On Age-of-Information Aware Resource Allocation for Industrial Control-Communication-Codesign

Scheuvens

On Age-of-Information Aware Resource Allocation for Industrial Control-Communication-Codesign

Dipl.-Ing.
Lucas Scheuvens

der Fakultät Elektrotechnik und Informationstechnik der
Technischen Universität Dresden

zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

– genehmigte Dissertation –

Vorsitzender	Prof. Dr. Stefan Mannsfeld
Gutachter	Prof. Dr.-Ing. Dr. h.c. Gerhard Fettweis
	Prof. Dr. Beatriz Soret
Tag der Einreichung	1. März 2022
Tag der Verteidigung	3. Mai 2022

Lucas Scheuvens
On Age-of-Information Aware Resource Allocation
for Industrial Control-Communication-Codesign
Dissertation

Technische Universität Dresden
Vodafone Chair Mobile Communications Systems
Institute of Communication Technology
Faculty of Electrical and Computer Engineering
01062 Dresden, Germany

Abstract

In industrial manufacturing, *Industry 4.0* refers to the ongoing convergence of the real and virtual worlds, enabled through intelligently interconnecting industrial machines and processes through information and communications technology. Ultra-reliable low-latency communication (URLLC) is widely regarded as *the* enabling technology for *Industry 4.0* due to its ability to fulfill highest quality-of-service (QoS) comparable to those of industrial wireline connections. In contrast to this trend, a range of works in the research domain of *networked control systems* have shown that URLLC's supreme QoS is not necessarily required to achieve high quality-of-control; the co-design of control and communication enables to jointly optimize and balance both quality-of-control parameters and network parameters through blurring the boundary between application and network layer. However, through the tight interlacing, this approach requires a fundamental (joint) redesign of both control systems and communication networks and may therefore not lead to short-term widespread adoption. Therefore, this thesis instead embraces a novel co-design approach which keeps both domains distinct but leverages the combination of control and communications by yet exploiting the age of information (AoI) as a valuable interface metric.

This thesis contributes to quantifying application dependability as a consequence of exceeding a given peak AoI with the particular focus on packet losses. The beneficial influence of negative temporal packet loss correlation on control performance is demonstrated by means of the automated guided vehicle use case. Assuming small-scale fading as the dominant cause of communication failure, a series of communication failures are mapped to an application failure through discrete-time Markov models for single-hop (e.g, only uplink or downlink) and dual-hop (e.g., subsequent uplink and downlink) architectures. This enables the derivation of application-related dependability metrics such as the *mean time to failure* in closed form. For single-hop networks, an AoI-aware resource allocation strategy termed *state-aware resource allocation* (SARA) is proposed that increases the application reliability by orders of magnitude compared to static multi-connectivity while keeping the resource consumption in the range of best-effort single-connectivity. This dependability can also be statistically guaranteed on a system level – where multiple agents compete for a limited number of resources – if the provided amount of resources per agent is increased by approximately 10 %. For the dual-hop scenario, an AoI-aware resource allocation optimization is developed that minimizes a user-

defined penalty function that punishes low application reliability, high AoI, and high average resource consumption. This optimization may be carried out offline and each resulting optimal SARA scheme may be implemented as a look-up table in the lower medium access control layer of future wireless industrial networks.

Kurzfassung

Unter dem Überbegriff *Industrie 4.0* wird in der industriellen Fertigung die zunehmende Digitalisierung und Vernetzung von industriellen Maschinen und Prozessen zusammengefasst. Die drahtlose, hoch-zuverlässige, niedrig-latente Kommunikation (engl. *ultra-reliable low-latency communication*, URLLC) – als Bestandteil von 5G – gewährleistet höchste Dienstgüten, die mit industriellen drahtgebundenen Technologien vergleichbar sind und wird deshalb als Wegbereiter von *Industrie 4.0* gesehen. Entgegen diesem Trend haben eine Reihe von Arbeiten im Forschungsbereich der vernetzten Regelungssysteme (engl. *networked control systems*, NCS) gezeigt, dass die hohen Dienstgüten von URLLC nicht notwendigerweise erforderlich sind, um eine hohe Regelgüte zu erzielen. Das Co-Design von Kommunikation und Regelung ermöglicht eine gemeinsame Optimierung von Regelgüte und Netzwerkparametern durch die Aufweichung der Grenze zwischen Netzwerk- und Applikationsschicht. Durch diese Verschränkung wird jedoch eine fundamentale (gemeinsame) Neuentwicklung von Regelungssystemen und Kommunikationsnetzen nötig, was ein Hindernis für die Verbreitung dieses Ansatzes darstellt. Stattdessen bedient sich diese Dissertation einem Co-Design-Ansatz, der beide Domänen weiterhin eindeutig voneinander abgrenzt, aber das Informationsalter (engl. *age of information*, AoI) als bedeutenden Schnittstellenparameter ausnutzt.

Diese Dissertation trägt dazu bei, die Echtzeitanwendungszuverlässigkeit als Folge der Überschreitung eines vorgegebenen Informationsalterschwellenwerts zu quantifizieren und fokussiert sich dabei auf den Paketverlust als Ursache. Anhand der Beispielanwendung eines fahrerlosen Transportsystems wird gezeigt, dass die zeitlich negative Korrelation von Paketfehlern, die in heutigen Systemen keine Rolle spielt, für Echtzeitanwendungen äußerst vorteilhaft ist. Mit der Annahme von schnellem Schwund als dominanter Fehlerursache auf der Luftschnittstelle werden durch zeitdiskrete Markovmodelle, die für die zwei Netzwerkarchitekturen *Single-Hop* und *Dual-Hop* präsentiert werden, Kommunikationsfehlerfolgen auf einen Applikationsfehler abgebildet. Diese Modellierung ermöglicht die analytische Ableitung von anwendungsbezogenen Zuverlässigkeitsmetriken wie die durschnittliche Dauer bis zu einem Fehler (engl. *mean time to failure*). Für Single-Hop-Netze wird das neuartige Ressourcenallokationsschema *State-Aware Resource Allocation* (SARA) entwickelt, das auf dem Informationsalter beruht und die Anwendungszuverlässigkeit im Vergleich zu statischer Multi-Konnektivität um Größenordnungen erhöht, während der Ressourcenverbrauch im Bereich von konventioneller Einzelkonnektivität bleibt.

Diese Zuverlässigkeit kann auch innerhalb eines Systems von Regelanwendungen, in welchem mehrere Agenten um eine begrenzte Anzahl Ressourcen konkurrieren, statistisch garantiert werden, wenn die Anzahl der verfügbaren Ressourcen pro Agent um ca. 10 % erhöht werden. Für das Dual-Hop Szenario wird darüberhinaus ein Optimierungsverfahren vorgestellt, das eine benutzerdefinierte Kostenfunktion minimiert, die niedrige Anwendungszuverlässigkeit, hohes Informationsalter und hohen durchschnittlichen Ressourcenverbrauch bestraft und so das benutzerdefinierte optimale SARA-Schema ableitet. Diese Optimierung kann offline durchgeführt und als Look-Up-Table in der unteren Medienzugriffsschicht zukünftiger industrieller Drahtlosnetze implementiert werden.

Acknowledgement

This thesis would not have been possible without many people and institutions.

First, I whole-heartedly thank may doctoral advisor Gerhard Fettweis for giving me the opportunity to pursue my PhD at the Vodafone Chair Mobile Communications Systems at the Technische Universität Dresden. You inspired me in your enthusiam, optimism, creativity, and willingness to drive things forward to generate impact. During the time at the chair, you placed trust in me and gave me the chance to prove myself with responsibilities that I had never taken before. Second, I express my gratitude to Beatriz Soret for her unhesitated willingness to act as the second reviewer.

I deeply thank my beloved wife Lea for the ever-lasting support, especially by taking over many family duties in particularly challenging and stressful times. This thesis would have been signficantly more challenging without you.

I consider myself lucky that the German Federal Ministry of Education and Research as well as the European Commission put me in the comfortable position to pursue this work on a full salary. I do not take this for granted and I am highly grateful for it. I thank the Technische Universität Dresden for supplying me with an environment that supports research excellence.

My direct supervisors have changed continuously during my thesis and I owe thanks to all. I especially thank André Noll-Barreto who continously encouraged me and greatly improved my work through his scientific rigour. In chronological order, I additionally thank Meryem, Henrik, David, and Arturo for offering me guidance when I needed it.

My time at the chair has been an exciting adventure. I generated many wonderful memories with the exceptional people I met there. In a non-exhaustive list, I would like to mention Andreas, Behnam, David, Eva, Friedrich, Henrik, Nick, Norman, Philipp, Robert, and Tom. We had a great time with lots of fun and amazing achievements, I am thankful to have been part of the team. I want to express my appreciation for Eva, Carolin, Kathrin, Raffael, Rüdiger, Sylvia, Steffen, and Yaning for keeping the chair running by helping us in all organizational matters.

Contents

1 Introduction

Wireless communication is ubiquitous in today's world. Mobile internet connectivity has increased tremendously in recent years, increasing from connecting 32 % of all people worldwide in 2014 to 49 % in 2019 [GSM20], with a projected penetration of 70 % by 2023 [Cis20]. The societal impact is immense and has manifested clearly during the Covid-19 pandemic [KP20]. Today, most connections still involve at least one human but this will change soon. Machine-to-machine-type communications (M2M) is the by far fastest growing connection type with a compound annual growth rate of 30 %, putting the second-place smartphone sector (7 %) in the shade. It is projected that by 2023 50 % of all wireless connections will not involve humans [Cis20].

Expecting this trend, it comes to no surprise that the integration of M2M communications to wireless networks has been a significant part in the standardization of the fifth generation of cellular networks (5G). Therein, three main pillars are defined [ITU15; Pop+18b].

1. Enhanced mobile broadband
2. Massive machine-type communications (mMTC)/Internet of things (IoT)
3. Ultra-reliable low-latency communications (URLLC)/Mission-critical machine-type communications (μMTC)

While the first is human-centric and mainly evolves around increasing data rates especially for augmented reality and virtual reality applications, the latter two are M2M. The domain of mMTC/IoT is expected to support a vast amount of wireless network devices, e.g., wireless sensors and actuators for smart city, smart logistics, and home applications [Sha+15], usually imposing loose latency requirements and are therefore non-mission-critical. On the other hand, *Tactile Internet* [Fet14; Sim+16; Sch+17] applications such as wireless factory automation, self-driving cars, and real-time remote control, are mission-critical and demand highly reliable connectivity at a minimal latency (URLLC).

In the context of wireless communications, the *reliability* of a connection is defined as the complement of the packet loss ratio (PLR), i.e., $1 - \text{PLR}$, whereby a packet is

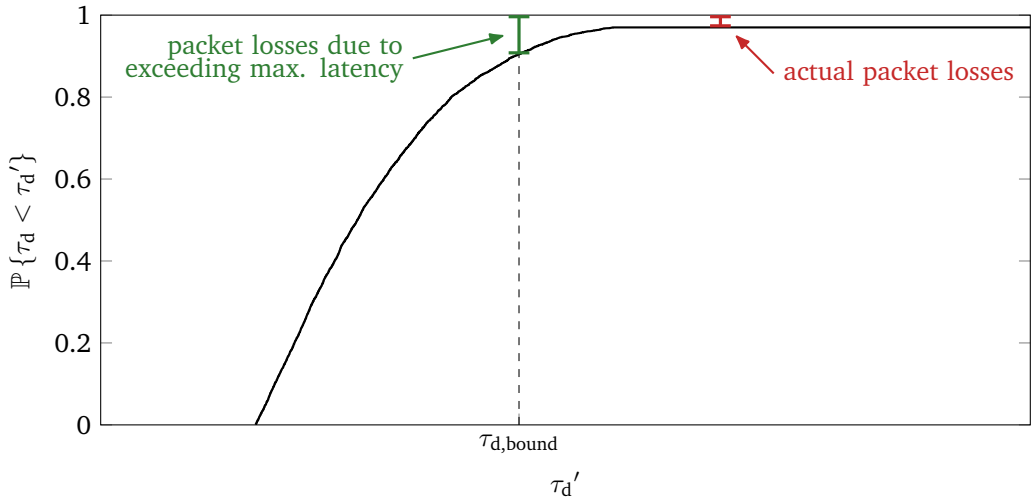


Fig. 1.1.: Latency and packet losses are related quantities (symbolic plot). Packets are also considered *lost* if they arrive after a given threshold $\tau_{d,bound}$ [Pop+18a].

considered lost when it is not correctly received within the time constraint required by the targeted service [Pop+18a; 3GP19b]. This definition relates the PLR to a latency requirement, compare Fig. 1.1.

Reliability may be generally enhanced through *diversity*. If the latency requirement is relaxed, ultra-reliability can be achieved through *time diversity* in the form of (infinitely many) retransmissions of erroneous data. This has been state-of-the-art in wireless standards for decades and is still valid for current wireless local area network (WLAN) and long term evolution (LTE) systems with their respective automatic repeat request (ARQ)/hybrid ARQ (HARQ) mechanisms. However, if the required latency decreases to a minimum, the time budget for a successful transmission is exhausted quickly, narrowing the possibility of retransmissions. For extremely low latencies in the range of <1 ms, only a single (initial) transmission may be viable. In this scenario, a reliability increase may only be achieved through concurrent diversity means, e.g., *frequency diversity* (in the form of multiple simultaneous transmissions in different frequency bands) or *spatial diversity* (in the form of multiple antennas). Concurrent diversity techniques fundamentally differ from time diversity in terms of cost. As the success of a specific wireless transmission cannot be predicted beyond long-term average metrics, wireless resources must be allocated preemptively (frequency diversity) or more antennas need to be installed (spatial diversity), which increases capital and operational expenditure. The authors of [ÖF15] have found that in realistic fading environments, an outage probability <0.001 % in a 1 ms grid may require more than 10 parallel links, which (a) is unrealistic from a hardware perspective and (b) severely challenges the scalability potential of such systems.

While the extreme quality of service (QoS) requirements of URLLC state a packet success probability of >99.999 % at a latency bound $\tau_{\mathrm{d,bound}} = 1\,\mathrm{ms}$, first 6G white papers envision “enhanced” [LL19] or “extreme” [NTT21] URLLC, which will signify the resource consumption problem further in future wireless networks.

1.1 The Need for an Industrial Solution

Although URLLC is also relevant for many other application domains, the subdomain of industrial manufacturing is of particular importance in Germany, playing a central strategic role for economic growth and development. The German federal government coined the term *Industry 4.0*, following the nomenclature of software versioning indicating a major upgrade. It describes a shift towards *smart manufacturing* that is broadly characterized by the following points. [Aca13]

1. Individual customer requirements can be met and even one-off items can be manufactured profitably.
2. Agile manufacturing processes pose resilience against unforeseen disruptions and maximize efficiency.
3. Value opportunities through new services are created, e.g., smart algorithms that leverage the resulting (big) data from smart sensors.

Technologically, this shift requires – among other things – an information technology system capable of exchanging data vertically (between manufacturing systems and business processes) and horizontally (within a layer) in the International Society of Automation 95 model (*automation pyramid*), see Fig. 1.2 [Roj17]. Within this model, the width of each layer typically describes the level of heterogeneity and data transmission becomes increasingly mission critical with lower layers.

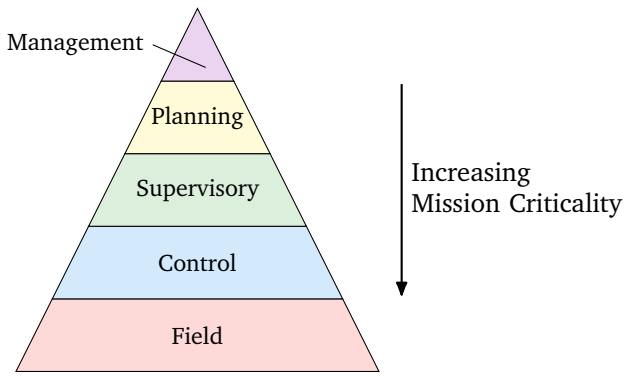


Fig. 1.2.: Automation pyramid according to International Society of Automation 95 model.

5G is often viewed as accelerator (or sometimes even enabler) for *Industry 4.0* as it is the first (wireless) technology to support a range of different QoS classes [Gun+18]. When connecting and operating applications on the factory floor (*field level*), tight requirements are formulated to avoid costly communication-induced production downtime. The motivation is straight-foward: no errors in the communication domain guarantee no communication-induced application failures, as is the case with prevalent production-grade wire-line communication such as Ethercat, SERCOS III, and Profinet IRT. However, the 1:1 replacement of wire-line technologies with wireless counterparts is infeasible due to their fundamentally different operating principles. Multiple key differences stand out:

1. For low-latency applications, the air interface is orders of magnitude less dependable due to path loss, shading, small-scale fading, and interference. Diversity may be used to increase reliability, however, at the cost of increased resource consumption (as discussed above).

2. The air interface is a shared medium while cables are not. An increased number of terminals introduces competition for resources, which not only reduces the individual transmission capacity but also increases the complexity of dynamic scheduling.

3. Commonly used wire-line ring topologies enable to "piggyback" data onto other packets, which reduces communication overhead by sharing header information. With wireless networks, header information must be transmitted separately for every data paket, which further accentuates the capacity issue in wireless networks.

These points indicate that wireless resources are extremely valuable and, therefore, spectral efficiency is crucial to enable real-time applications also in dense scenarios (*scalability*). The conventional URLLC approach does not address this necessity and, hence, might not be ideal to realize the *Industry 4.0* vision.

The need for *dynamic and dependable* industrial applications drives the need for *low-latency and dependable* wireless communication. This thesis contributes to paving the way towards the *Industry 4.0* vision without compromising on spectral efficiency and thereby enabling an increased application density.

1.2 Contributions

This thesis contributes to realizing dynamic, ultra-dependable, and scalable real-time applications over spectrally efficient wireless networks by fundamentally challenging the design target of low-level industrial networks. The age of information (AoI) as a time-based metric is embraced as primary indicator for dependability on an application level, rather than optimizing for long-term average metrics such as the PLR. In detail, the contributions are:

- Chapter 2 provides the foundation of the thesis by presenting the state of the art in the research fields of URLLC, networked control system (NCS), and AoI.
- In Chapter 3, the impact of packet losses on control applications is studied. First, a high-level introduction to linear, time-invariant (LTI) control theory is given. Second, two joint descriptions of control loops with packet losses are discussed to derive actual QoS requirements of closed-loop control applications. The first enables to derive conservative bounds within the LTI domain and therefore provides an easily accessible entry point for such an assessment. The second more advanced approach utilizes markov jump linear system (MJLS) theory to show that limiting the packet loss sequence length has a highly beneficial effect on control utility. The packet loss process is modeled as a discrete-time Markov chain and a coefficient is introduced that describes the temporal correlation between packet losses. The impact of positive and negative correlation on the control loop is studied by means of an example industrial automated guided vehicle (AGV) use case. Lastly, the theoretical results are validated through extensive simulation.
- Chapter 4 presents failure models of real-time applications due to communication errors. First, the (industrial) communication assumptions are presented and causes of failure are discussed. Two network architecture are introduced (single-hop and dual-hop) that build the foundation for all investigations and optimizations of this thesis. Markov chain failure models are developed for both network architectures that link the event of exceeding a certain AoI to an application failure. In the single-hop case, this translates to losing more than a given number of consecutive packets. Multiple key performance indicators (KPIs) are introduced to describe the gains in terms of network performance, resource consumption, and application dependability.
- Chapter 5 exploits the findings of Chapter 3 for an error-prone single-hop wireless network connection. Through the dynamic assignment of resources,

the network can enforce a negative temporal packet loss correlation, limiting the sequence length of consecutive packet losses. A range of different dynamic resource allocation schemes is presented, termed state-aware resource allocation (SARA) and the advantages of SARA with respect to application dependability and network resource consumption are demonstrated. Lastly, the impact of erroneous acknowledgments (ACKs) is investigated, which are otherwise considered ideal.

- In Chapter 6, the framework of Chapter 5 is extended to the multi-agent case, in which a network-wide resource pool must be shared. As system resources might be limited, individual agents may not receive the resources they require and may fail sooner than predicted in the single-agent case. The modeling is performed through a holistic discrete-time Markov chain approach, which includes all individual-agent state transitions. The computational complexity of the approach is investigated and contrasted with required extensive simulation efforts. The single-agent performance metrics of Chapter 4 are lifted to system-level performance metrics, in which the system is considered to fail as soon as a single agent fails. For the case of too few resources in the system, the strategies "random" and "cliff" are investigated, which differ in their prioritization of users. Also, the assignment of spare system resources is investigated. For better tractability, a low-complexity example is presented. The applicability of SARA to a system of agents is discussed thoroughly with a key focus on the required number of system resources that allows for unimpaired KPIs for the individual agent.

- Chapter 7 considers the dual-hop network architecture. An optimization approach is presented that derives an optimal SARA scheme depending on user- and application-specific penalty functions regarding the application dependability, the average resource consumption, and the AoI. Through multiple exemplary sets of penalty functions the dependence of the optimal SARA schemes on the input penalty functions is demonstrated. The interdependence of uplink (UL) and downlink (DL) stemming from a joint AoI optimization is studied as well as the impact of the individual link's packet loss probability.

- Finally, Chapter 8 concludes this thesis and highlights open questions for further research.

2 Related Work

This thesis contributes to bridging the gap between the two major engineering fields of communications and control in industrial environments. The broad vision of *Industry 4.0* is an all-connected factory of the future where (control) applications seamlessly interact with each other in a highly dependable fashion, i.e., with virtually zero failures. This chapter highlights state-of-the-art approaches to achieve this vision and identifies shortcomings that might hamper wide-spread adaption.

2.1 Communications

The vision to support real-time applications over wireless links has been pursued for many decades. In fact, many wireless standards exist that provide some degree of QoS to specifically serve such applications. Since the late 1990s, the research area of wireless sensor network has emerged and some customized technologies have been standardized, e.g., 802.15.4 upon which WirelessHART was built [Che+14]. These types of networks are optimized for low power consumption, low PLR, robustness against interference, and cheap/easy deployment. However, the real-time capabilities are fairly limited because (a) 802.15.4 radios have a maximum throughput of 250 kbit/s [Lu+16], and (b) the latency at which the reliability is ensured is typically in the order of hundreds of milliseconds to a few seconds [Chu+16; Li+17]. While this constitutes "real-time" for control systems in some industries, e.g., the process industry, it does not for more dynamic, higher-bandwidth control systems such as those requiring motion control. Apart from 802.15.4, other industy-focused networks evolve around the IEEE 802.11 family. For example, the commercially available iWLAN (Siemens) provides multiple so-called iFeatures that optimize QoS. Cyclic communication may be carried out through access point administrated polling (iPCF), data redundancy for increased reliability is supported (iPRP), and roaming times are decreased substantially compared to the standard 802.11g/n/ac/ax WLAN. Another example for enhanced QoS over the 802.11 protocol family is 802.11e, which was standardized in 2007. Within this standard, transmit opportunities may be granted to clients (once and also periodically), within which the radio channel can be used exclusively. For a more detailed overview on current (wireless) network

designs for (industrial) control systems, the interested reader may be referred to [Wil08] and [Par+18].

There are a range of shortcomings of these technologies that hamper wide-spread integration of wireless communications to industrial processes, many of which relate to *dependability* [5G 20]. First, the shared (and free-to-use) 2.4 GHz and 5 GHz industrial, scientific and medical spectrum is considered a major source of unreliability due to interference [5G 20; 5G 18], stemming from clients of the same radio access technology or different technologies in the same spectrum (Bluetooth, ZigBee, WLAN, WirelessHART, ...). To alleviate the interference issue, the German Federal Network Agency freed up spectrum for local usage in industrial *campus networks* in the range 3.7 GHz to 3.8 GHz (100 MHz) in 2019, with the special intention to provide industrial manufacturing with own dedicated bandwidth. Second, spectrum access must be deterministic and managed to regulate network usage and provide guaranteed transmit opportunities to clients; the QoS of a client must not decrease if other clients join the network, as is typically the case for carrier sense multiple access / collision avoidance (CSMA/CA)-based network access [LH90]. Third, time-synchronicity is regarded as crucial [3GP19a; 5G 20] as this is widely deployed in existing industrial communication networks.

URLLC, one of the main 5G pillars, addresses this need for dependability. In the domain of wireless factory automation, the requirements of URLLC were formulated to compete with well-established cable solutions (Ethercat, SERCOS III, and Profinet IRT) because the drop-in replacement of industrial wire-line communications solutions was seen as a major enabler [AMA19; Wei15; Pop+19]. Hence, PLR requirements in the range 10^{-9} to 10^{-5} whithin a deterministic latency bound of 250 μs to a couple of milli-seconds [Sch+17; Aij+17] and one-way payload data rates exceeding 90 Mbit/s [Bec18] were formulated. Consequently, great efforts were invested to achieve these QoS values. Two of the first publications to highlight the paradigm shift towards URLLC were [Fet14; Pop14]. To achieve the low latency, a reduction of the transmission time interval was proposed [Läh+14; Tul+14] and eventually also realized in 5G standardization [3GP21b]. M2M communications is often characterized by short packets, for which classical information theory yields unaccurate results. An overview for fundamental information theoretical results in the short-packet regime is given in [DKP16]. Subsequently, enhanced channel coding techniques for URLLC are discussed in [Syb+16; SS17]. Another fundamental issue is the wireless channel access, especially in the UL. Semi-persistent scheduling (SPS) is proposed for periodically operating devices for an effective reduction of control overhead, reduction of latency (due to instant access), and greater reliability since the control channel is eliminated as error source [3GP16]. Since the low-latency

constraint often entails that retransmissions such as ARQ and HARQ, which are widely established measures in broadband systems to ensure reliable data transmission, might not be applicable [Sac+18; AKC18], concurrent diversity techniques are proposed, e.g., interface diversity [NP16; NLP17], multi-antenna (spatial) diversity [Pop+18a], and frequency diversity [Wol+19; ÖF15]. The increase of signal-to-noise ratio (SNR) is obviously also a measure to increase reliability, as proposed through base-station densification in [Pop+18a]. On a system level, works have also investigated traffic multiplexing with other traffic classes [Gan+16], inaccurate link adaptation as a consequence of rapidly-varying intereference conditions [Poc+17], the spectral efficiency vs. network load trade-off [Poc+16], and optimized resource allocation [Sha+16]. Measures to support interruption-free mobility were investigated in [Sim+17; Tes+16a; Tes+16b]. Overviews on many of these topics are given in [Wei+14; Pop+18a; Li+17].

Although all mentioned works provide a valuable contribution to the development of future networks, the fundamental issue with URLLC is *scalability*. It is well-known from broadband communications that the level of redundancy used in a wireless transmission heavily impacts the spectral efficiency of the system [PDH15]. On the one extreme, a vast amount of redundancy, e.g., through excessive channel coding and/or low-order modulation, increases the probability of successful decoding, however, the spectral efficiency is low because (a) more channel bits are transmitted (coding) and/or (b) fewer channel bits are transmitted per unit time (low-order modulation). On the other extreme, too little redundancy decreases the probability of successfully decoding a packet, also reducing the spectral efficiency. The term *link adapation* refers to the process of optimizing the wireless link for maximum spectral efficiency and as a typical value, LTE systems target a block error rate in the order of 10 % [Eur17] through adaptive modulation and coding. In contrast, the high packet success probability requirement of URLLC requires a shift of the spectrally efficient operating point to a spectrally less efficient point. This might become the limiting factor in dense (industrial) scenarios and the aim of this thesis is to contribute to the vision of supporting highly dependable applications over spectrally efficient, i.e., *undependable* wireless links.

2.2 Control

The research domain of NCS aims at enabling closed-loop control over non-ideal communication channels and therefore views the mission criticality from a control

systems viewpoint. NCS research was initially driven by the idea to provide distributed and dependable control systems over low-cost, error-prone networks such as switched Ethernet [ZBP01] or WLAN [LH90]. These communication systems, which are mostly unable to provide QoS guarantees, are typically viewed as given [Liu+20] and communication impairments should be overcome through sophisticated, robust control algorithms. The considered impairments include but are not limited to delay [ZBP01], low SNR [BMF07], competition for resources [ZGK13], coarse quantization [EM01], and data rate limitations [Nai+07]. Extensive work on the trade-off between latency, reliability, and data rate and optimal operating points for NCS have been most recently performed in [Liu+20; GHP18; Hua+20; Hua+19; Pez+20]. Surveys on results in this domain were conducted in [HNX07; ZGK13; ZHY16].

It is important to understand that digital control systems are typically highly oversampled [FWP97] by a factor of 10-30 to linearize the control system response and ensure reasonable smoothness of operation. This makes them capable of tolerating a limited number of packet losses without significant performance degradation. A range of different works exist that quantify the packet loss tolerance for closed-loop control [Hee+10; Yu+04; YQC04; GC08], depending on the bandwidth of the application. Without taking the probability of packet losses into account, the works guarantee control system stability as long as a certain number of consecutive packet losses K is not exceeded. This number K depends upon the sampling period, the rate of successful transmissions and the network-induced delay. Other works also take the probability of packet losses into account, e.g., [ZBP01; Sin+04; EE11], and derive upper bounds on the packet loss tolerance depending on the transmission success probability. Furthermore, the authors in [WL11] study the effects of packet losses and latency in an *event-triggered* system of agents and derive the maximum allowable number of successive packet dropouts that ensures the asymptotic stability of the overall system, assuming exclusive channel access.

Most of the mentioned related work in this domain have a clear control theoretical focus. They derive mathematical bounds that explore theoretical operational boundaries and are thus difficult to apply in practical industrial setups that require a high level of flexibility and abstraction. As the main take-away message, it is important to understand that a basic tolerance of control systems against a few packet losses may also be shown by applying rather basic control engineering principles. These (conservative) bounds may be increased by the sophisticated control methods of NCS literature. This observation has also been recognized by industry-influenced organizations [5G 20; 3GP19a], which specifically state that – depending on the use case – a certain (small) number of consecutive packet losses may be tolerated.

As this thesis targets mainly readers from the communications domains, Chapter 3 is dedicated to providing high-level insights on how control systems are typically capable of tolerating a few consecutive packet losses and how bounds may be derived in a straight-forward fashion.

2.3 Codesign

With the *Industry 4.0* vision in mind to support a large number of control systems that (must) compete for limited network resources, the codesign of communication and control is as a natural next step: within a codesign framework, agents and network resources may be managed and optimized as a whole to co-optimize control performance and network usage. A considerable number of prior publications exist in this cross-layer research field. Most target the development of an optimal scheduler, i.e., deal with the question of who to optimally schedule when. Since almost exclusively single-queue time-division-multiple-access-based networks are considered, it was found that, intuitively, agents that feature the largest control state discrepancy should be scheduled first to maximize control utility [Aya+19; Wu+14; Don+11; CH08; Mam+14; Han+17; Eis+19]. This approach relies on (a) enough control awareness that the control sytem's *necessity* for a packet transmission can be estimated within the network and (b) that subsequently certain transmissions may be prioritized over others. The authors in [Kos+17] termed the potential impact of a packet's successful transmission on the reduction of the control state discrepancy as the *value of information*. The authors in [Gat+15; LCY20] take this even one step further by also incorporating current channel state information to the scheduling and power allocation for a truly holistic codesign.

2.3.1 The Need for Abstraction – Age of Information

The open systems interconnection (OSI) reference model – first established in 1983 by the International Telecommunication Union – characterizes a set of network functions without regard to their specific implementation. The model consists of seven abstraction layers, each of which serves a specific network purpose and provides a level of abstraction for the layer above. The model has been a huge success story and virtually all modern networks are inspired and interconnected through this approach. Through the abstraction layers, the problem is broken up into "manageable pieces" [DZ83], which supports flexibility and interoperability.

Similar to the OSI approach, industrial networks also require the layered approach and abstraction is needed to support a wide range of different applications over the same network interface. Although the results of value-of-information-related research demonstrate how performant an optimal codesign can be, they are not yet ready to be applied in an industrial context. The deep integration of the control application with low-level network functions such as the medium access control (MAC) and physical layer (PHY) layers is seen as a major obstacle and will henceforth be referred to as *deep codesign*. As an example, it is questionable whether application-level control state estimations may be interpreted and incorporated by the MAC for an optimal scheduling decision. In an attempt to avoid the deep codesign, some authors have decided to estimate the value of information through the AoI instead, which consitutes a *network-level* KPI and could be used as an interface metric between application and network layers. In today's networks, the AoI is completely unregarded.

By definition, the AoI is the time that has elapsed since the generation of the freshest information available at the receiver, monitoring a remote process at the transmitter, and indicates the *freshness* of information available at the receiver. Contrary to the value of information, the AoI can be monitored by the network alone and does therefore not require knowledge of the control system dynamics. Through AoI, the value of a packet can be approximated one-dimensionally and *control-unaware* network optimization can be performed that still provides a significant benefit for real-time applications. Although the AoI still constitutes a codesign approach, it keeps control (application) and communications (network) fields distinct, providing an abstraction between the domains, which has significant practical advantages.

The AoI was initially introduced by the authors in [KYG12], who derive the optimal rate at which a data source must generate its information to keep its states as timely as possible in a resource-constrained network through a first-come-first-serve queueing discipline. Since then, many more works have been published that differ in their assumptions regarding (a) the inclusion [KSM18; ZS19; Wan+20] vs. exclusion [Sun+16; Cha+19; Aya+19; SRP20] of packet losses, (b) slotted [KSM18; ZS19; Aya+19] vs. carrier-sense-multiple-access-based [Sun+16; Cha+19; Wan+20] channel access, (c) single-user [Sun+16; Cha+19] vs. multi-user [KSM18; ZS19; Wan+20; Aya+19; SRP20], (d) single-hop [Sun+16; KSM18; ZS19; Cha+19; Wan+20] vs. multi-hop [Aya+19; Ana+21; SRP20] network, and (e) uniform [Sun+16; KSM18; Cha+19; Wan+20; Aya+19] vs. non-uniform packet sizes [ZS19]. A survey on recent results can be found in [Yat+21]. These works almost exclusively assume event-based sampling at the sensor, meaning that sampling can occur anytime. Time synchronicity between control and communication system, i.e.,

a shared understanding of time, is not assumed. In most works, the primary objective is to minimize the *average* AoI present in the system through the design of optimal scheduling techniques for their respective system assumptions, which generally involves queuing-theoretic analysis. As sometimes the tail of the AoI distribution has a larger impact on application performance than the average, especially regarding the control application's dependability assessment, further works also attempted to minimize the probability of exceeding a certain AoI, which they term the peak AoI violation probability. The authors in [CCE16] introduce the term peak AoI and explore it in a queuing-theoretic context for M/M/1 queues. The authors in [Chi+21; Vik+20] extend this work to a range of tandem queues to also consider the impact of the edge node task queue or packet relaying. The authors in [Dev+19] incorporate more accurate models of channel coding, taking advantage of results in finite-blocklength information theory for memoryless channels. The authors in [Öst+19] proceeded to drop the memoryless assumption, also taking into account the temporal correlation of fading channels. The authors in [CSP21] explore the peak AoI in a M/M/2 fork-join system (static dual-connectivity). The AoI has most recently gained attention from influential system-level authors [Pop+21b; Pop+21a], stating that the AoI is a more suitable timing characteristic to characterize the overall real-time operation compared to latency. This is because the AoI more generally does not only incorporate *latency* but also *packet losses* and the application's *sampling period* as major influences on real-time application behavior and reliability, compare Fig. 2.1. In this thesis, only the sampling period and packet losses are considered as cause of AoI because in time-synchronous systems, packet losses encapsulate (variable) latency through conversion to (a) a constant latency if the latency is smaller than a certain threshold or (b) a packet loss if the threshold is exceeded.

2.4 Dependability

The authors of [Höß+17; HSF18b; HSF18a; Höß+19; Höß20] have transformed well-known metrics from dependability theory to wireless communications, such as mean time to failure, mean time between failures, mean down time, reliability, and availability. These metrics are particularly important because (a) humans are much more familiar with time-based metrics and are thus more capable to assign meaning to them, and (b) they more conclusively outline what statistical properties are to be expected from the wireless link and how expensive (in terms of network resources) an extreme PLR, e.g., 10^{-9}, is. Even then, e.g., the mean time between failures is only in the range of days rather than years. This thesis builds upon the

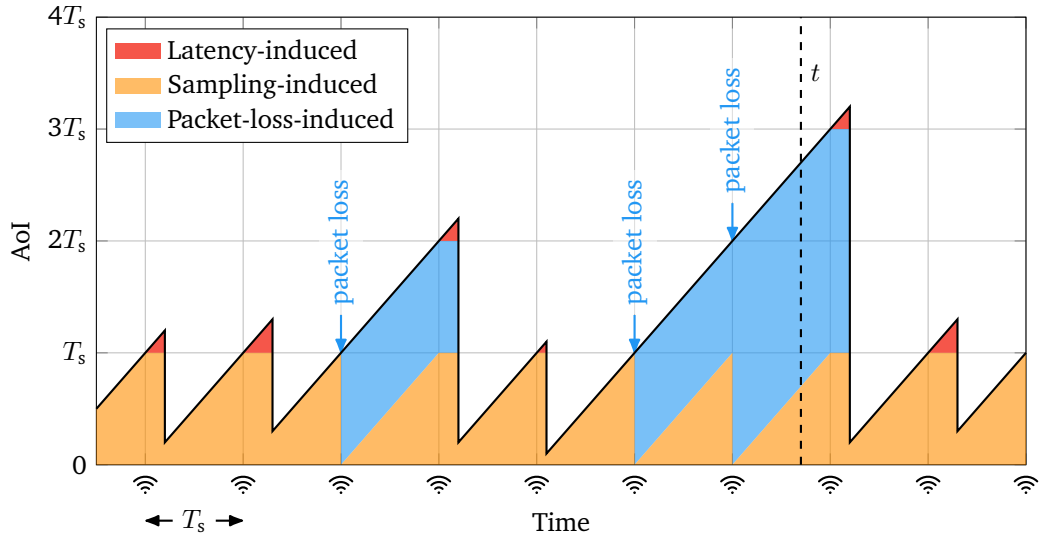


Fig. 2.1.: AoI "saw tooth", inspired by [Chi+22; Öst+19; Dev+19]. At each time instant t, the coloring specifies the AoI source composition. Note that in the diagram and this thesis in general, sampling and transmission are considered to be time-synchronized, i.e., there is no waiting (queueing) time. They occur periodically at every integer multiple of T_s, indicated by the symbol ⌔. A delayed packet translates into a linear AoI increase (latency-induced AoI). Upon packet reception, the AoI falls to the age of the received packet (the experienced latency). Between samples, the AoI increases as natural consequence of time-discrete systems (sampling-induced AoI). Upon packet losses, the AoI continues to increase. Note that with a try-once-discard approach (no immediate retransmissions), the AoI will keep increasing until (at least) the next sampling instant.

previous work by embracing the dependability theory framework and most centrally incorporating time-based metrics for assessment.

Dependability guarantees are of utmost importance for industrial applications as down-times are costly. In the early days of URLLC research, and as a siginficant motivator for reliable communications, down-times of the communications channel have been equated to down-times of the served application, which is clearly not necessary as the NCS-related research shows. However, a mapping of communication failures to application failures needs to be developed to assess the application-related dependability, which ultimately costs money. Mapping attempts from NCS research evaluate the asymptotic stability of the overall system, which is an inconclusive statement from a practical viewpoint because this mathematical property will not matter if in the meantime the application has caused damage. Rather, AoI-based failure evaluations must be applied [3GP19a], with the number of consecutive packet losses as a special case [5G 19]. **Proposition** [The closed-form AoI-based evaluation of control system failures is a main contribution of this thesis. This evaluation is

not only performed for a single agent but also for multiple coexisting agents that compete for network resources. As most prevalent network architectures, single-hop (UL *or* DL) and dual-hop connections (UL + DL) are investigated in this thesis.]

2.5 Conclusions

While this thesis touches on many topics covered in prior work, it also provides multiple new aspects, which haven't been studied before. First, periodic sampling and time-synchronous operation (a shared understanding of time [ITU20]) are applied because these are key requirements for industrial scenarios. This means that the sampling action is timed such that data transmission can be carried out directly thereafter and no random waiting times occur. Furthermore, a try-once-discard approach is employed, which avoids random-length, latency-prone ARQ or HARQ rounds and provides a great degree of determinism, which is also a key requirement in industrial environments. Dynamic multi-connectivity (MC) is applied on a per-transmission basis (adjustable number of channels for every transmission) within a pre-allocated resource pool and SPS is used for predictable resource scheduling. These assumptions specifically target *industrial* requirements and will be presented in greater detail in the connectivity assumptions of Chapter 4. The AoI is considered as a novel metric to support the vision of highly dependable applications over spectrally efficient wireless networks. Contrary to many AoI-related publications, the focus is not to minimize the average AoI but rather to ensure a pre-specified peak AoI violation probability that is expressed in terms of the control system's mean time to failure (MTTF) in this thesis. This analogy can be used because a control system failure is assumed as soon as the peak AoI is violated. Due to the connectivity assumptions, a queueing theoretic analysis is obsolete and the violation probability may be abstracted for a single-hop network as "How likely is it to lose more than K consecutive packets?". Also for the dual-hop network case, the AoI as the fundamental KPI may be approximated as an integer multiple of the sampling period. To the best of the author's knowledge, dynamic MC has not yet been utilized as a tool to ensure a low peak AoI violation probability while simultaneously attempting to achieve a spectral efficiency close to that of broadband networks.

3 Deriving Proper Communications Requirements

The work in this chapter was partially first published in [Sch+19]. The author's personal contributions comprise the derivation of the Markov model describing negative packet loss correlation for a maximum packet loss sequence length, its application to the AGV use case and the determination of an operating point for industrial wireless networks that is fundamentally different from current URLLC research suggestions.

The previous chapter put the contributions of this thesis into the context of the related work and motivated that both the URLLC approach as well as a *deep codesign* approach may not be suited to be applied in the ever-changing industrial environment, which requires a high degree of scalability (many concurrent applications), flexibility and reconfigurability. Tab. 3.1 compares both approaches (subjective classifications) with respect to various metrics and thereby illustrates possible areas of optimization. The *system redesign* row refers to the necessity to change existing control architectures in industrial environments, e.g., applying event-based sampling instead of periodic sampling. The *domain crossing* row addresses the "interweaving" of both domains, e.g., whether control model predictions can influence low-level network functions. The *network resources* row refers to how many resources are required on the air interface to support the respective codesign strategy.

Tab. 3.1.: KPI comparison of URLLC and the *deep codesign* approach.

	URLLC	*Deep Codesign*	*Target*
System Redesign	none	very high	none
Domain Crossing	none	very high	low
Network Resources	very high	very low	low

The aim of this chapter is to confirm that the *target* column of Tab. 3.1 is not wishful thinking and may be applied in near-future wireless networks. It will be demonstrated by means of a relevant example that a significant robustness against (isolated) packet losses already exists in control systems that are designed according to basic control theoretical principles. It will be shown that the PLR alone – as used

in URLLC – is a non-conclusive and non-ideal indicator to determine a network's suitability to support closed-loop wireless distributed control systems. Furthermore, the impact of packet loss correlation, which is closely related to the AoI, will be thoroughly investigated as the key to a low-complexity network optimization with extreme benefits in both control and communication domains. The packet loss correlation poses a design criterion that is not (yet) widely recognized and exploited in wireless network design for industrial environments and applications.

This chapter is structured as follows. First, a high-level introduction to system and control theoretical principles will be given, highlighting (the impact of) sampling, introducing z-domain and state-space representations, as well as typical performance requirements that together account for a certain quality of control (QoC). Second, a joint design of control loop with packet losses will be performed, highlighting the simplified yet effective *Reduced Sampling* approach and then the more general MJLS method. Third, both analysis methods are applied to the relevant AGV use case as focus application for *Industry 4.0*. The impact of packet loss correlation is analyzed in depth, building the foundation for the rest of this thesis. Extensive simulations of the AGV use case confirm the theoretically obtained results.

3.1 Fundamentals of Control Theory

The content of this subsection is gathered in large parts from the control system design fundamentals presented in [Nis00; FPE09; FWP97].

The behaviour of a many dynamic systems may be described through ordinary differential equations (ODEs). If these ODEs are linear, they can be broken down into a set of first-order ODEs and the system may be described via its linear *state-space representation*

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) \\ \boldsymbol{y}(t) &= \boldsymbol{C}(t)\boldsymbol{x}(t) + \boldsymbol{D}(t)\boldsymbol{u}(t) \ . \end{aligned} \tag{3.1}$$

Therein, $\boldsymbol{u}(t)$ is the *input vector* gathering all dynamic system inputs at time t, $\boldsymbol{y}(t)$ the *output vector* gathering all outputs, $\boldsymbol{x}(t)$ the *state vector* gathering all internal states, $\boldsymbol{A}(t)$ the *state matrix* capturing how internal states influence one another, $\boldsymbol{B}(t)$ the *input matrix* describing how inputs influence states, $\boldsymbol{C}(t)$ the *output matrix* relating the states to the system output, and $\boldsymbol{D}(t)$ the *feedthrough matrix* characterizing how inputs directly influence outputs. If the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ depend upon the absolute time t (as denoted above), the system is *time-variant*, otherwise it is *time-invariant* (all matrices are constant). Systems that

are both, linear and time-invariant build the class of LTI systems, for which a vast amount of methods exist to analyze their behavior. Their time domain representation (Eq. (3.1)) may be transformed to the frequency domain through

$$H(s) = \frac{Y(s)}{U(s)} = \boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D} \ , \tag{3.2}$$

where s denotes the Laplace variable, $Y(s)$ describes the Laplace-transform of the system output, $U(s)$ of the system input, and $H(s)$ denotes the *transfer function*.

3.1.1 Feedback Control

The continuous-time feedback control system of Fig. 3.1 is introduced. Its open-loop transfer function is given by

$$H_{\text{open-loop}}(s) = C(s)P(s) \tag{3.3}$$

and its closed-loop transfer function by

$$H_{\text{closed-loop}}(s) = \frac{C(s)P(s)}{1 + C(s)P(s)} \ . \tag{3.4}$$

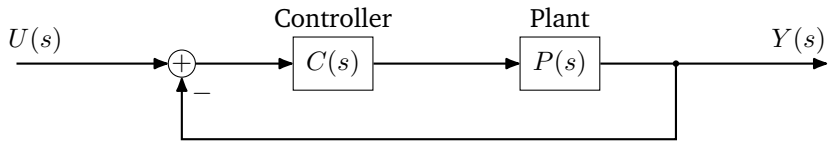


Fig. 3.1.: Feedback Control System.

3.1.2 Sampling

With the advent of ubiquitous computing, control systems have been digitalized to a great extent. Digitalization requires sampling of continuous-time processes, which must be frequent enough to incorporate all relevant frequency components of the signal under investigation. Usually, a target closed-loop system bandwidth is specified and an oversampling in the range 10 to 30 is chosen [FWP97].

While sampling can also be performed in an event-based fashion, which promises (a) increased control performance [AH14] and (b) a reduced number of required samples for a given control task [Wu+13; Han+15], the focus of this thesis lies on control systems with periodic sampling, i.e., with a constant inter-sampling

distance, also termed the *sampling period* T_s. The reasons are threefold: (a) Many of today's time-critical industrial control systems employ periodical sampling [5G 20; Aij20], (b) a vast amount of methods exists to describe and influence the behavior of periodically sampled systems, and (c) a networked system's proper operation can be ensured because of *plannable* network traffic [KYG12], which is particularly important for systems sharing a communication medium. With event-based communication, packet queues, collisions, and/or losses may occur, which might cause significant undesired latency overhead, which in turn reduces system reliability. Also, time-synchronicity, a major requirement for industrial manufacturing, cannot be guaranteed with such QoS degradations.

The choice of a proper sampling period and latency bound for networked control systems is mainly governed by the "single most important" [FWP97] effect of sampling: the delay introduced by the zeroth order hold (ZOH) that keeps the signal constant between samples. This operation results in the sampled signal lagging *on average* $T_\mathrm{s}/2$ behind the continuous signal [FWP97], in other words, a sampling period T_s has a comparable effect on the real-time application as a latency of $T_\mathrm{s}/2$. Formally, the effect of sampling is described by the phase shift

$$\mathrm{d}\phi = -2\pi f \frac{T_\mathrm{s}}{2} \,. \tag{3.5}$$

This relationship is particularly helpful to put network QoS requirements stated in many URLLC related publications into perspective. An end-to-end delay of $\tau_\mathrm{d} = 1\,\mathrm{ms}$ has approximately the same effect as periodic sampling with a sampling period $T_\mathrm{s} = 2\,\mathrm{ms}$.

Suppose a continuous-time LTI dynamic system – characterized by its state-space matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ – is discretized according to Fig. 3.2. Its discrete-time state-space representation is given by

$$\begin{aligned} \boldsymbol{x}[(n+1)T_\mathrm{s}] &= \boldsymbol{A}_\mathrm{d}\boldsymbol{x}[nT_\mathrm{s}] + \boldsymbol{B}_\mathrm{d}\boldsymbol{u}[nT_\mathrm{s}] \\ \boldsymbol{y}[nT_\mathrm{s}] &= \boldsymbol{C}_\mathrm{d}\boldsymbol{x}[nT_\mathrm{s}] + \boldsymbol{D}_\mathrm{d}\boldsymbol{u}[nT_\mathrm{s}] \,. \end{aligned} \tag{3.6}$$

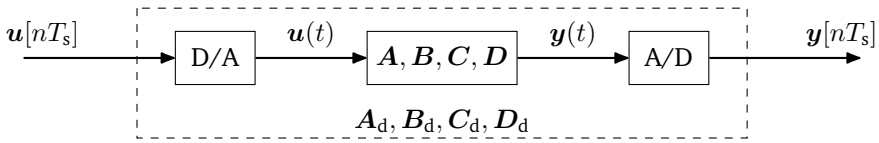


Fig. 3.2.: Discretization of a dynamic system [FWP97]. The D/A and A/D blocks are exerted at every integer multiple of T_s.

Therein, n refers to the sampling index and the following relations hold:

$$\begin{aligned} \boldsymbol{A}_\mathrm{d} &= e^{\boldsymbol{A}T_\mathrm{s}} & \boldsymbol{B}_\mathrm{d} &= \boldsymbol{B}\int_{\alpha=0}^{T_\mathrm{s}} e^{\boldsymbol{A}\alpha}\mathrm{d}\alpha \\ \boldsymbol{C}_\mathrm{d} &= \boldsymbol{C} & \boldsymbol{D}_\mathrm{d} &= \boldsymbol{D} \end{aligned} \tag{3.7}$$

Similar to Eq. (3.2), this sampled time-domain representation can be transformed to the frequency domain through

$$H(z) = \frac{Y(z)}{U(z)} = \boldsymbol{C}_\mathrm{d}(z\boldsymbol{I} - \boldsymbol{A}_\mathrm{d})^{-1}\boldsymbol{B}_\mathrm{d} + \boldsymbol{D}_\mathrm{d}\ . \tag{3.8}$$

Therein, the discrete-time frequency variable z relates to the continuous-time frequency variable s according to

$$z = e^{sT_\mathrm{s}}\ . \tag{3.9}$$

The inverse transformation from z-domain to state-space can be performed by constructing the canonical transfer function representation

$$H(z) = \frac{\beta_1 z^{N-1} + \beta_2 z^{N-2} + \cdots + \beta_N}{z^N + \alpha_1 z^{N-1} + \alpha_2 z^{N-2} + \cdots + \alpha_N} + \gamma \tag{3.10}$$

with N denoting the number of system states, and then determining the state-space equations through

$$\begin{aligned} \boldsymbol{A}_\mathrm{d} &= \begin{bmatrix} -\alpha_1 & -\alpha_2 & \cdots & -\alpha_{N-1} & -\alpha_N \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} & \boldsymbol{B}_\mathrm{d} &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ \boldsymbol{C}_\mathrm{d} &= \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_{N-1} & \beta_N \end{bmatrix} & \boldsymbol{D}_\mathrm{d} &= \begin{bmatrix} \gamma \end{bmatrix}\ . \end{aligned} \tag{3.11}$$

In this chapter, all system descriptions – discrete-time, continuous-time, state-space (time domain), and frequency domain – will be used.

3.1.3 Performance Requirements

The performance requirements of a control system may be manifold and in the following, a non-exhaustive list of typical requirements is presented.

Stability of a control system determines its fundamental usability. It states whether internal system states grow without bound (unstable), approach a finite value

(stable), or oscillate (boundary stable) during operation. Stability is a fundamental property of a dynamic system and does not depend upon its input. The stability of an LTI control system can be obtained in both the frequency domain and the time domain. In the frequency domain, the closed-loop system's stability is determined through applying the *Nyquist stability criterion* to the open-loop system. In the special case of a stable open-loop system, the phase margin ϕ_{m} of the open-loop system – defined as the difference in phase lag and $-180°$ at 0 dB gain – may be used as a simple criterion to evaluate closed-loop stability: A positive phase margin $\phi_{\mathrm{m}} > 0$ is synonomous with stability, a negative $\phi_{\mathrm{m}} < 0$ means instability. In the time domain, the eigenvalues of the closed-loop system's state matrix determine its stability. In the continuous-time case, the closed-loop system is stable if all eigenvalues lie in the complex left half-plane (negative real part) and in the discrete-time case, the system is stable if all eigenvalues lie within the unit cirlce. Otherwise, the closed-loop system is unstable.

Transient Response is the umbrella term for the dynamic properties of the control system and is usually evaluated by feeding a control system with an input step and analyzing the *step response*. The *rise time* of the step response determines how quickly the system is able to react to inputs. Too low rise times make the application slow and sluggish, potentially causing high state deviations, too fast rise times might make people uncomfortable, unnecessarily cause wear and tear or damage, or cause potentially unwanted oscillations due to overshoot.

Steady-State Response relates to the output of a control system after the transients have decayed. This is especially important for control systems that must be driven precisely to specific target positions.

Other performance requirements may be more application-specific and cannot be described by textbook control theoretical principles and definitions. E.g., an industrial robot may have the requirement of a maximum position error compared to a pre-calculated trajectory while not exceeding a maximum joint acceleration. In addition, it might be desired to limit the control effort, which typically relates to the control signal's energy, for a given task. These control performance criteria are not captured by any of the criteria above. Although computer tools such as MATLAB heavily aid the design process, the experience of well-trained engineers is indispensable to build highly performant control systems that meet all design criteria.

3.1.4 Packet Losses and Delay

Packet losses and delay are fundamental effects introduced by communication networks and usually undesired. They may result from a multitude of different factors including but not limited to network architecture, congestion, interference, medium access, resource allocation, scheduling, path loss, fading, and retransmissions. It is crucial to understand that packet losses and delay are tightly coupled quantities and not independent KPIs [Pop+18a; 3GP19b], as was illustrated in Fig. 1.1. Formally, a packet is considered *lost*, if it was not received successfully by the destination within a pre-defined delay bound $\tau_{\text{d,bound}}$. In a practical time-synchronized systems, any delay variations (latency *jitter*) will be eliminated by a receive buffer that stores any received signals until they are picked up by the next clock cycle on the application level. It is for this reason that approximating the system's delay with a constant value $\tau_{\text{d}} \widehat{=} \tau_{\text{d,bound}}$ is a reasonable assumption. If, furthermore, the delay is assumed significantly smaller than half of the control system's sampling period, $\tau_{\text{d,bound}} << T_{\text{s}}/2$, compare Eq. (3.5), it can be ignored alltogether. This simplification is assumed here, hence, in the following, only packet losses are assumed for joint modeling with control systems.

3.2 Joint Design of Control Loop with Packet Losses

Fig. 3.3 depicts a control system consisting of a discrete-time LTI controller $C(z)$ connected through a lossy wireless network to a continuous-time LTI plant $P(s)$. The random binary processes Λ_{dl} and Λ_{ul} generate successful packet transmissions and packet losses in the UL and DL, respectively. In case of a successful transmission, each switch forwards the transmitted data packet, in case of an unsuccessful transmission an estimation thereof is generated by a signal reconstruction filter. Note that while there exists a wide range of highly performant and sophisticated state estimators [Sch+07; SEM10; CL20; Bau+14], this chapter will be limited to ZOH and first order hold (FOH) for signal reconstruction for conciseness.

Because packet losses are inherently time-variant, the dynamic switching resembles a time-variant process, which transforms the former LTI system into a linear, time-variant (LTV) system. In the time domain, this means that the state-space matrices $\boldsymbol{A}_{\text{d}}[nT_{\text{s}}]$, $\boldsymbol{B}_{\text{d}}[nT_{\text{s}}]$, $\boldsymbol{C}_{\text{d}}[nT_{\text{s}}]$, and $\boldsymbol{D}_{\text{d}}[nT_{\text{s}}]$ of the closed-loop system depend upon whether at nT_{s} the packet transmissions in UL and DL have been successful or not. Frequency-domain analysis cannot be performed because neither the z- (discrete-time) nor the Laplace-transform (continuous-time) can be applied to LTV dynamic

systems. However, certain assumptions on the processes Λ_{dl} and Λ_{ul} as well as the signal reconstruction filters enable to also analyze such *switching systems*. In this chapter, the two methods (a) *Reduced Sampling* and (b) *MJLS* will be presented.

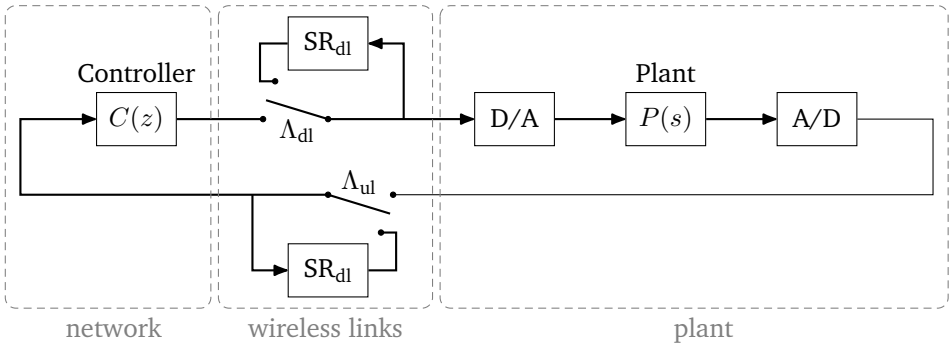


Fig. 3.3.: Joint design of control system with packet losses.

3.2.1 Method 1: *Reduced Sampling*

Let the packet loss sequence length (PLSL) describe the number of consecutively lost packets. For instance, a single isolated packet loss (a lost packet that is enclosed by two successful transmissions), is described by PLSL $= 1$. With this definition, the assumptions of

- *circular* and *synchronized* packet losses in UL and DL until a given PLSL and
- a ZOH signal reconstruction filter in both UL and DL (the previous known packet will be held as long as packets are lost)

provide a simplified yet impactful first impression regarding the influence of packet losses on the control system. *Circular* packet losses in this context mean that a successful packet transmission is always followed by exactly PLSL packet losses, which are then followed by exactly one successful packet transmission, and so on, see Fig. 3.4. *Synchronized* refers to packet losses occuring either in both, UL *and* DL

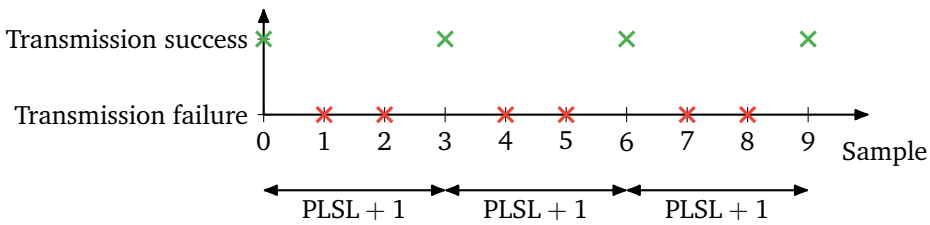


Fig. 3.4.: Depiction of circular packet losses, here with PLSL $= 2$.

simultaneously, or in *neither*, but never in only one transmission direction. These restrictions enable to examine the control system by standard LTI control theoretical principles as they relate to reducing the sampling rate by a factor of $(\text{PLSL} + 1)$. For example, at $\text{PLSL} = 1$, the effective sampling rate is reduced by a factor of 2 because successful and failed transmissions alternate. From a control theoretical perspective, this means a model transformation to a larger sampling period, which can be achieved through matrix powers in discrete-time state-space:

$$\begin{aligned} \boldsymbol{A}_\mathrm{d}{}' &= \boldsymbol{A}_\mathrm{d}{}^{\text{PLSL}+1} & \boldsymbol{B}_\mathrm{d}{}' &= \boldsymbol{B}_\mathrm{d} \sum_{k=0}^{\text{PLSL}} \boldsymbol{A}_\mathrm{d}{}^{k} \\ \boldsymbol{C}_\mathrm{d}{}' &= \boldsymbol{C}_\mathrm{d} & \boldsymbol{D}_\mathrm{d}{}' &= \boldsymbol{D}_\mathrm{d} \end{aligned} \tag{3.12}$$

As the closed-loop system is still LTI, the conventional open-loop phase margin and transient response characteristics can be calculated and evaluated. Thus, the *Reduced Sampling* method provides an easily applicable approach to obtain a rough estimation of a control system's packet loss tolerance. However, through its simple nature it fails to describe

1. signal reconstruction methods other than ZOH, which might yield better estimation performance in terms of stability and overall signal "smoothness",
2. aperiodic, non-circular, probabilistic packet losses, and
3. packet losses that occur in any transmission direction during one time step (both, neither, or only *one*).

Hence, also a more sophisticated analysis method is presented in the following section.

3.2.2 Method 2: *Markov Jump Linear System*

Large parts of this section are taken from [CFM05].

As pronounced earlier, packet losses in a networked control system transform any LTI control system to an LTV system. Originally, the motivation to also gain deeper insights on LTV systems stemmed from potential abrupt environmental disturbances, component failures, subsystem interconnections, or abrupt changes in the operation point for a non-linear plant [CFM05]. However, the same theory can also be applied to packet losses, which is a standard approach in NCS literature [SS05; SS01; TM04; MCF13; QAJ13; YFX11]. These works assume a two-state Markov chain with one *up*

(transmission success) and one *down* (transmission failure) state as network model. In [MCF13; QAJ13; YFX11; Par+18], it was found that burst errors, i.e., temporally correlated errors, are particularly harmful for the stabilizability of the markov jump linear system (MJLS).

The theory of MJLS is built around the assumption that a set of *operation modes* $\zeta \in \{0, \ldots, Z-1\}$ is defined. Each operation mode is associated with an own set of state-space matrices $\boldsymbol{A}_\mathrm{d}(\zeta)$, $\boldsymbol{B}_\mathrm{d}(\zeta)$, $\boldsymbol{C}_\mathrm{d}(\zeta)$, and $\boldsymbol{D}_\mathrm{d}(\zeta)$ that capture the system dynamics for that particular mode.

$$\begin{aligned} \boldsymbol{x}[(n+1)T_\mathrm{s}] &= \boldsymbol{A}_\mathrm{d}(\zeta[nT_\mathrm{s}])\boldsymbol{x}[nT_\mathrm{s}] + \boldsymbol{B}_\mathrm{d}(\zeta[nT_\mathrm{s}])\boldsymbol{u}[nT_\mathrm{s}] \\ \boldsymbol{y}[nT_\mathrm{s}] &= \boldsymbol{C}_\mathrm{d}(\zeta[nT_\mathrm{s}])\boldsymbol{x}[nT_\mathrm{s}] + \boldsymbol{D}_\mathrm{d}(\zeta[nT_\mathrm{s}])\boldsymbol{u}[nT_\mathrm{s}] \end{aligned} \tag{3.13}$$

It is assumed that the transitions between the operation modes can be modeled through a Markov model.*

Stability

The stability of an MJLS in the mean square sense is given if [CFM05]

$$\sum_{n=0}^{\infty} \mathbb{E}\left\{\boldsymbol{x}[nT_\mathrm{s}]\boldsymbol{x}^*[nT_\mathrm{s}]\right\} < \infty \quad \text{for} \quad \boldsymbol{u}[nT_\mathrm{s}] = 0 \quad \text{and any} \quad \boldsymbol{x}[0], \zeta[0]\,. \tag{3.14}$$

This means that for mean square stability**, the sum state energy must remain bounded independent of the initial control system state $\boldsymbol{x}[0]$ and initial operation mode $\zeta[0]$ of the MJLS, given zero input $\boldsymbol{u}[nT_\mathrm{s}] = 0$.

Let $\boldsymbol{P}_\mathrm{mjls}$ define the discrete-time transition probability matrix of the underlying Markov model, $\boldsymbol{I}_{(i\times i)}$ the identity matrix of size $(i \times i)$, and

$$\text{blockdiag}\left\{\boldsymbol{M}_i\right\} := \begin{bmatrix} \boldsymbol{M}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{M}_m \end{bmatrix} \quad \text{with} \quad \boldsymbol{M}_i \in \{\boldsymbol{M}_1, \ldots, \boldsymbol{M}_m\} \tag{3.15}$$

*Note that while this approach would also enable to modify controller behavior in case of packet losses (e.g., reducing the controller's aggressiveness), the focus here is to identify the boundaries of co-operation between existing control systems and error-prone networks. Therefore, controller co-optimization will not be performed.

**Note that in [Ji+91] it was proven that for systems (3.13), mean-square stability, stochastic stability, and exponential mean-square stability are equivalent.

as the block diagonal matrix formed with $\boldsymbol{M}_i$ in the diagonal and zero elsewhere. Further define

$$\mathcal{C} := \boldsymbol{P}_{\text{mjls}}{}^{\intercal} \otimes \boldsymbol{I}_{(n^2 \times n^2)} \tag{3.16}$$

$$\mathcal{N} := \text{blockdiag}\left\{\boldsymbol{A}_{\text{d}}(\zeta)^* \otimes \boldsymbol{A}_{\text{d}}(\zeta)\right\} \tag{3.17}$$

$$\mathcal{A} := \mathcal{C}\mathcal{N} \tag{3.18}$$

where $\boldsymbol{P}_{\text{mjls}}{}^{\intercal}$ denotes the transposition of $\boldsymbol{P}_{\text{mjls}}$, $\boldsymbol{A}_{\text{d}}(\zeta)$ the state matrix of the ζ-th operation mode, $\boldsymbol{A}_{\text{d}}(\zeta)^*$ the conjugate of $\boldsymbol{A}_{\text{d}}(\zeta)$, and $\otimes$ the Kronecker product. Similar to standard LTI control theory, the corresponding MJLS is mean square stable if the spectral radius (largest absolute eigenvalue) r_σ of the augmented state matrix $\mathcal{A}$ is smaller than 1, i.e., [CFM05]

$$r_\sigma(\mathcal{A}) < 1 \ . \tag{3.19}$$

The intuition of the above equations (3.16) – (3.18) is that if unstable states are visited more frequently, the overall MJLS system is more probably unstable.

Modeling Packet Losses

Building on the findings of previous works that temporal correlation of packet losses is particularly harmful for real-time applications, the MJLS modeling in this thesis is different from the cited works because the Markov states in this thesis do not refer to the network state (*up/down*) but rather to the AoI, thereby aiming at actively breaking the temporal packet loss correlation to increase stabilizability/dependability.

In any way, to incorporate packet losses into the MJLS description, all sets of state-space matrices must be derived, which depend on the signal reconstruction filters used in UL and DL, respectively. Four distinctions will be made here:

1. UL/DL both succeed:

$$\boldsymbol{A}_{\text{d}}{}^{(1,1)}, \boldsymbol{B}_{\text{d}}{}^{(1,1)}, \boldsymbol{C}_{\text{d}}{}^{(1,1)}, \boldsymbol{D}_{\text{d}}{}^{(1,1)} \tag{3.20}$$

   → Both transmission signals are forwarded to controller and plant, respectively.

2. UL fails / DL succeeds:

$$\boldsymbol{A}_{\text{d}}{}^{(0,1)}, \boldsymbol{B}_{\text{d}}{}^{(0,1)}, \boldsymbol{C}_{\text{d}}{}^{(0,1)}, \boldsymbol{D}_{\text{d}}{}^{(0,1)} \tag{3.21}$$

→ The UL signal is estimated through the UL signal history, while the successfully DL packet is forwarded to the plant.

3. UL succeeds / DL fails:

$$\boldsymbol{A}_{\mathrm{d}}{}^{(1,0)}, \boldsymbol{B}_{\mathrm{d}}{}^{(1,0)}, \boldsymbol{C}_{\mathrm{d}}{}^{(1,0)}, \boldsymbol{D}_{\mathrm{d}}{}^{(1,0)} \tag{3.22}$$

→ The successfully transmitted UL signal is forwarded to the controller, while the DL signal is estimated through the DL signal history.

4. UL/DL both fail:

$$\boldsymbol{A}_{\mathrm{d}}{}^{(0,0)}, \boldsymbol{B}_{\mathrm{d}}{}^{(0,0)}, \boldsymbol{C}_{\mathrm{d}}{}^{(0,0)}, \boldsymbol{D}_{\mathrm{d}}{}^{(0,0)} \tag{3.23}$$

→ Both transmission signals are estimated through the respective signal histories.

In the context of networked control, the system matrices given in (3.20) are stable (because UL and DL both succeed) and unstable for (3.21)–(3.23) (because the loop is open due to at least one transmission failing). MJLS theory is capable of determining "how much" of $\boldsymbol{A}_{\mathrm{d}}{}^{(0,1)}$, $\boldsymbol{A}_{\mathrm{d}}{}^{(1,0)}$, and $\boldsymbol{A}_{\mathrm{d}}{}^{(0,0)}$ the overall system can tolerate until it becomes unstable as a whole.

It should be noted that the four distinctions made above are a simplification and the modeling can be arbitrarily refined. The simplification mainly stems from previously estimated values being treated equally in the generation of the next estimation compared to successfully transmitted data. E.g., when using a FOH estimator as signal reconstruction filter, one might only want to include successfully transmitted data for the estimation of the next sample because estimated values carry a greater uncertainty. For this refined case, the number of distinctions would need to be increased as both estimator designs depend on the short-term history of packet transmission successes/failures.

After deriving the sets of applicable state-space representations, they must be mapped to an operation mode ζ of the MJLS, i.e., the state of the Markov chain that captures the switching behavior between operation modes. Three main objectives are pursued:

1. Highlight the effects of packet losses on the control loop.

2. Capture the impact of limiting the PLSL to a maximum value K. This could be seen as a suitable design goal for network engineers, i.e., making sure that K packet losses are not exceeded.

3. Show the impact of packet loss *correlation* [Par+18; MCF13; QAJ13; YFX11] on the closed-loop control performance. The prevention of such burst errors may also account for a suitable network optimization goal.

Hence, a temporal packet loss correlation model with a maximum PLSL of K is introduced, see Fig. 3.5.

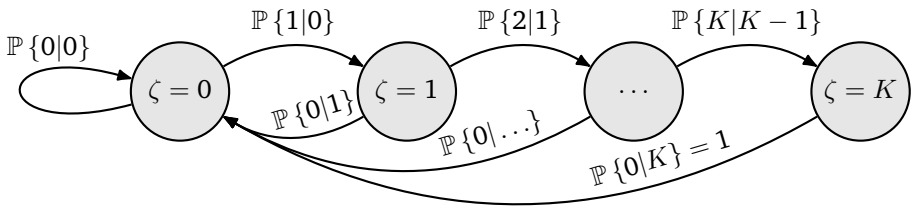


Fig. 3.5.: MJLS to describe packet loss correlation with a maximum PLSL of K.

This MJLS has the general transition probability matrix (all obvious zeros are omitted for readability)

$$\boldsymbol{P}_{\text{mjls}} = \begin{bmatrix} \mathbb{P}\{0|0\} & \mathbb{P}\{1|0\} & & & \\ \mathbb{P}\{0|1\} & & \mathbb{P}\{2|1\} & & \\ \vdots & & & \ddots & \\ \mathbb{P}\{0|K\} & & & & \mathbb{P}\{K|K-1\} \\ \mathbb{P}\{0|K\} & & & & 0 \end{bmatrix}. \tag{3.24}$$

It is crucial to understand that the implicit assumption of this Markov model is that every control system switches off or to an emergency mode in the moment the $(K+1)$-st consecutive packet is lost. For brevity, this event will be referred to as *control system failure*. The MJLS analysis presented in this section aids in finding this particular K value by stating whether the control system is stable *given* that every packet loss sequence is limited to K packet losses. As a next step (Chapter 4), the non-zero probability of exceeding K consecutive packet losses will be calculated and a control system failure probability will be derived. Here, this probability is assumed zero to analyze control performance *under given network conditions* to derive proper network requirements for control systems.

To summarize, the following novelties are captured by the MJLS of Fig. 3.5:

1. The MJLS is considered in operation mode $\zeta = 0$ if the last transmission was successful, else ζ is increased by one.

2. The maximum PLSL is limited to K. After K consecutive packet losses, the MJLS jumps back to operation mode $\zeta = 0$, i.e., a successful packet transmission is *guaranteed*.

3. By modifying the state transition probabilities, a temporal packet loss correlation can be introduced that increases/decreases the probability of having consecutive packet losses.

For the packet loss correlation, $-1 \leq \rho \leq 1$ is introduced. A value $\rho < 0$ states that the probability of consecutive packet losses is decreased (negatively correlated) with $\rho = -1$ resembling fully negative correlation and therefore a zero probability of losing consecutive packets (a single packet may still be lost). $\rho = 0$ means zero packet loss correlation, i.e., the probability of losing a packet is identical for every operation mode. $\rho > 0$ increases the probability of consecutive losses with $\rho = 1$ guaranteeing K consecutive packet losses as soon as the first packet loss occurs. It is emphasized again that in any case the maximum PLSL is limited to K, guaranteeing a successful packet transmission after K consecutive packet losses. The transition probability matrices that incorporate ρ are formally introduced as (all obvious zero entries are omitted for readability):

$$\boldsymbol{P}_{\text{mjls}} = \begin{cases} \begin{bmatrix} 1-(1+\rho)^0\tilde{p}_0 & (1+\rho)^0\tilde{p}_0 & & \\ \vdots & & \ddots & \\ 1-(1+\rho)^{K-1}\tilde{p}_0 & & & (1+\rho)^{K-1}\tilde{p}_0 \\ 1 & & & 0 \end{bmatrix} & -1 \leq \rho < 0 \\ \begin{bmatrix} (1-\rho)^0 p_0 & 1-(1-\rho)^0 p_0 & & \\ \vdots & & \ddots & \\ (1-\rho)^{K-1} p_0 & & & 1-(1-\rho)^{K-1} p_0 \\ 1 & & & 0 \end{bmatrix} & 0 \leq \rho \leq 1 \end{cases} \tag{3.25}$$

Therein, $\tilde{p}_0$ describes the probability of making the *first* packet loss and $p_0 = 1 - \tilde{p}_0$ the probability of a successful transmission after another successful transmission. $\tilde{p}_0$ may be adjusted such that the overall PLR remains constant and, hence, a fair comparison between different settings is ensured. This is done by deriving the steady-state state probabilities $\boldsymbol{\pi}$ of the Markov model presented in Fig. 3.5 by solving

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}_{\text{mjls}} \ , \tag{3.26}$$

which is a standard eigenvalue problem. Since every successful packet transmission leads to entering operation mode $\zeta = 0$, the PLR can derived in a straight-forward way as

$$\mathrm{PLR} = 1 - \pi_0 \ , \tag{3.27}$$

where π_0 denotes the first element of $\boldsymbol{\pi}$.

3.2.3 Conclusion

In this section, two joint design methods of control loop and packet losses have been present: the *Reduced Sampling* method is able to provide a first impression of the effect of packet losses on the control loop at low complexity because the control system remains LTI and therefore all LTI analysis tools remain available. The more advanced *MJLS* method enables to also capture (a) arbitrary stationary signal reconstruction filters [Sch+07], (b) arbitrary reconstruction strategies, and (c) arbitrary packet loss sequences as long as the switching behavior between operation modes can be modeled through a Markov chain. With the MJLS approach, former LTI control systems are transformed to LTV systems and, hence, many tools and KPIs, e.g., the phase margin, become unavailable. However, the most important *stability* property is still available within the theory of MJLS.

3.3 Focus Application: The AGV Use Case

To make the results of this chapter more tangible, an AGV focus application is introduced. The two modeling approaches of the last section will be applied to this use case for illustration purposes. Obviously, all other LTI systems may be investigated similarly.

AGVs are self-driving vehicles in industrial settings that usually autonomously transport goods. In 2018, the global AGV market size was valued at USD 2.49 billion and it is expected to grow at a compound annual growth rate of roughly 16 % from 2019 to 2025 [Gra18], which highlights the relevance of AGVs in the future. This use case is chosen here since AGVs will highly benefit from incorporating closed-loop wireless control algorithms due to their need for mobility, deeming wires useless. Today's guidance systems rely on infrastructure such as tape on the ground or inductive wiring below the floor surface and, consequently, are not flexible and don't feature real-time fleet management. With wireless systems in the loop, path reconfigurations, possible load-dependant control parameter adjustments, and live collision avoidance

can be performed solely in software and online, leading to full flexibility in fleet management, also in dynamically changing environments.

3.3.1 Control Loop Model

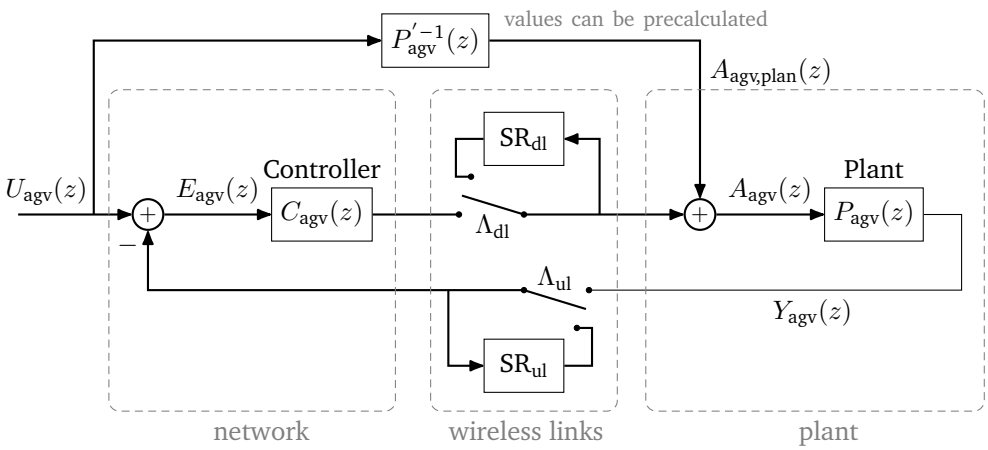


Fig. 3.6.: Control system model of AGV use case.

The considered position control loop model is depicted in Fig. 3.6. In this introductory section, the control model is introduced under the assumption of no packet losses, i.e., the random processes Λ_{ul} and Λ_{dl} are assumed to only generate successful packet transmissions (the switches are always in the horizontal position). The modeling of the error-prone wireless links consisting of Λ_{ul}, Λ_{dl}, SR_{ul}, and SR_{dl} will be presented later.

The control system is comprised of the following elements: The input $U_{\text{agv}}(z)$ constitutes a desired position, $Y_{\text{agv}}(z)$ the actual position. Consequently, $E_{\text{agv}}(z)$ denotes the position error. The controller $C_{\text{agv}}(z)$ is assumed to be located in the network, connected wirelessly to the discrete-time plant model $P_{\text{agv}}(z)$, which takes an acceleration command $A_{\text{agv}}(z)$ as control input. The trajectory planning is assumed to be capable of precalculating time-insensitive trajectory information $A_{\text{agv,plan}}(z)$ and forward it via a separate connection through $P'^{-1}_{\text{agv}}(z)$, which applies standard motion dynamics equations. The plant model $P_{\text{agv}}(z)$ resembles the motor used in the state-of-the-art KATE AGV by Götting GmbH and is detailed in Appendix A. For the outer-loop position control, a proportional-derivative (PD) controller is chosen for $C_{\text{agv}}(z)$ because of its simplicity and effectiveness. Standard control system design yields

$$C_{\text{agv}}(z) = K_{\text{p}} + \frac{K_{\text{d}}}{T_{\text{f}} + T_{\text{s}} z/(z-1)} = \frac{31.77z - 31.08}{z - 0.3077} \tag{3.28}$$

for a sampling period $T_\mathrm{s} = 45\,\mathrm{ms}$ (in the appropriate range for mobile robots [3GP19a]), $K_\mathrm{p} = 1$, $K_\mathrm{d} = 2$, and $T_\mathrm{f} = 0.02$. This results in an open-loop phase margin $\phi_\mathrm{m} = 55°$. As noted before, LTI control systems are stable if the phase margin is positive. Phase margins in the range $50° \leq \phi_\mathrm{m} \leq 80°$ are usually considered a good trade-off between stability and dynamic control behavior [FWP97]. The authors in [Kim17] state that particularly for direct current (DC) motors, phase margins in the range $30° \leq \phi_\mathrm{m} \leq 60°$ are desirable.

The overall closed-loop discrete-time state-space representation with the inputs $U_\mathrm{agv}(z)$, $A_\mathrm{agv,plan}(z)$ and the output $Y_\mathrm{agv}(z)$ is given by

$$
\begin{aligned}
\boldsymbol{A}_\mathrm{d,agv} &= \begin{bmatrix} 3.1\times10^{-1} & 0 & 0 & 0 & 6.9\times10^{-1} \\ 1.4 & 1.0 & 0 & 0 & -1.4 \\ 6.0\times10^{-2} & 1.2\times10^{-1} & 8.7\times10^{-1} & 5.2\times10^{-3} & -6.2\times10^{-2} \\ 2.1\times10^{1} & 2.7\times10^{1} & -2.6\times10^{1} & 3.5\times10^{-1} & -2.2\times10^{1} \\ 6.6\times10^{-4} & 1.9\times10^{-3} & 4.3\times10^{-2} & 1.3\times10^{-4} & 10\times10^{-1} \end{bmatrix} \\
\boldsymbol{B}_\mathrm{d,agv} &= \begin{bmatrix} -6.9 & 0 \\ 1.4\times10^{1} & 4.5\times10^{-1} \\ 6.2\times10^{-1} & 1.9\times10^{-2} \\ 2.2\times10^{2} & 6.8 \\ 6.8\times10^{-3} & 2.2\times10^{-4} \end{bmatrix} \\
\boldsymbol{C}_\mathrm{d,agv} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 1.0\times10^{-1} \end{bmatrix} \\
\boldsymbol{D}_\mathrm{d,agv} &= \begin{bmatrix} 0 & 0 \end{bmatrix} .
\end{aligned}
\tag{3.29}
$$

3.3.2 Control Performance Requirements

For the networked AGV control system, the following requirements are defined.

1. The system must be mean square stable.
2. The trajectory defined in Fig. 3.7 shall be traversed with a maximum deviation of $e_\mathrm{agv} < 2\,\mathrm{cm}$ at all times.
3. The control signal's root mean square (RMS) ratio is defined as

$$\eta_\mathrm{agv} = \mathrm{RMS}\{a_\mathrm{agv}\}/\mathrm{RMS}\{a_\mathrm{agv,ref}\} \ . \tag{3.30}$$

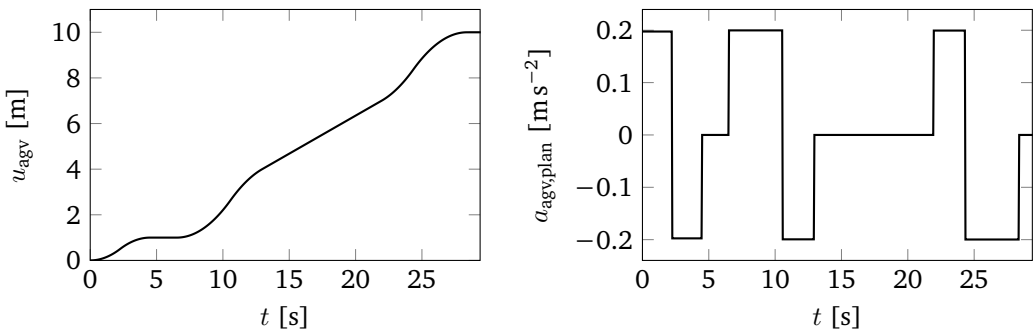
   This metric relates the average RMS value to the control signal's RMS value in case of zero packet losses and indicates the jerkiness of operation compared

to no packet losses. For the AGV use case, it is defined that $\eta_{\text{agv}} < 1.05$ is required.

4. The probability of exceeding a 30 % control signal RMS increase compared to the case with zero packet losses is defined as

$$P_{+30\,\%} = \mathbb{P}\left\{\text{RMS}\{a_{\text{agv}}\} > 1.3 \times \text{RMS}\{a_{\text{agv,ref}}\}\right\} \ . \tag{3.31}$$

This metric indicates the probability of certain packet loss statistics having a significant effect on the control performance and therefore acts as practical outlier protection. This protection is necessary because the assumed AGV model is only valid within certain operation bounds and unmeaningful (yet mathematically correct) results would otherwise be obtained. For the AGV use case, it is defined that $P_{+30\,\%} = 0$ is required.



(a) Desired position **(b)** Planned acceleration via motion dynamics

Fig. 3.7.: Reference trajectory for AGV use case.

3.3.3 Joint Modeling: Applying *Reduced Sampling*

An integer reduction of the sampling rate according to Eq. (3.12) yields the phase margins listed in Tab. 3.2. As expected, they generally decrease because of the additional delay associated with "holding the sample longer", compare Eq. (3.5). The interpretation of these phase margins is that – under the stated assumptions – the control system is still stable and operational if up to 2 circular packet losses occur ($\phi_{\text{m}} = 25.6°$). The system becomes marginally unstable ($\phi_{\text{m}} = -0.8°$) if 3 circular packet losses occur and significantly unstable if the circularly occuring packet loss sequences have a length of 4 or more. This simplified *Reduced Sampling* analysis method already demonstrates that digital control systems that are designed according to basic control theoretical principles inherently feature a tolerance against packet losses as long as not *too many* occur consecutively.

Tab. 3.2.: Resulting phase margins ϕ_m for K circular packet losses. $T_\mathrm{s,eff}$ describes the effective sampling period resulting from losing K packets in between two successful transmissions.

K	$T_\mathrm{s,eff}$	ϕ_m
0	45 ms	55.4°
1	90 ms	46.1°
2	135 ms	25.6°
3	180 ms	−0.8°
4	225 ms	−28.5°
5	270 ms	−55.8°

3.3.4 Joint Modeling: Applying *MJLS*

For the advanced analysis via MJLS theory, the state vector $\boldsymbol{x}$ must be extended by additional states that resemble the UL and DL signal reconstruction filters. Two cases will be investigated here:

1. Both transmission directions employ a ZOH signal reconstruction filter. This case requires to extend the state vector by one state per transmission direction and will be referred to as ZOH/ZOH.

2. The UL employs an FOH filter while the DL still employs a ZOH filter. This requires two additional states for the FOH filter and one additional state for the ZOH filter. The change from ZOH to FOH in the UL direction is motivated by the observation that the derivative component of the PD controller tends to dominate the control signal in the presence of packet losses, which manifests itself through a "spiky" control signal. An FOH filter reduces this effect. This case will be referred to as FOH/ZOH.

After appending the signal reconstruction states to the state vector and applying the appropriate state inter-dependencies in the state matrix $\boldsymbol{A}_\mathrm{d}$ for the cases discussed in Eq. (3.20)-(3.23), the resulting state matrices $\boldsymbol{A}_\mathrm{d}{}^{(1,1)}$, $\boldsymbol{A}_\mathrm{d}{}^{(1,0)}$, $\boldsymbol{A}_\mathrm{d}{}^{(0,1)}$, and $\boldsymbol{A}_\mathrm{d}{}^{(0,0)}$ must be mapped to the operation modes, which were presented in Fig. 3.5. For concise results, only the following mappings are considered:

1. For the ZOH/ZOH case, *synchronized* packet losses are assumed to compare the results from the MJLS analysis with the results obtained from the *Reduced Sampling* analysis. Both approaches should yield the same result regarding the

stability boundary for a maximum PLR and $\rho = 1$, which resembles *circular* packet losses. Formally:

$$\begin{aligned} \boldsymbol{A}_{\mathrm{d}}(\zeta = 0) &= \boldsymbol{A}_{\mathrm{d}}{}^{(1,1)} \\ \boldsymbol{A}_{\mathrm{d}}(\zeta \neq 0) &= \boldsymbol{A}_{\mathrm{d}}{}^{(0,0)} \end{aligned} \tag{3.32}$$

2. For the FOH/ZOH case, the DL is considered ideal as its detrimental effect is neglible compared to packet losses in the UL for the presented AGV control system due to the derivative component of the controller and, hence, presentation can be simplified significantly. Formally:

$$\begin{aligned} \boldsymbol{A}_{\mathrm{d}}(\zeta = 0) &= \boldsymbol{A}_{\mathrm{d}}{}^{(1,1)} \\ \boldsymbol{A}_{\mathrm{d}}(\zeta \neq 0) &= \boldsymbol{A}_{\mathrm{d}}{}^{(0,1)} \end{aligned} \tag{3.33}$$

The results for both cases will be discussed in the next sections.

Results for ZOH/ZOH case

Fig. 3.8 shows the stability boundary results for the ZOH/ZOH case and varying PLR, ρ, and K. Therein, the infeasibility region marks all impossible (PLR, ρ) combinations. For instance, for $\rho = -1$ and any K, the probability of having consecutive packet losses is zero, hence, the maximum valid PLR is limited to $\mathrm{PLR}_{\mathrm{max}} = 0.5$. Note also that for $\rho > 0$ the infeasibility slope is infinite because the PLSL is limited to K and, hence, $\mathrm{PLR} > \frac{K}{K+1}$ is impossible. The green area constitutes the stable region, in which the particular (PLR, ρ) combination leads to a stable AGV control system. Unstable combinations are summarized in the red area. For $K \in \{1, 2\}$, an unstable region does not exist, which means that any feasible (PLR, ρ) combination will lead to a stable system. For increasing K, the instability region increases. Similar to the results of previous works, the plots show that a high packet loss correlation negatively impacts the control loop as all solid curves are monotonically decreasing and exhibit a great negative slope for high ρ. Especially for high K, the packet loss correlation ρ must be strictly bounded to ensure overall mean square stability. Conversely, this means that higher average PLR can be tolerated if the packet loss correlation can be reduced.

The results of Sec. 3.3.3 (*Reduced Sampling*), which were derived by means of the phase margin ϕ_{m}, are confirmed. It was shown that at $K = 3$, the control system becomes marginally unstable, which is confirmed by the plot in Fig. 3.8 because there exists only a tiny instability region.

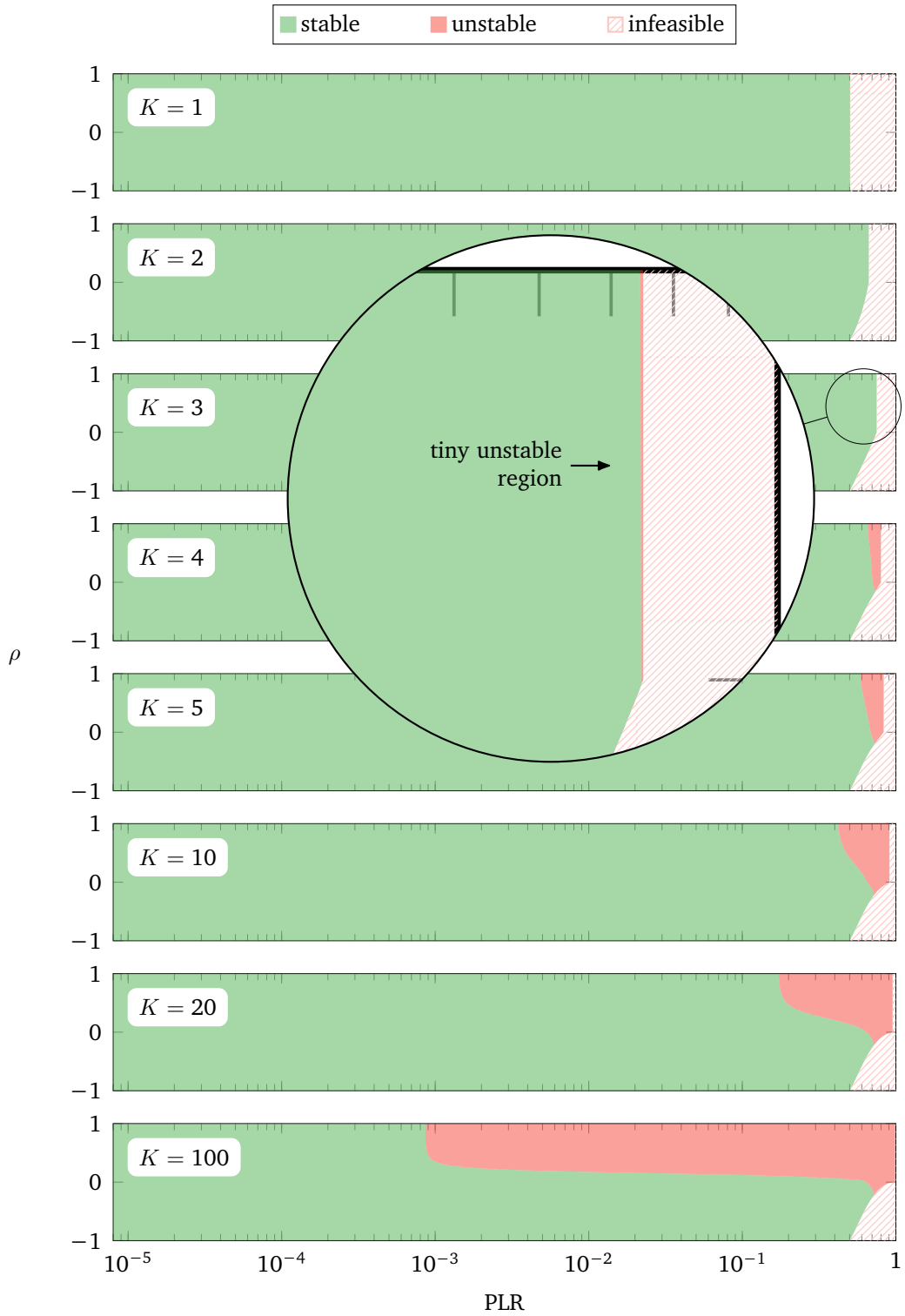


Fig. 3.8.: Stability boundaries for MJLS for the ZOH/ZOH case.

Results for FOH/ZOH case

For the FOH/ZOH case, the results are depicted in Fig. 3.9. Again for $K \in \{1, 2\}$, an unstable region does not exist, which means that all feasible (PLR, ρ) combinations will lead to a stable system. However, for $K > 2$, the system becomes unstable faster compared to the ZOH/ZOH case, which can be explained by the FOH signal reconstruction filter, which linearly extrapolates to estimate the next value. While this limits the "spikyness" of the control signal for short packet loss sequences – since the FOH filter is better at estimating the near future than the ZOH filter – the estimation accuracy deteriorates quickly for longer packet loss sequences because the estimation uncertainty grows. This is because for long PLSL, the FOH filter will estimate the next position value based on previously estimated values. While this behavior is expected and can be improved in a straight-forward fashion by, e.g, employing higher-order filters or other filter classes (Wiener, Kalman, etc.), the focus of this chapter is not to optimize the signal reconstruction filters but rather outline a path to derive proper network QoS for existing control architectures.

Fig. 3.9 emphasizes the performance improvement of networked closed-loop control systems that can be achieved by reducing packet loss correlation. The beneficial effect of low packet loss correlation can also be visualized in the time domain through extensive simulations of the AGV traversing the examplary trajectory shown in Fig. 3.7. 10^6 different runs are simulated, each with a different packet success/failure sequence realization generated in accordance with Eq. (3.25). For $K = 3$, the PLR is kept constant at a rather high value of PLR $= 30\,\%$, while $\rho \in \{1, -0.9\}$ is varied (see red marks in Fig. 3.9 for operating points). The results are displayed in Fig. 3.10a-3.10b. Therein, the red boundary constitutes the maximum value generated in any of the 10^6 simulations, the blue boundary the minimum value, and the black line the reference case of zero packet losses. Note that the trajectory deviation is not zero also for zero packet losses due to the AGV motor's inertia.

Tab. 3.3.: Comparison of the control signal RMS. Negative packet loss correlation reduces jerkiness significantly.

PLR [%]	ρ	$P_{+30\,\%}$	η_{agv}
30	1	0.998	2.01
30	−0.5	0.014	1.11
30	−0.9	0.000 06	1.046
10	1	0.025	1.17
10	−0.5	0	1.02
10	−0.9	0	1.015

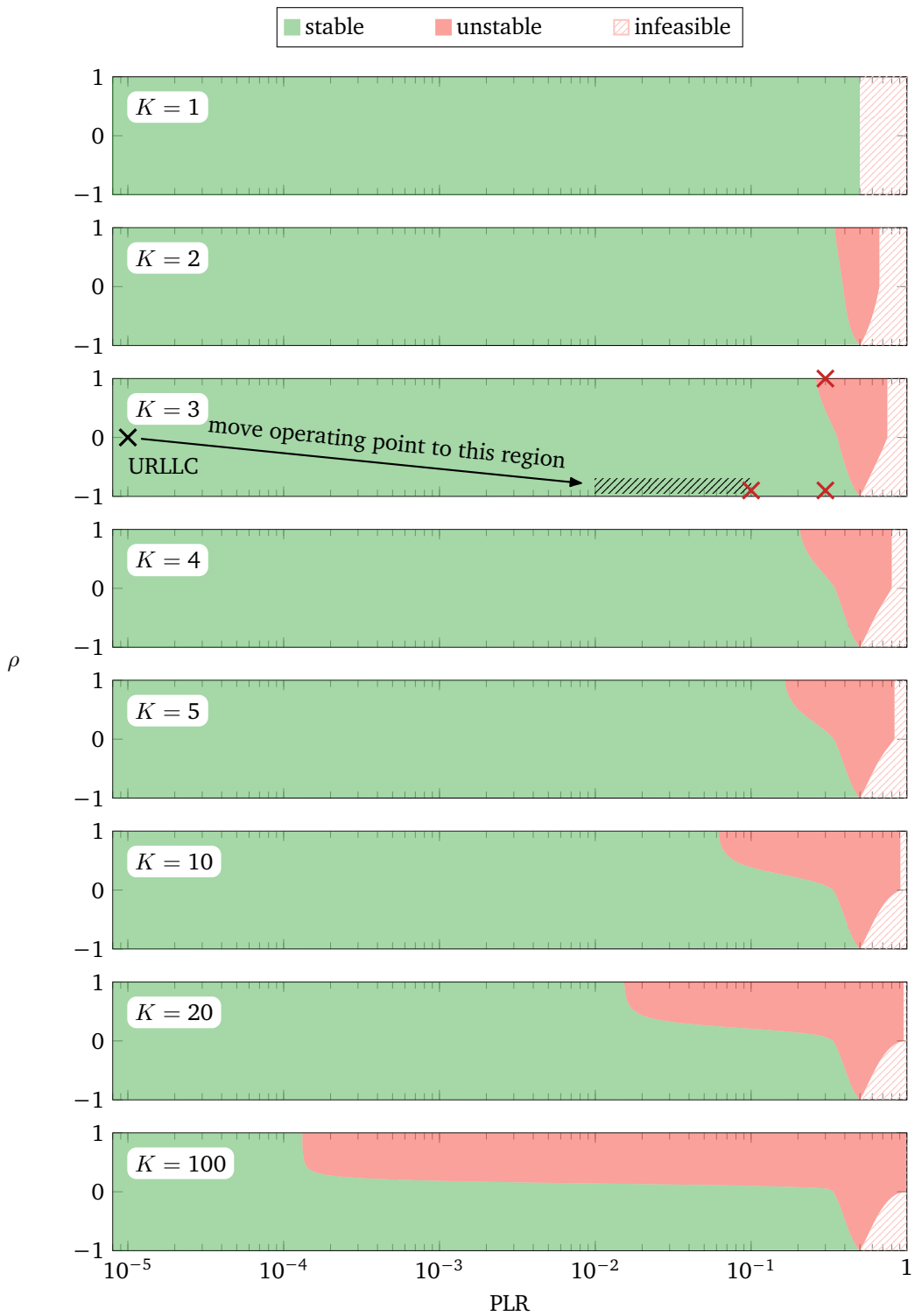
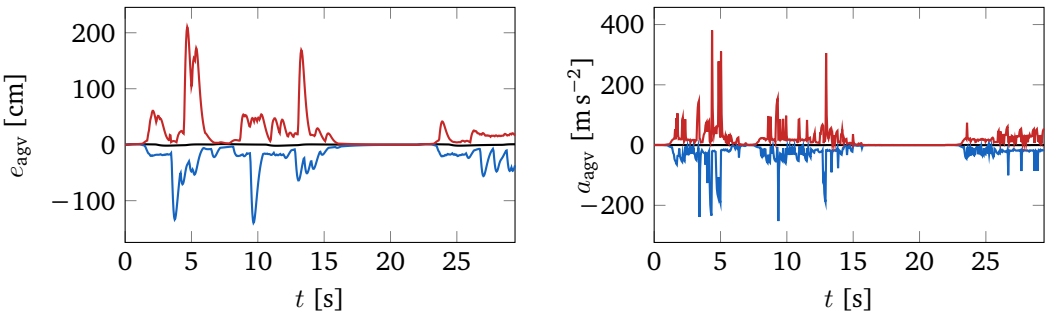
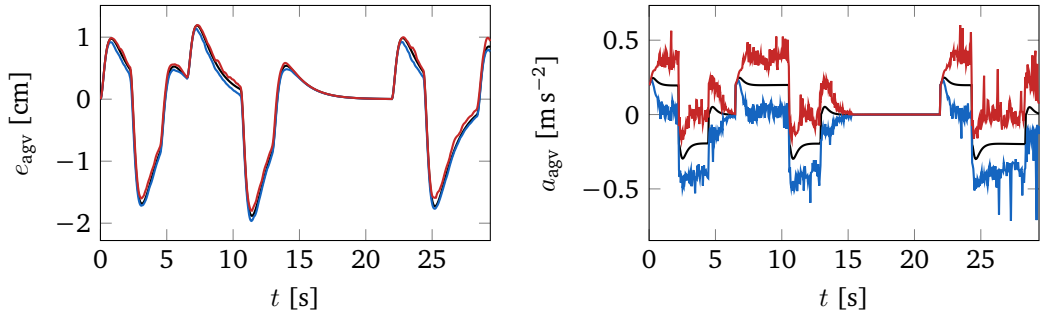


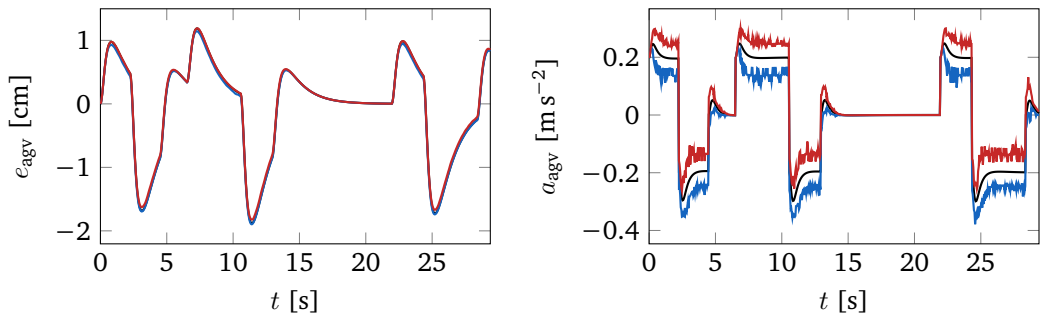
Fig. 3.9.: Stability boundaries for MJLS for the FOH/ZOH case. The settings marked in red are simulated extensively in the time-domain, see Fig. 3.10. The fundemantal idea of this thesis is to move the network's operating point from $(\text{PLR}, \rho) = (10^{-5}, 0)$ (URLLC design target) to the highlighted area with moderate PLR, yet highly negative packet loss correlation.



(a) PLR = 30 %, $\rho = 1$; the high packet loss correlation causes unrealistic control accelerations a_{agv}.



(b) PLR = 30 %, $\rho = -0.9$; a reduction of the packet loss correlation while keeping the PLR constant significantly reduces the control acceleration a_{agv} and position deviation e_{agv}.



(c) PLR = 10 %, $\rho = -0.9$; reducing the PLR makes the control performance nearly undistinguishable from the zero-error case (black curve).

Fig. 3.10.: Simulation results for the AGV use case. The position error e_{agv} and control input a_{agv} (acceleration) are depicted with their upper (red) and lower (blue) bound for 10^6 simulation runs.

As predicted by the MJLS theory, the combination $(\text{PLR}, \rho) = (30\,\%, 1)$ leads to high state deviations and the applied accelerations are unrealistic (Fig. 3.10a). For decreasing ρ, the "spikyness" of the control acceleration and the position uncertainty reduce significantly, compare Fig. 3.10b and note the axis. At $(\text{PLR}, \rho) = (30\,\%, -0.9)$, the position error is almost completely deterministic although still 30 % of all packets are lost. Naturally, the determinism is further increased at $\text{PLR} = 10\,\%$ (still with $\rho = -0.9$, see Fig. 3.10c), which constitutes a realistic long-term packet loss value for state-of-the-art best-effort networks and a good operating point for high spectral efficiency [Eur17; Fra+21].

Tab. 3.3 compares the RMS of the transient control signal a_{agv} with the RMS of the transient control signal without packet losses $a_{\text{agv,ref}}$ in terms of the metrics η_{agv} and $P_{+30\,\%}$, which were defined in Eqs. (3.30) and (3.31), respectively. For $\text{PLR} = 30\,\%$, it is shown that decreasing $\rho = 1 \rightarrow -0.9$ significantly reduces $\eta_{\text{agv}} = 2.01 \rightarrow 1.046$. At the same time, $P_{+30\,\%}$ is lowered from almost all simulation runs exceeding a +30 % RMS increase (99.8 %) to almost no runs (0.006 %). Lowering the packet loss ratio to $\text{PLR} = 10\,\%$, η_{agv} is only 1.5 % higher on average compared to the case of no packet losses and $P_{+30\,\%} = 0$.

3.4 Conclusions

The widely established design goal of URLLC to reduce packet losses to $\text{PLR} < 10^{-5}$ may result in significant spectral inefficiency. The analysis conducted in this chapter has shown that similar QoC can be achieved by state-of-the-art wireless systems that feature PLRs in the range 1 % to 10 % as long as packet losses are (highly) *negatively correlated*. This means that while the probability of the first packet loss can be moderately high, subsequent consecutive packet losses should be avoided. This relates strongly to keeping the AoI low, which is a network KPI not considered in current networks. The motivation for the subsequent chapters of this thesis is to increase the networked control system's stability margin and overall control performance by ensuring the negative packet loss correlation on a network level. **Proposition** [Graphically, this entails shifting the network's operating point from $(\text{PLR}, \rho) = (10^{-5}, 0)$, the design goal of URLLC, to the marked area in Fig. 3.9 of moderate packet losses with highly negative packet loss correlation, e.g, $(\text{PLR}, \rho) = (0.1, -0.9)$.]

4 Modeling Control-Communication Failures

Parts of the work in this chapter were first published in [Sch+20b] and [Sch+21]. The author's personal contributions comprise (a) the development of the Markov models, which translate communication failures to control system failures, and (b) the selection of related reliability KPIs.

The previous chapter showed that a negative packet loss correlation, i.e., reducing consecutive packet losses, significantly increases control performance KPIs because it can continuously keep the information within the loop "fresh". This might enable to operate closed-loop control systems over spectrally efficient wireless networks with PLRs in the range of 1 % to 10 % instead of 10^{-5} with only minimal impact on control performance.

This chapter first highlights the potential causes of packet loss and possible mitigation techniques. Then, a network architecture is presented that can realize such abstract negative packet loss correlation with existing wireless network technologies, focused on industrial environments. The boundaries of applicability are outlined and the usefulness of the stated assumptions is validated. With the network assumptions in mind, two AoI-based models are presented that map communication failures (packet losses) to control system failures. The first relates to single-hop networks consisting only of one transmit (TX)/receive (RX) pair; the second extends this to the dual-hop case, where two TX/RX pairs exist. Lastly, reliability KPIs are introduced that adequately describe relevant system performance aspects.

4.1 Communication Assumptions

As opposed to wire-line communication, the wireless channel suffers from multiple effects that hamper data transmission.

Path loss describes the RX power attenuation that results from the distance between TX and RX antenna as well as the general environment, e.g., indoor vs. outdoor.

Large-scale fading describes the medium-term variations in RX power due to blockage of dominant multipath components, e.g., by moving obstacles like trucks or people.

Small-scale fading describes the short-term variations in RX power caused by multi-path propagation and Doppler shifts.

Interference is caused by multiple users transmitting at the same time, at the same place, over the same frequency.

There exist a wide range of engineering solutions to mitigate their impact in today's communication networks. For instance, path loss may be addressed by the deployment of small cells that significantly reduce the distance between base station and user equipment [LW16]. Large-scale fading can be effectively addressed through automatic gain control in the analog circuitry of the RX device. Interference may be addressed through a channel access scheme, which might be done centrally by the base station (BS) or multiple BSs (cellular) or decentrally through schemes like carrier-sense multiple access (WLAN, Bluetooth, etc.). The focus of this thesis, however, is small-scale fading as a cause of failure because it is particularly difficult to mitigate and has a significant impact on reliability in the context of URLLC and targeted applications [Höß20].

4.1.1 Small-Scale Fading as a Cause of Failure

When radio waves propagate, they are typically subject to multi-path propagation (caused by reflections and scattering) and the Doppler effect. This causes the amplitude, phase, and multipath delay to rapidly fluctuate over a short period of time and distance [Rap02]. The constructive interference of multipath signals at a receiver leads to a strong RX signal while destructive interference causes signal attenuation. Extreme signal attenuation is termed *deep fade*, which degrades the RX signal by several orders of magnitude and might disrupt connectivity. The variation of small-scale fading in time is characterized by the coherence time T_c, which is predominantly determined by the Doppler effect, i.e., the temporal change of the communication channel.

$$T_\mathrm{c} \approx \frac{1}{f_\mathrm{d}}\sqrt{\frac{9}{16\pi}} \approx \frac{0.423}{f_\mathrm{d}} \quad \text{[Rap02]} \tag{4.1}$$

Therein, $f_\mathrm{d} = \frac{v_\mathrm{max}}{c} f_\mathrm{c}$ denotes the maximum Doppler shift, v_max the maximum relative velocity of scatterers, c the speed of light, and f_c the carrier frequency. The coherence time approximates the time it takes until the communication channel's autocorrelation function decays to 0.5 and therefore indicates how fast the channel changes. The variation of small-scale fading in frequency is characterized by the coherence bandwidth

$$B_\mathrm{c} \approx \frac{1}{5\sigma_\tau} \,. \quad \text{[Rap02]} \tag{4.2}$$

Therein, σ_τ denotes the RMS delay spread, which can be calculated from the power delay profile. The coherence bandwidth approximates the bandwidth over which the communication channel can be considered "flat", i.e., all frequencies exhibit approximately equal gain and phase. Additionally to the equations from basic communications literature, extensive measurement campaigns have been carried out to characterize the communication channel for specific environments in more detail. The work conducted in [Bur+21] is especially noteworthy as this campaign focuses specifically on industrial environments that usually feature more metallic surfaces and therefore reflections. The channel characteristics presented therein yield the "ball park numbers"

$$T_\mathrm{c} = 10\,\mathrm{ms} \qquad \text{and} \qquad B_\mathrm{c} = 10\,\mathrm{MHz}\,. \tag{4.3}$$

Failure Mitigation

The rapid fluctuations of small-scale fading usually make it impossible to dynamically react to poor channel conditions, e.g., by switching channels, waiting for a better transmission opportunity (which might not be an option for low-latency traffic) or adapting the modulation and coding scheme. Therefore, URLLC literature has proposed in recent years to add *diversity*, which is a broad term referring to parallelizing transmissions to increase the probability of success. This can be done through the use of frequency, time, code, and/or space diversity. For closed-loop control systems, diversity in time, i.e., successive transmissions of the same data, spaced by at least the coherence time (which also extends to retransmission schemes like HARQ) – is often not feasible due to tight timing restrictions and/or the data being outdated fast [ZBP01; Öst+19], compare Fig. 1.1. Hence, in this thesis, only diversity schemes supporting *simultaneous* transmissions are considered. To keep the discussion concise, a frequency diversity scheme is assumed [ÖTF15] and the amount of diversity is quantified through the *number of channels*. A generalization to other diversity schemes is possible if the success probability of any given transmission attempt is quantifiably adjustable.

4.1.2 Connectivity Models

The connectivity models presented in this section bundle the assumptions made for (a) small-scale fading as the main cause of failure, (b) the control system architecture of Chapter 3, and (c) current feasibility for diversity-enabled communication in (industrial) standardized 5G networks.

Two major distinctions will be made with regards to the network architecture,displayed in Figs. 4.1 and 4.2.

1. A single-hop architecture is considered, composed of either a UL connection or a DL connection, but not both. This network mode is relevant for wireless sensor networks (only UL), the distribution of control commands (only DL) or closed-loop control systems with one transmission direction that may be approximated as ideal.

2. A dual-hop architecture is considered with two transmissions in either UL or DL, or one successive UL and DL connection. This network mode is relevant for wireless sensor networks with relaying nodes and networked closed-loop control applications.

The term *agent* is introduced to refer to a sensor/plant/actuator in any network configuration.

Generally, the UL is assumed to connect the agent to a network function (e.g., the networked controller) and the DL connects the network function (back) to the agent (in the dual-hop case through an intermediate node). The term *transmission cycle* is introduced to describe the data processing chain

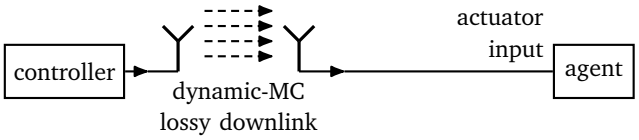
single-hop: data generation $\rightarrow$ UL/DL transm.

dual-hop: data generation $\rightarrow$ UL/DL transm. $\rightarrow$ processing $\rightarrow$ UL/DL transm. .

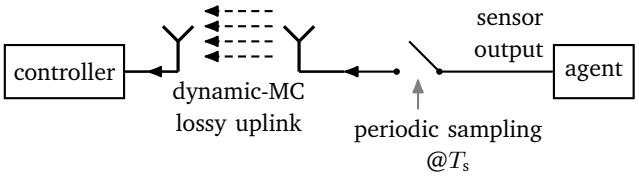
Each transmission cycle is periodically triggered with the fixed sampling period T_s, which is common for industrial applications [5G 20] and has been thoroughly motivated in Sec. 3.1.2.

It is assumed that the network manages channel access and that a time-frequency grid exists similar to that in 5G.

The agent's sensor sampling action is assumed to be synchronized with the time-frequency grid, fulfilling the time synchronicity requirement stated in Sec. 1. Consequently, the waiting time between sampling and transmitting is negligible, contrary to many available publications that were mentioned in Chapter 2. Signal processing

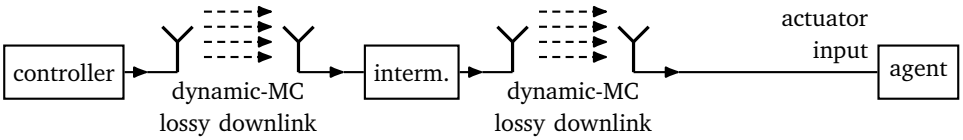


(a) DL transmission.

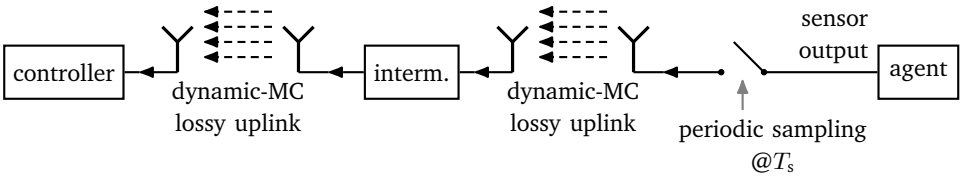


(b) UL transmission.

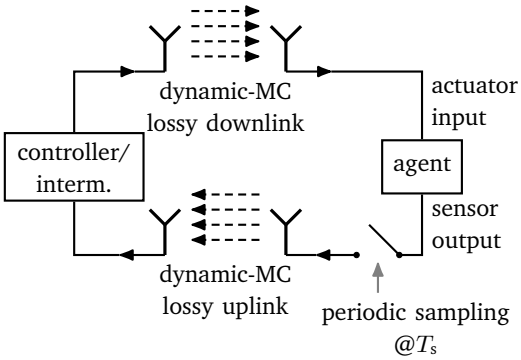
Fig. 4.1.: Single-hop connectivity models.



(a) Two successive DL transmissions.



(b) Two successive UL transmissions.



(c) Closed-loop example with one UL and one successive DL transmission.

Fig. 4.2.: Dual-hop connectivity models.

and computation times are assumed to be constant and also the DL transmission is assumed to be time-synchronized through pre-reserved resources. Hence, for the dual-hop network case, the overall network latency – given that both transmissions succeed – can be approximated through adding two transmission time intervals and the intermediate node's computation time budget, which also includes a *known* frame alignment waiting time in the second-hop transmission. This *fixed* latency will not be of further interest because it can be parameterized to be very low, as can be seen, e.g., in 5G networks [3GP21a].

Furthermore, it is assumed that wireless resources are assigned to the agent through SPS [Pop+19; 3GP21c] from a pre-reserved resource pool [CSP14] that needs to be shared among multiple co-existing agents. The term *channel* shall henceforth be used to describe one transmission opportunity within this resource pool. It is assumed that an agent is capable to connect to the nwtork through $l \in \{1, \ldots, \hat{l}\}$ parallel channels in both, UL and DL, which constitutes a well-known MC approach. The number of channels for a given transmission may be adjusted on a per-transmission basis, as assigned by the network. Any set of channels that are assigned to the agent for a single transmission are assumed to have a frequency spacing larger than the coherence bandwidth B_c and the packet interarrival time is assumed larger than the coherence time T_c of the underlying small-scale fading process. Given today's high available bandwidths for local usage [Bun20; Bun21], the coherence bandwidth given in Eq. (4.3) can be readily exceeded. The coherence time, on the other hand, can be artificially mitigated through, e.g., channel hopping between channels that have a channel spacing larger than the coherence bandwidth [Höß+20]. Consequently, all transmissions can be regarded as *independent in both frequency and time*, which is essential for the derivation of appropriate control system failure models.

For analytical tractability it is assumed that all channels feature a fixed per-channel packet loss probability p_loss. To combine all l parallel data streams, a selection combining scheme is considered because of its low complexity, which allows the combination of channels in higher network layers, e.g., through cyclic redundancy checks. More sophisticated combining schemes that require combination on low network layers, such as maximum ratio combining or joint decoding [Wol+17], are expected to yield even better results than shown in this thesis.

To ensure correct operation, control information must be exchanged over the network. Since the resource allocation is managed by the network, it must know whether transmission attempts succeeded or not. In the UL, the success/failure of packet transmissions can be evaluated in a straight-forward manner by the network-side receiver. In the DL, however, the network relies on ACKs for this information. In

a similar fashion, the network is able to control changes to the resource assignment in the DL easily, yet relies on reliable transmissions regarding the currently valid UL resource assignments, which the agent needs to receive before initiating the next transmission cycle. For the beginning, ideal transmissions are assumed for both, because in today's networks, signaling information can be protected with low-rate powerful codes and have thus a PLR orders of magnitude lower than that of data packets. Hence, the impact of erroneous network control information is expected to be minimal.

All mentioned assumptions are in accordance with current 5G standardization [3GP21a].

4.2 Failure Models

As has been motivated in Chapters 2 and 3 and contrary to many other works related to co-design of control application and (wireless) communication (CoCoCo), the network of the previous section is unaware of the control system's dynamic behavior (i.e., its state-space matrices are unknown to the network). This means that the network is incapable of making a *control-aware* resource allocation optimization that incorporates the information on the system's dynamic behavior into an estimation of how valuable a certain data packet is, compare, e.g., [Aya+19; Eis+19]. Instead, the AoI – a *network* KPI – is utilized as a direct indicator. While this is a "low-hanging fruits" approach and does not yield an optimal control-communication trade-off, the AoI constitutes a possible compromise between performance on the one hand and network complexity and implementability on the other hand.

It is particularly noteworthy that for the stated connectivity assumptions, the AoI and the number of consecutive packet losses are identical for the single-hop network because there is only a single point of failure. For the dual-hop network, the more general AoI metric must be applied directly. The following sections detail the AoI-aware failure modeling for both network architectures.

4.2.1 Single-hop network

The joint modeling of control systems and packet losses in Chapter 3 revealed that packet losses, i.e., failures in the communications domain, do not necessarily lead to a control system failure. Many control systems inherently tolerate packet losses to some extent as long as not *too many* occur consecutively. What constitutes *too*

many is application-specific and described through the *survival time* τ_{surv} [3GP19b; 5G 20], which defines the time period that a control system or sensor network may prolong operation without receiving an expected packet (with the expectation stemming from the periodicity of packet transmissions). With the sampling period T_{s}, a maximum number of tolerable packet losses K may be defined:

$$K = \left\lfloor \frac{\tau_{\text{surv}}}{T_{\text{s}}} \right\rfloor \quad \text{with} \quad K \in \mathbb{N}_0 \,. \tag{4.4}$$

This constitutes a mapping of (communications-related) packet losses and (control-system-related) failures. It is assumed that if $K+1$ packets are lost consecutively, the control system switches off or enters an emergency mode to avoid damage or maybe, more severely, human harm. The specific value of K varies depending on (a) the control system under investigation, (b) the required control performance, and (c) the sampling period. As has been demonstrated in Sec. 3.3.4, the value of K might also increase dramatically if certain network QoS may be guaranteed, e.g., low packet loss rates and/or low packet loss correlation. However, from a practical viewpoint, it is advisable to settle for conservative K as not all causes of failure can be modeled and safety is crucial. Conservative estimations for K may be generated by the *Reduced Sampling* approach, which was presented in Sec. 3.3.3.

Based on the connectivity assumptions of Sec. 4.1 and the proposed mapping above, the Markov failure model depicted in Fig. 4.3 is proposed.* Analogously to the Markov chain presented in Sec. 3.5, which modeled the operation modes of an MJLS, the agent is considered in state s_k with $k \in \{0, \ldots, K\}$, where k denotes the number of consecutive packet losses that occurred immediately before entering s_k. For instance, in s_2 the last two transmissions failed. Consequently, when a transmission succeeds (green transitions), the state s_0 is entered. Whenever a packet is lost (red transitions) with transition probability $\tilde{p}_k$, the state index is incremented by one. The rightmost state s_{d} denotes the failure state and refers to $K+1$ consecutive packet losses. All other states are considered *up*. Note that in the Markov model of Sec. 3.5 there was no failure state because the focus was to determine control system stability *given* that the maximum number of consecutive losses is K, whereas here, the focus is to determine (and ultimately decrease) the probability of losing $K+1$ consecutive packets. The failure state s_{d} is *absorbing* since it is assumed that the control system switches off after $K+1$ consecutive packets have been lost.

*Please note that henceforth, the term *state* will refer to a state in the Markov chain failure model unless otherwise mentioned. We highlight this explicitly to avoid confusion with the *control states*, which were frequently mentioned in Chapter 3.

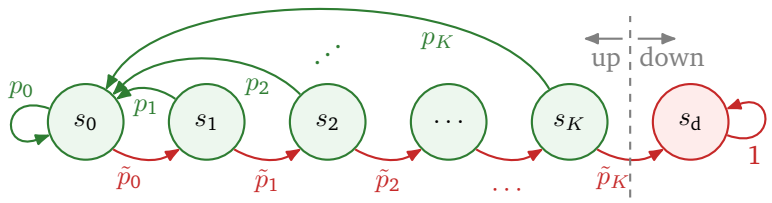


Fig. 4.3.: Single-agent failure model for single-hop network.

4.2.2 Dual-hop network

The failure model of a multi-hop network may also be modeled by the failure model of Fig. 4.3 if it is assumed that information can only traverse the communication system if all of the single transmissions have been successful, i.e., is expressed through a logical AND. For the dual-hop network case, this would mean that data packets are only received by the destination if both transmit links succeed within a transmission cycle. The modeling approach considered here instead offers a higher AoI granularity because the intermediate node is assumed to always try forwarding the most current data in an attempt to decrease the AoI at the destination. For example, if at t the first-hop transmission was successful but the second-hop transmission has failed and at $t+1$, the first-hop transmission has failed, the second-hop transmission will attempt to forward data based on the information provided at t (the most current).

The resulting failure model is captured by the Markov model depicted in Fig. 4.4 and a low-order example ($K = 1$) is given in Fig. 4.5. Therein, each state $s_{k_{1\mathrm{h}}|k_{2\mathrm{h}}}$ refers to having an AoI of $k_{1\mathrm{h}}T_\mathrm{s}$** at the intermediate node and an AoI of $k_{2\mathrm{h}}T_\mathrm{s}$** at the destination. E.g., in $s_{1|1}$, the currently available information at both, intermediate node and destination, is one sampling period old. As before, green transitions refer to successful transmissions, red transitions refer to unsuccessful transmissions. States encircled with a solid line refer to the AoI situation right before a first-hop transmission attempt and are termed *first-hop states* in the following. States encircled with a dashed line describe the AoI situation within a transmission cycle, i.e., *after* the first hop but *before* the second hop, and are termed *second-hop states*. All second-hop states are marked additionally through a dash in the second index to highlight that a transmission cycle is in progress and the data is about to be transmitted on the second hop, which will update the AoI at the destination.

**Due to the periodic sampling assumption, the multiplication with T_s is omitted in the remainder of this thesis and the AoI will be referred to as a multiple of the sampling period T_s for brevity.

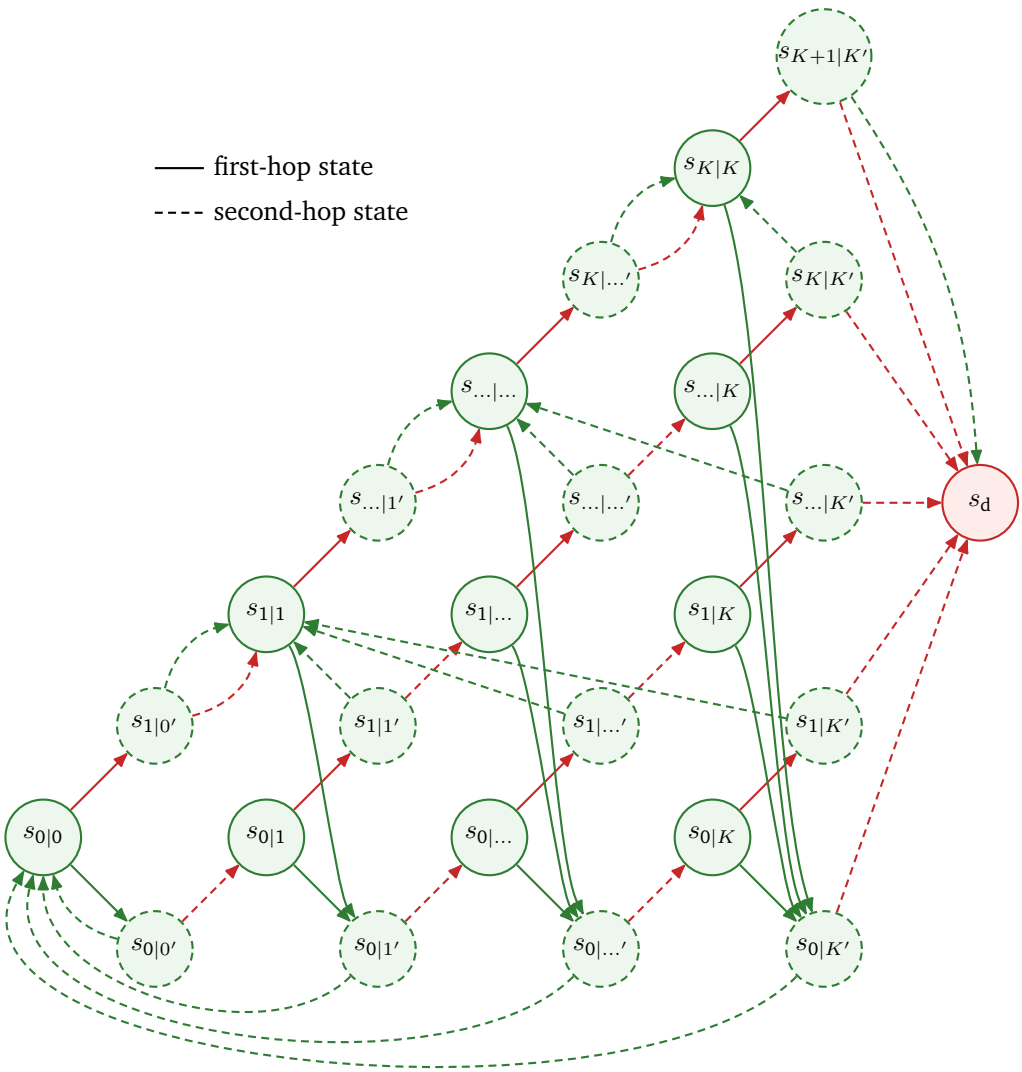


Fig. 4.4.: Single-agent failure model for dual-hop network.

The transition probabilities are omitted in Fig. 4.4 for readability, but are indexed similarly to the single-hop case. Successful transmissions originating in $s_{i|j}$ and $s_{i|j'}$, respectively, are denoted by a transition probability of $p_{i|j}$ and $p_{i|j'}$, failed transmissions occur with a probability $\tilde{p}_{i|j} = 1 - p_{i|j}$ and $\tilde{p}_{i|j'} = 1 - p_{i|j'}$.

The intermediate second-hop states are motivated by the possibility that the information about the success/failure of the first-hop transmission may be used to adjust the success probability (by providing more/fewer channels) for the subsequent second-hop transmission with the same transmission cycle. This is possible in this model because it is assumed that the time between the first and second hop, during which the intermediate node processes the signal, suffices to make a resource assignment decision for the second-hop transmission.

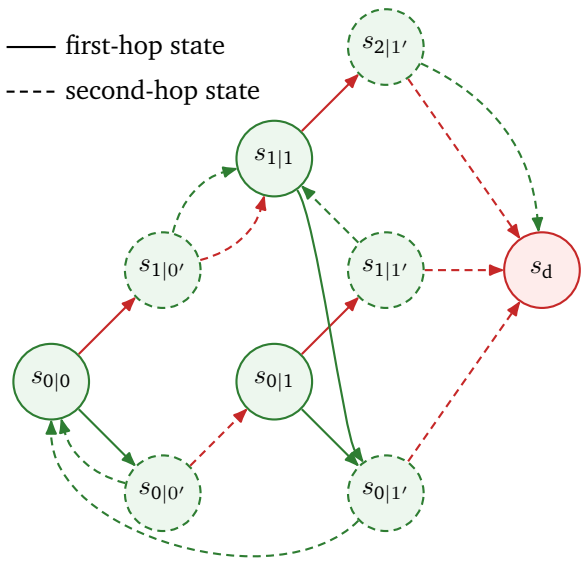


Fig. 4.5.: Single-agent failure model example for dual-hop network with $K = 1$.

It is apparent that the AoI at the intermediate node cannot exceed the AoI at the destination in between transmission cycles (i.e., for all first-hop states) because the destination is behind the intermediate node in terms of information flow. Hence, the Markov model is composed of a triangle structure. The AoI dependency of the second hop on the first hop also manifests itself through the fact that after a successful second-hop transmission, the destination AoI $k_{2\mathrm{h}}$ can only drop to the current AoI at the intermediate node $k_{1\mathrm{h}}$ instead of zero.

The determining KPI for an agent's failure is the AoI at the destination $k_{2\mathrm{h}}$. Originally, the peak AoI is defined as the maximum time elapsed since the last received update at the destination [CCE16]. With this definition, a peak AoI violation probability may be calculated [Dev+19]. Following this motivation, the peak AoI is adopted as a KPI in this thesis and refers to the control system's maximum tolerable AoI and, hence, the system is assumed to enter the absorbing down state s_{d} if

$$k_{2\mathrm{h}} > \text{peak AoI} \ , \tag{4.5}$$

at which point the control system is assumed to switch off.

It is noteworthy that in certain states, there is no updated information available at the intermediate node, which makes any subsequent second-hop transmission obsolete. For example, in $s_{1|0'}$ the most recent first-hop transmission has failed and the destination already has all the information that the intermediate node has. This is indicated in Fig. 4.4 through a double transition (green+red) as the second-hop transmission success/failure has no impact on the AoI at the destination. This is also

true for $s_{K+1|K'}$, at which point the application is already condemned to fail after the next transmission (the state, however, is included for completeness).

Proposition [Through this extended modeling, it becomes clear how the AoI extends the single-hop concept of "consecutive packet losses" to dual-hop/multi-hop networks and is therefore the more general concept. While the AoI is able to fully capture the consecutiveness of packet losses when applied to single-hop networks, it generalizes the concept to "information freshness" in dual-hop and multi-hop systems.]

4.3 Performance Metrics

In dependability-related wireless communications literature, the term *availability* is used to describe the ratio of time in which a particular radio service works as intended, which commonly resembles the PLR [Höß+19] or, more generally, the probability of successfully receiving a transmitted packet. This terminology is especially useful for processes that converge to a meaningful steady-state, which are characterized by a non-zero probability of leaving a *down* (failure) state (and re-entering an *up* state). For instance, the stochastic nature of outages experienced due to small-scale fading, which are most commonly investigated as cause of failure in URLLC research [Höß+17], leads to the classical PLR as a meaningful steady-state performance metric. Processes, however, that do not transition out of the failure state such as the models presented in Fig. 4.3 and 4.4 will evidently not lead to meaningful steady-state results. Processes of this kind are better described through time-based metrics, such as the MTTF.

Hence, the agent-related MTTF will be introduced in the following as well as – for comparison – the classical communications-related PLR. Building on this, a closed-form solution for the average number of assigned channels $\bar{l}$ will be derived. In addition, the status update age (SUA) and PLSL will be introduced to characterize the AoI first from the agent's viewpoint and then from the network's viewpoint.

4.3.1 Mean Time to Failure

Through the failure model of Fig. 4.3, the agent-related MTTF for the single-hop case can be derived as

$$\mathrm{MTTF}_{\mathrm{sh}} = T_{\mathrm{s}} \boldsymbol{e}_0 \boldsymbol{N}_{\mathrm{sh}} \mathbf{1}^{\intercal} \ , \tag{4.6}$$

when initialized in s_0, with $\boldsymbol{e}_0 = [1\ 0\ \dots\ 0]$, $\mathbf{1} = [1\ 1\ \dots\ 1]$ and $\boldsymbol{N}_{\text{sh}} = (\boldsymbol{I} - \boldsymbol{Q}_{\text{sh}})^{-1}$ as the fundamental matrix of the absorbing Markov chain [GS12]. Therein, $\boldsymbol{Q}_{\text{sh}}$ denotes the transition probability matrix of all transient states, i.e., all states whose exit probability is greater than zero. It is extracted from the transition probability matrix $\boldsymbol{P}_{\text{sh}}$ of the Markov model via (all zeros are omitted for readability)

$$\boldsymbol{P}_{\text{sh}} = \left[\begin{array}{cccc|c} p_0 & \tilde{p}_0 & & & \\ p_1 & & \tilde{p}_1 & & \\ \vdots & & & \ddots & \\ p_K & & & & \tilde{p}_K \\ \hline & & & & 1 \end{array}\right] = \begin{bmatrix} \boldsymbol{Q}_{\text{sh}} & \boldsymbol{R}_{\text{sh}} \\ \mathbf{0} & 1 \end{bmatrix} . \tag{4.7}$$

As each state s_k is associated with a number of channels l_k, the transition probabilities can be calculated as

$$\tilde{p}_k = {p_{\text{loss}}}^{l_k} \quad \text{and} \quad p_k = 1 - \tilde{p}_k \ . \tag{4.8}$$

For the dual-hop case (Fig. 4.4), a factor of $\frac{1}{2}$ must be introduced to Eq. (4.6) to capture the fact that the Markov model in Fig. 4.4 is comprised of both, first-hop states and second-hop states and, hence, two transitions occur within one sampling period T_{s}.

$$\text{MTTF}_{\text{dh}} = \frac{1}{2} T_{\text{s}} \boldsymbol{e}_0 \boldsymbol{N}_{\text{dh}} \mathbf{1}^{\intercal} \tag{4.9}$$

4.3.2 Packet Loss Ratio

For comparison, also the traditional, communications-related long-term average PLR is introduced. Note that the PLR refers to the loss rate after combining, thus different from p_{loss}. For the single-hop case, the vector

$$\boldsymbol{\pi}_{\text{sh}} = \frac{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sh}}}{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sh}} \mathbf{1}^{\intercal}} \tag{4.10}$$

captures the probability of being in each transient state during operation (after a short initial settling time). Via

$$\text{PLR} \approx 1 - \pi_0 \ , \tag{4.11}$$

the PLR can be derived, where π_0 is the probability of being in state s_0, hence, the first element of $\boldsymbol{\pi}_{\text{sh}}$. The intuition is that whenever a state different from s_0 is

entered, a packet must have been lost. The approximation stems from initializing the Markov chain in s_0.

For the dual-hop case, the equations differ slightly. While the state probabilities may be derived analogously,

$$\boldsymbol{\pi}_{\text{dh}} = \frac{\boldsymbol{e}_0 \boldsymbol{N}_{\text{dh}}}{\boldsymbol{e}_0 \boldsymbol{N}_{\text{dh}} \mathbf{1}^\intercal} \ , \tag{4.12}$$

the individual first-hop and second-hop state probabilities are of particular interest. The first-hop state probabilities, which denote the long-term probability of being in a certain first-hop state after a transmission cycle has ended, may be directly derived from $\boldsymbol{\pi}_{\text{dh}}$.

$$\boldsymbol{\pi}_{\text{1h}} = 2\boldsymbol{S}_{\text{1h}} \boldsymbol{\pi}_{\text{dh}} \tag{4.13}$$

Therein, $\boldsymbol{S}_{\text{1h}}$ is a selection matrix for all first-hop states composed of 1's and 0's to select all first-hop state probabilities $\pi_{k_{\text{1h}}|k_{\text{2h}}}$ (solid states in Fig. 4.4) from $\boldsymbol{\pi}_{\text{dh}}$. The factor of 2 conditions the state probabilities on being in a first-hop state. Similarly, the second-hop state probabilities may be derived, which state the long-term probability of being in a certain second-hop state during a transmission cycle.

$$\boldsymbol{\pi}_{\text{2h}} = 2\boldsymbol{S}_{\text{2h}} \boldsymbol{\pi}_{\text{dh}} \tag{4.14}$$

$\boldsymbol{\pi}_{\text{2h}}$ comprises all individual second-hop state probabilities $\pi_{k_{\text{1h}}|k_{\text{2h}}'}$ (dashed states in Fig. 4.4).

The PLR for the first-hop transmissions may be calculated through

$$\text{PLR}_{\text{1h}} \approx 1 - \sum_{i=0}^{K} \pi_{0|i} \ , \tag{4.15}$$

where the summation is performed over those elements of $\boldsymbol{\pi}_{\text{1h}}$ that correspond to the first-hop states in the bottom row of Fig. 4.4. These states are reached if and only if a successful UL transmission occurred. The PLR for the second-hop transmissions may be calculated similarly through

$$\text{PLR}_{\text{2h}} \approx 1 - \sum_{i=0}^{K} \pi_{i|i} \ . \tag{4.16}$$

Here, the summation is performed over those elements of $\boldsymbol{\pi}_{\text{1h}}$ (sic!) that correspond to the first-hop states of the top left diagonal as these states are reached if and only if a successful second-hop transmission occurred.

4.3.3 Average Number of Assigned Channels

For the single-hop case, let $\boldsymbol{l}$ denote a cost vector that gathers the number of channels l_k assigned in each state s_k. Then, the average number of parallel channels $\bar{l}$ used throughout operation can be calculated as

$$\bar{l} = \boldsymbol{l}\boldsymbol{\pi}^{\mathsf{T}} \ . \tag{4.17}$$

For the dual-hop case, let $\boldsymbol{l}_{1\mathrm{h}}$ and $\boldsymbol{l}_{2\mathrm{h}}$ denote cost vectors that gather the number of channels for each first-hop and second-hop transmission, respectively. Then, the average number of channels assigned in first-hop and second-hop transmissions, respectively, can be calculated as

$$\bar{l}_{1\mathrm{h}} = \boldsymbol{l}_{1\mathrm{h}}\boldsymbol{\pi}_{1\mathrm{h}}{}^{\mathsf{T}} \tag{4.18}$$

$$\bar{l}_{2\mathrm{h}} = \boldsymbol{l}_{2\mathrm{h}}\boldsymbol{\pi}_{2\mathrm{h}}{}^{\mathsf{T}} \ . \tag{4.19}$$

4.3.4 Age of Information

Two AoI-related metrics are introduced as KPIs:

Status update age (SUA) is defined as [CCE14]

$$\mathrm{SUA} = t - t' \ , \tag{4.20}$$

where t represents the current absolute time and t' denotes the generation time of the last successfully received packet. For multi-hop transmissions, t' denotes the packet generation time at the origin. The SUA's probability mass function (PMF) describes the age distribution of the most recent successfully received packet. Originally, SUA was defined as a time-continuous random variable [CCE14], however, as our connectivity model assumes periodic operation, the SUA is only evaluated immediately after a transmission cycle and, therefore, it is a discrete-time quantity. For the single-hop case, the PMF of SUA is equal to the state probabilities $\boldsymbol{\pi}$, i.e.,

$$f_{\mathrm{SUA,sh}}(k) = \pi_k \quad \text{with} \quad k \in \{0, \ldots, K\} \ , \tag{4.21}$$

since the Markov model was specifically designed to incorporate the SUA information. For the dual-hop case, the PMF may be constructed through

$$f_{\text{SUA,dh}}(k_{\text{2h}}) = \sum_{i=0}^{K} \pi_{i|k_{\text{2h}}} \quad \text{with} \quad k_{\text{2h}} \in \{0, \ldots, K\} \,, \tag{4.22}$$

which consitutes a column-wise summation of first-hop state probabilities.

Packet loss sequence length (PLSL) refers to the number of consecutive packet losses. The PLSL's PMF describes the number of failed transmissions between two successful transmissions and may also take the value 0 for consecutive successful transmissions. It can be derived directly from the Markov model of Fig. 4.3 by calculating the loop probabilities of all loops originating in s_0.

$$f_{\text{PLSL}}(k) = \begin{cases} p_k \prod_{\kappa=0}^{k-1} \tilde{p}_\kappa & \text{for} \quad k \in \{0, \ldots, K\}, \\ \prod_{\kappa=0}^{K} \tilde{p}_\kappa & \text{for} \quad k = K+1. \end{cases} \tag{4.23}$$

The PLSL is only defined for the single-hop case because the concept of consecutiveness is unintuitive for multiple hops.

SUA evaluates the AoI from the agent's standpoint while PLSL views it from a communication perspective.

4.4 Conclusions

In this chapter, an industrial network architecture is presented in a realistic *Industry 4.0* scenario. Two possible network variations are considered: (a) single-hop (one TX/RX pair) and (b) dual-hop (two TX/RX pairs). As packet losses do not necessarily lead to a control system failure, a relation between the two must be established to analyze control system failures due to communication failures. For this, the AoI is used, which constitutes a network metric that captures information freshness and is, therefore, a highly relevant metric for control systems. With the assumed network architecture, the number of consecutive packet losses directly translates to the AoI for the single-hop case. Any non-zero tolerance against consecutive packet losses offers an exploitable degree of freedom. The relation between the number of consecutive packet losses and a control system failure, which ultimately occurs if the number of consecutive packet losses exceeds the mentioned packet loss tolerance, is formulated through a Markov model. For the dual-hop case, the AoI must be used

directly as a design KPI. Another Markov model is introduced that captures the AoI interdependencies between first-hop and second-hop transmissions. Building upon these models, a range of KPIs is derived: the state-dependent availability, the agent's mean time to failure, the long-term average packet loss ratio, the average number of assigned channels, the status update age, and the packet loss sequence length.

Using the models and KPIs of this chapter, an optimization towards increasing the mean time to failure and reducing the required number of parallel channels is presented and performed in the next chapter.

5 Single Hop – Single Agent

The work in this chapter was previously published in [Sch+19], [Sch+20b] and [Sch+21]. The author's personal contributions comprise the derivation of sensible resource allocation schemes and evaluation of the resulting data.

In the previous chapter, a failure mapping of (communications-related) packet losses to (application-related) failures was presented. Two main distinctions regarding the network architecture were made: (a) a single-hop network and (b) a dual-hop network. The single-hop network will be the focus of this chapter while the dual-hop case will be covered in Chapter 7. As all discussions and results in this chapter refer to the single-hop case, the index will be dropped from any variables for readability without loss of conciseness.

The main goal of this chapter is to find dynamic MC resource allocation schemes that increase the MTTF of the control system while ensuring a low average number of MC channels, ideally kept as close to $\bar{l} = 1$ (single-connectivity) as possible. Only a single agent is considered, i.e., system-induced resource shortages are not considered here but rather in Chapter 6.

This chapter is structured as follows. First, SARA will be presented as dynamic MC approach to realize negative temporal packet loss correlation. As the suitability of channel assignment schemes highly depends on the targeted KPIs values, a range of sensible SARA schemes will be introduced. Subsequently, the (closed-form) results will be presented and discussed. Lastly, the impact of erroneous ACKs on the KPIs is investigated, which are otherwise assumed ideal.

5.1 State-Aware Resource Allocation

The fundamental idea behind SARA is to dynamically alter the number of parallel channels depending on the current SUA, which effectively temporally negatively correlates packet losses. An increasingly negative correlation increasingly reduces the probability of long packet loss sequences, which has a significantly beneficial effect on control performance, as was demonstrated in Chapter 3.

The number of parallel channels that can be assigned simultaneously to a single agent is considered to be limited to $\hat{l} = 4$ in this thesis because each parallel demodulation at the receiver requires a costly PHY signal processing chain in hardware. Also, limiting $\hat{l}$ facilitates to achieve uncorrelated packet transmissions as all parallel channel assignments must be spaced at least by the coherence bandwidth for the failure models to be valid, see Sec. 4.1.

The following notation is applied: Adaptation schemes that follow a regular pattern are denoted as $\boldsymbol{l}_i^j$, with i indicating the base number of channels, i.e., the number of channels allocated after a successful transmission; and j indicating the number of channels added for each consecutively lost packet. Whenever a packet is transmitted successfully, the number of channels is reset to the base value of the scheme. $\boldsymbol{l}_{l_{\text{fix}}}^0$ corresponds to a MC approach with l_{fix} fixed channels, termed *static* schemes in the following (URLLC baseline for comparison). The schemes that are considered in this thesis are summarized in Tab. 5.1. Note that also two schemes are considered that do not follow a regular pattern. The first is termed $\boldsymbol{l}_{\text{rf}}$, with the subscript standing for *relax-full* since the first packet loss does not trigger a channel increase, while the second consecutive packet loss triggers the full amount of parallel channels available. The second is termed $\boldsymbol{l}_{1\text{f}}$. Therein, a full allocation (the maximum number of parallel channels) is applied when $k = K$ and for decreasing k, the number of parallel channels are steadily decreased by one; however in any case, a single channel will be assigned for $k = 0$.

Also note that as k increases, the number of assigned channels for $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_2^1$ is clipped to $\hat{l} = 4$ at $k = 4$ and $k = 3$, respectively. In Tab. 5.1, when $K < 4$, the schemes are accordingly truncated, e.g., when $K = 2$ for $\boldsymbol{l}_1^1$, the link assignment will be $1 \rightarrow 2 \rightarrow 3$. To ensure a fair comparison regarding the MTTF, the sampling period T_{s} is adjusted to K such that $(K+1)T_{\text{s}} = 180\,\text{ms}$, which roughly corresponds to the maximum sampling period of Sec. 3.3.3, see Tab. 5.2. While this ensures a fair performance comparison in terms of dependability, it is emphasized that the control system smoothness will decrease when T_{s} increases and it is up to the control engineer to specify an upper bound on T_{s} (lower bound on the smoothness) [FWP97]. In digital control, the guideline for good stability and smoothness is to oversample by 10-30 times, see [FWP97] for details. While this thesis is limited to the schemes presented in Tab. 5.1, the analysis can be performed for any values of i, j or K and also for other arbitrary schemes that are not considered here.

Tab. 5.1.: State-aware resource allocation – example schemes.

		consec. lost packets k					K	class
		0	1	2	3	4		
number of channels l_k	$\boldsymbol{l}_1^0$	1					0	static
		1	1				1	
		1	1	1			2	
		1	1	1	1		3	
		1	1	1	1	1	4	
	$\boldsymbol{l}_2^0$	2					0	
		2	2				1	
		2	2	2			2	
		2	2	2	2		3	
		2	2	2	2	2	4	
	$\boldsymbol{l}_3^0$	3					0	
		3	3				1	
		3	3	3			2	
		3	3	3	3		3	
		3	3	3	3	3	4	
	$\boldsymbol{l}_1^1$	1	2				1	dynamic
		1	2	3			2	
		1	2	3	4		3	
		1	2	3	4	4	4	
	$\boldsymbol{l}_2^1$	2	3				1	
		2	3	4			2	
		2	3	4	4		3	
		2	3	4	4	4	4	
	$\boldsymbol{l}_{\mathrm{rf}}$	1	1	4			2	
		1	1	4	4		3	
		1	1	4	4	4	4	
	$\boldsymbol{l}_{1\mathrm{f}}$	1	4				1	
		1	3	4			2	
		[corresponds to $\boldsymbol{l}_1^1$]					3	
		1	1	2	3	4	4	

Tab. 5.2.: Sampling period vs. packet loss tolerance inter-dependency. [Sch+19]

K	0	1	2	3	4
T_{s}	180 ms	90 ms	60 ms	45 ms	36 ms

5.2 Results

The dynamic resource allocation of SARA can improve the control system's MTTF by orders of magnitude compared to the static multi-connectivity baseline, while simultaneously keeping the average channel consumption low. Tab. 5.3 shows the analytical results of the introduced performance metrics PLR, MTTF, and $\bar{l}$ for $K = 3$ tolerable consecutive packet losses at a sampling period $T_s = 45$ ms and a per-channel packet loss probability $p_{\text{loss}} = 10\,\%$, which constitutes a realistic block error rate for spectrally efficient, state-of-the-art best-effort networks [Eur17; Fra+21]. It is observed that the MTTF improves by 100x from 8 weeks to 16 years (which is the appropriate range for mobile robots according to [3GP19a]) for the SARA scheme $\boldsymbol{l}_1^1$ compared to static dual-connectivity $\boldsymbol{l}_2^0$, while only using approximately half the amount of parallel channels on average (1.09 compared to 2). For $\boldsymbol{l}_1^1$, the resulting PLR remains over 9 %. This clearly shows that when the network is AoI-aware, even high resulting PLR values do not necessarily lead to an AoI-induced control system failure. While the PLR for $\boldsymbol{l}_2^0$ is approximately one order of magnitude lower than for $\boldsymbol{l}_1^1$, the MTTF is 100x lower. **Proposition** [This shows that a low number of base channels and a dynamic increase of channels *only when it matters* simultaneously decreases the average resource consumption and increases the control system's MTTF by *orders of magnitude*. Another interpretation of this is that SARA *temporally negatively correlates* the occurrence of packet losses.] That is, the packet loss sequence length is kept short by reacting to packet losses and dynamically increasing the number of channels used in the next transmission. While this does not reduce the PLR significantly, it reduces the probability of long packet loss sequences by orders of magnitude.

This potential is also demonstrated in Fig. 5.1, which depicts the performance of different adaptation schemes with varying per-channel packet loss probabilities p_{loss} for $K \in \{0, 1, 3\}$. At $K = 1$ and a target MTTF of 10 years, $\boldsymbol{l}_2^1$ requires only $p_{\text{loss}} = 10\,\%$ ($\bar{l} = 2.01$) compared to static MC scheme $\boldsymbol{l}_2^0$, which requires $p_{\text{loss}} = 3\,\%$ ($\bar{l} = 2.00$) and therefore might lead to a significant coding overhead. Intuitively, at $K = 3$ (or in general at higher K values), the susceptibility to high p_{loss} values decreases as all curves move further to the right. **Proposition** [For low p_{loss} values, the shown curves diverge, which shows that SARA gains can be enhanced through combination with moderately low p_{loss}.] For example, when comparing $\boldsymbol{l}_1^0$ ($\bar{l} = 1$) and $\boldsymbol{l}_{\text{rf}}$ ($\bar{l} \approx 1.03$) at $K = 1$ and $p_{\text{loss}} = 1\,\%$, the MTTF can be increased from $\approx$15 minutes to $\approx$28 years, a factor of $\approx$1 million through increasing the average channel consumption by only 3 %.

Tab. 5.3.: Resulting performance metrics for $K = 3$ tolerable packet losses and $p_{\text{loss}} = 10\,\%$. For MTTF, the sampling period is $T_s = 45$ ms.

scheme	average packet loss rate PLR	mean time to failure MTTF	average cost $\bar{l}$
$\boldsymbol{l}_1^0$	100.0×10^{-3}	8 minutes	1.00
$\boldsymbol{l}_2^0$	10.0×10^{-3}	8 weeks	2.00
$\boldsymbol{l}_3^0$	1.0×10^{-3}	1.4k years	3.00
$\boldsymbol{l}_1^1$	91.7×10^{-3}	16 years	1.09
$\boldsymbol{l}_2^1$	9.9×10^{-3}	14 millennia	2.01
$\boldsymbol{l}_{\text{rf}}$	99.1×10^{-3}	16 years	1.03

The diagram shows that when the control system's tolerance against packet losses is high ($K = 3$) and $\boldsymbol{l}_1^1$ is used, even p_{loss} values between 20 % and 30 % enable MTTF $\in$ [2 weeks, 10 years]. To achieve MTTF = 2 weeks for the static single-connectivity scheme $\boldsymbol{l}_1^0$, which exhibits a similar average channel consumption $\bar{l} = 1$ compared to $\bar{l} = 1.09$ ($\boldsymbol{l}_1^1$), the per-channel packet loss probability would need to take values $p_{\text{loss}} < 3\,\%$, which might require a significant amount of coding overhead, and therefore spectral inefficiency. Additionally, it becomes clear that there exists a large benefit from increasing the control system's tolerance K when comparing $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_2^0$. For example, increasing K from 3 → 4 at $p_{\text{loss}} = 20\,\%$ increases the MTTF from 1 week → 10 years (500x) for $\boldsymbol{l}_1^1$ while it increases only from 8 hours → 5 days (15x) for $\boldsymbol{l}_2^0$.

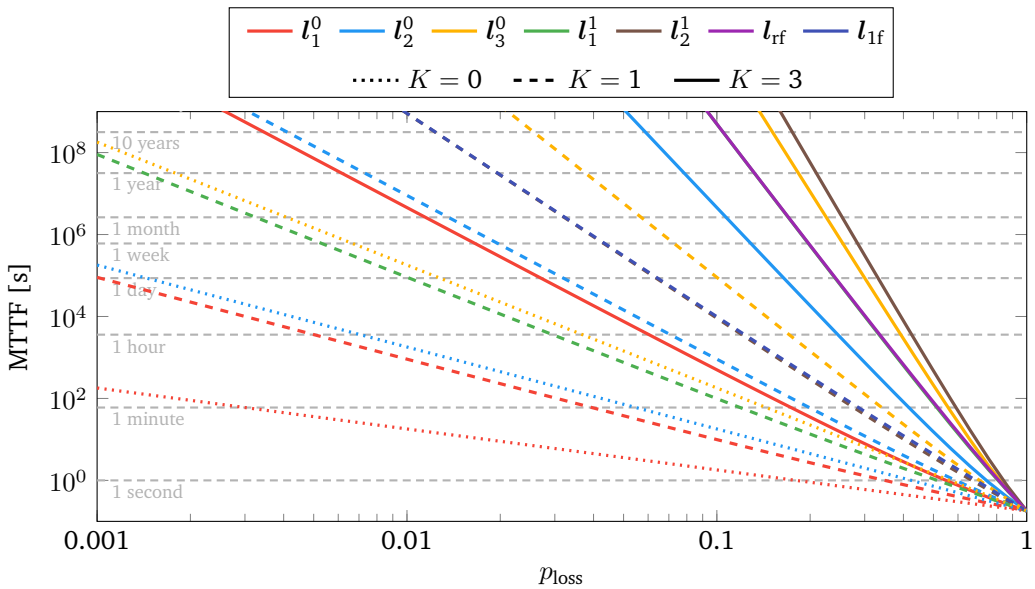
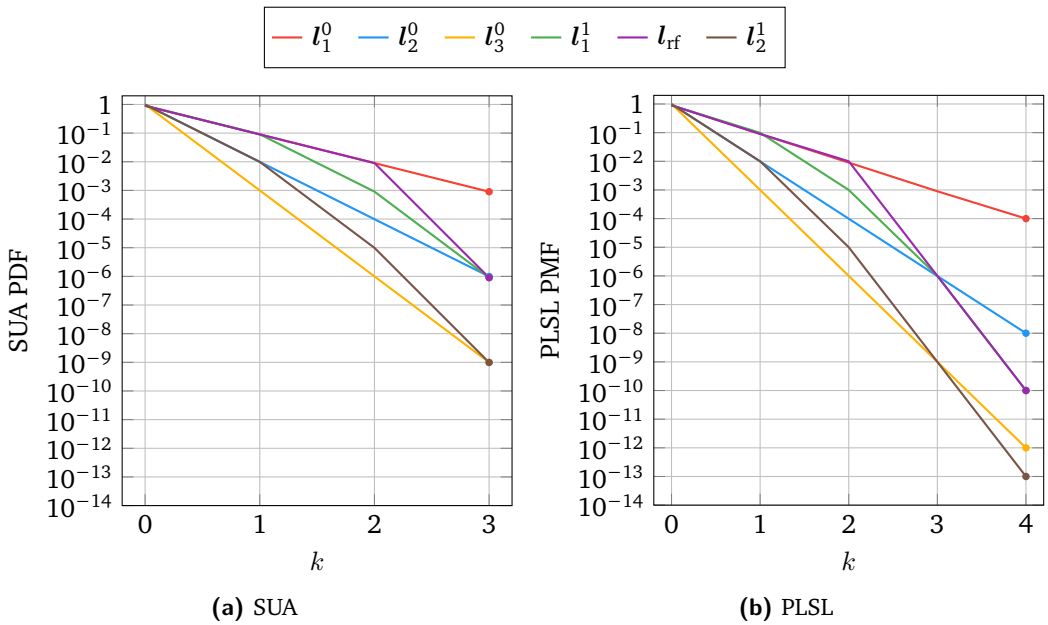


Fig. 5.1.: Performance of different SARA schemes regarding p_{loss} and MTTF.

The SUA and PLSL PMFs of all considered resource allocation schemes are depicted in Fig. 5.2 for $K = 3$. Note that while PLSL also includes a value for $k = K + 1$, the SUA is only defined until $k = K$. It is observed that first, the difference of SUA and PLSL is negligible because $p_{\text{loss}} = 10\,\%$ is significantly smaller than 1 and, hence, all state visits of states s_k for $k \in \{1, \ldots, K\}$ transition into s_0 with probability $1 - 10^{-l_k}$, which is always close to one. Second, the ratio of $\text{SUA}(k)/\text{SUA}(0)$ and $\text{PLSL}(k)/\text{PLSL}(0)$ for $k \neq 0$ can be explained approximately through the number of transmission opportunities. For example for $\boldsymbol{l}_1^1$, $\frac{\text{SUA}(3)}{\text{SUA}(0)} \approx 10^{-6}$ because to reach s_3, $1 + 2 + 3 = 6$ transmissions must have failed. Hence, through setting the number of parallel channels l_k in each state s_k, a certain distribution for SUA and PLSL will be achieved. For instance when comparing $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_{\text{rf}}$, packet loss sequences of length $k = 2$ are obtained in 1/100 for $\boldsymbol{l}_{\text{rf}}$ while they only occur in 1/1000 (ten times less frequent) for $\boldsymbol{l}_1^1$ due to assigning two channels in s_1 instead of one. **Proposition** [It can also be observed that SARA in general performs a PMF reshaping that allows a few number of consecutive packet losses but ensures a limit with an extraordinarily high probability (concave curves).]



(a) SUA **(b)** PLSL

Fig. 5.2.: AoI metrics for $K = 3$. (Note that for better readability, the presentation is slightly inaccurate; the SUA probability density function (PDF) should take the form of a stair plot while the PLSL PMF consists of a set of Dirac functions.)

5.3 Erroneous Acknowledgments

For the failure model of Sec. 4.2 to be valid, it was assumed that the scheduler in the network is aware of the success/failure of each downlink transmission, i.e., acknowledgments were considered ideal. This section is dedicated to determining the impact of erroneous ACKs on the resource allocation scheme and therefore determine the validity of the assumption. The updated connectivity model is depicted in Fig. 5.3. A fixed ACK outage probability $\tilde{p}_{\text{ack}}$ is considered.

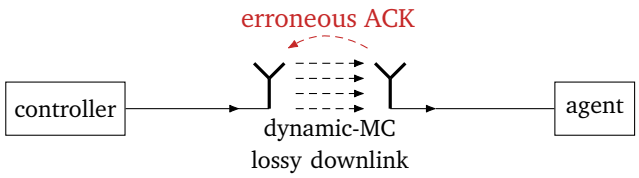


Fig. 5.3.: Single-hop connectivity model with erroneous ACKs.

Transmitting ACK information instead of negative acknowledgment (NACK) constitutes a conservative approach. While this increases communication overhead, ACKs provide a better "reactivity" of the network against errors in both, the payload and ACK transmission. In case of a lost NACK, a lost packet may remain undetected and the network would not be able to respond to a transmission failure. As a consequence, the MTTF would drastically decrease, which is unacceptable for industrial use cases. Missing/erroneous ACKs have the same effect to the network as a lost packet in terms of resource allocation. Hence, choosing ACKs instead of NACKs always yields a worse SUA assumption than the actual SUA.

The Markov failure model depicted in Fig. 5.4 captures these thoughts. An agent is modeled to be in $s_{i,j}$ if the network assumes that i consecutive packets have been lost while in reality j consecutive packets have been lost. Red transitions indicate that a transmission has failed while blue arrows indicate that the transmission has succeeded but the correponding ACK has failed. In both cases, the network assumes that the packet was not received successfully (*downward* direction in the model), triggering more resources for the next transmission according to the underlying SARA scheme. The conservativeness of the approach is visualized through the triangular shape of the model because $i \geq j, \forall i, j$. The undermost row comprises a set of states $s_{\infty,j}$, in which the network assumes that the agent has already failed while it actually has not. It is assumed that the network will continue to provide the agent with the maximum number of parallel channels as long as it does not actively confirm that it has failed.

The extreme cases are:

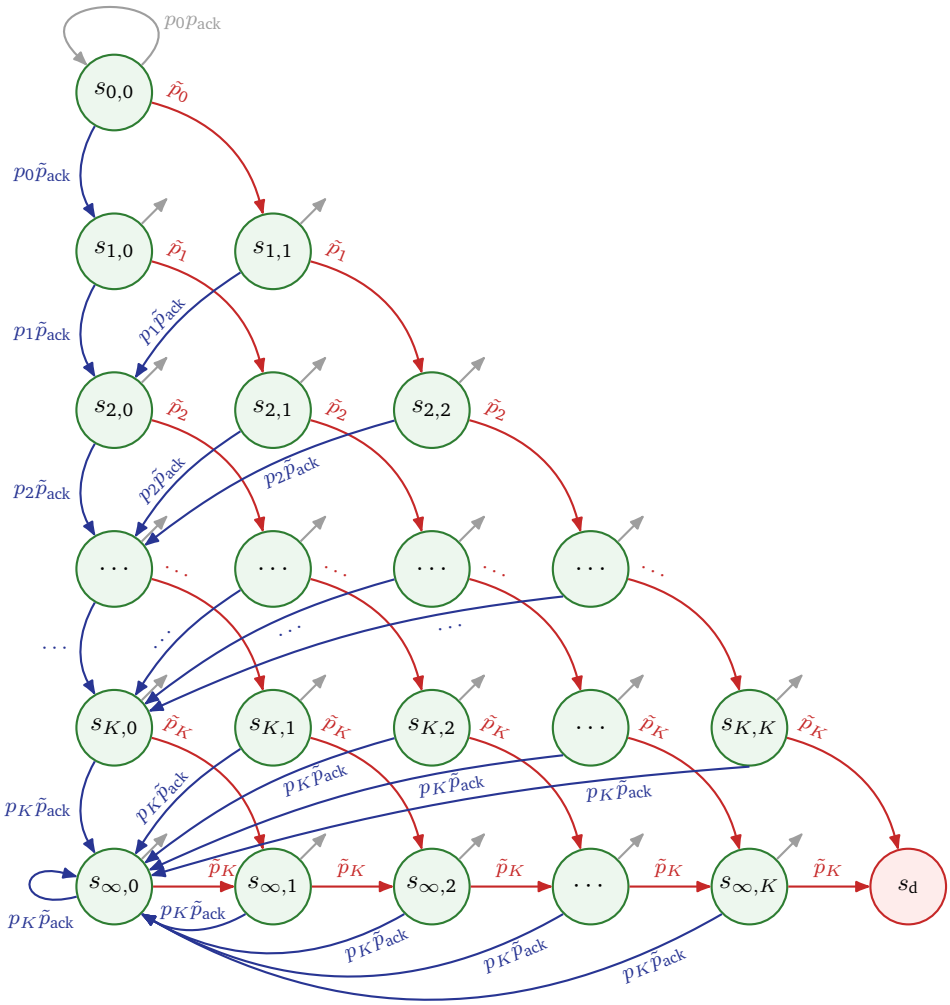
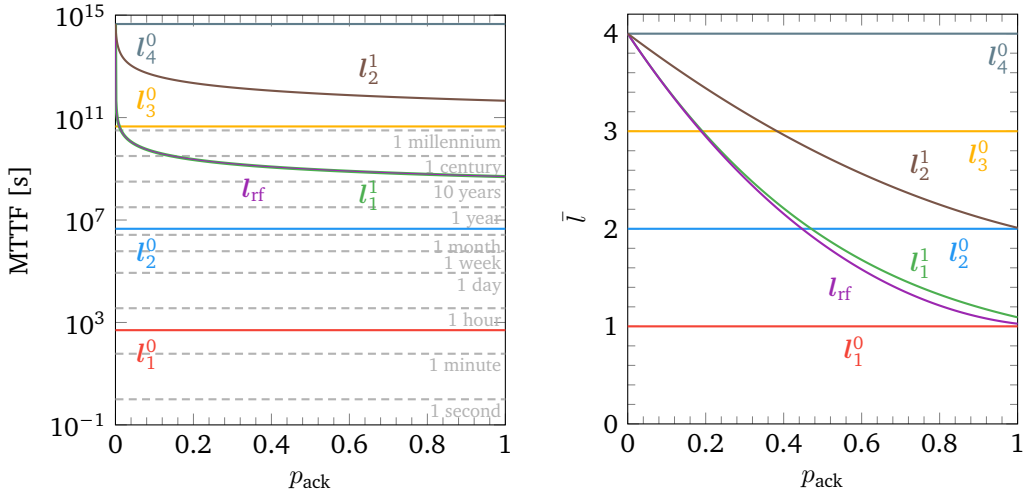


Fig. 5.4.: Single-hop failure model with erroneous ACKs. The transitions for successful transmissions are only hinted at (gray) to avoid visual clutter: all *up* states $s_{i,j}$ (green) also feature a transition to $s_{0,0}$ with probability $p_i p_{\text{ack}}$.

1. If $p_{\text{ack}} = 1$, i.e., all ACKs are successfully received, the model in Fig. 5.4 devolves into the model depicted in Fig. 4.3, only passing through the states with $i = j$ (the top-right diagonal).

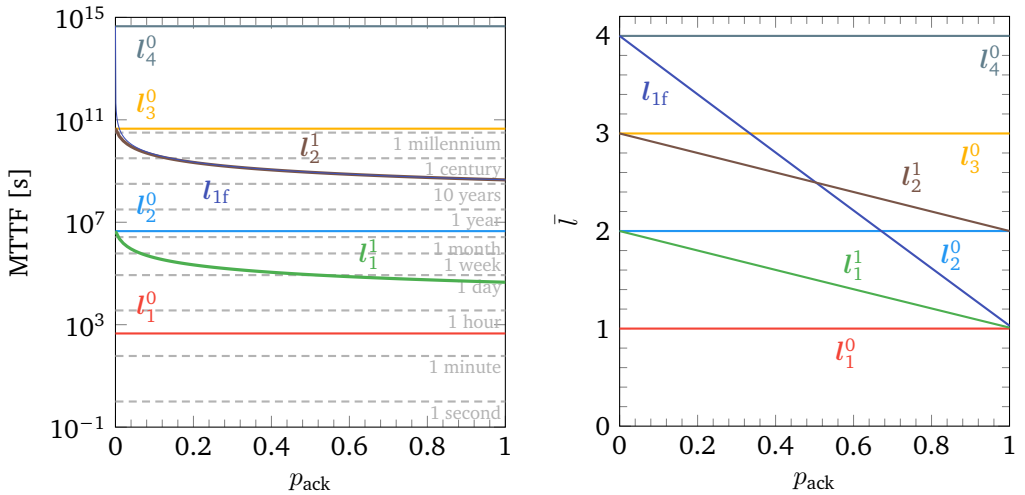
2. If $p_{\text{ack}} = 0$, i.e., all ACKs are erroneous, the system devolves into static MC, always assigning the maximum number of parallel channels $\hat{l}$. The system, which is initialized in $s_{0,0}$ will initially traverse downwards and eventually become stuck in the undermost row.

The methodology of solving the Markov chain is very similar to the case with ideal ACKs and will not be repeated. Selected results are depicted in Fig. 5.5. The probability of successfully receiving an ACK is varied from $0.5 \leq p_{\text{ack}} \leq 1$ in all



(a) MTTF development for $p_{\text{loss}} = 10\,\%$, $K = 3$

(b) $\bar{l}$ development for $p_{\text{loss}} = 10\,\%$, $K = 3$



(c) MTTF development for $p_{\text{loss}} = 1\,\%$, $K = 1$

(d) $\bar{l}$ development for $p_{\text{loss}} = 1\,\%$, $K = 1$

Fig. 5.5.: MTTF and $\bar{l}$ development for $0 \leq p_{\text{ack}} \leq 1$. As expected, the increase in $\bar{l}$ is low especially for $p_{\text{ack}} \geq 0.9$, which is easily achievable through low-rate powerful codes [Sha+18].

cases. On the left side, the MTTF development is shown, on the right side the $\bar{l}$ development. Two observations stand out:

1. Because of the conservative approach, which assumes a transmission failure in case of erroneous acknowledgments and consequently increases the number of assigned channels, the MTTF is lowest for $p_{\text{ack}} = 1$. Consequently, the average channel consumption $\bar{l}$ is also lowest at this point.

2. The sensitivity towards erroneous ACKs is low. Even at a comparatively low $p_{\text{ack}} = 90\,\%$, $\bar{l}$ only increases by approximately 10 % for all presented SARA schemes at $K = 3$ and $p_{\text{loss}} = 10\,\%$. **Proposition** [This underlines SARA's robustness against erroneous acknowledgments and, therefore, state uncertainties.] At $p_{\text{loss}} = 1\,\%$ and a relatively low packet loss tolerance $K = 1$, the sensitivity is slightly higher, however, still low. Because ACKs can be successfully transmitted with high probability through the use of low-rate powerful codes (easily $> 90\,\%$ [Sha+18]), the simplification of neglecting the impact of erroneous acknowledgments is justified well and upheld within all further chapters of this thesis.

5.4 Conclusions

SARA performs an AoI-aware resource allocation fulfilling two major goals. First, the dependability of the real-time application is increased by orders of magnitude compared to best-effort single-connectivity. Second, the average number of assigned resources is kept very close to single-connectivity. For instance, at $K = 3$, the average number of assigned channels is $\bar{l} = 1.09$ and therefore remains within a 10 % increase compared to single-connectivity. At the same time, SARA's MTTF is 16 years in this scenario, which constitutes a 10^6x improvement over single-connectivity (MTTF = 8 minutes) and still a 100x gain over dual-connectivity (MTTF = 8 weeks at two channels on average). When assuming common ACK transmission reliabilities from current wireless networks, SARA's sensitivity to erroneous ACKs is low, increasing the resource consumption only by a few percent. All in all, this demonstrates the high potential of negatively correlating packet losses for low-cost dependability improvements in industrial scenarios.

6 Single Hop – Multiple Agents

The work in this chapter was previously published in [Sch+20a] and [Sch+21]. The author's personal contributions comprise the mapping of the single-agent failure model to a system failure model, the computational complexity considerations, the derivation of meaningful system-level KPIs, the verification through extensive system-level simulations, as well as the evaluation of all (closed-form) results.

While for a single-hop network and a single agent the system's dependability benefits of SARA were demonstrated to be profound (see previous chapter), the question arises how SARA performs when applied on the system level where (a) the number of agents $M > 1$ and (b) the number of channels, which are available for assignment to M agents, is limited to L_{av}. As all agents in the system traverse their individual Markov chain of Fig. 4.3, a situation might occur in which the total number of channels in the system does not suffice to assign all requested* channels, which might drastically reduce the system-wide MTTF.

One strategy to avoid any situation with too few channels is deploying $\hat{l}M$ or more channels as this will guarantee that all agents will always be assigned the number of channels they request. However, this is certainly wasteful, especially considering the wide range of up to $\hat{l} = 4$ parallel channels for each agent. Additionally, if $\hat{l}M$ channels were available in the system, one might raise the question why the agents do not employ static MC with $\hat{l}$ channels to fully utilize the available channels. Reducing the number of available channels in the whole system to $L_{\text{av}} < \hat{l}M$ will entail not being able to assign all requested channels sometimes, which consequently will affect the system-wide MTTF. Hence in this chapter, a system extension of the model presented in Fig. 4.3 is developed that incorporates multiple agents and a limited number of available channels in the system. It demonstrates how SARA's applicability with regards to statistical multiplexing can be exploited to ensure a low number of required channels L_{av} while maximizing the system-wide MTTF.

*Here, the term "request" is used to refer to the number of channels l_k that each agent in s_k is supposed to receive according to the ideal SARA schemes presented in Tab. 5.1. This should not be confused with a classical transmission request of cellular networks.

6.1 Failure Model

A system of M homogeneous agents is considered, each of which operates according to the Markov chain depicted in Fig. 4.3. This means that each agent is able to tolerate up to K consecutive packet losses before failing. A superimposed Markov chain is introduced that agglomerates all individual agent states to a single system state

$$S_{|s_0|,|s_1|,\ldots,|s_\mathrm{d}|} \ . \tag{6.1}$$

Thereby, $|s_k|$ denotes the number of agents that currently reside in state s_k, consequently, $|s_\mathrm{d}| + \sum_{k=0}^{K} |s_k| = M$. Please note that for the system state only the number of agents in each state s_k matters and not the specific set of agents. While each agent transitions between its individual states according to the referenced Markov chain (and thereby whether packets are received successfully or not), this is also reflected in the system state.

The system is defined *down* as soon as the first agent fails, i.e., $|s_\mathrm{d}| > 0$, else it is *up*. All *down* states are collapsed to a single *down* state S_d and the last index in Eq. (6.1), corresponding to $|s_\mathrm{d}|$, is dropped from the notation as it will take the value 0 for all *up* states. For more concise notation, a linear index $i \in \{0, \ldots, Z_\mathrm{up} - 1\}$ is introduced for all *up* system states, with Z_up denoting the number of *up* system states. Thereby, all system states $S_{|s_0|,|s_1|,\ldots,|s_K|}$ are sorted in descending order by their subscripts from left to right and mapped to S_i, resulting in

$$\begin{aligned} S_{M,0,\ldots,0} &= S_0 \\ S_{M-1,1,0,\ldots,0} &= S_1 \\ S_{M-1,0,1,0,\ldots,0} &= S_2 \\ &\vdots \\ S_{0,\ldots,0,M} &= S_{Z_\mathrm{up}-1} \qquad . \end{aligned} \tag{6.2}$$

6.1.1 Admission Control

The system is considered to feature L_av available channels. The assumption that up to $\hat{l}$ parallel channels can be assigned to any agent remains unaffected. As L_av might not suffice to serve all agents according to the underlying SARA scheme, some agents must be underserved in some system states. The system does not have enough parallel channels if $L_\mathrm{av} < L_\mathrm{req}(S_i)$, with $L_\mathrm{req}(S_i)$ as the total number of requested channels (by all agents) in state S_i. Hence, $L_\mathrm{req}(S_i) - L_\mathrm{av}$ channels need

to be "denied" by admission control, constituting a system-induced deviation from the presented SARA resource allocation in Tab. 5.1. Two admission control schemes are considered.

random From the set of all requested channels, L_{av} are randomly selected and assigned; the rest will be denied. From an agent's perspective, this implies a weighting by the number of channels that each agent is requesting, i.e., agents with more requested channels are more likely to be denied channels. Note that whenever an agent was denied a channel, it is still susceptible to be denied another (if it still has at least one) as long as $L_{\text{req}}(S_i) > L_{\text{av}}$. Hence, there might be agents that are denied all requested channels, leading to a guaranteed packet loss and, consequently, a transition $s_k \rightarrow s_{k+1}$ for these particular agents.

cliff Agents in higher individual states (k large) are prioritized over agents in lower individual states (k small). That is, agents that have lost more consecutive packets most recently are prioritized over agents that successfully received a packet more recently. The "cliff" metaphor refers to prioritizing agents closer to "falling off the cliff".

6.1.2 Transition Probabilities

To derive the transition probabilities of the system Markov chain, the individual agent's Markov chain in Fig. 4.3 is revisited. Therein, the state transition probabilities p_k were determined by the selected resource allocation scheme summarized in Tab. 5.1. In the system case, the number of channels for each agent is not only determined through the chosen scheme and its personal state s_k, but also through admission control that might assign fewer channels if $L_{\text{av}} < L_{\text{req}}(S_i)$. Thus, the transition probabilities and, consequently, all performance metrics additionally depend upon L_{av}. The approach to derive the transition probabilities is applied combinatorics and the details shall be omitted for conciseness. The process is summarized by the following steps.

1. Fix M, K, L_{av}, and the resource allocation scheme. Keep in mind that the resource allocation scheme can only be implemented for *every* agent in a particular time step if enough channels are available. Channels may be denied according to the *random* or *cliff* admission control scheme if, in total, too many are requested.

2. Calculate the set of all *up* system states according to Eq. (6.1).

3. For every *up* system state, calculate all possible sink states.
4. For every *up* system state, calculate all possible channel allocations. For the *cliff* admission control scheme the channel allocation is deterministic, i.e., given the state, the channel allocation is fixed. For the *random* admission control scheme however, there are a multitude of possible channel allocations, each with their respective probability.
5. Determine the probability of reaching each possible sink state for each possible channel allocation via Eq. (4.8) and applied combinatorics.
6. Combine each channel allocation probability with the probability of reaching a given sink state with this particular channel allocation.
7. Combine these probabilities for each sink state.

6.1.3 Computational Complexity

The number of *up* states can be derived through combinatorics as

$$Z_{\text{up}} = \frac{(M+K)!}{M!K!} \tag{6.3}$$

and therefore scales approximately with M^K for large M and small fixed K. The number of transitions N_{trans} originating in a system state $S_{|s_0|,\ldots,|s_K|}$ can be derived as

$$N_{\text{trans}}(S_{|s_0|,\ldots,|s_K|}) = \begin{cases} \prod_{k=0}^{K} (|s_k|+1) & \text{if } |s_K| = 0 \\ \prod_{k=0}^{K-1} (|s_k|+1) + 1 & \text{otherwise} \end{cases} \tag{6.4}$$

with the case discrimination stemming from merging all *down* states to S_{d}. The number of transitions $N_{\text{trans,total}}$ in the whole Markov chain scales with M^{2K}.

Fig. 6.1 plots the number of *up* states and the total number of transitions in the Markov chain, both exact according to Eqs. (6.3) and (6.4), outlining the complexity of the approach.

As mentioned earlier, the *random* admission control scheme results in a multitude of possible channel allocations for each particular system state S_i, consequently leading to the same number of transitions to each possible sink state S_j, each with respective probability. Although still being closed-form, this leads to a severe increase in computing complexity. Therefore, the number of agents is limited to $M \leq 20$ for the *random* assignment case within the analysis here.

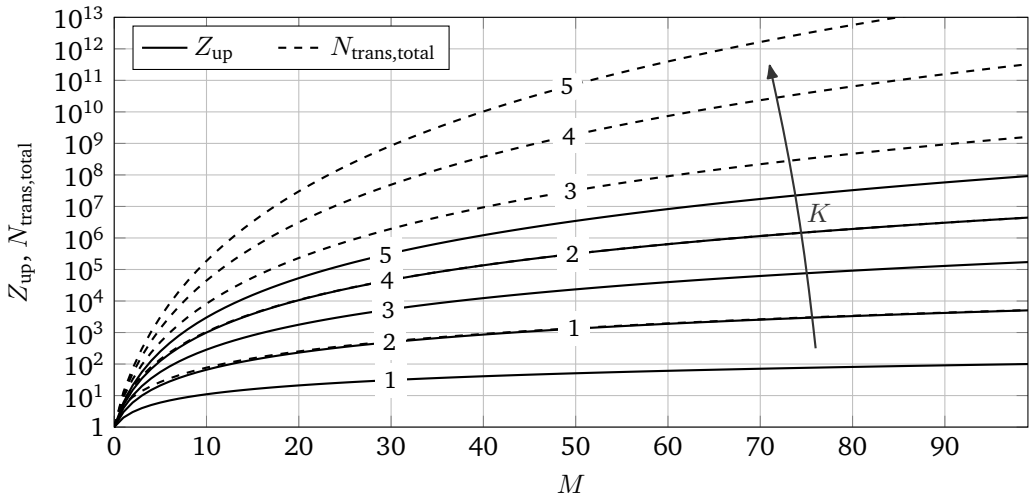


Fig. 6.1.: Computational complexity assessment with inputs M and K.

6.1.4 Performance Metrics

Analogously to the individual agent's MTTF, the MTTF_{sys} on the system level is introduced. Recall that the system is defined as *down* as soon as at least one agent fails, i.e., reaches its individual s_{d} state. It is known that the time to failure (TTF) distribution for a Markov chain with one absorbing (failure) state is phase-type, which can be approximated by an exponential distribution [Neu94]. It can be shown that the minimum of M independent and identically distributed exponential random variables TTF_m with $m \in \{0, \ldots, M-1\}$, which refers to the earliest failing agent, reduces the expectation by a factor M, i.e., $\mathbb{E}\left\{\min_m \text{TTF}_m\right\} = \frac{1}{M}\mathbb{E}\left\{\text{TTF}_m\right\}$. Hence, MTTF_{sys} without any limitation of channels, $\text{MTTF}_{\text{sys,max}}$, is approximately a factor M smaller compared to the single-agent MTTF, i.e.,

$$M \approx \frac{\text{MTTF}}{\text{MTTF}_{\text{sys,max}}} \leq \frac{\text{MTTF}}{\text{MTTF}_{\text{sys}}} \,. \tag{6.5}$$

This is intuitive because instead of only one agent potentially failing, there are M agents potentially failing simultaneously.

The derivation of MTTF_{sys} is performed analogously to Eqs. (4.6)-(4.7). Hence, when initializing in S_0 (all agents in s_0),

$$\text{MTTF}_{\text{sys}} = T_{\text{s}} \boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \mathbf{1}, \tag{6.6}$$

where $\boldsymbol{N}_{\text{sys}}$ denotes the fundamental matrix of the system Markov chain. Also analogously,

$$\boldsymbol{\pi}_{\text{sys}} = \frac{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}}}{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \mathbf{1}^\intercal} \tag{6.7}$$

comprises the probabilities of being in each transient state during operation, i.e., the system version of Eq. (4.10). Additionally, let $\mathbf{1}(L_{\text{av}})$ denote a $(1 \times Z_{\text{up}})$ vector indicating all states, in which all available channels or more are requested, and, hence, whose elements are composed through

$$\mathbf{1}(L_{\text{av}})_i = \begin{cases} 1 \text{ if } L_{\text{av}} \leq L_{\text{req}}(S_i) \\ 0 \text{ else} \end{cases} . \tag{6.8}$$

Then, let the *channel saturation ratio* η denote the proportion of time in which all L_{av} channels are in use as an indicator for the value of adding an additional channel.

$$\eta(L_{\text{av}}) = \frac{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \mathbf{1}(L_{\text{av}})^\intercal}{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \mathbf{1}^\intercal} \tag{6.9}$$

6.2 Illustration Scenario

An illustration scenario with purposefully low numbers M, K, and L_{av} is considered in this section to facilitate comprehending the introduced system model.

The $\boldsymbol{l}_1^1$ resource allocation scheme is assumed, i.e., 1 channel will be requested per agent in s_0, and 2 channels per agent in s_1. Furthermore, it is established that this scenario has $M = 3$ agents, each of which can tolerate only one single isolated packet loss, $K = 1$. There are $L_{\text{av}} = 4$ available channels and the *random* admission control strategy is applied for the case that too many channels are requested in total. The resulting system Markov chain is depicted in Fig. 6.2. According to Eqs. (6.3) and (6.4), it has 4 *up* states and 14 state transitions, 4 of which lead to the absorbing state S_{d}. With $L_{\text{av}} = 4$, the transient system states $S_{1,2}$ and $S_{0,3}$ are subject to being denied 1 and 2 channels, respectively, see Tab. 6.1a. Taking $S_{1,2}$ as an example, the agents request $(1, 2, 2)$ channels and hence, in total 1 channel needs to be denied. With the *random* admission control strategy, this leads to three possible channel assignments and corresponding probabilities as illustrated in Tab. 6.1b. Note that if the *cliff* admission control strategy (prioritizing agents in higher states) were chosen instead, this table would always consist of only a single row, e.g., in this scenario $(0, 2, 2)$ (with probability 1). All possible channel assignment probabilities are subsequently required to be combined with the probability of reaching a given

sink state with that particular channel assignment. In Tab. 6.1c, $S_{2,1}$ is chosen as an exemplary sink state. The final transition probability is determined through probability combination. Creating the system's transition probability matrix $\boldsymbol{P}_{\text{sys}}$ from all transitions in Fig. 6.2, and subsequently the submatrix $\boldsymbol{Q}_{\text{sys}}$ (which only includes only the transition probabilities among transient states), the system's fundamental matrix $\boldsymbol{N}_{\text{sys}}$ and finally MTTF_{sys} through Eq. (6.6) results in $\text{MTTF}_{\text{sys}} \approx 20\,\text{s}$ for $T_{\text{s}} = 90\,\text{ms}$ and a per-channel packet loss probability $p_{\text{loss}} = 10\,\%$. While this MTTF_{sys} is certainly too low for any realistic industrial application, this illustration scenario clarifies the presented procedure. In the next sections, more realistic values will be evaluated.

Tab. 6.1.: System states and exemplary channel assignments for the illustration scenario.

(a)

i	source state	L_{req}
0	$S_{3,0}$	3
1	$S_{2,1}$	4
2	$S_{1,2}$	5
3	$S_{0,3}$	6

(b)

requested channels	assigned channels	prob.
$(1,2,2)$	$(0,2,2)$	0.2
	$(1,1,2)$	0.4
	$(1,2,1)$	0.4

(c)

sink state	prob.
$S_{3,0}$	0
$S_{2,1}$	0.9801
S_{d}	0.0199
$S_{3,0}$	0.8019
$S_{2,1}$	0.0891
S_{d}	0.109
$S_{3,0}$	0.8019
$S_{2,1}$	0.0891
S_{d}	0.109

Transition probability $S_{1,2} \rightarrow S_{2,1}$:

$0.2 \cdot 0.9801 + 0.4 \cdot 0.0891 + 0.4 \cdot 0.0891 = 0.2673$

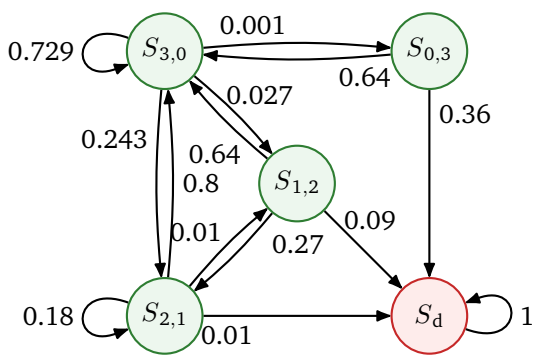


Fig. 6.2.: System model of the illustration scenario with $M = 3$ and $K = 1$. All green states are *up*, only S_{d} is *down*.

6.3 Results

A modeling approach for SARA on the system level based on a Markov chain was presented in the previous section; closed-form expressions for the MTTF_{sys} and the channel saturation ratio η were derived, the complexity regarding increasing M and K was highlighted and an illustration scenario clarified the methodology by means of an oversimplified example with $M = 3$, $K = 1$. In this section, more realistic values are chosen.

The structure is as follows. First, the presented closed-form model will be verified with system-level simulations. Second, the derived performance metrics MTTF_{sys} and η are evaluated for all resource allocation schemes of this thesis, also highlighting the impact of the introduced admission control strategies *random* and *cliff*. Apart from the system-level verification, the per-channel packet loss probability is set to $p_{\text{loss}} = 10\,\%$, following the throughput-optimizing design recommendations in [WJ11], which enable a high spectral efficiency. Only for the system-level verification, p_{loss} will be increased to 30 % because otherwise the simulations would take months or even years for a reasonable statistical significance.

6.3.1 Verification through System-Level Simulation

Fig. 6.3 depicts the results of extensive system-level simulation for the schemes $\boldsymbol{l}_1^0$, $\boldsymbol{l}_2^0$, and $\boldsymbol{l}_1^1$ at $M = 20$ for $p_{\text{loss}} = 30\,\%$. Each simulation was run until at least one agent

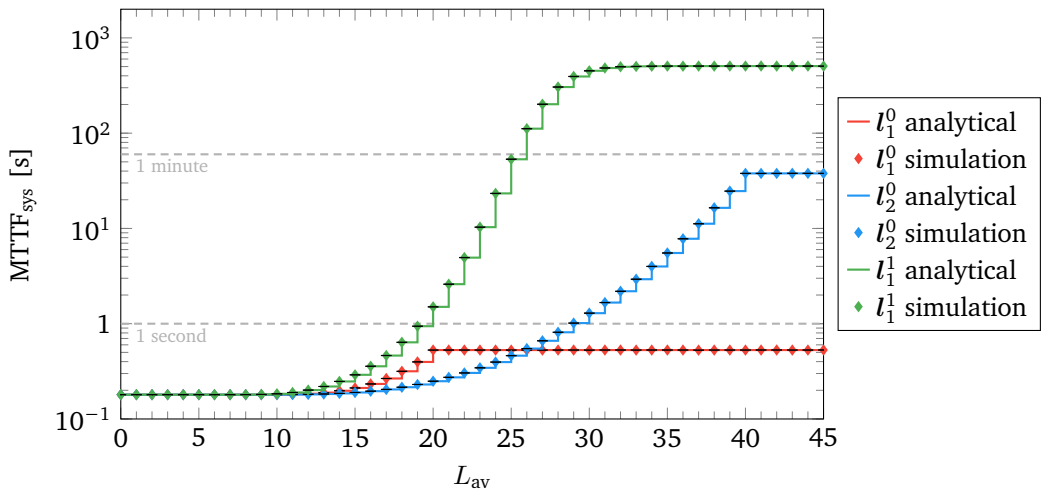


Fig. 6.3.: Verification of closed-form MTTF_{sys} expression through extensive system-level simulation.

failed and the number of runs was 10^6 for each data point. The simulated $\mathrm{MTTF_{sys}}$ values are shown as colored markers and for comparison, the closed-form $\mathrm{MTTF_{sys}}$ from Eq. (6.6) is shown as stair plot. The 99 % confidence intervals determined via the *large sample confidence interval* method is also plotted (in black). They are barely visible because the simulations match the analytical results very well.

6.3.2 Applicability on the System Level

From now on, only closed-form results will be presented. Fig. 6.4 compares $\mathrm{MTTF_{sys}}$ vs. L_{av} for $M = 20$ and $K = 3$. Please note the logarithmic ordinate axis and the human-readable time units. It is stressed that the computation time of $\mathrm{MTTF_{sys}}$ for each data point solely depends on M and K and amounts to approximately 3 seconds in this scenario. Because the alternative system-level simulation also depends upon p_{loss}, which is set to 10 % here, the simulation time with reasonable statistical significance amounts to years for this parameter set, underlining the need for an analytical solution.

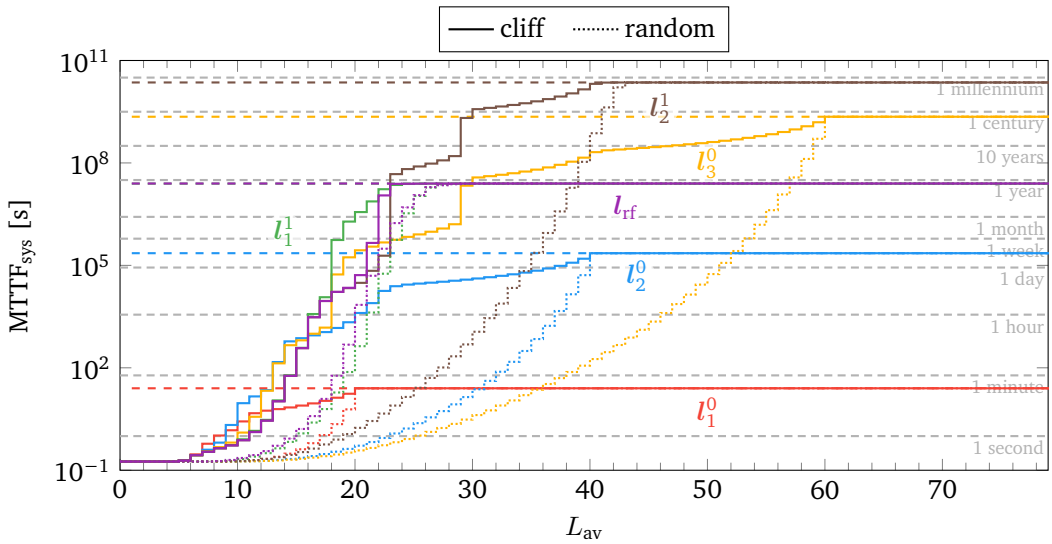


Fig. 6.4.: Dependency of $\mathrm{MTTF_{sys}}$ on L_{av} for $M = 20$ and $K = 3$ at $p_{\mathrm{loss}} = 10\,\%$.

It can be observed that reducing the number of available channels towards $L_{\mathrm{av}} \to 0$, each resource allocation scheme features an $\mathrm{MTTF_{sys}} = 180\,\mathrm{ms} = (K+1)T_{\mathrm{s}}$ because at least one agent is not allocated any channel during this time and, thus, passes straight through the single-agent Markov chain (see Fig. 4.3), reaching the *down* state in the shortest possible time. On the other extreme, for $L_{\mathrm{av}} \to \infty$, the $\mathrm{MTTF_{sys}}$ reaches its maximum value (displayed with a colored dashed line), which can be

calculated through Eq. (6.5). Please note that the graphs for $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_{\text{rf}}$ overlap to a significant extent.

The SARA schemes $\boldsymbol{l}_1^1$, $\boldsymbol{l}_2^1$, and $\boldsymbol{l}_{\text{rf}}$ show clear benefits over static MC, i.e., $\boldsymbol{l}_1^0$, $\boldsymbol{l}_2^0$, and $\boldsymbol{l}_3^0$. Comparing $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_2^0$ (static dual-connectivity) for example, multiple advantages stand out:

1. $\boldsymbol{l}_1^1$ features a 100x improvement in terms of $\text{MTTF}_{\text{sys,max}}$ over $\boldsymbol{l}_2^0$.
2. SARA can also be applied to a multi-user system with limited resources in an efficient way. Reaching the identical MTTF_{sys} requires fewer available channels L_{av} for $\boldsymbol{l}_1^1$. $\boldsymbol{l}_2^0$ reaches its $\text{MTTF}_{\text{sys,max}} \approx 3$ days at $L_{\text{av}} = 40$ (for both admission control schemes), while the same MTTF_{sys} is reached for $\boldsymbol{l}_1^1$ at only $L_{\text{av}} = 23$ (*random*) or even $L_{\text{av}} = 18$ (*cliff*).
3. Comparing the respective $\text{MTTF}_{\text{sys,max}}$, $\boldsymbol{l}_1^1$ also proves to scale better. 99 % of the maximum value ($0.99 \times \text{MTTF}_{\text{sys,max}} \approx 9$ months) is reached at $L_{\text{av}} = 29$ (*random*) and $L_{\text{av}} = 25$ (*cliff*), respectively, for $\boldsymbol{l}_1^1$ and therefore earlier than for $\boldsymbol{l}_2^0$ ($0.99 \times \text{MTTF}_{\text{sys,max}} \approx 3$ days) at $L_{\text{av}} = 40$. That is, the 100x MTTF_{sys} increase requires only between 63 % and 73 % of the channels, depending on the admission control scheme under consideration.

In general, the static schemes perform slightly better than the dynamic SARA schemes only for a low number of available channels ($L_{\text{av}} < 16$). However, in these cases there are less available channels than agents, such that in fact all static schemes become dynamic due to the *cliff* admission control scheme, supporting the advantages of dynamic and state-aware resource allocation. Furthermore, these scenarios rather reflect situations of peak load with a higher number of agents than usual, since a reliable system that does not involve a reduction of the effective sampling rate by design should provide at least $L_{\text{av}} = M$ channels. Overall, an analysis as the one shown in Fig. 6.4 provides a valuable tool for a reasonable choice of the best resource allocation scheme with respect to the $\text{MTTF}_{\text{sys,max}}$ depending on the situation at hand.

6.3.3 Comparison of Admission Control Schemes

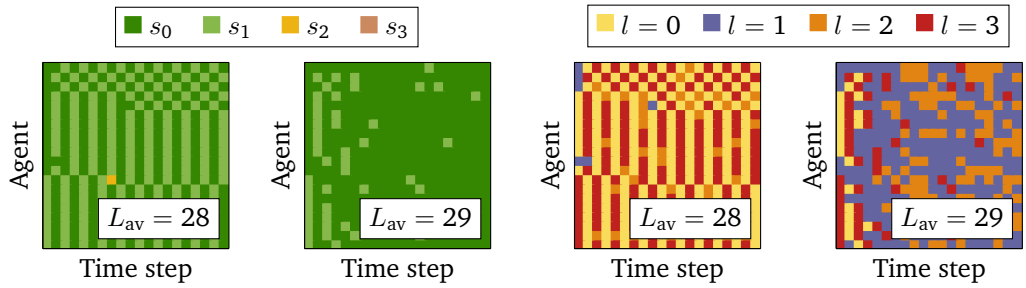
Comparing the *random* and *cliff* admission control schemes in Fig. 6.4, the *random* scheme for all static resource allocation schemes is shown to produce convex curves without sudden jumps, yielding an increasing gain in the high-L_{av} regime until the MTTF_{sys} saturates at 20 ($\boldsymbol{l}_1^0$), 40 ($\boldsymbol{l}_2^0$), and 60 ($\boldsymbol{l}_3^0$) available channels, respectively.

This comparatively late gain increase might be undesirable in the sense that it makes choosing an operating point difficult, as every channel added to the system increases the performance not only absolutely but also relatively. In other words, every channel added to the system is more valuable to the system than the last. The graphs of all dynamic resource allocation schemes in comparison also feature curves without sudden jumps for the *random* admission control scheme, however, they are convex only in the beginning, but become concave and flatten out slightly when approaching $\text{MTTF}_{\text{sys,max}}$.

In comparison overall, it is demonstrated that the *random* scheme is inferior to the *cliff* scheme for all L_{av}, confirming the intuitive assumption that in case of too few assignable channels, it is beneficial to prioritize agents in higher individual states over lower individual states. Since the *cliff* admission control scheme has proven to be superior over the whole range of L_{av}, all resource allocation schemes, and also all values M and K, the *random* admission control scheme will be omitted in the following for better readability.

All MTTF_{sys} curves with *cliff* admission control feature an earlier, steeper and coarser development compared to the *random* scheme. Fig. 6.4 shows clearly that in terms of aiming at an MTTF_{sys} increase there are certain regions where adding another channel to the system is extremely beneficial, e.g., the 29th channel for the static three-fold connectivity scheme $\boldsymbol{l}_3^0$, whereas in other ranges, adding another channel does not provide much gain, e.g., the 44th channel for $\boldsymbol{l}_3^0$. **Proposition** [The locations featuring a steep MTTF_{sys} increase can be explained through a termination of a system-induced oscillation that results as a combination of a particular resource allocation scheme, the *cliff* admission control scheme and a particular L_{av} value.] This will be explained by means of an example, however, a similar explanation can be found for all major steps in Fig. 6.4. Consider the $\boldsymbol{l}_3^0$ scheme at $L_{\text{av}} = 28$ and $L_{\text{av}} = 29$ that differ by one order of magnitude in terms of MTTF_{sys}. Fig. 6.5 shows a section of a simulation conducted for both cases. In Fig. 6.5a, the individual agent states are displayed and in Fig. 6.5b, the allocated channels are shown. For $L_{\text{av}} = 28$, when $S_{12,8,0,0}$ is reached, i.e., 12 agents are in s_0 and 8 agents are in s_1, the 28 available channels are distributed the following way: Since 8 agents will be prioritized by the *cliff* admission control scheme, they receive all channels that they request, i.e., 3, totalling $3 \times 8 = 24$ channels. This leaves 4 channels. Since the other 12 agents are in the same (lower) state s_0, these remaining 4 channels will be distributed equally, resulting in 4 agents receiving one channel while 8 agents receive none. Hence, 8 agents are completely excluded from transmission, consequently being forced to transition into s_1. This completes the cycle because in the next time step, also at least 8 will be in s_1 (even more when packets are lost), being prioritized

over the other agents, and so forth. This cycle is broken up by introducing one more channel to the system, i.e., $L_{\text{av}} = 29$, because one more agent in s_0 can be assigned one channel. Hence, only 7 agents are forced into s_1, slowly breaking down the oscillation. Consequently, the probability of being in the higher state s_1, which is closer to failing, is reduced dramatically, which improves the MTTF_{sys}.

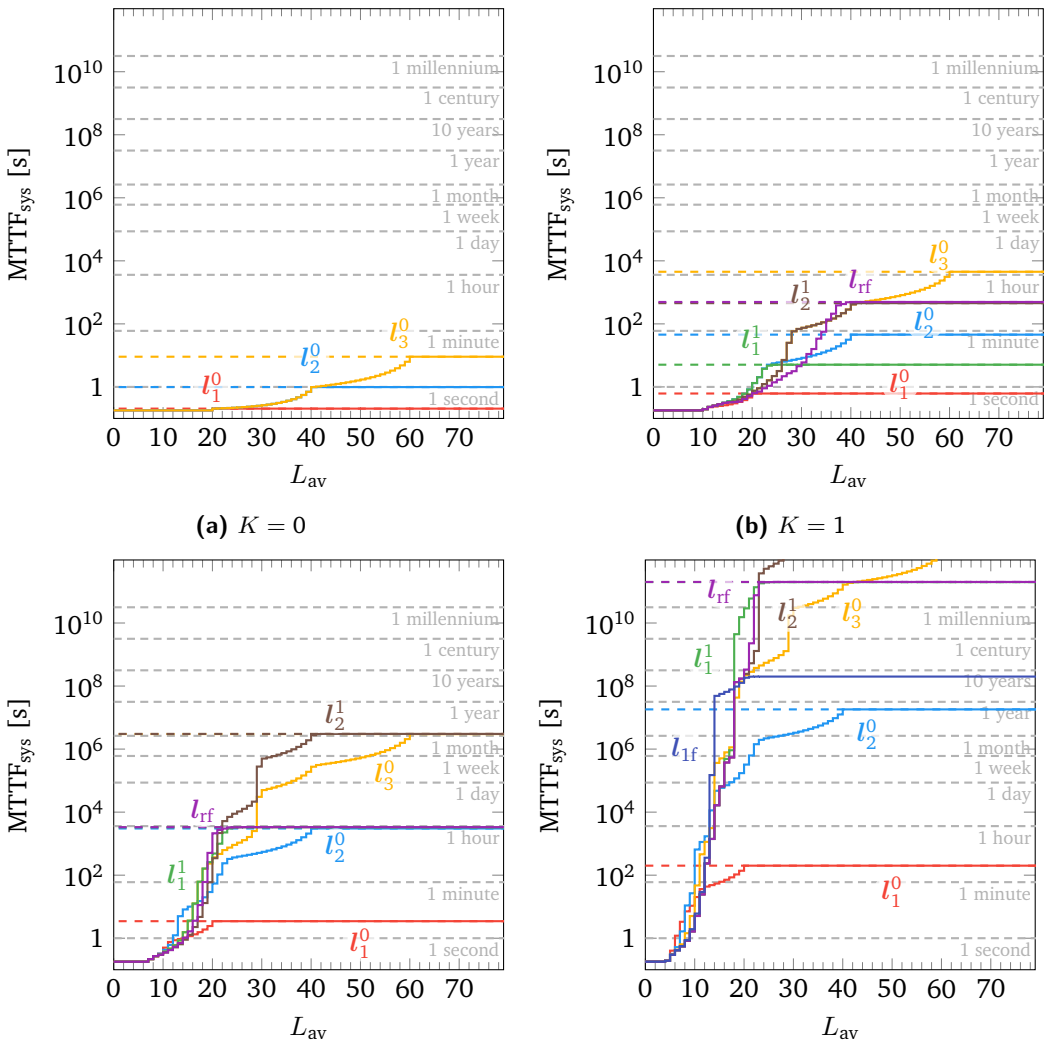


(a) Agent states. **(b)** Number of channels l assigned to each agent.

Fig. 6.5.: Sample system simulation for $\boldsymbol{l}_3^0$ with $M = 20$, highlighting the state oscillations in $L_{\text{av}} = 28$ (left, respectively) that can be terminated by adding one more channel, $L_{\text{av}} = 29$ (right, respectively)

6.3.4 Impact of the Packet Loss Tolerance

The packet loss tolerance K has a great influence on the achievable MTTF_{sys} since it corresponds to the system's temporal diversity. Fig. 6.6 shows the MTTF_{sys} for $K \in \{0, 1, 4\}$ (the case $K = 3$ was already evaluated in Fig. 6.4), highlighting the trade-off between the necessary number of channels in the system vs. K when targeting a specific MTTF_{sys}. For instance, when $K = 0$ in Fig. 6.6a, triple-connectivity ($\boldsymbol{l}_3^0$) will only yield $\text{MTTF}_{\text{sys}} < 1$ minute. Since every added static channel increases the MTTF_{sys} by approximately one order of magnitude given the assumptions of Sec. 4.1, 10 channels will be required *per agent* to ensure a $\text{MTTF}_{\text{sys}} \approx 1$ year. Ignoring the fact that 10 parallel channels violate the limit of $\hat{l} = 4$, this also requires $L_{\text{av}} = 200$ system channels for $M = 20$ agents, underlining the enormous resource usage of static MC with no packet loss tolerance. Please note that in the baseline case $K = 0$, SARA cannot be applied because the system is already considered down after the first lost packet and, hence, there is no headroom for reacting to packet losses. For $K = 1$ in Fig. 6.6b, three static parallel channels ($\boldsymbol{l}_3^0$) lead to $\text{MTTF}_{\text{sys}} \approx 1$ hour. Due to the increased temporal diversity from tolerating $K = 1$ consecutive packet losses, every additional static channel increases the MTTF_{sys} by roughly two orders of magnitude instead of one. Hence, only 5 static parallel channels will be required to reach a $\text{MTTF}_{\text{sys}} \approx 1$ year. Also, $K = 1$ enables



(a) $K = 0$ **(b)** $K = 1$

(c) $K = 2$ **(d)** $K = 4$

Fig. 6.6.: Dependency of MTTF$_{sys}$ on the packet loss tolerance K.

to apply a one-stage SARA, already highlighting SARA's potential by increasing the MTTF$_{sys}$ by roughly $K = 1$ order of magnitude (compare $l_1^0 \leftrightarrow l_1^1$ and $l_2^0 \leftrightarrow l_2^1$), only with marginally increased resource consumption in terms of L_{av}. In Fig. 6.6d, in which it is assumed that the control application is able to tolerate $K = 4$ consecutive packet losses, SARA's enormous potential is highlighted. The MTTF$_{sys}$ increases towards (theoretically) thousands of years while still only requiring $L_{av} = 24$ channels in total. **Proposition** [Keeping in mind that the design guideline for digital control applications is to oversample in the range 10 to 30 to achieve a reasonable smoothness [FWP97], a value of $K = 4$ is reasonable and often available *for free* if isolated smoothness interruptions are acceptable (which might depend on the control application under consideration).]

6.3.5 Impact of the Number of Agents

A central objective of this chapter is to identify SARA's dependability limitations on the system level. For the unconstrained case, i.e., where the number of channels is unlimited, the system's dependability is determined by the underlying SARA scheme and the number of agents (according to Eq. 6.5). In this case, each agent's individual dependability is not influenced. For the constrained case, this changes because channels will be statistically denied if other agents require them (more urgently). **Proposition** [It is important to understand that – as is generally the case in scenarios that multiplex resources – also the absolute number of agents is important when evaluating SARA's performance. This is because the fewer agents exist, the more probable an individual agent will be affected by other agents' (preferred) resource allocation.] This is shown in Fig. 6.7, which plots the required number of channels per agent that are required to achieve a dependability of 99 % of the (unimpaired) $\mathrm{MTTF}_{\mathrm{sys,max}}$ for three example SARA schemes. It can be seen that – as expected – the number of channels per agent is large for small M and decreases quickly if more agents are added to the system. In the extreme case $M \rightarrow \infty$, the average number of channels per agent approaches the single-agent asymptote of Chapter 5. **Proposition** [In a practical scenario with $M = 20$ agents, the system's effect on the additionally required number of channels is shown to be in the order of 10 %.]

6.3.6 Age of Information

The SUA and PLSL can be evaluated not only for the single-agent case, see Sec. 5.2, but also on the system level; $\mathrm{SUA}_{\mathrm{sys}}$ and $\mathrm{PLSL}_{\mathrm{sys}}$ denote the system-related quantities. From a network perspective, $\mathrm{PLSL}_{\mathrm{sys}}$ provides particularly valuable insight regarding the frequency of PLSLs, including system effects such as the state oscillations that were highlighted in Sec. 6.3.3. Fig. 6.8 depicts the PMF dependency of $\mathrm{SUA}_{\mathrm{sys}}$ and $\mathrm{PLSL}_{\mathrm{sys}}$ on L_{av}, respectively, for the resource allocation schemes $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_{\mathrm{rf}}$ when employed on the system level. Please note that L_{av} is not the argument of the PMFs but rather a parameter such that at each L_{av} value, all graph values sum up to one. In other words, the diagrams depict the development of the distributions as L_{av} increases. As expected, when $L_{\mathrm{av}} \rightarrow 30$, the same values as in Sec. 5.2 are obtained because the agents almost always are assigned the number of channels they request and, hence, there exists no system-induced limitation. However for lower values of L_{av}, the aforementioned state oscillations can be observed. For example at $L_{\mathrm{av}} = 16$ for both $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_{\mathrm{rf}}$, there exists a state oscillation that loops through the states s_0, s_1, and s_2 as can be seen in both diagrams. $\mathrm{SUA}_{\mathrm{sys}}$ shows an equal probability

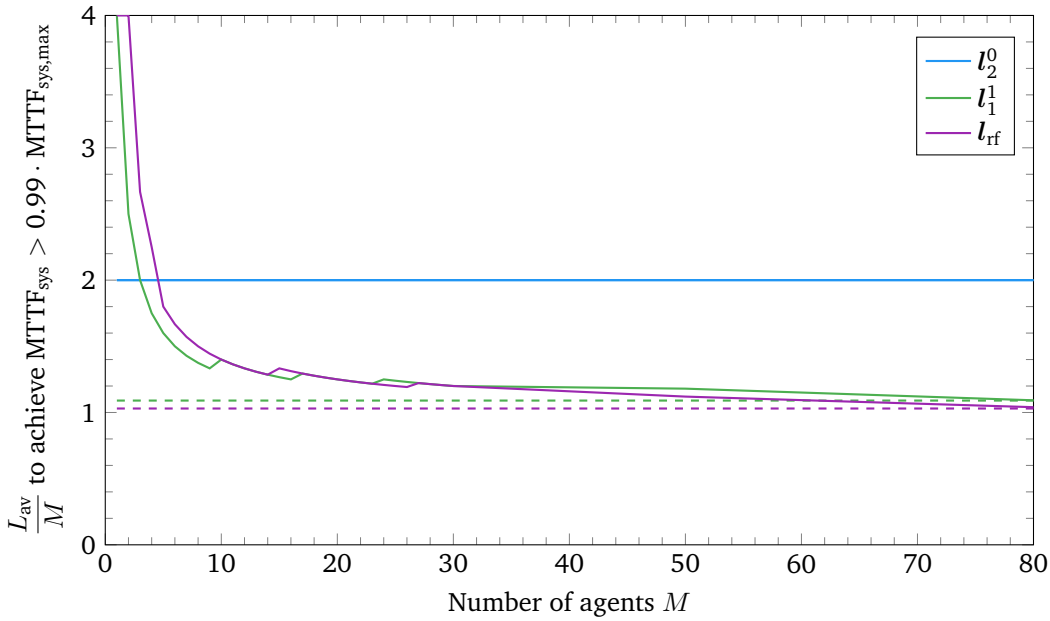
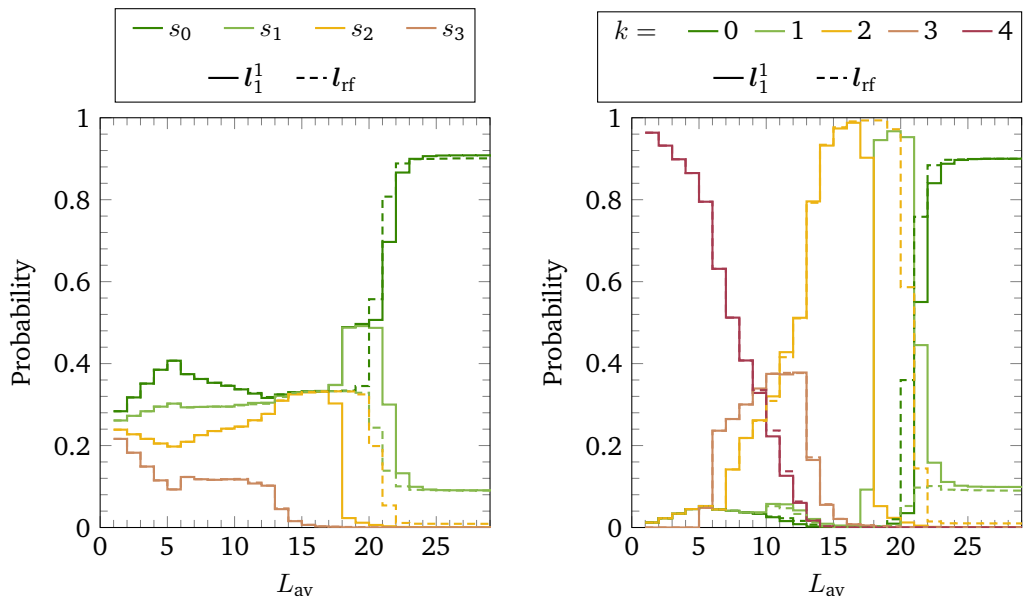


Fig. 6.7.: The fewer agents a system has, the more probable an individual agent will be affected by the (preferred) resource allocation of other agents if the number of channels L_{av} is limited. As $M \to \infty$, the average number of channels per agent approaches the single-agent asymptote.

of being in each of the involved states, indicating that entering s_0 from s_0 and s_1, respectively, is virtually impossible (due to lack of channels that are assigned in those states). This is confirmed by $PLSL_{sys}$, which shows that s_0 is almost always entered coming from s_2 (the yellow graph is close to 1). This state oscillation is broken up at $L_{av} = 18$ for $\boldsymbol{l}_1^1$, however, a state oscillation involving s_0 and s_1 (without s_2) is entered immediately. Finally, at $L_{av} = 21$, state oscillations are entirely broken up. For $\boldsymbol{l}_{rf}$, the state oscillation incorporating the states s_0, s_1, and s_2 is broken up later compared to $\boldsymbol{l}_1^1$ at $L_{av} = 20$, however, the state oscillation involving only s_0 and s_1 is avoided.

It is emphasized that the state oscillations resulting from the *cliff* admission control scheme should be avoided because they increase the effective sampling period T_s of the application by a factor corresponding to the oscillation length (in the examples above first three and then two, respectively) and thereby also increase the overall AoI in the system. Consequently for $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_{rf}$, $L_{av} \geq 23$ should be chosen.



(a) Development of SUA_{sys} PMF versus L_{av}.

(b) Development of PLSL_{sys} PMF versus L_{av}.

Fig. 6.8.: AoI PMF development over L_{av} for $\boldsymbol{l}_1^1$ and $\boldsymbol{l}_{\text{rf}}$ resource allocation schemes.

6.3.7 Channel Saturation Ratio

The sum of requested channels in each system state S_i was introduced earlier as $L_{\text{req}}(S_i)$. Moreover, the channel saturation ratio η was introduced, indicating the proportion of time spent in any state S_i that requests all L_{av} available channels. η is displayed in Fig. 6.9 for $M = 20$ and $K = 3$. Obviously, the 21st (for $\boldsymbol{l}_1^0$), 41st (for $\boldsymbol{l}_2^0$), and 61st (for $\boldsymbol{l}_3^0$) channel do not offer any benefit to the system, explaining the abrupt drop in η. As expected, the dynamic schemes result in graphs that feature a more gradual decrease because starting at the 21st channel (for $\boldsymbol{l}_1^1$) and the 41st channel (for $\boldsymbol{l}_2^1$), respectively, not all system states request all available channels, i.e., there is at least one S_i with $L_{\text{req}}(S_i) < L_{\text{av}}$. In other words, this indicates that additional channels are used by some (but not by all) system states and, thus, they are being wasted in these particular system states. Hence, in the following, an overprovisioning scheme is introduced that, in addition to all requested channels, also assigns all spare channels.

6.3.8 Enforcing Full Channel Saturation

To exploit $\eta = 1$ for all resource allocation schemes throughout the entire operation time, an overprovisioning scheme is introduced that distributes spare channels. For conciseness, only an overprovisioning strategy is considered that prioritizes

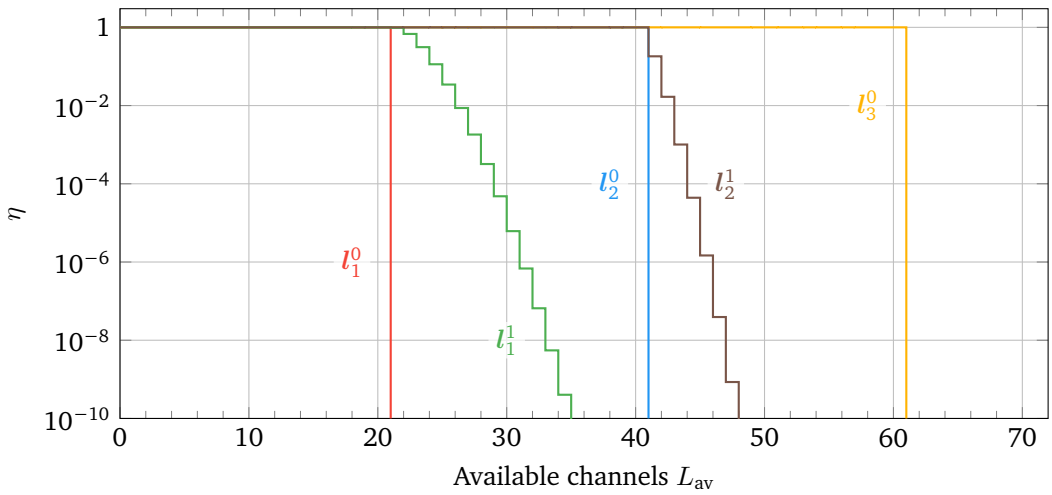


Fig. 6.9.: Channel saturation ratio η for $M = 20$ agents.

agents in higher individual states. Similar to the *cliff* admission control scheme, analysis has shown that prioritizing agents in high individual states maximizes the MTTF_{sys} compared to schemes that prioritize agents in lower states or distribute the resources randomly. As before, at most $\hat{l}$ channels can be assigned to each agent simultaneously and whenever multiple agents are equally close to failing, the channels are distributed equally. Fig. 6.10 shows the results for $M = 20$ and $K = 3$. Please note that since every added channel until $L_{\text{av}} = \hat{l}M = 80$ will be assigned regardless of being requested or not, there exist no individual $\text{MTTF}_{\text{sys,max}}$ values anymore. Multiple observations stand out:

1. All static MC resource allocation schemes are effectively turned into dynamic SARA schemes through overprovisioning. The MTTF_{sys} increase in the range $21 \leq L_{\text{av}} \leq 23$ for the $\boldsymbol{l}_1^0$ scheme is roughly two orders of magnitude per additional channel, clearly highlighting the added benefit of dynamic MC. When compared to $\boldsymbol{l}_2^0$, $\boldsymbol{l}_1^0$ with overprovisioning surpasses the performance of $\boldsymbol{l}_2^0$ at only 2 added channels, i.e., $L_{\text{av}} = 22$. This extreme increase can also be observed for $\boldsymbol{l}_2^0$ (at $L_{\text{av}} = 41$) and $\boldsymbol{l}_3^0$ (at $L_{\text{av}} = 61$), clearly showing that statically distributing the channels is sub-optimal for increasing application dependability.

2. For illustration purposes, the $\boldsymbol{l}_0^0$ scheme is introduced, which corresponds to each agent requesting no channels. Consequently, when overprovisioning is deactivated, $\text{MTTF}_{\text{sys}} = 180\,\text{ms}$, corresponding to a guaranteed system failure after $K+1$ time slots (the minimum value, see dashed line on the bottom of the diagram). However, when overprovisioning is activated, $\boldsymbol{l}_0^0$ behaves similarly

to the SARA schemes, underlining the fundamental idea of prioritizing agents in high individual states. Especially in the $L_{\text{av}} \leq 18$ domain, l_1^1, l_{rf}, l_2^1, and l_0^0 perform identically because they result in the same final resource allocation for each system state, just approached from different directions (denying channels versus overprovisioning).

3. As soon as there are enough channels available that the underlying scheme is always overridden by the overprovisioning, all schemes lead into the same graph because they do not influence the allocation anymore; in our case when $L_{\text{av}} \geq 61$. Please note that the (theoretical) $\text{MTTF}_{\text{sys,max}} \approx 800\text{k}$ years at $L_{\text{av}} = 80$ is the absolute maximum achievable MTTF_{sys} with the boundary conditions $\hat{l} = 4$ and $K = 3$, corresponding to static MC with four parallel channels and enough channels to serve all agents all the time.

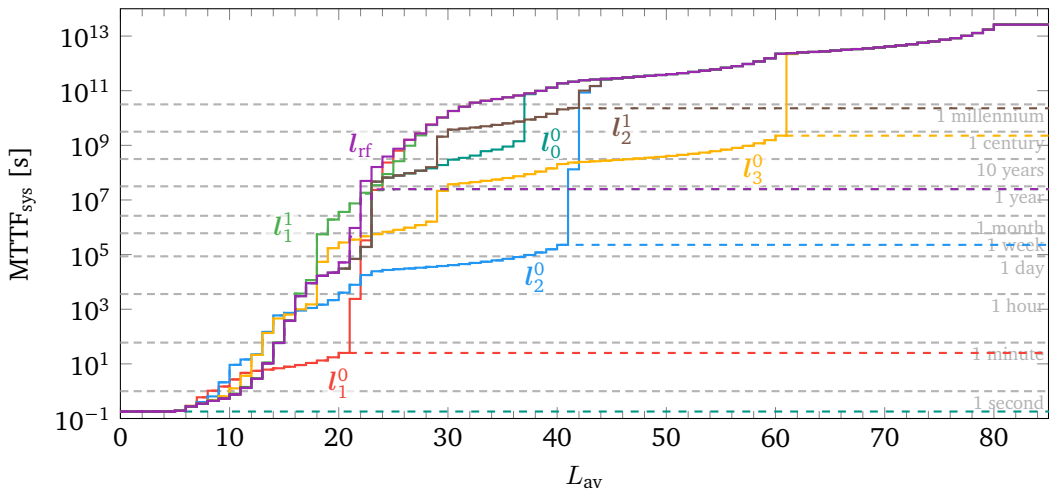


Fig. 6.10.: MTTF_{sys} development while ensuring full channel saturation ($\eta = 1$) through overprovisioning, i.e., distributing spare channels while prioritizing agents in higher individual states.

6.4 Conclusions

This chapter presented a closed-form system model – verified through extensive system-level simulation – which extends the single-agent analysis to multiple agents that compete for a limited number of channels. The system-wide MTTF_{sys} is considered in addition to the single-agent MTTF, which leads to much stronger statements, as each agent is considered as a single point of failure. At 20 agents in the system, the number of required channels to support 99 % of the maximum possible MTTF_{sys}

is 24, or 1.2 channels per agent. This constitutes a slight increase from 1.09 channels, which were determined in the single-agent case (Chapter 5). **Proposition** [This demonstrates that using closed-loop control systems in conjunction with networks that use SARA is also highly feasible in a multi-agent system with competition for resources, because of a significant statistical multiplexing gain.] The AoI within this SARA-enabled system was studied by means of the SUA and PLSL, enabling to employ SARA schemes that are able to statistically guarantee certain PMFs, also on the system level. These AoI metrics enable to identify state oscillations that reduce the effective sampling rate of the served control systems and thereby systematically harm control performance and also decrease the $\mathrm{MTTF}_{\mathrm{sys}}$ by orders of magnitude. It was shown that assigning unrequested/remaining channels for transmission, further drastically increases the $\mathrm{MTTF}_{\mathrm{sys}}$, again by orders of magnitude.

7 Dual Hop – Single Agent

The work in this chapter was not previously published.

In this chapter, SARA is extended to the dual-hop network case, i.e., two consecutive transmissions are considered error-prone and hence, the dependability is influenced by both.

This chapter is structured as follows. First, the derivation of optimal SARA schemes is formalized. Then, example optimization target functions are introduced for the AGV use case introduced in Chapter 3 to perform the optimization over a range of different K and p_{loss}. The resulting optimal SARA schemes are presented and throughly discussed, also assessing the suitability of each scheme to fulfill the desired requirements. The superior dependability, network resource consumption, as well as application-level performance of SARA (compared to static multi-connectivity) is demonstrated through extensive simulation. As a brief outlook, the optimization is repeated without the integer constraint on the number of parallel channels to assess its impact on SARA. Lastly, the chapter is concluded.

7.1 State-Aware Resource Allocation

The dual-hop failure model of Fig. 4.4 is considered. Similar to SARA in the single-hop case, the number of channels used for first-hop (solid states) and second-hop transmissions (dashed states) shall be adjusted according to the history of packet losses. However, as the SUA at the destination is now affected by two transmissions, a holistic resource allocation including the AoI at both, the intermediate node and the destination, may be optimal. For example, it might be beneficial to increase the number of first-hop channels even though only the second hop has lost the packet in the most recent transmission. This approach requires a formalization of the optimization problem.

Three objectives are pursued: (a) increase the MTTF, (b) reduce the average number of channels $\bar{l}$ in both transmissions, and (c) ensure that high SUAs only occur with a low probability. As these objectives partly counteract another, the derivation of

appropriate SARA schemes may be understood as a multi-objective optimization problem, constrained to an integer number of links in this thesis. Formally:

$$\boldsymbol{l}^{\text{opt}} = \arg\min_{\boldsymbol{l}} \left\{ \Gamma \left(g_{\text{MTTF}} \left(\text{MTTF} \left(\boldsymbol{l} \right) \right), g_{\text{SUA}} \left(f_{\text{SUA}} \left(\text{SUA} \left(\boldsymbol{l} \right) \right) \right), g_{\bar{l}} \left(\bar{l} \left(\boldsymbol{l} \right) \right) \right) \right\} \quad \text{subject to } l_k \in \left\{ 1, \ldots, \hat{l} \right\} \tag{7.1}$$

All functions $g_{(\cdot)}$ constitute penalty functions, which must be carefully designed by control system engineers, as MTTF and $\bar{l}$ directly influence operational expenditure, while the PMF of the SUA f_{SUA} influences control performance. The function Γ combines all penalty functions to a single value.

For a given K, the objective is to find the optimal solution among a set of $(Z_{1\text{h}} + Z_{2\text{h}})^{\hat{l}}$ possible resource assignment schemes, where $Z_{1\text{h}}$ and $Z_{2\text{h}}$ denote the number of first-hop and second-hop states, respectively, of the failure model depicted in Fig. 4.4.

$$Z_{1\text{h}} = \sum_{i=1}^{K+1} i \tag{7.2}$$

$$Z_{2\text{h}} = \sum_{i=1}^{K+2} i - 1 \tag{7.3}$$

Since the problem is (a) non-convex, (b) integer-constrained, and (c) grows quickly with K, therefore rendering *exhaustive search* infeasible for $K > 3$, MATLAB's *Surrogate Optimization* from the *Global Optimization Toolbox* is used.

7.2 Optimization Targets

Due to the heterogeneity of potential requirements for control systems, there is not one single optimal SARA solution and the choice of all functions $g_{(\cdot)}$ and Γ is highly subjective. The form (exponential, polynomial, etc.) and weights (prefactors, exponents, etc.) of all functions $g_{(\cdot)}$ and their relation to one another must be carefully balanced to reflect the trade-off between operational expenditure and control performance. The contribution of this thesis is not a single optimal SARA solution but should be understood as a framework that yields optimized AoI-based resource allocation for dependable real-time applications.

In this thesis, the optimization process is illustrated by means of the example penalty functions depicted in Fig. 7.1.

Tab. 7.1.: A set of reference values for the presented punishment functions define a set of optimization targets.

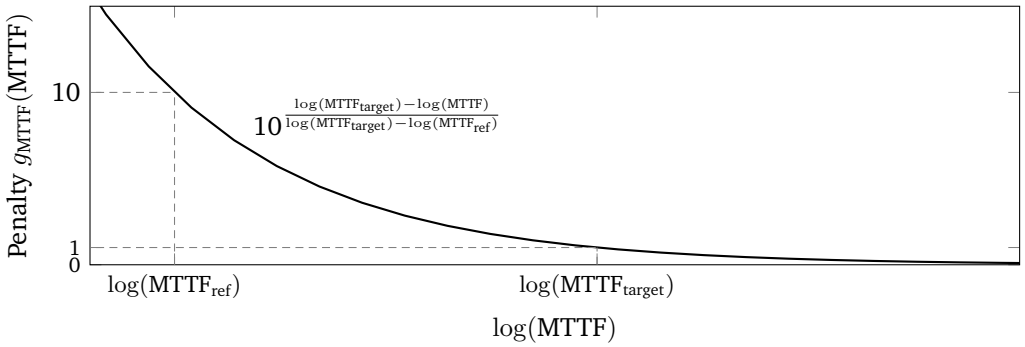
Name	$\text{MTTF}_{\text{target}}$	MTTF_{ref}	$\bar{l}_{\text{ref}}$	SUA_{ref}
T_{ref}	10 years	1.0 year	2	45 ms
$\text{T}_{\text{MTTF}\downarrow}$	**1 year**	**0.1 years**	2	45 ms
$\text{T}_{\bar{l}\uparrow}$	10 years	1.0 year	**1.2**	45 ms
$\text{T}_{\text{SUA}\uparrow}$	10 years	1.0 year	2	**20 ms**

- The penalty function $g_{\text{MTTF}}(\text{MTTF})$ is chosen to be exponentially decaying in the log-domain, compare Fig. 7.1a. It is fully described by $\text{MTTF}_{\text{target}}$ (the target MTTF) and a reference penalty of 10 at a variable reference $\text{MTTF}_{\text{ref}} < \text{MTTF}_{\text{target}}$.
- The penalty function $g_{\bar{l}}(\bar{l})$ is chosen to be zero for $\bar{l} < 1$ because the minimum number of assignable channels is one. Similar to $g_{\text{MTTF}}(\text{MTTF})$, a reference penalty of 10 at a reference $\bar{l}_{\text{ref}}$ defines the slope of the exponential penalty function for $\bar{l} > 1$. The penalty values of the first and second hop are calculated independently for $\bar{l}_{1\text{h}}$ and $\bar{l}_{2\text{h}}$, respectively, according to Eqs. (4.18) and (4.19), and the average value is used for combining.
- The penalty function $g_{\text{SUA}}(\text{SUA})$ is also assumed to take exponential form. Negative SUAs are impossible, hence, $g_{\text{SUA}}(\text{SUA}) = 0, \forall \text{SUA} < 0$. The steepness is determined by the reference SUA_{ref}, at which the penalty is 10.
- The combination function $\Gamma(\cdot)$ is chosen to add all penalty values from the individual penalty functions.

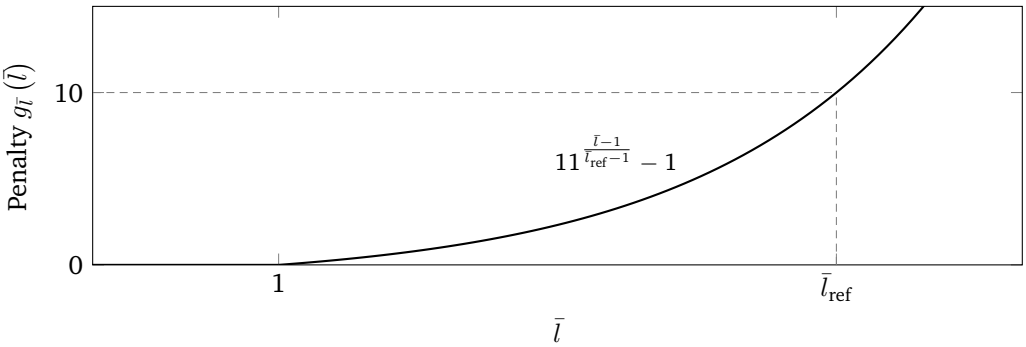
Tab. 7.1 summarizes the reference values used in the penalty functions and thereby defines a set of optimization targets. The T_{ref} target sets a reasonable baseline for the reference AGV use case presented in Chapter 3. The other optimization targets differ slightly in one KPI target: decreased MTTF requirement ($\text{T}_{\text{MTTF}\downarrow}$), increased $\bar{l}$ requirement ($\text{T}_{\bar{l}\uparrow}$), increased SUA requirement ($\text{T}_{\text{SUA}\uparrow}$).

Notation

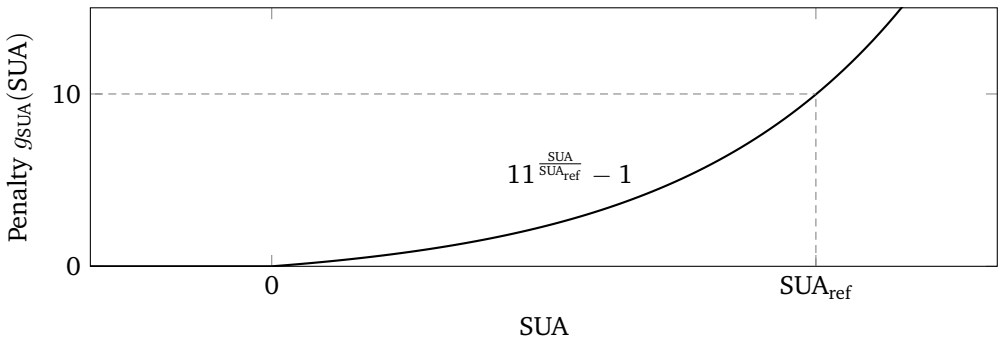
The chosen set of penalty functions is an integral part of solving Eq. (7.1). However, the optimal solution also depends on the peak AoI denoted by K and the per-channel packet loss probabilities $p_{\text{loss,1h}}$ and $p_{\text{loss,2h}}$ for the first and second hop, respectively. In this chapter, it is assumed that $p_{\text{loss,1h}} = p_{\text{loss,2h}} = p_{\text{loss}}$, however, the framework also allows for $p_{\text{loss,1h}} \neq p_{\text{loss,2h}}$. Furthermore, the sampling period remains constant



(a) Exponentially decaying function in the log domain for MTTF



(b) Exponential function for $\bar{l}$



(c) Exponential function for SUA

Fig. 7.1.: Example penalty functions for SARA optimization.

at $T_s = 45\,\text{ms}$ and therefore does not require indication. An optimal resource allocation solution will subsequently be referred to as

$$\boldsymbol{l}^{\text{opt}}(\text{opt. target}, K, p_{\text{loss}}) \ .$$

7.3 Results

The optimal SARA schemes for $p_{\text{loss}} \in \{10\,\%, 1\,\%\}$ and $K \in \{0, 1, \ldots, 5\}$ resulting from 10^4 function evaluations* during the optimization process are displayed in Tab. 7.2. The correspondig KPIs are listed in Tab. 7.3 together with the associated sum penalty and the individual penalty contributions. Multiple observations stand out.

1. It becomes evident once more that per-channel packet loss probabilities p_{loss}, which are typically associated with spectrally efficient best-effort networks (p_{loss} in the range 1 % to 10 %) require some degree of packet loss tolerance $K > 0$ on the application level for realistic industrial reliability requirements. For the assumed penalty functions and $p_{\text{loss}} = 10\,\%$, at least $K = 2$ is required to approach $\text{MTTF}_{\text{target}}$ and ensure an average channel consumption $\bar{l} < 2$. For $p_{\text{loss}} = 1\,\%$, the peak AoI may be set to $K \geq 1$ for useable results. This underlines the importance of tolerating sporadic occurences of $peakAoI > 0$, which translates to isolated packet losses in both, first-hop and second-hop, transmissions.
2. The optimal number of channels generally increases for larger SUAs, which confirms the feasibility of the single-hop SARA schemes presented in Chapter 5. **Proposition** [It can be observed by assessing the individual entries of Tab. 7.2 that the interdependence of the SUA-dependent optimal first-hop and second-hop allocation is lower as one might expect; for nearly all entries, the optimal number of links in the first hop depends solely on the SUA at the intermediate node, which is the KPI the first-hop transmission can directly influence (index $k_{1\text{h}}$). This also holds for the optimal number of channels assigned to the second-hop transmission, which (almost) only depends on the index $k_{2\text{h}}$.] It is noteworthy that for all optimal schemes, the number of channels assigned in any second-hop state $s_{i+1|i'}$ is one, which constitutes the minimum number of assignable channels. This is intuitive

*This limited number of evaluations together with numerical precision errors cause some oddities in the channel allocations for $K \in \{4, 5\}$ in states with high SUA and extreme MTTF > 200 mio. years. For example, $\boldsymbol{l}^{\text{opt}}(\text{T}_{\text{SUA}\uparrow}, 4, 10\,\%)$ features a varying sequence of $(l_{0|4}, l_{1|4}, l_{2|4}, l_{3|4}, l_{4|4}) = (1, 4, 3, 1, 4)$, which does not make sense. However, as this is not important for the main message of this chapter, the results are presented uncorrected.

because any successful transmission in this state does not contribute to minimizing the SUA at the destination and therefore, as few resources as possible are assigned.

3. The sum penalty Γ may function as an indicator whether the used penalty reference values (Tab. 7.1) constitute a reasonable combination or are incompatible/unachievable. For example, a combination of $\text{MTTF}_{\text{target}} = 10$ years is unrealistic for $p_{\text{loss}} = 10\,\%$ and $K \leq 1$, indicated by a minimum penalty $\Gamma \approx 10^3$. For the set of penalty functions used in this thesis, $\Gamma < 10$ can be considered as a good feasibility boundary. For easier interpretation, the penalty values are color-coded in Tab. 7.3.
4. The intuitive assumption is confirmed that optimal SARA schemes for low K are dominated by a high MTTF penalty, whereas for high SUA tolerances (large K), the optimal solution is determined by a $\bar{l} \leftrightarrow$ SUA trade-off (low $\bar{l}$ causes high SUA and vice versa).
5. The impact of modifying individual penalties is clearly reflected in the resulting KPIs. For example, $\text{T}_{\text{SUA}\uparrow}$, which penalizes high SUAs more severely than the reference T_{ref}, enforces minimum 2 channels in both successive transmissions to ensure the high SUA requirement. As a consequence, the penalty does not drop lower than $\Gamma = 15$ for any K, which constitutes a relatively high value and therefore indicates that $p_{\text{loss}} = 10\,\%$ might be too high for $\text{T}_{\text{SUA}\uparrow}$. In a practical optimization, a certain maximum penalty may be defined to infer a required p_{loss} for suitable network design. This is demonstrated for $K = 1$ and the reference values of $\text{T}_{\text{SUA}\uparrow}$ in Fig. 7.3, which displays the change of MTTF, $\bar{l}$, SUA PMF, as well as the associated minimum penalty for varying p_{loss}. It can be seen that all curves except the sum penalty Γ feature discontinuities as a result of the integer constraint. The largest steps in Fig. 7.3b correspond to adding one additional channel in $s_{0|0}$ (first hop) and $s_{0|0'}$ (second hop), respectively, while more subtle changes correspond to a channel increase in higher states. The increase of channels obviously increases the MTTF (Fig. 7.3a) and reduces the probability of having a $\text{SUA} = T_{\text{s}}$ (Fig. 7.3c). The individual penalty components also feature these discontinuities, while the sum penalty function Γ (Fig. 7.3d) is monotonically and steadily increasing per its construction. Assuming a reasonable feasibility boundary of $\Gamma = 10$, a per-channel packet loss probability $p_{\text{loss}} < 1\,\%$ should be ensured.

7.3.1 Extensive Simulation

In Chapter 3, it was shown that a good control performance may be achieved over an error-prone wireless network as long as the temporal packet loss correlation is highly negative. To describe this temporal correlation, an artificial packet loss correlation variable ρ was introduced. Here, the extensive simulations displayed in

Tab. 7.2.: Optimal dual-hop SARA schemes.

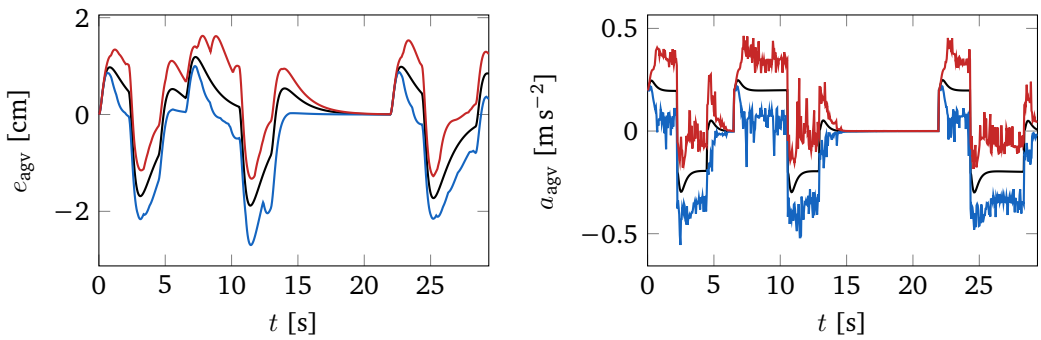
	p_{loss}	10 %																								1 %											
	target	T_{ref}						$T_{\text{MTTF}\downarrow}$						$T_{\bar{l}\uparrow}$						$T_{\text{SUA}\uparrow}$						T_{ref}			$T_{\text{MTTF}\downarrow}$			$T_{\bar{l}\uparrow}$			$T_{\text{SUA}\uparrow}$		
	K	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	0	1	2	0	1	2	0	1	2
	$l_{0\|0}$	4	4	2	1	1	1	4	3	1	1	1	1	2	2	1	1	1	1	4	4	2	2	2	2	4	1	1	4	1	1	2	1	1	4	1	1
number of first-hop channels $l_{k_{1\text{h}}\|k_{2\text{h}}}$	$l_{0\|1}$		3	2	1	1	1		3	2	1	1	1		1	1	1	1	1		3	2	2	2	2		2	1		1	1		1	1		2	2
	$l_{1\|1}$		4	4	3	2	2		4	4	2	2	2		4	4	2	2	2		4	4	4	4	4		4	2		4	2		4	1		4	3
	$l_{0\|2}$			2	1	1	1			2	1	1	1			1	1	1	1			2	2	2	2			1		1			1			2	
	$l_{1\|2}$			4	3	2	2			4	2	2	2			4	2	1	1			4	4	4	4			2		2			1			3	
	$l_{2\|2}$			4	4	3	3			4	4	3	3			4	4	2	2			4	4	4	4			4		4			4			4	
	$l_{0\|3}$				1	1	1				1	1	1				1	1	1				3	2	2												
	$l_{1\|3}$				3	2	2				2	2	2				2	1	1				4	4	4												
	$l_{2\|3}$				4	4	3				4	3	4				4	2	2				3	4	4												
	$l_{3\|3}$				4	4	4				4	4	4				4	4	3				4	4	4												
	$l_{0\|4}$					3	1					3	3					1	1					1	2												
	$l_{1\|4}$					4	2					3	2					1	1					4	3												
	$l_{2\|4}$					4	3					4	3					2	2					3	2												
	$l_{3\|4}$					4	4					4	4					4	3					1	3												
	$l_{4\|4}$					4	4					4	4					4	4					4	4												
	$l_{0\|5}$						1						1						1						1												
	$l_{1\|5}$						2						4						1						1												
	$l_{2\|5}$						4						4						2						4												
	$l_{3\|5}$						4						3						4						1												
	$l_{4\|5}$						1						1						3						1												
	$l_{5\|5}$						4						4						4						1												
	$l_{0\|0'}$	4	4	2	1	1	1	4	3	1	1	1	1	2	1	1	1	1	1	4	4	2	2	2	2	4	1	1	4	1	1	2	1	1	4	1	1
	$l_{1\|0'}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
number of second-hop channels $l_{k_{1\text{h}}\|k_{2\text{h}}'}$	$l_{0\|1'}$		4	4	2	2	2		4	4	2	2	2		4	3	1	1	1		4	4	4	4	4		4	2		4	2		4	1		4	3
	$l_{1\|1'}$		4	4	2	2	2		4	4	2	2	2		4	4	2	2	2		4	4	4	4	4		4	2		4	1		4	1		4	3
	$l_{2\|1'}$		1	1	1	1	1		1	1	1	1	1		1	1	1	1	1		1	1	1	1	1		1	1		1	1		1	1		1	1
	$l_{0\|2'}$			4	4	3	3			4	4	3	3			4	4	3	3			4	4	4	4			4		4			4			4	
	$l_{1\|2'}$			4	4	3	3			4	4	3	3			4	4	3	3			4	4	4	4			4		4			4			4	
	$l_{2\|2'}$			4	4	4	3			4	4	3	3			4	4	3	3			4	4	4	4			4		4			4			4	
	$l_{3\|2'}$			1	1	1	1			1	1	1	1			1	1	1	1			1	1	1	1			1		1			1			1	
	$l_{0\|3'}$				4	4	4				4	4	4				4	4	4				4	4	4												
	$l_{1\|3'}$				4	4	4				4	4	4				4	4	4				4	4	4												
	$l_{2\|3'}$				4	4	4				4	4	4				4	4	4				4	4	4												
	$l_{3\|3'}$				4	4	4				4	4	4				4	4	4				4	4	4												
	$l_{4\|3'}$				1	1	1				1	1	1				1	1	1				1	1	1												
	$l_{0\|4'}$					4	4					4	4					4	4					4	4												
	$l_{1\|4'}$					4	4					4	3					4	4					4	4												
	$l_{2\|4'}$					4	4					4	4					4	4					1	4												
	$l_{3\|4'}$					4	4					4	4					4	4					2	4												
	$l_{4\|4'}$					4	3					4	4					4	4					3	1												
	$l_{5\|4'}$					1	1					1	1					1	1					1	1												
	$l_{0\|5'}$						4						4						4						4												
	$l_{1\|5'}$						4						4						4						4												
	$l_{2\|5'}$						4						4						4						4												
	$l_{3\|5'}$						4						4						4						1												
	$l_{4\|5'}$						4						2						4						1												
	$l_{5\|5'}$						4						3						4						1												
	$l_{6\|5'}$						1						1						1						1												

Fig. 3.10, which plots the path deviation and acceleration signal (control value) of an AGV traversing a given path, are repeated in Fig. 7.2 with the optimized SARA results of this chapter. Figs. 7.2a and 7.2b build the reference static single- and dual-connectivity cases. Note that the MTTF is only 2 minutes (single-connectivity) and 10 days (dual-connectivity), which does not constitute ultra-dependable operation. With the optimization target functions associated with $\mathrm{T}_{\mathrm{ref}}$, the extensive simulation results for the optimal SARA scheme $\boldsymbol{l}^{\mathrm{opt}}(\mathrm{T}_{\mathrm{ref}}, K = 3, p_{\mathrm{loss}} = 10\,\%)$ are plotted in Fig. 7.2c. Recall that from 10^6 simulation runs, the red graph displays the maximum value among all runs and the blue graph the minimum value. It can be seen that the path deviation (left side) is significantly more deterministic than that of single-connectivity and comparable to dual-connectivity. The control value (right side) is most deterministic for the SARA scheme and least deterministic for single-connectivity. At the same time, SARA features an average channel consumption of $\bar{l} = 1.18$, which is 40 % less than $\bar{l} = 2$ for dual-connectivity. Simultaneously, the MTTF is improved by factor of 2700x from 10 days (dual-connectivity) to 74 years and a factor 2×10^7x over single-connectivity (2 minutes).

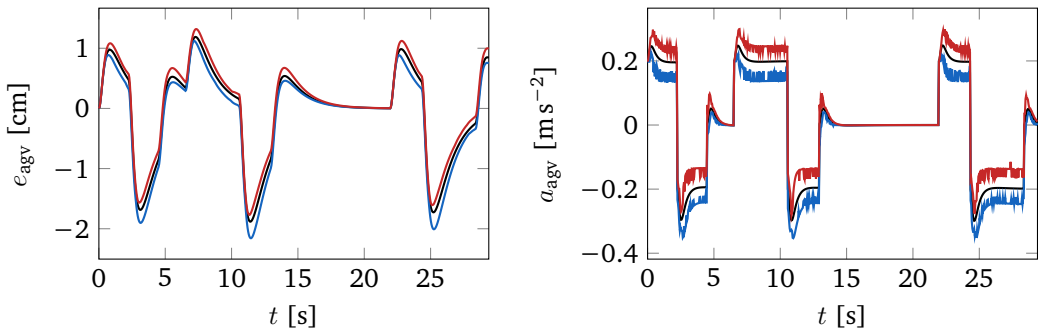
7.3.2 Non-Integer-Constrained Optimization

In the previous section, it was shown how the AoI-based SARA framework of this thesis may be utilized to derive a per-channel packet loss probability requirement p_{loss} that ensures desired KPIs relevant to real-time operation in industrial scenarios, namely MTTF, $\bar{l}$, and SUA. Thereby, the number of channels, which can be assigned to an agent in parallel, was assumed to be integer-constrained, i.e., $l \in \{1, \ldots, \hat{l}\}$, which causes discontinuities in all functions of Fig. 7.3.

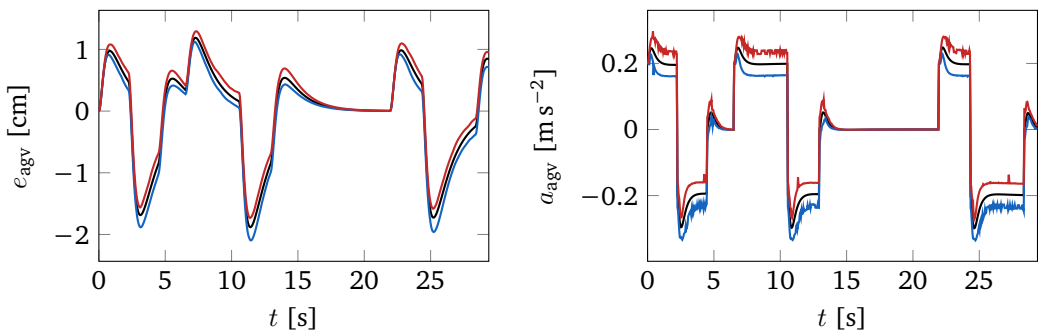
This integer constraint was motivated in Chapter 4 by MC as a tool to increase a transmission's success probability, enabled by selectively combining an integer number of uncorrelated system channels that individually feature an identical packet loss probability p_{loss}. This way, the combined packet loss probability for a transmission could be controlled by the number of parallel channels assigned for one transmission. In this section, this assumption is replaced by allowing to divide the available spectrum arbitrarily. The derivation of a PHY and MAC that allows such arbitrary distribution will be left for future work and the results of this section shall be understood as a teaser. For conciseness and comparability, the term *channels* will be further used despite the blurred meaning in this new context. The new channel constraint is formally described by $l \in [1, \hat{l}\,]$.



(a) Static single-connectivity: with a single channel ($\bar{l} = 1$) the acceleration is jerky and the agent does not operate dependably (MTTF = 2 minutes).

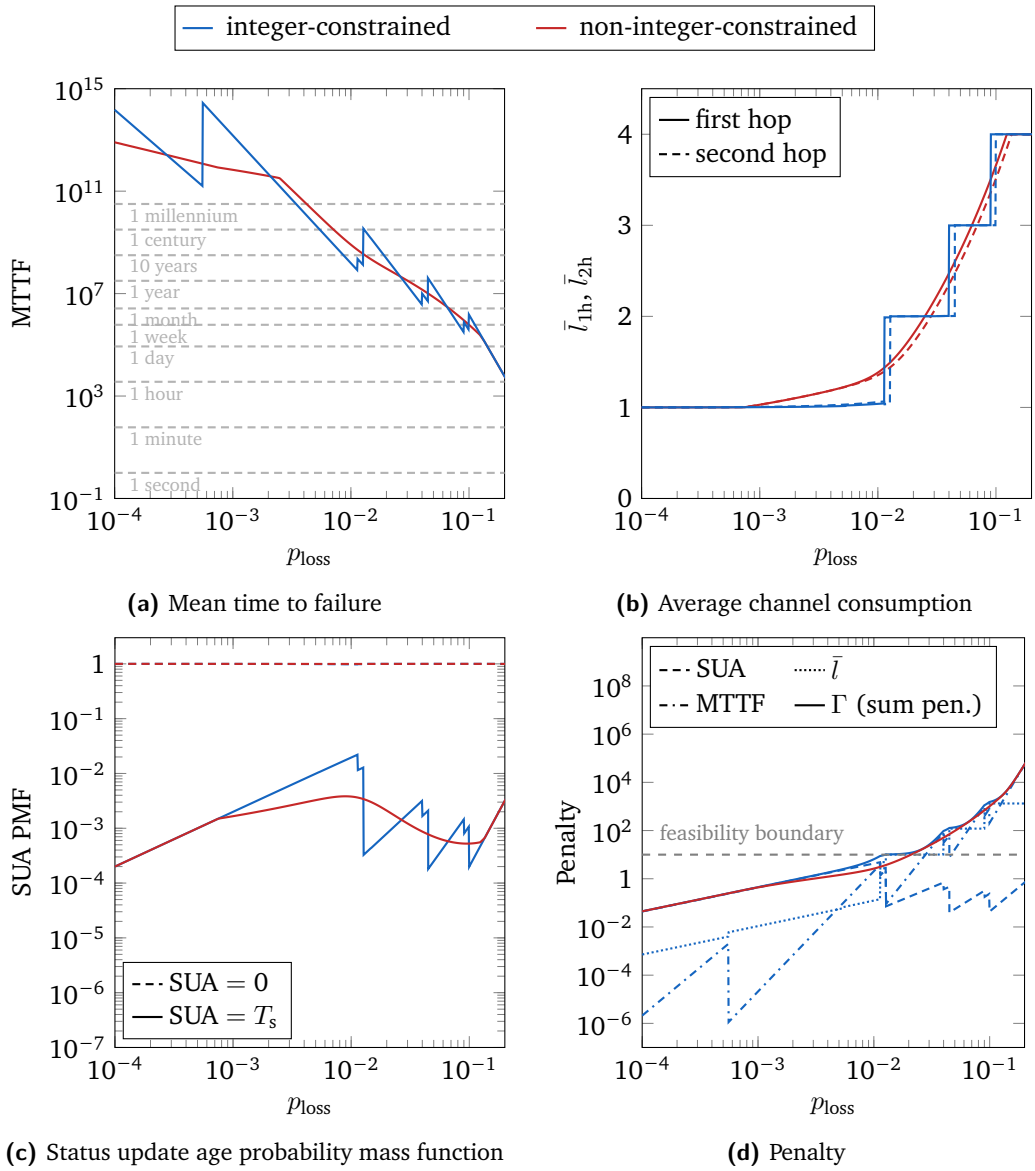


(b) Static dual-connectivity: doubling the number of channels ($\bar{l} = 2$) results in significantly more determinism, however, the agent is still not very dependable (MTTF = 10 days).



(c) $p_{\text{loss}} = 10\,\%$, T_{ref}, $K = 3$: With SARA, the average number of channels reduces to $\bar{l} = 1.18$, the determinism is comparable to that of static dual-connectivity, and the agent's dependability increases drastically to MTTF = 74 years.

Fig. 7.2.: Simulation results for the AGV use case, which was introduced in Chapter 3. The position error e_{agv} and control input a_{agv} (acceleration) are depicted with their upper (red) and lower (blue) bound for 10^6 simulation runs. The SARA scheme outperforms static single- and dual-connectivity in determinism and dependability on the application level, while simultaneously using 40 % fewer network resources than dual-connectivity.



(a) Mean time to failure

(b) Average channel consumption

(c) Status update age probability mass function

(d) Penalty

Fig. 7.3.: KPIs development at varying packet loss probabilities

The optimization of Eq. (7.1) is repeated without the integer constraint and the results are plotted in Fig. 7.3 (red) for a direct comparison with the constrained case (blue). As expected, the discontinuities are resolved and intuitively, the optimal non-constrained solution lies "in between" the constrained solutions as the constrained graphs zig-zag above and below the unconstrained. The sum penalty Γ of the non-integer-constrained optimal resource allocation lower-bounds the integer-constrained sum penalty, which is also expected.

This demonstrates that the presented framework is not restricted to networks with an integer-constrained number of parallel channels, but that it can also be applied to networks that are able to dynamically (on a per-transmission basis) alter the transmission success probability by more elaborated means than MC. It remains an open research question, how channels may be mapped to physical resources (in the time-frequeny-coding space) to yield optimal spectral efficiency for a given target transmission success probability. With such a mapping, the penalty function for the average channel consumption $\bar{l}$ in Eq. (7.1) could be replaced by an advanced penalty function that punishes actual PHY resource usage.

7.4 Conclusions

This chapter presents a formal approach to AoI-based SARA optimization for the dual-hop network case, i.e., AoI-sensitive information is transmitted over two hops (twice either UL or DL or once UL+DL). Penalty functions for low MTTF, high average channel consumption $\bar{l}$, and high SUAs are combined to solve a global optimization problem and derive the best AoI-dependent channel allocation holistically for both hops. Due to the heterogeneity of control systems and therefore a wide range of possible KPI requirements, the validaty of the approach is demonstrated by means of a set of example exponential penalty functions. However, the approach is also applicable to other penalty function classes that define different KPI requirements. Within the chosen set of assumed exponential penalty functions, multiple KPI prioritizations (weightings) are set by varying the individual penalty function steepness.

It is shown that SARA can also be applied to control systems in a dual-hop network, enabling extreme dependability through (on average) highly spectrally efficient transmissions on both hops. Intuitively, low SUA penalties paired with high peak AoI values lead to extreme control-related reliability while keeping the average channel consumption close to single-connectivity. Moreover, it was shown how

different optimization emphases influence the optimal channel allocation, e.g., that a high SUA penalty already triggers multiple channels in the base states and thereby increases the average channel consumption significantly.

It is generally observed that within the transmission cycle, the number of channels on the first hop is mainly determined by the SUA at the intermediate node while the number of channels on the second hop mainly depends upon the SUA at the destination, indicating only a limited cross-dependency. In other words, the optimal resource allocation for a transmission mainly depends on the SUA at the *immediate receiver*.

The presented framework also allows to derive required network QoS to ensure a given set of control-related KPIs. For example, the maximum per-channel packet loss probability p_{loss} may be derived ensuring a limit on the sum penalty, which is crucial for appropriate network design.

Lastly, it was shown that the fundamental SARA idea may also be applied to network architectures that do not restrict the number of parallel channels to an integer value. Although the physical meaning of the term *channel* is blurred and a more appropriate description needs to be developed in future work, it was shown that networks featuring more elaborated reliability enhancement techniques than MC with an integer number of selectively combined channels may also be subjected to the same optimization framework.

Tab. 7.3.: Resulting KPIs.

p_{loss}	opt. target	K	MTTF	$\bar{l}_{1\text{h}}$	$\bar{l}_{2\text{h}}$	Penalty	Penalty Share MTTF	Penalty Share $\bar{l}$	Penalty Share SUA
10 %	T_{ref}	0	4 min.	4	4	1.4×10^{6}			
		1	17 days	4	4	1.5×10^{3}			
		2	4 years	2.02	2.03	1.4×10^{1}			
		3	74 years	1.18	1.17	2.53			
		4	27k years	1.09	1.17	2.34			
		5	230M years	1.09	1.17	2.34			
	$\text{T}_{\text{MTTF}\downarrow}$	0	4 min.	4	4	1.4×10^{5}			
		1	2 days	3	3	3.3×10^{2}			
		2	5 months	1.33	1.48	5.6			
		3	33 years	1.09	1.17	2.38			
		4	27k years	1.09	1.17	2.34			
		5	229M years	1.09	1.17	2.34			
	$\text{T}_{\bar{l}\uparrow}$	0	2 seconds	2	1.99	1.4×10^{8}			
		1	1 hour	1.95	1.32	1.3×10^{5}			
		2	1 month	1.27	1.34	1.7×10^{2}			
		3	8 years	1.09	1.06	6.5			
		4	3k years	1.09	1.04	4.99			
		5	4.7M years	1.09	1.04	4.99			
	$\text{T}_{\text{SUA}\uparrow}$	0	4 min.	4	4	1.4×10^{6}			
		1	17 days	4	4	1.5×10^{3}			
		2	4 years	2.02	2.03	1.8×10^{1}			
		3	29k years	2.02	2.03	1.5×10^{1}			
		4	241M years	2.02	2.03	1.5×10^{1}			
		5	14G years	2.02	2.03	1.5×10^{1}			
1 %	T_{ref}	0	26 days	4	4	1.5×10^{3}			
		1	5 years	1.04	1.06	2.38			
		2	36k years	1.01	1.02	2.3×10^{-1}			
	$\text{T}_{\text{MTTF}\downarrow}$	0	26 days	4	4	1.3×10^{3}			
		1	5 years	1.03	1.06	5.1×10^{-1}			
		2	29k years	1.01	1.02	2.3×10^{-1}			
	$\text{T}_{\bar{l}\uparrow}$	0	4 min.	2	2	1.6×10^{6}			
		1	5 years	1.03	1.06	3			
		2	362 years	1.0003	1.0009	2.7×10^{-1}			
	$\text{T}_{\text{SUA}\uparrow}$	0	26 days	4	4	1.5×10^{3}			
		1	5 years	1.04	1.06	6.5			
		2	3.6M years	1.03	1.04	4.37			

8 Conclusions and Outlook

This thesis provides new insights into enabling dependable, mission-critical industrial applications over spectrally efficient, error-prone wireless networks. As a core element, the novel AoI metric is adopted, which adequately captures the freshness of data for the underlying time-sensitive application and constitutes a novel design target for future real-time-capable wireless networks. It is confirmed by means of a relevant industrial control application use case that keeping the AoI low is beneficial for real-time applications. The typical oversampling, which ensures reasonable smoothness of operation and linearity, allows for a small number of consecutive packet losses until an application-specific maximum AoI threshold is exceeded (= the application fails). This small degree of packet loss tolerance may be exploited to achieve low AoI (strictly limiting the number of consecutive packet losses) without sacrificing on spectral efficiency. As a result, the occurrence of packet losses with temporally highly negative correlation is recommended, making it increasingly unlikely to lose increasingly many consecutive packets. This negative packet loss correlation may be achieved with only a fractional decrease of the overall packet loss ratio, enabling high spectral efficiency. Through the development of AoI-based application failure models that map packet losses to application failures, the (un-)reliability of the application resulting as a direct consequence of packet losses is assessed. A framework to jointly optimize a range of appropriate target KPIs is presented and the scalability potential within a resource-constrained system is evaluated. In the following, the key results are summarized and conclusions are drawn. Lastly, possible extensions to the work of this thesis are discussed.

8.1 Key Results and Conclusions

To demonstrate the effect of high AoI on control applications, the AGV use case as relevant industrial real-time application is modeled. The two methods *Reduced Sampling* and *MJLS* are presented to quantify the packet loss tolerance of LTI control applications. The Reduced Sampling method is a simple yet effective approach to assess the impact of circularly recurring packet loss sequences of given length by deriving the well-known phase margin (as stability criterion) of the underlying LTI

control system, which constitutes sampling at an effective lower rate. As a typical result, the packet loss tolerance is non-zero, contrary to frequently postulated URLLC requirements. Further refining the understanding of the control-communication interdependence, MJLS theory enables a more general joint modeling for any packet loss processes that can be described by a discrete-time Markov chain. It is subseqently shown that highly reliable real-time applications may be achieved also for error-prone wireless networks if a negative temporal packet loss correlation can be ensured, which effectively limits the number of consecutive packet losses. As a consequence, the recommendation is to shift the network's operating point from (a) extremely low-PLR and no temporal packet loss correlation to (b) best-effort-PLR and highly negative correlation, which constitutes **the first key contribution** of this thesis.

As foundation for meaningful results, a network architecture is assumed that is able to fulfill common industrial network requirements, more specifically: tight time-synchronization, semi-persistent scheduling (support of periodical sampling), and MC as reliability enabler. On this basis, **the second key contribution** of this thesis is the quantification of the (un-)reliability of real-time applications as a consequence of packet losses, where an application failure is defined as exceeding an application-specific AoI threshold, termed the peak AoI. As mapping between packet losses (communication errors) and application failure, Markov chain failure models are derived for single-hop (one of either uplink or downlink) and dual-hop (two successive uplink/downlink transmissions) real-time data transmission. For the single-hop case and the simplified network assumptions, the probability of exceeding the peak AoI is linked via the sampling period with the probability of exceeding a certain number of consecutive packet losses. The developed failure models enable to derive closed-form expressions for the MTTF, the average channel consumption, and the AoI PMF, all of which are confirmed through extensive simulation.

Building upon the developed Markov failure models, an AoI-aware resource allocation approach is developed as **third key contribution** of this thesis for both the single-hop and dual-hop case, termed SARA. SARA describes the AoI-dependent optimal number of channels that should be chosen in any transmission slot, penalizing (a) low application MTTF, (b) high average channel consumption, and (c) high AoI through freely selectable, application-specific penalty functions. This optimization may be performed offline to yield a simple look-up table, which may be integrated into the low-level MAC layer in a straight-forward way. As a generally observed result, the optimal number of channels gradually increases for increasing AoI, which is intuitively expected. Exploiting the AoI-dependent dynamic MC capabilities of the network, an MTTF gain of *multiple orders of magnitude* can be

observed while at the same time keeping the average resource consumption in the range of single-connectivity, i.e., a single best-effort channel.

As a **fourth key contribution** of this thesis, the single-hop failure model is lifted to the system level, where multiple users share a finite set of channels. In this scenario (different from the single-user case), the optimal SARA might not be applicable in time slots in which multiple users lose packets and consequently request more channels at the same time. It is shown in closed-form that a roughly 10 % over-provisioning of channels – compared to the average number of channels required in the single-user case – ensures an unimpaired system reliability. If this over-provisioning is not implemented, system-induced state oscillations may occur that (a) significantly degrade control performance due to a systematic sampling rate reduction and (b) reduce system reliability by orders of magnitude. Hence, the presented framework allows to accurately dimension required network resources that do not compromise individual user KPIs.

Overall, this thesis provides a new perspective of how a CoCoCo with the relatively novel AoI as "glue metric" may substantially transform how dependable applications may be realized over spectrally efficient, unreliable wireless links. While highly dependable links are certainly crucial sometimes, they are not needed all the time, which can be exploited on the network level. Consequently, this contributes to realizing the vision of *Industry 4.0*.

8.2 Future Work

The work of this thesis may be extended in many directions.

The obvious first step is to extend the single-agent dual-hop failure model to a system failure model of multiple agents that compete for a given number of first-hop and second-hop channels. Then, the scalability evaluation of SARA for the dual-hop network case can be performed. Due to the low cross-effects between first and second hop with respect to the optimal number of assigned channels, it is expected that the results regarding the required number of system channels to ensure a maximum $\mathrm{MTTF}_{\mathrm{sys}}$ is comparable to the results of Chapter 6, i.e., close-to-optimal performance can be achieved by providing 1.1–1.2 channels per agent in the system, on both the first and second hop.

Second, a more realistic PHY model could be employed for a more accurate assessment of the required amount of PHY resources to achieve a certain target packet

loss probability. These characteristics might include the (agent-specific) mid-term average SNR, modulation, coding, and also statistical channel characteristics (because it is not only important how many resources are spent but also how correlated they are), which lead to a variable set of agent-specific PLRs, which heavily impacts the given modeling approach. This improved model could yield a more realistic network cost evaluation than the abstract *number of channels* used in this thesis. Also, the integer constraint of this thesis would be dropped, which resulted from the assumption of utilizing dynamic MC to enhance the packet delivery success probability. Moreover, varying distances of agents would be factored in, expressed in terms of mid-term average SNR. As a further step, it might be useful to evaluate the sensitivity of a wrong mapping (required resources $\rightarrow$ target packet loss probability) since it was shown in this thesis that the resulting reliability KPIs heavily depend on p_{loss}.

Third, acknowledging different sampling periods for agents would also provide more realistic insights of SARA when operated in a realistic networked system. In this thesis, a strict time synchronization requirement was assumed as this is repeatedly stated as a key requirement for industrial automation. However, guaranteeing transmission slots for a set of fixed (periodic) but different intertransmission intervals is a challenging sampling-period-constrained packing problem, which certainly limits network accessibility because new applications that violate a non-overlapping solution to the packing problem would be rejected. Conversely, if the network is loaded beyond this threshold, network resources will be systematically "overbooked", which leads to a system-induced reduction of packet success rates that are considerably more challenging to assess than the channel limitations in Chapter 6.

Fourth, more than two hops may be of interest for ultra-distributed applications. This extension could be performed in a straight-forward fashion by extending the dual-hop failure model by a third, forth, ..., i-th dimension and applying the same SARA optimization to the i-dimensional-hop case.

Fifth, ideal ACKs were assumed in the DL and ideal transmissions for network control information were assumed in the UL (the set of channels to be used for UL transmission) in this thesis. It was shown in Sec. 5.3 that erroneous ACKs only have a small effect on SARA efficiency. State estimation failures only cause a minimal resource consumption increase while maintaining desired control-related MTTF values when they are treated conservatively, i.e., a packet is considered lost if an ACK is not received. Such a conservative failure strategy is challenging to implement in the UL for the channel allocation control information as all further UL network resources other than the default channel assigned via SPS are managed by the

network and cannot be used without an explicit grant. A possible extension is to evaluate the impact of this effect on all presented KPIs as well as to develop a UL channel assignment procedure that limits the negative effects.

Sixth, the integration of SARA in a multi-cell system could be evaluated. For this thesis, it was assumed that wireless resources can be assigned only once in the whole network. However, in realistic distributed scenarios multiple cells exist that require a frequency reuse. In future work, the inter-cell interference should be taken into account as well because this significantly impacts the packet success probability.

Lastly, a joint AoI-based optimization can be performed that also factors in the base sampling period of the control system. Taking the peak AoI as the upper bound for the AoI, the fixed sampling period T_s directly yields the number of packets that can be lost until a control system failure occurs. It was observed for high packet loss tolerances K that not constraining the number of parallel channels $l \geq 1$ yielded $l_k = 0$ for small k during SARA optimization, which effectively translates into a change of the *base sampling period* T_s. In other words, the optimization result was to systematically not assign any channels if the last transmission was successful. In this thesis, this was prevented for conciseness and also because the base sampling period crucially determines the phase margin and overall smoothness of operation [FWP97]. In future work, however, it might be beneficial for control systems to introduce oversampling as an additional degree of freedom for optimal AoI-based SARA solutions. This increases the possibility for dynamic adjustments as more consecutive packets can be lost until the peak AoI is exceeded, however, the PHY layer resource requirements may be kept low by transmitting highly aggressively (high modulation, high coding rate) in low-SUA states. As a side effect, higher sampling periods increase the predictability of signals on the application level in case of packet losses, compare Chapter 3. Furthermore, a dynamic adjustment of the sampling period might also aid in solving the packing problem mentioned above because the usage of sampling period integer multiples simplifies avoiding unwanted overlaps.

A DC Motor Model

Open-loop motor dynamics may be modeled through the continuous-time state-space representation [Kri01]

$$\boldsymbol{A}_\mathrm{m} = \begin{bmatrix} -(B_\mathrm{m}+B_\mathrm{l})/(J_\mathrm{m}+J_\mathrm{l}) & K_\mathrm{t}/(J_\mathrm{m}+J_\mathrm{l}) \\ -K_\mathrm{e}/L_\mathrm{m} & -R_\mathrm{m}/L_\mathrm{m} \end{bmatrix} \qquad \boldsymbol{B}_\mathrm{m} = \begin{bmatrix} 0 \\ 1/L_\mathrm{m} \end{bmatrix}$$
$$\boldsymbol{C}_\mathrm{m} = \begin{bmatrix} 1 & 0 \end{bmatrix} \qquad \boldsymbol{D}_\mathrm{m} = \begin{bmatrix} 0 \end{bmatrix} \ . \tag{A.1}$$

The parameters used in this thesis are described in Tab. A.1 and their corresponding values are given. The system states are $\dot{\theta}_\mathrm{m}$ (rotational velocity) and I_m (armature current), the input is V_m (input voltage), and the output is $\dot{\theta}_\mathrm{m}$. For this open-loop motor model, a linear-quadratic regulator controller K_LQR is designed for a minimum rise time while still complying with the motor specifications, which are also given in Tab. A.1. Subsequently, a precompensation factor $\bar{N}$ is required to ensure a zero steady-state error. The continuous-time model relating a desired input acceleration $A_\mathrm{agv}(s)$ to the actual AGV position $Y_\mathrm{agv}(s)$ is detailed in Fig. A.1.

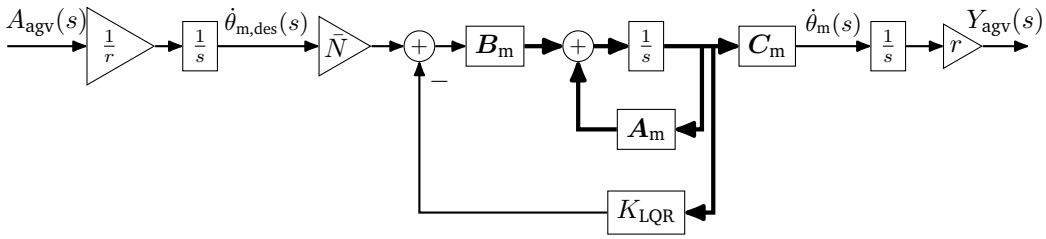


Fig. A.1.: The motor model includes an inner control loop that follows a given desired $\dot{\theta}_\mathrm{m,des}$. The external circuitry converts from acceleration to rotational velocity (input) and from rotational velocity to position (output).

Tab. A.1.: Summary of DC motor and AGV physical properties. [ebm19]

Param.	Value	Unit	Description
J_m	3.42×10^{-6}	$\mathrm{kg\,m^2}$	rotor moment of inertia
B_m	1.2669×10^{-4}	$\mathrm{N\,m\,s}$	motor viscous friction
K_e	237.2×10^{-3}	V/rad/s	electromotive force const.
K_t	0.0378	Nm/A	motor torque const.
R_m	0.85	Ω	electr. resistance
L_m	1100×10^{-3}	H	electr. inductance
m_l	50	kg	load mass
μ_l	0.001	–	friction coefficient
r	10	cm	wheel radius
B_l	$9.81 \times \mu_\mathrm{l} m_\mathrm{l} r = 49$	$\mathrm{mN\,m}$	wheel friction
J_l	$m_\mathrm{l} r^2 = 0.5$	$\mathrm{kg\,m^2}$	load inertia
$M_\mathrm{m,n}$	110	$\mathrm{mN\,m}$	nominal torque
$V_\mathrm{m,n}$	24	V	nominal voltage
$I_\mathrm{m,n}$	2.5	A	nominal current
$M_\mathrm{m,start}$	480	$\mathrm{mN\,m}$	starting torque
$n_\mathrm{m,n}$	4000	$\mathrm{min^{-1}}$	rated speed
$I_\mathrm{m,max}$	14	A	maximum current

The conversion of this continuous-time dynamic system model to discrete time is outlined in Sec. 3.1.2 and the final discrete-time state-space representation of the plant P_agv, sampled at $T_\mathrm{s} = 45\,\mathrm{ms}$, is given by

$$\boldsymbol{A}_\mathrm{d,P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.1241 & 0.8715 & 0.005\,19 & 0 \\ 26.51 & -25.96 & 0.3497 & 0 \\ 1.94 \times 10^{-3} & 4.3 \times 10^{-2} & 1.34 \times 10^{-4} & 1 \end{bmatrix} \qquad \boldsymbol{B}_\mathrm{d,P} = \begin{bmatrix} 0.045 \\ 1.94 \times 10^{-3} \\ 0.6828 \\ 2.16 \times 10^{-5} \end{bmatrix}$$
$$\boldsymbol{C}_\mathrm{d,P} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{D}_\mathrm{d,P} = \begin{bmatrix} 0 \end{bmatrix} . \tag{A.2}$$

Bibliography

[3GP16] 3GPP, "Specification # 36.881 - study on Latency Reduction Techniques for LTE", Technical Specification, 2016. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2901 (visited on Apr. 2, 2021).

[3GP19a] 3GPP, "Specification # 22.104 - Service requirements for cyber-physical control applications in vertical domains", Technical Specification, 2019. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3528 (visited on Jul. 8, 2019).

[3GP19b] 3GPP, "Specification # 22.261 - Service requirements for next generation new services and markets", Technical Specification, 2019. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3107 (visited on Sep. 25, 2019).

[3GP21a] 3GPP, "Specification # 38.101 - NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone", Technical Specification, 2021. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3283 (visited on Apr. 2, 2021).

[3GP21b] 3GPP, "Specification # 38.214 - NR; Physical layer procedures for data", Technical Specification, 2021. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3216 (visited on Apr. 2, 2021).

[3GP21c] 3GPP, "Specification # 38.300 - NR; NR and NG-RAN Overall description; Stage-2", Technical Specification, 2021. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3191 (visited on Apr. 2, 2021).

[5G 18] 5G Alliance for Connected Industries and Automation, "5G for Connected Industries and Automation", White Paper, 2018. [Online]. Available: https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/Whitepaper_5G_for_Connected_Industries_and_Automation/WP_5G_for_Connected_Industries_and_Automation_Download_19.03.19.pdf (visited on Apr. 21, 2021).

[5G 19] 5G Alliance for Connected Industries and Automation, "5G for Automation in Industry - Primary use cases, functions and service requirements", White Paper, 2019. [Online]. Available: https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_5G_for_Automation_in_Industry/WP_5G_for_Automation_in_Industry_final.pdf (visited on Apr. 21, 2021).

[5G 20] 5G Alliance for Connected Industries and Automation, "Key 5G Use Cases and Requirements", White Paper, 2020. [Online]. Available: https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_Key_5G_Use_Cases_and_Requirements/Key_5G_Use_Cases_and_Requirements_DOWNLOAD.pdf (visited on Apr. 21, 2021).

[Aca13] Acatech – Deutsche Akademie der Wissenschaften, "Recommendations for implementing the strategic initiative INDUSTRIE 4.0", Technical Report, 2013. [Online]. Available: https://en.acatech.de/publication/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/ (visited on Feb. 20, 2022).

[AMA19] N. Ahmed, A. Mansoor, and R. Ahmad, "Mission-Critical Machine-Type Communications: An Overview and Perspectives towards 5G", *IEEE Access*, 2019. DOI: 10.1109/ACCESS.2019.2894263.

[Aij20] A. Aijaz, "Private 5G: The Future of Industrial Wireless", *IEEE Industrial Electronics Magazine*, 2020. DOI: 10.1109/MIE.2020.3004975.

[Aij+17] A. Aijaz, M. Simsek, M. Dohler, and G. Fettweis, "Shaping 5G for the Tactile Internet", in *5G Mobile Communications*, Springer International Publishing, 2017. DOI: 10.1007/978-3-319-34208-5_25.

[Ana+21] P. M. d. S. Ana, N. Marchenko, P. Popovski, and B. Soret, "Age of Loop for Wireless Networked Control Systems Optimization", *arXiv:2106.00415*, 2021.

[AH14] D. Antunes and W. P. M. H. Heemels, "Rollout Event-Triggered Control: Beyond Periodic Control Performance", *IEEE Transactions on Automatic Control*, 2014. DOI: 10.1109/TAC.2014.2351932.

[AKC18] A. Avranas, M. Kountouris, and P. Ciblat, "Energy-Latency Tradeoff in Ultra-Reliable Low-Latency Communication With Retransmissions", *IEEE Journal on Selected Areas in Communications*, 2018. DOI: 10.1109/JSAC.2018.2874143.

[Aya+19] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems", in *IEEE International Conference on Cyber-Physical Systems*, 2019. DOI: 10.1145/3302509.3311050.

[Bau+14] N. W. Bauer, S. B. van Loon, N. van de Wouw, and W. M. Heemels, "Exploring the Boundaries of Robust Stability Under Uncertain Communication: An NCS Toolbox Applied to a Wireless Control Setup", *IEEE Control Systems Magazine*, 2014. DOI: 10.1109/MCS.2014.2320394.

[Bec18] Beckhoff Automation GmbH, "EtherCAT – The Ethernet Fieldbus", Technical Report, 2018. [Online]. Available: https://www.ethercat.org/download/documents/ETG_Brochure_EN.pdf (visited on Oct. 2021).

[BMF07] J. H. Braslavsky, R. H. Middleton, and J. S. Freudenberg, "Feedback Stabilization Over Signal-to-Noise Ratio Constrained Channels", *IEEE Transactions on Automatic Control*, 2007. DOI: 10.1109/TAC.2007.902739.

[Bun20] Bundesnetzagentur, "Zahlreiche Frequenzzuteilungen für 5G-Campusnetze", Press Statement, 2020. [Online]. Available: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Presse/Pressemitteilungen/2020/20200921_Campusnetze.pdf?__blob=publicationFile&v=2 (visited on Oct. 28, 2021).

[Bun21] Bundesnetzagentur, "WLAN-Nutzungen nun auch im 6 GHz-Bereich", Press Statement, 2021. [Online]. Available: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Presse/Pressemitteilungen/2020/20200921_Campusnetze.pdf?__blob=publicationFile&v=2 (visited on Oct. 28, 2021).

[Bur+21] F. Burmeister, N. Schwarzenberg, T. Hößler, and G. Fettweis, "Measuring Time-Varying Industrial Radio Channels for D2D Communications on AGVs", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2021.

[CL20] S. Cai and V. K. N. Lau, "MSE Tail Analysis for Remote State Estimation of Linear Systems Over Multiantenna Random Access Channels", *IEEE Transactions on Automatic Control*, 2020. DOI: 10.1109/TAC.2019.2937797.

[CH08] A. Cervin and T. Henningsson, "Scheduling of event-triggered controllers on a shared network", in *IEEE Conference on Decision and Control*, 2008. DOI: 10.1109/CDC.2008.4738939.

[Cha+19] 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 (INFOCOM WKSHPS)*, 2019. DOI: 10.1109/INFCOMW.2019.8845114.

[Che+14] D. Chen, M. Nixon, S. Han, A. K. Mok, and X. Zhu, "WirelessHART and IEEE 802.15.4e", in *IEEE International Conference on Industrial Technology (ICIT)*, 2014. DOI: 10.1109/ICIT.2014.6895027.

[Chi+22] F. Chiariotti, J. Holm, A. E. Kalør, B. Soret, S. K. Jensen, T. B. Pedersen, and P. Popovski, "Query Age of Information: Freshness in Pull-Based Communication", *IEEE Transactions on Communications*, 2022. DOI: 10.1109/TCOMM.2022.3141786.

[CSP21] F. Chiariotti, B. Soret, and P. Popovski, "Peak Age of Information Distribution Bounds for Multi-Connectivity Transmissions", in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021. DOI: 10.1109/SPAWC51858.2021.9593271.

[Chi+21] F. Chiariotti, O. Vikhrova, B. Soret, and P. Popovski, "Peak Age of Information Distribution for Edge Computing With Wireless Links", *IEEE Transactions on Communications*, 2021. DOI: 10.1109/TCOMM.2021.3053038.

[Chu+16] T. D. Chung, R. Ibrahim, V. S. Asirvadam, N. Saad, and S. M. Hassan, "Latency analysis of WirelessHART control message with variable payload", in *IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*, 2016. DOI: 10.1109/ROMA.2016.7847798.

[Cis20] Cisco, Inc., "Cisco Annual Internet Report (2018–2023)", Technical Report, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf (visited on Aug. 24, 2021).

[CSP14] G. Corrales Madueño, Č. Stefanović, and P. Popovski, "Reliable Reporting for Massive M2M Communications With Periodic Resource Pooling", *IEEE Wireless Communications Letters*, 2014. DOI: 10.1109/LWC.2014.2326674.

[CCE16] M. Costa, M. Codreanu, and A. Ephremides, "On the Age of Information in Status Update Systems With Packet Management", *IEEE Transactions on Information Theory*, 2016. DOI: 10.1109/TIT.2016.2533395.

[CCE14] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management", in *2014 IEEE International Symposium on Information Theory*, 2014. DOI: 10.1109/ISIT.2014.6875100.

[CFM05] O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete-Time Markov Jump Linear Systems*. London: Springer-Verlag, 2005. DOI: 10.1007/b138575.

[DZ83] J. Day and H. Zimmermann, "The OSI reference model", *Proceedings of the IEEE*, 1983. DOI: 10.1109/PROC.1983.12775.

[Dev+19] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal, "Reliable Transmission of Short Packets Through Queues and Noisy Channels Under Latency and Peak-Age Violation Guarantees", *IEEE Journal on Selected Areas in Communications*, 2019. DOI: 10.1109/JSAC.2019.2898760.

[Don+11] M. C. F. Donkers, W. P. M. H. Heemels, N. van de Wouw, and L. Hetel, "Stability Analysis of Networked Control Systems Using a Switched Linear Systems Approach", *IEEE Transactions on Automatic Control*, 2011. DOI: 10.1109/TAC.2011.2107631.

[DKP16] G. Durisi, T. Koch, and P. Popovski, "Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets", *Proceedings of the IEEE*, 2016. DOI: 10.1109/JPROC.2016.2537298.

[ebm19] ebmpapst, "Bürstenlose Innenläufer Servomotoren Baureihe ECI", Technical Specification, 2019. [Online]. Available: https://www.ebmpapst.com/content/dam/ebm-papst/media/catalogs/products/Katalog_ECI_DE.pdf (visited on Jul. 29, 2021).

[Eis+19] M. Eisen, M. M. Rashid, K. Gatsis, D. Cavalcanti, N. Himayat, and A. Ribeiro, "Control Aware Radio Resource Allocation in Low Latency Wireless Control Systems", *IEEE Internet of Things Journal*, 2019. DOI: 10.1109/JIOT.2019.2909198.

[EE11] N. Elia and J. N. Eisenbeis, "Limitations of Linear Control Over Packet Drop Networks", *IEEE Transactions on Automatic Control*, 2011. DOI: 10.1109/TAC.2010.2080150.

[EM01] N. Elia and S. Mitter, "Stabilization of linear systems with limited information", *IEEE Transactions on Automatic Control*, 2001. DOI: 10.1109/9.948466.

[Eur17] European Telecommunications Standards Institute (ETSI), "TS 36.101 – LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception", Technical Specification, 2017. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/136100_136199/136101/14.03.00_60/ts_136101v140300p.pdf (visited on Mar. 1, 2022).

[Fet14] G. P. Fettweis, "The Tactile Internet: Applications and Challenges", *IEEE Vehicular Technology Magazine*, 2014. DOI: 10.1109/MVT.2013.2295069.

[Fra+21] N. Franchi, T. Hößler, L. Scheuvens, N. Schwarzenberg, W. Anwar, A. Traßl, and G. P. Fettweis, "Selected Aspects and Approaches on Improving Dependability in Industrial Radio Networks", in *Wireless Networks and Industrial IoT: Applications, Challenges and Enablers*, Springer International Publishing, 2021. DOI: 10.1007/978-3-030-51473-0_2.

[FPE09] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*. Upper Saddle River, NJ, USA: Prentice Hall, 2009.

[FWP97] G. F. Franklin, M. L. Workman, and J. D. Powell, *Digital Control of Dynamic Systems*. Addison-Wesley Longman Publishing Co., Inc., 1997.

[Gan+16] K. Ganesan, T. Soni, S. Nunna, and A. R. Ali, "Poster: A TDM approach for latency reduction of ultra-reliable low-latency data in 5G", in *IEEE Vehicular Networking Conference (VNC)*, 2016. DOI: 10.1109/VNC.2016.7835946.

[GC08] H. Gao and T. Chen, "Network-Based H_∞ Output Tracking Control", *IEEE Transactions on Automatic Control*, 2008. DOI: 10.1109/TAC.2008.919850.

[GHP18] K. Gatsis, H. Hassani, and G. J. Pappas, "Latency-Reliability Tradeoffs for State Estimation", *arXiv:1810.11831*, 2018.

[Gat+15] K. Gatsis, M. Pajic, A. Ribeiro, and G. J. Pappas, "Opportunistic Control Over Shared Wireless Channels", *IEEE Transactions on Automatic Control*, 2015. DOI: 10.1109/TAC.2015.2416922.

[Gra18] Grand View Research, "Automated Guided Vehicles (AGV) Market Analysis Report By Type (Tow Vehicle, Unit Load Carrier), By Navigation Technology, By Application (Logistics & Warehousing), By End-use Industry, And Segment Forecasts, 2018 - 2025", Market Analysis, 2018. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/automated-guided-vehicle-agv-market (visited on Mar. 1, 2019).

[GS12] C. Grinstead and J. Snell, *Introduction to Probability*. American Mathematical Society, 2012.

[GSM20] GSM Association, "The State of Mobile Internet Connectivity 2020", Technical Specification, 2020. [Online]. Available: https://www.gsma.com/r/wp-content/uploads/2020/09/GSMA-State-of-Mobile-Internet-Connectivity-Report-2020.pdf (visited on Aug. 24, 2021).

[Gun+18] M. Gundall, J. Schneider, H. D. Schotten, M. Aleksy, D. Schulz, N. Franchi, N. Schwarzenberg, C. Markwart, R. Halfmann, P. Rost, D. Wübben, A. Neumann, M. Düngen, T. Neugebauer, R. Blunk, M. Kus, and J. Grießbach, "5G as Enabler for Industrie 4.0 Use Cases: Challenges and Concepts", in *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2018. DOI: 10.1109/ETFA.2018.8502649.

[Han+15] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic Event-Triggered Sensor Schedule for Remote State Estimation", *IEEE Transactions on Automatic Control*, 2015. DOI: 10.1109/TAC.2015.2406975.

[Han+17] D. Han, J. Wu, H. Zhang, and L. Shi, "Optimal sensor scheduling for multiple linear dynamical systems", *Automatica*, 2017. DOI: 10.1016/j.automatica.2016.09.015.

[Hee+10] W. P. M. H. Heemels, A. R. Teel, N. van de Wouw, and D. Nešić, "Networked Control Systems With Communication Constraints: Tradeoffs Between Transmission Intervals, Delays and Performance", *IEEE Transactions on Automatic Control*, 2010. DOI: 10.1109/TAC.2010.2042352.

[HNX07] J. Hespanha, P. Naghshtabrizi, and Y. Xu, "A Survey of Recent Results in Networked Control Systems", *Proceedings of the IEEE*, 2007. DOI: 10.1109/JPROC.2006.887288.

[HSF18a] T. Hößler, M. Simsek, and G. P. Fettweis, "Mission Reliability for URLLC in Wireless Networks", *IEEE Communications Letters*, 2018. DOI: 10.1109/LCOMM.2018.2868956.

[Höß20] T. Hößler, "On Dependable Wireless Communications through Multi-Connectivity", Ph.D. dissertation, Technische Universität Dresden, 2020.

[Höß+17] T. Hößler, L. Scheuvens, N. Franchi, M. Simsek, and G. P. Fettweis, "Applying reliability theory for future wireless communication networks", in *IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017. DOI: 10.1109/PIMRC.2017.8292773.

[Höß+20] T. Hößler, L. Scheuvens, P. Schulz, A. Noll Barreto, M. Simsek, and G. P. Fettweis, "Dynamic Connectivity for Robust Applications in Rayleigh-Fading Channels", *IEEE Communications Letters*, 2020. DOI: 10.1109/LCOMM.2019.2954093.

[Höß+19] T. Hößler, P. Schulz, M. Simsek, and G. P. Fettweis, "Mission Availability for Wireless URLLC", in *IEEE Global Communications Conference (GLOBECOM)*, 2019. DOI: 10.1109/GLOBECOM38437.2019.9013362.

[HSF18b] T. Hößler, M. Simsek, and G. P. Fettweis, "Joint Analysis of Channel Availability and Time-Based Reliability Metrics for Wireless URLLC", in *IEEE Global Communications Conference (GLOBECOM)*, 2018. DOI: 10.1109/GLOCOM.2018.8647801.

[Hua+20] K. Huang, W. Liu, Y. Li, A. Savkin, and B. Vucetic, "Wireless Feedback Control with Variable Packet Length for Industrial IoT", *arXiv:2005.13321*, 2020.

[Hua+19] K. Huang, W. Liu, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Real-Time Remote Estimation with Hybrid ARQ in Wireless Networked Control", *arXiv:1903.-12472*, 2019.

[ITU15] ITU-R, "IMT Vision - Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond", Technical Report, 2015. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf (visited on Aug. 24, 2021).

[ITU20] ITU-T, "Definitions and Terminology for Synchronization in Packet Networks, G.8260", Technical Report, 2020. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.8260-202003-I!!PDF-E&type=items (visited on Aug. 24, 2021).

[Ji+91] Y. Ji, H. Chizeck, X.-Y. Feng, and K. Loparo, "Stability and control of discrete-time jump linear systems", 1991.

[KSM18] I. Kadota, A. Sinha, and E. Modiano, "Optimizing Age of Information in Wireless Networks with Throughput Constraints", in *IEEE Conference on Computer Communications*, 2018. DOI: 10.1109/INFOCOM.2018.8486307.

[KYG12] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?", in *Proceedings IEEE INFOCOM*, 2012. DOI: 10.1109/INFCOM.2012.6195689.

[Kim17] S.-H. Kim, *Electric Motor Control*. Elsevier, 2017.

[KP20] E. Koeze and N. Popper, "The Virus Changed the Way We Internet", News Article, 2020. [Online]. Available: https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html (visited on Aug. 26, 2021).

[Kos+17] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case", in *IEEE International Symposium on Information Theory (ISIT)*, 2017. DOI: 10.1109/ISIT.2017.8006543.

[Kri01] R. Krishnan, *Electric Motor Drives: Modeling, Analysis, and Control*. Pearson, 2001.

[Läh+14] E. Lähetkangas, K. Pajukoski, J. Vihriälä, G. Berardinelli, M. Lauridsen, E. Tiirola, and P. Mogensen, "Achieving low latency and energy consumption by 5G TDD mode optimization", in *IEEE International Conference on Communications Workshops (ICC)*, 2014. DOI: 10.1109/ICCW.2014.6881163.

[LL19] M. Latvo-Aho and K. Leppänen, "Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence", White Paper, 2019. [Online]. Available: http://urn.fi/urn:isbn:9789526223544 (visited on Oct. 20, 2021).

[LCY20] V. K. N. Lau, S. Cai, and M. Yu, "Decentralized State-Driven Multiple Access and Information Fusion of Mission-Critical IoT Sensors for 5G Wireless Networks", *IEEE Journal on Selected Areas in Communications*, 2020. DOI: 10.1109/JSAC.2020.2980914.

[Li+17] X. Li, D. Li, J. Wan, A. V. Vasilakos, C.-F. Lai, and S. Wang, "A review of industrial wireless networks in the context of Industry 4.0", *Wireless Networks*, 2017. DOI: 10.1007/s11276-015-1133-7.

[LW16] C.-H. Liu and L.-C. Wang, "Optimal Cell Load and Throughput in Green Small Cell Networks With Generalized Cell Association", *IEEE Journal on Selected Areas in Communications*, 2016. DOI: 10.1109/JSAC.2016.2520218.

[Liu+20] 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", *arXiv:2007.00256*, 2020.

[LH90] E. Lo and R. Hardy, "Indoor wireless LAN access methods for factories", in *IEEE Conference on Vehicular Technology*, 1990. DOI: 10.1109/VETEC.1990.110305.

[Lu+16] C. Lu, A. Saifullah, B. Li, M. Sha, H. Gonzalez, D. Gunatilaka, C. Wu, L. Nie, and Y. Chen, "Real-Time Wireless Sensor-Actuator Networks for Industrial Cyber-Physical Systems", *Proceedings of the IEEE*, 2016. DOI: 10.1109/JPROC.2015.2497161.

[Mam+14] M. H. Mamduhi, D. Tolić, A. Molin, and S. Hirche, "Event-triggered scheduling for stochastic multi-loop networked control systems with packet dropouts", in *IEEE Conference on Decision and Control*, 2014. DOI: 10.1109/CDC.2014.7039815.

[MCF13] P. Minero, L. Coviello, and M. Franceschetti, "Stabilization Over Markov Feedback Channels: The General Case", *IEEE Transactions on Automatic Control*, 2013. DOI: 10.1109/TAC.2012.2212616.

[Nai+07] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback Control Under Data Rate Constraints: An Overview", *Proceedings of the IEEE*, 2007. DOI: 10.1109/JPROC.2006.887294.

[Neu94] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Courier Corporation, 1994.

[NLP17] J. J. Nielsen, R. Liu, and P. Popovski, "Optimized Interface Diversity for Ultra-Reliable Low Latency Communication (URLLC)", in *IEEE Global Communications Conference (GLOBECOM)*, 2017. DOI: 10.1109/GLOCOM.2017.8254053.

[NP16] J. J. Nielsen and P. Popovski, "Latency analysis of systems with multiple interfaces for ultra-reliable M2M communication", in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2016. DOI: 10.1109/SPAWC.2016.7536857.

[Nis00] N. S. Nise, *Control Systems Engineering*. John Wiley & Sons, Inc., 2000.

[NTT21] NTT DOCOMO, INC., "5G Evolution and 6G", White Paper, 2021. [Online]. Available: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_v3.0.pdf (visited on Oct. 20, 2021).

[ÖF15] D. Öhmann and G. P. Fettweis, "Minimum duration outage of wireless Rayleigh-fading links using selection combining", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2015. DOI: 10.1109/WCNC.2015.7127552.

[ÖTF15] D. Öhmann, W. Teske, and G. P. Fettweis, "Combining Nakagami-m Fading Links for High Availability in Wireless Networks", in *IEEE Vehicular Technology Conference (VTC Spring)*, 2015. DOI: 10.1109/VTCSpring.2015.7145753.

[Öst+19] J. Östman, R. Devassy, G. Durisi, and E. Uysal, "Peak-Age Violation Guarantees for the Transmission of Short Packets over Fading Channels", in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019. DOI: 10.1109/INFCOMW.2019.8845098.

[Par+18] 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*, 2018. DOI: 10.1109/COMST.2017.2780114.

[PDH15] S. Park, R. C. Daniels, and R. W. Heath, "Optimizing the Target Error Rate for Link Adaptation", in *IEEE Global Communications Conference (GLOBECOM)*, 2015. DOI: 10.1109/GLOCOM.2015.7417770.

[Pez+20] M. Pezzutto, F. Tramarin, S. Dey, and L. Schenato, "Adaptive transmission rate for LQG control over Wi-Fi: A cross-layer approach", *Automatica*, 2020. DOI: 10.1016/j.automatica.2020.109092.

[Poc+16] G. Pocovi, K. I. Pedersen, B. Soret, M. Lauridsen, and P. Mogensen, "On the impact of multi-user traffic dynamics on low latency communications", in *International Symposium on Wireless Communication Systems (ISWCS)*, 2016. DOI: 10.1109/ISWCS.2016.7600901.

[Poc+17] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC layer enhancements for ultra-reliable low-latency communications in cellular networks", in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017. DOI: 10.1109/ICCW.2017.7962790.

[Pop+19] P. Popovski, Č. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A. Bana, "Wireless Access in Ultra-Reliable Low-Latency Communication (URLLC)", *IEEE Transactions on Communications*, 2019. DOI: 10.1109/TCOMM.2019.2914652.

[Pop14] P. Popovski, "Ultra-reliable communication in 5G wireless systems", in *International Conference on 5G for Ubiquitous Connectivity*, 2014. DOI: 10.4108/icst.5gu.2014.258154.

[Pop+21a] P. Popovski, F. Chiariotti, V. Croisfelt, A. E. Kalør, I. Leyva-Mayorga, L. Marchegiani, S. R. Pandey, and B. Soret, "Internet of Things (IoT) Connectivity in 6G: An Interplay of Time, Space, Intelligence, and Value", *arXiv:2111.05811*, 2021. DOI: 10.48550/arXiv.2111.05811.

[Pop+21b] P. Popovski, F. Chiariotti, K. Huang, A. E. Kalør, M. Kountouris, N. Pappas, and B. Soret, "A Perspective on Time towards Wireless 6G", *arXiv:2106.04314*, 2021.

[Pop+18a] P. Popovski, J. J. Nielsen, C. Stefanovic, E. de Carvalho, E. Strom, K. F. Trillingsgaard, A.-S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks", *IEEE Network*, 2018. DOI: 10.1109/MNET.2018.1700258.

[Pop+18b] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View", *IEEE Access*, 2018. DOI: 10.1109/ACCESS.2018.2872781.

[QAJ13] D. E. Quevedo, A. Ahlen, and K. H. Johansson, "State Estimation Over Sensor Networks With Correlated Wireless Fading Channels", *IEEE Transactions on Automatic Control*, 2013. DOI: 10.1109/TAC.2012.2212515.

[Rap02] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 2002.

[Roj17] A. Rojko, "Industry 4.0 Concept: Background and Overview", *International Journal of Interactive Mobile Technologies*, 2017.

[Sac+18] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai, "5G Radio Network Design for Ultra-Reliable Low-Latency Communication", *IEEE Network*, 2018. DOI: 10.1109/MNET.2018.1700232.

[Sch+07] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of Control and Estimation Over Lossy Networks", *Proceedings of the IEEE*, 2007. DOI: 10.1109/JPROC.2006.887306.

[Sch+19] L. Scheuvens, T. Hößler, A. Noll Barreto, and G. P. Fettweis, "Wireless Control Communications Co-Design via Application-Adaptive Resource Management", in *IEEE 5G World Forum (5GWF)*, 2019. DOI: 10.1109/5GWF.2019.8911713.

[Sch+21] L. Scheuvens, T. Hößler, P. Schulz, N. Franchi, A. Barreto, and G. P. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems", *IEEE Transactions on Communications*, 2021. DOI: 10.1109/TCOMM.2021.3095953.

[Sch+20a] L. Scheuvens, P. Schulz, T. Hößler, N. Franchi, A. Barreto, and G. P. Fettweis, "System Analysis of State-Aware Resource Allocation for Closed-Loop Control Systems", in *IEEE Global Communications Conference (GLOBECOM)*, 2020.

[Sch+20b] L. Scheuvens, P. Schulz, T. Hößler, N. Franchi, A. Noll Barreto, and G. P. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control based on Multi-Connectivity", in *IEEE Global Communications Conference (GLOBECOM)*, 2020.

[Sch+17] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture", *IEEE Communications Magazine*, 2017. DOI: 10.1109/MCOM.2017.1600435CM.

[SS01] P. Seiler and R. Sengupta, "Analysis of communication losses in vehicle control problems", in *Proceedings of the 2001 American Control Conference*, 2001. DOI: 10.1109/ACC.2001.945935.

[SS05] P. Seiler and R. Sengupta, "An H_∞ approach to networked control", *IEEE Transactions on Automatic Control*, 2005. DOI: 10.1109/TAC.2005.844177.

[Sha+18] A. Shapin, K. Kittichokechar, N. Andgart, M. Sundberg, and G. Wikström, "Physical Layer Performance for Low Latency and High Reliability in 5G", in *International Symposium on Wireless Communication Systems (ISWCS)*, 2018. DOI: 10.1109/ISWCS.2018.8491084.

[Sha+16] H. Shariatmadari, S. Iraji, Z. Li, M. A. Uusitalo, and R. Jäntti, "Optimized transmission and resource allocation strategies for ultra-reliable communications", in *IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016. DOI: 10.1109/PIMRC.2016.7794801.

[Sha+15] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: Current status and future perspectives toward 5G systems", *IEEE Communications Magazine*, 2015. DOI: 10.1109/MCOM.2015.7263367.

[SS17] A. Sharma and M. Salim, "Polar Code: The Channel Code contender for 5G scenarios", in *International Conference on Computer, Communications and Electronics (Comptelix)*, 2017. DOI: 10.1109/COMPTELIX.2017.8004055.

[SEM10] L. Shi, M. Epstein, and R. M. Murray, "Kalman Filtering Over a Packet-Dropping Network: A Probabilistic Perspective", *IEEE Transactions on Automatic Control*, 2010. DOI: 10.1109/TAC.2009.2039236.

[Sim+16] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. P. Fettweis, "5G-Enabled Tactile Internet", *IEEE Journal on Selected Areas in Communications*, 2016. DOI: 10.1109/JSAC.2016.2525398.

[Sim+17] M. Simsek, D. Zhang, D. Öhmann, M. Matthé, and G. Fettweis, "On the Flexibility and Autonomy of 5G Wireless Networks", *IEEE Access*, 2017. DOI: 10.1109/ACCESS.2017.2715318.

[Sin+04] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations", *IEEE Transactions on Automatic Control*, 2004. DOI: 10.1109/TAC.2004.834121.

[SRP20] B. Soret, S. Ravikanti, and P. Popovski, "Latency and timeliness in multi-hop satellite networks", in *IEEE International Conference on Communications (ICC)*, 2020. DOI: 10.1109/ICC40277.2020.9149009.

[Sun+16] Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh", in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2016. DOI: 10.1109/INFOCOM.2016.7524524.

[Syb+16] M. Sybis, K. Wesolowski, K. Jayasinghe, V. Venkatasubramanian, and V. Vukadinovic, "Channel Coding for Ultra-Reliable Low-Latency Communication in 5G Systems", in *IEEE Vehicular Technology Conference (VTC-Fall)*, 2016. DOI: 10.1109/VTCFall.2016.7880930.

[TM04] S. Tatikonda and S. Mitter, "Control under communication constraints", *IEEE Transactions on Automatic Control*, 2004. DOI: 10.1109/TAC.2004.831187.

[Tes+16a] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Evaluation of adaptive active set management for multi-connectivity in intra-frequency 5G networks", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016. DOI: 10.1109/WCNC.2016.7564823.

[Tes+16b] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. Fettweis, "Evaluation of Context-Aware Mobility Robustness Optimization and Multi-Connectivity in Intra-Frequency 5G Ultra Dense Networks", *IEEE Wireless Communications Letters*, 2016. DOI: 10.1109/LWC.2016.2604817.

[Tul+14] H. Tullberg, Z. Li, A. Hoglund, P. Fertl, D. Gozalvez-Serrano, K. Pawlak, P. Popovski, G. Mange, and Ö. Bulakci, "Towards the METIS 5G concept: First view on Horizontal Topics concepts", in *European Conference on Networks and Communications (EuCNC)*, 2014. DOI: 10.1109/EuCNC.2014.6882694.

[Vik+20] O. Vikhrova, F. Chiariotti, B. Soret, G. Araniti, A. Molinaro, and P. Popovski, "Age of Information in Multi-hop Networks with Priorities", in *IEEE Global Communications Conference (GLOBECOM)*, 2020. DOI: 10.1109/GLOBECOM42002.2020.9348175.

[Wan+20] X. Wang, C. Chen, J. He, S. Zhu, and X. Guan, "AoI-Aware Control and Communication Co-design for Industrial IoT Systems", *IEEE Internet of Things Journal*, 2020. DOI: 10.1109/JIOT.2020.3046742.

[WL11] X. Wang and M. D. Lemmon, "Event-Triggering in Distributed Networked Control Systems", *IEEE Transactions on Automatic Control*, 2011. DOI: 10.1109/TAC.2010.2057951.

[Wei15] M. Weiner, "Low-Latency, High-Reliability Wireless Networks for Control Applications", Ph.D. dissertation, UC Berkeley, 2015.

[Wei+14] M. Weiner, M. Jorgovanovic, A. Sahai, and B. Nikolié, "Design of a low-latency, high-reliability wireless communication system for control applications", in *IEEE International Conference on Communications (ICC)*, 2014. DOI: 10.1109/ICC.2014.6883918.

[Wil08] A. Willig, "Recent and Emerging Topics in Wireless Industrial Communications: A Selection", *IEEE Transactions on Industrial Informatics*, 2008. DOI: 10.1109/TII.2008.923194.

[Wol+19] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. Santos Filho, and G. Fettweis, "How Reliable and Capable is Multi-Connectivity?", *IEEE Transactions on Communications*, 2019. DOI: 10.1109/TCOMM.2018.2873648.

[Wol+17] A. Wolf, P. Schulz, D. Öhmann, M. Dörpinghaus, and G. Fettweis, "On the Gain of Joint Decoding for Multi-Connectivity", in *IEEE Global Communications Conference (GLOBECOM)*, 2017. DOI: 10.1109/GLOCOM.2017.8254232.

[Wu+14] C. Wu, M. Sha, D. Gunatilaka, A. Saifullah, C. Lu, and Y. Chen, "Analysis of EDF scheduling for Wireless Sensor-Actuator Networks", in *IEEE International Symposium of Quality of Service (IWQoS)*, 2014. DOI: 10.1109/IWQoS.2014.6914298.

[Wu+13] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, "Event-Based Sensor Data Scheduling: Trade-Off Between Communication Rate and Estimation Quality", *IEEE Transactions on Automatic Control*, 2013. DOI: 10.1109/TAC.2012.2215253.

[WJ11] P. Wu and N. Jindal, "Coding versus ARQ in Fading Channels: How Reliable Should the PHY Be?", *IEEE Transactions on Communications*, 2011. DOI: 10.1109/TCOMM.2011.102011.100152.

[Yat+21] 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. DOI: 10.1109/JSAC.2021.3065072.

[YFX11] K. You, M. Fu, and L. Xie, "Mean square stability for Kalman filtering with Markovian packet losses", *Automatica*, 2011. DOI: 10.1016/j.automatica.2011.09.015.

[Yu+04] M. Yu, L. Wang, G. Xie, and T. Chu, "Stabilization of networked control systems with data packet dropout via switched system approach", in *IEEE International Conference on Robotics and Automation*, 2004. DOI: 10.1109/CACSD.2004.1393903.

[YQC04] D. Yue, Qing-Long Han, and Chen Peng, "State feedback controller design of networked control systems", *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2004. DOI: 10.1109/TCSII.2004.836043.

[ZGK13] L. Zhang, H. Gao, and O. Kaynak, "Network-Induced Constraints in Networked Control Systems — A Survey", *IEEE Transactions on Industrial Informatics*, 2013. DOI: 10.1109/TII.2012.2219540.

[ZBP01] W. Zhang, M. S. Branicky, and S. M. Phillips, "Stability of networked control systems", *IEEE Control Systems Magazine*, 2001. DOI: 10.1109/37.898794.

[ZHY16] X. Zhang, Q. Han, and X. Yu, "Survey on Recent Advances in Networked Control Systems", *IEEE Transactions on Industrial Informatics*, 2016. DOI: 10.1109/TII.2015.2506545.

[ZS19] B. Zhou and W. Saad, "Minimizing Age of Information in the Internet of Things with Non-Uniform Status Packet Sizes", in *IEEE International Conference on Communications (ICC)*, 2019. DOI: 10.1109/ICC.2019.8761311.

Publications of the Author

[Fet+19] G. P. Fettweis, K. Leo, B. Voit, U. Schneider, and **L. Scheuvens**, "Venturing Electronics into Unknown Grounds", in *Technical Digest - International Electron Devices Meeting (IEDM)*, 2019. DOI: 10.1109/IEDM.2018.8614545.

[Fra+21] N. Franchi, T. Hößler, **L. Scheuvens**, N. Schwarzenberg, W. Anwar, A. Traßl, and G. P. Fettweis, "Selected Aspects and Approaches on Improving Dependability in Industrial Radio Networks", in *Wireless Networks and Industrial IoT: Applications, Challenges and Enablers*, Springer International Publishing, 2021. DOI: 10.1007/978-3-030-51473-0_2.

[Höß+17] T. Hößler, **L. Scheuvens**, N. Franchi, M. Simsek, and G. P. Fettweis, "Applying reliability theory for future wireless communication networks", in *IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017. DOI: 10.1109/PIMRC.2017.8292773.

[Höß+20] T. Hößler, **L. Scheuvens**, P. Schulz, A. Noll Barreto, M. Simsek, and G. P. Fettweis, "Dynamic Connectivity for Robust Applications in Rayleigh-Fading Channels", *IEEE Communications Letters*, 2020. DOI: 10.1109/LCOMM.2019.2954093.

[Kle+17] H. Klessig, H. Kuntzschmann, **L. Scheuvens**, B. Almeroth, P. Schulz, and G. Fettweis, "Twitter as a source for spatial traffic information in big data-enabled self-organizing networks", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2017. DOI: 10.1109/WCNC.2017.7925565.

[Mah+20] A. H. Mahdi, T. Hößler, **L. Scheuvens**, N. Franchi, and G. Fettweis, "Multi-Connectivity Management for Mobile Ultra-Reliable Low-Latency Communications", in *International ITG Workshop on Smart Antennas*, 2020.

[Sch+19a] **L. Scheuvens**, T. Hößler, A. Noll Barreto, and G. P. Fettweis, "Wireless Control Communications Co-Design via Application-Adaptive Resource Management", in *IEEE 5G World Forum (5GWF)*, 2019. DOI: 10.1109/5GWF.2019.8911713.

[Sch+21] **L. Scheuvens**, T. Hößler, P. Schulz, N. Franchi, A. Barreto, and G. P. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems", *IEEE Transactions on Communications*, 2021. DOI: 10.1109/TCOMM.2021.3095953.

[Sch+20a] **L. Scheuvens**, P. Schulz, T. Hößler, N. Franchi, A. Barreto, and G. P. Fettweis, "System Analysis of State-Aware Resource Allocation for Closed-Loop Control Systems", in *IEEE Global Communications Conference (GLOBECOM)*, 2020.

[Sch+20b] **L. Scheuvens**, P. Schulz, T. Hößler, N. Franchi, A. Noll Barreto, and G. P. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control based on Multi-Connectivity", in *IEEE Global Communications Conference (GLOBECOM)*, 2020.

[Sch+19b] **L. Scheuvens**, M. Simsek, A. Noll-Barreto, N. Franchi, and G. P. Fettweis, "Framework for Adaptive Controller Design Over Wireless Delay-Prone Communication Channels", *IEEE Access*, 2019. DOI: 10.1109/ACCESS.2019.2911120.

[Tra+19a] A. Traßl, T. Hößler, **L. Scheuvens**, N. Franchi, and G. P. Fettweis, "Deriving an Empirical Channel Model for Wireless Industrial Indoor Communications", in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019. DOI: 10.1109/PIMRC.2019.8904382.

[Tra+21a] A. Traßl, T. Hößler, **L. Scheuvens**, N. Schwarzenberg, and G. P. Fettweis, "Outage Prediction for URLLC in Rician Fading", in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021. DOI: 10.1109/PIMRC50174.2021.9569370.

[Tra+19b] A. Traßl, **L. Scheuvens**, T. Hößler, N. Franchi, and G. P. Fettweis, "On Dependability Metrics for Wireless Industrial Communications - Applied to IEEE 802.11ax", in *IEEE 5G World Forum (5GWF)*, 2019.

[Tra+20] A. Traßl, **L. Scheuvens**, T. Hößler, E. Schmitt, N. Franchi, and G. P. Fettweis, "Outage Prediction for URLLC in Rayleigh Fading", in *European Conference on Networks and Communications (EuCNC)*, 2020. DOI: 10.1109/EuCNC48522.2020.9200953.

[Tra+21b] A. Traßl, E. Schmitt, T. Hößler, **L. Scheuvens**, N. Franchi, N. Schwarzenberg, and G. Fettweis, "Outage prediction for ultra-reliable low-latency communications in fast fading channels", *EURASIP Journal on Wireless Communications and Networking*, 2021. DOI: 10.1186/s13638-021-01964-w.

[Tra+22] A. Traßl, P. Schulz, **L. Scheuvens**, N. Schwarzenberg, and G. Fettweis, "On the Outage Probability of Channel Prediction Enabled Max-Min Radio Resource Allocation", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, *accepted*.

[Wat+18] S. Watzek, Y. Zou, **L. Scheuvens**, and G. Fettweis, "Opportunity Report - Towards Enhanced EU-US ICT Pre-competitive Collaboration", PICASSO Project, 2018. [Online]. Available: http://www.picasso-project.eu/wp-content/uploads/2018/05/PICASSO-Opportunity-Report-5G-05-2018.pdf (visited on Oct. 22, 2021).

[Zou+20] Y. Zou, G. Fettweis, A. Ghosh, G. Ricart, M. Latva-Aho, and **L. Scheuvens**, "Challenges and Potential for EU–US Collaboration in 5G and Beyond Networks", in *ICT Policy, Research, and Innovation*, John Wiley & Sons, Ltd, 2020. DOI: 10.1002/9781119632481.ch6.

List of Figures

List of Tables

List of Operators and Constants

$\mathbf{0}$	All-zeros matrix of appropriate dimensions
$\mathbf{1}$	All-ones matrix of appropriate dimensions
$\operatorname{blockdiag}\{\cdot\}$	Block diagonal
c	Speed of light
$(\cdot)^*$	Matrix conjugate
$\mathbb{E}\{\cdot\}$	Expectation value
e	Euler's number, $e = 2.718\,281\ldots$
$\in$	Element relation
$\boldsymbol{e}_i$	Unit vector in dimension i
$!$	Factorial
$\boldsymbol{I}$	Identity matrix of appropriate dimensions
∞	Infinity
$\otimes$	Kronecker product of matrices
$\mathbb{N}_0$	Set of natural numbers, incl. 0
$\mathbb{P}\{\cdot\}$	Probability value
π	Mathematical constant, $\pi = 3.141\,592\ldots$
$\prod$	Product of elements
$\sum$	Sum of elements
$(\cdot)^\intercal$	Matrix transpose

List of Symbols

B_{c}	Coherence bandwidth [Hz]
η	Ratio of time in which all channels are in use
f	Frequency [Hz]
f_{c}	Carrier frequency [Hz]
f_{d}	Maximum Doppler shift [Hz]
f_{PLSL}	PMF of the PLSL
f_{SUA}	PMF of the SUA
g	Penalty function
Γ	Combination function for individual penalty functions
k	Number of consecutive packet losses
k_{1h}	AoI after the first hop
k_{2h}	AoI after the second hop
K	Maximum number of tolerable consecutive packet losses
l	Number of channels simultaneously assigned to an agent
$\boldsymbol{l}_{(\cdot)}^{(\cdot)}$	Channel assignment scheme
$\bar{l}$	Longterm average number of channels assigned to an agent
l_{fix}	Number of fixed channels
$\hat{l}$	Maximum number of simultaneously assignable channels
L_{av}	Number of available channels in the system
L_{req}	Total number of requested system channels by all agents
M	Number of agents in the system
MTTF	Mean time to failure
n	Discrete time index
p	Probability of packet transmission success
p_0	Probability of not losing the first packet in an MJLS
p_{loss}	Packet loss probability for a single link
$\tilde{p}$	Probability of packet transmission failure
$\tilde{p}_0$	Probability of losing the first packet in a MJLS
PLR	Long-term ratio of lost packets divided by total number of packets
PLSL	Number of consecutive packet losses

π_i	Probability of being in a certain Markov state during operation
$\boldsymbol{\pi}$	State probability vector of an absorbing Markov chain
r_σ	Spectral radius (largest absolute eigenvalue) of a matrix
RMS	Root mean square
ρ	Temporal packet loss correlation
s	Laplace transform parameter
$s_{(\cdot)}$	Single-agent Markov state
$S_{(\cdot)}$	Multi-agent/system Markov state
SNR	Ratio of signal power to noise power
SUA	Time that has passed since the last update was received
σ_τ	RMS delay spread [s]
t	Time [s]
t'	Time at which the last successfully received packet was generated [s]
T_c	Coherence time [s]
T_s	Sampling period; time between two consecutive samples [s]
$T_\mathrm{s,eff}$	Effective sampling period; time between two consecutive successful packet transmissions [s]
TTF	Time until the next failure [s]
τ_d	Delay [s]
τ_surv	Survival time of a control system [s]
v_max	Maximum velocity
z	z-transform parameter
ζ	Operation mode in MJLS

List of Acronyms

5G	The fifth generation of cellular networks
ACK	Acknowledgment
AGV	Automated guided vehicle
AoI	Age of information
ARQ	Automatic repeat request
BS	Base station
CoCoCo	Co-design of control application and (wireless) communication
CSMA/CA	Carrier sense multiple access / collision avoidance
DC	Direct current
DL	Downlink
FOH	First order hold
HARQ	Hybrid ARQ
IoT	Internet of things
KPI	Key performance indicator
LTE	Long term evolution
LTI	Linear, time-invariant
LTV	Linear, time-variant
M2M	Machine-to-machine-type communications
MAC	Medium access control
MATLAB	*Matrix Laboratory*, numeric computing environment developed by *The Mathworks*
MC	Multi-connectivity
MJLS	Markov jump linear system
mMTC	Massive machine-type communications
μMTC	Mission-critical machine-type communications
MTTF	Mean time to failure
NACK	Negative acknowledgment
NCS	Networked control system
ODE	Ordinary differential equation
OSI	Open systems interconnection
PD	Proportional-derivative

PDF	Probability density function
PHY	Physical layer
PLR	Packet loss ratio
PLSL	Packet loss sequence length
PMF	Probability mass function
QoC	Quality of control
QoS	Quality of service
RMS	Root mean square
RX	Receive
SARA	State-aware resource allocation
SNR	Signal-to-noise ratio
SPS	Semi-persistent scheduling
SUA	Status update age
TTF	Time to failure
TX	Transmit
UL	Uplink
URLLC	Ultra-reliable low-latency communications
WLAN	Wireless local area network
ZOH	Zeroth order hold