where wiring is not possible or high flexibility is required. However, due to the nature of the wireless medium, reliability of WNCS remains an open challenge, in particular for low-latency applications.

In a conventional approach, one would separately derive worst case requirements from the control system and impose them on the communication system. Communication latency metric is then often used as a benchmark metric to design and evaluate the communication system. However, such decoupling in system design for low-latency and high reliability leads to over-provisioning communication network resources.

Furthermore, actions taken at a control system level can have a direct impact on the communication system and vice-versa, as formulated by Witsenhausen as counterexample for distributed control problem in [1] and exemplified for an autonomous guided vehicle (AGV) use case in [2].

For systems with a sense-compute-actuate cycle like the ones considered in this paper, where the receiver is interested in fresh knowledge of the remotely controlled system, rather than the individual packet delay, the notion of Age of Information (AoI) [3] has been proposed as more representative than communication latency. The AoI defines the time that has elapsed since the newest update available at the destination was generated at the source, and it captures not only the communication delays but also the impact of the packet generation at the controlled process.

In recent work on WNCS, authors have been increasingly exploring the inter-relation

between the control and communication systems with help of AoI. In [4] and [5], for example, the authors demonstrate how the latency and reliability trade-off directly impacts the system level stability, proposing a co-design of both control and communication entities. Specifically in [4], authors have demonstrated a counter-intuitive conclusion that the plant can still be stabilized with an arbitrarily large delay under certain channel conditions. Another interesting finding was presented in [6], where authors elucidate an example how one can optimize long-term system performance by assuming more risks with less reliable transmissions in exchange for lower latency.

Despites its benefits, the drawback of AoI is that it has been used so far to separately optimize transmissions in uplink (UL) and downlink (DL). However, WNCS applications are closed-loop applications, where the UL communication can affect the DL and vice-versa, resulting in system performance changes or in the use of network resources. In this context, we propose to explicitly address the two-way nature of the control-communications interplay.

This paper contains two main contributions:

1. We propose a new metric, the Age of Loop (AoL), which extends the current AoI definition to take into consideration both UL and DL of the control loop in WNCS, and thus can provide a more precise system state estimation.

2. We demonstrate how to apply the AoL metric for joint WNCS optimizations with the application example of a remotely controlled inverted pendulum system [7]. With a Reinforcement Learning (RL) approach, we find the bandwidth allocation policy based on the AoL state, which significantly outperforms policies based on fixed latency requirements.

The rest of this paper is organized as following: in the next two sections, we introduce the system and WNCS model, respectively. In Section 3, we define the AoL and show how we can evaluate the control system performance using the proposed metric. Finally, in Section 5, we formulate the bandwidth allocation problem and propose a solution, where the results are analyzed in Section 6.

2 System Model

We consider the classical inverted pendulum system model, a widely used benchmark problem in both control and RL domain. As illustrated in Figure D.1, a pole is attached by a joint to a cart, which can be moving along a frictionless track. The pendulum starts upright at a random initial angle $\theta_0 \in (\theta_{0,min}, \theta_{0,max})$, and the goal is to prevent it from falling over by applying a force to the cart. While conceptually simple, the system dynamics are highly unstable and continuously requires fast control cycles to keep stability. When, in turn, being controlled over a wireless channel, the problem becomes an illustrative model of strict timing requirement.

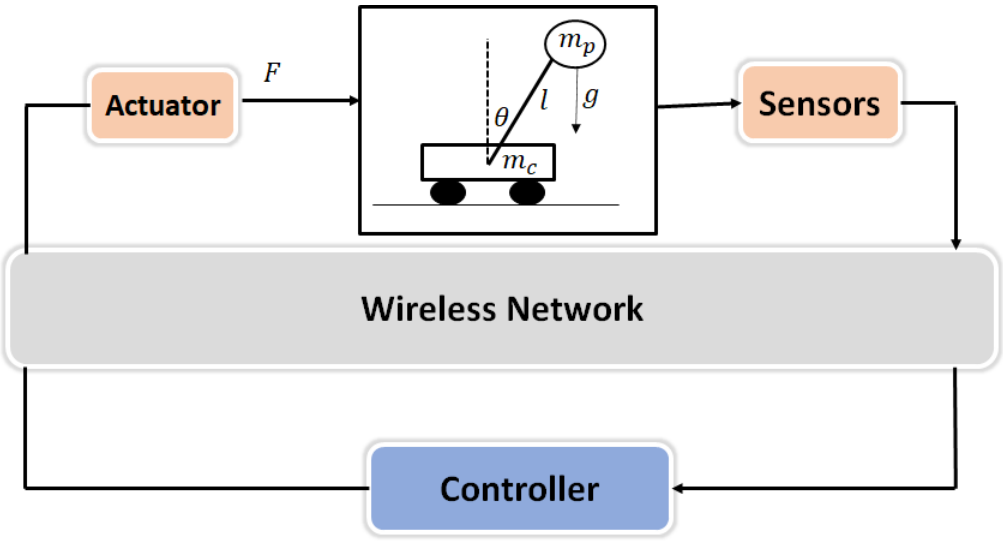


Fig. C.1: Inverted pendulum system model.

2.1 Control System Model

The system dynamics can be represented by the differential equations [8]:

$$\begin{aligned} \ddot{\theta} &= \frac{g \cdot \sin(\theta) + \cos(\theta)\left(\frac{-F - m_p l \dot{\theta}^2 \sin(\theta)}{m_c + m_p}\right)}{l\left(\frac{4}{3} - \frac{m_p \cos^2(\theta)}{m_c + m_p}\right)}, \\ \ddot{x} &= \frac{F + m_p l(\dot{\theta}^2 \sin(\theta) - \ddot{\theta}\cos(\theta))}{m_c + m_p}, \end{aligned} \tag{C.1}$$

where x and θ are, respectively, the cart position coordinates and the pole angle from vertical reference. The mass of the cart is m_c, and the mass of the pendulum is m_p, while l is the length of the pendulum, and F is the force applied to the cart under gravity g. We use the Newton's notation ($\dot{\square}, \ddot{\square}$) to represent derivatives w.r.t time.

By defining a state space vector $X = [x, \dot{x}, \theta, \dot{\theta}]$, we can design a standard optimal controller in two steps. First, computing the Jacobian of (D.1) around the operating point $X = [0, 0, 0, 0]$ to linearize the plant, so that the system dynamic takes the linear time invariant form:

$$\begin{cases} \dot{X} = AX + Bu + w, \\ u = -KX, \end{cases} \tag{C.2}$$

where u is the linear state feedback control policy of gain K, w is a process disturbance modeled as a zero-mean and one-standard deviation Gaussian white noise, A and B are

the system transition matrix, respectively calculated as [9]:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-12m_p g}{13m_c+m_p} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{12(m_p g+m_c g)}{l(13m_c+m_p)} & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ \frac{13}{13m_c+m_p} \\ 0 \\ \frac{-12}{l(13m_c+m_p)} \end{bmatrix}. \tag{C.3}$$

The second step consists of finding the optimal control policy, u^*, subject to (D.2) that minimizes the cost function,

$$J(u) = \int_0^\infty \left(X^T Q X + u^T R u\right) dt, \tag{C.4}$$

where R and Q are arbitrary positive defined matrices in which we can assign weights to state space variables and control signal. In control theory this kind of problem formulation is known as Linear-Quadratic-Regulator (LQR) [10].

The optimal control policy then can be defined by solving the Algebraic Riccati Equation [10] as:

$$\begin{aligned} &A^T P + PA - PBR^{-1}B^T P + Q = 0, \\ &K^* = R^{-1}B^T P, \\ &u^* = K^* X. \end{aligned} \tag{C.5}$$

For (A, B) controllable, the infinite horizon LQR with $Q, R > 0$ gives a convergent closed-loop system [10], where the stability can be easily guaranteed.

2.2 Networked Control Model

As defined in [2], we adopt a similar NCS model to define the system behavior over the wireless medium operating in Frequency Division Duplexing (FDD) mode with separated frequency bands for the uplink (UL) and downlink (DL) directions, which makes the medium access for UL and DL independent from each other in time domain. Figure D.2 illustrates the proposed model, showing the details of the interaction between the communication and application control loop. First, the sensor readings of the application describe the system states, X_i, which are stored in memory and communicated to the controller over the uplink channel. The readings and transmissions of sensor values are done strictly periodically with the cycle time ΔT_{in}, as it is commonly done across various control systems [11].

At the controller, the received sensor values are also stored into the memory. The control application gets the recent values, and produces a control signal u_i according to (E.4). Immediately after producing a control command, the controller sends it over a downlink channel to the controlled system. After finishing the current transmission,

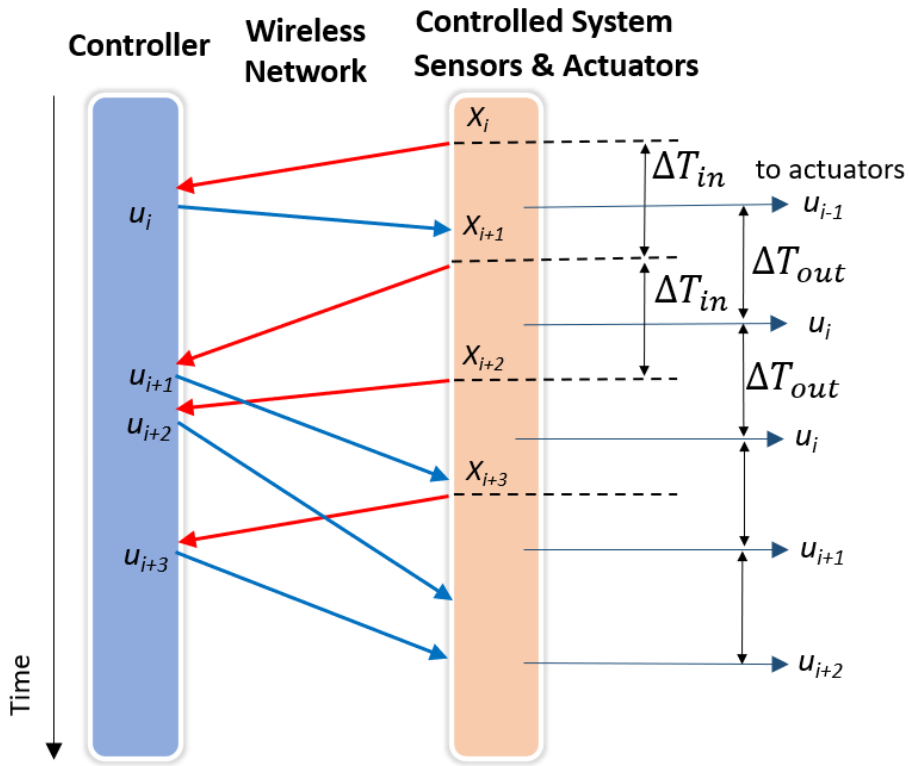


Fig. C.2: WNCS Model.

the controller keeps waiting for the next state update from the controlled device, and starts the procedure once again.

At the controlled system side, the received command u_i is stored in the memory. The output application for actuators control (e.g., motor drives) is called periodically with the time interval ΔT_{out}, calls the most recently stored command values from the memory and applies them to the application drives, producing the system dynamics of (D.1).

2.3 Wireless channel model

Both the DL and UL transmissions can suffer latency while delivering the information, which, in this model, depends on two main factors: the current channel quality and the total bandwidth allocated for the transmission. To evaluate this behavior, we consider the 3GPP 4-bit CQI Table 7.2.3-1 [12], where we can estimate the amount of time to deliver the data considering both the channel quality indicator (CQI) and the total bandwidth allocated at the transmission. The following two assumptions have been adopted: a) the UL finishes its transmission within ΔT_{in}, and b) the DL only starts a new transmission after finishing the current one. The details of the bandwidth allocation problem are discussed later in Section 5.

2.4 System Model Discussion

It is important to emphasize that (F.7) is guaranteed to be bounded according to the Riccati-equation [10]. However, the combination of two main factors might affect the system LQR performance. The first is the uplink effect, which represents the level of knowledge the controller has about the plant, meaning that, if ΔT_{in} is too high or the uplink takes overly long to deliver sensor data, the controller will compute the control signal using old state feedback, causing the control command to be ineffective or even harmful for the plant. The second is the downlink effect, or simply the overall delay to deliver the control signal. This is important because if the plant applies outdated control commands for too long, the stability of the controlled system might also be compromised.

Each of these factors might affect the plant in different ways and cannot be independently decoupled, which means that a delay in the UL will impact the DL transmission, provoking cumulative effects at the plant and at the network resources.

3 Age of Information and Age of Loop

Age of information (AoI) provides a measure for quantifying the freshness of the knowledge we have about the status of a remote system. It represents the time duration between the generation time of the freshest received data and the current time. We can

refer to its formal definition as in [3, 13], where, at time t, if the newest data (i.e., with the largest generation time) received at the destination was generated at time $U(t)$, the AoI $\Delta(t)$ is defined as $\Delta(t) = t - U(t)$.

The formal AoI definition, however, is inherited to a single communication link. Papers which so far explored WNCS related problems using AoI are limited to specific analysis over only the UL [5, 14, 15] or DL [4, 6] transmissions. However, wireless networked control systems, as the one considered in this paper, rely intrinsically on both DL and UL with a closed-loop, where the UL communication can affect the DL and vice-versa, impacting system performance and the use of network resources. A simple intuitive example that can illustrate this idea is that a high UL AoI implicates less knowledge that the controller has about the plant, which demands more urgency to deliver the control signal and, as a consequence, more network resources usage by the DL link. To address this implications, we propose a new metric to evaluate the overall age of a WNCS closed-loop, which we refer to as Age-of-Loop (AoL).

Specifically, we can first verify that the state values X_i are periodically generated and transmitted at time intervals of $t_i = \{i \cdot \Delta T_{\text{in}}\}, \forall i \in \mathbb{N}^+$, , where we can define $\{X_i, t_i\}$ the sequence of generated state values and its respective time step. The control signal, in turn, is asyncrhonous and must finish the current DL transmission to start a new one upon reception of a new status update. We can define a sequence $\{u_j, \hat{t}_j\} \forall j \in \mathbb{N}^+$ with aperiodically generated control commands u_j at time step $\hat{t}_j$. If $\{X_i, t_i\}$ is the freshest state feedback that spawned a new control signal, we can extend the DL transmission definition to include state time information, i.e., DL : $\{u_j, \hat{t}_j, t_i\}$. Reciprocally, every state feedback also occurs under the input of the freshest control command, so that we can also extend the UL transmission definition to include control time information, i.e., UL : $\{X_i, t_i, \hat{t}_j\}$.

We consider two plausible definitions of the AoL depending on the selected time origin: the DL-AoI is DL-initiated, meaning that the time origin is a new control command; the UL-AoI is UL-initiated, i.e., the time origin is a new status update in the UL. The DL AoL metric captures the time elapsed since the control command that led to the latest received update in the controller was generated. Analogously, the UL AoL metric refers to the time elapsed since the status update that caused the latest applied control command was generated at the sensor. Mathematically, if the origin is the DL, the current AoL is the difference between the current time t and the time when the freshest received control command was generated:

$$\text{DL AoL}(t) = t - \hat{t}_j. \tag{C.6}$$

Likewise, if the time origin is the UL, the AoL is calculated as the difference between the current time and the time when the freshest received state was spawned:

$$\text{UL AoL}(t) = t - t_i. \tag{C.7}$$

Essentially, the main idea of AoL is to establish a metric that encompasses the behavior

of two separated and locally measured entities (DL and UL) into a single instance, in which we can observe from different perspectives. It is important to note that, in the case of two independent AoI links, we inherently need an instantaneous and perfect feedback channel to the source to know the instantaneous age at the destination, thus making complex and potentially imprecise the union of two directions; AoL fixes this. In practice, it also offers the possibility to design solutions that enclose the whole closed-loop behavior by checking the loop age from either an UL or DL perspective. For example, we can potentially design a power allocation policy for the UL by observing the current UL AoL status. Likewise, we are able to define a modulation coding scheme algorithm for the DL transmissions by observing the DL AoL. It will be proven that they are both valid to optimize the stability of the WNCS. To illustrate the proposed concept, Figure C.3 shows a representative time diagram of the AoL behavior.

4 AoL Evaluation

We aim to estimate the performance of the control system measured according to (F.7) using the current AoL status calculated at the controller (DL AoL). More formally, we can use the value function definition [16] to estimate the expected LQR cost, i.e.,

$$V(\Delta_{\text{AoL}}(t)) = \int_{t}^{\infty} (X^T Q X + u^T R u) dt. \tag{C.8}$$

Since the control policy is unchangeable over time and the plant operation is sampled at cycles of ΔT_{out}, (C.8) becomes the recursion problem:

$$\begin{aligned} V(\Delta_{\text{AoL}}(t)) = & \int_{t}^{t+\Delta T_{\text{out}}} (X^T Q X + u^T R u) dt \\ & + V(\Delta_{\text{AoL}}(t + \Delta T_{\text{out}})), \end{aligned} \tag{C.9}$$

where we can solve iteratively using a temporal difference (TD) learning algorithm [16] with actual state transitions, such that:

$$\begin{aligned} V(\Delta_{\text{AoL}}(t)) \leftarrow V(\Delta_{\text{AoL}}(t)) + \alpha \Bigg[& \int_{t}^{t+\Delta T_{\text{out}}} (X^T Q X + u^T R u) dt \\ & + \gamma V(\Delta_{\text{AoL}}(t + \Delta T_{\text{out}})) - V(\Delta_{\text{AoL}}(t)) \Bigg], \end{aligned} \tag{C.10}$$

where α and γ are, respectively, the learning rate and the discount factor of future values. We can emphasize that (C.10) converges asymptotically to the correct predictions with probability 1 if the step-size α decreases according to the usual stochastic approximation conditions [16].

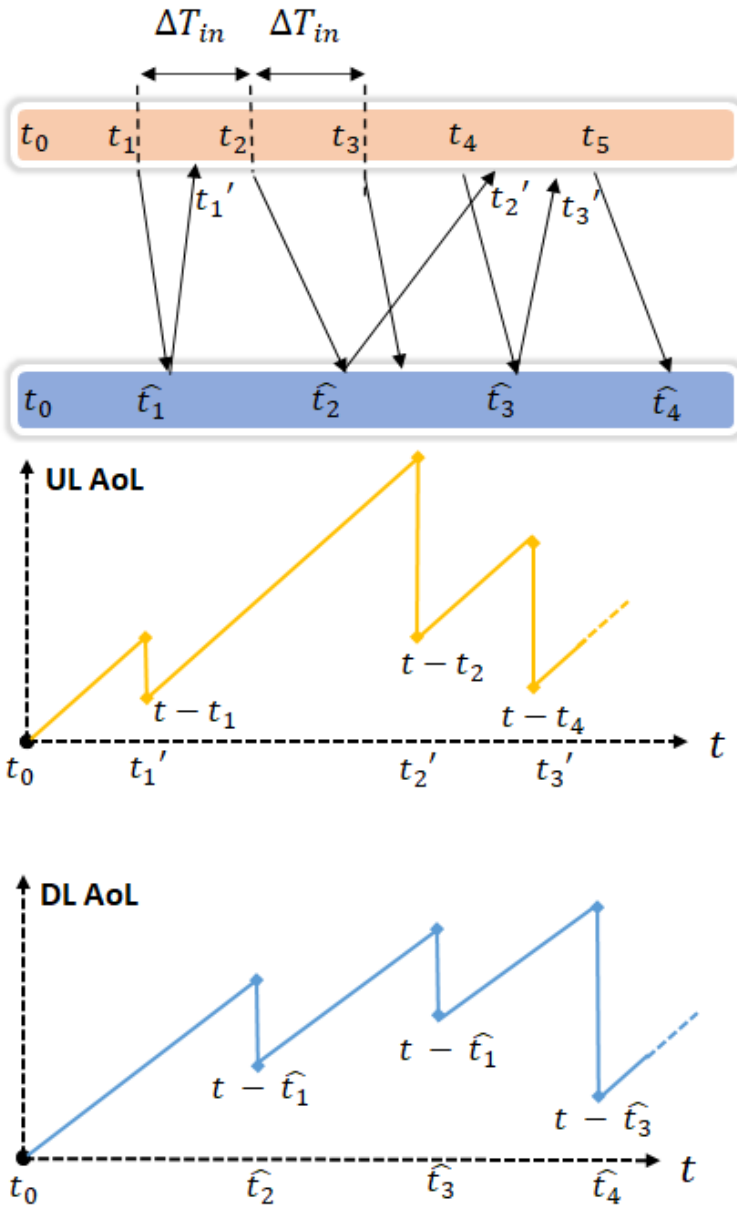


Fig. C.3: Time Diagram of AoL Behavior.

4.1 Numerical Evaluation

Considering the following inverted pendulum configuration: $m_c = 1.0\,\text{kg}$, $m_p = 0.1\,\text{kg}$, $l = 0.5\,\text{m}$, $g = 9.8\,m/s^2$ and $\Delta T_{\text{out}} = 1\,\text{ms}$, we evaluated the expected LQR behavior for different AoL states using (C.10).

Figure C.4 illustrates the obtained result, where we can emphasize three extensive conclusions. First, low AoL values, as expected, provide the best system performance, such that the theoretical LQR upper bound is achieved if the AoL is close to zero. Second, prior to an AoL around 40 ms, the LQR slightly decrease. After that point, however, the system starts to progressively become less tolerable to additional AoL delays. The third and most relevant conclusion is the fact that between 10 and 40 ms, there is no considerable variation at the system performance, meaning that we can avoid over-provisioning network resources by learning the system robustness.

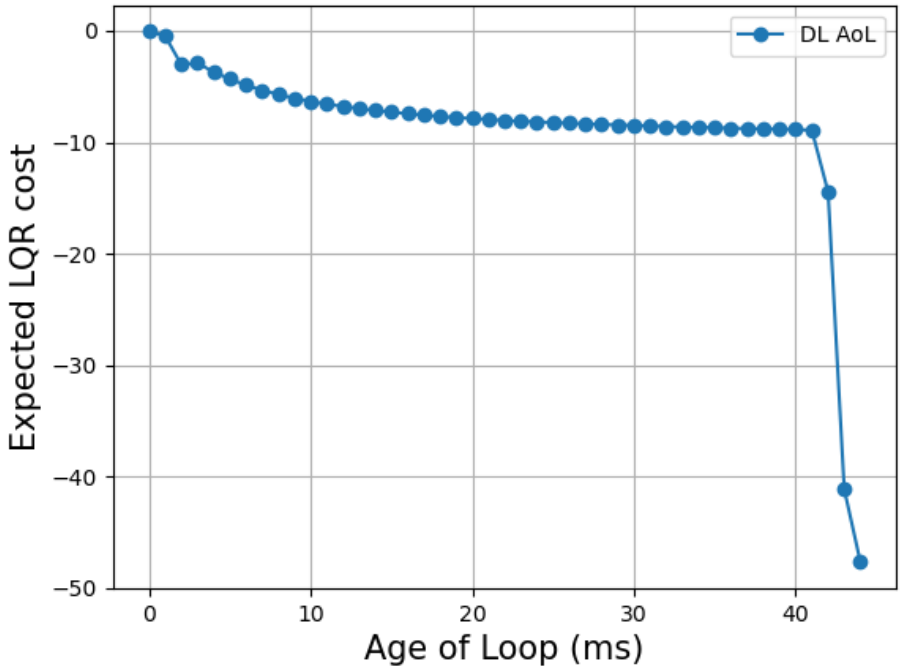


Fig. C.4: Expected LQR Cost vs Age of Loop.

4.2 AoI vs AoL

We performed the same numerical evaluation using a state space comprised of a single DL AoI or a single UL AoI. The goal is to verify the estimation error of the value function when we change the state space for a single AoI metric instead of AoL. We analyze the value function estimation by verifying the Temporal Difference (TD) error, given by:

$$\int_{t}^{t+\Delta T_{\text{out}}} (X^T Q X + u^T R u) dt + \gamma V(\Delta_{\text{AoL}}(t + \Delta T_{\text{out}})) - V(\Delta_{\text{AoL}}(t)), \tag{C.11}$$

which indicates, for each state, how far the predicted value function deviates from the actual value. For example, the learning rule in (C.10) adjusts state value in a direction that tends to reduce the TD error.

Lower TD errors indicates better accuracy about the value function estimation over each state, which is ultimately important for learning better policies, especially in RL context. Figure C.5 illustrates the obtained result along training episodes. We can verify that, as expected, the estimated values are more precise when the whole loop age is considered. The DL and UL AoL values can be merely different, especially because of ΔT_{in}. After generation, the DL data might spend time between state transmissions before finishing the loop, which does not happen in UL case. Thus explaining the slight different behavior of both values in Figure C.5.

5 Bandwidth Allocation Problem

As discussed in section 4, the AoL status of a WNCS can provide an estimation of the system LQR performance, so that we can use the learned value function to build a policy. In this work, we explore the bandwidth allocation problem of a remote controller, where two main objectives must be satisfied: minimize the LQR cost while using the minimum amount of bandwidth.

In more details, we can define $B = \{b_1, b_2, \ldots, b_i, \ldots, b_N\}, b_{i+1} > b_i$ a set of bandwidths in which the controller, for every DL transmission, must decide for a certain bandwidth allocation $b \in B$ given the current AoL state information and the current channel quality. So, for $T = \{t_1, t_2, \ldots, t_i, \ldots t_N\}\, t_{i+1} > t_i$ the time instances where control packets starts transmission and $C = \{c_1, c_2, \ldots, c_i, \ldots, c_N\}$ the corresponding CQI of each transmission, the goal is to find an allocation policy $\pi : \{\Delta_{\text{AoL}}(t_i), c_i \rightarrow b_i\}$, $\forall t_i \in T$, $\forall c_i \in C$, $\forall b_i \in B$ that minimizes the infinite-horizon LQR cost plus the amount of

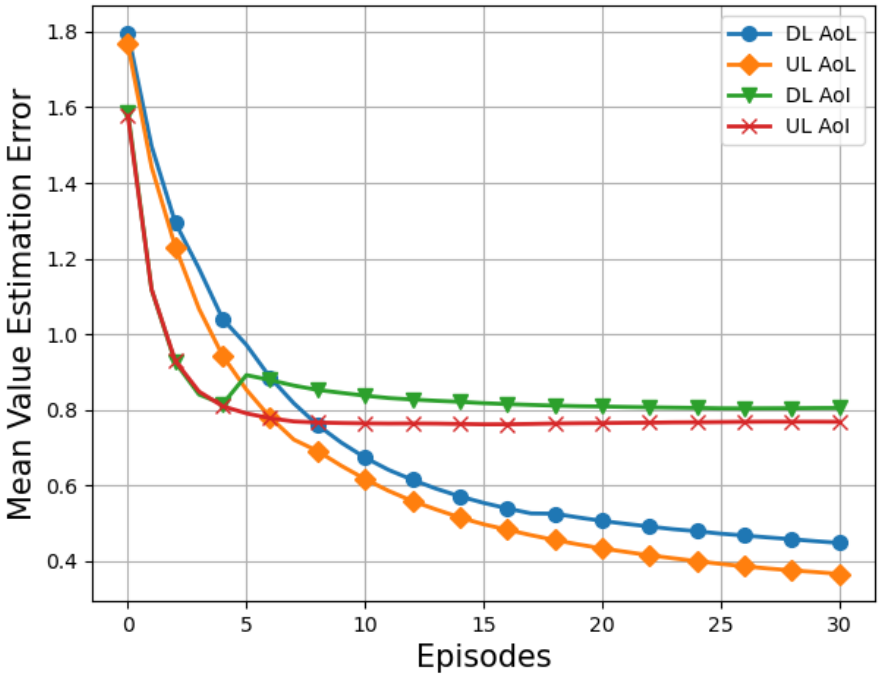


Fig. C.5: Mean value function estimation error over training episodes.

bandwidth usage over the system trajectory, i.e.,

$$\pi^* = \arg\min_{\pi} \left(\int_0^{\infty} (X^T Q X + u^T R u) dt + \sum_{i=1}^{N} \frac{b_i}{b_N} \right),$$
$$\text{s.t. } (D.1), (E.4). \tag{C.12}$$

5.1 Solution Proposal

We can decompose the problem in (C.12) into sub-problems, where between two consecutive control transmissions $[t_i, t_{i+1}), \forall t_i \in T$, we select at t_i a bandwidth $b_i \in B$ based on the AoL and CQI state $\{\Delta_{\text{AoL}}(t_i), c_i\}$. Receiving, as consequence, a one-stage decision cost of:

$$\int_{t_i}^{t_{i+1}} (X^T Q X + u^T R u) dt + \frac{b_i}{b_N}, \tag{C.13}$$

which depends only on the present state and the decision taken on that state. Such decision-making model is a typical Markov Decision Proces (MDP) [16], where we can optimally solve each sub-problem with actual state transitions and overlap those solutions to build the overall optimal solution. In this context, we can define the following MDP configuration:

State Space

Comprised of 20 AoL values, each representing regions of 5 ms from 0 to 100 ms. In addition, 15 possible CQI values for each AoL, resulting in a total of 300 states.

Action Space

Represented by the bandwidth set with ten possible values: $B = \{100, 200, 300, \ldots, 1000\}$ kHz.

Reward

The immediate cost as defined in (F.18).

Scenario

We evaluate the proposed MDP considering the NCS model described in section 2.2, assuming the following inverted pendulum configuration: $m_c = 1.0\,\text{kg}$, $m_p = 0.1\,\text{kg}$, $l = 0.5\,\text{m}$, $g = 9.8\,\text{m/s}^2$, control packet size of 1024 bits and $\Delta T_{\text{out}} = 1\,\text{ms}$. For each run, the CQI is randomly chosen $\{1, 2, 3, \ldots, 15\}$. The evaluation is also performed under different sensor feedback $\Delta T_{\text{in}} = 1, 5, 10, 15$ and 20 ms.

To solve the proposed MDP, we advocate a RL methodology for two main reasons. First, the MDP transitions probabilities are not easily tractable since the AoL variation will simultaneously depend on the channel and resource allocation of both UL and DL links. So, the UL behavior might be analytically unpredictable from the DL perspective and vice-versa. Second, learning a value function from the AoL states means that we have a prediction of system performance given the current AoL condition. In other words, this methodology offers the possibility for the network to essentially learn the control system behavior, where the bandwidth allocation policy is just one of multiple network functions in which it can serve. We could easily extend the learned values to find optimal polices, for example, in terms of packet length, power allocation, antenna configuration and so on.

Hence, we solved the proposed MDP by applying a TD RL algorithm, based on a ϵ-greedy decision making during training, with exponential learning and exploration rate decay [16], as represented in Algorithm 3.

Algorithm 3 Algorithmic description for the RL methodology.

Initialize: Hyperparameters: $\alpha, \gamma, \epsilon, N_{episodes}$
Set state-action value function $Q(states, actions)$ to initial values.
for $\Delta T_{\text{in}} \in \{1, 5, 10, 15, 20 \text{ ms}\}$ **do**
  for $e \in \{1, \ldots, N_{episodes}\}$ **do**
    Initialize the CartPole Environment;
    dl_{CQI} = Random_Integer$(1, 2, \ldots, 15)$;
    ul_{CQI} = Random_Integer$(1, 2, \ldots, 15)$;
    Set $(\Delta T_{\text{in}}, dl_{CQI}, ul_{CQI})$ to the Environment;
    Get the initial state s from the Environment;
    for each control packet transmission **do**
      Select action A = Epsilon-greedy(s, ϵ) for bandwidth allocation;
      Run System Dynamics and Control according to (D.1) and (E.4), respectively, until end of transmission.
      Observe the next state s' and the corresponding reward R according to (F.18);
      Update state-action value as following:
      TD_error $= R + \gamma \cdot \max_a[Q(s', :)] - Q(s, A)$;
      $Q(s, A) \leftarrow Q(s, A) + \alpha \cdot$ TD_error
      $s \leftarrow s'$
    end for
  end for
end for

6 Results

We compare the proposed solution with a bandwidth allocation scheme based on predefined delay requirements, which is the general solution currently used in industry. In more details, given an arbitrary requirement of T_r ms for the control packet to be delivered, we can directly calculate the minimum amount of bandwidth to achieve the necessary requirement using the 3GPP 4-bit CQI Table 7.2.3-1 [12] and the total packet size. These baseline approaches, as well as the RL scheme, were evaluated on the scenario described in Section 5.

We analyze the results for three common network requirements, $T_r = 1$ ms, $T_r = 5$ ms and $T_r = 10$ ms. In each case, we analyzed the total bandwidth usage and the total LQR cost, which are respectively illustrated in Figure C.6 and Figure C.7.

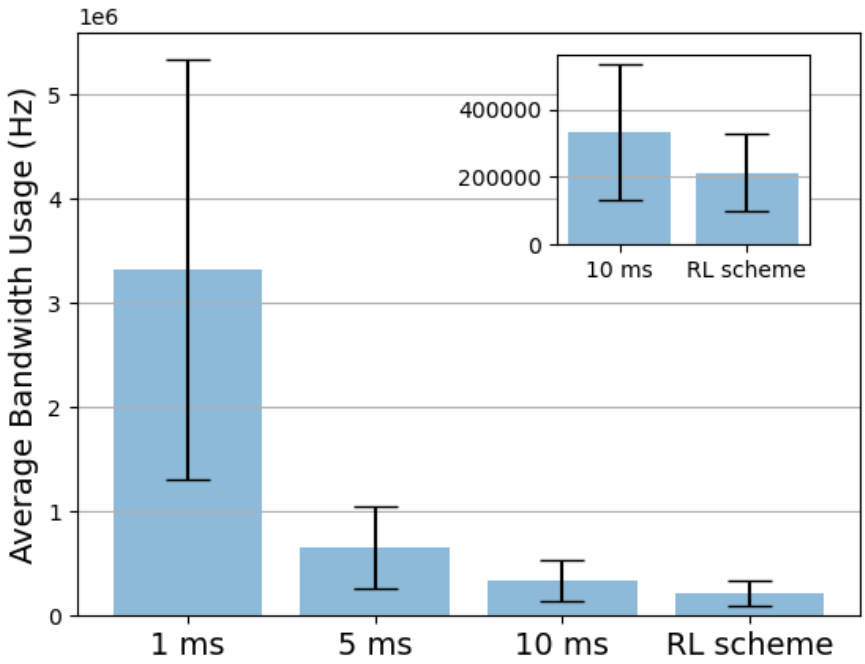


Fig. C.6: Total amount of bandwidth usage for each method.

The immediate conclusion we can verify is that the RL scheme was capable to learn the system delay requirement, such that we can relate the LQR cost in Figure C.7 with

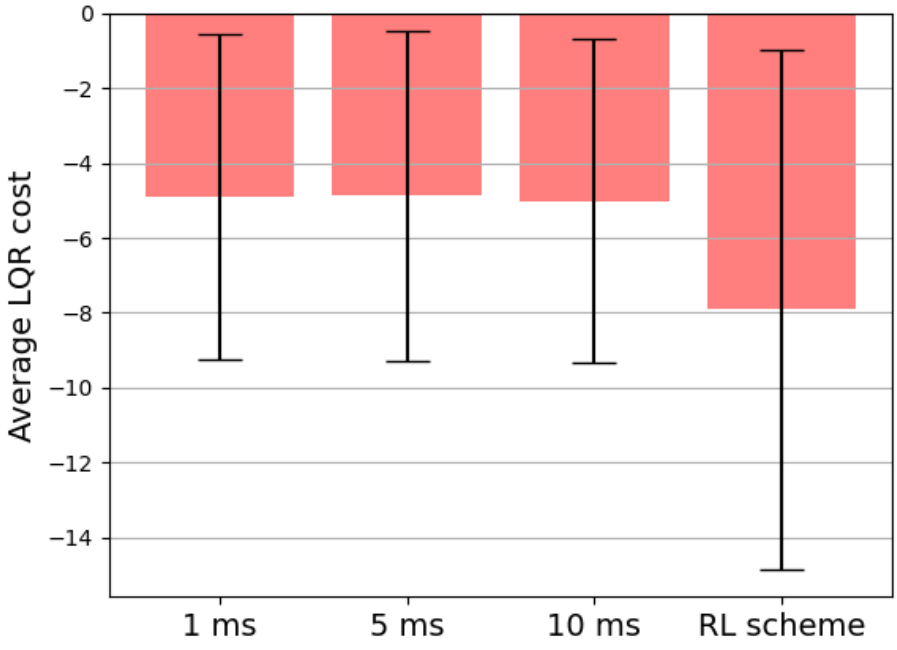


Fig. C.7: Total amount of LQR cost for each method.

the result in Figure C.4 to show that it is operating around the LQR edge performance (around -8) in order to save bandwidth. The second conclusion is that, as expected, strict latency requirement ($T_r = 1$ ms) demands more bandwidth usage. Compared to $T_r = 10$ ms, however, the RL scheme could still save 36% more bandwidth, which is an indication that 10 ms is still a sub-optimal requirement, but we can learn it from the RL algorithm.

7 Conclusions

In this work, we proposed a new metric to evaluate the age of an WNCS closed-loop, and we applied this metric, the Age of Loop, to track the LQR performance of a inverted pendulum control system. Furthermore, we also propose a bandwidth allocation policy based on the age of loop and channel quality information, showing that we can learn the system robustness in order to avoid over-provisioning of network resources on a networked control system.

As future works, we intend to explore a joint DL and UL RL methodology where both cooperate to optimize system performance and network resources.

References

[1] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, 1968.

[2] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Wireless control of autonomous guided vehicle using reinforcement learning," in *IEEE GLOBECOM*, 2020.

[3] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, 2019.

[4] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the latency, rate and reliability tradeoff in wireless networked control systems for IIoT," *IEEE IoT Journal*, 2020.

[5] K. Gatsis, H. Hassani, and G. J. Pappas, "Latency-reliability tradeoffs for state estimation," *IEEE Transactions on Automatic Control*, 2020.

[6] Huang, Kang *et al.* , "Wireless feedback control with variable packet length for industrial IoT," *IEEE Wireless Communications Letters*, 2020.

[7] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, 1983.

[8] R. V. Florian, "Correct equations for the dynamics of the cart-pole system," *Center for Cognitive and Neural Studies (Coneural), Romania*, 2007.

[9] Z. M. Wang, D. F. Yang, K. Yang, L. Y. Guo, and J. M. Tan, *Machine Tool Technology, Mechatronics and Information Engineering*. TransTech Publications Ltd, 2014.

[10] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[11] P. Park, S. C. Ergen, C. Fischione, C. Lu, and K. H. Johansson, "Wireless network design for control systems: A survey," *IEEE Communications Surveys & Tutorials*, 2017.

[12] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 10 2014, version 12.3.0.

[13] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE Journal on Selected Areas in Communications*, 2021.

[14] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using AoI in a single-loop networked control system," in *IEEE Conference on Computer Communications Workshops*, 2019.

[15] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "AoI-penalty minimization for networked control systems with packet loss," in *IEEE Conference on Computer Communications Workshops*, 2019.

[16] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

Paper D

Control-Aware Scheduling Optimization of Industrial IoT

Pedro M. de Sant Ana, Nikolaj Marchenko, Beatriz Soret and Petar Popovski

The paper has been published in the
IEEE 95th Vehicular Technology Conference 2022.

The layout has been revised.

Abstract

In this paper, we elaborate on the frequency resource allocation problem of Wireless Networked Control Systems (WNCS). We consider a multi user wireless environment (e.g., factory) where the users are remote industrial Internet of Things (IIoT) devices competing for network resources and a centralized network base station needs to assign the resources to each device accordingly in order to keep the overall control system stability. We design a joint network and control scheduler solution, where we can estimate the degradation of the control system for a given network state and use this information to assign the frequency resource to each device. We show that the proposed solution outperforms traditional scheduling baselines, including genetic algorithms, assuming polynomial complexity in worst case scenario and generalizing for different control and network configurations.

1 Introduction

Wireless networked control systems (WNCS) are an essential part of cyber-physical applications, such as factory automation, logistics or transportation, enabling mobile control operation in scenarios where high flexibility is required. Nevertheless, due to the nature of the wireless medium, reliability of WNCS remains an open research challenge, especially considering that many WNCS use cases involve large scale setting in scenarios where the network resources are limited. As consequence, multiple control system devices may be required to simultaneously share the wireless medium, where we can often encounter scenarios where the amount of devices sharing the communication resources is higher than the channel capacity.

For this reason, recent works [1–4] have been exploring the issue of allocating communication resources in WNCS, traditionally referred as scheduling problem. We can basically visualize this problem into two approaches. The first is the most common in industry and consists of defining a set of network requirements, for example based on the standards of the 3rd Generation Partnership Project (3GPP), that are established according to the specifications of the control system. Such conventional approach completely decouples the control and communication entities, such that, as demonstrated in [5], might lead to high over-provisioning of network resources. The second approach consists of a joint network and control design, where the scheduling solutions take into consideration details of the control application. For example, the authors in [1] propose a new method to find periodic scheduling policies under the assumptions of global system stability. In [3], authors propose a joint network and control optimization, targeting to design a network resource allocation policy that optimizes the control performance using the least amount of network resources. Likewise, the authors in [4] elaborate on a similar idea of [3] but also considering fading channels. All the results obtained in [1–4]

clearly provide an indication that, considering details of the control application while designing the scheduler, can considerably improve both the network and the control efficacy.

One fundamental assumption, however, that we find in the literature of network/control scheduler design [1–4], is the resource allocation without considering the time and frequency dynamics of a real network scheduler. Essentially, from a practical perspective, the 3GPP standardizes and defines the MAC scheduling scheme based on the definition of a frame structure over time transmission intervals (TTI) [6]. Not considering such dynamics can be relatively critical, since the frequency allocation is constrained by the maximum amount of frequency resources we can allocate per TTI, which represents a direct impact at the strict timing requirements of WNCS. Therefore, aiming to elaborate on such constraint, we propose, in this work, the following contributions:

- We introduce a novel control-aware scheduler mechanism, where we estimate the control performance of each IIoT device based on its current network conditions. To achieve that, we formulate the problem as a Markov Decision Process (MDP), such that we use the obtained value function to provide a network resource allocation policy according to the the 3GPP MAC scheduling scheme in [6].
- We show that the proposed solution generalizes for any new IIoT device entering the network, as long as the control information does not change, thus considerably reducing the overall system complexity to obtain an optimal scheduler, especially in highly dense scenarios.

The rest of this paper is organized as following: in the next two sections, we introduce the details about the control systems and WNCS model. In section 3 we formulate the problem and propose a solution in section 4. Finally, in section 5, we provide the details of the numerical analysis and the results.

2 System Model

We consider the classical inverted pendulum system model, a widely used benchmark problem in control theory domain. As illustrated in Figure D.1, a pole is attached by a joint to a cart, which can be moving along a frictionless track. The pendulum starts upright at a random initial angle $\theta_0 \in (\theta_{0,min}, \theta_{0,max})$, and the goal is to prevent it from falling over by applying a force to the cart. While conceptually simple, the system dynamics are highly unstable and continuously requires fast control cycles to keep stability. When, in turn, being controlled over a wireless channel, the problem becomes an illustrative model of strict timing requirement, capturing the important tradeoffs.

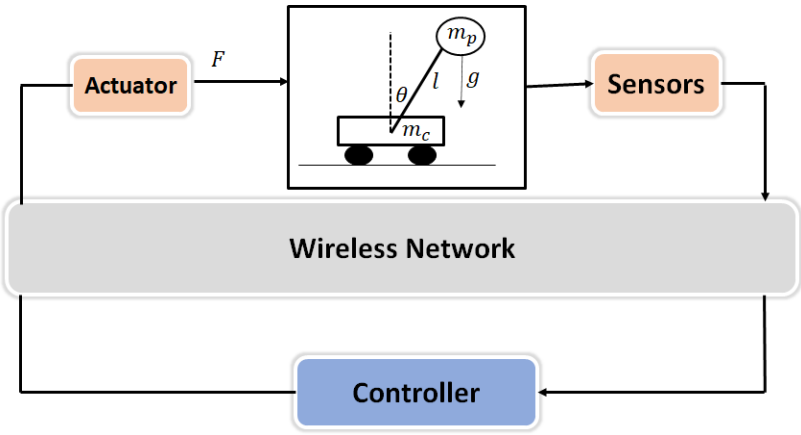


Fig. D.1: Inverted pendulum system model.

2.1 Control System Model

The system dynamics can be represented by the differential equations [7]:

$$\begin{aligned} \ddot{\theta} &= \frac{g \cdot \sin(\theta) + \cos(\theta)\left(\frac{-F - m_p l \dot{\theta}^2 \sin(\theta)}{m_c + m_p}\right)}{l\left(\frac{4}{3} - \frac{m_p \cos^2(\theta)}{m_c + m_p}\right)}, \\ \ddot{x} &= \frac{F + m_p l(\dot{\theta}^2 \sin(\theta) - \ddot{\theta}\cos(\theta))}{m_c + m_p}, \end{aligned} \tag{D.1}$$

where x and θ are, respectively, the cart position coordinates and the pole angle from vertical reference. The mass of the cart is m_c, and the mass of the pendulum is m_p, while l is the length of the pendulum, and F is the force applied to the cart under gravity g. We use the Newton's notation ($\dot{\square}, \ddot{\square}$) to represent derivatives w.r.t time.

By defining a state space vector $X = [x, \dot{x}, \theta, \dot{\theta}]$, we can design a standard optimal controller in two steps. First, computing the Jacobian of (D.1) around the operating point $X = [0, 0, 0, 0]$ to linearize the plant, so that the system dynamic takes the linear time invariant form:

$$\begin{cases} \dot{X} = AX + Bu + w, \\ u = -KX, \end{cases} \tag{D.2}$$

where $u \in \mathbb{R}$ is the linear state feedback control policy of gain K, w is a process disturbance modeled as a zero-mean and covariance $W \in \mathbb{R}^4$, A and B are the system

transition matrix, respectively calculated as [8]:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-12m_p g}{13m_c+m_p} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{12(m_p g+m_c g)}{l(13m_c+m_p)} & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ \frac{13}{13m_c+m_p} \\ 0 \\ \frac{-12}{l(13m_c+m_p)} \end{bmatrix}. \tag{D.3}$$

The second step consists of finding the optimal control policy, u^*, subject to (D.2) that minimizes the cost function,

$$J(X,u) = \int_0^\infty \left(X^T Q X + u^T R u\right) dt, \tag{D.4}$$

where R and Q are arbitrary positive defined matrices in which we can assign weights to state space variables and control signal. In control theory this kind of problem formulation is known as Linear-Quadratic-Regulator (LQR) [8].

The optimal control policy can then be defined by solving the Algebraic Riccati Equation as [8]:

$$\begin{aligned} &A^T P + PA - PBR^{-1}B^T P + Q = 0, \\ &K^* = R^{-1}B^T P, \\ &u^* = K^* X. \end{aligned} \tag{D.5}$$

For (A, B) controllable, the infinite horizon LQR with $Q, R > 0$ gives a convergent closed-loop system [8], where the stability can be easily guaranteed.

2.2 Wireless Networked Control Model

As defined in [9], we adopt a similar WNCS model to define the system behavior over the wireless medium operating in Frequency Division Duplexing (FDD) mode with separated frequency bands for the uplink (UL) and downlink (DL) directions, which makes the medium access for UL and DL independent from each other in the time domain. Figure D.2 illustrates the proposed model, showing the details of the interaction between the communication and application control loop.

First, the sensor readings of the application describe the system states, X_i, with D bytes of sensor data stored in the transmission buffer and communicated to the controller over the uplink channel.

The readings and transmissions of sensor values are done strictly periodically with the cycle time ΔT_{in}, as it is commonly done across various control systems [10].

At the controller, the received sensor values are also stored into the memory. The control application gets the recent values, and produces a control signal u_i according to (E.4). Immediately after producing a control command, the controller sends it over a downlink channel to the controlled system. After finishing the current transmission,

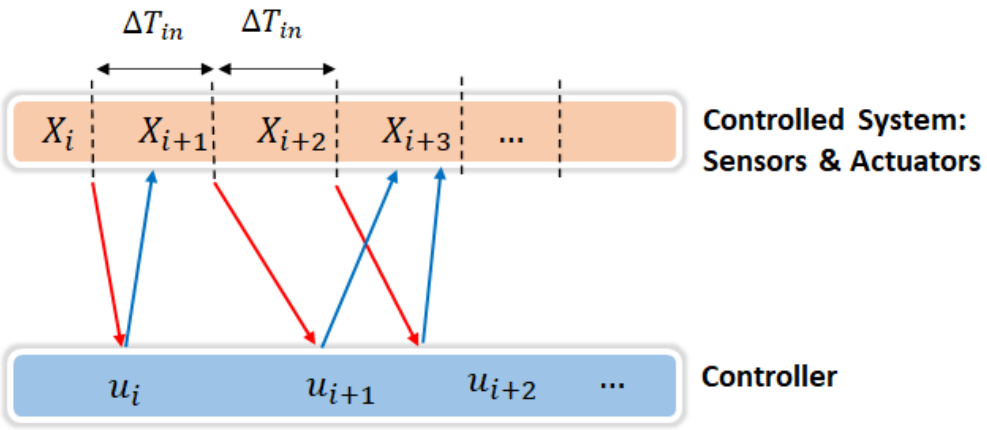


Fig. D.2: WNCS Model.

the controller keeps waiting for the next state update from the controlled device, and starts the procedure once again.

At the controlled system side, we assume the well known zero-order hold (ZOH) strategy, where the received command u_i is stored in the memory. The output application for actuators control (e.g., motor drives) calls the most recently stored command values from the memory and applies them to the application drives, producing the system dynamics of (D.1).

2.3 Wireless channel model

In most WNCS applications, it has been observed that control data has relatively much smaller size compared to the sensor data, thus bringing a strong imbalance between downlink and uplink data traffic [11]. For example, while some WNCS might be required to send image or video data, LiDAR cloud points, AR/VR information and so on, control packets are always based on sending simple control direction, such as acceleration, force or vehicle heading angle that can be either based on an optimal control algorithm as in (E.4) or an human control (e.g., haptic or AR/VR use cases). It is, therefore, plausible to assume that radio resource management (RRM) is not the primarily problem for such downlink communication.

For this reason, in this work, we plausibly design a wireless communication model where we can manage network resources at the uplink, while we do not have control over the downlink delay itself, but we might have some statistical inference about its behavior and use that information to potentially enhance the uplink resource allocation. Hence, the model details are defined as following:

Uplink

We evaluate the uplink behavior considering the 3GPP 4-bit CQI Table 7.2.3-1 [6], where, for each TTI, the amount of data delivered depends on both the current channel quality indicator (CQI) and the total bandwidth allocated to the transmission. So, the more Resource Blocks (RBs) the scheduler assign to the plant in a given TTI, faster the sensor information can be delivered. This approach has been traditionally used for benchmark purposes [12], including the Nokia Open source project in [13].

Downlink

As widely adopted in the literature [14], random and variable delays of communication systems, such as Wi-Fi and LTE, present an uni-modal and asymmetric distribution with long tail to the right in which can be characterized by a Gamma probability density function f, as

$$f(x;k,\theta) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}, \tag{D.6}$$

where $\Gamma(\cdot)$, k and θ are, respectively, the gamma function, the shape and the scale parameter. So, each time a new control command is generated, we randomly select x from $f(x;k,\theta)$ to characterize the downlink latency duration.

2.4 System Level Considerations

It is important to emphasize that (F.7) is guaranteed to be bounded according to the Riccati-equation [8]. However, the combination of two main factors might affect the system LQR performance. The first is the uplink effect, which represents the level of knowledge the controller has about the plant, meaning that, if the uplink takes overly long to deliver sensor data, the controller will compute the control signal using old state feedback, causing the control command to be ineffective or even harmful for the plant. The second is the downlink effect, or simply the overall delay to deliver the control signal. This is important because if the plant applies outdated control commands for too long, the stability of the controlled system might also be compromised.

Each of these factors might affect the plant in different ways and cannot be independently decoupled, which means that a delay in the UL will impact the DL transmission, provoking cumulative effects at the plant and at the network resources.

3 Problem Description

Consider the system model described in section 2, where a set of $K \in \mathbb{N}^+$ inverted pendulum plants are constantly updating its system states to be delivered to a remote

centralized controller. Each plant might have specific control parameters, channel quality or sensor data size. As described by Fig. D.3, for N_f the number of uplink frequency resources, e.g. Resource Blocks (RBs), the network, for each time transmission interval (TTI), must allocate the frequency resources to each plant. This recreates a well-known case of OFDM resource allocation, where a MAC scheduler assigns RBs to control system devices under different radio conditions.

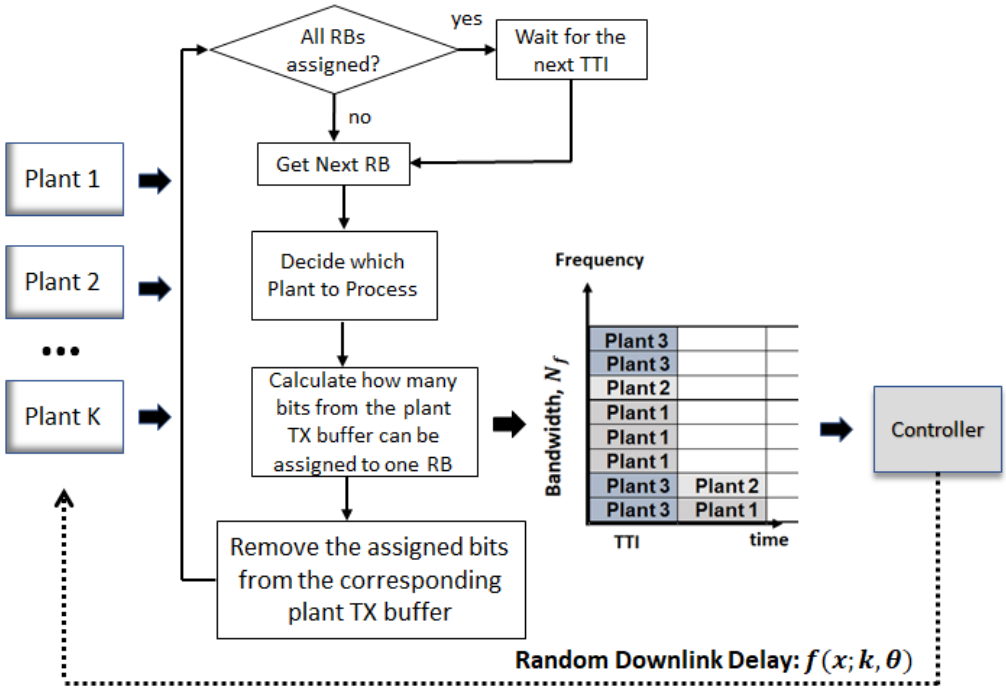


Fig. D.3: The main steps of the resource allocation process.

We can formalize the uplink resource allocation problem by first defining, for each plant $k \in K$, $S_k(t) = [D_k(t), C_k(t), A_k(t)]$ a state space at TTI t, where D_k is the current amount of bits at the uplink transmission buffer, C_k is the corresponding channel quality indicator and A_k the current uplink age of information (AoI). Hence, the scheduler, for every TTI t, must define a frequency allocation policy $\pi : [S_1(t), S_2(t), \ldots, S_K(t)] \rightarrow \mathcal{R}$, that maps the state space of all K plants to a vector $\mathcal{R} \in \mathbb{N}^K$, representing the amount of resource blocks assigned to each user $k \in K$, such that $\sum_{k=1}^{K} r_k = N_f$, $\forall r_k \in \mathcal{R}$.

To define the objective function, we must first emphasize that, in control problems, we are mostly interested in defining a feasibility region in which each plant must correctly operate. It basically means that, to guarantee safety, the physical state of every plant

must not achieve predefined regions. More formally, consider $\tilde{X} = [x, \dot{x}, \theta, \dot{\theta}]$ a plant physical state space where $|x| \geq \mathcal{K}_1$ and $|\theta| \geq \mathcal{K}_2$ for any $\mathcal{K}_1, \mathcal{K}_2 \in \mathbb{R}^+$. So, for $X_k(t)$ the current physical state of plant k at TTI t, we define:

$$\mathbb{1}_{\tilde{X}}(k,t) = \begin{cases} 1 & X_k(t) \in \tilde{X}, \\ 0 & X_k(t) \notin \tilde{X}, \end{cases} \tag{D.7}$$

a function indicating whether the given plant k has achieved the undesirable physical state $\tilde{X}$ at TTI t. So, the overall scheduler goal is to find the optimal policy π^*, such that:

$$\begin{aligned} &\pi^* = \arg\min_{\pi} \sum_{t=0}^{\infty} \sum_{k=1}^{K} \mathbb{1}_{\tilde{X}}(k,t) \\ &\text{s.t. } (D.2), (E.4). \end{aligned} \tag{D.8}$$

We can observe that (D.8) represents an example of application oriented communication, where the main goal is to keep the values of the system states of all plants within the desired feasibility region by controlling the network allocation policy.

4 Solution Proposal

Consider the system dynamics (D.2) defined for any plant $k \in K$, where we can introduce $P_k(t)$ the state covariance matrix at the current TTI t, such that, as in [15], we can describe it as a function of the current AoI value $A_k(t)$, as:

$$P_k(t) \triangleq \mathbb{E}[X_k^T(t) X_k(t)] = H_k(A_k(t)) \triangleq \sum_{i=0}^{A_k(t)-1} A_k^t W_k (A_k^i)^T,$$

for $(\cdot)^T$ the matrix transpose operation.

For a given control system k, we can measure its performance with the quadratic cost of the physical states [], as:

$$\mathcal{J}_k = \lim_{T\to\infty} \sum_{t=1}^{T} \mathbb{E}[X_k^T(t) Q_k X_k(t)] = \lim_{T\to\infty} \sum_{t=1}^{T} \mathrm{Tr}(Q_k P_k(t)), \tag{D.9}$$

for $\mathrm{Tr}(\cdot)$ the matrix trace operation. We can observe that (D.9) represents a degradation metric for the control performance that we can map as a function of the current AoI value and the system parameters. Minimizing (D.9) guarantees that system states $X_k(t)$ are bounded when $t \to \infty$ [8], such that we can potentially apply (D.9) to observe how far or close the control system is from the feasibility region defined in (D.7).

Hence, the main idea is that, by observing the state S_k, we can evaluate, for each plant $k \in K$, the expected control performance $\mathcal{J}_k$ according to (D.9) and use this

information to prioritize the plant resource allocation accordingly. More formally, we can model such estimation of performance for a given state space by applying the value function definition [8, 16], where, for any TTI t, we define:

$$\mathcal{V}_k(S_k) = \mathbb{E}_{\tilde{\pi}}\left[\sum_{\tau=0}^{\infty}\gamma^{\tau}\mathrm{Tr}(Q_k P_k(t+\tau+1))\big|S_k(t)=S_k\right], \tag{D.10}$$

where we can also extend to a state-action value function [16]:

$$\mathcal{Q}_k(S_k, r) = \mathbb{E}_{\tilde{\pi}}\left[\sum_{\tau=0}^{\infty}\gamma^{\tau}\mathrm{Tr}(Q_k P_k(t+\tau+1))\big|S_k(t)=S_k, r_k(t)=r\right], \tag{D.11}$$

$\forall r \in [0; N_f]$ where $r_k(t)$ is the number of resources blocks allocated for user k at TTI t, such that $\sum_{k=1}^{K} r_k(t) = N_f$, $\forall t$, $\gamma \in [0, 1]$ is called discount factor and determines the long-term impact of future costs, and $\tilde{\pi}$ is the probability function $\Pr\{r_k(t) = r \,|\, S_k(t) = S_k\}$ of selecting r for a given state S_k.

To obtain the state value function in (D.10), as well as the state-action value function in (D.11) for a plant k, we formulate the problem as a Markov Decision Process (MDP) [8], defined as:

- The state space is describe by $S_k = [D_k, C_k, A_k]$, where, for every TTI t, the state transition is given by:

$$\begin{bmatrix} D_k(t+1) \\ C_k(t+1) \\ A_k(t+1) \end{bmatrix} = \begin{bmatrix} \min\{0, D_k(t) - g(C_k(t), r_k(t))\} \\ h(C_k(t)) \\ A_k(t)+1 \end{bmatrix}, \tag{D.12}$$

  where g is a function that describe the amount of bits transmitted for a given CQI $C_k(t)$ and the number of resource blocks allocated $r_k(t)$ according to the 3GPP table [6], while h is a function to map the CQI variation over t. The terminal state is given $\forall S_k$ where $D_k = 0$.

- The action space is simply the amount of resource blocks allocated to user $k \in K$ at TTI t given by $r_k(t)$.

- The one stage cost is designed for two conditions: A) For each non-terminal state transition (i.e., $D_k(t+1) > 0$), we calculate the control cost in (D.9) incurred from the current AoI value $A_k(t)$:

$$\mathcal{C}_k(t) = \mathrm{Tr}(Q_k H_k(A_k(t))). \tag{D.13}$$

  For a terminal state transition (i.e., $D(t+1) = 0$), we account to the control cost in (D.9) plus the downlink delay effect with its corresponding probability density function:

$$\mathcal{C}_k(t) = \sum_{l=1}^{\infty} f(l; k, \theta)\mathrm{Tr}(Q_k H_k(A_k(t)+l)) \tag{D.14}$$

Algorithm 4 Evaluation of the Bellman Equation.

Initialize: $\gamma, D_{\max}, C_{\max}, N_f, A_{\max}, N_{\text{iterations}}, \tilde{\pi}$

$\mathcal{V}$ = array of zeros with size $(D_{\max}, C_{\max}, A_{\max})$
$\mathcal{Q}$ = array of zeros with size $(D_{\max}, C_{\max}, A_{\max}, N_f)$
for $n \in [1, N_{iterations}]$ **do**
  for each $d \in [1, D_{\max}], c \in [1, C_{\max}], a \in [1, A_{\max}]$ **do**
    for each $r \in [0, N_f]$ **do**
      $d' = \min\{0, d - g(c, r)\}$; $a' = a + 1$; $c' = h(c)$
      if $d' == 0$ **then**
        Calculate $\mathcal{C}$ as in (D.14).
        $\mathcal{Q}[d, c, a, r] \leftarrow \mathcal{C}$ // Terminal State
      else
        Calculate $\mathcal{C}$ as in (D.13).
        $\mathcal{Q}[d, c, a, r] \leftarrow \mathcal{C} + \gamma \mathcal{V}[d', c', a']$
      end if
    end for
    $\mathcal{V}[d, c, a] \leftarrow \mathbb{E}_{\tilde{\pi}}\left[\mathcal{Q}[d, c, a, :]\right]$
  end for
end for

We can note that the proposed MDP is fully observable and we can analytically track, by the one-stage cost, the random effects from the downlink delay f and the control process disturbance w for each plant. Hence, we can optimally obtain (D.11) and (D.10) using the Bellman equation [16], in which the steps can be described according to Algorithm 4.By obtaining (D.11), we can now establish the resource allocation as a function of the expected control performance for any given state. We design, therefore, the resource block assignment as described in Algorithm 5, to generate the policy π for solving (D.8).

The primarily idea of Algorithm 5 is to prioritize the plants with the highest current $\mathcal{Q}$ value, considering the corresponding amount of RBs already assigned. It is important to emphasize that, as long as the control dynamics of the IoT device does not change, which is broadly the case, the network only needs to evaluate the Bellman equation once, requiring polynomial complexity $O(n^3)$ [16] in the worst case scenario for n the number of states. Similarly, adding new IoT devices to the network will only demand the the step from Algorithm 4 if the control dynamics are not previously known by the network. So, in case of multiple similar IoT devices entering the network, which involves, for example, typical use cases regarding factory automation, the step of Algorithm 4 is performed only once and generalize for all other devices, thus considerably reducing the overall system complexity to obtain the optimal scheduler, especially in highly dense scenarios.

Algorithm 5 Proposed RB assignment policy.

Initialize: K, N_f and Q_k for each $k \in K$
1: $\mathcal{R}$ = array of zeros with size K
2: **for** each TTI t **do**
3: Observe the states $S_k(t) = [D_k(t), C_k(t), A_k(t)], \forall k$
4: **repeat**
5: $i = \arg\min_{k \in K} \mathcal{Q}_k(D_k(t), C_k(t), A_k(t), \mathcal{R}[k])$
6: $\mathcal{R}[i] \leftarrow \mathcal{R}[i] + 1$
7: **until** All RBs assigned.
8: **end for**
Output: $\mathcal{R}$

5 Numerical Analysis

To evaluate the proposed solution, we consider a simulation scenario as summarized in Table D.1, where we can further emphasize the following points:

- At the beginning of every simulation episode, each plant uniformly random selects a sensor data size, CQI and pole length among the provided options. The higher the data size and the lower the CQI, more resource blocks are needed to deliver the sensor data to the controller.
- A pole length of 0.25 and 0.75 provide system dynamics with maximum absolute eigenvalues of, respectively, 5.61 and 3.24. This means that the control requirements are also different, such that the plants with the highest eigenvalues are more unstable, i.e. the cost in (D.9) grows faster, and thus demand faster control cycles.
- For evaluating the Bellman equation in Algorithm 4, we assume $\tilde{\pi}$ with equal probability for every action, i.e. $\Pr\{r_k(t) = r \,|S_k(t) = S_k\} = 1/N_f$, $\forall r, \forall S_k$. For the evaluation of the cost in (D.14), we bound l up to three standard deviations.

5.1 Baselines

We compare the results with two sets of baselines. The first set encompass the most traditional scheduling algorithms, where the allocation of RBs is intuitively clear, such as Round Robin (RR), Maximum CQI (MaxCQI), Minimum CQI (MinCQI) and Age-based scheduler (Age). The RR equally distribute the RBs to each plant following a rational order. The MaxCQI prioritize the users with highest CQI and represents a direct maximization of the network throughput and spectral efficiency. The MinCQI, on the other hand, prioritize the users with the lowest CQI, thus maximizing the fairness

Table D.1: System Parameters

Bandwidth, B	10 MHz
Number of Physical Resource Blocks, N_f	50
Sensor Data sizes, D	[512, 768, 1024] bits
CQI	[1,2,3,4,5]
Carrier frequency	2.4 GHz
DL latency pdf shape k and scale θ	[2.0; 3.5]
m_c, m_p, g	1.0 kg, 0.1 kg, 9.8 m/s^2
Pole Length, l	[0.25, 0.75] m
Control matrices, R and Q	(0.1) and I_{4x4}
Stability threshold $\mathcal{K}_1$ and $\mathcal{K}_2$ from (D.7)	2 m and 20^o
Sensor cycle time, ΔT_{in}	5 ms

regarding the resource distribution among the users. Finally, the Age scheduler prioritize resources for the users with the current highest age of information.

The second set consists of state-of-the-art genetic algorithms that implement a parameter search at the state space and return weights that are optimally assigned to the elements of the state space based on a black-box optimization. This methodology was recently applied at a similar problem in [12], where the proposed solution using Co-variance matrix adaptation evolution strategy (CMA-ES) outperformed RL approaches. We compare our proposed solution against CMA-ES and another evolutionary strategy based on finite differences, named Policy Gradients with Parameter Exploration (PGPE) [17]. The implementation of both strategies were strictly attained based on the Pybrain public library [17].

5.2 Evaluation

From a control design perspective, we are mostly interested in analyzing the stability level of the overall system over time. More formally, consider a stability target given by:

$$\Pr\{\sum_{t=0}^{\infty}\sum_{k=1}^{K} \mathbb{1}_{\bar{X}}(k,t) \geq 1\} < 0.01, \tag{D.15}$$

which essentially describes that all control system must not violate the stability condition and, in an infinity-horizon perspective, the probability of this violation should not surpass 1%. This measure can afterwards be used as a basis to evaluate metrics commonly used in industry, such as mean time to failure (MTTF), Mean time between failures (MTBF), etc.

5.3 Results

The Fig. D.4 shows the maximum amount of plants the network can tolerate while respecting the stability target defined in (D.15). The Age scheduler performing surprisingly good is an interesting evidence of how much the timing requirements can affect real-time control applications. The RR as the worst performer only evidence that the intuitive idea of equally dividing the resources can provide harmful results, while the proposed scheduler solution achieved the highest amount of users for the same 10 MHz bandwidth. In Fig. D.5, we check the 99th percentile of the LQR performance according to (F.7), for a fixed number of $K = 30$ plants. The proposed solution, as we expect, achieved a resource allocation policy that prioritize the plants with the highest expected LQR, thus reducing the probability of stability violation in (D.15).

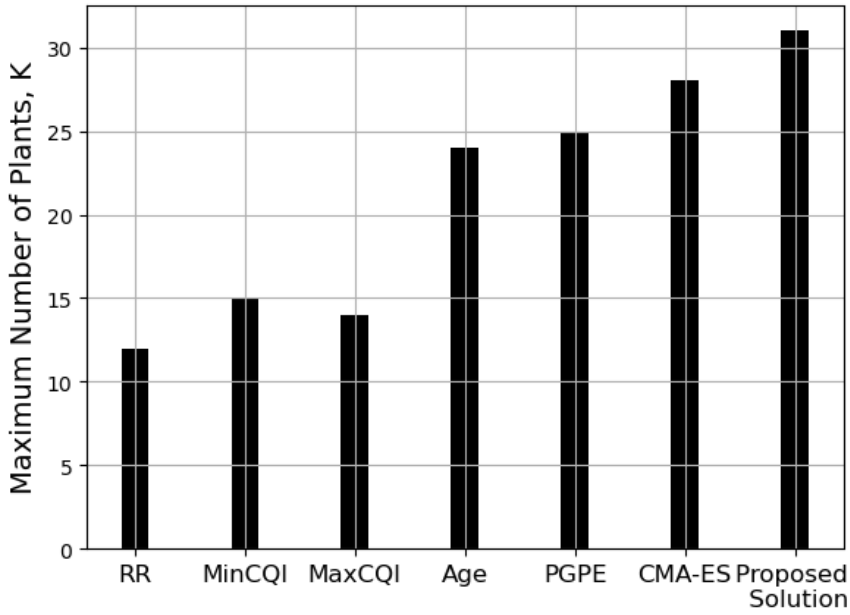


Fig. D.4: Number of plants supported for the stability target in (D.15).

6 Conclusions

In this work, we elaborated on an uplink resource allocation problem, where a set of control systems with distinct control and network requirements must share frequency resources to transmit their sensor data to a centralized edge controller. The goal is

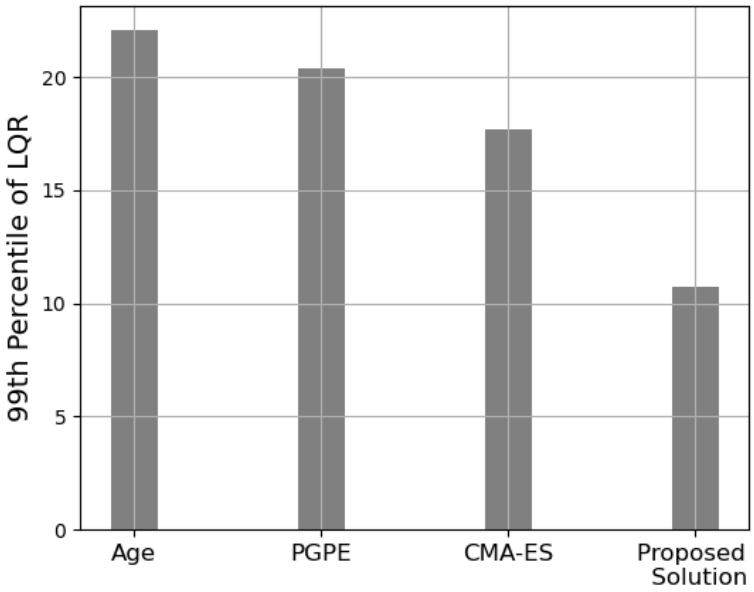


Fig. D.5: 99th percentile of the LQR performance according to (F.7).

to design a resource allocation policy that can achieve the long-term stability of the overall system. We have proposed a resource allocation scheme based on the estimation of the value function of each control device. The proposed approach generalizes for multiple control dynamics and can be solved in polynomial time, while also minimizing the overall control cost and increasing the system capacity compared to traditional scheduler baselines and evolutionary approaches.

References

[1] A. Kundu and D. E. Quevedo, "Stabilizing scheduling policies for networked control systems," *IEEE Transactions on Control of Network Systems*, 2019.

[2] G. Karadag and Iqbal, "Optimal power control, scheduling, and energy harvesting for wireless networked control systems," *IEEE Transactions on Communications*, 2020.

[3] K. Huang, W. Liu, Y. Li, B. Vucetic, and A. Savkin, "Optimal downlink–uplink scheduling of wireless networked control for industrial iot," *IEEE Internet of Things Journal*, 2019.

[4] Y. Sadi and S. C. Ergen, "Energy and delay constrained maximum adaptive schedule for wireless networked control systems," *IEEE Transactions on Wireless Communications*, 2015.

[5] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Age of loop for wireless networked control systems optimization," in *IEEE PIMRC*, 2021.

[6] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 10 2014, version 12.3.0.

[7] R. V. Florian, "Correct equations for the dynamics of the cart-pole system," *Center for Cognitive and Neural Studies (Coneural), Romania*, 2007.

[8] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[9] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Wireless control of autonomous guided vehicle using reinforcement learning," in *IEEE GLOBECOM*, 2020.

[10] P. Park, S. C. Ergen, C. Fischione, C. Lu, and K. H. Johansson, "Wireless network design for control systems: A survey," *IEEE Communications Surveys & Tutorials*, 2017.

[11] X.-M. Zhang, Q.-L. Han, X. Ge, D. Ding, L. Ding, D. Yue, and C. Peng, "Networked control systems: A survey of trends and techniques," *IEEE/CAA Journal of Automatica Sinica*, 2019.

[12] P. M. de Sant Ana and N. Marchenko, "Radio Access Scheduling using CMA-ES for Optimized QoS in Wireless Networks," in *IEEE Globecom Workshops*, 2020.

[13] A. Valcarce, "Wireless Suite: A collection of problems in wireless telecommunications," 2020. [Online]. Available: https://github.com/nokia/wireless-suite

[14] S. Yasuda and H. Yoshida, "Prediction of round trip delay for wireless networks by a two-state model," in *IEEE WCNC*, 2018.

[15] Huang, Kang *et al.* , "Wireless feedback control with variable packet length for industrial IoT," *IEEE Wireless Communications Letters*, 2020.

[16] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[17] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber, "PyBrain," *Journal of Machine Learning Research*, 2010.

Paper E

Goal-Oriented Wireless Communication for a Remotely Controlled Autonomous Guided Vehicle

Pedro M. de Sant Ana, Nikolaj Marchenko, Beatriz Soret and Petar Popovski

The paper has been published in the
IEEE Wireless Communications Letters 2023.

The layout has been revised.

Abstract

This letter considers a remote control Autonomous Guided Vehicle (AGV), where packets carrying sensory data and control information are sent through a time-correlated wireless fading channel. We illustrate there is an inherent dependency between the data rate and the resulting control accuracy for such a system. We propose a goal-oriented wireless solution, in which data transmission rate is continuously adapted according to Age-of-Loop metric to achieve precise remote trajectory control of an AGV. The numerical analysis shows that the proposed solution provides higher control accuracy at the AGV trajectory compared to straightforward fixed-rate transmission policies, and to a reference solution based on Age-of-Information metric, more commonly used in research of wireless networked control systems.

1 Introduction

Precise remote trajectory tracking control [1] of Autonomous Guided Vehicles (AGV) over time-varying wireless channel dynamic factory environments is a challenging problem. Age-of-Information (AoI) was introduced as a helpful metric indicating the freshness of transmitted data, which is used for goal-oriented optimization of such Wireless Networked Control Systems (WNCS) [2, 3]. However, with AoI the analysis and optimization are constrained to a single communication link, either downlink (DL) or uplink (UL). In [4], we showed that such optimization can lead to sub-optimal behavior for close-loop control problems, and introduced the Age-of-Loop (AoL) metric to account for both DL and UL effects, as well as their interplay. In addition, most related control solutions are made with simplified assumption of the strong channel stationarity [5, 6]. In contrast, in [7] we showed that correlated fading plays a significant role in remote path tracking and needs to be considered.

Building on our previous works [4] and [7], the contributions of this letter are the following:

- We analytically show how a physical AGV process can vary as a function of the AoL, providing an important theoretical insight about the relationship between the control planning and the radio resource allocation. Such analysis handles a joint network-control design within the context of the goal-oriented communication paradigm [8], where the contextual value of the information is taken into account to leverage the synergy between a control system application and the wireless medium.

- We propose a model of a remote AGV control that can dynamically adapt the transmission data rate, aiming to optimize the AGV trajectory. To find the optimal transmission policy, the problem is formulated as a semi-Markov Decision

Process where the channel correlation is evaluated over time to address the fading issue.

- We show that with the proposed approach we can outperform fixed-data rate policies, as well as state-of-the-art solutions that are purely based on AoI, thus achieving the goal of higher system trajectory accuracy, a result that so far has not been obtained in current literature.

2 AGV System Model

2.1 AGV Model

We define the vehicle state vector $\mathbf{X} \in \mathbb{R}^4$ and control vector $\mathbf{U} \in \mathbb{R}^2$ over time t as in [1], where:

$$\mathbf{X}(t) = [x(t), y(t), v(t), \theta(t)], \ \mathbf{U} = [a(t), \delta(t)]. \tag{E.1}$$

Here, x, y are the 2D coordinates, v is the velocity, and θ is the heading orientation angle of the vehicle, respectively. The control is based on manipulating the vehicle acceleration a and its front wheel angle δ in order to manage its trajectory along a desired path. Implementation details of the control system model, design and structure can be found in [7].

The control system dynamics are defined according to the kinematic vehicle model [1], such that $\dot{x}(t) = v(t)\cos\theta(t)$, $\dot{y}(t) = v(t)\sin\theta(t)$, $\dot{v}(t) = a(t)$ and $\dot{\theta}(t) = v(t)\tan\delta(t)/L$, for the vehicle inter-axle distance L. The notation $\dot{\square}$ corresponds to derivatives w.r.t time. The kinematic vehicle model is linearized by dynamically applying the first-order Taylor expansion around the current vehicle position as operational point [1], generating the ordinary differential equation:

$$\dot{\mathbf{X}}(t) = f(\mathbf{X}(t), \mathbf{U}(t)) = \mathbf{A}(t)\mathbf{X}(t) + \mathbf{B}(t)\mathbf{U}(t) \tag{E.2}$$

where the transition matrices $\mathbf{A}(t)$ and $\mathbf{B}(t)$ are given as [1]:

$$\mathbf{A}(t) = \begin{bmatrix} 0 & 0 & \cos\theta(t) & -v(t)\sin\theta(t) \\ 0 & 0 & \sin\theta(t) & v(t)\cos\theta(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\tan\delta(t)}{L} & 0 \end{bmatrix} \mathbf{B}(t) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & \frac{v(t)}{L\cos^2\delta(t)} \end{bmatrix}$$

2.2 AGV Control and Performance

To control the AGV, consider a trajectory planner that arranges a sequence of desired states X^* that the AGV has to follow. To achieve that, the controller must provide the corresponding control commands u according to which the vehicle adjusts its position

and minimizes the error $\mathbf{X_e}(t) = \mathbf{X}(t) - \mathbf{X}^*(t)$ between the current state X and the desired state X^* at each time instant t along the path. More formally, we can define a cost function whose goal is to minimize $X_e(t)$ as

$$J(\mathbf{X_e}(t), \mathbf{U}(t)) = \int_0^\infty \left[\mathbf{X_e}(t)^T \mathbf{Q} \mathbf{X_e}(t) + \mathbf{U}(t)^T \mathbf{R} \mathbf{U}(t)\right] dt, \tag{E.3}$$

subject to the system dynamics in (F.5), where R and Q are arbitrary positive defined matrices in which we can specify weights among the different components of the state space and the control signal. In control theory, such design formulation is known as Linear-Quadratic-Regulator (LQR) and the optimization to find the optimal control command $\mathbf{U}^*(t) = [a^*(t), \delta^*(t)]$ is done in an online fashion each time a new control command is generated, by solving the Algebraic Riccati Equation [1]:

$$\begin{aligned} &\mathbf{A}^T(t)\mathbf{P} + \mathbf{P}\mathbf{A}(t) - \mathbf{P}\mathbf{B}(t)\mathbf{R}^{-1}\mathbf{B}^T(t)\mathbf{P} + \mathbf{Q} = \mathbf{0}, \\ &\mathbf{K}^* = \mathbf{R}^{-1}\mathbf{B}^T(t)\mathbf{P}, \\ &\mathbf{U}^*(t) = \mathbf{K}^*\mathbf{X}(t). \end{aligned} \tag{E.4}$$

For $\mathbf{A}(t)$ and $\mathbf{B}(t)$ controllable, the infinite time horizon LQR with $\mathbf{Q}, \mathbf{R} > \mathbf{0}$ gives a convergent closed-loop system [1], where the stability can be guaranteed.

To evaluate the AGV control performance, we analyze the cross track error (XTE), a commonly used metric that measures the distance deviation from the planned path. We can derive the XTE as in [1], where we first obtain the error state vector $X_e(t)$ for a desired state $\mathbf{X}^*(t) = [x^*(t), y^*(t), v^*(t), \theta^*(t)]$, as

$$\begin{bmatrix} x_e(t) \\ y_e(t) \\ v_e(t) \\ \theta_e(t) \end{bmatrix} = \begin{bmatrix} \cos\theta^*(t) & \sin\theta^*(t) & 0 \\ -\sin\theta^*(t) & \cos\theta^*(t) & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x(t) - x^*(t) \\ y(t) - y^*(t) \\ v(t) - v^*(t) \\ \theta(t) - \theta^*(t) \end{bmatrix},$$

and use the kinematic model to calculate its derivatives:

$$\begin{aligned} \dot{\theta_e}(t) &= \dot{\theta}(t) - \dot{\theta^*}(t) = w(t) - w^*(t), \\ \dot{v_e}(t) &= \dot{v}(t) - \dot{v^*}(t) = a(t) - a^*(t), \\ \dot{x_e}(t) &= y_e(t)w(t) + v^*(t) - v(t)\cos\theta_e(t), \\ \dot{y_e}(t) &= -x_e(t)w^*(t) + v(t)\sin\theta_e(t). \end{aligned} \tag{E.5}$$

The XTE is given by the lateral distance y_e between the selected tracing point and the point of the path that is closest to it. Hence, we are interested at the car position as its projection on the path. More formally:

$$\begin{bmatrix} y_e(t) \\ \theta_e(t) \end{bmatrix} = \begin{bmatrix} -\sin\theta^*(t) & \cos\theta^*(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x(t) - x^*(t) \\ y(t) - y^*(t) \\ \theta(t) - \theta^*(t) \end{bmatrix}, \tag{E.6}$$

and the corresponding error derivatives:

$$\begin{aligned} \dot{y_e}(t) &= v(t)\sin\theta_e(t), \\ \dot{\theta_e}(t) &= w(t) - w^*(t). \end{aligned} \tag{E.7}$$

It follows from (E.6) and (E.7) that the XTE increases over time as a function of the current vehicle speed, the error between the current vehicle orientation, and the corresponding angle at the desired path. By properly designing the optimal LQR control in (F.7), the XTE is minimized along the path [1]. On the other hand, if the communication between the plant and the controller (or vice-versa) fails to deliver the information, we expect to see some impact at the XTE even with the optimal control policy. We elaborate on these effects in next sections.

3 Wireless communication model

We consider the two directions of the communication model: the sensor-control communication, here referred to as the *uplink* direction, and the control-actuator communication, which is the *downlink*. In most control applications, including the AGV, it has been observed that control data has relatively small size as compared to the sensor data [9], thus bringing a strong imbalance between downlink and uplink data traffic. Thus, AGVs are often required to send large image or video data, for example light detection and ranging (LiDAR) cloud points and augmented or virtual reality (AR/VR) information. However, the control packets in the opposite direction are based on sending simple control direction, such as acceleration or vehicle heading angle that can be either based on a control algorithm as in (E.4) or human control as in haptic or AR/VR use cases. In this context, we design a communication model where the uplink network resources, being the bottleneck, are dynamically adapted, but considering not only the uplink itself but also statistical inferences about the downlink behavior. The next two sub-sections provide more details about each model.

3.1 Uplink Model

We model the uplink communication as a first-order Markov process [10], also known as Gilbert-Elliot channel. The time-correlation property is represented by two states: the good state G if the packet can be successfully received; and the bad state B, otherwise. The corresponding transition probability matrix is defined by a stationary Markov distribution:

$$\mathbf{M} = \begin{bmatrix} P_{gg} & 1 - P_{gg} \\ 1 - P_{bb} & P_{bb} \end{bmatrix}. \tag{E.8}$$

where P_{gg} is the probability that the current transmission is successful, given that the previous transmission was successful. Same logic applies for P_{bb} but for unsuccessful

transmissions. Given the matrix M, the channel properties are completely characterized [10]. In particular, the marginal probability of a packet error, ε_p, is then given by:

$$\varepsilon_p = 1 - \frac{1 - P_{bb}}{2 - (P_{bb} + P_{gg})} \tag{E.9}$$

For Rayleigh fading, $\varepsilon_p = 1 - e^{\gamma_{th}}$, where γ_{th} is the minimum threshold SNR required to successfully decode the received signal. We can evaluate γ_{th} as a function of the utilized data rate R, bandwidth B and the average SNR γ, as:

$$\gamma_{\text{th}} = \frac{2^{R/B} - 1}{\gamma}, \tag{E.10}$$

We define the Jakes's channel correlation coefficient to be $\rho = J_0(2\pi f_d T_s)$, where $J_0()$ is the zero-order Bessel function of the first kind, f_d is the Doppler frequency shift and T_s is the sampling time. The error probability of a single back to back failure, P_{bb}, is then written as [10]

$$P_{bb} = 1 - \frac{Q(\theta, \rho\theta) - Q(\rho\theta, \theta)}{e^{\gamma_{th}} - 1}, \tag{E.11}$$

where $Q(.,.)$ is the Marcum Q function and $\theta = \sqrt{\frac{2\gamma_{th}}{1-\rho^2}}$.

For any number of bits D to be transmitted, we have a total data transmission delay calculated as D/R, being a direct function of the selected data rate R. If the transmission fails, we assume a re-transmission will occur immediately. A retransmission automatically increases the resulting latency of the data, or data delivery time. By increasing R, one reduces the single-packet transmission delay but increases the packet error probability according to (F.9)-(F.12). This demonstrates the dependency between the data rate R and the control accuracy (i.e., XTE) in Section 6.

3.2 Downlink Model

Due to the potential imbalance between DL and UL data traffic discussed along this section, we assume that transmission rate adaptation is not meaningful in the DL communication. However, DL packet delays can also happen due to other factors not necessary related to communications, such as the processing time of the control application [9]. For the sake of generality, we consider a general probability distribution $\mathcal{G}(x)$, where $l \sim \mathcal{G}(x)$ is the DL packet delivery time. In the evaluation of Section 6, we take two exemplary distributions to illustrate the impact of the DL latency in the global results: (1) fixed time intervals representing the simplest model; (2) a Gamma distribution, considered in previous literature [11] to represent a heavy-tailed distribution of the DL packets delay.

In summary, we consider the following assumptions in the proposed model: 1) the DL behavior is given by a random latency selected from a probability distribution, while

the UL is modeled as a first-order Markov process under Rayleigh fading. 2) the AGV is given according to the kinematic vehicle model under optimal LQR control. More details are described in Table E.1, in which summarizes the main system parameters used in the numerical evaluation presented in Section 6.

4 AGV behavior under imperfect communication

The sensor generation process is assumed to be periodic at the AGV. However, a new data transmission can start only when the previous is complete. By adopting the wireless channel model of Section 5, the wireless communication becomes non-deterministic and leads to delays and packet losses, which has an impact on the control behavior. Fig. F.4 shows a demonstrative data flow diagram for the AGV control.

The AoI is defined as the time that has elapsed since the latest information update at the destination. It captures not only the communication delays but also the impact of the packet generation [2]. However, AoI considers only one transmission direction. Therefore, for the closed-loop control of the AGV, instead of AoI, we use the Age of Loop, $\Delta_{\mathrm{L}}(t)$, a continuous and bounded timing metric first proposed in [4], to capture the overall aging from the sensor to actuator, (or from actuator to sensor), including the effect of packet losses. As illustrated in Fig. F.4, the AoL grows linearly over time and drops at the time instances where the control loop is closed, $(tc_1, tc_2, \dots)$, to the corresponding timestamp in which the state feedback that spawned a new control signal was generated, i.e. $\Delta_{\mathrm{L}}(t) = t - ts_i, \forall i \in \{1, 2, 3, \dots\}$.

We can now model the dynamics in (F.5) using AoL and the stochastic system approach [12]:

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{U}(t - \Delta_{\mathrm{L}}(t)). \tag{E.12}$$

With communication delay and packet losses (i.e., $\Delta_{\mathrm{L}}(t) > 0$), we are especially interested in the asymptotic behavior of the vehicle tracking error:

Proposition 1: The tracking error (E.7) is asymptotically non-negative w.r.t $\Delta_{\mathrm{L}}(t)$.

Proof: Consider the system dynamics in (F.13), where (F.5) becomes a special condition of $\Delta_{\mathrm{L}}(t) = 0$, $\forall t$. Assuming that the AGV is moving (i.e., $v(t) > 0$), we show in (E.7) that its tracking error y_e will increase over time for any $\sin\theta_e(t) > 0$. For $\mathcal{U} = [a, \delta]$ where $a \in [0, a_{max}]$ and $\delta \in [-\delta_{max}, +\delta_{max}]$, $\forall \delta_{max}, a_{max} \in \mathbb{R}$ the set of all possible control actions in which the AGV can take, applying the optimal control policy in (E.4), will provide:

$$\mathbf{U}^*(t) = \underset{\mathbf{U}(t)\in\mathcal{U}}{\operatorname{argmin}} \int_0^\infty \left(\mathbf{X_e}(t)^T \mathbf{Q} \mathbf{X_e}(t) + \mathbf{U}(t)^T \mathbf{R} \mathbf{U}(t)\right) dt,$$

such that, for $(\mathbf{A}, \mathbf{B})$ controllable, there exists a unique solution, $\mathbf{P}$, in the class of positive semi-definite matrices, to the algebraic Riccati equation [1]. Hence, there will

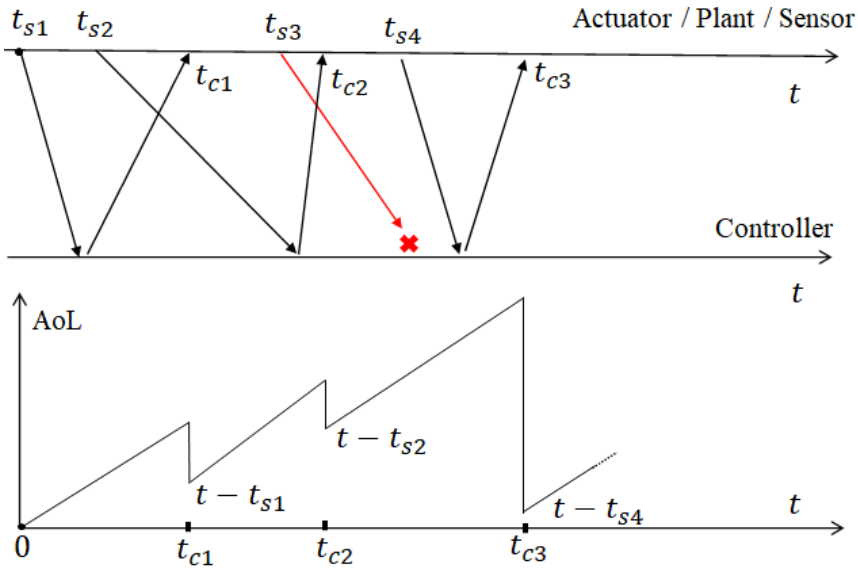


Fig. E.1: Timing diagram of signals transmitted with corresponding AoL.

be a unique solution $\delta^*(t) \in \mathbf{U}^*(t)$ that applied to $\dot{\theta}(t) = \frac{v(t)}{L} \tan \delta^*(t)$ will minimize the difference $\theta_e(t) = \theta(t) - \theta^*(t)$ over time. As a consequence, the only possible way for any $\Delta_{\mathrm{L}}(t) > 0$ to not affect $y_e(t)$ is i.f.f. $\delta^*(t) = \delta^*(t - \Delta_{\mathrm{L}}(t))$, which in turn requires that the desired path does not change within the given $\Delta_{\mathrm{L}}(t)$ interval, i.e., $\theta^*(t) = \theta^*(t - \Delta_{\mathrm{L}}(t))$.

Remark 1: The AGV tracking error will not necessarily increase when $\Delta_{\mathrm{L}}(t) > 0$, which directly depends on how much the planned path varies compared to the current path within $\Delta_{\mathrm{L}}(t)$ interval. One could potentially exploit such behavior while designing the AGV wireless controller, where regions of the planned path may be considered for specifying the network resource allocation.

5 AGV Control with variable data rate

5.1 Problem Formulation

Let $\mathcal{R} = \{r_1, r_2, \ldots, r_i, \ldots, r_N\}, r_{i+1} > r_i$ be the set of available rates. For each transmission, the AGV needs to pick a a certain data rate $r \in R$ to transmit D bits with sensory data. Fundamentally, depending on the selected r, the transmission duration D/r can be larger or smaller in exchange for, respectively, lower or higher packet error probability, which also depends on the current channel correlation state (G or B). Hence, the fundamental problem is to select a transmission rate that min-

imizes the resulting packet latency, and, correspondingly, to minimize the XTE. Let $T = \{t_1, t_2, \ldots, t_i, \ldots t_N\}$, $t_{i+1} > t_i$ be the time instances in which the AGV starts transmitting packets, and $\{[\Delta_{\mathrm{L}}(t_1), c(t_1)], [\Delta_{\mathrm{L}}(t_2), c(t_2)], \ldots, [\Delta_{\mathrm{L}}(t_N), c(t_N)]\}$ the sequence of measured age of loop and channel correlation states $c \in \{G, B\}$ at the corresponding transmission time, the goal is to find a rate selection policy $\pi : [\Delta_{\mathrm{L}}(t_i), c(t_i)] \rightarrow r_i$, $\forall t_i \in T$, $\forall c \in \{G, B\}$, $\forall r_i \in R$ that minimizes the infinite-horizon track-error:

$$\begin{aligned} &\pi^* = \arg\min_{\pi} \int_0^{\infty} y_e(t) dt \\ &\text{s.t. } (F.13), (E.4). \end{aligned} \tag{E.13}$$

5.2 Solution Proposal

The problem in (F.17) can be decomposed into sub-problems, such that between two consecutive AGV transmissions $[t_i, t_{i+1}), \forall t_i \in T$, the data rate $r_i \in R$ at t_i is selected based on the corresponding AoL and channel states $\{\Delta_{\mathrm{L}}(t_i), c(t_i)\}$, using a policy π. If l is the downlink delay randomly selected from the selected distribution for $\mathcal{G}(x)$ and D/r the total duration to transmit the uplink data, the one-stage decision cost $\mathcal{C}$ for selecting r at AoL state $\Delta_{\mathrm{L}}(t_i)$ is:

$$\mathcal{C}(\Delta_{\mathrm{L}}(t_i), r) = \int_{\Delta_{\mathrm{L}}(t_i)}^{\Delta_{\mathrm{L}}(t_i)+D/r+l} y_e(t) dt, \tag{E.14}$$

which depends only on the current Δ_{L} and the decision r taken on that state. The overall goal is to find an update rule for the AGV transmission data rate that minimizes the track error cost of its short and long-term rate decisions along the path. Such decision-making process occurring in continuous and irregular times represents a typical Semi Markov Decision Process (semi-MDP) [13], where we can model as:

State space

The state is composed by the current AoL value at the beginning of each transmission and the corresponding channel correlation state $\{\Delta_{\mathrm{L}}(t_i), c(t_i)\}$, $\forall t_i \in T$, where the state transition and the corresponding probabilities depend on the downlink latency distribution f and the packet error probabilities (F.9), according to (E.15). We can extend the state space by including SNR values for addressing the cases where instantaneous SNRs are given instead of the average.

$$\begin{aligned} &P(\Delta_{\mathrm{L}}(t_{i+1}),\, c(t_{i+1}) \,|\, c(t_i) = G,\, r_i) = \\ &\begin{cases} (1 - P_{gg}) f(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \Delta_{\mathrm{L}}(t_i) + \frac{D}{r_i} + l,\; c(t_{i+1}) = B \\ P_{gg} f(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \frac{D}{r_i} + l,\; c(t_{i+1}) = G \end{cases} \end{aligned}$$

$$P(\Delta_{\mathrm{L}}(t_{i+1}),\, c(t_{i+1}) \,|\, c(t_i) = B,\, r_i) = \\ \begin{cases} P_{bb} f(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \Delta_{\mathrm{L}}(t_i) + \frac{D}{r_i} + l,\; c(t_{i+1}) = B \\ (1 - P_{bb}) f(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \frac{D}{r_i} + l,\; c(t_{i+1}) = G \end{cases} \tag{E.15}$$

Action

The action decision is the corresponding data rate $r_i \in R$ selected at the beginning of each transmission t_i.

Reward

The immediate reward is given by the one-stage cost in (F.18), where we can evaluate over $l \sim f(k, \theta)$, as:

$$\mathcal{C}(\Delta_{\mathrm{L}}(t_i), r) = \mathbb{E}_f \left[\int_{\Delta_{\mathrm{L}}(t_i)}^{\Delta_{\mathrm{L}}(t_i) + D/r + l} y_e(t) dt \right]. \tag{E.16}$$

Solution

We have solved the proposed semi-MDP to obtain the optimal policy π^* using a classical dynamic programming strategy, named value iteration. This method is guaranteed to provide an optimal solution by iteratively solving the Bellman equation combining sweeps of policy evaluation and policy improvement [13].

Proposition 2: Using a wireless transmission rate policy π, the total track error in (F.17) is bounded and the convergence rate is a function of P_{bb} for a given downlink latency $l \sim f(k, \theta)$.

Proof: Consider $\mathcal{S}$ the overall set of possible AoL values where $\Delta_{\mathrm{L}} \in \mathcal{S}$ and a fixed rate transmission policy $\pi : [\Delta_{\mathrm{L}}, c] \to r$ where r is selected $\forall c \in \{G, B\}$. As in [2], we can define the overall cost in (F.17) as a function of the intermediate costs (F.18):

$$\sum_{\Delta_{\mathrm{L}} \in \mathcal{S},\, c \in \{G,B\}} \mathcal{C}(\Delta_{\mathrm{L}}, \pi(\Delta_{\mathrm{L}}.c)) \Phi_\pi(\Delta_{\mathrm{L}}, c), \tag{E.17}$$

with $\Phi_\pi(\Delta_{\mathrm{L}}, c)$ denoting the stationary probability of AoL state $[\Delta_{\mathrm{L}}, c]$ using the policy π. Since the rate is fixed by r, we will have $\mathcal{S} \in \{i \cdot D/r + l\}\ \forall i \in \mathbb{N}^+,\ l \sim (f(k, \theta))$, where $i \cdot D/r$ is achieved after a sequence of $c = \{\underbrace{B, B, \ldots, B}_{i-1}, G\}$, thus obtaining Φ_π as:

$$\Phi_\pi(\Delta_{\mathrm{L}}, c) = \sum_{i=1}^{i \cdot \frac{D}{r} < \Delta_{\mathrm{L}}} f(\Delta_{\mathrm{L}} - i \cdot D/r)(1 - P_{bb}) P_{bb}^{i-1}, \tag{E.18}$$

Table E.1: System Parameters

Carrier frequency	2.4 GHz
Bandwidth, B	1 MHz
Mean SNR, γ	1 dB
Data message size, D	1 KByte
DL latency Gamma pdf shape k and scale θ	(2.0; 3.5) [11]
DL latency fixed time intervals	5 ms
AGV velocity, v	10 km/h
Doppler frequency, f_D	22.16 Hz
Inter-axle distance, L	0.5 m
Control matrices R and Q	I_{2x2}, I_{4x4}
Maximum Steering angle, δ	45^o
Path Planner	Cubic spline [14]

From (E.7), we can verify that the track error grows as a function of $v(t)\sin\theta_e(t)$. By plausibly assuming the worst case scenario where $\sin\theta_e(t) = 1$ under a maximum speed $v(t) = \mathcal{V}_{max}$ starting at $y_e(t_0) = 0$, the one-stage cost $\mathcal{C}(\Delta_{\mathrm{L}}, r)$ becomes the integral over a right trapezoid, where:

$$\mathcal{C}(\Delta_{\mathrm{L}}, r) = \frac{1}{2}(\frac{D}{r} + l)(\mathcal{V}_{max})(2\Delta_{\mathrm{L}} + \frac{D}{r} + l).$$

So, as $\Delta_{\mathrm{L}} \to \infty$ for any given downlink latency l, we have:

$$\sum_{\Delta_{\mathrm{L}}=1}^{\infty} \mathcal{C}(\Delta_{\mathrm{L}}, r)\Phi_{\pi}(\Delta_{\mathrm{L}}, c) =$$
$$\sum_{\Delta_{\mathrm{L}}=1}^{\infty} \frac{\mathcal{V}_{max}(D/r + l)(2\Delta_{\mathrm{L}} + D/r + l)}{2P_{bb}^{l/D/r}} f(l)(1 - P_{bb})P_{bb}^{\Delta_{\mathrm{L}}/D/r},$$

since $|P_{bb}| < 1$, both infinity series $\sum_{\Delta_{\mathrm{L}}=1}^{\infty} P_{bb}^{\Delta_{\mathrm{L}}/D/r}$ and $\sum_{\Delta_{\mathrm{L}}=1}^{\infty} \Delta_{\mathrm{L}} P_{bb}^{\Delta_{\mathrm{L}}/D/r}$ are convergent. For a variable transmission rate policy, where r is optimally selected along different states, we can refer to [2] to demonstrate that the sufficiency of current proof is enough since the optimal variable rate policy results in an average cost no higher than that of a fixed rate policy, thus finishing the proof of Proposition 2.

6 Numerical Results

Table E.1 summarized the main system parameters from the proposed model used for the numerical evaluation of XTE.

Fig. E.2 represents the policy derived by the proposed solution. There is a different optimal rate selection depending on the channel correlation state. When the channel is in a good state, the policy chooses the maximum rate, i.e., to deliver the information as fast as possible. However, in a bad state, the transmission rate is gradually lowered. This increases the packet transmission delay while increasing the chances to bring the communication to a good state again. As a result, this reduces the overall data latency and increases the control accuracy. The figure also shows when the AoL is low, even though in the bad state, the policy chooses a higher transmission rate despite the higher risk for failure. Here, it makes sense to take more risk of transmission failure, but on high AoL levels, risking failure is not the best option.

Fig. E.3 provides a performance comparison between the obtained policy, fixed transmission data rate policies and the solution in [2] as baseline. We observe the maximum AGV track error captured along the path for different simulation rounds, as steep path deviations are centrally critical for the AGV operation. Essentially, we can observe that the optimal fixed data rate performance is approaching the proposed solution, which, in turn, offers the best track error result, especially for the Gamma latency. The baseline relies uniquely on the use of AoI, causing the optimization to be inherently limited to a single communication link and, thus, sub-optimal. Our proposed method tackles such limitation by using AoL to consistently improve the AGV performance, a gain that can be relatively meaningful, especially for critical control applications, such as AGV control within factory halls.

7 Conclusions

In this letter we have proposed a model of a remote AGV control that dynamically adapts the wireless transmission rate aiming to optimize the AGV trajectory. We account both effects of DL and UL using the age of loop concept, as well as the channel correlated fading effect. Our numerical results have endorsed the importance of the proposed solution for goal-oriented AGV applications. As future work, we plan to tackle more complex factory scenarios by including AGV path control planning at the proposed semi-MDP problem.

References

[1] Z. Sun, R. Wang, Q. Ye, Z. Wei, and B. Yan, “Investigation of intelligent vehicle path tracking based on longitudinal and lateral coordinated control,” *IEEE Access*, 2020.

[2] Huang, Kang *et al.* , “Wireless feedback control with variable packet length for industrial IoT,” *IEEE Wireless Communications Letters*, 2020.

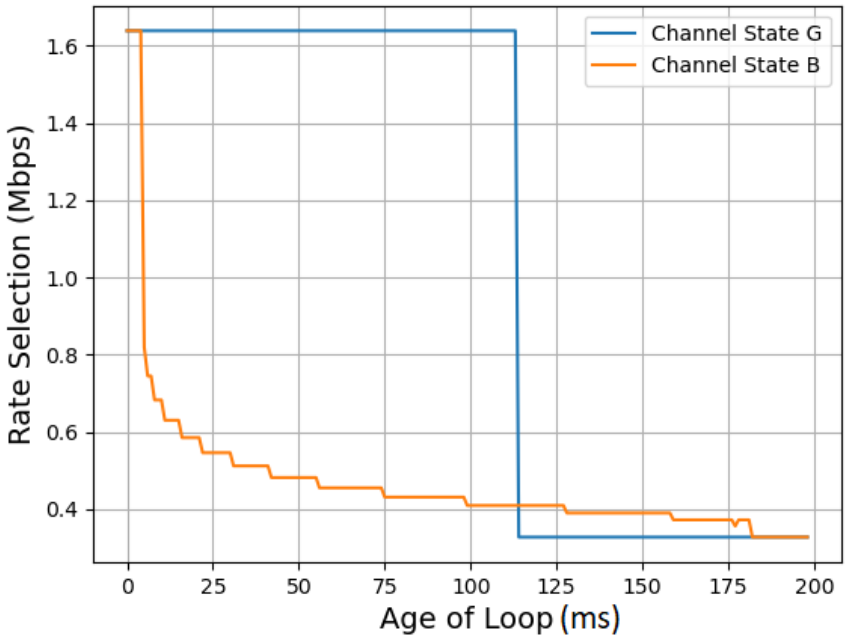


Fig. E.2: Policy behavior for data rate selection according to each state.

[3] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the latency, rate and reliability tradeoff in wireless networked control systems for IIoT," *IEEE IoT Journal*, 2020.

[4] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Age of loop for wireless networked control systems optimization," in *IEEE PIMRC*, 2021.

[5] M.-Y. Lee and B.-S. Chen, "Robust h network observer-based attack-tolerant path tracking control of autonomous ground vehicle," *IEEE Access*, 2022.

[6] A. J. Rojas, "Modified algebraic Riccati equation closed-form stabilizing solution," *IEEE Access*, 2021.

[7] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Wireless control of autonomous guided vehicle using reinforcement learning," in *IEEE GLOBECOM*, 2020.

[8] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, 2021.

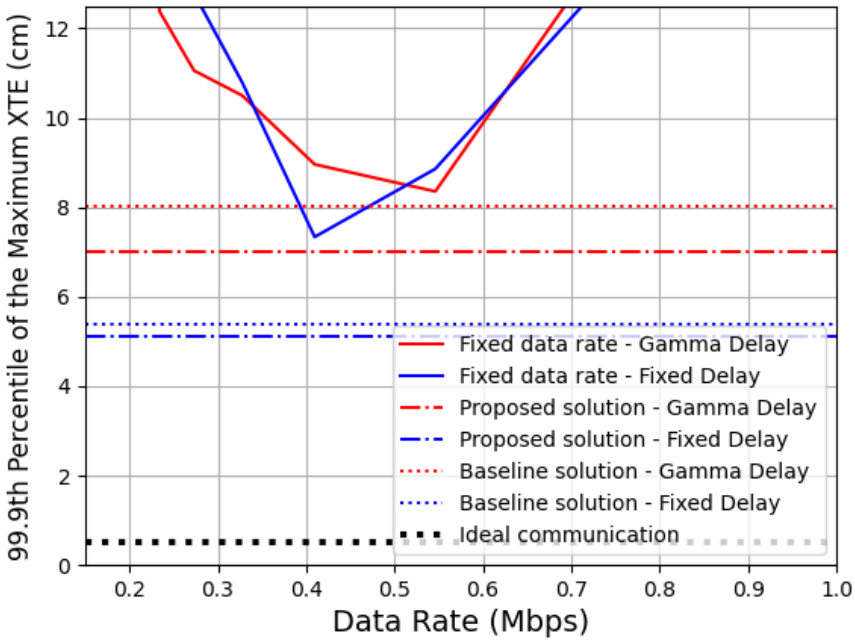


Fig. E.3: Results for the 99.9th XTE percentile. Baseline taken from [2].

[9] L. Chettri and R. Bera, "A comprehensive survey on internet of things (IoT) toward 5G wireless systems," *IEEE IoT Journal*, 2019.

[10] S. Nayak, S. Yeotikar, E. Carrillo, E. Rudnick-Cohen, M. K. M. Jaffar, R. Patel, S. Azarm, J. W. Herrmann, H. Xu, and M. Otte, "Experimental comparison of decentralized task allocation algorithms under imperfect communication," *IEEE Robot. Autom. Lett.*, 2020.

[11] S. Yasuda and H. Yoshida, "Prediction of round trip delay for wireless networks by a two-state model," in *IEEE WCNC*, 2018.

[12] M. S. Mahmoud and M. M. Hamdan, "Fundamental issues in networked control systems," *IEEE/CAA Journal of Automatica Sinica*, 2018.

[13] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[14] A. Sakai, D. Ingram, J. Dinius, K. Chawla, A. Raffin, and A. Paques, "Pythonrobotics: a python code collection of robotics algorithms," *arXiv preprint:1808.10703*, 2018.

Paper F

Experimental Study of Information Freshness for Goal-Oriented Wireless Communication in a Factory

Pedro M. de Sant Ana, Nikolaj Marchenko, Beatriz Soret and Petar Popovski

The paper has been submitted to
IEEE Internet of Things Journal 2023.

The layout has been revised.

Abstract

Traditionally, transmission of wireless data targets rate and reliability as generic requirements. The new paradigm of goal-oriented wireless communication focuses on the effectiveness*: the impact the received bits have on the actual goal of the application that relies on wireless connectivity. Recently, we have introduced the Age-of-Loop (AoL) metric for goal-oriented wireless optimization in Wireless Networked Control Systems (WNCS). AoL is an an alternative to the well-known Age-of-information (AoI) that is suited for networked control and is simpler to measure in practical systems as compared to AoI. We have set up a measurement campaign within a factory environment to collect and analyze AoL data and experimentally verify its significance for WCNS problems. We propose a numerical model to characterize the AoL behavior in this scenario and utilize it to address the problem of wirelessly controlled Autonomous Guide Vehicle (AGV). We formulate this problem as a semi-Markov-Decision Process (MDP) where the AGV trajectory is optimized by controlling network model parameters. The obtained optimized policy using AoL provides a better performance for the AGV as compared to the solutions that rely on AoI.*

1 Introduction

Goal-oriented wireless communication is a new design paradigm and it is seen among the key 6G enablers of critical and massive machine-type communications [1] [2]. The idea is to establish new communication principles, which are no longer ruled by traditional radio link performance indicators, but rather by the ultimate application goal. More precisely, the usefulness of the information with respect to the goal of the data exchange is considered, seeking the *effectiveness* of the communication [1]. Although being a recent research topic, this approach has demonstrated numerous advantages [3–5], and its potential for the joint design of network and control systems is to be unleashed.

Understanding the application effectiveness from a network design perspective can be a puzzling problem, especially if we aim at a high granularity level. The questions of defining and quantifying the attributes of effectiveness are still open in current literature. Age of Information (AoI), a metric that quantifies information freshness, has been seen as a generic attribute that represents the relevance of information [5]. In fact, the AoI has been explored in recent literature as a potential metric for analyzing and optimizing wireless networked controlled systems (WNCS) [6–10]. The main shortcoming of such approaches, however, is that designing AoI-based methodologies requires the solutions to be constrained to a single communication link, such that all the optimizations separately occur either in the downlink (DL) or the uplink (UL) direction.

In [11], we demonstrated that such constraint can lead to sub-optimal behavior for closed-loop control problems, and introduced the Age-of-Loop (AoL) as an alternative

to AoI. The AoL metric is defined to account for both DL and UL effects, as well as their interplay, which enabled us to fundamentally learn the network requirements of a closed-loop control system and use this information to optimally allocate network resources.

In this paper, we build on [11] and demonstrate the practical application of AoL in closed-loop control systems with wireless links. Specifically, the paper contributions are: (a) We set up a measurement campaign using a 5G standalone network to measure and analyze the AoL behavior within a Bosch factory environment in Stuttgart-Feuerbach [12], showing its practical edge over AoI, since no time synchronization is needed for the end-to-end evaluation; (b) By analysing the AoL data, we design a Gilbert-Elliot communication model to numerically characterize the obtained AoL behavior in the factory, where we further validate the proposed model using the empirical AoL data; (c) We use the obtained AoL model to build on a wireless controlled AGV application, where we can statistically verify the network impact over the AGV track-error performance; and (d) We formulate a problem of AGV trajectory optimization by controlling the network model parameters as Semi-Markov Decision Process, where the obtained optimized policy using AoL provide better AGV performance than AoI baseline solutions.

The rest of the paper is organized as follows. In Section 2, we elaborate on wireless networked control systems and introduce the problem of wireless AGV control. In Section 3, we present details about AoL, explaining the main definitions and related constraints compared to AoI. In Sections 4 and 5, we define a model for the AGV control and for the wireless communication, respectively, and study the impact of the wireless channel on the AGV performance. We validate the proposed model with AoL measurements obtained in a factory setting using a 5G-SA network, where the measurement scenario is described in Section 6 and its corresponding analysis and methodology in Section 7. Finally, in Section 8, we formulate a problem of AGV trajectory optimization, where we design a semi-MDP to establish a goal-oriented wireless transmission policy. We evaluate the proposed approach, comparing with baseline methodologies that relies on the use of AoI. Concluding remarks are given in Section 9.

2 Wireless Networked Control Systems

WCNS represents a dynamic unstable plant (e.g., AGV, robotic arm, drone) where the control loop is closed through a wireless network, as illustrated in Fig. F.1. There are many advantages of WNCS deployment in comparison with traditional control systems in terms of flexibility, maintainability, and lower costs. In this context, WNCS is considered an essential enabler for future industrial, logistics, and transport applications, where a high level of flexibility, data fusion, resource sharing, and cost reduction are desired [13].

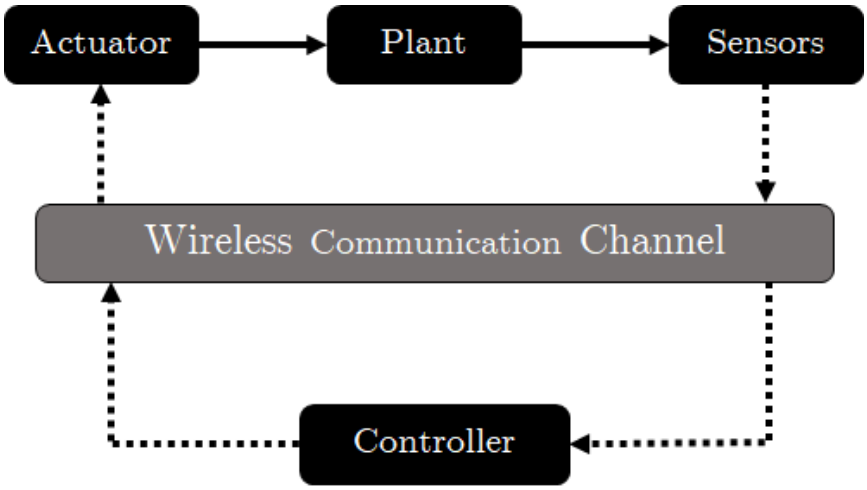


Fig. F.1: General WNCS model example.

Nevertheless, the dynamic nature of wireless networks introduces a series of challenges not present in wired control loops, which can affect the behavior of the WNCS by degrading its performance or causing instability. For example, data transmissions from the sensor to the controller or from the controller to the actuator are subject to communication constraints (e.g., limited bandwidth or power, interference), causing variable delays and packet dropouts. The current literature on WNCS can be classified into two main categories [13]:

1. Control *of* network [14–16]: considers the problem of communication network only, such as radio resource management, routing or congestion protocols, network topology, etc. The focus is essentially achieving network configurations that attain the quality-of-service (QoS) constraints in WNCSs.

2. Control *over* network [17–19]: focuses on the design of control systems by assuming predictable constraints about the network behavior. Here the focus is on the optimization of the quality-of-control (QoC) in stochastic environment caused by the network behavior.

In both categories there is a clear separation between the control and the network entities, which are separately optimized. This leads to proposed solutions that rely on (strong) assumptions about the behavior of either the network or the control. Such decoupling in systems with low-latency and high reliability leads to over-provisioning of the communication resources [11]. Furthermore, actions taken at a control system level can have a direct impact on the communication system and vice-versa, as formulated by

Witsenhausen as counterexample for distributed control problem in [20] and exemplified for an autonomous guided vehicle (AGV) use case in [21]. For this reason, in this paper, we focus on a third category, which is the *control-oriented network*, where the decisions taken by the network do not rely only on network QoS KPIs, but actually on the application goal itself. We illustrate the proposed approach through the use-case of a wireless controlled AGV, as elaborated next.

2.1 Wireless AGV Tracking Problem

Among different WNCS applications, the real-time wireless control of autonomous vehicles is seen as an essential enabler for future industrial and logistics operation [21]. It is an attractive solution for reducing hardware costs and providing higher transportation flexibility and coordination. As illustrated in Fig. F.2, the idea is to deploy an edge cloud infrastructure that constantly receives sensor updates from a remote located AGV, followed by control input commands to the AGV actuators. The main goal is to make the AGV to follow a desired path by minimizing the track-error, which is the difference between the current path and the desired path. However, both communication directions (i.e., sensor to controller and controller to actuator) are subject to wireless impairments, such as network-induced delays and packet dropouts. Such uncertainty caused by the network can potentially harm the control system performance, thus directly impacting the AGV trajectory, which is risky in time-critical applications, such as manufacturing. This leads to the broader problem of: considering all the communications constrains, how to adapt the network parameters in order to minimize the AGV trajectory error? As suggested in [9], the answer for this question is not so obvious, even leading to counter-intuitive examples, such as system improved stabilization under higher latency scenarios. We elaborated further on this problem in [21], where we demonstrated that the AGV track-error is primarily impacted by two main factors: 1) the status of the AGV physical states, such as its velocity, and 2) the potential burst of communication errors. Furthermore, in [22] we introduced a new metric, Age-of-Loop (AoL), showing that the can use machine learning to grasp the control system behavior and subsequently use the obtained information to take optimal decisions for both the network and the control system. In this paper, we build on the concepts proposed in both works. Specifically, we address the optimization of the AGV track-error based on AoL and validate the approach with experimental data.

3 Age of Information and Age of Loop

The AoI is defined as the time elapsed since the last information update at the destination. It represents a metric that quantifies the freshness of the knowledge we have

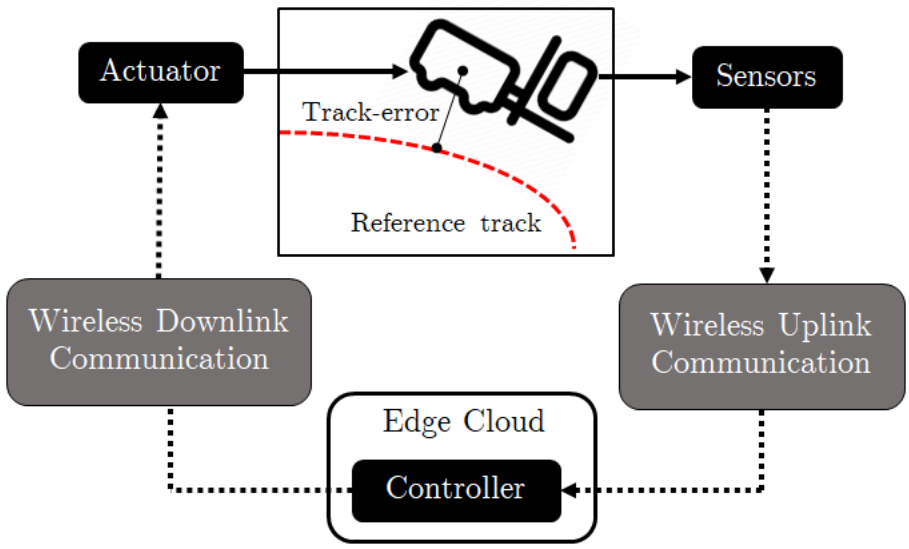


Fig. F.2: Wireless AGV control problem.

about the status of a remote system. Mathematically [23],

$$\Delta(t) = t - U(t), \tag{F.1}$$

where $U(t)$ is the timestamp of the most recent data received (i.e., with the largest generation time) and $\Delta(t)$ represents the AoI process at time t. The functions of AoI to acquire alternative target metrics have been examined in several publications [6, 7, 23]. Other work used AoI as a state information metric for control optimization [8–10].

However, there is a limitation of AoI as a metric, as it is originally suited for one-way communication. For this reason, previous work that has addressed WNCS-related issues using AoI has been limited to specific analysis for UL [6–8] or DL [9, 10] transmissions only. However, wirelessly networked control systems inherently rely on both DL and UL because it is a closed-loop control problem where UL communications can affect DL and vice versa, affecting system performance and network resource utilisation. Consider the following for illustration: A high AoI in the UL (sensor-control) implies that the controller knows less about the status of the system, which requires more urgency to transmit the control signal and consequently consumes more network resources in the DL (control-sensor).

A possible alternative that can be used in a closed-loop WNCS is the Age-of-Loop (AoL) [11], defined as a metric that captures the elapsed time of the entire loop process and considers both UL and DL links. This is clarified through the WNCS example on

Fig. F.3. Consider the closed-loop control system where an AGV sends its states update and expect a feedback control command from an edge controller. In this context, denote $\mathcal{S} = \{t_{s_1}, t_{s_2}, \ldots, t_{s_i}, \ldots\}, t_{s_{i+1}} \geq t_{s_i}$, a sequence of time instances where data packets at the AGV are generated and subsequently sent to the edge controller, while $\mathcal{C} = \{t_{c_1}, t_{c_2}, \ldots, t_{c_i}, \ldots\}, t_{c_{i+1}} \geq t_{c_i}$, the time instances where the corresponding feedback control command is received by the AGV from the controller, satisfying an arbitrary arrival process where $t_{c_i} > t_{s_i}$. Additionally, we can define $\delta_i = t_{c_i} - t_{s_i}$ the total time encompassing the packet generation until the control loop is closed, which can include the time spent for transmissions (DL and UL), queuing and control computation. For a given timestamp t, the latest control loop cycle was initiated at the timestamp:

$$s(t) = \sup\{t_{s_i} \in \mathcal{S} : t_{s_i} + \delta_i \leq t\}, \tag{F.2}$$

such that we can define the Age of Loop as:

$$\Delta_{\mathrm{L}}(t) := t - s(t). \tag{F.3}$$

Remark: The AoI is defined for two independent links, requiring instantaneous and perfect feedback channel for the sender to know the age at the receiver. This is a practical limitation often neglected in the literature. In contrast, AoL can be easily implemented and exploited in a real network, as it captures the behavior of both UL/SL directions into a single metric that can be easily measured at either of the end points.

4 AGV Control System Model

To model the AGV, we define the vehicle state vector $\mathbf{X} \in \mathbb{R}^4$ and control vector $\mathbf{U} \in \mathbb{R}^2$ over time t as in [24]:

$$\mathbf{X}(t) = [x(t), y(t), v(t), \theta(t)], \ \mathbf{U}(t) = [a(t), \delta(t)]. \tag{F.4}$$

where x, y are the 2D coordinates, v is the velocity, and θ is the heading orientation angle of the vehicle. The control is based on manipulating the vehicle acceleration a and front wheel angle δ to control the trajectory of the vehicle along a desired path. Details of the model implementation, design, and structure of the control system can be found in [21]. The dynamics of the control system is defined according to the kinematic model [24], such that $\dot{x}(t) = v(t)\cos\theta(t)$, $\dot{y}(t) = v(t)\sin\theta(t)$, $\dot{v}(t) = a(t)$ and $\dot{\theta}(t) = v(t)\tan\delta(t)/L$, for the vehicle inter-axle distance L. The notation $\dot{\square}$ corresponds to the derivatives w.r.t time. The kinematic vehicle model is linearized by dynamically applying the first-order Taylor expansion around the current vehicle position as the operating point [24], leading to the ordinary differential equation:

$$\dot{\mathbf{X}}(t) = f(\mathbf{X}(t), \mathbf{U}(t)) = \mathbf{A}(t)\mathbf{X}(t) + \mathbf{B}(t)\mathbf{U}(t) \tag{F.5}$$

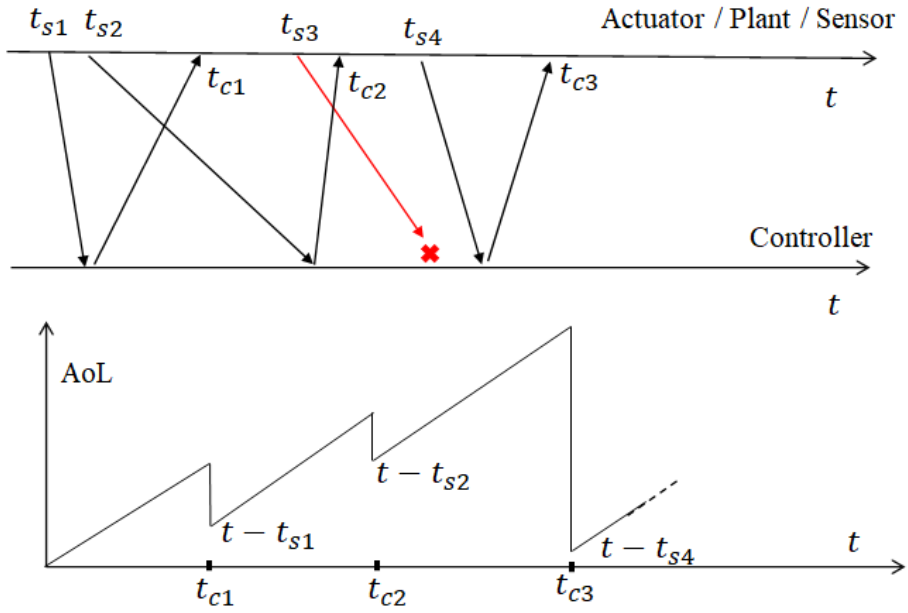


Fig. F.3: WNCS time diagram.

where the transition matrices $\mathbf{A}(t)$ and $\mathbf{B}(t)$ are given as [24]:

$$\mathbf{A}(t) = \begin{bmatrix} 0 & 0 & \cos\theta(t) & -v(t)\sin\theta(t) \\ 0 & 0 & \sin\theta(t) & v(t)\cos\theta(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\tan\delta(t)}{L} & 0 \end{bmatrix} \quad \mathbf{B}(t) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & \frac{v(t)}{L\cos^2\delta(t)} \end{bmatrix}$$

4.1 AGV Control and Performance

To control the AGV, consider a trajectory planner that specifies a sequence of desired states $\mathbf{X}^*(t)$ that the AGV must follow. So, we must adjust the vehicle position by minimizing the error state vector, $\mathbf{X_e} = \mathbf{X}(t) - \mathbf{X}^*(t)$, between the current state $\mathbf{X}(t)$ and the desired state $\mathbf{X}^*(t)$ along the planned path, as given by [25]:

$$\mathbf{X_e}(t) = \begin{bmatrix} \cos\theta^*(t) & \sin\theta^*(t) & 0 \\ -\sin\theta^*(t) & \cos\theta^*(t) & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x(t) - x^*(t) \\ y(t) - y^*(t) \\ v(t) - v^*(t) \\ \theta(t) - \theta^*(t) \end{bmatrix}, \tag{F.6}$$

To achieve this, we can formally define a cost function J whose objective is to minimize $\mathbf{X_e}(t)$ by applying the control commands $\mathbf{U}(t)$ subject to the system dynamics in (F.5). In control theory, this type of control design is known as Linear-Quadratic-Regulator (LQR) [26], formulated as

$$J(\mathbf{X_e}(t), \mathbf{U}(t)) = \int_0^\infty \left[\mathbf{X_e}(t)^T \mathbf{Q} \mathbf{X_e}(t) + \mathbf{U}(t)^T \mathbf{R} \mathbf{U}(t)\right] dt, \tag{F.7}$$

where the arbitrary positive defined matrices $\mathbf{Q}$ and $\mathbf{R}$ serve to weight the impact of the state space and the control signal, respectively. For $(\mathbf{A}, \mathbf{B})$ controllable, we can find the optimal control command $\mathbf{U}^*(t) = [a^*(t), \delta^*(t)]$ that minimizes (F.7) by solving the Algebraic Riccati Equation [26]. This solution for the LQR problem provides a convergent closed-loop system, where the control-system stability can be guaranteed [26].

The performance of the AGV control is evaluated according to the cross track error (XTE), a commonly used metric that measures the distance the AGV has deviated from the planned path. More formally, the XTE is given by the lateral distance between the selected tracing point and the point of the path that is closest to it. Hence, we can derive the XTE from the error state vector in (F.6), where we are interested in the distance y_e between the virtual car position as its projection on the path:

$$y_e(t) = -\sin\theta^*(t)[x(t) - x^*(t)] + \cos\theta^*(t)[y(t) - y^*(t)] \tag{F.8}$$

By properly designing the optimal LQR control in (F.7), the XTE is minimized along the path [27]. If the communication between the plant and the controller (or vice-versa) cannot provide the information, we expect some impact on the XTE performance even with an optimal control policy. We discuss this impact in more detail in next section.

5 Wireless communication model

5.1 Uplink and Downlink Transmissions

We consider the two directions of the communication model: the uplink (UL) for communication between sensor and controller, and the downlink (DL), for communication between controller and actuator. In most control applications, including AGVs, it has been found that control data is relatively small compared to sensor data [28], resulting in a strong imbalance between DL and UL data traffic. Consequently, AGVs often need to send large image or video data, such as light detection and ranging (LiDAR) cloud points and augmented or virtual reality (AR/VR) information. However, the control packets in the opposite direction are based on sending a short control command, such as acceleration or vehicle heading angle, which can either be based on a control algorithm or on human control as in haptic or AR/VR use cases. In this context, we design a

communication model in which the UL network resources, representing the bottleneck, are dynamically adjusted. We consider not only the UL, but also statistical inferences about the behavior of DL.

We model UL communication as a first-order Markov process [29], also known as the Gilbert-Elliot channel. The time-correlation property is represented by two states: the good state G if the packet can be successfully received; and the bad state B, otherwise. The corresponding transition probability matrix is defined by a stationary Markov distribution:

$$\mathbf{M} = \begin{bmatrix} \mathcal{P}_{gg} & 1 - \mathcal{P}_{gg} \\ 1 - \mathcal{P}_{bb} & \mathcal{P}_{bb} \end{bmatrix}. \tag{F.9}$$

where $\mathcal{P}_{gg}$ is the probability that the current transmission is successful, if the previous transmission was successful. The same logic applies to $\mathcal{P}_{bb}$, but for unsuccessful transmissions. Given the matrix M, the channel properties are fully characterized [29]. In particular, the marginal probability of a packet error, $\mathcal{P}_{\varepsilon}$, is then given by:

$$\mathcal{P}_{\varepsilon} = 1 - \frac{1 - \mathcal{P}_{bb}}{2 - (\mathcal{P}_{bb} + \mathcal{P}_{gg})} \tag{F.10}$$

For Rayleigh fading, $\mathcal{P}_{\varepsilon} = 1 - e^{\gamma_{th}}$, where γ_{th} is the minimum threshold SNR required to successfully decode the received signal. We can evaluate γ_{th} as a function of the utilized data rate R, bandwidth B and the average SNR γ, as:

$$\gamma_{\text{th}} = \frac{2^{R/B} - 1}{\gamma}, \tag{F.11}$$

We define the Jakes's channel correlation coefficient to be $\rho = J_0(2\pi f_d T_s)$, where $J_0()$ is the zero-order Bessel function of the first kind, f_d is the Doppler frequency shift and T_s is the sampling time. The error probability of a single back to back failure, $\mathcal{P}_{bb}$, is then written as [29]

$$\mathcal{P}_{bb} = 1 - \frac{Q(\theta, \rho\theta) - Q(\rho\theta, \theta)}{e^{\gamma_{th}} - 1}, \tag{F.12}$$

where $Q(.,.)$ is the Marcum Q function and $\theta = \sqrt{\frac{2\gamma_{th}}{1-\rho^2}}$.

The transmission latency for sending D bits at a data rate of R is D/R. If the transmission fails, we assume that a retransmission occurs immediately. Retransmission automatically increases the resulting data latency or data delivery time. Increasing R decreases the latency for the transmission of a single packet, but increases the error probability of the packet according to (F.9)-(F.12). This illustrates the dependence between the data rate R and the control accuracy (i.e., XTE) in the Section 8.

Given the potential imbalance described in this section between DL and UL traffic, we assume that transmission rate adjustment is not useful in DL communications. However, DL packet delays can be caused by factors unrelated to communication, such as the

control application's processing time [28]. We consider a probability distribution $f_\tau(x)$ for generality, where $l \sim f_\tau(x)$ is the DL packet delivery time. Based on experimental data from a factory environment, we derive a model for $f_\tau(x)$ in section 6.

5.2 Wireless Channel impact over AGV Control

When using the proposed wireless channel model, communication becomes non-deterministic, with delays and packet losses. This has an impact on the behavior of the control system in the equation (F.5). Fig. F.4 shows a timing diagram of the signal transmissions at the remotely controlled AGV. The corresponding wireless transmissions are independent of the periodic sampling T_s, so that a new transmission starts immediately after the previous one, allowing the possibility of multiple transmissions within the sampling time window. Such a model describes the operation of many industrial use cases where the communication system is completely decoupled from the application, so that learning its behavior becomes part of the network adaptation problem.

For this scenario, we can model the system dynamics in (F.5) using the well-known stochastic system approach [13], in which we can consider as a function of AoL the delay from sensor to actuator, from actuator to sensor, and the effect of packet loss, so that

$$\dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{B}(t - \Delta_{\mathrm{L}}(t)). \tag{F.13}$$

In an ideal communication scenario, we would have $\Delta_{\mathrm{L}}(t) \to 0$, $\forall t$, such that (F.13) would be equivalent to (F.5) and the sufficient condition for minimizing the track error is given by the control optimization. More formally, we can refer to [30] (Proposition 1) to verify that the AGV tracking error has an asymptotically non-negative w.r.t $\Delta_{\mathrm{L}}(t)$. Hence, from a control perspective, we must ideally close the system loop as fast as possible to guarantee an optimal XTE performance.

According to Section 5, the total transmission time in the UL for sensory data of size D is D/R, where the probability of packet loss is defined by (F.9). The latency for the DL must necessarily follow a probability density function. In this case, we have the option of transmitting at a high UL data rate in order to quickly minimize AoL, but at the risk of a high packet loss probability. Alternatively, we can transmit at a lower data rate, which reduces the possibility of error but increases the time required to complete the sensory data transmission. In other words, increasing the data rate decreases the delay in transmitting a single packet while increasing the likelihood of transmission error and the need for re-transmissions. This also increases data latency, which is similar to the well-known trade-off issue observed in [31].

6 AGV Measurements in Factory Environment

In this Section, we provide details of a measurement campaign of a private 5G NR Stand-Alone (SA) network deployed in a Bosch factory. Using the collected data we

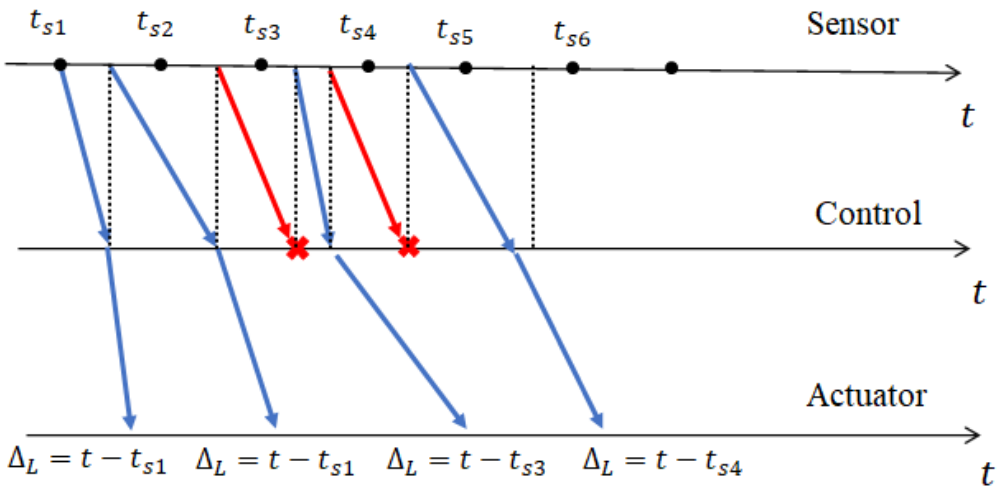


Fig. F.4: Timing diagram of signals transmitted.

later derive corresponding AoL values and emulate the resulting performance of the the AGV control application.

6.1 Private 5G Network Setup

Fig. F.5 shows schematically the main physical and some functional components of the network installed in the plant. This includes the following equipment:

- 8xAirScaIe Pico Remote Radio Heads (pRRHs) for 5G NR SA operation in band n78 according to 3GPP Release 15 specification [32].
- Edge server aiming at the offload of designated UE tasks, such as image processing and control.
- 5G BBU: 5G Baseband unit, for radio baseband signal processing.
- One Wireless Device (UE).
- Router: For Traffic Routing to Intranet and Internet, as well as the 5G Network Core.

The network operates 100 MHz of bandwidth in band n78 (Time Division Duplexing), on the 3.7 GHz carrier frequency. It supports maximum 256QAM modulation and uses 4x4 MIMO in Uplink and Downling. The RF Output is from 50 mW to 250 mW per TX path. The DL to UL ratio is configured as 7/3.

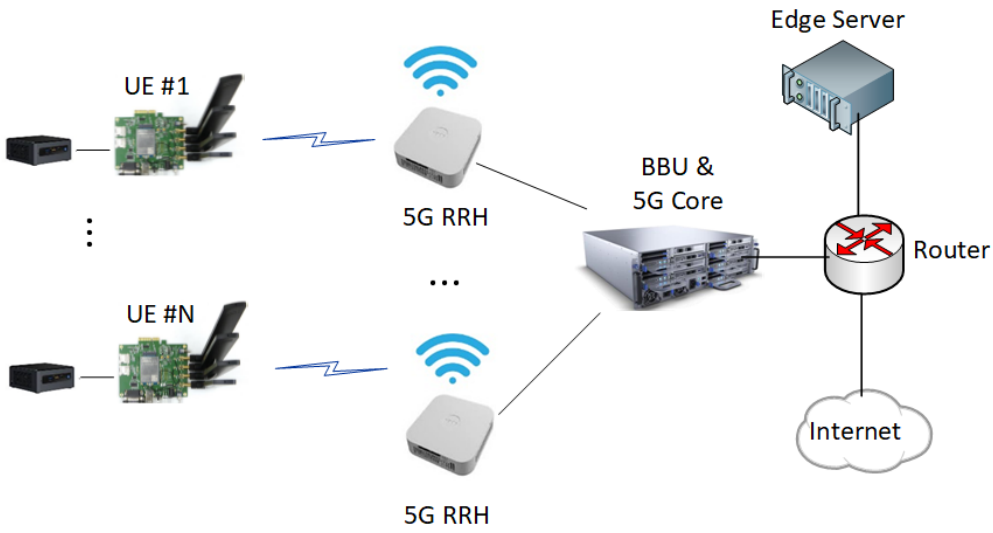


Fig. F.5: Main network components.

Fig. G.2 shows the shopfloor map, where the private 5G NR SA network has been deployed. The eight purple dots mark the locations across the production hall, where the pRRHs were mounted close to the ceiling. The pRRHs serve as smarter antennas served by the singly BBU unit. In the current setup, all antennas send the same information in DL and can receive the same information in UL from UEs. The BBU takes care of properly decoding the strongest signals. In this way, it is possible to configure all eight pRRHs in the factory in one 5G cell, providing the following benefits for this architecture:

- The whole 100 MHz bandwidth available everywhere where the radio coverage is available.
- No handover events.
- Lower costs for the pRRHs and BBUs.

As a UE device for the measurement over the 5G network we used a Quectel RM500Q M2 module (based on the Qualcomm Snapdragon X55 5G modem) with a Quectel development board connected via USB to a Linux mini-PC, which is then placed on the AGV (see Fig. G.2). Essentially, the mini-PC serves as actuator/sensor according to the model shown in Fig. F.1. A powerful Linux-based edge server is Ethernet-connected to the 5G network servers, and takes the role of a controller.

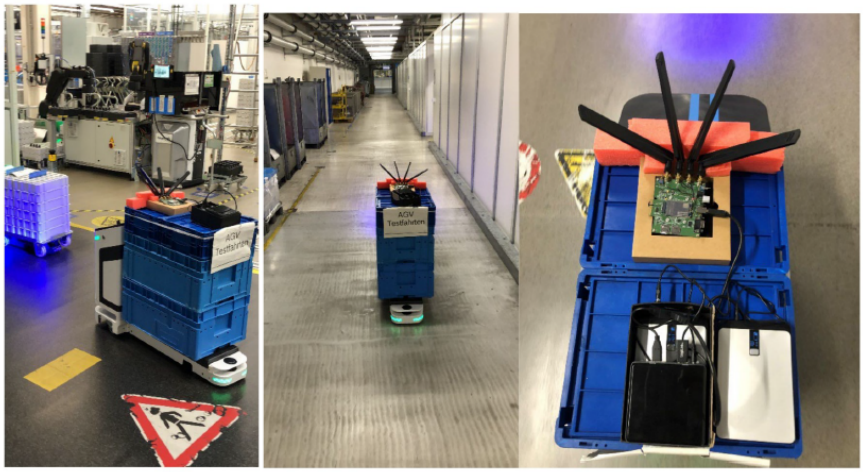


Fig. F.6: AGV drive test across factory hall.

6.2 AGV Drive Test

Bosch Rexroth ActiveShuttle is a commercial AGV used for transporting parts and goods in the production. In our setup, an ActiveShuttle is assigned three pick-up/drop-off jobs in such an order that it has to go through all corridors in the 5G test area. The ActiveShuttle is equipped with LiDAR sensor which is used to obtain its current location on the shopfloop. Fig. F.7 shows a trace of the AGV drive around the production hall. Numbers marked yellow show the time and the location the AGV was at after the start. The maximum speed of the AGV is set to 1 m/s due to safety requirements.

The ping command is executed at the UE-side to trigger a ping response from the destination edge server, emulating periodic status update of the control loop cycle represented in Fig. F.4. The ping request interval time is set to 10 ms. In addition, to log all traffic going through 5G network interfaces, Wireshark was running at the mini-PC and at the edge server. This enabled logging of the exact time each ping packet enters and leaves network interfaces on the client and server side.

7 Evaluation of AoL

Using the measurements of the AGV drive test, we empirically evaluated the AoL under the proposed drive test scenario. The Fig. F.8 illustrates the AoL measured during the AGV drive test at the first 500 ms. From the AoL definition, the bottom peaks correspond equivalently to the round-trip time (RTT). The top peaks represent the maximum

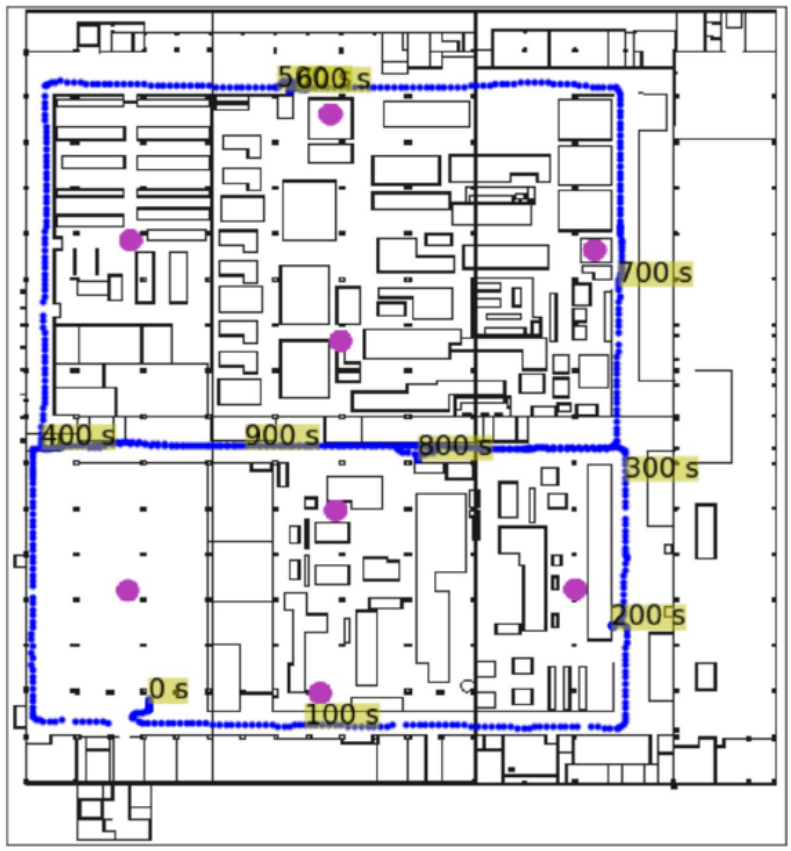


Fig. F.7: Factory shopfloor plan, deployment of the 5G RRHs, and an AGV route across the factory hall.

amount of time where the control loop remains opened. As elaborated on [21], such peak values can be used to evaluate the performance degradation of the control system, as well as its stability. Fig. G.3 illustrates an histogram of the AoL behavior along the test, where we can observe an average of around 20 ms and a standard deviation of approximately 7.5 ms. Nevertheless, besides the low frequency of occurrence, we can also identify tail values that can achieve more than 120 ms. From a practical perspective, such tail values are essentially critical, especially when it comes offloading highly unstable control applications in the edge cloud. For instance, the inverted pendulum control example explored by the authors in [21] demonstrates that the control stability is compromised after the AoL surpass values of 40 ms.

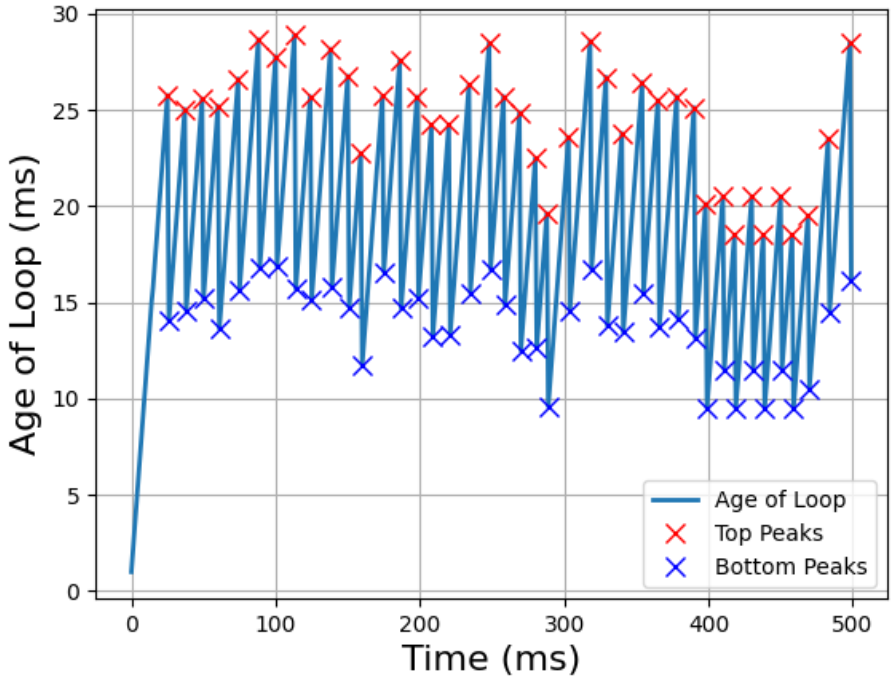


Fig. F.8: Age of Loop over time.

7.1 Probability Density Estimation

Using once again the measurements performed during the AGV drive test, we estimate the probability density function from the empirical AoL data. We start with a parametric estimation approach [33], where we calculate the Residual Sum of Squares (RSS) [33] of multiple theoretical distributions, as well as its corresponding parameters, in order to find the best fitted distribution. More formally, consider $\mathcal{F} = \{f_1(z|\boldsymbol{w_1}), f_2(z|\boldsymbol{w_2}), \ldots, f_i(z|\boldsymbol{w_i}), \ldots, f_n($ a finite set of $n \in \mathbb{N}^+$ different probability distributions (e.g., Gaussian, Gamma, uniform, etc) over the explanatory variable z, where $\boldsymbol{w_i} \in \mathbb{R}^{p_i}, \forall i \in \{1, \ldots, n\}$ is a vector that characterizes the corresponding parameter space of the distribution f_i, containing a fixed number of parameters $p_i \in \mathbb{N}^+$. So, the goal is to find the optimal distribution f^* and the corresponding parameters $\boldsymbol{w^*}$ that minimizes the RSS, as:

$$f^*(z|\boldsymbol{w}^*) = \underset{\substack{\forall i \in \{1,2,\ldots,n\} \\ \forall \theta_i \in \mathbb{R}^{p_i}}}{\arg\min} \left\{ \sum_{j=1}^{N} \Big(y_j - f_i(x_j|w_i) \Big)^2 \right\}, \tag{F.14}$$

for N the total number of empirical samples, y_j the j-th value of the variable to be

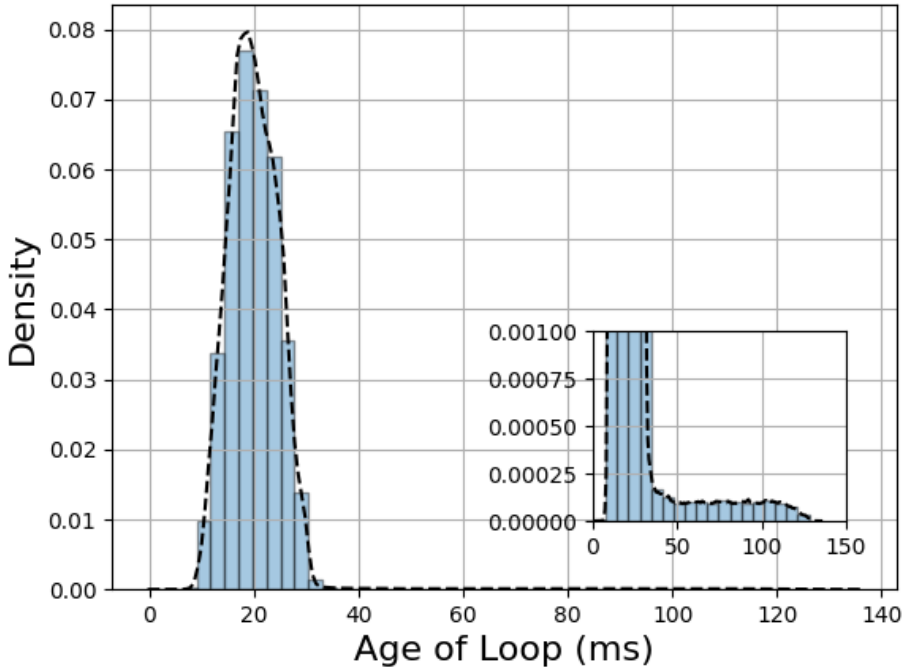


Fig. F.9: Empirical Age of Loop distribution during the AGV drive test.

predicted and $f_i(z_j|w_i)$ the predicted value of y_j using the distribution i with parameters w_i.

In general terms, the RSS is a widely used statistical method that measures the amount of variance of a given data set that is not explained by a regression model. More specifically, since we are summing the square values of the residuals (i.e., the deviation between the predicted and empirical values), we are basically evaluating the dispersion of the data points in comparison to each proposed distribution model. Small RSS values represent a strong indication that the proposed model correctly fits the empirical data points. Fig. F.10 illustrates the minimum RSS value obtained according to each probability distribution model.

As we can note, the Gamma distribution provided the best fit for the AoL data values. To emphasize the comparison between the empirical and theoretical models, the Fig. F.11 illustrates the Cumulative Distribution Function (CDF) of both distributions. An interesting conclusion about the obtained result is its direct relation with the related literature [34–36], where the authors analyze RTT measurements of LTE and Wi-Fi, showing that a Gamma distribution also represents a good fit for the RTT estimation and design.

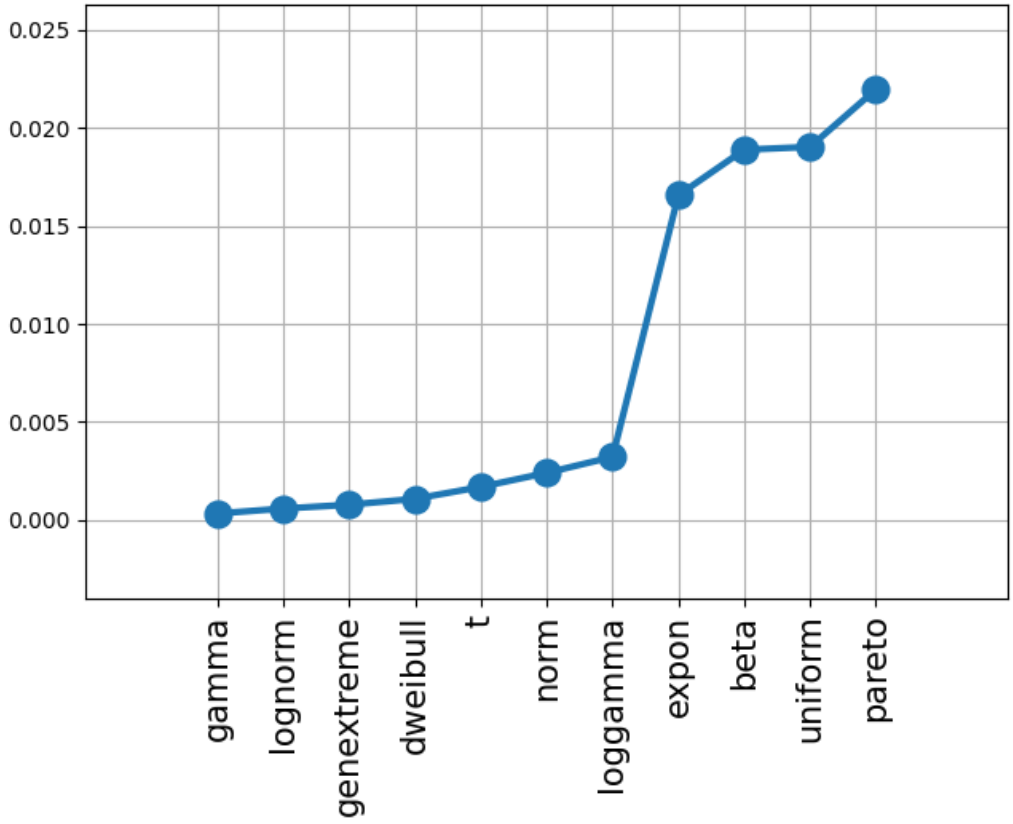


Fig. F.10: RSS comparison

7.2 AoL Configuration

Using the logs collected from incoming and outgoing traffic of both the AGV and the edge, we investigate four different effects that mainly contribute to the AoL: the ping sampling process, the edge computation delay and the communication latency in DL and UL. More formally, for every packet $k \in \mathbb{N}^+$, consider $\{s_1^1, s_2^1, \dots, s_k^1, \dots\}, s_k^1 \geq s_{k-1}^1$ the time instances of outgoing packets at the AGV and $\{s_1^2, s_2^2, \dots, s_k^2, \dots\}, s_k^2 \geq s_{k-1}^2$ the time instances of outgoing packets at the edge. Likewise, we can separately define the time instances of incoming packets as $\{d_1^1, d_2^1, \dots, d_k^1, \dots\}, d_k^1 \geq d_{k-1}^1$ and $\{d_1^2, s_2^2, \dots, d_k^2, \dots\}, d_k^2 \geq d_{k-1}^2$ for, respectively, the AGV and the edge.

By considering the proposed formulation, the ping sampling process is simply evaluated as $s_{k+1}^1 - s_k^1$ and the edge latency as $s_k^2 - d_k^2$, $\forall k \in \mathbb{N}^+$, which can be directly measured using the corresponding local traces of each communication module. For the

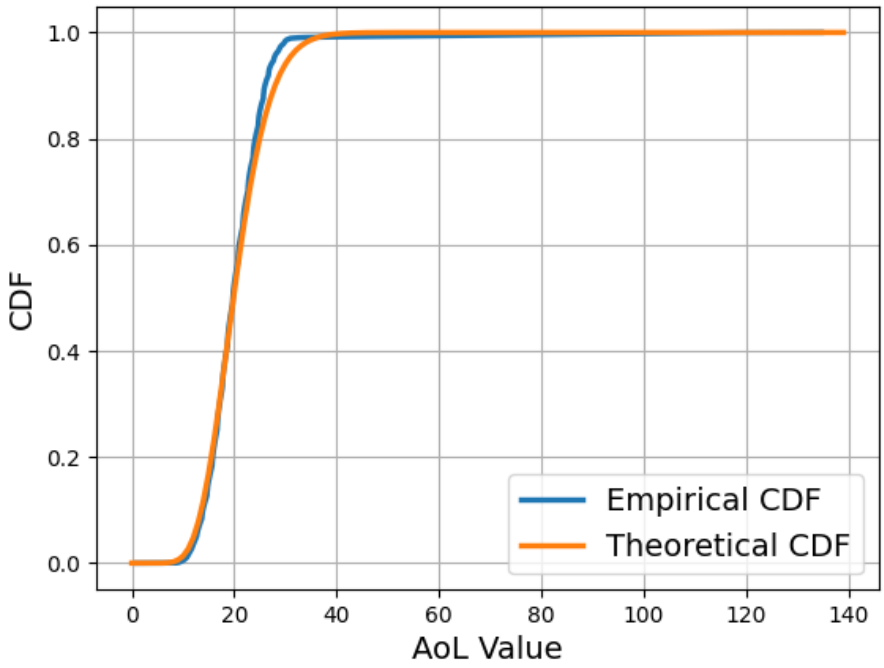


Fig. F.11: CDF comparison

evaluation of the latency, we can define $\tau_k^1 = d_k^1 - s_k^1$ as the latency for the UL direction and $\tau_k^2 = d_k^2 - s_k^2$ as the latency for the DL direction. The problem, however, is that, since we lack precise timing synchronization between both communication entities, we cannot directly evaluate τ_k^1 and τ_k^2 by simply subtracting the incoming and outgoing time of a given packet k. So, we propose a recursive strategy to evaluate the latency, which can provide a satisfying approximation while totally preserving the main statistical properties of the latency probability distribution. To achieve that, consider $\Delta_k^1 = \tau_{k+1}^1 - \tau_k^1$ and $\Delta_k^2 = \tau_{k+1}^2 - \tau_k^2$ as the increment of latency between each packet transmission, where we can re-arrange the equations as following:

$$\Delta_k^1 = \tau_{k+1}^1 - \tau_k^1 = (d_{k+1}^1 - d_k^1) - (s_{k+1}^1 - s_k^1),$$
$$\Delta_k^2 = \tau_{k+1}^2 - \tau_k^2 = (d_{k+1}^2 - d_k^2) - (s_{k+1}^2 - s_k^2),$$

such that τ_k^1 and τ_k^2 then can be rewritten as:

$$\tau_{k+1}^1 = \Delta_k^1 + \tau_k^1,$$
$$\tau_{k+1}^2 = \Delta_k^2 + \tau_k^2.$$

Note that Δ_k^1 and Δ_k^2 can be evaluated using the time differences of incoming and outgoing packets between each communication module without the need of timing synchronization. Hence, we can recursively calculate the latency values, considering, however, a first guess for the initial values τ_1^1 and τ_1^2, in which we assume as half of the first RTT measurement:

$$\tau_1^1 = \tau_1^2 = \frac{d_1^1 - s_1^1}{2}.$$

We summarize the obtained results of each effect contributing to the AoL according to Fig.s F.12 and F.13, such that we can fundamentally emphasize three main conclusions. First, besides the ping triggering process being configured to exactly every 10 ms, we can still encounter a random behavior at the sampling process, which is possibly due to internal hardware interrupts at the NUC CPU. Second, the edge latency, as expected, is quite negligible (< 0.1 ms), also showing less variance at the generation of ping replies compared to the NUC, which is likely expected since the hardware capability at the edge exceed the one at the NUC. This result, however, could definitely worsen given a more complex task to be performed at the edge (e.g., AGV path planning or path tracking control), so it is still important to monitor and track the edge behavior. Finally, we can observe a more interesting conclusion regarding the DL and UL latency results. In general (up to the 99-th percentile), the DL latency is slightly better than the UL, which is somehow expected given that the network capabilities between both entities are different. The problem, however, is that we can also observe very high latency values (> 100 ms) at the DL direction that comes with very low probability (after the 99.99-th percentile), a behavior that is not observed at the UL, which has in turn its highest value at around 30 ms. As a consequence from this observation, we can potentially infer that the tail values observed for the AoL in Fig. G.3 arise as a direct effect of the latency behavior in DL communication.

7.3 Empirical Downlink Model

As discussed in Section 5, we did not consider transmission rate adaptation at the DL communication, as DL data packets have considerably smaller size compared to the UL. However, DL packet delays can also unavoidably happen, as we could grasp from the obtained measurement results in Fig.s F.13 and F.12.

In this context, aiming at designing a general and realistic model for the DL behavior, we propose an approach as elaborated in sub-section 7.1. Using the empirical data obtained from the DL latency variable τ_k^2, we estimate its probability density function by finding the distribution model and the corresponding parameters that minimizes

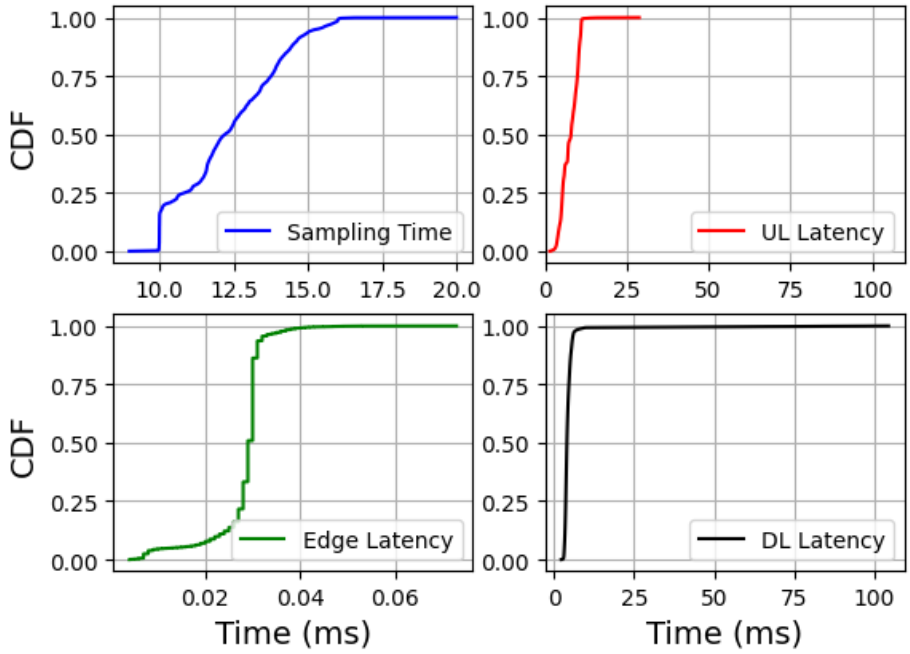


Fig. F.12: CDF comparison of each effect contributing to the total AoL.

the RSS according to (F.14). Assuming that $f_\tau(z|\mathbf{w})$ is the optimal PDF model and corresponding parameters describing the distribution of τ_k^2, the DL latency will follow a beta distribution, which is described by two parameters, $\mathbf{w} = [w_1, w_2]$, and formally defined as:

$$f_\tau(z, w_1, w_2) = \frac{\Gamma(w_1 + w_2) x^{w_1 - 1}(1 - z)^{w_2 - 1}}{\Gamma(w_1)\Gamma(w_2)}, \tag{F.15}$$

where we provide w_1 and w_2 as a function of the shift, μ, and scale, φ, variables over (F.15). More specifically, consider $f_\tau(z, w_1, w_2, \mu, \varphi)$ an equivalent form to $f_\tau(y, w_1, w_2)/\varphi$, with $y = (z - \mu)/\varphi$, so the obtained optimal values are: $w_1 = 2.319$, $w_2 = 2335.509$, $\mu = 2.259$ and $\varphi = 2355.869$. The Fig. F.14 illustrates the CDF and PDF comparison between the empirical data and the theoretical beta distribution.

7.4 Simulation of AoL behavior

Considering the wireless communication model described in last subsections, we aim finding the model parameters that provide an approximate AoL behavior according to the empirical results illustrated in Fig. G.3. Since we assume an unavoidable DL

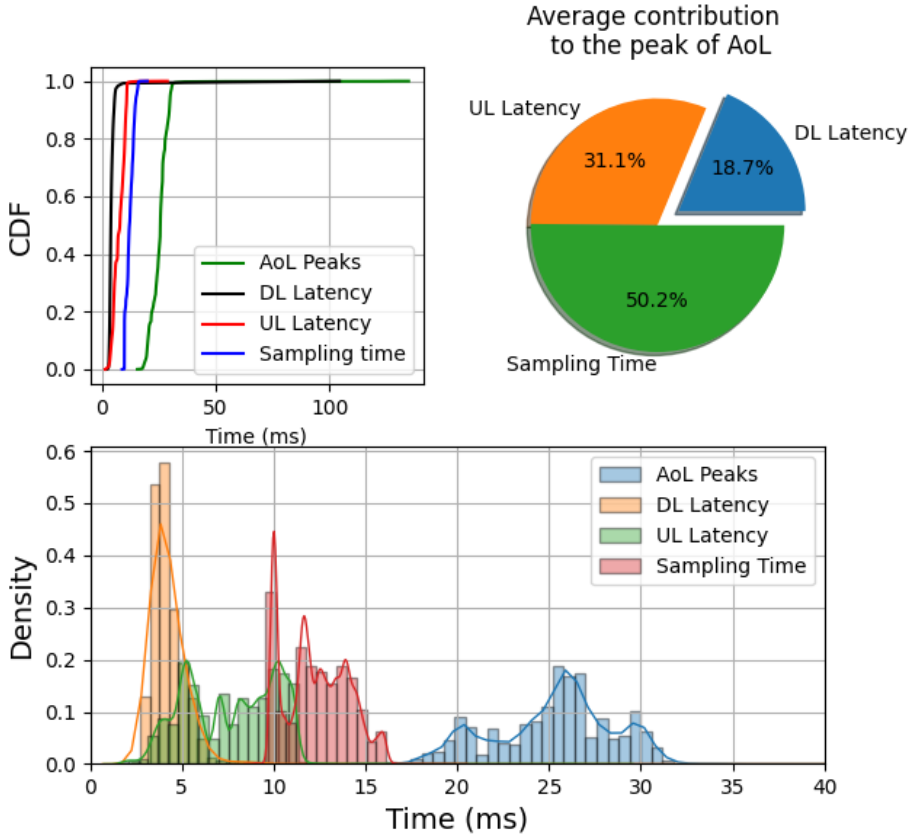


Fig. F.13: Comparison of each effect contributing to the total AoL.

latency modeled according to the empirical measurements, the goal then is to find the corresponding parameters of the first-order Markov process at the UL that, in addition to the DL behavior and sampling process, can statistically provide a simulated AoL behavior that fits the empirical AoL results.

This problem represents a typical hyperparameter searching, where we approach a solution by using a grid search methodology, based on numerically evaluating the Kolmogorov-Smirnov (KS) test between the empirical AoL data and the generated AoL data from the model as a criteria to select the optimal UL parameters. The algorithmic description 6 provides more details about the proposed methodology, where we must emphasize three main aspects. First, we fix parameters that are usually not manageable by the network, such as the AGV velocity, sampling time and sensor data size, as well as

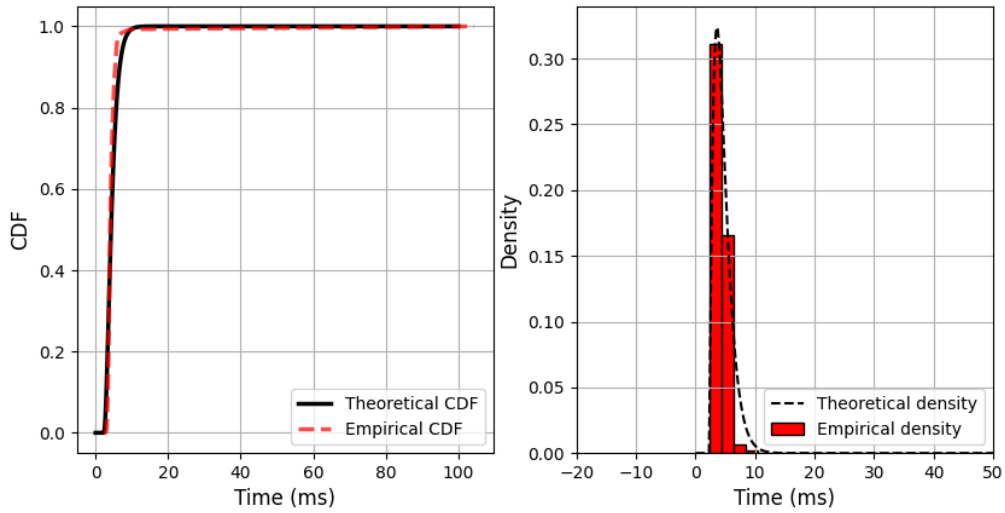


Fig. F.14: Probability density estimation for modeling the DL latency.

the carrier frequency. Consequently, the search space is basically only composed by the bandwidth allocation and the SNR, thus justifying the use of grid search instead of more complex methodologies such as evolutionary or gradient-based optimization. Second, per each bandwidth and SNR search points, we evaluate the generated AoL data for a range of different UL data rate values. This generalization of the AoL behavior will play an important role to the problem of data rate adaptation that will be introduced in the next section. Finally, we apply the KS test as a measure to quantify the statistical distance between the simulated and empirical AoL data values, where we can accept or reject the hypothesis that the generated data is from the specified empirical distribution at certain level of significance. The KS test has shown to provide superior estimates of error in curve fitting models [37].

We summarize the network and control parameters according to Table F.1. The Fig. F.15 illustrates the CDF and PDF comparison between the simulated AoL results and the empirical data, showing that the proposed communication model can provide a good approximation to the real measurements, more specifically at the 0.05 significance level.

Algorithm 6 Hyperparameter searching for the UL communication model.

Initialize: Velocity v, carrier frequency f_c, sensor data size D, sensor sampling T_{sensor}, and empirical AoL data Δ_L.

1: Define search space for bandwidth allocation $\mathcal{BW}$, the signal-to-noise ratio γ and range of data rates R.
2: **for** each $i \in \mathcal{BW}$ **do**
3: **for** each $j \in \gamma$ **do**
4: Init empty array of simulated AoL values $\tilde{\Delta_L} = [\,]$
5: **for** each $k \in R$ **do**
6: Evaluate the Distribution in (F.9).
7: **for** $t \in [0, \ldots, N_{iterations}]$ **do**
8: Generate new sensor data every T_{sensor} seconds.
9: UL communication with latency D/k and packet error probability according to (F.9).
10: DL communication latency according to (F.15)
11: Collect AoL measurements and append to $\tilde{\Delta_L}$.
12: **end for**
13: **end for**
14: Evaluate the Kolmogorov-Smirnov statistic distance d_{KS} between $\tilde{\Delta_L}$ and Δ_L.
15: Save parameters space $[i, j]$ for minimum d_{KS}.
16: **end for**
17: **end for**

Table F.1: Network Parameters

Carrier frequency, f_c	3.5 GHz
Sensor sampling time, T_{sensor}	10 ms
Sensor data size, D	1 KByte
AGV velocity, v	7 km/h
Doppler frequency, f_D	22.68 Hz
Bandwidth, B	1.3 MHz
SNR	10 dB
Data rates, R	$[0.2, 0.4, 0.6, \ldots, 5.0]$ Mbps
Inter-axle distance, L	0.5 m
Control matrices R and Q	I_{2x2}, I_{4x4}
Maximum Steering angle, δ	45^o
Path Planner,	Cubic spline [38]

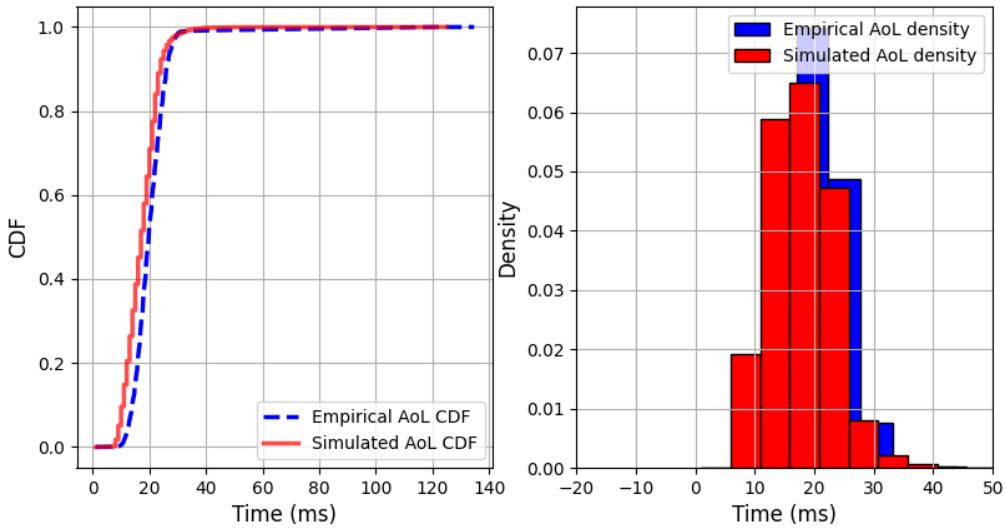


Fig. F.15: Comparison between the simulated and empirical AoL data after tuning the communication model.

$$
\begin{aligned}
&P(\Delta_{\mathrm{L}}(t_{i+1}),\, c(t_{i+1}) \,|\, c(t_i) = G,\, r_i) = \\
&\begin{cases}
(1-\mathcal{P}_{gg}) f_\tau(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \Delta_{\mathrm{L}}(t_i) + (t_i \bmod T_s) + \frac{D}{r_i} + l,\; c(t_{i+1}) = B \\
\mathcal{P}_{gg} f_\tau(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \frac{D}{r_i} + (t_i \bmod T_s) + l,\; c(t_{i+1}) = G
\end{cases} \\
&P(\Delta_{\mathrm{L}}(t_{i+1}),\, c(t_{i+1}) \,|\, c(t_i) = B,\, r_i) = \\
&\begin{cases}
\mathcal{P}_{bb} f_\tau(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \Delta_{\mathrm{L}}(t_i) + (t_i \bmod T_s) + \frac{D}{r_i} + l,\; c(t_{i+1}) = B \\
(1-\mathcal{P}_{bb}) f_\tau(l), & \Delta_{\mathrm{L}}(t_{i+1}) = \frac{D}{r_i} + (t_i \bmod T_s) + l,\; c(t_{i+1}) = G
\end{cases}
\end{aligned}
\tag{F.16}
$$

8 AGV Control with variable data rate

8.1 Problem Formulation and Solution

Using the experimental model obtained in the last section, where the behavior of DL is given by the PDF in (F.15) corresponding to the network parameters obtained in Table F.1, we formulate the following problem. Let $\mathcal{R} = \{r_1, r_2, \ldots, r_i, \ldots, r_N\}, r_{i+1} > r_i$ be the set of available rates. For each transmission, the AGV must choose a specific data rate $r \in R$ to transmit D bits of sensor data. In principle, the transmission duration D/r can be larger or smaller depending on the chosen r, in exchange for a lower or higher

packet error probability, respectively, which also depends on the current state of the channel correlation (G or B). Thus, the fundamental problem is to choose a transmission rate that minimizes the resulting packet latency, and accordingly minimize the XTE. Let $T = \{t_1, t_2, \ldots, t_i, \ldots t_N\}$, $t_{i+1} > t_i$ be the time instances in which the AGV starts transmitting packets, and $\{[\Delta_{\mathrm{L}}(t_1), c(t_1)], [\Delta_{\mathrm{L}}(t_2), c(t_2)], \ldots, [\Delta_{\mathrm{L}}(t_N), c(t_N)]\}$ the sequence of measured age of loop and channel correlation states $c \in \{G, B\}$ at the corresponding transmission time, the goal is to establish a rate selection policy $\pi : [\Delta_{\mathrm{L}}(t_i), c(t_i)] \rightarrow r_i$, $\forall t_i \in T$, $\forall c \in \{G, B\}$, $\forall r_i \in R$ that minimizes the infinite-horizon tracking error:

$$\begin{aligned} &\pi^* = \arg\min_{\pi} \int_0^{\infty} y_e(t)dt \\ &\text{s.t. } (F.13). \end{aligned} \tag{F.17}$$

The problem in (F.17) can be decomposed into sub-problems, such that between two consecutive AGV transmissions $[t_i, t_{i+1}), \forall t_i \in T$, the data rate $r_i \in R$ at t_i is selected based on the corresponding AoL and channel states $\{\Delta_{\mathrm{L}}(t_i), c(t_i)\}$, using a policy π. If l is the downlink delay randomly selected from the selected distribution for $f_\tau(x)$ and D/r the total duration to transmit the uplink data, the one-stage decision cost $\mathcal{C}$ for selecting r at AoL state $\Delta_{\mathrm{L}}(t_i)$ is:

$$\mathcal{C}(\Delta_{\mathrm{L}}(t_i), r) = \int_{\Delta_{\mathrm{L}}(t_i)}^{\Delta_{\mathrm{L}}(t_i)+(t_i \bmod T_s)+\frac{D}{r}+l} y_e(t)dt, \tag{F.18}$$

which depends only on the current Δ_{L} and the decision r made in that state. The overall goal is to find an update rule for the AGV transmission data rate that minimizes cost of tracing errors of its short- and long-term rate decisions along the path. Such decision process, which runs at continuous and irregular times, represents a typical Semi Markov Decision Process (semi-MDP) [39], which we can model as follows:

State space

The state is composed by the current AoL value at the beginning of each transmission and the corresponding channel correlation state $\{\Delta_{\mathrm{L}}(t_i), c(t_i)\}$, $\forall t_i \in T$, where the state transition and the corresponding probabilities depend on the downlink latency distribution f_τ and the packet error probabilities (F.9), according to (F.16). We can extend the state space by including SNR values for addressing the cases where instantaneous SNRs are given instead of the average.

Action

The action decision is the corresponding data rate $r_i \in R$ selected at the beginning of each transmission t_i.

Reward

The immediate reward is given by the one-stage cost in (F.18).

Solution

We have solved the proposed semi-MDP to obtain the optimal policy π^* using a classical dynamic programming strategy, named value iteration. This method is guaranteed to provide an optimal solution by iteratively solving the Bellman equation combining sweeps of policy evaluation and policy improvement [39].

8.2 Results

The strategy derived from the proposed solution is depicted in Fig. F.16. There is a different optimal rate selection depending on the state of the channel correlation. When the channel is in good condition, the policy selects the maximum rate, which means that the information is transmitted as quickly as possible. In a bad state, however, the transmission rate gradually decreases. This increases the delay in packet transmission while increasing the chances of restoring communication. This reduces the overall latency of the data and improves control accuracy. The Figure also shows that when the AoL is low, the policy chooses a higher transmission rate undeterred by the increased risk of failure, despite the fact that the state is bad. Here, taking a higher risk of transmission failure makes sense, but at high AoL values, risking failure is not the best option.

Figure F.17 compares the performance of the obtained strategy, fixed transmission rate strategies, and the solution in [9] as a baseline. Because steep path deviations are crucial for AGV operation, we observe the maximum AGV tracking error recorded along the path for different simulation rounds. In essence, we can draw two major conclusions: 1) We can see the dependency: at low fixed data rates, we see negative effects on the XTE because the latency (D/R) is high, as expected. Similarly, a high data rate has an impact on XTE performance due to the increased likelihood of packet loss. 2) Essentially, the optimal fixed data rate performance approaches the proposed solution, which provides the best track error result in turn. Because the baseline is solely based on the use of AoI, the optimization is inherently limited to a single communication link and thus sub-optimal. Our proposed method overcomes this limitation by using AoL to consistently improve AGV performance, a gain that can be relatively significant, particularly for critical control applications like AGV control within factory halls.

9 Conclusions

In this paper, we performed an experimental setup to model and analyze the wireless channel behavior of an AGV in a Bosch factory environment in Stuttgart-Feuerbach. In

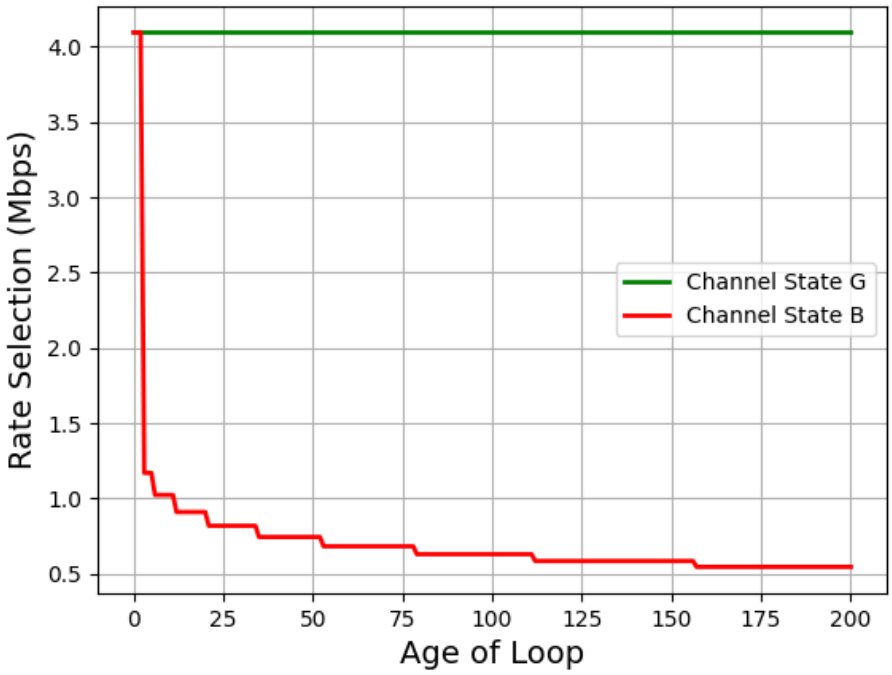


Fig. F.16: Policy behavior for data rate selection according to each state.

this context, we proposed a model for a remote AGV control system that dynamically adjusts the wireless transmission rate to optimize the trajectory of the AGV. We consider both the effects of DL and UL using the concept of loop age and the channel-dependent fading effect. We numerically analyze the targeted methodology for AGV applications and validate the proposed solution using experimental data collected in a 5G standalone network in a factory environment. The obtained results have supported the importance of the proposed solution for AGV applications, especially in factory floors. In the future, we plan to tackle more complex factory scenarios by incorporating AGV path control scheduling into the proposed semi-MDP problem.

References

[1] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Computer Networks*, 2021.

[2] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad,

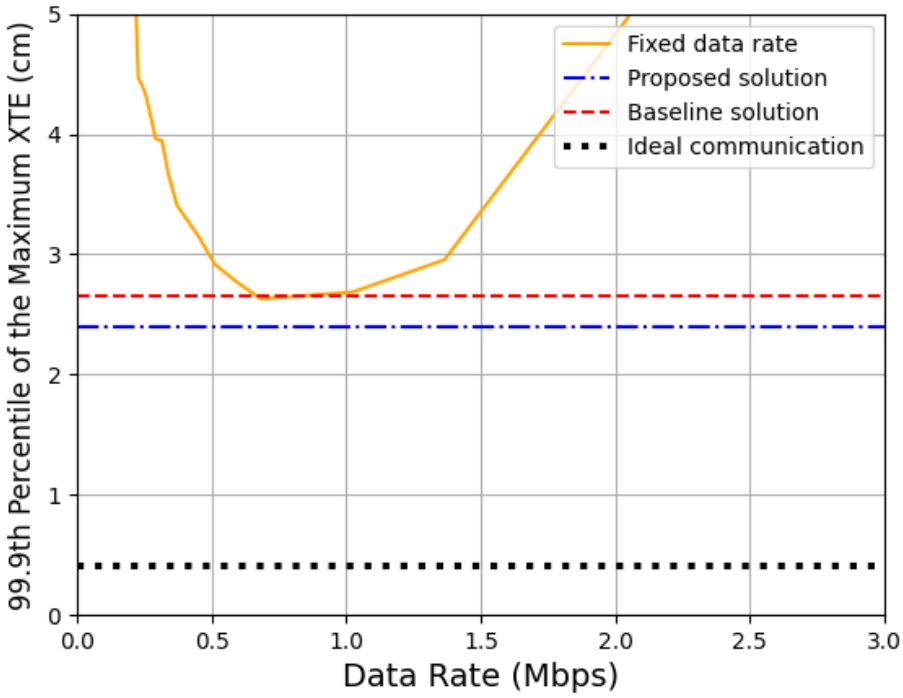


Fig. F.17: Results for the 99th XTE percentile.

G. Liva, A. Munari, B. Soret, T. Soleymani, and K. H. Johansson, "Semantic communications in networked systems: A data significance perspective," *IEEE Network*, 2022.

[3] N. Pappas and M. Kountouris, "Goal-oriented communication for real-time tracking in autonomous systems," in *IEEE International Conference on Autonomous Systems*, 2021.

[4] F. Binucci, P. Banelli, P. Di Lorenzo, and S. Barbarossa, "Dynamic resource allocation for multi-user goal-oriented communications at the wireless edge," in *European Signal Processing Conference*. IEEE, 2022.

[5] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, 2021.

[6] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using AoI in a single-loop networked control system," in *IEEE Conference on Computer Communications Workshops*, 2019.

[7] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "AoI-penalty minimization for networked control systems with packet loss," in *IEEE Conference on Computer Communications Workshops*, 2019.

[8] K. Gatsis, H. Hassani, and G. J. Pappas, "Latency-reliability tradeoffs for state estimation," *IEEE Transactions on Automatic Control*, 2020.

[9] Huang, Kang *et al.* , "Wireless feedback control with variable packet length for industrial IoT," *IEEE Wireless Communications Letters*, 2020.

[10] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the latency, rate and reliability tradeoff in wireless networked control systems for IIoT," *IEEE IoT Journal*, 2020.

[11] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Age of loop for wireless networked control systems optimization," in *IEEE PIMRC*, 2021.

[12] "The Bosch plant in Feuerbach - where tradition meets high-tech," https://www.bosch-presse.de/pressportal/de/en/bosch-in-feuerbach-tradition-meets-modernity-177029.html, accessed: 2023-03-06.

[13] M. S. Mahmoud and M. M. Hamdan, "Fundamental issues in networked control systems," *IEEE/CAA Journal of Automatica Sinica*, 2018.

[14] T. Zhu, Z. Cai, X. Fang, J. Luo, and M. Yang, "Correlation aware scheduling for edge-enabled industrial Internet of things," *IEEE Transactions on Industrial Informatics*, 2022.

[15] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward computing resource reservation scheduling in industrial internet of things," *IEEE Internet of Things Journal*, 2020.

[16] Q. Yu and M. Gu, "Adaptive group routing and scheduling in multicast time-sensitive networks," *IEEE Access*, 2020.

[17] Y. Jiang, J. Fan, T. Chai, F. L. Lewis, and J. Li, "Tracking control for linear discrete-time networked control systems with unknown dynamics and dropout," *IEEE Transactions on Neural Networks and Learning Systems*, 2017.

[18] J. Zhang and E. Fridman, "Dynamic event-triggered control of networked stochastic systems with scheduling protocols," *IEEE Transactions on Automatic Control*, 2021.

[19] X. Wang, J. Sun, G. Wang, and L. Dou, "A mixed switching event-triggered transmission scheme for networked control systems," *IEEE Transactions on Control of Network Systems*, 2021.

[20] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, 1968.

[21] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Wireless control of autonomous guided vehicle using reinforcement learning," in *IEEE GLOBECOM*, 2020.

[22] A. M. de Oliveira, V. S. Varma, R. Postoyan, I.-C. Morărescu, J. Daafouz, and O. L. V. Costa, "Network-aware controller design with performance guarantees for linear wireless systems," *IEEE Transactions on Automatic Control*, 2020.

[23] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, 2019.

[24] Z. Sun, R. Wang, Q. Ye, Z. Wei, and B. Yan, "Investigation of intelligent vehicle path tracking based on longitudinal and lateral coordinated control," *IEEE Access*, 2020.

[25] T. C. Martin, M. E. Orchard, and P. V. Sánchez, "Design and simulation of control strategies for trajectory tracking in an autonomous ground vehicle," *IFAC Journal of Systems and Control*, 2013.

[26] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[27] G. Rigatos and K. Busawon, *Robotic manipulators and vehicles: control, estimation and filtering*. Springer, 2018.

[28] L. Chettri and R. Bera, "A comprehensive survey on internet of things (IoT) toward 5G wireless systems," *IEEE IoT Journal*, 2019.

[29] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Transactions on Vehicular Technology*, 1997.

[30] P. M. de Sant Ana, N. Marchenko, B. Soret, and P. Popovski, "Goal-oriented wireless communication for a remotely controlled autonomous guided vehicle," *IEEE Wireless Communications Letters*, 2023.

[31] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, 2010.

[32] 3GPP, "Release 15 Description; Summary of Rel-15 Work Items," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 21.915, 10 2019, version 15.0.0.

[33] C. Giraud, *Introduction to high-dimensional statistics*. Chapman and Hall/CRC, 2021.

[34] D. Chen, X. Fu, W. Ding, H. Li, N. Xi, and Y. Wang, "Shifted gamma distribution and long-range prediction of round trip timedelay for Internet-based teleoperation," in *IEEE ROBIO*, 2009.

[35] S. Yasuda and H. Yoshida, "Prediction of round trip delay for wireless networks by a two-state model," in *IEEE WCNC*, 2018.

[36] Z. Juanping and G. Xianwen, "Time-delay analysis and estimation of Internet-based robot teleoperation system," in *Chinese Control and Decision Conference*. IEEE, 2009.

[37] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association*, 1951.

[38] A. Sakai, D. Ingram, J. Dinius, K. Chawla, A. Raffin, and A. Paques, "Pythonrobotics: a python code collection of robotics algorithms," *arXiv preprint:1808.10703*, 2018.

[39] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

Paper G

Goal-Oriented Wireless Communication and Control using Age of Loop

Pedro M. de Sant Ana, Nikolaj Marchenko, Beatriz Soret and Petar Popovski

The paper in under review in
IEEE Communications Magazine 2023.

The layout has been revised.

Abstract

Goal-oriented communication is a novel communication paradigm where application end-goals, often time-constrained, serve as the optimization basis of the communication process across all layers. This requires timely data transmission, usually expressed through the Age-of-information (AoI). Yet, in practice it is difficult to measure AoI due to synchronization requirements. We present Age-of-Loop (AoL) as a metric that is suitable for goal-oriented communication in Wireless Networked Control Systems (WNCS) that rely on sensing and control. We showcase an application of remote controlled Autonomous Guided Vehicle (AGV), where AoL is the key metric used to optimize the AGV trajectory. We apply Reinforcement Learning (RL) to capture the contextual value of the information through AoL, enabling solutions that optimize the overal sense-connect-control cycle.

1 Introduction

Goal-oriented communication represents a paradigm in which communication is not regarded only as a way to reproduce information correctly at a distance, but rather as a means of achieving a certain goal. Today's wireless systems are fundamentally built upon principles of reliable communications over noisy channels, where the main goal is to optimize traditional network performance indicators, such as throughput, latency, packet loss, etc. Such approaches have been demonstrated to have some critical flaws, especially in the context of Wireless Networked Control Systems (WNCS), which can lead to extreme cases of over-provisioning of network resources [1], [2].

To go beyond content-agnostic wireless connectivity, a number of recent works have turned towards the principle to tie the communication with the usefulness of the received bits for the application goal itself. This is challenging, since it may require a high granularity level of understanding the application effectiveness. Fundamental questions, such as how to define and how to quantify effectiveness attributes from a network perspective, are still open in the current literature.

In this context, many authors have been investigating the Age of Information (AoI) as a representative attribute that quantifies the information freshness, classified by the authors in [2] within a level of mesoscopic semantic scale. AoI has also been explored as a potential metric for analyzing and optimizing WNCS related problems [3–7]. Their main limitation, however, is that with AoI the proposed solutions are constrained to a single communication link, either downlink (DL) or uplink (UL). We have shown in [1] that such approach can lead to sub-optimal behavior for closed-loop control problems, and introduced the Age-of-Loop (AoL) as a metric that accounts for both DL and UL, as well as their interplay.

In this paper we provide a brief tutorial of AoL and demonstrate its practical ap-

plication in closed-loop control systems with wireless links. Specifically, we consider an experimental setup with a private 5G-NR network within a factory environment, used for remote control of AGVs. We use the measurements from this setup to analyze the AoL behavior and show its practical edge over AoI, since no time synchronization is needed for the end-to-end evaluation (Section 3). Furthermore, we showcase the use of AoL as key metric to optimize the performance of a remote controlled AGV, where the contextual value of AoL information is learned through a reinforcement learning (RL) methodology. Finally, we investigate the proposed approach from two different perspectives. First, in the context of a control application, the AGV must learn how to adjust its speed in a scenario of random and uncontrollable network behavior (Section 5). Second, in the context of a communication system, we design a scheduler to manage the bandwidth resource allocation that fundamentally targets AGV application goals (Section 6).

2 Measuring Age in Communications

Age of information is defined as the time that has elapsed since the latest information update at the destination. It represents a metric that quantifies the freshness of the knowledge we have about the status of a remote system. We can refer to its formal definition as in [8], where, at time t, if the newest data (i.e., with the largest generation time) received at the destination was generated at time $U(t)$, the AoI $\Delta(t)$ is defined as $\Delta(t) = t - U(t)$. For this reason, AoI is inherited to a single communication link. Prior works that have explored WNCS-related problems using AoI are limited to specific analysis over only the UL [3–5] or DL [6, 7] transmissions. However, wireless networked control systems relies intrinsically on both DL and UL as a closed-loop control problem, where the UL communication can affect the DL and vice-versa, impacting system performance and the use of network resources. An intuitive example to illustrate this idea is that a higher AoI value of sensor-to-control communication implies less accurate knowledge that the controller has about the plant. This demands more urgency to deliver the control signal and, as a consequence, more network resources usage by the control-sensor network link.

To address this implications, one possible alternative to evaluate the overall age of a WNCS closed-loop is the Age-of-Loop. The AoL, as defined in [1], is used as a metric that captures the time elapsed of the whole loop process, thus considering both links (UL and DL). We clarify the proposed idea with a remote AGV control example, as illustrated in Figure G.1. The AGV is sending its sensor information to a remote located controller through a wireless channel, receiving a control action as feedback. In this scenario, the AoL grows linearly over time and drops at the time instances where the control loop is closed, $(tc_1, tc_2, \dots)$, to the corresponding timestamp in which the state feedback that spawned a new control signal was generated, i.e. $\Delta_{\mathrm{L}}(t) = t - ts_i, \forall i \in \{1, 2, 3, \dots\}$.

The moment of generating data is a key factor that differentiates AoL from Round

Trip Time (RTT). While RTT measures the end-to-end delay between sender and receiver, AoL measures the time it takes for a packet to complete a loop in the network, including the time it takes for the packet to be generated and transmitted. In this sense, AoL is to RTT what AoI is to latency, since both AoL and AoI take into account the moment when data is generated, while RTT and latency are measures of the time it takes for data to travel through the network. One of main advantages of AoL is that it encompasses the behavior of two separated and locally measured entities (DL and UL) into a single instance. In contrast, for the case of two independent AoI links, we inherently need an instantaneous and perfect feedback channel to the source to know the instantaneous age at the destination, thus making complex and potentially imprecise the union of two AoI directions. In the next two sections, we show how we can potentially utilize AoL as a metric for optimizing, respectively, control decisions and network resource allocation.

3 AoL measurement in factory setting

In this section we provide an example of AoL calculation and behavior based on measurements in a private 5G NR SA network deployed in a factory production hall. This setting is illustrated in Figure G.2, as well as the AGV route across the factory hall. The network operates in band n72 (TDD) on 3.7 GHz carrier frequency, and 100 MHz bandwidth. Eight Remote Radio Heads (RRHs) (purple dots) were deployed across the production hall. The RRHs build a single cell and share radio resources so that there is no inter-cell handover required in such a deployment. RRHs are connected to a single baseband unit, a resource scheduler and a 5G core, which are deployed on the network edge servers.

As a UE device we used a Quectel RM500Q M2 module with a Quectel development board connected via USB to a Linux mini-PC, which is attached to the mobile AGV under test. Essentially, the mini-PC serves as actuator/sensor according to the model shown in Figure G.1. A powerful Linux-based edge server is Ethernet-connected to the 5G network servers, and takes the role of a controller. In addition to the ping and RF measurements, the location of the AGV was recorded.

A common way to practically measure latency in a communication system is to perform ping tests to a server. In our case we generate a new ping request from a mini-PC to the edge server every 10 ms. Based on the timestamps of the ping replies, we measure the AoL by verifying the corresponding timestamp in which the received packet at the client was generated. As we use round-trip time measurement, there is no need for high precision client-server time synchronization to evaluate the AoL. In contrast, at least millisecond-level precision would be required to measure the one-way communication delay for AoI.

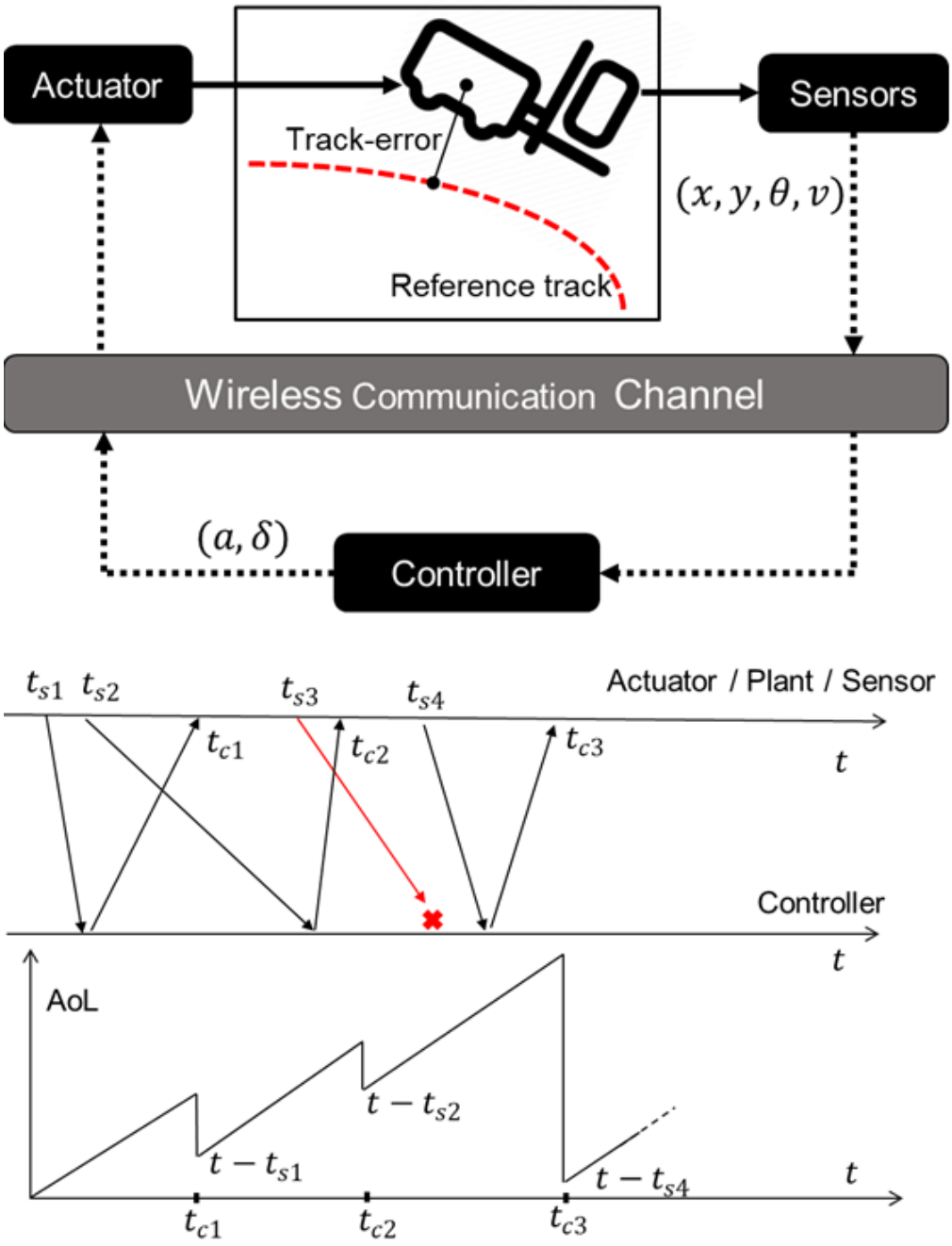


Fig. G.1: Timing diagram of signals transmitted with corresponding AoL from the AGV.

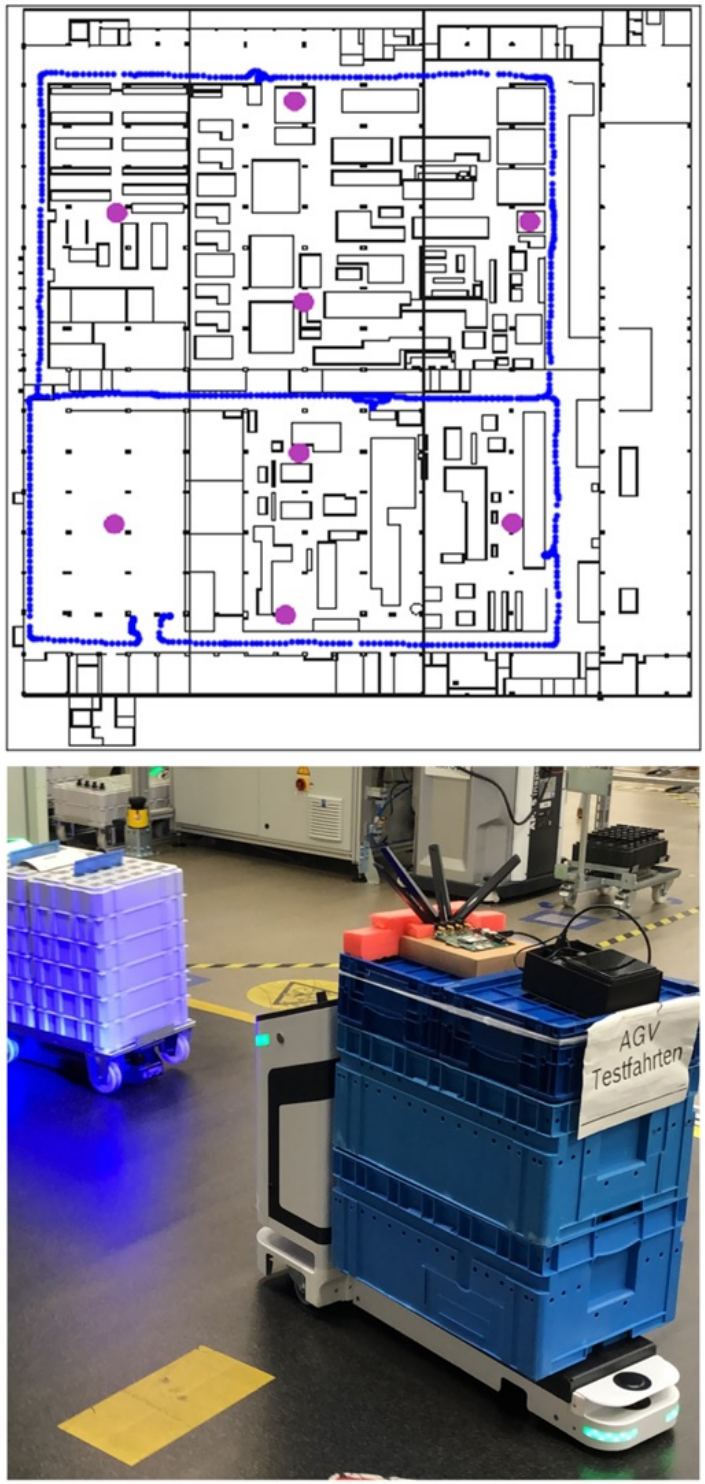


Fig. G.2: AGV drive test across factory hall.

3.1 AoL Evaluation and Estimation

Using the measurements of the AGV drive test, we experimentally evaluated the AoL under the proposed drive test scenario. The Figure G.3 illustrates the AoL empirical behavior along the test, where we can observe an average of around 20 ms and a standard deviation of approximately 7.5 ms. Nevertheless, we can also identify tail values that can achieve more than 60 ms, which essentially represent the maximum amount of time where the control loop remains opened. From a practical perspective, this result can be critical, especially when it comes offloading highly unstable control applications in the edge cloud. For instance, the inverted pendulum control example explored by the authors in [9] demonstrates that the control stability is compromised after the AoL surpass values of 40 ms.

We also estimate the probability density function from the empirical AoL data. We approach that with a parametric estimation, where we calculate the Residual Sum of Squares (RSS) of multiple theoretical distributions (e.g, gaussian, gamma, pareto, dweibull, beta, etc.), as well as its corresponding parameters around a level of significance of 0.01, in order to find the best fitted distribution. Essentially, we are basically evaluating the dispersion of the data points in comparison to each proposed distribution model by summing the square values of the residuals, i.e. the deviation between the predicted and empirical values. Using this methodology, we have obtained, as illustrated in Figure G.3, a three parameter Gamma distribution [10] as the best fit for the AoL data values, where a is the shape parameter, while *loc* and *scale* respectively represent the shift and scale of the Gamma distribution. An interesting conclusion about the obtained result is its direct relation with the related literature [10], where the authors analyze RTT measurements of LTE and Wi-Fi, showing that a Gamma function also represents a good fit for the RTT estimation. In the next section, we will use the obtained model from the experiments to further evaluate the AGV control performance under the proposed AoL behavior.

4 Modeling a Remotely Controlled AGV

We define a vehicle control model as illustrated in Figure G.1, where the controller is remote located through a wireless communication channel. As in [11], the vehicle states are defined by the 2D coordinates x and y, the velocity v and the heading angle orientation θ. Those states are periodically collected by the sensors and sent to the controller through the wireless channel. The controller, in turn, after receiving the AGV state information, provides a control command based on manipulating the vehicle acceleration a and its front wheel angle δ in order to manage its trajectory along a reference path. Both sensor sampling and control actuation are not affected by the wireless channel behavior. Implementation details of the control system model, design and structure can be found in [9].

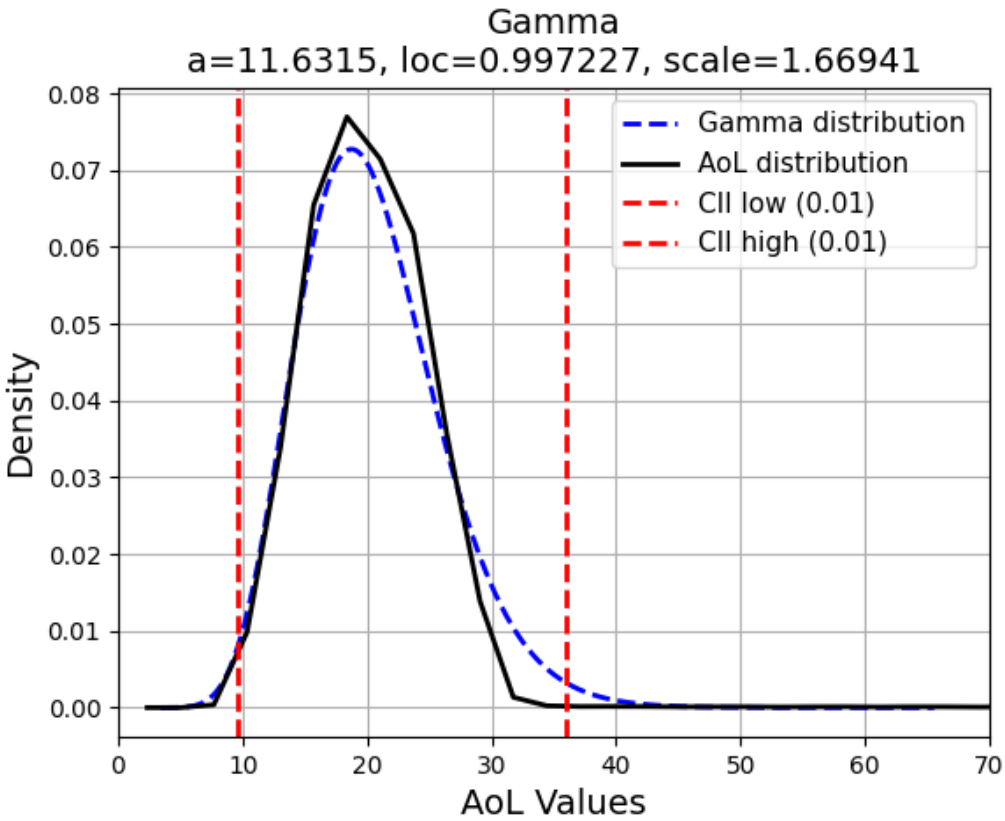


Fig. G.3: Distribution of AoL empirical data and its probability density approximation using a Gamma function.

The control system dynamics are defined according to the kinematic vehicle model [12], such that $\dot{x} = v\cos\theta$, $\dot{y} = v\sin\theta$, $\dot{v} = a$ and $\dot{\theta} = v\tan\delta/L$, for the vehicle inter-axle distance L. The notation $\dot{\square}$ corresponds to derivatives w.r.t time. To evaluate the AGV control performance, we analyze the cross track error (XTE) [12], a commonly used metric that measures the distance which the AGV has deviated from the planned path. As illustrated in Figure G.1, the XTE is given by the lateral distance between the vehicle position point and the point of the path that is closest to it. Hence, we are interested at the car position as its projection on the path. if the communication between the AGV and the controller (or vice-versa) fails to deliver the information, the controller will take decisions based on delayed sensor information in addition to incorrect actuation control commands, potentially harming the XTE performance.

To analyze this behavior, we investigate the AGV track error together with its corresponding mission time, which is the total duration taken by the AGV to complete its trajectory. The wireless channel will behave as a random process, where the time to complete the control loop is selected from the Gamma pdf in Figure G.3. We expect that the faster the AGV tries to complete the path, the higher the path deviation can be due to the increased speed and erroneous transmissions over the wireless channel. As an example, we evaluated the maximum AGV track error captured along the path for different simulation rounds, as steep path deviations are centrally critical for the AGV operation, comparing the XTE performance and its corresponding mission time over the proposed wireless channel using two different baseline controllers, the LQR [12] and Stanley [11]. The results are illustrated in Figure G.4, where we can note the 99th percentile of the XTE exponentially declining as we loose the mission time requirement.

5 Controlling the AGV speed using AoL

We can observe from Figure G.4 that for a given wireless channel (and AoL) distribution the XTE is directly impacted by the mission time requirement, and, correspondingly, to the speed of the AGV. More precisely, there is a time constraint in which the AGV must complete its path, where the controller, on the other hand, is subject to a completely unknown behavior for the uplink and downlink transmissions. Hence, the main question is how can we control the AGV speed such that, for a given mission time constraint, we can minimize its trajectory error.

By analyzing the AoL from the control perspective, we can intuitively measure how old the control loop is and use this information to accelerate or decelerate the vehicle speed accordingly. Since the dynamics are unknown to the controller, we propose a learning methodology to empirically evaluate the control decision according to the current AoL state, by modeling the problem as a Markov Decision Process [1] as:

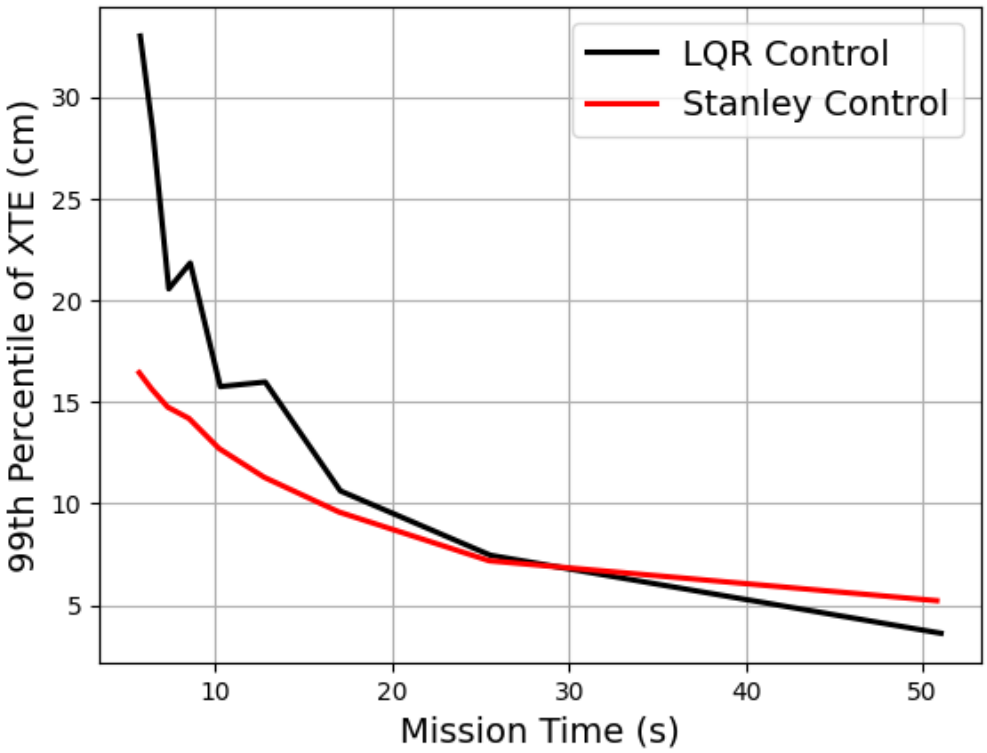


Fig. G.4: Trade-off between track-error and mission time of a wireless controlled AGV application.

State Space

The current Age of Loop value measured at the controller.

Action Space

Composed by two possible actions, increasing or decreasing the target vehicle speed by 0.5 m/s, which is taken immediately after receiving sensory data information.

Reward

For every step of the MDP while the goal is not reached the reward is 0.0. We consider, however, three special conditions: 1) if the mission time constraint $\mathcal{T}_{\max}$ is achieved (i.e., the AGV violates the time budget), we penalize with a fixed value and the episode is ended. 2) If AGV reaches the goal within the time constraint, we penalize by the maximum track error measured along the path. 3) If the AGV track error achieves a maximum predefined threshold $\epsilon_{\max}$ value along the path, the episode is ended and penalized with its corresponding value. This is a typical case in RL of delayed rewards approach, where specific events (e.g, time or XTE constraint violation) mostly determine the agent reward.

Solution

We solve the proposed MDP using Q-Learning [1]. As widely studied in current literature, the main idea is to provide a policy matrix Q indicating, for each AoL state, the decision of increasing or decreasing the vehicle speed based on the expected track-error along the path. Parameters, such as the learning rate and the discount factor, were empirically tuned during training.

5.1 Results

We analyzed the proposed solution with two arbitrary mission time requirements of 15 s and 25 s, which is the time budget for the AGV to complete the path, comparing the 99th percentile of the XTE against the controller baselines in Figure G.4. The main difference of proposed solution lies in the control decisions which are based on the current AoL status instead of the physical AGV states, as strictly assumed in the baselines. The obtained results are illustrated in Figure G.5, where we can observe that the RL solution, for the 15 s mission time requirement, provides a track-error reduction of 47% compared to the LQR and 27% compared to the Stanley controller. Likewise, for the 25 s mission time, the XTE reduction for the RL solution lies around 30.3% compared to the LQR and 27.6% compared to the Stanley controller. These results essentially show that we can derive a learning-based control policy in which exclusively

take control decisions based on AoL and can notably provide a better performance than traditional control baselines.

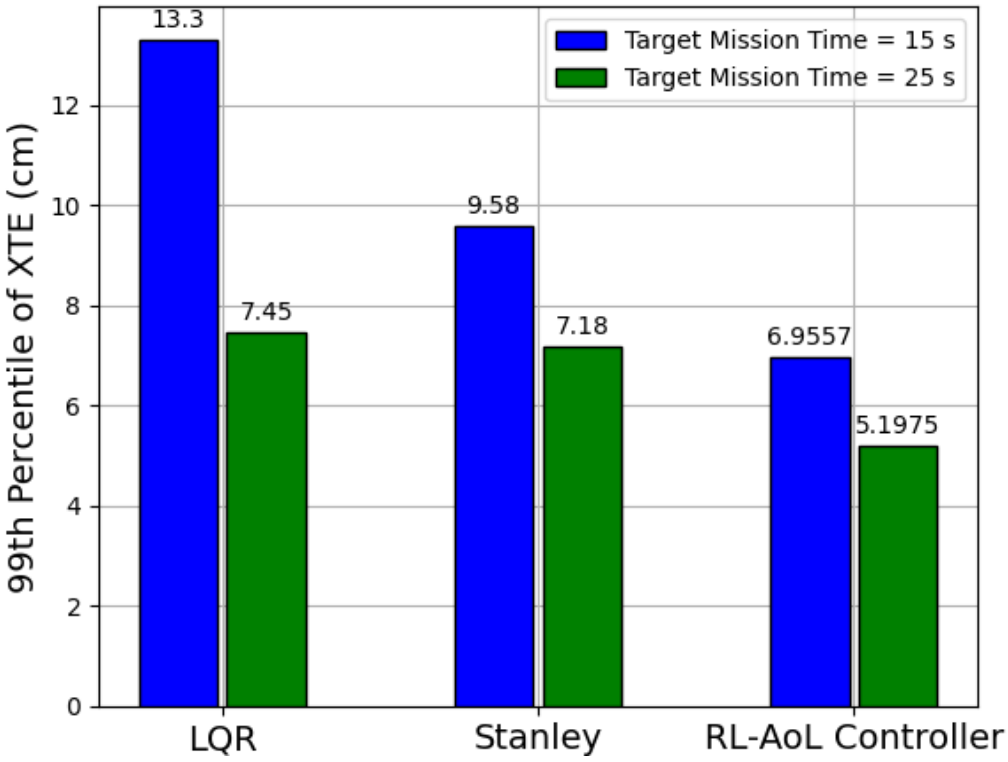


Fig. G.5: Performance comparison of controllers under different mission time target and similar communication AoL.

6 Controlling AGV bandwidth usage using AoL

As another example of AoL purpose, we show an application of network resource allocation for the same AGV model described in Section 4. To model the network behavior, we consider the work in [13], where we can observe a considerable traffic imbalance between DL and UL in WNCS problems. For example, some AGVs might be required to send image or video data, LiDAR cloud points, AR/VR information and so on, while the control packets are always based on sending simple control direction, such as acceleration, force or vehicle heading angle. For this reason, we can plausibly design a

wireless communication model where we can manage network resources at the uplink, while we do not have control over the downlink delay itself, but we might have some statistical inference about its behavior and use that information to potentially enhance the uplink resource allocation. Hence, the model details are defined as following:

- For the sensor-control (uplink) communication, we consider the 3GPP 4-bit CQI Table 7.2.3-1 [14], where, for each TTI, the amount of data delivered depends on both the current channel quality indicator (CQI) and the total bandwidth allocated to the transmission. So, the more Resource Blocks (RBs) the scheduler assign to the plant in a given TTI, faster the sensor information can be delivered. This approach has been traditionally used for benchmark purposes, including the Nokia Open source project [15].
- For the control-actuator (downlink) communication, we assume random and variable delays, defined by a Gamma distribution as defined by the literature in [10] and [13]. So, each time a new control command is generated, we randomly select a sample from the distribution to characterize the downlink latency duration.

In such scenario, by analyzing the AoL from the AGV perspective, we aim to provide the minimum amount of RBs capable to make the AGV system complete the path for a given track-error constraint. More concretely, we wish to minimize the total bandwidth utilized by the AGV considering that the track-error does not surpass a certain threshold value along the path. To achieve that, we design a RL problem where we analyze the RL agent performance not only from the control perspective by using the track-error, but also from the communication perspective by the total bandwidth consumption. So, different from the RL solution in Section 5, the agent decisions now can directly impact and control the AoL behavior. In this context, we formulate again the problem as a Markov Decision Process, defined as:

State Space

The current Age-of-Loop value measured at the plant.

Action Space

The amount of resource block utilized in each transmission. In this case, from 1 to 100, which is the maximum amount of RBs for 20 MHz total bandwidth according to the 3GPP specification in [14].

Reward

We divide the reward into two parts: First, for every transmission, we penalize the AGV proportionally to the bandwidth consumption as a function of the maximum amount

of RBs available. For example, if current transmission utilizes b from a total of $b_{\max}$ RBs, the agent receives $b/b_{\max}$. Second, we penalize the agent and finish the episode whenever the AGV surpass a given predefined track-error threshold of $\epsilon_{\max} = 10$ cm along the path.

Solution

We solve the proposed MDP again using a temporal difference learning methodology. The idea is that at the end we will have a matrix Q indicating, for each AoL state, what is the expected long-term reward in terms of track-error and bandwidth consumption that the assignment of each number of RBs can cause.

6.1 Results

We compare the proposed solution with a bandwidth allocation scheme based on predefined delay requirements, which is the general solution currently used in industry. In more details, given an arbitrary requirement of T_r ms for the sensor packet to be delivered, we can directly calculate the minimum amount of bandwidth to achieve the necessary requirement using the 3GPP 4-bit CQI Table 7.2.3-1 [14] and the total sensor data size.

We analyze the results for three common network requirements, $T_r = 1$ ms, $T_r = 5$ ms and $T_r = 10$ ms. In each case, we analyzed again the 99th percentile of the track-error, the average bandwidth usage and the average AoL observed, which are respectively illustrated in Table G.1.

Table G.1: Performance comparison of different networking solutions.

	Avg Bandwidth Usage	**99th-Track-error**	**Avg AoL**
$T_r =$ **1 ms**	46.70 MHz	4.41 cm	11.02 ms
$T_r =$ **5 ms**	9.34 MHz	8.23 cm	16.10 ms
$T_r =$ **10 ms**	4.67 MHz	12.15 cm	23.47 ms
RL Solution	4.72 MHz	9.97 cm	20.61 ms

The immediate conclusion we can verify from Table G.1 is that the RL scheme was capable to learn the system track-error requirement ($\epsilon_{\max} = 10$ cm), providing a worse case performance exactly on the limit of the track error budget, while simultaneously providing the lowest bandwidth consumption. Comparing extreme cases, at the 1 ms requirement, we have the lowest track error result, which, however, comes at the cost of 10 times higher bandwidth consumption. At 10 ms requirement, we have similar bandwidth consumption, but a track-error 20% above the limit. In conclusion, the RL

scheme learned a bandwidth allocation policy that is focused on the physical process of the control system, exploiting the application requirements to save more network resources. It is interesting to highlight that, recently, 3GPP has defined in its Release 17 the concept of survival time as part of the service requirements for industrial 5G systems, which represents a maximum duration of the communication link outage which a given application can handle without any failure. In this context, we can infer two feasible interpretations from the proposed RL scheme: 1) We can essentially learn the survival time of the application. 2) We can understand the dynamics that leads to the application survivability. More concretely, when the behavior of the control dynamics is clearly defined, such as the inverted pendulum in [1], we can essentially learn granularity levels of AoL states that can cause the system to lose stability. In this case, the AoL value itself becomes a representation of the survival time concept established by 3GPP. However, with the RL scheme we can actually go beyond such strict requirement definition. The key factor for the decision making is not the current age itself, but a sequence of steps that lead to this current age. That is exactly the reason why the RL scheme and the fixed 10 ms timing requirement, besides having the same bandwidth consumption, provided different results. The decisions learned by the policy are not strictly based on the current AoL itself, but the sequence of steps (state-transitions) that lead to this AoL value. Such concept goes beyond the traditional timing requirements definition we usually find in the standards, where the main question for the network decision-maker is not only about the immediate information state, but actually the dynamics of the information.

7 Conclusions

In this work, we elaborated on Age of Loop, a recent proposed metric that targets wireless control systems problems by jointly accounting effects of donwlink and uplink data. We have set up a measurement campaign using a 5G-SA network to empirically evaluate the AoL behavior within a factory environment. Using the obtained measures, we have developed a model to analyze and optimize the trajectory of an AGV using AoL, where two solutions were proposed: 1) an AGV speed controller, and 2) a network radio resource (bandwidth) scheduler. The obtained results provide an indication that using AoL we can approach goal-oriented solutions that can be managed by either the application or the network.

References

[1] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, “Age of loop for wireless networked control systems optimization,” in *IEEE PIMRC*, 2021.

[2] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, 2021.

[3] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using AoI in a single-loop networked control system," in *IEEE Conference on Computer Communications Workshops*, 2019.

[4] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "AoI-penalty minimization for networked control systems with packet loss," in *IEEE Conference on Computer Communications Workshops*, 2019.

[5] K. Gatsis, H. Hassani, and G. J. Pappas, "Latency-reliability tradeoffs for state estimation," *IEEE Transactions on Automatic Control*, 2020.

[6] Huang, Kang *et al.* , "Wireless feedback control with variable packet length for industrial IoT," *IEEE Wireless Communications Letters*, 2020.

[7] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the latency, rate and reliability tradeoff in wireless networked control systems for IIoT," *IEEE IoT Journal*, 2020.

[8] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, 2019.

[9] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Wireless control of autonomous guided vehicle using reinforcement learning," in *IEEE GLOBECOM*, 2020.

[10] S. Yasuda and H. Yoshida, "Prediction of round trip delay for wireless networks by a two-state model," in *IEEE WCNC*, 2018.

[11] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt *et al.*, "Towards fully autonomous driving: Systems and algorithms," in *IEEE Intelligent Vehicles Symposium*, 2011.

[12] T. C. Martin, M. E. Orchard, and P. V. Sánchez, "Design and simulation of control strategies for trajectory tracking in an autonomous ground vehicle," *IFAC Journal of Systems and Control*, 2013.

[13] P. M. de Sant Ana, N. Marchenko, P. Popovski, and B. Soret, "Control-aware scheduling optimization of industrial iot," in *IEEE VTC*. IEEE, 2022.

[14] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 10 2014, version 12.3.0.

[15] P. M. de Sant Ana and N. Marchenko, “Radio Access Scheduling using CMA-ES for Optimized QoS in Wireless Networks,” in *IEEE Globecom Workshops*, 2020.

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-674-4

AALBORG UNIVERSITY PRESS