This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953

State-Aware Resource Allocation for Wireless Closed-Loop Control Systems

Lucas Scheuvens, Tom Hößler, Philipp Schulz, Norman Franchi, André Barreto, and Gerhard P. Fettweis

***Abstract*—Wireless closed-loop control is of major significance for future industrial manufacturing. However, control applications pose stringent quality of service requirements for reliable operation. Contrary to traditional ultra-reliable low-latency communications design goals such as low packet loss rates and low latency, research results in the domain of networked control systems (NCS) state that depending on the sampling period, control applications inherently tolerate a few consecutive packet losses. This translates into a better-suited metric to capture control application requirements and therefore a more conclusive design goal for wireless networks: ensuring a maximum age of information (AoI). With a Markov modeling approach, we propose to exploit the tolerance through a novel dynamic multi-connectivity scheme that we term state-aware resource allocation (SARA), which temporally negatively correlates packet losses, thus avoiding long packet loss sequences. Through statistical multiplexing, SARA enables a mean time to failure (MTTF) in the order of years while keeping the per-agent average channel usage close to one, also in a multi-agent setting with competition for resources. Compared with static dual-connectivity, the MTTF can be increased 100-fold whereas the number of required channels reduces by 40%. Our approach also statistically guarantees system-wide AoI distributions, which aid to ensure control performance.**

***Index Terms*—Control Communications Co-Design, Age of Information, Wireless Control, Adaptive Resource Management, System Analysis, Availability**

I. Introduction

The fifth generation of cellular networks (5G) is a major enabler for digitalization across a wide spectrum of industries, including healthcare, manufacturing, automotive and energy. In manufacturing alone, the global business opportunities for service providers are expected to grow 76 % annually until 2030 (reaching $150bn), underlining the high demand in upcoming years [1]. In this context, industrial wireless closed-loop control systems are of major significance, since they will enable a hitherto unknown degree of flexibility and productivity [2], creating an immense societal impact. The shift towards high-performing and all-connected manufacturing is summarized under the umbrella term *Industry 4.0*. Obviously, this will require powerful networking capabilities: [3]–[6]

Performance Ultra-reliable and low-latency wireless connections must be provided to not disrupt running processes. Industrial manufacturers demand assertions on the expected duration of error-free operation of control applications originating from the uncertainty of the wireless connection. This includes potential connectivity disruption originating from interference in the same wireless channel, leading to the demand for dedicated spectrum for industrial usage.

Time synchronicity Low-latency, deterministic, and reliable communication requires a shared understanding of time between communicating devices, ensuring *predictable* behaviour.

Mobility Distributed real-time applications might span an area that cannot be served with a single access point (AP), also demanding real-time connectivity at cell boundaries and managed hand-offs between cells.

Integration Industrial manufacturing requests a direct and seamless wireless communication across the whole automation pyramid (from the field level to the cloud).

Diagnosis The network must offer transparent diagnostic information on the connectivity and expose this information to third parties. This is essential for root-cause analysis.

Standardized 5G communications systems build a solid foundation for many of these requirements, while common off-the-shelf, unmanaged, and unoptimized communication systems such as wireless local area network (WLAN) 802.11, Bluetooth, ZigBee, and WirelessHart do not. Especially ultra reliable low latency communications (URLLC) - as one of the main 5G pillars - is still considered to be the enabling technology for industrial real-time applications such as wireless closed-loop control [7], as it achieves extraordinary quality of service (QoS): latency lower than 1 ms and reliability larger than 99.999 %. URLLC is furthermore embedded in an ecosystem that also addresses many of the other industrial requirements. It is therefore not surprising that associations like the 5G Alliance for Connected Industries and Automation (5G ACIA) have formed in Germany with the target of ensur-

This work was in part funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden. This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101015956 - "Hexa-X". It was co-financed by public funding of the state of Saxony/Germany. We thank the Center for Information Services and High Performance Computing (ZIH) at TU Dresden for generous allocations of computer time. The authors alone are responsible for the content of the paper. Upon request, the authors are glad to share the MATLAB® code that was used to generate the results of this article.

L. Scheuvens, P. Schulz, and G. P. Fettweis are with the Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, Dresden, Germany, Email: {lucas.scheuvens, philipp.schulz2, gerhard.fettweis}@tu-dresden.de

Tom Hößler was with the Barkhausen Institut, Dresden, Germany. He is now with Xilinx Dresden GmbH, Dresden, Germany. E-Mail: tom.hoessler@xilinx.com

N. Franchi is with the Institute for Electrical Smart City Systems, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, Email: norman.franchi@fau.de

A. Barreto is with the Barkhausen Institut, Dresden, Germany, Email: andre.nollbarreto@barkhauseninstitut.org.

ing that industrial connectivity requirements find their way into standardization for future deployment. At the same time, the control engineering research community has made major progress investigating so-called networked control systems (NCSs), dealing with the question of how to cope with communications imperfections (unreliable, latency-prone, data-rate-limited, ...) and how to reduce their impact on control performance through developing robust control algorithms. These NCS research results stand in stark contrast to previously assumed QoS requirements for closed-loop control [8], which eventually sparked URLLC development. They demonstrate that while real-time applications profit from network capabilities beyond enhanced mobile broadband (eMBB), ultra-reliable communication (in terms of low packet loss rate (PLR)) is usually *not* required for ultra-reliable applications. In this sense, we believe that the interpretation of URLLC should be updated from providing ultra-reliable *connections* to providing ultra-reliable *applications* through exploiting the insights gained from fundamental NCS research. This co-design of control application and (wireless) communications (CoCoCo) reformulates communication requirements, which has the potential to drastically reduce operational expenditure (OpEx) for manufacturers.

Our work targets to exploit basic NCS research results towards quantifying the achievable reliability gains for real-time applications that are designed towards tolerating a few consecutive packet losses. The spirit of our CoCoCo approach is centered around the industrial requirements listed above, which is novel to the best of our knowledge. In detail, our contributions are:

- Based on a set of connectivity assumptions aligned with industrial needs, we derive a single-agent failure model that translates a sequence of packet losses into a failure in the control application domain by means of an absorbing Markov chain, incorporating age of information (AoI). The number of allocated resources in each state defines a so-called state-aware resource allocation (SARA) scheme.
- Building on the failure model, we derive closed-form expressions for the key performance metrics mean time to failure (MTTF), the average number of utilized channels $\bar{l}$, the status update age (SUA) probability mass function (PMF), and the packet loss sequence length (PLSL) PMF.
- We formulate the optimization problem to obtain the optimal resource allocation scheme through assigning a cost to low MTTF, high $\bar{l}$ and high SUA and minimizing the cost function, yielding optimal SARA.
- After considering only the single-agent case, we extend the failure model to a system failure model, which also includes a limited number of available channels. While the scheduler still attempts to assign the available resources according to an underlying SARA scheme, the number of channels may not suffice to satisfy all agents with the SARA resources that correspond to their individual single-agent state. We present two admission control schemes termed *random* and *cliff* to decide which agents will be underserved.

II. Related Work

The prior work, upon which we base the motivation and approach of this article, is classifiable into three groups: (a) URLLC, (b) NCS, and (c) value of information (VoI) / AoI.

A. URLLC

In the context of URLLC, 3rd Generation Partnership Project (3GPP) defines *reliability* as the complement of the PLR, i.e., $1 - \text{PLR}$, whereby a packet is considered lost when it is not correctly received within the time constraint required by the targeted service [9]. URLLC was primarily motivated as drop-in replacement of industrial wire-line communications solutions [10]–[12] such as Ethercat, SERCOS III, and Profinet IRT. This requires achieving a PLR lower than 10^{-5} while simultaneously featuring a deterministic latency lower than 1 ms [8] and one-way payload data rates exceeding 90 Mbit/s [13]. Consequently, great efforts were invested to achieve these QoS values [14]–[16]. The low latency constraint often entails that retransmissions such as automatic repeat request (ARQ) and hybrid automatic repeat request (HARQ) – the most-established measures in current wireless systems to ensure reliable data transmission – might not be feasible [17], [18] and reliability must instead be achieved through decreasing spectral efficiency (higher TX power, lower code rate, lower modulation order, multi-connectivity [19]), which might limit scalability.

B. NCS

NCS research is aimed at enabling closed-loop control over non-ideal communication channels through optimization of control algorithms in order to cope with these non-idealities. They include but are not limited to delay [20], low signal-to-noise ratio (SNR) [21], competition for resources [22], coarse quantization [23], and data rate limitations [24]. Surveys on results in this domain were conducted in [22], [25], [26]. While the above factors are also relevant aspects of communications networks that need to be considered when employing real-time wireless control, they will not be the focus of this article because they are less relevant for industrial 5G networks. Instead, we put our attention on the impact of packet losses on control applications, since they are an intrinsic property of wireless networks and can only be mitigated through increasing wireless resource consumption. A fact that is often neglected in URLLC-related literature is that control applications are – proper design implied – capable of tolerating a limited number of packet losses. This is particularly true for oversampled control applications, as is often the case to guarantee reasonable smoothness of operation [27]. Many works exist that attempt to quantify the packet loss tolerance for closed-loop control. E.g., the authors in [28] propose a controller design method for periodic sampling that includes packet dropouts (and network-induced latency) in the design process. The authors in [29] extend the analysis to model reference control. Both do not take the probability of packet losses into account and guarantee the existence of a stabilizing controller as long as a certain number of consecutive packet

losses K is not exceeded, which 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., [20], [30], [31], and derive upper bounds on the packet loss tolerance depending on the transmission success probability. Furthermore, the authors in [32] 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. Together with the (recent) specifications given in [5] and [33], this leads to the important foundation in this article that control applications can be designed to have an inherent tolerance against (a limited number of consecutive) packet losses.

C. VoI and AoI

A considerable number of prior publications exist that jointly consider the design of control applications and a communications network, most of which target the development of an optimal scheduler. When considering the best scheduling policy, it was found that, intuitively, agents that feature the largest state discrepancy should be scheduled first in order to maximize control utility, e.g., [34]–[40]. This approach relies on enough control awareness that the control application's *necessity* for a packet transmission can be estimated in the network and that it will subsequently be prioritized in case of a high control state discrepancy. The authors in [41] termed the potential impact of a packet's successful transmission on the reduction of the control state discrepancy as the VoI. The results of VoI-related research look highly promising, however, gaining access to lower-layer networking decisions (such as the scheduler) is a difficult endeavor because this network functionality is usually highly integrated. Thus, many authors have decided to estimate the VoI through the AoI instead. 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 VoI, the AoI can be monitored by the network alone and does therefore not require detailed knowledge (in terms of state-space representations) about the underlying control application. Through AoI, the value of a packet can be approximated and *control-unaware* network optimization can be performed. This is a compromise that keeps control and communications fields distinct, which has practical advantages in industrial environments, yet it provides a high-level network key performance indicator (KPI) that provides a significant benefit for real-time applications. The term was introduced by the authors in [42]. Since then, many more works have been published that differ in their assumptions regarding (a) the inclusion [43]–[45] vs. exclusion [34], [46], [47] of packet losses, (b) slotted [34], [43], [44] vs. carrier-sense multiple access (CSMA)-based [45]–[47] channel access, (c) single-user [46], [47] vs. multi-user [34], [43]–[45], (d) single-hop [43]–[47] vs. multi-hop [34] network, and (e) uniform [34], [43], [45]–[47] vs. non-uniform packet sizes [44]. All of these works assume event-based sampling at the sensor, meaning that sampling can occur anytime. Hence, time synchronicity, i.e., transmission directly after sampling through a shared understanding of time, is not assumed. Mostly, the main goal of these works is to minimize the *average* AoI present in the system through the design of optimal scheduling techniques for their respective system assumptions, generally involving queuing-theoretic analysis because of the assumption of event-based sampling. As sometimes the tail of the AoI distribution has a larger impact on application performance than the average, especially regarding the control application's reliability assessment, further works also attempted to minimize the probability of exceeding a certain AoI, which they term the peak age of information (PAoI) violation probability. The authors in [48] introduce the term PAoI and explore in a queuing-theoretic context. The authors in [49] extended this work to incorporate more accurate models of channel coding, taking advantage of results in finite-blocklength information theory for memoryless channels. The authors in [50] proceeded to drop the memoryless assumption, also taking into account the temporal correlation of fading channels.

While this article touches on many topics covered in prior work, it also provides multiple new aspects, which haven't been studied before. First, we apply a periodic sampling and time-synchronous operation because these are key requirements by Industry 4.0. This means that the sampling action is timed such that data transmission can be carried out directly thereafter and no random waiting times occur. We further ensure time synchronicity through a try-once-discard (TOD) approach, which avoids random-latency-prone ARQ or HARQ rounds. We assign time-synchronous dynamic multi-connectivity (MC) on a per-transmission basis within a pre-allocated resource pool and use semi-persistent scheduling (SPS) for predictable resource scheduling. These assumptions all target *industrial* requirements and will be presented in greater detail in Section III. While the VoI-related literature shows that the best trade-off between network usage versus control utility is achieved when the network is aware of the control application's dynamics and capable of predicting plant behavior, early-stage industrial networks will most probably not feature interfaces that may be programmed by control application engineers to make (optimal) scheduling decisions in the lower medium access control (MAC) network layer. The abstraction of using the AoI as a cross-layer metric to approximate information value is suboptimal, yet may prove to be an elegant compromise. AoI can be incorporated to network design more easily and will therefore be considered in this article. Contrary to many AoI-related publications, we do not attempt to minimize the average AoI but rather target to ensure a pre-specified PAoI violation probability that we express in terms of the MTTF of the control application. This analogy can be used because we assume that a control application outage occurs as soon as the PAoI is violated. Due to our connectivity assumptions, a queueing theoretic analysis is obsolete and the violation probability may be abstracted as "How likely is it to lose more than K consecutive packets?". Contrary to all prior publications, to the best of our knowledge, we are

the first to utilize dynamic MC as a tool to ensure a low PAoI violation probability while simultaneously attempting to achieve a resource consumption that is close to common bandwidth-efficient single-connectivity. Moreover, we note that manufacturing businesses demand application availability guarantees when deploying wireless control solutions as this relates strongly to OpEx. To the best of our knowledge, no works exist that provide (closed-form) availability guarantees with classical performance metrics from dependability theory, in particular for multiple coexisting agents.

III. System Model

We assume a set of M agents that need to be served over the network in the uplink (sensor $\rightarrow$ controller) and downlink (controller $\rightarrow$ actuator) direction. For this article, packet losses may only occur in the downlink direction, i.e., the uplink is considered ideal. Future work will also include erroneous uplink transmissions. All agents feature periodic sampling, which is common for industrial applications [5], leading to a constant packet inter-arrival time equal to the control application's sampling period T_s. Each agent's sensor sampling action is assumed to be synchronized with the time grid of the 5G base station (BS), fulfilling the time synchronicity requirement stated in Section I. Consequently, the waiting time between sampling and transmitting is negligible, contrary to many available publications that were mentioned in Section II. Hence, the latency between the sensor sampling instant and the plant receiving an actuator command – if the downlink transmission succeeds – can be approximated through adding 2 transmission time intervals (TTIs) (uplink+downlink) and the controller's computation time. This *fixed* latency will not be of further interest in this article because it can be parameterized to be very low for 5G networks [51]. Furthermore, we assume that all agents are served wireless resources (in terms of resource blocks (RBs) in the 5G time-frequency grid) through SPS [12], [52] from a pre-reserved resource pool, which is assumed to contain L_av channels. The term *channel* shall henceforth be used to describe one transmission opportunity within this resource pool. We assume that an agent is capable to connect to its networked controller through $l \in \{0, \ldots, \hat{l}\}$ parallel channels. Any set of channels that are assigned to an agent for a single transmission are assumed to have a frequency spacing larger than the coherence bandwidth and the packet interarrival time is assumed larger than the coherence time of the underlying small-scale fading process, which is easily achievable through, e.g., channel hopping [53]. Consequently, all transmissions can be regarded as *independent in both frequency and time*. This line of argument can be found in greater detail in our previous work [54]. For analytical tractability we assume that all channels feature a fixed per-channel packet loss probability p_loss. An extension for agent-specific p_loss values will be conducted in future work. To combine all l parallel data streams for one agent, a selection combining (SC) scheme is considered because of its low complexity, which allows the combination of channels in higher network layers, e.g., through cyclic redundancy checks (CRCs). More sophisticated combining schemes that require combination on low network layers, such as maximum ratio combining (MRC), will yield even better results than shown in this article. Additionally, ideal acknowledgments (ACKs) are assumed because signaling information (such as ACKs) 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 ACKs is expected to be minimal. The chosen communication architecture allows for (a) minimal communication overhead due to SPS (as only dynamic resource assignments need to be communicated *on demand*) and (b) for predictability and time synchronicity of data flow, which is essential for Industry 4.0. All mentioned assumptions are in accordance with current 5G standardization [51].

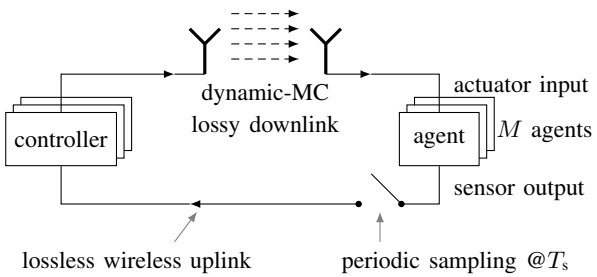


Figure 1: System Model

IV. Single-Agent Analysis

A single agent will be considered in this section. We structure this section as follows: (a) The single-agent failure model will be derived, (b) the performance metrics will be introduced, (c) the optimization problem will be formulated, leading to SARA, and, (d) the (closed-form) results will be presented and discussed. While the goal is to minimize the average number of required channels, we do not assume resource scarcity in this section. The case of a limited number of channels will be studied in Section V when the failure model is mapped to the multi-agent case.

A. Failure Model

The related work has shown that packet losses, i.e., failures in the communications domain, do *not necessarily* lead to a failure in the control application domain. While there are control applications that truly require ultra-high packet transmission success, many tolerate packet losses to some extent as long as not *too many* occur consecutively. The authors in [5], [9] refer to the term "survival time" to describe the time period that an application may prolong operation without receiving an expected packet. PAoI and survival time relate to each other according to

$$\text{survival time} + T_\mathrm{s} = \text{PAoI} \ , \tag{1}$$

where T_s is the control application's sampling period. A control application, for which PAoI $> (K+1)T_\mathrm{s}$ with $K \in \mathbb{N}_0$ holds true, is able to tolerate K consecutive packet losses, which can subsequently be exploited on a network level. This constitutes a mapping of (communications-related) packet

This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953

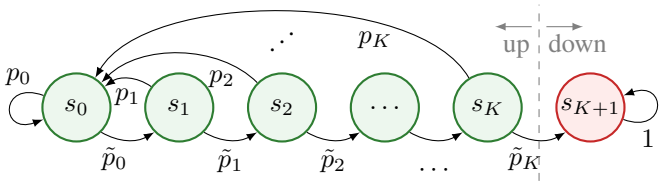


Figure 2: Single-agent failure model.

losses and (control-application-related) failures and is the first step towards a joint optimization yielding *application-related* performance metrics, which are essential for industrial manufacturing. For clarity, we will henceforth refer to this mapping as *CoCo-mapping*. In this model it is assumed that if $K+1$ packets are lost consecutively, the control application will be shut down to avoid damage or, more severely, human harm. The specific value of K varies depending on the control application under investigation, the required control performance and also the chosen sampling period. In our previous work [54], we determined $K = 3$ for an example automated guided vehicle (AGV) use case.

Based on the connectivity assumptions of Section III and the CoCo-mapping, we propose a Markov chain [55] failure model, which is depicted in Fig. 2. 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 other CoCoCo-related publications that commonly refer to control states (or estimations thereof) of a state-space description. In the presented model, an agent is considered in state s_k with $k \in \{0, \ldots, K+1\}$, 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, the state s_0 is entered. Whenever a packet is lost with transition probability $\tilde{p}_k$, the state index is incremented by one. The rightmost state s_{K+1} of Fig. 2 denotes the only failure state because it refers to $K+1$ consecutive packet losses. All other states are considered "up". The failure state s_{K+1} is *absorbing* since it is unclear how to proceed with operation after $K+1$ consecutive packet losses have occurred. Note that with the stated connectivity assumptions (periodic transmissions, negligible latency, lossy downlink, ideal uplink), the number of consecutive packet losses linearly translates into the AoI, and consequently losing $K+1$ consecutive packet translates into the AoI exceeding the PAoI.

B. 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 [56]. This terminology is especially useful for processes that converge to a meaningful steady-state. Such processes 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 ultra-reliable communications (URC) research [57], 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 model presented in Fig. 2 will evidently not lead to meaningful steady-state results. Processes of this kind are better described through time-based metrics, such as the mean time to failure. Hence, in the following we will introduce the communications-related availability and the application-related MTTF. Building on this, we will also derive closed-form solutions for the state probabilities $\boldsymbol{\pi}$ and the average number of assigned channels $\bar{l}$. In addition, we introduce two metrics which characterize the AoI.

1) Availability: With the connectivity assumptions stated in Sec. III, the availability is given as

$$p_k = 1 - p_{\text{loss}}^{l_k} \ , \tag{2}$$

where l_k characterizes the number of parallel transmission channels assigned in state s_k and, consequently, p_k describes the success probability of that transmission. The failure probability is therefore $\tilde{p}_k = 1 - p_k$.

2) Mean Time to Failure: Through the CoCo-mapping of the failure model, the application-related MTTF can be derived for a single agent as

$$\text{MTTF} = T_\text{s} \boldsymbol{e}_0 \boldsymbol{N} \boldsymbol{1}^\intercal \ , \tag{3}$$

when initialized in s_0, with $\boldsymbol{e}_0 = [1\ 0\ \ldots\ 0]$, $\boldsymbol{1} = [1\ \ldots\ 1]$ and $\boldsymbol{N} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}$ as the fundamental matrix of the absorbing Markov chain [58]. Therein, $\boldsymbol{Q}$ 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 $\boldsymbol{P}$ of the Markov chain in Fig. 2 via (all zeros are omitted for readability)

$$\boldsymbol{P} = \left[\begin{array}{ccc|c} p_0 & \tilde{p}_0 & & \\ p_1 & & \tilde{p}_1 & \\ \vdots & & & \ddots \\ p_K & & & \tilde{p}_K \\ \hline & & & 1 \end{array}\right] = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{R} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} . \tag{4}$$

3) Packet Loss Ratio: For comparison, we also introduce the traditional, communications-related PLR after combining, thus different from p_loss. With the probability vector

$$\boldsymbol{\pi} = \frac{\boldsymbol{e}_0 \boldsymbol{N}}{\boldsymbol{e}_0 \boldsymbol{N} \boldsymbol{1}^\intercal} \tag{5}$$

of being in each transient state during operation, it can be derived via

$$\text{PLR} \approx 1 - \pi_0 \ , \tag{6}$$

where π_0 is the probability of being in state s_0, hence, the first element of $\boldsymbol{\pi}$. The intuition is that whenever a state different from s_0 is entered, a packet transmission failure must have occurred. The approximation stems from initializing the Markov chain in s_0.

4) Average Number of Assigned Channels: With $\boldsymbol{l}$ as a cost vector that gathers the number of channels l_k assigned in each state s_k, the average number of parallel channels $\bar{l}$

used throughout operation can be calculated as

$$\bar{l} = \boldsymbol{l}\boldsymbol{\pi}^{\mathsf{T}} \ . \tag{7}$$

5) Age of Information: As introduced before, the number of consecutive packet losses may be interpreted as the AoI for the stated connectivity assumptions. Two AoI-related metrics are introduced as KPIs: The first is the status update age (SUA) τ, whose PMF describes the age distribution of the most recent successfully received packet. Formally, τ is defined as

$$\tau = t - t' \ , \tag{8}$$

where t represents the current absolute time and t' stands for the time of the last successful packet arrival at the receiver. Originally, τ was defined as a time-continuous random variable [59], however, as our connectivity assumptions state slotted operation, we evaluate τ only at the sampling instants and thereby transform τ into a discrete-time quantity. The PMF of τ is equal to the state probabilities $\boldsymbol{\pi}$, i.e.,

$$f_\tau(k) = \pi_k, \quad k \in \{0, ..., K\} \ , \tag{9}$$

since the Markov model was specifically designed to incorporate the SUA information. The second is the packet loss sequence length (PLSL) θ whose PMF describes the number of unsuccessful transmissions between two successful transmissions. τ evaluates the AoI from a control application standpoint while θ views it from a communications perspective. The PMF of θ can be derived directly from the Markov model presented in Fig. 2 by calculating the loop probabilities of all loops originating in s_0.

$$f_\theta(k) = \begin{cases} p_k \prod_{\kappa=0}^{k-1} \tilde{p}_\kappa, & k \in \{0, ..., K\}, \\ \prod_{\kappa=0}^{K} \tilde{p}_\kappa, & k = K+1. \end{cases} \tag{10}$$

In the next section, it will be shown that by changing the number of parallel channels l_k assigned in each state s_k of the individual agent's Markov model, see (2), the transition probabilities can be tuned such that a high MTTF can be achieved, while at the same time the average number of assigned channels $\bar{l}$ is kept low.

C. State-Aware Resource Allocation

The fundamental idea behind our presented SARA approach is to dynamically alter the number of parallel channels depending on the agent's SUA. This can be understood as multi-objective optimization, constrained to an integer number of links (in this article). Formally:

$$\begin{aligned} \boldsymbol{l}_{\text{opt}} = \arg\min_{\boldsymbol{l}} \left\{ g_{\text{MTTF}}(\text{MTTF}(\boldsymbol{l})), g_\tau(f_{\tau(\boldsymbol{l})}), g_{\bar{l}}(\bar{l}(\boldsymbol{l})) \right\} \\ \text{subject to } l_k \in \{0, \dots, \hat{l}\} \end{aligned} \tag{11}$$

All functions $g_{(\cdot)}$ constitute penalty functions, which must be carefully designed by control application engineers, as MTTF and $\bar{l}$ directly influence OpEx, while f_τ influences control performance. Due to the heterogeneity of control applications, there is not one single optimal SARA solution and the choice of all functions $g_{(\cdot)}$ 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 OpEx and control performance. However, in general it was observed that SARA schemes featuring a low number of channels in the first states and a large number of channels in high states (increasingly many parallel channels for increasing SUA) exhibit a good MTTF $\leftrightarrow$ f_τ $\leftrightarrow$ $\bar{l}$ trade-off, making it decreasingly likely for packet losses to occur consecutively the longer the current number of consecutive packet losses is. This enables to significantly increase the MTTF while keeping $\bar{l}$ low [60].

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 article because each parallel demodulation at the receiver requires a costly physical layer (PHY) signal processing chain in hardware. Also, choosing a lower $\hat{l}$ facilitates achieving uncorrelated packet transmissions, as was assumed in Section III.

Since the optimization problem in (11) highly depends on the chosen penalty functions $g_{(\cdot)}$, we present a set of sensible solutions in this article, which stem from different quadratic penalty functions g_{MTTF}, g_τ, and $g_{\bar{l}}$. We apply the following notation: Adaptation schemes that follow a regular pattern are denoted as S_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. $S^0_{l_{\text{fix}}}$ 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 article are summarized in Tab. I. Please note that also two schemes are considered that do not follow a regular pattern. We term the first one S_{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. We term the second one $\text{S}_{1\text{f}}$, with the subscript standing for one channel allocated after a successful transmission, but making sure that the maximum number of parallel channels are allocated when $k = K$, i.e., $l_K = \hat{l}$. Then, with decreasing k, the number of channels is always decreased by one; however for $k = 0$, always one channel will be assigned. Please also note that as k increases, the number of assigned channels for S_1^1 and S_2^1 is clipped to $\hat{l} = 4$ at $k = 4$ and $k = 3$, respectively. In Tab. I, when $K < 4$, the schemes are accordingly truncated, e.g., when $K = 2$ for S_1^1, the link assignment will be $1 \to 2 \to 3$. In order 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 corresponds to the determined PAoI of the control application in [54]. While this ensures a fair performance comparison in terms of dependability, we emphasize that the control application 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) [27]. In digital control, the guideline for good stability and smoothness is to oversample

TABLE I: STATE-AWARE RESOURCE ALLOCATION – EXAMPLE SCHEMES.

		consec. lost packets k					K	class
		0	1	2	3	4		
number of channels l_k	S_1^0	1	1	1	1	1	{0,...,4}	static
	S_2^0	2	2	2	2	2		
	S_3^0	3	3	3	3	3		
	S_1^1	1	2	3	4	4	{1,...,4}	dynamic
	S_2^1	2	3	4	4	4		
	S_{rf}	1	1	4	4	4	{2,...,4}	
	S_{1f}	1	4				1	
		1	3	4			2	
		[corresponds to S_1^1]					3	
		1	1	2	3	4	4	

TABLE II: SAMPLING PERIOD VS. PACKET LOSS TOLERANCE INTERDEPENDENCY. [54]

K	0	1	2	3	4
T_s	180 ms	90 ms	60 ms	45 ms	36 ms

by 10-20 times, see [27] for details. While within this article we limit ourselves to the schemes presented in Tab. I, other schemes may be optimal for different solutions of (11).

D. Results

The dynamic resource allocation of SARA can improve the control application's MTTF by orders of magnitude compared with the static multi-connectivity baseline, while simultaneously keeping the average channel consumption low, as we have previously shown in [60]. Tab. III shows the analytical results of the introduced performance metrics PLR, MTTF, and $\bar{l}$ for $K = 3$ tolerable consecutive packet losses at an exemplary per-channel packet loss probability $p_{\text{loss}} = 10\,\%$ and a sampling period $T_s = 45\,\text{ms}$. We observe that the MTTF improves by 100x from 8 weeks to 16 years for the SARA scheme S_1^1 compared to static dual-connectivity S_2^0, while only using approximately half the amount of parallel channels on average (1.09 compared to 2). For S_1^1, the resulting PLR remains over $9\,\%$. This clearly shows that when the network is capable to adjust its operation to better serve the needs of closed-loop control applications, even high resulting PLR values do not necessarily lead to an AoI-induced control application outage. While the PLR for S_2^0 is approximately one order of magnitude lower than for S_1^1, the MTTF is 100x greater. 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 application'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 through reacting to packet losses and dynamically increasing the number of channels used in the subsequent transmission. While this does not reduce the PLR significantly, it reduces the probability of long packet loss sequences by orders of magnitude.

TABLE III: RESULTING PERFORMANCE METRICS FOR $K = 3$ AND $p_{\text{LOSS}} = 10\,\%$. FOR MTTF, THE SAMPLING PERIOD IS $T_s = 45\,\text{ms}$.

scheme	average packet loss rate	mean time to failure	average cost
	PLR	MTTF	$\bar{c}$
S_1^0	100.0×10^{-3}	8 minutes	1.00
S_2^0	10.0×10^{-3}	8 weeks	2.00
S_3^0	1.0×10^{-3}	1.4k years	3.00
S_1^1/S_{1f}	91.7×10^{-3}	16 years	1.09
S_2^1	9.9×10^{-3}	14 millennia	2.01
S_{rf}	99.1×10^{-3}	16 years	1.03

This potential is also demonstrated in Fig. 3, 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, S_2^1 requires only $p_{\text{loss}} = 10\,\%$ ($\bar{l} = 2.01$) compared to static MC scheme S_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. For low p_{loss} values, the shown curves diverge, which shows that SARA gains are enhanced. For example, when comparing S_1^0 ($\bar{l} = 1$) and S_{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 $3\,\%$.

The PMFs of τ and θ of all resource allocation schemes are depicted in Fig. 4 for $K = 3$. To display all information in one diagram, the data are presented columnwise, i.e., each τ ($-$) and θ set ($+$) sums up to one. Please note that while θ also includes a value for $k = K + 1$, the SUA is only defined until $k = K$. We observe that first, the difference of τ and θ 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 $\tau(k)/\tau(0)$ and $\theta(k)/\theta(0)$ for $k \neq 0$ can be explained approximately through the number of transmission opportunities. For example for S_1^1, $\tau(3)/\tau(0) \approx 10^{-6}$ because in order 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 τ and θ will be achieved. For instance when comparing S_1^1 and S_{rf}, we observe that packet loss sequences of length $k = 2$ are obtained in 1/100 for S_{rf} while they only occur in 1/1000 (ten times less frequent) for S_1^1 due to assigning two channels in s_1 instead of one.

This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953

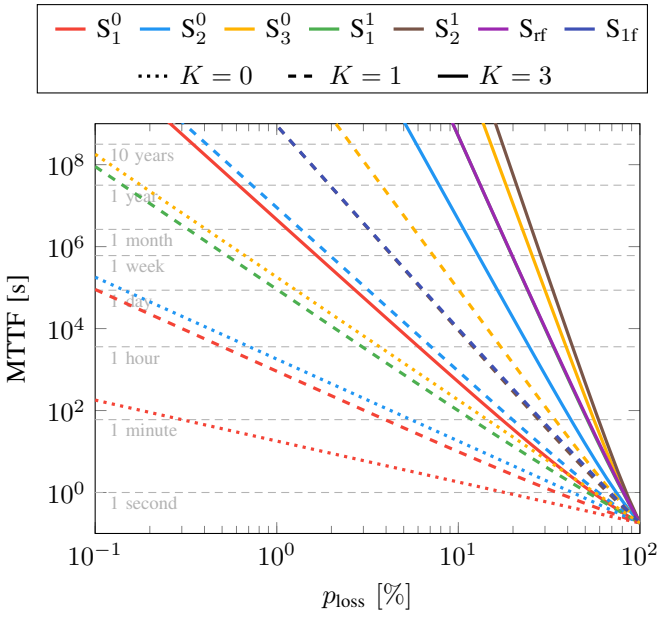


Figure 3: Performance of different SARA schemes regarding p_{loss} and MTTF.

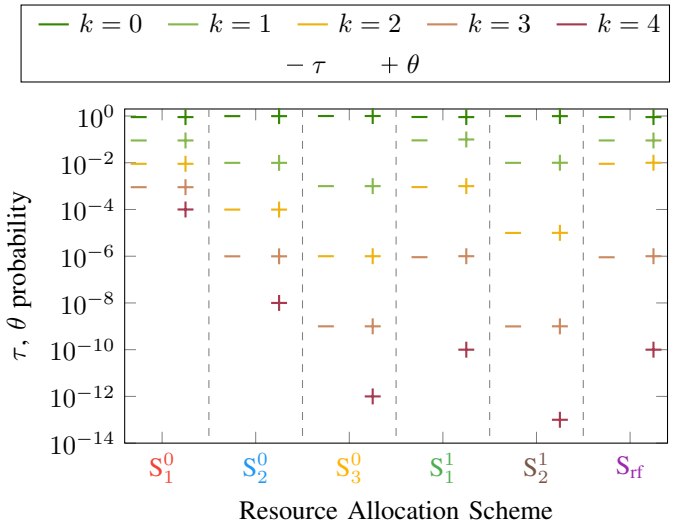


Figure 4: AoI metrics τ and θ for all resource allocation schemes with a single agent.

V. Multi-Agent Analysis

In the last section, the application's dependability benefits of a single agent were demonstrated. Naturally, the question arises how SARA performs when applied on system level where (a) the number of agents is greater than one and (b) the number of available channels is limited.

A. Failure Model

We introduce a system of M homogeneous agents of which each individually operates according to the Markov chain depicted in Fig. 2. This means that each agent is able to tolerate K consecutive packet losses before ultimately failing, i.e., reaching the absorbing state s_{K+1}. We introduce a superimposed Markov chain that agglomerates all individual agent states to a single system state

$$S_{|s_0|,|s_1|,\ldots,|s_{K+1}|} \,. \tag{12}$$

Thereby, $|s_k|$ denotes the amount of agents that currently reside in state s_k, consequently, $\sum_{k=0}^{K+1} |s_k| = M$. 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.

We define that the system is *down* as soon as the first agent fails, i.e., $|s_{K+1}| > 0$, else it is *up*. All *down* states are collapsed to a single down state S_{d} and the last index in (12) corresponding to $|s_{K+1}|$ is dropped from the notation as it will take the value 0 for all *up* states. For more concise notation, we introduce a linear index $i \in \{0, \ldots, Z_{\text{up}} - 1\}$ for all *up* system states, with Z_{up} describing 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.

1) Admission Control: We introduce a number of L_{av} available channels to the system. 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 $L_{\text{av}} < L_{\text{req}}(S_i)$, with $L_{\text{req}}(S_i)$ as the total number of requested[1] channels (by all agents) in state S_i. Hence, $L_{\text{req}}(S_i) - L_{\text{av}}$ channels need to be "denied" by admission control, constituting a system-induced deviation from the presented SARA resource allocation in Tab. I. In this article, two admission control schemes are considered, which extends our work in [61].

random From the set of all requested channels, L_{av} are randomly selected and assigned; the rest will be denied. We stress that whenever an agent was denied a channel, it is still susceptible to be denied another 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 from $s_k \to s_{k+1}$ 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".

2) Transition Probabilities: To derive the transition probabilities of the system Markov chain, we revisit the individual agent's Markov chain in Fig. 2. Therein, the state transition probabilities p_k were determined by the chosen resource allocation scheme summarized in Tab. I. 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}.

[1]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. I. This should not be confused with a classical transmission request of cellular networks.

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 (12).
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 (2) 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.

B. Performance Metrics

Analogously to the individual agent's MTTF, we introduce the MTTF_{sys} on the system level. Please recall that the system is defined as *down* as soon as at least one agent fails, i.e., reaches its individual s_{K+1} 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 [62]. It can be shown that the minimum of M independent and identically distributed exponential random variables TTF_m with $m \in \{0, \dots, M-1\}$, which refers to the earliest failing agent, reduces the expectation by a factor M, i.e., $\mathbb{E}[\min_m \text{TTF}_m] = \frac{1}{M}\mathbb{E}[\text{TTF}_m]$. Hence, the maximum system-wide MTTF_{sys}, $\text{MTTF}_{\text{sys,max}}$, which is achieved without any limitation of channels, approximately reduces by a factor M compared to the single-agent MTTF, i.e.,

$$M \approx \frac{\text{MTTF}}{\text{MTTF}_{\text{sys,max}}} \leq \frac{\text{MTTF}}{\text{MTTF}_{\text{sys}}} \ . \tag{13}$$

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

The derivation of MTTF_{sys} is a known procedure for absorbing Markov chains and is performed analogously to (3)–(4). Hence, when initializing in S_0 (all agents in s_0),

$$\text{MTTF}_{\text{sys}} = T_{\text{s}} \boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \boldsymbol{1}^{\intercal}, \tag{14}$$

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

$$\boldsymbol{\pi}_{\text{sys}} = \frac{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}}}{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \boldsymbol{1}^{\intercal}} \tag{15}$$

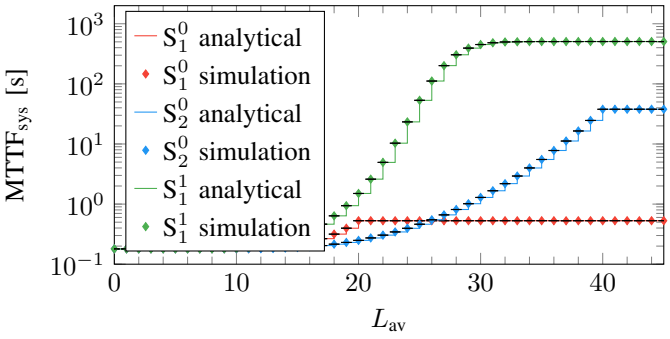


Figure 5: Verification of closed-form MTTF_{sys} expression through extensive system-level simulation.

comprises the probabilities of being in each transient state during operation, i.e., the system version of (5). Additionally, we introduce $\boldsymbol{1}(L_{\text{av}})$ denoting 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

$$\boldsymbol{1}(L_{\text{av}})_i = \begin{cases} 1 \text{ if } L_{\text{av}} \leq L_{\text{req}}(S_i) \\ 0 \text{ else} \end{cases} \ . \tag{16}$$

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

$$\eta(L_{\text{av}}) = \frac{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \boldsymbol{1}(L_{\text{av}})^{\intercal}}{\boldsymbol{e}_0 \boldsymbol{N}_{\text{sys}} \boldsymbol{1}^{\intercal}} \tag{17}$$

C. Results

In this section, we will first verify the presented closed-form model with system-level simulations. We then proceed to evaluate the derived performance metrics MTTF_{sys} and η for all resource allocation schemes of this article, 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 [63], 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.

1) Verification through System-Level Simulation: Fig. 5 shows the results of extensive system-level simulation for the schemes S_1^0, S_2^0, and S_1^1 at $M = 20$ for $p_{\text{loss}} = 30\,\%$. Each simulation was run until at least one agent failed and the number of runs was 10^6 for each data point. The simulated MTTF_{sys} values are shown as colored markers and for comparison, the closed-form MTTF_{sys} from (14) 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.

2) Applicability on the System Level: From now on, only closed-form results will be presented. Fig. 6 compares MTTF_{sys} vs. L_{av} for $M = 20$ and $K = 3$. Please note

This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953

the logarithmic ordinate axis and the human-readable time units. We stress 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.

As expected, each resource allocation scheme features an $\mathrm{MTTF_{sys}} = 180\,\mathrm{ms} = (K+1)T_{\mathrm{s}}$ for $L_{\mathrm{av}} \to 0$ 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. 2), 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 (13). Please note that the graphs for S_1^1 and $\mathrm{S_{rf}}$ overlap to a significant extent.

The SARA schemes S_1^1, S_2^1, and $\mathrm{S_{rf}}$ show clear benefits over static MC, i.e., S_1^0, S_2^0, and S_3^0. Comparing S_1^1 and S_2^0 (static dual-connectivity) for example, multiple advantages stand out:

1) S_1^1 features a 100x improvement in terms of $\mathrm{MTTF_{sys,max}}$ over S_2^0.
2) SARA can also be applied to a multi-user system with limited resources in an efficient way. Reaching the identical $\mathrm{MTTF_{sys}}$ requires fewer available channels L_{av} for S_1^1. S_2^0 reaches its $\mathrm{MTTF_{sys,max}} \approx 3$ days at $L_{\mathrm{av}} = 40$ (for both admission control schemes), while the same $\mathrm{MTTF_{sys}}$ is reached for S_1^1 at only $L_{\mathrm{av}} = 23$ (random) or even $L_{\mathrm{av}} = 18$ (cliff).
3) Comparing the respective $\mathrm{MTTF_{sys,max}}$, S_1^1 also proves to scale better. $99\,\%$ of the maximum value ($0.99 \times \mathrm{MTTF_{sys,max}} \approx 9$ months) is reached at $L_{\mathrm{av}} = 29$ (random) and $L_{\mathrm{av}} = 25$ (cliff), respectively, for S_1^1 and therefore earlier than for S_2^0 ($0.99 \times \mathrm{MTTF_{sys,max}} \approx 3$ days) at $L_{\mathrm{av}} = 40$. That is, the 100x $\mathrm{MTTF_{sys}}$ 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 ($L_{\mathrm{av}} < 16$). 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 $L_{\mathrm{av}} = M$ channels. Overall, an analysis as the one shown in Fig. 6 provides a valuable tool for a reasonable choice of the best resource allocation scheme with respect to the $\mathrm{MTTF_{sys,max}}$ depending on the situation at hand.

3) Comparison of Admission Control Schemes: Comparing the random and cliff admission control schemes in Fig. 6, 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 $\mathrm{MTTF_{sys}}$ saturates at 20 (S_1^0), 40 (S_2^0), and 60 (S_3^0) available

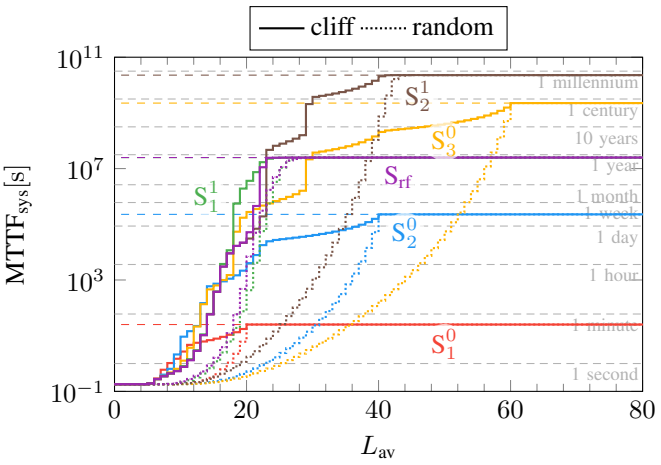


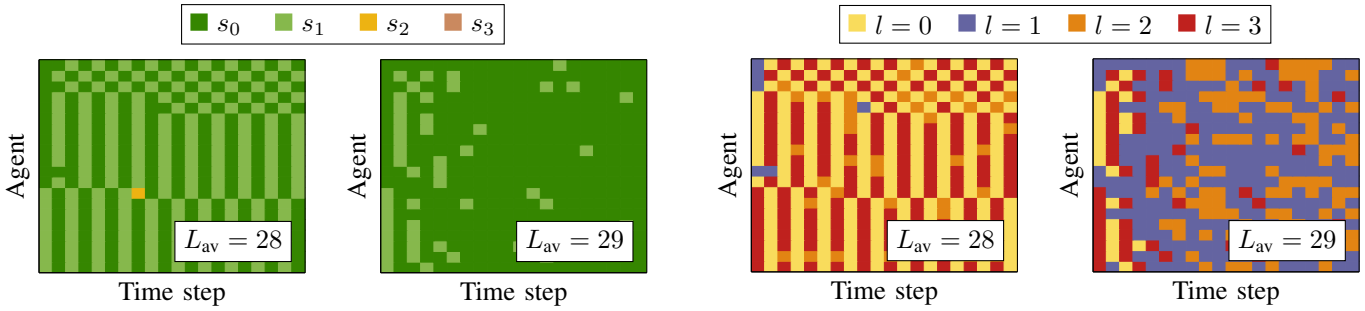
Figure 6: Dependency of $\mathrm{MTTF_{sys}}$ on L_{av} for $M = 20$ and $K = 3$ at $p_{\mathrm{loss}} = 10\,\%$.

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 $\mathrm{MTTF_{sys,max}}$.

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 $\mathrm{MTTF_{sys}}$ curves with cliff admission control feature an earlier, steeper and coarser development compared to the random scheme. Fig. 6 shows clearly that in terms of $\mathrm{MTTF_{sys}}$ 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 S_3^0, whereas in other ranges, adding another channel does not provide much gain, e.g., the 44th channel for S_3^0. The locations which feature a steep $\mathrm{MTTF_{sys}}$ 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. We will explain this by means of an example, however, a similar explanation can be found for all major steps in Fig. 6. Consider the S_3^0 scheme at $L_{\mathrm{av}} = 28$ and $L_{\mathrm{av}} = 29$ that differ by one order of magnitude in terms of $\mathrm{MTTF_{sys}}$. Fig. 7 shows a section of a simulation conducted for both cases. In Fig. 7a, the individual agent states are displayed and in Fig. 7b, the

This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953



(a) Agent states.

(b) Number of channels l assigned to each agent.

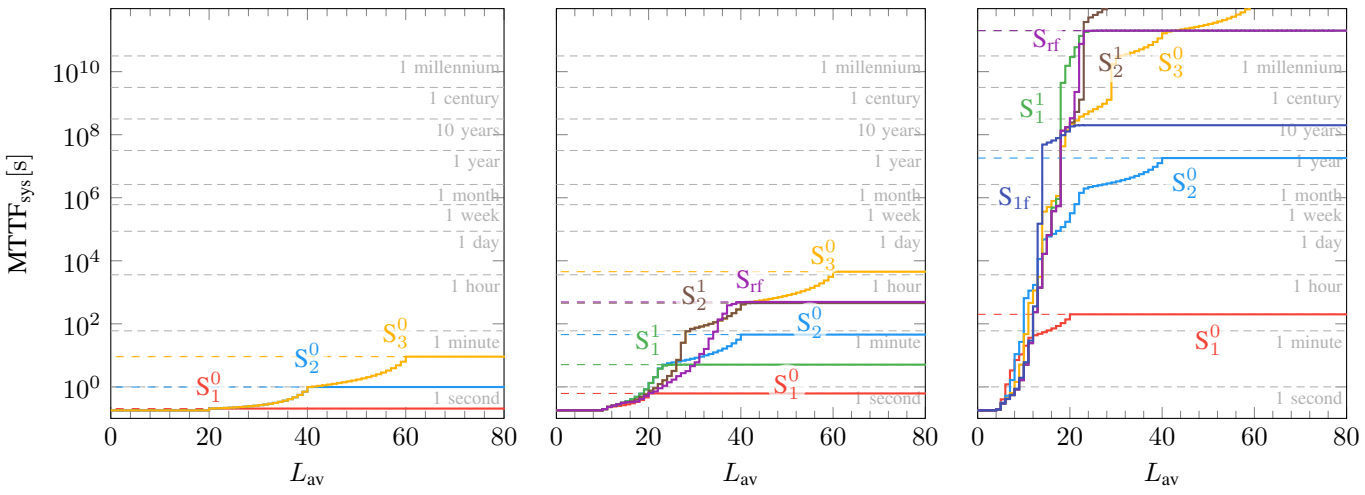
Figure 7: Sample system simulation for S_3^0 with $M = 20$, highlighting the state oscillations in $L_{\text{av}} = 28$ (left, respectively) that can be terminated by adding one more channel, $L_{\text{av}} = 29$ (right, respectively)

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}.

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. 8 shows the MTTF_{sys} for $K \in \{0, 1, 4\}$ (the case $K = 3$ was already evaluated in Fig. 6), 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. 8a, triple-connectivity (S_3^0) will only yield $\text{MTTF}_{\text{sys}} < 1\,\text{minute}$. Since every added static channel increases the MTTF_{sys} by approximately one order of magnitude given the assumptions of Section III, 10 channels will be required *per agent* to ensure a $\text{MTTF}_{\text{sys}} \approx 1\,\text{year}$. 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. 8b, three static parallel channels (S_3^0) lead to $\text{MTTF}_{\text{sys}} \approx 1\,\text{hour}$. 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 $\text{MTTF}_{\text{sys}} \approx 1\,\text{year}$. Also, $K = 1$ enables to apply a one-stage SARA, already highlighting SARA's potential by increasing the MTTF_{sys} by roughly $K = 1$ order of magnitude (compare $\mathrm{S}_1^0 \leftrightarrow \mathrm{S}_1^1$ and $\mathrm{S}_2^0 \leftrightarrow \mathrm{S}_2^1$), only with marginally increased resource consumption in terms of L_{av}. In Fig. 8c, 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 $L_{\text{av}} = 24$ channels in total. Keeping in mind that the design guideline for digital control applications is to oversample in the range 10 to 20 in order to achieve a reasonable smoothness (corresponding to 20 to 40 times the closed-loop system bandwidth) [27], a value of $K = 4$ is reasonable and often available *for free* if isolated smoothness interruptions are acceptable (which depends on the control application under consideration).

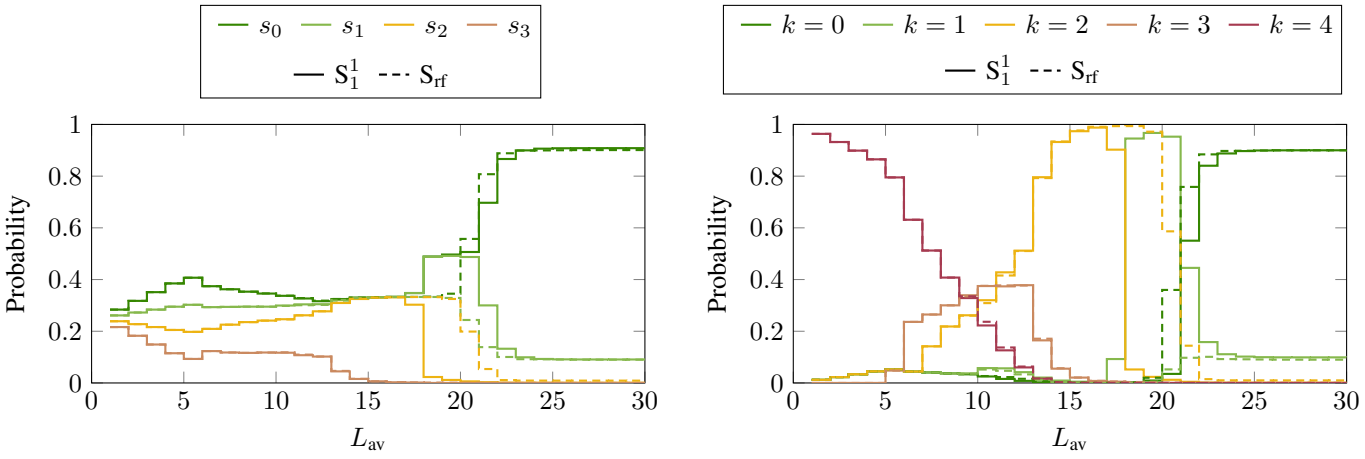
5) Age of Information: The SUA and PLSL can be evaluated not only for the single-agent case, see Section IV-D, but also on system level; τ_{sys} (SUA) and θ_{sys} (PLSL) denote the system-related quantities. From a network perspective, θ_{sys} provides particularly valuable insight regarding the frequency of packet loss sequence lengths, including system effects such as the state oscillations that were highlighted in Section V-C3. Fig. 9 shows the PMF dependency of τ_{sys} and θ_{sys} on L_{av}, respectively, for the resource allocation schemes S_1^1 and S_{rf} when employed on 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 $L_{\text{av}} \to 30$, the same values as in Section IV-D 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 $L_{\text{av}} = 16$ for both S_1^1 and S_{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. τ_{sys} shows an equal probability 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 θ_{sys}, which shows that s_0 is almost

This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953



(a) $K = 0$ (b) $K = 1$ (c) $K = 4$

Figure 8: Dependency of MTTF$_{sys}$ on the packet loss tolerance K.



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

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

Figure 9: AoI PMF development over L_{av} for S_1^1 and S_{rf} resource allocation schemes.

always entered coming from s_2 (the yellow graph is close to 1). This state oscillation is broken up at $L_{av} = 18$ for S_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 S_{rf}, the state oscillation incorporating the states s_0, s_1, and s_2 is broken up later compared to S_1^1 at $L_{av} = 20$, however, the state oscillation involving only s_0 and s_1 is avoided.

We stress 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 S_1^1 and S_{rf}, $L_{av} \geq 23$ should be chosen.

6) Channel Saturation Ratio: The sum of requested channels in each system state S_i was introduced earlier as $L_{req}(S_i)$. 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. 10 for $M = 20$ and $K = 3$. Obviously, the 21st (for S_1^0), 41st (for S_2^0), and 61st (for S_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 S_1^1) and the 41st channel (for S_2^1), respectively, not all system states request all available channels, i.e., there is at least one S_i with $L_{req}(S_i) < L_{av}$. 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, we introduce an overprovisioning scheme that, in addition to all requested channels, also assigns all spare channels.

7) Enforcing Full Channel Saturation: To exploit $\eta = 1$ for all resource allocation schemes throughout the entire operation time, we introduce an overprovisioning scheme that distributes spare channels. For conciseness, we limit ourselves to an overprovisioning strategy that prioritizes agents in higher indi-

This document is a preprint of: L. Scheuvens, T. Hossler, P. Schulz, N. Franchi, A. Barreto and G. Fettweis, "State-Aware Resource Allocation for Wireless Closed-Loop Control Systems," in IEEE Transactions on Communications, Jul 2021. DOI:10.1109/TCOMM.2021.3095953

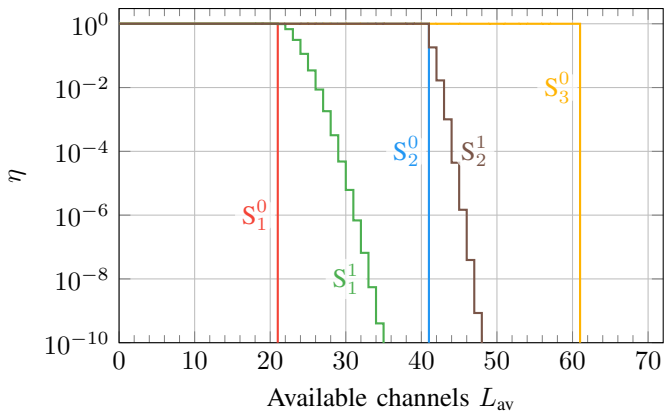


Figure 10: Channel saturation ratio η for $M = 20$.

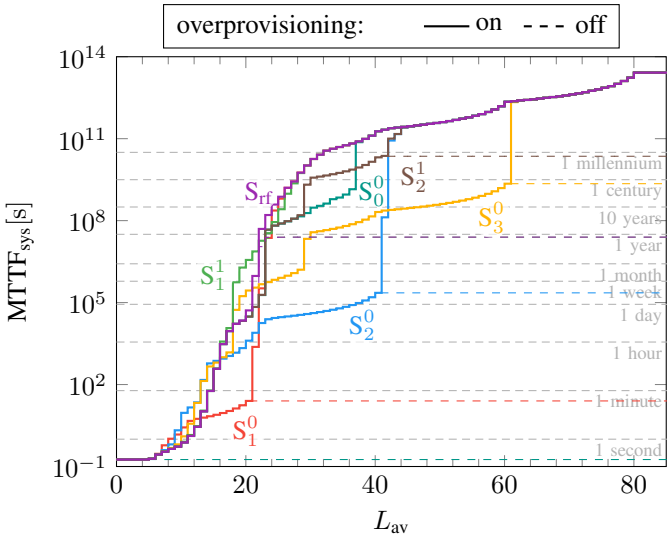


Figure 11: MTTF_{sys} development while ensuring full channel saturation ($\eta = 1$) through overprovisioning, i.e., distributing spare channels while prioritizing agents in higher individual states.

vidual states. Similar to the cliff admission control scheme, our 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. 11 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 S_1^0 scheme is roughly two orders of magnitude per additional channel, clearly highlighting the added benefit of dynamic MC. When compared to S_2^0, S_1^0 with overprovisioning surpasses the performance of S_2^0 at only 2 added channels, i.e., $L_{\text{av}} = 22$. This extreme increase can also be observed for S_2^0 (at $L_{\text{av}} = 41$) and S_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, we introduce the S_0^0 scheme, corresponding 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, S_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, S_1^1, S_{rf}, S_2^1, and S_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.

VI. Conclusion

Building on the knowledge that most control applications are able to tolerate a certain number of consecutive packet losses as long as a given maximum bound on the AoI is not exceeded, this article introduces the dynamic multi-connectivity concept SARA that is able to adaptively change the number of parallel channels that are used for transmission depending on the number of packets that were lost previously. This enables to increase the mean time to failure by orders of magnitude while keeping the average resource consumption in the range of single-connectivity. This article furthermore develops 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. It demonstrates that using closed-loop control applications in conjunction with networks that use SARA is highly feasible in terms of the amount of channels required, also in a multi-user system with competition for resources, because of a significant statistical multiplexing gain. The optimal resource allocation can be determined through solving a given optimization problem. As an example scheme, dynamically adding one link per lost packet (S_1^1 scheme) has proven extensive reliability gains for control applications that

are able to tolerate three consecutive packet losses. With 20 agents in the system, each agent only requires 1.2 channels on average to reach $99\,\%$ of the maximum $\mathrm{MTTF}_{\mathrm{sys}}$, yielding a system-wide mean time to failure of approximately 1 year until the first agent fails. S_1^1 outperforms static single-connectivity by 6 orders of magnitude in terms of MTTF (at $\sim$1.2x average channel usage) and static dual-connectivity by 2 orders of magnitude (at $\sim$0.6x average channel usage). The AoI within a SARA system was studied by means of the status update age and the packet loss sequence length, enabling to employ a SARA scheme that is able to statistically guarantee a certain PMF, also on system level. These AoI metrics enable to identify state oscillations that effectively reduce the effective sampling rate and thereby systematically harm control performance and also decrease the $\mathrm{MTTF}_{\mathrm{sys}}$ by orders of magnitude. Furthermore it was shown for a large number of available channels, that assigning unrequested/remaining channels for transmission further drastically increases the $\mathrm{MTTF}_{\mathrm{sys}}$.

REFERENCES

[1] Ericsson, "5G for Business: A 2030 Market Compass," Ericsson, Tech. Rep., Oct. 2019.

[2] German Federal Ministry of Education and Research, *Industrie 4.0*. Bundesministerium für Bildung und Forschung (BMBF), 2017.

[3] 5G Alliance for Connected Industries and Automation, "5G for Connected Industries and Automation," 5G ACIA, Tech. Rep., 2018.

[4] 5G Alliance for Connected Industries and Automation, "A 5G Traffic Model for Industrial Use Cases," 5G ACIA, Tech. Rep., 2019.

[5] 5G Alliance for Connected Industries and Automation, "Key 5G Use Cases and Requirements," 5G ACIA, Tech. Rep., Mar. 2020.

[6] A. Aijaz, "Private 5G: The Future of Industrial Wireless," *IEEE Industrial Electronics Magazine*, vol. 14, no. 4, pp. 136–145, Dec. 2020.

[7] 5G Alliance for Connected Industries and Automation, "5G for Automation in Industry - Primary use cases, functions and service requirements," 5G ACIA, Tech. Rep., Jul. 2019.

[8] A. Aijaz *et al.*, "Shaping 5G for the Tactile Internet," in *5G Mobile Communications*, W. Xiang, K. Zheng, and X. S. Shen, Eds. Cham: Springer International Publishing, 2017, pp. 677–691.

[9] 3rd Generation Partnership Project, "Specification # 22.261 - Service requirements for next generation new services and markets," 3rd Generation Partnership Project, Technical Specification, 2019.

[10] N. Ahmed, A. Mansoor, and R. Ahmad, "Mission-Critical Machine-Type Communications: An Overview and Perspectives towards 5G," *IEEE Access*, vol. 7, pp. 127 198–127 216, Jan. 2019.

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

[12] P. Popovski *et al.*, "Wireless Access in Ultra-Reliable Low-Latency Communication (URLLC)," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[13] Beckhoff Automation GmbH, "EtherCAT – The Ethernet Fieldbus," Beckhoff Automation GmbH, Tech. Rep., Nov. 2018.

[14] X. Li *et al.*, "A review of industrial wireless networks in the context of Industry 4.0," *Wireless Networks*, vol. 23, no. 1, pp. 23–41, Jan. 2017.

[15] M. Weiner *et al.*, "Design of a low-latency, high-reliability wireless communication system for control applications," in *2014 IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 3829–3835.

[16] P. Popovski *et al.*, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[17] J. Sachs *et al.*, "5G Radio Network Design for Ultra-Reliable Low-Latency Communication," *IEEE Network*, vol. 32, no. 2, pp. 24–31, Mar. 2018.

[18] 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*, vol. 36, no. 11, pp. 2475–2485, Nov. 2018.

[19] 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.

[20] W. Zhang, M. S. Branicky, and S. M. Phillips, "Stability of networked control systems," *IEEE Control Systems Magazine*, vol. 21, no. 1, pp. 84–99, Feb. 2001.

[21] J. H. Braslavsky, R. H. Middleton, and J. S. Freudenberg, "Feedback Stabilization Over Signal-to-Noise Ratio Constrained Channels," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1391–1403, Aug. 2007.

[22] L. Zhang, H. Gao, and O. Kaynak, "Network-Induced Constraints in Networked Control Systems—A Survey," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 403–416, Feb. 2013.

[23] N. Elia and S. Mitter, "Stabilization of linear systems with limited information," *IEEE Transactions on Automatic Control*, vol. 46, no. 9, pp. 1384–1400, Sep. 2001.

[24] G. N. Nair *et al.*, "Feedback Control Under Data Rate Constraints: An Overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, Jan. 2007.

[25] J. Hespanha, P. Naghshtabrizi, and Y. Xu, "A Survey of Recent Results in Networked Control Systems," *Proceedings of the IEEE*, vol. 95, pp. 138–162, 2007.

[26] X. Zhang, Q. Han, and X. Yu, "Survey on Recent Advances in Networked Control Systems," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1740–1752, Oct. 2016.

[27] G. F. Franklin, M. L. Workman, and D. Powell, *Digital Control of Dynamic Systems*, 3rd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1997.

[28] 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*, vol. 51, no. 11, pp. 640–644, Nov. 2004.

[29] H. Gao and T. Chen, "Network-Based H∞ Output Tracking Control," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 655–667, Apr. 2008.

[30] B. Sinopoli *et al.*, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[31] N. Elia and J. N. Eisenbeis, "Limitations of Linear Control Over Packet Drop Networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 4, pp. 826–841, Apr. 2011.

[32] X. Wang and M. D. Lemmon, "Event-Triggering in Distributed Networked Control Systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 3, pp. 586–601, Mar. 2011.

[33] 3rd Generation Partnership Project, "Specification # 22.104 - Service requirements for cyber-physical control applications in vertical domains," 3rd Generation Partnership Project, Technical Specification, 2019.

[34] O. Ayan *et al.*, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, ser. ICCPS '19. New York, NY, USA: Association for Computing Machinery, Apr. 2019, pp. 109–117.

[35] C. Wu *et al.*, "Analysis of EDF scheduling for Wireless Sensor-Actuator Networks," in *2014 IEEE 22nd International Symposium of Quality of Service (IWQoS)*, May 2014, pp. 31–40.

[36] M. C. F. Donkers *et al.*, "Stability Analysis of Networked Control Systems Using a Switched Linear Systems Approach," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2101–2115, Sep. 2011.

[37] A. Cervin and T. Henningsson, "Scheduling of event-triggered controllers on a shared network," in *2008 47th IEEE Conference on Decision and Control*, Dec. 2008, pp. 3601–3606.

[38] M. H. Mamduhi *et al.*, "Event-triggered scheduling for stochastic multi-loop networked control systems with packet dropouts," in *53rd IEEE Conference on Decision and Control*, Dec. 2014, pp. 2776–2782.

[39] D. Han *et al.*, "Optimal sensor scheduling for multiple linear dynamical systems," *Automatica*, vol. 75, pp. 260–270, Jan. 2017.

[40] M. Eisen *et al.*, "Control Aware Radio Resource Allocation in Low Latency Wireless Control Systems," *IEEE Internet of Things Journal*, vol. 6, Apr. 2019.

[41] A. Kosta *et al.*, "Age and value of information: Non-linear age case," in *2017 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 326–330.

[42] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, Mar. 2012, pp. 2731–2735.

[43] I. Kadota, A. Sinha, and E. Modiano, "Optimizing Age of Information in Wireless Networks with Throughput Constraints," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Apr. 2018, pp. 1844–1852.

[44] B. Zhou and W. Saad, "Minimizing Age of Information in the Internet of Things with Non-Uniform Status Packet Sizes," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.
[45] X. Wang *et al.*, "AoI-Aware Control and Communication Co-design for Industrial IoT Systems," *IEEE Internet of Things Journal*, pp. 1–1, 2020.
[46] Y. Sun *et al.*, "Update or wait: How to keep your data fresh," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, Apr. 2016, pp. 1–9.
[47] J. P. Champati *et al.*, "Performance Characterization Using AoI in a Single-loop Networked Control System," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 197–203.
[48] M. Costa, M. Codreanu, and A. Ephremides, "On the Age of Information in Status Update Systems With Packet Management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897–1910, Apr. 2016.
[49] R. Devassy *et al.*, "Reliable Transmission of Short Packets Through Queues and Noisy Channels Under Latency and Peak-Age Violation Guarantees," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 721–734, Apr. 2019.
[50] J. Östman *et al.*, "Peak-Age Violation Guarantees for the Transmission of Short Packets over Fading Channels," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 109–114.
[51] 3rd Generation Partnership Project, "Specification # 38.101 - NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone," 3rd Generation Partnership Project, Technical Specification, 2021.
[52] ——, "Technical Specification 38.300 - NR; NR and NG-RAN Overall description; Stage-2," 3rd Generation Partnership Project, Technical Specification, 2021.
[53] T. Hößler *et al.*, "Dynamic Connectivity for Robust Applications in Rayleigh-Fading Channels," *IEEE Communications Letters*, vol. 24, no. 2, pp. 456–460, Feb. 2020.
[54] L. Scheuvens *et al.*, "Wireless Control Communications Co-Design via Application-Adaptive Resource Management," in *IEEE 5G World Forum*, Dresden, Germany, Oct. 2019.
[55] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, ser. London Mathematical Society Student Texts. Cambridge: Cambridge University Press, 2002.
[56] T. Hößler *et al.*, "Mission Availability for Wireless URLLC," in *2019 IEEE Global Communications Conference (GLOBECOM)*. Waikoloa, HI, USA: IEEE, Dec. 2019, pp. 1–6.
[57] ——, "Applying reliability theory for future wireless communication networks," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct. 2017, pp. 1–7.
[58] C. Grinstead and J. Snell, *Introduction to Probability*. American Mathematical Society, 2012.
[59] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *2014 IEEE International Symposium on Information Theory*, Jun. 2014, pp. 1583–1587.
[60] L. Scheuvens *et al.*, "State-Aware Resource Allocation for Wireless Closed-Loop Control based on Multi-Connectivity," in *IEEE Global Communications Conference (GLOBECOM)*, Taipei, Taiwan, 2020.
[61] ——, "System Analysis of State-Aware Resource Allocation for Closed-Loop Control Systems," in *IEEE Global Communications Conference (GLOBECOM)*, Taipei, Taiwan, 2020.
[62] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Courier Corporation, Jan. 1994.
[63] P. Wu and N. Jindal, "Coding versus ARQ in Fading Channels: How Reliable Should the PHY Be?" *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3363–3374, Dec. 2011.