Age of Information in Multi-Hop Connections with Tributary Traffic and no Preemption

Federico Chiariotti, *Member, IEEE,* Olga Vikhrova, *Member, IEEE,* Beatriz Soret, *Member, IEEE,* and Petar Popovski, *Fellow, IEEE*

***Abstract*—Age of Information (AoI) has gained significant attention from the research community because of its applications to Internet of Things (IoT) monitoring and control. In this work, we treat multihop connections over queuing networks with tributary flows and non-preemptive service: packets cannot be discarded because they are utilized for other system objectives, such as data analytics. Without preemption, the key tool for optimizing AoI is then the scheduling policy between the different data flows at each intermediate node. This is the subject of our analysis, along with the impact of packet erasure on the age. We derive upper and lower bounds for the average AoI considering several queuing policies in arbitrary network topologies, and present the results in different scenarios. Network topology, tributary traffic load, and link characteristics such as packet erasure generate complex trade-offs, which affect the optimal operation point and the age performance. The scheduling strategy at each node can also affect performance and fairness among users, particularly at critical bottleneck links, which have a significant impact on the overall performance of the whole network.**

I. Introduction

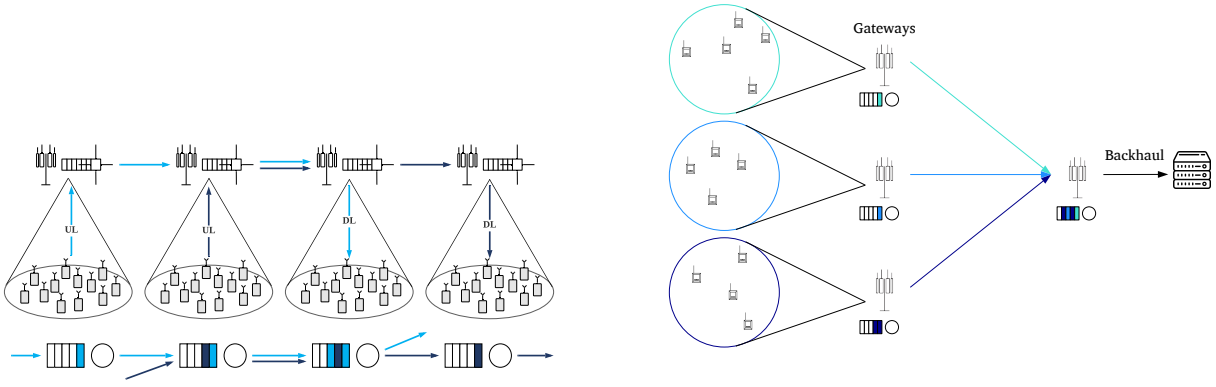
Age of Information (AoI) [1] is a metric that has recently attracted significant attention in Internet of Things (IoT) systems, particularly with respect to monitoring and remote control applications [2]. It is a process that measures the time elapsed since the generation of the last received update, i.e., the freshness of the information available to the receiver at any given time. By measuring the time since the generation of the *last* received packet at any given time, instead of the latency for each individual packet, it can provide some additional nuance, allowing system designers to calibrate the activation rate and duty cycle of sensors.

While latency is minimized in a scenario in which the traffic load on the network is very low, as communication buffers are almost always empty and there are no medium access issues, such a scenario will have a high AoI due to the longer interarrival time between packets. On the other hand, a high traffic load leads to a high transmission latency, which becomes the dominant component of the age. If we plot the AoI as a function of the update frequency, and consequently of the generated load, we often see a U-shaped curve, where the optimal update frequency is the value that strikes the best balance between update frequency and network congestion.

Most of the early works analyzing AoI dealt with single queuing systems. In this paper, we consider a less studied scenario of queuing *networks*, with multihop connections as well as tributary data flows at intermediate nodes. Two examples from Fig. 1 illustrate why this scenario is of interest in IoT systems. The first example, which is rapidly gaining traction in the real world, is given by Low Earth Orbit (LEO) satellite constellations [3], which form an *ad hoc* network such as the one shown in Fig. 1a. Each satellite is connected to the ground through an uplink (UL) and downlink (DL), and satellites have Inter-Satellite Links (ISLs) between each other [4]. The second example involves a terrestrial IoT network, and is shown in Fig. 1b. Multiple gateways can gather data from sensors, with independent wireless access, and then share a backhaul link to a server, which is often the bottleneck [5]. In both cases, the complex interactions between different flows of information, such as the light blue and dark blue ones in the figure, can give rise to interesting patterns in the AoI and change the optimization considerations in the system design.

A widely used tool in relation to AoI is service preemption [6], where some packets are dropped from the queues if such an action contributes to the AoI minimization. However, in many cases preemption is not desirable due to other system objectives. A recent work [7] has proven that there is an unavoidable trade-off between AoI and reliability, and that preemptive systems are optimal for AoI, but cause significant packet loss due to the dropping policy. This can become an issue in any case in which the AoI is not the only objective, but older packets still need to be delivered: one example would be a live monitoring and logging system, which requires data to be fresh, but also aims at reconstructing the historical trajectory of the system to train a machine learning model or analyze long-term patterns. This dual objective is common to many IoT scenarios, such as underwater networks [8]. While there have been some works in the literature analyzing arbitrary queuing networks [9]–[11], non-preemptive, continuous-time networks are still mostly unexamined. We will give more details on the differences between our work and the ones mentioned in Sec. II. In absence of preemption, the tributary flows provide a new flavor to the problem of AoI minimization, as now the intermediate nodes can make scheduling decisions that are affecting multiple data flows. In fact, it is important to note that, although we analyze a single multihop connection, the results are applicable to all involved connections of the tributary flows, as we explicitly address their interactions through

F. Chiariotti (corresponding author, email: fchi@es.aau.dk), B. Soret, and P. Popovski are with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark. O. Vikhrova is with Faculty of Information Technology and Communication Sciences, Tampere University, 33101 Tampere, Finland. B. Soret is also with the Telecommunication Research Institute (TELMA), Universidad de Málaga, 29010 Málaga, Spain. This work was partly funded by the IntellIoT project under the H2020 framework grant ID 957218, and partly by the Villum Investigator Grant "WATER" from the Velux Foundation, Denmark.



(a) Satellite network with four nodes and two exemplary path routes.

(b) Terrestrial IoT network with three gateways sharing a backhaul.

Fig. 1: Examples of multihop networks with tributary traffic flows.

the scheduling policy. Tandem structures with different rates and intersecting flows are one of the most general and complex models in queuing theory, and the additional degree of freedom in the queuing policy makes the AoI computation even more complex. We consider the Oldest Packet First (OPF) [12] and Maximum Age First (MAF) [13] policies and show that they can lead to non-trivial changes in the behavior of AoI. Finally, our model also includes unreliable wireless links, which are modeled as factors in the stochastic service time for each queue that works as a relay for the received data packets.

Our analysis derives tight upper and lower bounds of the average AoI for this model, which is useful in many applications. The main contributions of the paper are:

- We model a queuing network as a connected graph, with an arbitrary number of nodes and flows that can have any pair of nodes as their source and destination. The nodes have Markovian service and non-preemptive queuing, and all flows have Poisson arrivals.
- We compute tight upper and lower bounds to the average AoI for any flow in the network with non-preemptive, First Come First Serve (FCFS) queuing, which corresponds to a Jackson network [14] model. The bounds are derived for each individual flow, aggregating others as a single tributary traffic flow. However, the tributary traffic on each link depends on the topology and routing scheme, and the bounds depend on the routing and data rate of all other flows in the network.
- We show the performance of the system and the tightness of the bounds in two case studies, which are simulated with a Monte Carlo approach. The first is a line network in which the tributary traffic all has the same destination, and the second is a classic dumbbell topology with a single shared bottleneck link.
- We go beyond Jackson networks, which require the FCFS queuing policy to be applied at all nodes, and examine two other policies: OPF and MAF. Both policies rely on packet timestamps to schedule packets: OPF selects the packet with the lowest timestamp, i.e., the one that was generated first, independently of the arrival time at the node, while MAF maintains a measure of the AoI at the following node, transmitting packets from the source with the highest age first. These policies can reduce the average AoI and the unfairness between flows with different paths by considering the whole connection instead of each individual link.
- We consider the impact of packet erasures due to fluctuations of the wireless channel on the AoI. Naturally, losing packets has a negative impact on the reliability, but it can counterintuitively reduce the AoI: in high load scenarios, removing some packets from later links can help ease congestion, having a positive impact on the overall age. This trade-off is yet another instance of the dilemma *"how often should one update?"* [1], although the answer in a network with multiple flows is far from trivial.

A simplified version of our model was presented in [15], which analyzed a simple line network with no cross-traffic, or a case with 2 sources sharing a bottleneck in a tandem network with 2 links. A preliminary version of this work computed a looser upper bound for the AoI under the same conditions and was uploaded as a preprint in [16]: while the system model was similar, this version has an extended analysis, using more advanced mathematical tools to find a tighter upper bound and represent a more general scenario. Furthermore, the analysis of the results and of their potential application to AoI-oriented optimization is more complete in this work. Overall, this paper expands significantly on both previous versions, with a more complete analysis and improved results.

The rest of this paper has the following organization: first, an analysis of the relevant literature and the gaps in the state of the art is presented in Sec. II. Secondly, the system model and analytical tools to compute the AoI bounds are presented in Sec. III. The actual calculation of the upper and lower bounds is described in detail in Sec. IV, and numerical results are presented in Sec. V. Finally, Sec. VI concludes the paper, along with a discussion of potential future directions of this work.

II. Related work

While less than a decade old, AoI has been the subject of intense study: the first paper to define it [1] characterized it in $G/G/1$ queues, focusing on the two special cases with exponential and deterministic service, which are more analytically tractable. Several works have then considered different queuing models, attempting to characterize real wireless scenarios by including packet erasures [17] and retransmissions [18], or verifying the realism of existing models with experiments over live connections [19].

Another widely studied problem is represented by the optimization of the senders and relays when multiple sources share a link or connection, with the goal of maintaining a low age for all flows [20]. If we consider non-trivial networks, optimizing all the senders has been shown to be NP-hard [21], but near-optimal heuristics can be defined [22]. In particular, source optimization is crucial for uncoordinated systems such as slotted and unslotted ALOHA: the average AoI for such systems has been derived in [23] and [24]. Another recent work [25] combines slotted ALOHA and source control policies, optimizing transmission patterns to reduce the average AoI.

The scheduled case is a different problem: if a link is shared by multiple flows, and a coordinator (e.g., a base station) can decide which packets to prioritize, the scheduling policy becomes a major factor. The performance of this kind of system has been compared with uncoordinated systems in [26]. Finding the optimal scheduling policy [20] is a complex problem, but there are some valid heuristics: one such example is represented by the MAF policy [27], which selects the packet from the source with the highest age. If nodes are allowed to discard packets, considering a limited transmission period for packets, after which they are dropped, can also be beneficial for the overall performance of the system [28].

However, while most of the AoI literature concentrates on single queuing systems, more complex networks with multiple hops have recently drawn significant attention. In particular, the $M/M/1$ case with 2 queues in tandem has been studied extensively due to its analytical tractability: the average Peak Age of Information (PAoI) with multiple sources was derived in [29], and other works have tried to derive the full distribution of the age, either analytically [30] or using the Chernoff bound [31]

More general queuing networks, with multiple hops and multiple sources of traffic, are harder to analyze, and most works in the literature assume that each node applies service preemption, i.e., discards queued packets to make room for fresher updates. In fact, preemption is optimal in any tandem of $M/M/1$ queuing systems [6]: as queuing can cause a significant additional delay and AoI optimization does not require updates to be reliable, it is better to drop the packet in service and try to send the new one. A similar result holds for $M/M/k$ queues [9] and different arrival processes [32]. The effect of preemption on the AoI in tandem queues is analyzed in [10], which characterizes the moment generation function of the aging process. If service is not Markovian, the decision over whether to preempt becomes tougher, as time already spent in service reduces the remaining delay [33]. Queue replacement, in which only the freshest update from each source is kept in the queue, but the packet already in service is not dropped, can avoid the issue altogether, reducing the queue size without affecting time in service and leading to a $G/G/1/2$ queuing model for each node. These systems are analyzed in [34]–[36], which give results for single and multiple sources. Finally, a transport protocol to control the generation rate of status updates to minimize the AoI over the Internet is presented in [37].

In our work, we consider non-preemptive systems, which have a higher age but enable other data analytics tasks by presenting a more accurate picture of the process evolution, as packet losses are only due to the wireless channel and not due to the queuing policy.

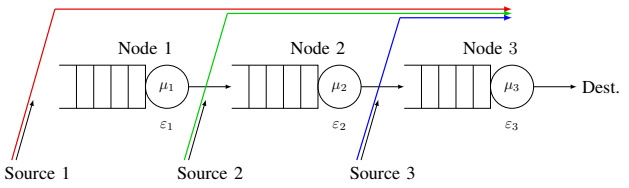
More extensive results can be obtained in discrete-time models, in which each hop takes 1 unit of time and there is some limitation on how often the nodes can transmit, usually due to interference or energy constraints. In [38], the problem of multihop networks with many source-destination pairs and interference constraints is addressed, and the optimal policy is reduced to solving the equivalent problem in which all source-destination pairs are just a single hop away. A recent extension [39] also considers the question of routing, which can be optimized together with scheduling at each node. The contention case was examined in [11], which gave upper and lower bounds for the average AoI and PAoI in a multi-source multihop wireless network with explicit channel contention. Like our work and [10], these works also consider arbitrary network topologies, but with a strong assumption of slotted time and universal synchronization: to the best of our knowledge, the scenario we consider in this work is still unexplored.

III. System model

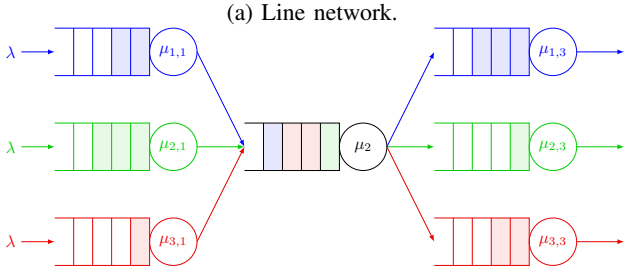
In the following, we will indicate vectors with bold letters (e.g., $\mathbf{x}$), random variables with capital letters (e.g., X), and sets with cursive capitals (e.g., $\mathcal{E}$). Additionally, $\mathbb{E}[X]$ stands for the expected value of the random variable X, $p_X(x)$ denotes its Probability Density Function (PDF), and $P_X(x)$ its Cumulative Density Function (CDF).

Let us consider a multihop communication system with multiple sources of packets and erasure channels between the nodes. Two simple examples of such networks are shown in Fig. 2. The system can be modeled as a connected directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ defined by its set of nodes $\mathcal{V}$, set of edges $\mathcal{E}$, and set of flows $\mathcal{F}$. The probability of packet erasure at link $e = (u, v) \in \mathcal{E}$ is given by $\varepsilon_e \in [0, 1)$, and the packet transmission delay over link e is an exponentially distributed random variable with parameter μ_e. In turn, each flow $f \in \mathcal{F}$ is defined by a tuple $f = (s_f, d_f, \lambda_f)$, with $s_f, d_f \in \mathcal{V}$ and $\lambda_f > 0$. Update packets of flow f are generated at the source node s_f according to a Poisson process with intensity λ_f, and the packets need to be delivered to node d_f. All packets then traverse the system following the routing path $\mathcal{P}_f$:

$$\mathcal{P}_f = \big\{(\mathbf{u}, \mathbf{v}) \in \mathcal{E}^{K_f}, u_1 = s_f, v_{K_f} = d_f, v_{i-1} = u_i, \\ u_i \neq u_j,\ \forall i, j \in \{1, \dots, K_f\} : i \neq j\big\}, \tag{1}$$



(a) Line network.



(b) Dumbbell network.

Fig. 2: Examples of networks supported by our model definition.

where K_f denotes the number of links in the routing path, $f \in \{1, \ldots, |\mathcal{F}|\}$. We can consider a line network in Fig. 2a as a first example. The queueing network has three nodes, $K = 3$, and the destination is denoted as d. We have three flows, represented with three different colors, with $\mathcal{P}_1 = \{(1,2),(2,3),(3,d)\}$ in red, $\mathcal{P}_2 = \{(2,3),(3,d)\}$ in green, and $\mathcal{P}_3 = \{(3,d)\}$ in blue. In the dumbbell network in Fig. 2b, the path from source i is $\mathcal{P}_i = \{(s_i,b_1),(b_1,b_2),(b_2,d_i)\}$, where the link (b_1,b_2) is the shared bottleneck. Note that we imposed a condition that there are no loops in the route. We denote the k-th link in $\mathcal{P}_f$ as $\mathcal{P}_f(k)$. Let us also define the partial path over the first k links for flow f as follows:

$$\mathcal{P}_f(1,\ldots,k) = \big\{(\mathbf{u},\mathbf{v}) \in \mathcal{E}^k, u_1 = s_f, v_{i-1} = u_i, \\ u_i \neq u_j, \forall i,j \in \{1,\ldots,K_f\} : i \neq j\big\}. \tag{2}$$

In the following, we consider one flow f and its communication path as a queuing network with K_f nodes in tandem, where each communication link is modeled by an $M/M/1$ system, as shown in Fig. 3. The choice of the $M/M/1$ queue as a model for each link is motivated by analytical tractability, as introducing complexity in tandem queues often creates limits to the mathematical solution. While the model is theoretical, it has been used successfully in wireless networks [40], [41], and can approximate resource-constrained satellites with multiple tasks and applications [42] or an ALOHA system with a sufficiently large number of users and exponential backoff, as argued in [30]. These two scenarios correspond to the examples presented in Fig. 1.

As other flows may share some links with flow f, their packets will compete for resources at the nodes of the given queuing network. Such flows are treated as sources of cross traffic in the queuing network for flow f. By repeating the analysis for all flows in the communication system, we can get the AoI and latency for the whole set of sources. In the line network, all flows share the same destination, which is at the end of the line, but they have different sources, and they all share the last communication link. On the other hand, the flows in the dumbbell network act as cross-traffic for other flows only at the central link shared by all flows. Let $\mathcal{L}_k^f$ define a set of flows, whose routes include link $\mathcal{P}_f(k)$:

$$\mathcal{L}_k^f = \Big\{f' \in \mathcal{F} \setminus \{f\} : \mathcal{P}_f(k) \in \mathcal{P}_{f'}\Big\}. \tag{3}$$

We can then consider a scenario in which the k-th link in path $\mathcal{P}_f$ is shared with flow f'. We define the index $x(f,f',k)$ as the index of that specific link on path $\mathcal{P}_{f'}$, i.e., the number of links in $\mathcal{P}_{f'}$ that flow f' traverses before the one shared with path $\mathcal{P}_f$:

$$x(f,f',k) = \Big\{\ell \in \{1,\ldots,K_{f'}\} : \mathcal{P}_{f'}(\ell) = \mathcal{P}_f(k)\Big\}. \tag{4}$$

In the case of the line network, flow 1 shares its second link with flow 2 (that same second link would be the first one in flow 2's path). On the other hand, in the dumbbell network, the index of the shared link is the same for all flows. The rate of the tributary traffic θ_k^f that arrives at node k of the queuing network can be defined as

$$\theta_k^f = \sum_{f' \in \mathcal{L}_k^f} \lambda_{f'} \prod_{i \in \mathcal{P}_{f'}(1,\ldots,x(f,f',k)-1)} (1-\varepsilon_i). \tag{5}$$

In the line network case, this corresponds to the sum of the rates of all sources that contribute to the node, thinned out by the packet erasures at each link, while in the dumbbell network, there is no tributary traffic except at the bottleneck, where all the traffic (excepting the packets generated by the considered source) is tributary traffic. Note that the arriving tributary traffic is Poisson, as it is a simple sum of several thinned Poisson flows, and thus follows Burke's theorem [43]. The rate of the tributary traffic π_k that leaves node k yields

$$\pi_k^f = \sum_{f' \in \mathcal{L}_k^f} \theta_k^{f'} \varepsilon_k + \sum_{f' \in \mathcal{L}_k^f \setminus \mathcal{L}_{k+1}^f} \theta_k^{f'} (1-\varepsilon_k). \tag{6}$$

As all sources in the line network share the same destination d, this value is 0 in all the nodes of the line network. On the other hand, in the dumbbell network all tributary traffic leaves immediately after the bottleneck. We can also compute the rate ψ_k^f of the new tributary traffic enters at node k, i.e., traffic from flows that did not share the $k-1$-th link:

$$\psi_k^f = \sum_{f' \in \mathcal{L}_k \setminus \mathcal{L}_{k-1} \setminus \{f\}} \lambda_{f'} \prod_{i \in \mathcal{P}_{f'}(1,\ldots,x(f,f',k)-1)} (1-\varepsilon_i). \tag{7}$$

In the line network, this corresponds to the source entering the system at node k, i.e., the one for which the link is the first one in the line. In the dumbbell network, this value is 0 for all links except the bottleneck. If the error rate ε is 0 for all the links in the network, and all nodes apply the FCFS queuing policy, our formulation is equivalent to the well-known Jackson network model [14]. The line and dumbbell networks are two simple examples of this class of network, but it can be extremely wide, with flows going from arbitrary points to arbitrary points in complex network topologies.

We will later analyze a slightly different network class, in which nodes apply the OPF and MAF queuing policies:

1) OPF [12] is a version of FCFS which sorts packets by their generation time (which can be read from a timestamp included in the header) instead of the time of their

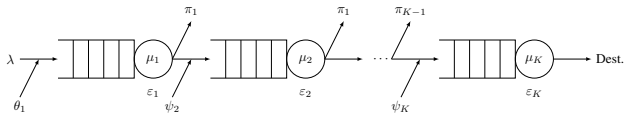


Fig. 3: General multihop queuing system. The considered source has rate λ, while other sources in the network (potentially at each node) are modeled as tributary traffic.

arrival at the node. In a way, OPF is FCFS from a system perspective, considering the whole network instead of a single link. It enhances fairness between different flows, as packets that have already gone through longer paths can traverse later links faster, at the expense of fresher packets from sources closer to their destination.

2) MAF [13] is explicitly aimed at reducing the AoI, and is more complex, as each node needs to keep track of the age of each source (at the node itself, not at the destination, as the node cannot know if and when packets reach their destination). Packets are then prioritized based on the current age of the source, with the source with the highest age being served first.

These policies can help control the competition for resources among different flows: while FCFS privileges flows with shorter paths, a distributed monitoring application might care equally about sensors that are close to it and ones farther away, and OPF and MAF can partially counterbalance the difference in their paths, as well as improving overall AoI.

As the tributary traffic rates and network parameters are all defined considering flow f, we omit the index f for the sake of brevity in what follows. Fig. 3 illustrates a general case of the reference queuing network, while Fig. 2a and Fig. 2b are special cases of the general model. In the line network example in Fig. 2a, all flows have the same destination. Some of the sources are farther away than others, so they have to traverse more links. Each node receives all traffic from its predecessor in the chain and from additional sources. In this kind of network, the bottleneck is often the last link, as it needs to deal with the traffic generated by all sources. Fig. 2b shows a classic dumbbell topology, in which flows follow separate paths that share one single link and then diverge again. It is easy to see how the model in Fig. 3 can represent both systems, as seen from the perspective of a single flow between source and destination, and how the analysis can be repeated for any of the flows in the network.

Finally, we do not consider the retransmission of packets, but rather a pure erasure channel. Considering retransmissions over a single link would change the analysis slightly, introducing a correlation between packet erasure and longer service times. Additionally, retransmissions might not be useful to reduce the AoI, as long as λ is not too small. End-to-end retransmission is also not modeled, as in most practical cases it would have a negative effect on AoI: if a new packet from the same source has already been sent, the retransmitted packet does not reduce the age, and in fact can even contribute to congestion, increasing the overall age for all sources.

A. System delay

The probability of delivering a packet through the first k links correctly is given by

$$p_s(k) = \prod_{j=1}^{k} (1 - \varepsilon_j)\,. \tag{8}$$

The total arrival rate at each node k is given by the surviving packets from the source and the tributary traffic, and it is $p_s(k-1)\lambda + \theta_k$. We then define the response rate at node k as:

$$\alpha_k = \mu_k - (p_s(k-1)\lambda_f + \theta_k). \tag{9}$$

We can just consider α_k instead of taking ψ_k and π_k separately by applying the Poisson Arrivals See Time Averages (PASTA) property [44] of open $M/M/1$ networks. If all systems apply the FCFS queuing policy, the total system time in the k-th node in steady-state is a Poisson process with rate α_k, according to Little's law. The overall service time and waiting time of the connection then follow a Hypoexponential distribution [45]. The vector $\boldsymbol{\alpha}$, containing the response rates for the K links, has N unique elements. The multiplicity of the i-th element is denoted as n_i. If $N = 1$ and $n_1 = K$, all rates are the same and the Hypoexponential distribution is reduced to the Erlang distribution. In fact, the Hypoexponential distribution is the convolution of several Erlang distributions. The connection as a whole can be considered as a single $M/PH/1$ queue [46] with errors. The PDF $p_T(t)$ of the total network time is given in [47] as:

$$p_T(t) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \gamma_{ij} \frac{t^{j-1}}{(j-1)!} e^{-\alpha_i t}, \tag{10}$$

where γ_{ij} is a coefficient defined as:

$$\gamma_{ij} = \prod_{\ell=1}^{N} \sum_{\mathbf{m} \in \mathcal{M}_{ij}} \prod_{\ell=1, \ell \neq i}^{N} \binom{n_\ell + m_\ell - 1}{m_\ell} \frac{(\alpha_\ell^{n_\ell})\,(-1)^{n_i - j}}{(\alpha_\ell - \alpha_i)^{n_\ell + m_\ell}}, \tag{11}$$

where $\mathcal{M}_{ij}$ is the set defined as:

$$\mathcal{M}_{ij} = \left\{ \mathbf{m} : m_i = 0, \sum_{\ell=1}^{N} m_\ell = n_i - j \right\}. \tag{12}$$

The average system delay is then $\bar{T} = \sum_{j=1}^{N} \frac{n_j}{\alpha_j^{n_j}}$.

B. Modeling the age

In order to model the AoI, we consider communication flow f and its associated line network as discussed above. Due to the exponential service time at each node, the rate of departure for flow f from node k is a thinned Poisson process with rate $\lambda_f(1 - \varepsilon_k)$. If θ_k denotes the total tributary traffic arriving at node k, therefore, the departure rate of the tributary traffic from node k equals to $\theta_k - \pi_k$. For the reader's convenience, the complete notation is given in Table I.

We follow standard practice and define the AoI $\Delta(t)$ as the difference between the current time t and the generation time of the last received packet. Without loss of generality, the system is first observed at $t = 0$ when the queue is empty with

TABLE I: Relevant notation

Notation	Definition	Notation	Definition
$\mathcal{G}$	Network graph	$\mathcal{V}$	Set of nodes
$\mathcal{E}$	Set of links	$\mathcal{P}$	Set of paths in $\mathcal{G}$
$R(s,d)$	Routing function	$\mathcal{F}$	Set of communication flows
K	Number of links in a route	W_i	Packet i total waiting time
λ	Packets generation rate at source	$W_{i,j}$	Packet i waiting time at node j
S_i	Packet i total service time	θ_k	tributary traffic rate at node k
$S_{i,j}$	Packet service time at node j	$\Delta(t)$	Aging process
$\boldsymbol{\theta}$	Vector of tributary traffic rates	$N(\mathcal{T})$	Number of arrivals from source by time $\mathcal{T}$
$\boldsymbol{\psi}$	Vector of new tributary traffic	$\Delta_{\mathcal{T}}$	Time average AoI over $\mathcal{T}$
$\boldsymbol{\pi}$	Vector of thinning rates	$\bar{\Delta}$	Average AoI
$p_s(k)$	Packet delivery success probability over k links	ρ_k	Traffic load at node k
$p_e(n;k)$	Probability of n consecutive errors over k links	ρ	Error-free load
ε_k	Channel error probability for the k-th link	S_k	Average service time at node k
$\boldsymbol{\varepsilon}$	Vector of channel error probabilities	$\boldsymbol{\alpha}$	Vector of response rates
μ_k	Packet service rate at node k	α_k	Response rate at node k
$\boldsymbol{\mu}$	Vector of service rates	p_c	Uplink collision probability
δ_{ij}	Hypoexponential distribution coefficient of packet service time	γ_{ij}	Hypoexponential distribution coefficient of packet total network time
r_i	Status update i time at destination	g_i	Status update i generation time
Y_i	Packet interarrival time	Q_i	Area under the $\Delta(t)$ AoI process
Z_i	Packet interdeparture time	Q'_i	Additional area below the $\Delta(t)$ process after a missed packet
T_i	Packet i network time		
$T_{i,j}$	Packet i system time at node j	$Q_i^{(n)}$	Total area around $\Delta(t)$ process after n missed packets
Ω_j	Time difference between arrival and departure time of two consecutive packets at node j	ν_{ij}	Hypoexponential distribution coefficient of packet service time (excluding the last link)

age $\Delta(0) \geq 0$ as illustrated in Fig. 4. The status update i is generated at time g_i and is received at the destination at time r_i. We define Y_i as the interarrival time $Y_i = g_i - g_{i-1}$ between two packets, Z_i as the interdeparture time $Z_i = r_i - r_{i-1}$, and T_i as the total network time in the system $T_i = r_i - g_i$. The latter includes the time spent in all the nodes (queuing and service time) until departure from the system at node K. The AoI at time t is then formally defined as:

$$\Delta(t) = t - \max_{i \in \mathbb{N}: r_i \leq t} (g_i). \tag{13}$$

Our definitions follow the work in [48], which considered a single queuing system, but in our case, the connection is modeled as a tandem of $M/M/1$ systems, each of which has to deal with tributary traffic. We remind the reader that, while other sources are abstracted as tributary traffic, the same analysis can be performed for any source in the network.

C. Derivation of the average age

Following the method from [48], we can now evaluate the average AoI by calculating the area under $\Delta(t)$ over a period $\mathcal{T}$. Other recent works on AoI, such as [32] and [49], use he Stochastic Hybrid Systems (SHS) method to derive the average AoI in simpler context with good results, but its high computational complexity makes it hard to apply in a general case with no preemption and potentially very large networks with arbitrary traffic. The evolution of the AoI exhibits the saw-tooth pattern plotted in Fig. 4, and we can then divide the area under the curve in a number of non-overlapping trapezoids Q_i. The average AoI $\Delta_{\mathcal{T}}$ is then given by the sum of the areas of the trapezoids, divided by the observation period $\mathcal{T}$:

$$\Delta_{\mathcal{T}} = \frac{1}{\mathcal{T}} \left(Q_{\text{ini}} + Q_{\text{last}} + \sum_{i=2}^{N(\mathcal{T})} Q_i \right), \tag{14}$$

where $N(\mathcal{T})$ is the number of arrivals from the source by time $\mathcal{T}$. The average AoI $\bar{\Delta}$ is given by the limit $\bar{\Delta} = \lim_{\mathcal{T}\to\infty} \Delta_{\mathcal{T}}$.

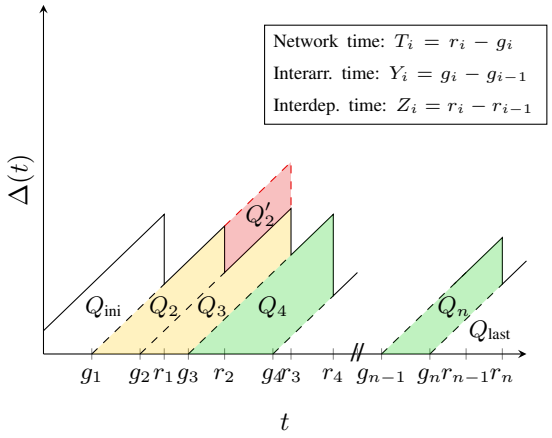


Fig. 4: Evolution of the AoI in a queue network with K nodes and errors. The network times T_i are defined as the total time spent in the system, since arrival in node 1 until departure in node K. The additional area Q'_2, highlighted in red, shows the increase in the AoI in case a packet is lost.

Let us consider an error-free transmission, i.e., $\varepsilon_k = 0 \, \forall k$. Each $Q_i^{(0)}$ (with $i > 1$) is a trapezoid whose area can be calculated as the difference between two isosceles right triangles, the larger of which has a cathetus $T_i + Y_i$, while the smaller has a cathetus T_i. Applying the Pythagorean theorem, the area of trapezoid $Q_i^{(0)}$ is given by:

$$Q_i^{(0)} = \frac{1}{2}(T_i + Y_i)^2 - \frac{1}{2}T_i^2 = Y_i T_i + \frac{Y_i^2}{2}. \tag{15}$$

We remark that the geometric analysis with error-free transmission is not in itself new, while to the best of our knowledge, the analysis with errors in the following is an original contribution. If there is an error and a packet is lost, the AoI computation changes slightly. As Fig. 4 shows, if packet 2 is lost, the additional area Q'_2, highlighted in red in the figure, needs to be included in the computation. If we consider the

$Q_i^{(1)} = Q_i^{(0)} + Q_{i-1}^{(0)} + Q'_{i-1}$ trapezoid resulting from a single lost packet, its area is given by:

$$\begin{aligned} Q_i^{(1)} &= \frac{1}{2}\left(T_i + Y_i + Y_{i-1}\right)^2 - \frac{1}{2}T_i^2 \\ &= Y_i T_i + Y_{i-1}T_i + Y_i Y_{i-1} + \frac{Y_i^2}{2} + \frac{Y_{i-1}^2}{2}. \end{aligned} \tag{16}$$

The trapezoid is the sum of the red and yellow areas in the figure. We can generalize this result to the case with n consecutive errors, denoting the trapezoid that ends as the i-th packet arrives after n lost packets as $Q_i^{(n)}$:

$$Q_i^{(n)} = \sum_{j=0}^{n}\left[Y_{i-j}T_i + \frac{1}{2}Y_{i-j}^2\right] + \sum_{j=0}^{n}\sum_{\ell=0,\ell\neq j}^{n}\frac{Y_{i-j}Y_{i-\ell}}{2}. \tag{17}$$

We note here that, since each node is an $M/M/1$ queue, the interarrival times Y_i and Y_{i-j} are independent for any $0 < j \leq i$. We can then compute the average area of the trapezoid, $\bar{Q}_i^{(n)}$:

$$\bar{Q}_i^{(n)} = \mathbb{E}\left[Q_i^{(n)}\right] = \sum_{j=0}^{n}\left[Y_{i-j}T_i\right] + \frac{(n+1)^2}{\lambda^2}. \tag{18}$$

The average AoI in the case of n consecutive errors can then be computed using the well-known geometric method, computing the average area of the trapezoid and dividing it by its side, i.e., by its extension on the x-axis:

$$\Delta_i^{(n)} = \frac{\bar{Q}_i^{(n)}}{\sum_{j=0}^{n} Y_{i-j}}\xi_i = \left(\frac{\lambda\sum_{j=0}^{n}\left[Y_{i-j}T_i\right]}{n+1} + \frac{(n+1)}{\lambda}\right)\xi_i, \tag{19}$$

where ξ_i is an indicator variable that is equal to 1 if the transmission of packet i was successful and 0 otherwise. When the system reaches a steady state, the system times are stochastically identical, i.e., $T =^{st} T_i =^{st} T_{i-1}$, and the same holds for the interarrival times. The probability of having n consecutive lost packets from the considered source after k links follows the geometric distribution with parameter $p_s(k)$:

$$p_e(n;k) = p_s(k)(1-p_s(k))^n. \tag{20}$$

In the following, we denote $p_s(k)$ simply as p_s to increase readability. We can then apply the law of total probability to (19) to compute the average age, assuming that packet i finds the system in steady state:

$$\begin{aligned} \bar{\Delta} &= \sum_{n=0}^{\infty} p_e(n;K)\Delta_i^{(n)} \\ &= \sum_{n=0}^{\infty} p_e(n;K)\left(\frac{\lambda\sum_{j=0}^{n}\left[Y_{i-j}T_i\right]}{n+1} + \frac{(n+1)}{\lambda}\right)\xi_i. \end{aligned} \tag{21}$$

The sum converges for $p_s \in (0,1]$, due to the properties of the geometric sum. The $\mathbb{E}\left[Y_{i-j}T_i\right]$ term cannot be easily computed, as the sequence of arrival times affects the queue at each link, leading to a complex effect on the system time. Deriving the AoI analytically is too cumbersome for non-trivial networks, but we can find relatively compact upper and lower bounds on the average AoI for each flow. The tributary traffic is considered in the $\boldsymbol{\alpha}$ vector, which represents the response rate of each node given the total load. In the case of more complex erasure models, such as the Gilbert-Elliott channel, the exponential series becomes a more complex summation, affecting the following results. However, the basic principles we use in the following to compute the upper and lower bounds remain the same, and the derivation can be adapted to these cases with some computational effort.

Intuitively, Y_i and T_i have a negative correlation: if packet i from a given source arrives a long time after packet $i-1$, it will often find fewer packets in the queue, as all older packets from the same source will be already delivered, and it will only have to deal with the tributary traffic. On the other hand, a packet arriving right after its predecessor from the same source will often find a longer queue, and thus, experience a longer waiting time. We would like to highlight again that the geometric analysis is not in and of itself novel, but to the best of our knowledge, it has never been applied for error-prone networks with multiple flows. The following derivations are then the main novel element of this work, as we will derive bounds on the average AoI in networks of arbitrary size, which has only been done for preemptive queues.

IV. Age of Information bounds

In this section, we analyze the average AoI by finding upper and lower bounds to the value of the series in (21). Naturally, these bounds will depend on the queuing policy applied in the network.

First, we try to expand the first term in the summation in (21). The correlation between the arrival process and the system times makes the series complicated to solve:

$$\begin{aligned} \sum_{n=0}^{\infty}\frac{\lambda p_e(n;K)}{n+1}\sum_{j=0}^{n}\mathbb{E}\left[Y_{i-j}T_i\right] &= \sum_{j=0}^{\infty}\sum_{n=j}^{\infty}\frac{\lambda p_e(n;K)}{n+1}\mathbb{E}\left[Y_{i-j}T_i\right] \\ &= \sum_{j=0}^{\infty} p_s(1-p_s)^j\Phi(1-p_s,1,j+1)\lambda\mathbb{E}\left[Y_{i-j}T_i\right], \end{aligned} \tag{22}$$

where $\Phi(z,s,a)$ is the Hurwitz-Lerch transcendent function [50], [51]. We can then get the expression of the average AoI as the series:

$$\bar{\Delta} = \sum_{j=0}^{\infty}\lambda p_s(1-p_s)^j\Phi(1-p_s,1,j+1)\mathbb{E}\left[Y_iT_i\right] + \frac{1}{\lambda p_s}. \tag{23}$$

If $p_s = 1$, i.e., the connection is error-free, this simplifies to:

$$\bar{\Delta} = \lambda\mathbb{E}\left[Y_{i-j}T_i\right] + \frac{1}{\lambda}. \tag{24}$$

The $\mathbb{E}\left[T_iY_i\right]$ term in (23) is the most complicated part. We first decompose the system time T_i by considering each of the K nodes in the connection separately, and decompose it further by dividing the system time $T_{i,k}$ into the waiting time $W_{i,k}$ and the service time $S_{i,k}$:

$$T_i = W_{i,1} + S_{i,1} + W_{i,2} + S_{i,2} + \ldots + W_{i,K} + S_{i,K}. \tag{25}$$

As commonly done in the literature, we rewrite this term to get:

$$\mathbb{E}\left[T_i Y_{i-j}\right] = \mathbb{E}\left[(W_i + S_i)Y_{i-j}\right] = \mathbb{E}\left[W_i Y_{i-j}\right] + \mathbb{E}\left[S_i\right]\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[W_i Y_{i-j}\right] + \sum_{k=1}^{K} \frac{\mathbb{E}\left[S_{i,k}\right]}{\lambda}. \tag{26}$$

As the service times $S_{i,k}$ are independent from each other, the distribution $p_S(t)$ of the service time $\sum_{j=1}^{N'} S_{i,j}$ is another Hypoexponential, defined as in (10) but with parameter μ_i instead of α_i:

$$p_S(t) = \sum_{i=1}^{N'} \sum_{j=1}^{n'_i} \delta_{ij} \frac{t^{j-1}}{(j-1)!} e^{-\mu_i t}, \tag{27}$$

In order to distinguish the coefficients of the distribution, which can be computed as in (11) by using the service rate vector $\boldsymbol{\mu}$ instead of $\boldsymbol{\alpha}$, we denote them as δ_{ij}. N' and n'_i are the equivalents of N and n_i for the vector $\boldsymbol{\mu}$, indicating the number of distinct elements and the multiplicity of each element, respectively. We then give the average service time, $\bar{S}$:

$$\bar{S} = \sum_{j=1}^{N'} \frac{n'_j}{\mu_j^{n'_j}}. \tag{28}$$

The service and arrival process are independent, so the second term in (26) can be simplified, but the correlation between W_i and Y_{i-j} makes the first term complex to calculate. Maintaining $\mathbb{E}\left[Y_{i-j}W_i\right]$ as an unknown variable, the average AoI then becomes:

$$\bar{\Delta} = \sum_{j=0}^{\infty} (1-p_s)^j p_s \Phi(1-p_s, 1, j+1)\left(\lambda \mathbb{E}\left[Y_{i-j}W_i\right] + \bar{S}\right) + \frac{1}{\lambda p_s}. \tag{29}$$

A. Upper bound on the average AoI

We can first give an upper bound on the average AoI. It is easy to prove that $W_{i,k} \leq T_{i-1,k}$, as packet i arrives to the k-th queuing system after packet $i-1$, and enters service immediately after packet $i-1$ is served. Moreover, Y_i is independent from T_{i-1} by definition, so we have $\mathbb{E}\left[T_{i-1}Y_i\right] = \mathbb{E}\left[T_{i-1}\right]\mathbb{E}\left[Y_i\right]$. We can then exploit the fact that $\mathbb{E}\left[T_{i-1}Y_i\right] \geq \mathbb{E}\left[W_i Y_{i-j}\right] \forall j$:

$$\mathbb{E}\left[W_i Y_{i-j}\right] \leq \mathbb{E}\left[T_{i-1}\right]\mathbb{E}\left[Y_i\right] = \frac{\bar{T}}{\lambda}. \tag{30}$$

We also know that the Hurwitz-Lerch transcendent is upper-bounded by the exponential sum:

$$\Phi(1-p_s, 1, j+1) = \sum_{n=0}^{\infty} \frac{(1-p_s)^n}{n+1} \leq \sum_{n=0}^{\infty} (1-p_s)^n \leq \frac{1}{p_s}. \tag{31}$$

We can then substitute the transcendent and $\mathbb{E}\left[W_i Y_{i-j}\right]$ with their upper bounds into (23):

$$\bar{\Delta} \leq \sum_{j=0}^{\infty} (1-p_s)^j \bar{T} + \frac{1}{\lambda p_s} = \frac{\lambda \bar{T} + 1}{\lambda p_s}. \tag{32}$$

Interestingly, the upper bound is equivalent to the average AoI assuming that Y_{i-j} and T_i are independent. In fact, this upper bound was presented as an approximation of the AoI in our preliminary work [16], while the derivation above proves that it is actually a tight upper bound in all cases. As finding $\bar{T}$ involves the calculation of the response rate α_k for each link in the path, which depends on the cross-traffic, the worst-case complexity for the upper bound calculation is $O(K|\mathcal{V}||\mathcal{F}|)$, but it will be much lower in most practical cases.

B. Lower bound on the average AoI

In order to give a lower bound on the average age, we can divide the waiting time W_i into K components by considering each node in the connection separately, getting $W_i = W_{i,1} + \ldots + W_{i,K}$. Each $W_{i,k}$ depends on the time elapsed from the arrival of packet i to system k to the departure of packet $i-1$ from it. As packet i might arrive after packet $i-1$ has left the system, this time difference, which we denote as $\Omega_{i,k}$, can be negative:

$$\Omega_{i,k} = \begin{cases} T_{i-1,1} - Y_i & \text{if } k = 1; \\ T_{i-1,k} - W_{i,k-1} - S_{i,k-1} & \text{if } k > 1. \end{cases} \tag{33}$$

If $\Omega_{i,k} < 0$, the waiting time $W_{i,k}$ is 0, i.e., $W_{i,k} = (\Omega_{i,k})^+$, where $(x)^+ = \max(x, 0)$ is the positive part function. The total waiting time is then simply given by:

$$W_i = \sum_{k=1}^{K} (\Omega_{i,k})^+ = (T_{i-1,1} - Y_i)^+ + \sum_{k=2}^{K} \left(T_{i-1,k} - W_{i,k-1} - S_{i,k-1}\right)^+. \tag{34}$$

It is trivial to prove that the sum of positive parts is larger than the positive part of the sum:

$$\sum_{i=1}^{n} (x_i)^+ \geq \left(\sum_{i=1}^{n} x_i\right)^+ \quad \forall \mathbf{x} \in \mathbb{R}^n, \forall n \in \mathbb{N}. \tag{35}$$

Thus, we can write a lower bound on the total waiting time of packet i as:

$$W_i \geq \left(\sum_{k=1}^{K} \Omega_{i,k}\right)^+ = \left(T_{i-1} - Y_i - \sum_{k=1}^{K-1} S_{i,k}\right)^+ = \left(T_{i-1} - Y_i - S_{\setminus K}\right)^+, \tag{36}$$

where $S_{\setminus K} = \sum_{k=1}^{K-1} S_{i,k}$. The bound becomes exact if and only if packet i is queued at each node, i.e., $\Omega_{i,k} > 0 \, \forall k$. If packet i finds an empty queue on the k-th system, the actual waiting time is longer than the bound in (36) by $|\Omega_{i,k}|$. We can now derive $\mathbb{E}[W_i Y_{i-1}]$ by taking (36) and decomposing it further:

$$\begin{aligned} \mathbb{E}\left[W_i Y_{i-1}\right] &\geq \mathbb{E}\left[Y_{i-1}\left(T_{i-1} - Y_i - S_{\setminus K}\right)^+\right] \\ &= \mathbb{E}\left[\left(Y_{i-1}T_{i-1} - Y_{i-1}Y_i - Y_{i-1}S_{\setminus K}\right)^+\right] \\ &\geq \mathbb{E}\left[W_i Y_i\right] + \frac{1}{\lambda \mu_K} - \frac{2}{\lambda^2}. \end{aligned} \tag{37}$$

We then apply the same procedure iteratively to get $\mathbb{E}[W_i Y_{i-j}]$:

$$\mathbb{E}\left[W_i Y_{i-j}\right] \geq \mathbb{E}\left[W_i Y_i\right] + j\left(\frac{1}{\lambda\mu_K} - \frac{2}{\lambda^2}\right). \tag{38}$$

We can now go back to (21) and use the values we have derived in the summation:

$$\begin{aligned}\bar{\Delta} =& \sum_{n=0}^{\infty} \lambda p_e(n;K)\left(\frac{\mathbb{E}\left[Y_i^2\right](n+1)}{2} + \sum_{j=0}^{n}\frac{\mathbb{E}\left[Y_{i-j}T_i\right]}{n+1}\right)\\ \geq& \frac{1}{\lambda p_s} + \bar{S} + \lambda\mathbb{E}\left[W_i Y_i\right] + \sum_{n=1}^{\infty}\frac{np_s(1-p_s)^n}{n+1}\left(\frac{1}{\mu_K} - \frac{2}{\lambda}\right).\end{aligned} \tag{39}$$

Because of the stability condition, we know that $\mu_K > \lambda$, so the term in the sum is always negative. To compute the lower bound on the AoI, we then need to find an upper bound to the sum. We can then simply use the fact that $n+1 > n$ to simplify the sum and get a closed-form lower bound, which holds for all three considered policies:

$$\bar{\Delta} \geq \frac{1}{\lambda p_s} + \bar{S} + \lambda\mathbb{E}\left[W_i Y_i\right] + (1-p_s)\left(\frac{1}{\mu_K} - \frac{2}{\lambda}\right). \tag{40}$$

C. Lower bound for the FCFS policy

The bound in (40) still depends on the unknown term $\mathbb{E}\left[W_i Y_i\right]$. We now compute the lower bound for it in the FCFS case. In order to do so, we condition the expectation on Y_i and $S_{\setminus K}$:

$$\mathbb{E}\left[W_i | Y_i = y, S_{\setminus K} = s\right] \geq \mathbb{E}\left[(T_i - y - s)^+\right]. \tag{41}$$

We can exploit the fact that the departure process from any node k is a Poisson process [43] to analyze each node separately [14] and apply the network decomposition method. First, we derive the lower bound on $\mathbb{E}\left[W_i Y_i | S_{\setminus K} = s\right]$ by applying the law of total probability:

$$\begin{aligned}&\mathbb{E}\left[W_i Y_i | S_{\setminus K} = s\right] \geq \int_0^{\infty} y\mathbb{E}\left[W_i | Y_i = y, S_{\setminus K} = s\right] p_{Y_i}(y)dy\\ &= \int_0^{\infty} y\int_{y+s}^{\infty}(t-y-s)p_T(t)p_{Y_i}(y)dtdy\\ &\geq \int_0^{\infty}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{\ell=0}^{j}\frac{y\lambda e^{-\lambda y}\gamma_{ij}e^{-\alpha_i(y+s)}(y+s)^{\ell}(j-\ell)}{(\ell!)\alpha_i^{j-\ell+1}}dy\\ &\geq \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{\ell=0}^{j}\frac{\gamma_{ij}\lambda(j-\ell)}{(\ell!)\alpha_i^{j-\ell+1}}e^{-\alpha_i s}\int_0^{\infty} y(y+s)^{\ell}e^{-(\alpha_i+\lambda)y}dy\\ &\geq \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{\ell=0}^{j}\sum_{m=0}^{\ell+1}\frac{\lambda\gamma_{ij}(j-\ell)(\ell-m+1)s^m e^{-\alpha_i s}}{(m!)\alpha_i^{j-\ell+1}(\alpha_i+\lambda)^{\ell-m+2}}.\end{aligned} \tag{42}$$

We now define vector $\boldsymbol{\mu}_{\setminus K} = (\mu_1, \ldots, \mu_{K-1})$, denoting the Hypoexponential coefficients linked to it, which can be derived as we did above, by ν_{ij}. As before, to avoid confusion, we define N'' as the number of unique elements of the vector and n_i'' as the multiplicity of the i-th such unique element. We then remove the condition on $S_{\setminus K}$:

$$\begin{aligned}&\mathbb{E}\left[W_i Y_i\right] \geq \int_0^{\infty}\mathbb{E}\left[W_i Y_i | S_{\setminus K} = s\right] p_{S_{\setminus K}}(s)ds\\ &\geq \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{\ell=0}^{j}\sum_{m=0}^{\ell+1}\sum_{o=1}^{N''}\sum_{p=1}^{n_\ell''}\frac{\gamma_{ij}\nu_{\ell m}\lambda(j-\ell)(\ell-m+1)}{\alpha_i^{j-\ell+1}(\alpha_i+\lambda)^{\ell-m+2}}\\ &\quad\times\frac{1}{m!(p-1)!}\int_0^{\infty}s^{m+p-1}e^{-(\alpha_i+\mu_o)s}ds\\ &\geq \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{\ell=0}^{j}\sum_{m=0}^{\ell+1}\sum_{o=1}^{N''}\sum_{p=1}^{n_\ell''}\frac{\gamma_{ij}\nu_{\ell m}\lambda(j-\ell)(\ell-m+1)}{\alpha_i^{j-\ell+1}(\alpha_i+\lambda)^{\ell-m+2}}\\ &\quad\times(\alpha_i+\mu_o)^{-(m+p)}\binom{m+p-1}{m},\end{aligned} \tag{43}$$

We can then obtain the lower bound on $\bar{\Delta}$ by substituting the value of (43) into (40). The overall worst-case complexity for the bound calculation is $O(K^5|\mathcal{V}||\mathcal{F}|)$, which is considerably higher than the upper bound complexity, but still polynomial in nature. However, if the response rates for different links in a flow's path are all different, as would be common in practical networks, the complexity is reduced to $O(K|\mathcal{V}||\mathcal{F}|)$. We note that the difference between the two bounds, which we denote as η_{FCFS}, is also computable:

$$\begin{aligned}\eta_{\text{FCFS}} =& \bar{T} - \bar{S} - (1-p_s)\left(\frac{1}{\mu_K} - \frac{2}{\lambda}\right)\\ &- \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{\ell=0}^{j}\sum_{m=0}^{\ell+1}\sum_{o=1}^{N''}\sum_{p=1}^{n_\ell''}\binom{m+p-1}{m}\\ &\times\frac{\gamma_{ij}\nu_{\ell m}\lambda^2(j-\ell)(\ell-m+1)}{\alpha_i^{j-\ell+1}(\alpha_i+\lambda)^{\ell-m+2}(\alpha_i+\mu_o)^{m+p}}.\end{aligned} \tag{44}$$

D. Lower bound for the OPF and MAF policies

If we consider the OPF policy, FCFS is not applied at each separate node, but rather to the whole connection: packets that were generated first at their source are served first at each node, regardless of the order in which packets arrived at that specific node. The MAF policy [27] takes this one step further by considering age directly.

The lower bound in (40) is still valid for these policies, but the value of $\mathbb{E}\left[W_i Y_i\right]$ is different due to the different scheduling. The lower bound on it for OPF and MAF relies on two bounds, which remove some possible cases from the calculation. First, as we did for FCFS, we consider the case in which packet i is queued at each link, i.e., $\Omega_{i,k} > 0\,\forall k$. In this case, the transmission of a packet begins immediately after the previous packet was sent. Secondly, we only consider the case in which all packets that arrive at the node after packet i also have a higher timestamp, i.e., are "younger": in this case, the packet is only queued behind the packets that had already arrived when it got to system k. In reality, cross traffic might have a longer path to traverse, and later packets might well be older: as we do not consider this case in the

bound calculation, this makes the bound looser, particularly if the connection is congested. As we show in the following sections, these assumptions still result in a very tight bound if the network is not too congested, which is the same for both OPF and MAF, as they behave in the same way in those cases.

The steady-state distribution $\Pi_k(n)$ of the number of waiting packets at node k is given by $(1-\rho_k)\rho_k^n$, with $\rho_k = \frac{p_s(k)\lambda+\theta_k}{\mu_k}$. This is true for all nodes, as the overall output of an $M/M/1$ queuing system is always a Poisson process, independently of the queuing policy. If we denote the number of packets in the queue as L_k, we can give the conditional expectation of $W_{i,k}$:

$$\begin{aligned}
&\mathbb{E}\left[W_{i,k}|Y_i = y, S_{i,k-1} = s\right] \\
&\geq \sum_{n=0}^{\infty} \Pi_k \mathbb{E}\left[W_{i,k}|Y_i = y, S_{i,k-1} = s, L_k = n\right] \\
&\geq \sum_{n=0}^{\infty} (1-\rho_k)\,\rho_k^n \int_s^{\infty} \frac{\mu_k^n t(t-s)^{n-1} e^{-\mu_k(t-s)}}{(n-1)!} dt \\
&\geq (1-\rho_k)\rho_k s e^{-\alpha_k s}.
\end{aligned} \tag{45}$$

We now apply the law of total probability for $k > 1$, i.e., the nodes past the first one:

$$\begin{aligned}
\mathbb{E}\left[W_{i,k}|Y_i = y\right] &\geq \int_0^{\infty} P_{S_{k-1}}(s)\mathbb{E}\left[W_{i,k}|Y_i = y, S_{i,k-1} = s\right] ds \\
&= \int_0^{\infty} \mu_{k-1} e^{-(\mu_{k-1}+\alpha_k)s}(1-\rho_k)\rho_k s ds \\
&\geq \frac{(1-\rho_k)\rho_k\mu_{k-1}}{\alpha_k + \mu_{k-1}}.
\end{aligned} \tag{46}$$

We now remove the condition over Y_i in the same way:

$$\mathbb{E}\left[W_{i,k}Y_i\right] \geq \frac{(1-\rho_k)\rho_k\mu_{k-1}}{(p_s(k)\lambda+\theta_k)\,(\alpha_k+\mu_{k-1})}. \tag{47}$$

As OPF and MAF are equivalent to FCFS for the first node in the connection, the value of the expected queuing time $\mathbb{E}[W_{i,1}]$ is:

$$\mathbb{E}\left[W_{i,1}|Y_i = y\right] = \int_0^{\infty} \frac{\lambda y e^{-(\alpha_1+\lambda)y}}{\alpha_1} dy = \frac{\lambda}{\alpha_1\mu_1^2}. \tag{48}$$

The lower bound on $\mathbb{E}\left[W_i Y_i\right]$ is then given by the sum of (47) and (48):

$$\mathbb{E}\left[W_i Y_i\right] \geq \frac{\lambda}{\alpha_1\mu_1^2} + \sum_{k=2}^{K} \frac{(1-\rho_k)\rho_k\mu_{k-1}}{(p_s(k)\lambda+\theta_k)\,(\alpha_k+\mu_{k-1})}. \tag{49}$$

We then get the lower bound on the age by substituting (49) into (40):

$$\begin{aligned}
\bar{\Delta} \geq &\frac{1}{\lambda p_s} + \bar{S} + \frac{\lambda^2}{\alpha_1\mu_1^2} + (1-p_s)\left(\frac{1}{\mu_K} - \frac{2}{\lambda}\right) \\
&+ \sum_{k=2}^{K} \frac{\lambda(1-\rho_k)\rho_k\mu_{k-1}}{(p_s(k)\lambda+\theta_k)\,(\alpha_k+\mu_{k-1})}.
\end{aligned} \tag{50}$$

TABLE II: Simulation parameters.

Parameter	Value	Description
K_{line}	$\{2, 6, 10\}$	Number of relay nodes (line)
$\mu_{1,\dots,K-1}$	1	Service rate of the first $K-1$ links
μ_K	0.8	Service rate of the last link (line)
μ_K	1	Service rate of the last link (dumbbell)
ψ	0	Cross traffic rate (line)
ε	0.02	Error probability for all links
N_{pkt}	100000	Total number of packets for each source
K_{db}	4	Number of links (dumbbell)
N_{db}	$\{2, 6, 10\}$	Number of cross traffic sources (dumbbell)

In this case, the worst-case complexity for the bound calculation is $O(K|\mathcal{V}||\mathcal{F}|)$, exactly the same as the upper bound complexity. The difference η_{OPF} between the two bounds is:

$$\begin{aligned}
\eta_{\text{OPF}} = &\frac{\bar{T}}{p_s} - \bar{S} - (1-p_s)\left(\frac{1}{\mu_K} - \frac{2}{\lambda}\right) - \frac{\lambda}{\alpha_1\mu_1^2} \\
&- \sum_{k=2}^{K} \frac{(1-\rho_k)\rho_k\mu_{k-1}}{(p_s(k)\lambda+\theta_k)\,(\alpha_k+\mu_{k-1})}.
\end{aligned} \tag{51}$$

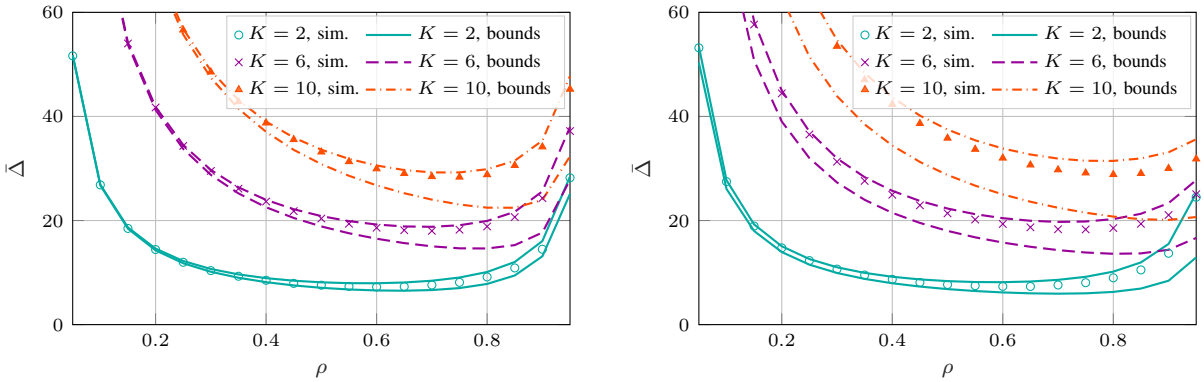
V. Numerical evaluation

We now present two case studied, which were simulated[1] using a Monte Carlo approach. In each simulation, we discarded the first 1000 packets from each source, which represented the initial transition to the steady state, and the final 1000 packets, to ensure that the results reflected the steady state behavior of the system. Excluding these initial and final transitory periods, we simulated $N_{\text{pkt}} = 10^5$ packets for each source. The full list of parameters we used is in Table II. The scenarios we considered correspond to the line network in Fig. 2a and the dumbbell network Fig. 2b. Both scenarios are special cases of the general system model depicted in Fig. 3, and they represent two common network configurations.

- *Line network* (Fig. 2a): this is an extension of the tandem, as this scenario involves a sequence of K links one after the other. The destination for all flows is placed at the end of the line, and there is a source placed before each link, with rate θ_i. As all sources share the common destination, $\psi_i = 0\,\forall i \leq K$. Naturally, the last link is the bottleneck, as packets from all sources converge on it. This kind of model can represent a satellite network in which ground nodes placed in remote areas report to a ground station through a chain of K satellites.
- *Dumbbell network* (Fig. 2b): this topology represents a case with a shared backhaul like the one we discussed in the introduction, with several gateways with independent wireless access using the same link to connect to the server. The N sources have N different destinations, but they share the k-th link in their path, which becomes the bottleneck: we have $\theta_j = 0\ \forall j \neq k$ and $\psi_k = 1$.

These two scenarios represent two extremes among the networks that the general model can represent: in the former, tributary traffic accumulates all the way to the final link, while in the other, it is concentrated in a single link in the middle,

[1]The Matlab code for the simulations and bounds is available at https://github.com/AAU-CNT/network_aoi.



(a) Error-free network ($\varepsilon = 0$)

(b) Error-prone network ($\varepsilon = 0.02$)

Fig. 5: Average AoI as a function of the maximum load for the line network scenario using the FCFS policy.

while all other links are less loaded. However, the graph model is flexible and can represent any topology, with multiple sources and destinations. Jackson networks are a subclass of the possible networks that can be represented, as they require no channel errors and the use of the FCFS queuing policy.

A. *Line network*

We first consider the line network scenario. We consider a single source for each link, so that we have K sources, each with rate λ_k. In the following, we consider all sources to have the same packet generation rate λ, plotting the AoI as a function of the maximum load ρ:

$$\rho = \frac{\lambda_1 + \sum_{j=2}^{K} \theta_j}{\mu_K}. \tag{52}$$

We analyzed the error-free and error-prone case, considering the same erasure probability ε for all links. We do not consider packet errors in our computation of ρ to provide a meaningful comparison between the error-free and error-prone cases. In the line network, there is one source per node, and the error-free load is the same for all sources, as all connections share the bottleneck (placed in the downlink).

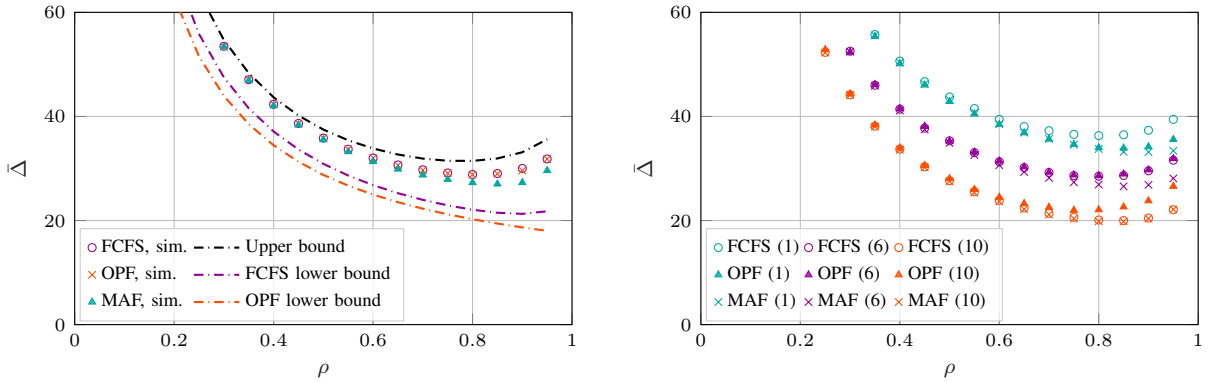
We first consider the FCFS policy, whose performance is shown in Fig. 5. The average AoI is computed over all sources in the network, and the bounds are averaged for each source, as they are computed separately. We can easily see that AoI follows a U-shaped curve: if the traffic load is very low, the average AoI is mostly given by the long interarrival times, while if the traffic increases too much, network congestion becomes the dominant factor in the age. For instance, when $\rho = 0.05$ and $K = 10$, the arrival rate λ_k for each source is just 0.004, as $\mu_K = 0.8$. This means that the average interarrival time is 250. On the other hand, a load above 0.9 can cause significant queuing delays, and AoI tends to infinity as ρ approaches 1. Moreover, the effect of packet erasures can be easily seen by comparing the results in the error-free scenario in Fig. 5a and in the error-prone scenario in Fig. 5b. While channel errors have a negligible negative effect if the traffic load is low, the reduced congestion can actually have a positive impact on AoI if the load is very high: since packets are frequent, one loss does not significantly increase the AoI, and the reduced load on the downlink can improve the congestion and decrease queuing delays. This effect is particularly noticeable for $\rho > 0.8$, with a significant difference in the AoI.

The bounds we derived for the average AoI are very tight for low values of ρ, as the approximation on the queuing time has a limited effect on the overall AoI. The upper bound is particularly tight, both in the error-free and error-prone scenarios. We can see that the lower bound becomes looser for longer line networks, but the upper bound remains tight in all cases.

We can also consider the impact of the scheduling policy, comparing the three policies in the error-prone scenario. In this case, we only consider $K = 10$, the setting with the largest performance differences. Fig. 6a shows the performance of the three policies in this scenario: OPF performs slightly better than FCFS on average, while the difference between them and MAF is more significant. As discussed above, the lower bound is looser for OPF and MAF, while the upper bound is the same for all policies and is tighter overall.

Interestingly, as Fig. 6b shows, FCFS actually outperforms the other policies if we only consider source 10, i.e., the one closest to the destination: on the other hand, the farthest source performs significantly worse. For all policies, the first source is the one with the highest AoI, as it has to traverse more links, but prioritizing older packets reduces this effect, allowing the packets from sources farther away to jump to the front of the line if they have already suffered significant delays.

The difference between FCFS and OPF is then not in the average performance, but in their fairness: while OPF considers the accumulated delay for packets from farther sources, giving them a higher priority, FCFS considers each link individually, leading to a lower fairness. However, MAF can combine the best aspects of the two policies, as it performs as well as the best of the two for the first, sixth and tenth source, i.e., across the whole chain of links. Furthermore, MAF can easily prevent a single source with a bad connection from increasing the age for all others, maintaining age fairness because it uses age directly as a scheduling metric. This might be useful for several IoT applications, which monitor a process over a wide



(a) Average AoI for all sources.

(b) AoI for the first, sixth and last source.

Fig. 6: Average AoI as a function of the maximum load for the FCFS, OPF, and MAF policies in a line network with $K = 10$ and $\varepsilon = 0.02$.

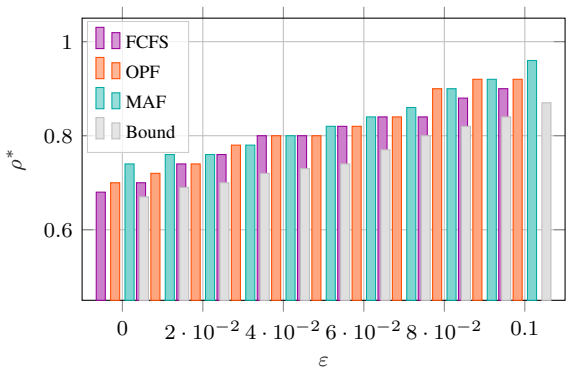


Fig. 7: Optimal ρ^* as a function of ε for the three policies in a line network, $K = 6$.

geographical area and require a measure of fairness among different sensors: to benefit the application, the freshness of the information from a sensor should not depend on its location, as all sensors have equal importance.

We can also examine the effect of increasing the error probability on the links on the optimal value of ρ, i.e., the one that minimizes the average AoI, which we denote as ρ^*. As Fig. 7 shows, the three policies are very similar, with OPF and MAF having a slightly higher optimal load at which the average AoI is minimized. As the error probability increases, so does the optimal load for all policies. This is intuitively correct, as the packets that are erased by the errors on earlier queues reduce the actual load on the last link. We remind the reader that ε is the error rate for each link, and the source farthest from the destination will have to go through all K links, which means that only a fraction $(1-\varepsilon)^{K-1}$ of its packets will reach the last link. We can also use the upper bound in (32) to determine the optimal value ρ^* directly: while the result is a K-th degree polynomial, the minimum is easy to evaluate numerically, and extremely close to the actual optimum for the FCFS policy. The optimization problem to find the minimum in the case in which all sources have the same λ is:

$$\rho^* = \arg\min_{\rho} \frac{K + \rho \sum_{j=1}^{N} \frac{n_j}{\alpha_j^{n_j}}}{\rho p_s}. \tag{53}$$

Finally, we can take a deeper look at the fairness by using a well-known fairness measure, the Jain Fairness Index (JFI) $\mathcal{J}(\mathbf{x})$:

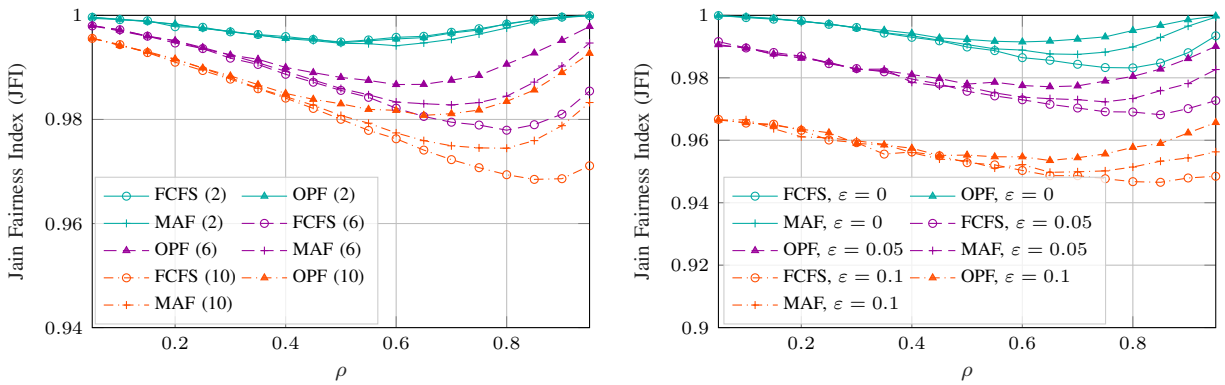
$$\mathcal{J}(\mathbf{x}) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2}, \tag{54}$$

where n is the length of vector $\mathbf{x}$. In this case, the metric on which we compute the JFI is the average age for each source. Fig. 8 shows the JFI as a function of the load, varying the number of links and error probability. Interestingly, the JFI highly depends on the load, and values of ρ closer to the optimum have a lower fairness. However, there is a stark difference between the three policies, particularly in longer line network, as shown in Fig. 8a: using MAF can significantly improve the performance over OPF, which in turn is better than FCFS. The packet error rate also plays an important role in fairness, with an even stronger effect than the length of the line network. However, the effect of the different policies is still clear in Fig. 8b.

B. *Dumbbell topology*

We now look at the dumbbell network scenario. We have N sources, each of which has to go through $K = 5$ links to reach the destination. The third link in each connection is the shared bottleneck, i.e., $\theta_3 > 0$ and $\theta_4 = 0$. As for the line network, we consider the packet generation rate for each source to be λ, so that the total error-free load on the bottleneck is $\rho = N\lambda$. The system is symmetric for all sources, so its expected fairness is always perfect for all policies.

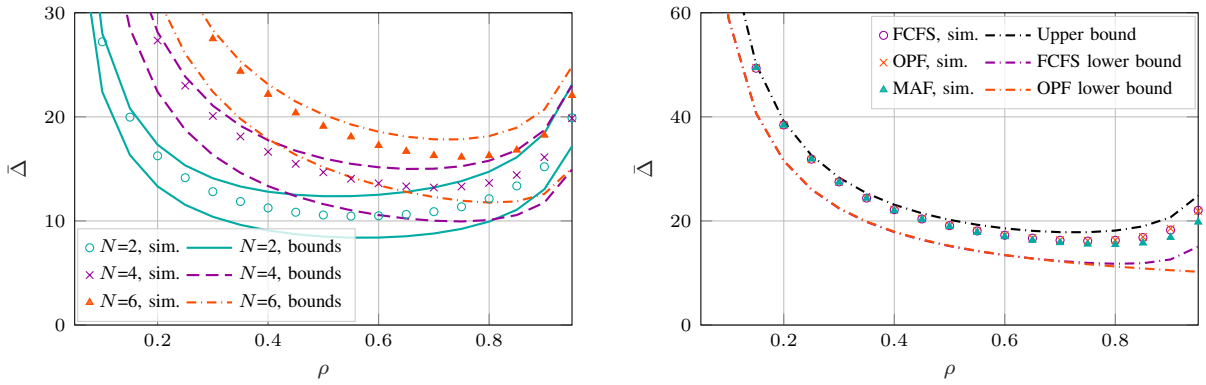
As above, we first consider the FCFS policy, considering the error-prone network. Fig. 9a shows the average AoI, considering three different values of N. In this case, the actual performance of the system is farther from the upper bound, but the two bounds taken together can give system designers a reliable indication of expected performance. Interestingly,



(a) JFI in a line network with $\varepsilon = 0.02$ (the numbers in the legend indicate the value of K).

(b) JFI in a line network with $K = 6$ with different values of ε.

Fig. 8: JFI as a function of the maximum load for the FCFS, OPF, and MAF policies in a line network.



(a) FCFS policy, $N \in \{2, 4, 6\}$ sources.

(b) All policies, $N = 6$ sources.

Fig. 9: Average AoI as a function of the maximum load in a dumbbell network with errors, $K = 5$.

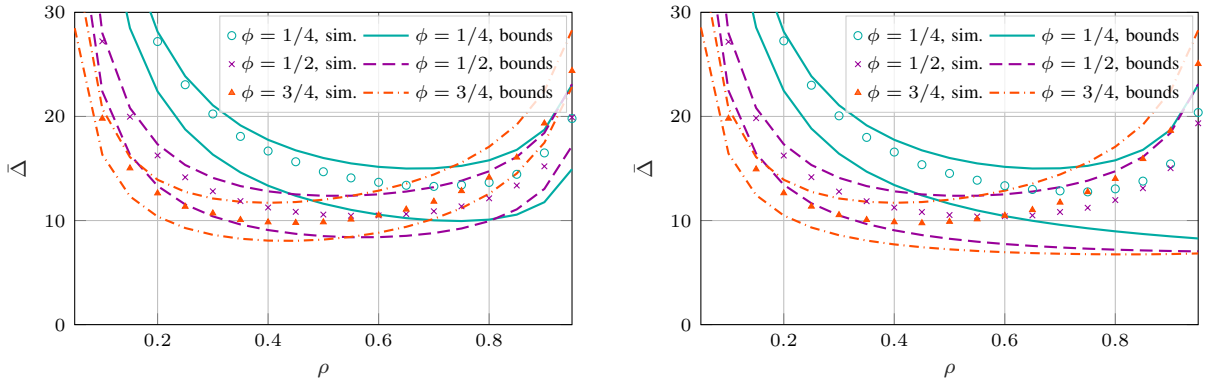
networks with more sources have a higher optimal operating load, as the traffic is split among more sources: if we set $\rho = 0.6$, $\lambda = 0.3$ when $N = 2$, but $\lambda = 0.1$ for $N = 6$. Interestingly, if there are more sources, the other links in each connection almost never experience queuing, while queuing before the bottleneck can happen for high values of ρ if $N = 2$. This leads us to draw an interesting conclusion on the optimal routing strategy: in order to minimize the average AoI, the bottleneck should be placed as early as possible in the connection, as the links before the bottleneck might also get congested, but packets exiting the bottleneck are already spread by the slower link, and there is almost no queuing for the links after the bottleneck even with a very high load. The line network example we presented above, with gradually increasing load until the bottleneck in the downlink, is the worst possible scenario for AoI.

As the flows all have similar connections, the benefits of OPF are negligible in this scenario, as Fig. 9b shows. On the other hand, MAF can slightly reduce the average AoI, particularly for higher loads. As we mentioned above, fairness is not a consideration in this case, as the symmetry of the scenario makes it perfectly fair with all three policies.

We can also introduce an element of asymmetry by changing the rate of the two sources if we consider $N = 2$: in this case, the source with the most frequent updates has an advantage. In the following, we consider the first source to have a fraction ϕ of the total load. This means that $\lambda_1 = \phi\rho\mu$ and $\lambda_2 = (1-\phi)\rho\mu$, where μ is the bottleneck rate ($\mu = 1$ in our simulations).

Fig. 10 shows the results and the bounds for three different values of ϕ for the FCFS and MAF policies. While the difference between the two policies is small, MAF has a slight advantage, but the lower bound is tighter for FCFS. We also note that the source with $\phi = 3/4$ has a lower AoI for small values of ρ, but actually a higher one if the load on the system is high. This is because, while the less active source only has queuing at the bottleneck, packets from the source generating more traffic might also be queued at the other links, as it generates enough traffic to cause delays. On the other hand, if ρ is low, the more active source can reduce interarrival times, which are the main component in the AoI.

Interestingly, the MAF policy's optimal operating load is higher, as Fig. 11 shows: as the number of sources grows, the traffic that each source generates shrinks, and the critical task of the scheduler is to choose the correct source to minimize the AoI. If the rate increases, the scheduler almost always has packets from all sources, and it can then choose the one with the highest age, optimizing the overall performance and



(a) FCFS policy.

(b) MAF policy.

Fig. 10: Average AoI as a function of the total load for different values of the relative traffic ϕ in a dumbbell network with errors for $N = 2$ and $K = 5$.

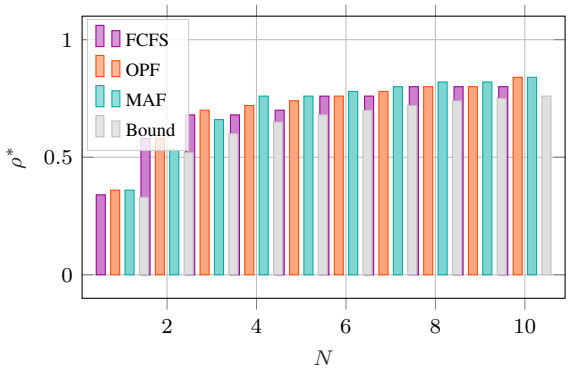


Fig. 11: Optimal ρ^* as a function of the number of sources for the three policies in a dumbbell network with errors, $K = 4$.

increasing fairness. On the other hand, FCFS can be more vulnerable to surges of packets from a subset of sources, as it does not prioritize sources with a high AoI. As for the line network, the upper bound in (32) can be used to compute ρ^*, as the figure shows. Although the value computed with the bound tends to be slightly pessimistic, sending less traffic than the optimum, the difference in the AoI between the setting obtained with the bound and the actual setting is minimal.

VI. Conclusions and Future Work

In this work, we have modeled AoI for multihop connections in queuing networks with tributary flows and non-preemptive queuing service, deriving tight upper and lower bounds. We consider fairness among different flows and several queuing policies, and present some general considerations for system design. The bounds we found are also computationally efficient, and can be used to find the optimal load of a system or to quickly dimension the design: while the resulting AoI would be a range between the two bounds, whose width is given by the value η we give in (44) and (51) for the different policies, the calculation is orders of magnitude faster than a Monte Carlo simulation. Interestingly, our results have also shown that the MAF policy only improves the average AoI by 5% over simple FCFS in both considered scenarios. However, this might partly be due to the sources having the same rate λ, and the gains might be more significant in the case of highly unbalanced systems in which some sources transmit often and others only sporadically.

Future research on the subject might involve link and source models that are not Markovian, complicating the analysis by making each link a $G/G/1$ queuing system. Another possibility is the derivation of tight bounds on the tail of the AoI distribution, which would be extremely useful for reliability purposes, as it can model worst-case scenarios for the IoT application. More advanced control mechanisms, both in terms of scheduling and packet dropping at relay nodes and of congestion control at the sources, are the final goal of all these studies, which passively characterize AoI to enable further optimization.

References

[1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *International Conference on Computer Communications (INFOCOM)*. IEEE, Mar. 2012, pp. 2731–2735.

[2] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the Internet of Things," *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, Dec. 2019.

[3] M. Handley, "Delay is not an option: Low latency routing in space," in *17th Workshop on Hot Topics in Networks*. ACM, Nov. 2018, pp. 85–91.

[4] I. Leyva-Mayorga, B. Soret, M. Röper, D. Wübben, B. Matthiesen, A. Dekorsy, and P. Popovski, "LEO small-satellite constellations for 5G and beyond-5G communications," *IEEE Access*, vol. 8, pp. 184 955–184 964, Oct. 2020.

[5] D. M. Kim, R. B. Sø rensen, K. Mahmood, O. N. Osterbo, A. Zanella, and P. Popovski, "Data aggregation and packet bundling of uplink small packets for monitoring applications in lte," *IEEE Network*, vol. 31, no. 6, pp. 32–38, 2017.

[6] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1248–1257, Jun. 2019.

[7] R. Talak and E. H. Modiano, "Age-delay tradeoffs in queueing systems," *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1743–1758, Mar. 2021.

[8] Z. Fang, J. Wang, C. Jiang, Q. Zhang, and Y. Ren, "Aoi-inspired collaborative information collection for AUV-assisted Internet of Underwater Things," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14 559–14 571, Jan. 2021.

[9] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5215–5232, Apr. 2019.

[10] R. D. Yates, "Age of information in a network of preemptive servers," in *Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, Apr. 2018, pp. 118–123.

[11] S. Farazi, A. G. Klein, and D. R. Brown, "Fundamental bounds on the age of information in multi-hop global status update networks," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 268–279, Jul. 2019.

[12] T. M. Chen, J. Walrand, and D. G. Messerschmitt, "Dynamic priority protocols for packet voice," *IEEE journal on selected areas in communications*, vol. 7, no. 5, pp. 632–643, Jun. 1989.

[13] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *International Symposium on Information Theory (ISIT)*. IEEE, Jun. 2017, pp. 576–580.

[14] J. R. Jackson, "Jobshop-like queueing systems," *Management science*, vol. 10, no. 1, pp. 131–142, Oct. 1963.

[15] B. Soret, S. Ravikanti, and P. Popovski, "Latency and timeliness in multi-hop satellite networks," in *International Conference on Communications (ICC)*. IEEE, Jun. 2020.

[16] F. Chiariotti, O. Vikhrova, B. Soret, and P. Popovski, "Information freshness of updates sent over LEO satellite multi-hop networks," *arXiv preprint arXiv:2007.05449*, Jul. 2020.

[17] K. Chen and L. Huang, "Age-of-information in the presence of error," in *International Symposium on Information Theory (ISIT)*. IEEE, Jul. 2016, pp. 2579–2583.

[18] 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*, vol. 37, no. 4, pp. 721–734, Feb. 2019.

[19] H. B. Beytur, S. Baghaee, and E. Uysal, "Measuring age of information on real-life connections," in *27th Signal Processing and Communications Applications Conference (SIU)*. IEEE, Apr. 2019.

[20] I. Kadota and E. Modiano, "Minimizing the age of information in wireless networks with stochastic arrivals," *IEEE Transactions on Mobile Computing*, Dec. 2019.

[21] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[22] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *International Symposium on Information Theory (ISIT)*. IEEE, Jun. 2015, pp. 3008–3012.

[23] R. D. Yates and S. K. Kaul, "Status updates over unreliable multiaccess channels," in *International Symposium on Information Theory (ISIT)*. IEEE, Jun. 2017, pp. 331–335.

[24] R. D. Yates and S. K. Kaul, "Age of information in uncoordinated unslotted updating," *ArXiV Preprint arXiv:2002.02026*, Feb. 2020.

[25] X. Chen, K. Gatsis, H. Hassani, and S. S. Bidokhti, "Age of information in random access channels," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, Jun. 2020, pp. 1770–1775.

[26] R. Talak, S. Karaman, and E. Modiano, "Distributed scheduling algorithms for optimizing information freshness in wireless networks," in *19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, Jun. 2018.

[27] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the Age of Information in broadcast wireless networks," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, Sep. 2016, pp. 844–851.

[28] J. Li, Y. Zhou, and H. Chen, "Age of information for multicast transmission with fixed and random deadlines in IoT systems," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8178–8191, Sep. 2020.

[29] C. Xu, H. H. Yang, X. Wang, and T. Q. Quek, "Optimizing information freshness in computing-enabled IoT networks," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 971–985, Oct. 2019.

[30] F. Chiariotti, O. Vikhrova, B. Soret, and P. Popovski, "Peak Age of Information distribution for edge computing with wireless links," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3176–3191, Jan. 2021.

[31] J. P. Champati, H. Al-Zubaidy, and J. Gross, "Statistical guarantee optimization for aoi in single-hop and two-hop fcfs systems with periodic arrivals," *IEEE Transactions on Communications*, Sep. 2020.

[32] C. Kam, J. P. Molnar, and S. Kompella, "Age of information for queues in tandem," in *Military Communications Conference (MILCOM)*. IEEE, Oct. 2018.

[33] B. Wang, S. Feng, and J. Yang, "To skip or to switch? minimizing age of information under link capacity constraint," in *19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, Jun. 2018.

[34] N. Pappas, J. Gunnarsson, L. Kratz, M. Kountouris, and V. Angelakis, "Age of information of multiple sources with queue management," in *International Conference on Communications (ICC)*. IEEE, Jun. 2015, pp. 5935–5940.

[35] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Queue management for age sensitive status updates," in *International Symposium on Information Theory (ISIT)*. IEEE, Jul. 2019, pp. 330–334.

[36] ——, "Age of information performance of multiaccess strategies with packet management," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 244–255, Jun. 2019.

[37] T. Shreedhar, S. K. Kaul, and R. D. Yates, "An age control transport protocol for delivering fresh updates in the internet-of-things," in *20th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, Jun. 2019.

[38] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *55th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, Oct. 2017.

[39] V. Tripathi, R. Talak, and E. Modiano, "Information freshness in multi-hop wireless networks," *arXiv preprint arXiv:2111.09217*, Nov. 2021.

[40] H. Bobarshad, M. van der Schaar, and M. R. Shikh-Bahaei, "A low-complexity analytical modeling for cross-layer adaptive error protection in video over WLAN," *IEEE Transactions on Multimedia*, vol. 12, no. 5, pp. 427–438, May 2010.

[41] N. T. Le, S. W. Choi, and Y. M. Jang, "Approximate queuing analysis for IEEE 802.15.4 sensor network," in *2010 Second International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2010, pp. 193–198.

[42] D. Chen, C. Yang, P. Gong, L. Chang, J. Shao, Q. Ni, A. Anpalagan, and M. Guizani, "Resource cube: Multi-virtual resource management for integrated satellite-terrestrial industrial IoT networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 963–11 974, Jul. 2020.

[43] P. J. Burke, "The output of a queueing system," *Operations Research*, vol. 4, no. 6, pp. 699–704, Dec. 1956.

[44] R. W. Wolff, "Poisson arrivals see time averages," *Operations Research*, vol. 30, no. 2, pp. 223–231, Apr. 1982.

[45] S. V. Amari and R. B. Misra, "Closed-form expressions for distribution of sum of exponential random variables," *IEEE Transactions on Reliability*, vol. 46, no. 4, pp. 519–522, Dec. 1997.

[46] F. Gillent and G. Latouche, "Semi-explicit solutions for $M/PH/1$-like queuing systems," *European Journal of Operational Research*, vol. 13, no. 2, pp. 151–160, Jun. 1983.

[47] H. Jasiulewicz and W. Kordecki, "Convolutions of Erlang and of Pascal distributions with applications to reliability," *Demonstratio Mathematica*, vol. 36, no. 1, pp. 231–238, Jan. 2003.

[48] Q. Kuang, J. Gong, X. Chen, and X. Ma, "Age-of-information for computation-intensive messages in mobile edge computing," in *11th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, Oct. 2019.

[49] J. Doncel and M. Assaad, "Age of information in a decentralized network of parallel queues with routing and packets losses," *Journal of Communications and Networks*, vol. 24, no. 1, pp. 17–20, Feb. 2022.

[50] M. Lerch, "Note sur la fonction $\mathfrak{K}(w,x,s) = \sum_{k=0}^{\infty} \frac{e^{2k\pi ix}}{(w+k)^s}$," *Acta Mathematica*, vol. 11, pp. 19–24, Jan. 1887.

[51] D. H. Bailey and J. M. Borwein, "Crandall's computation of the incomplete Gamma function and the Hurwitz zeta function, with applications to Dirichlet L-series," *Applied Mathematics and Computation*, vol. 268, pp. 462–477, Oct. 2015.

Federico Chiariotti (S'15–M'19) is currently an assistant professor at the Department of Electronic Systems, Aalborg University, Denmark. He received his Ph.D. in information engineering in 2019 from the University of Padova, Italy. He received the bachelor's and master's degrees in telecommunication engineering from the University of Padova in 2013 and 2015, respectively. He has authored over 50 published papers on wireless networks and the use of artificial intelligence techniques to improve their performance. He was a recipient of the Best Paper Award at several conferences, including the IEEE INFOCOM 2020 WCNEE Workshop. His current research interests include network applications of machine learning, transport layer protocols, Age of Information, bike sharing system optimization, and adaptive video streaming.

Olga Vikhrova (M'21) is a postdoctoral researcher at Tampere University, Finland. She is currently a holder of the Academy of Finland Postdoctoral fellowship. She received her Ph.D. (2021) in Information Engineering from University Mediterranea of Reggio Calabria, Italy, and M.Sc. (2014) in Information and Computer Science from Peoples' Friendship University of Russia (RUDN University), Russia. Her current research interests include analysis of the wireless multi-hop networks, applied mathematical modeling, and machine learning.

Beatriz Soret (M'11) received her M.Sc. and Ph.D. degrees in Telecommunications from the University of Malaga, Spain, in 2002 and 2010, respectively. She is currently an associate professor at the Department of Electronic Systems, Aalborg University, and a Senior Research Fellow at the Communications Engineering Department, University of Malaga. Her research interests include LEO satellite communications, AoI and semantic communications, and 5G and 6G systems.

Petar Popovski (S'97–A'98–M'04–SM'10–F'16) is a Professor at Aalborg University, where he heads the section on Connectivity and a Visiting Excellence Chair at the University of Bremen. He received his Dipl.-Ing and M. Sc. degrees in communication engineering from the University of Sts. Cyril and Methodius in Skopje and the Ph.D. degree from Aalborg University in 2005. He is a Fellow of the IEEE. He received an ERC Consolidator Grant (2015), the Danish Elite Researcher award (2016), IEEE Fred W. Ellersick prize (2016), IEEE Stephen O. Rice prize (2018), Technical Achievement Award from the IEEE Technical Committee on Smart Grid Communications (2019), the Danish Telecommunication Prize (2020) and Villum Investigator Grant (2021). He is a Member at Large at the Board of Governors in IEEE Communication Society, Vice-Chair of the IEEE Communication Theory Technical Committee and IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is currently an Editor-in-Chief of IEEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. Prof. Popovski was the General Chair for IEEE SmartGridComm 2018 and IEEE Communication Theory Workshop 2019. His research interests are in the area of wireless communication and communication theory. He authored the book "Wireless Connectivity: An Intuitive and Fundamental Guide", published by Wiley in 2020.