Age-Delay Tradeoffs in Queueing Systems

Rajat Talak and Eytan Modiano

***Abstract*—We consider an m server system in which each server can service at most one update packet at a time. The system designer controls (1) *scheduling* - the order in which the packets get serviced, (2) *routing* - the server that an arriving update packet joins for service, and (3) the *service time distribution* with fixed service rate. Given a fixed update generation process, we prove a strong age-delay and age-delay variance tradeoff, wherein, as the average AoI approaches its minimum, the packet delay and its variance approach infinity. In order to prove this result, we consider two special cases of the m server system, namely, a single server system with last come first served with preemptive service and an infinite server system. In both these cases, we derive sufficient conditions to show that three heavy tailed service time distributions, namely Pareto, log-normal, and Weibull, asymptotically minimize the average AoI as their tail gets heavier, and establish the age-delay tradeoff results. We provide an intuitive explanation as to why such a seemingly counter intuitive age-delay tradeoff is natural, and that it should exist in many systems.**

***Index Terms*—Age of Information (AoI), Queueing Systems, Packet Delay, Age-Delay Tradeoff.**

I. Introduction

Information freshness and low latency communication is gaining increasing relevance in many futuristic communication systems, such as industrial automation, autonomous driving, tele-surgery, financial markets, and virtual reality [4]–[7]. The latency requirements vary depending on the application. While applications such as autonomous driving, tele-surgery, virtual reality, financial markets are envisioned to require a latency of a few milliseconds, other systems such as industry automation, control signalling in power grids aim at a latency of 10-100 milliseconds [4], [5]. In many of these applications, seeking the most recent status update is crucial to the overall system performance. In a network of unmanned aerial vehicles, for example, exchanging the most recent position, speed, and other control information [8], [9]; in operations monitoring systems, accessing the most recent sensor measurement; and in cellular systems, obtaining the timely channel state information from the mobile users [10], is important and can lead to significant performance improvements.

Age of information (AoI) is a metric for information freshness that measures the time that elapses since the last received fresh update was generated at the source [11], [12]. It is, therefore, a destination-centric measure, and is suitable for applications that necessitate timely updates. A typical evolution of AoI for a single source-destination system is shown in Figure 1. The AoI increases linearly in time, until the destination receives a fresh update packet. Upon reception of a fresh update packet i, at time $t_i^{'}$, the AoI drops to the time since the packet i was generated, which is $t_i^{'} - t_i$; here t_i is the time of generation of the update packet i. Unlike the traditional latency metrics such as packet delay, AoI only accounts for the update packets that deliver fresh updates to the destination – such packets are called *informative packets* [13]. For example, in Figure 1, packet 3 is an informative packet but packet 2 is not. This is because packet 3 reaches the destination before packet 2, which is therefore rendered stale by time $t_2^{'}$.

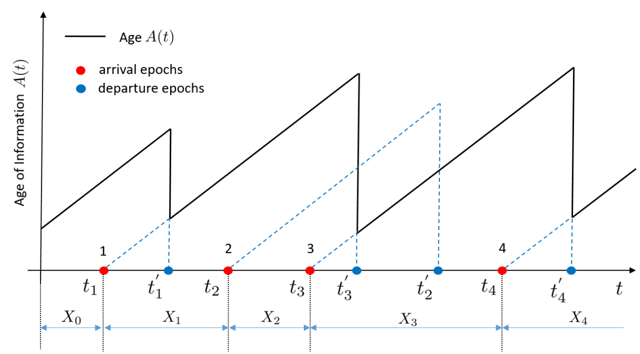

Fig. 1. Age evolution in time. Update packets generated at times t_i and received, by the destination, at times $t_i^{'}$. Packet 2 is received out of order, and thus, doesn't contribute to age.

AoI was first studied for the first come first serve (FCFS) M/M/1, M/D/1, and D/M/1 queues in [12]. Since then, AoI has been analyzed for several queueing systems with the goal to minimize AoI [11]–[34]. Two time average metrics of AoI, namely, peak and average age are generally considered. The analysis mostly relies on the specificities of the queueing model under consideration. Typically, a peak or average age expression is derived and then optimized over the update generation rate. However, progress has been made recently towards a more general analysis of AoI. A general formula for the stationary distribution of AoI for a single-server systems was recently developed in [27], [35], while [36] used the theory of stochastic hybrid systems to systematically derive expressions for the average age and its higher moments.

Despite the difficulty in analyzing age for general queueing systems several approaches that reduce or minimize AoI have been brought to light. The advantage of having parallel servers in reducing AoI was demonstrated in [13], [26], [29]. Methods such as limiting the buffer sizes [11], [33] or introducing packet deadlines [31]–[33] have also been shown to reduce

The manuscript has been accepted, on November 30, 2020, for publication in the IEEE Transactions on Information Theory. This work was supported by NSF Grants AST-1547331, CNS-1713725, and CNS-1701964, and by Army Research Office (ARO) grant number W911NF- 17-1-0508.

A preliminary version of the this work was available on arXiv [1] and appeared in ISIT 2019 [2], [3].

The authors are with the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT), Cambridge, MA. {talak, modiano}@mit.edu

AoI. In a general queueing system, with exponentially distributed service times, the last come first serve (LCFS) queue scheduling discipline with preemptive service was proved to be age optimal in [22]–[25]. In [37], an optimal update generation policy was investigated, and it was discovered that an intuitively apt zero-wait policy, which sends the next update right after the previous one is received, is not always age optimal.

More recently, minimizing age metrics over update generation and service time distribution has been of interest [1], [35]. Service time variations in real systems can be caused by many factors such as channel fading, sudden drops in signal-to-noise ratio, packet collisions due to contention for a link layer packet. For a transport layer packet, on the other hand, the service time variability is affected by the underlying routing mechanisms, packet management, packet fragmentation used, ARQ protocols used by all the underlying links, exogenous traffic etc. Service time variability may also be attributed to the fact that the server is busy serving other high priority packets. It is hard to pin down how each of these factors, in conjunction, affect the service time of a generated update. It is therefore necessary to fully understand how age and delay behave in variation to the service time distribution, and how the service time distribution affects age and delay. When is age and delay minimized, and when is it at its worst? are important questions to answer. In a related work [1], we analyzed age and delay over the update generation and service time distribution variations, and showed that determinacy in packet generation and/or service does not necessarily lead to minimum age. Similar results were independently obtained in [35].

Packet delay and delay variance (jitter), on the other hand, have traditionally been considered as measures of communication latency. Minimizing packet delay in a queueing system is known to be a hard problem. Several works have focused on the problem of reducing or minimizing the packet delay and its variance [38]–[53]. It is widely believed that AoI and delay are closely related, and hence, can be minimized simultaneously. For example, in a simple FCFS queue under Poisson arrivals, less variability in service time distribution minimizes both packet delay and peak age [1], [35]. For a system of m parallel servers with exponential service times, minimum age and delay can be simultaneously attained by resorting to the LCFS with preemptive service [22], [23].

Is it then always possible to minimize age and packet delay simultaneously, or are there systems in which minimizing one does not imply minimizing the other?

In this work, we answer this question by considering an m server queueing system. We show that as we tailor the queue scheduling discipline, routing, and the service time distributions to minimize average age, the packet delay and its variance approach infinity. As an example, consider a single server queue with a fixed update generation and service rate. Updates are generated according to a Poisson process. In Figure 2, we plot the achieved packet delay and average age attained under various queue scheduling policies (FCFS and LCFS with preemptive service[1]) and service time distributions (deterministic, exponential, log-normal, Pareto, and Weibull). For the three heavy tailed service time distributions, namely, log-normal, Pareto, and Weibull we plot the age-delay points for various values of the free parameter; see Table I. It appears that there is a strong age-delay tradeoff, i.e. a lower average age can be achieved at the cost of higher delay.

Intuitively, this tradeoff can be explained as follows: *In order to achieve minimum age the system has to prioritize informative packets. In doing so, the non-informative updates lag behind in the system thereby incurring a large waiting time cost. The delay and its variance get dominated by the large delays incurred by the non-informative update packets, thus leading to the tradeoff curve in Figure 2.*

Given this intuition, we suspect that an age-delay tradeoff should exist in many systems. In this paper, we prove it for an m server queueing system.

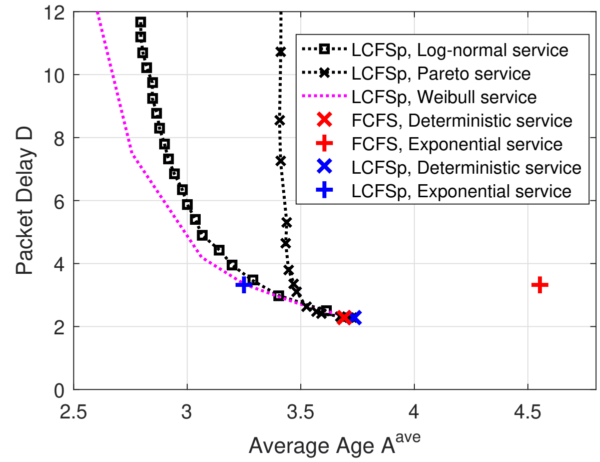

Fig. 2. Plot of achieved age-delay points for various single server systems, Poisson packet generation at rate $\lambda = 0.5$, and service at rate $\mu = 0.8$. Scheduling disciplines: FCFS, LCFS with preemptive service. Service time distributions: Deterministic, Exponential, and Heavy Tailed distributions in Table I.

TABLE I
HEAVY TAILED SERVICE TIME DISTRIBUTIONS WITH MEAN $\mathbb{E}[S] = 1/\mu$.

Name	Distribution	Free Parameter
Log-normal	$S = \exp\left(-\log\mu - \frac{\sigma^2}{2} + \sigma N\right)$ $N \sim \mathcal{N}(0,1)$	$\sigma > 0$
Pareto	$F_S(x) = 1 - (\theta(\alpha)/x)^{\alpha}\,\mathbb{I}_{\{x>\theta(\alpha)\}}$ $\theta(\alpha) = (\alpha-1)/(\mu\alpha)$	$\alpha > 1$
Weibull	$\mathbf{P}[S > x] = \exp\left\{-(x/\beta(k))^k\right\}$ $\beta(k) = [\mu\Gamma(1+1/k)]^{-1}$	$k > 0$

[1]In Figure 2, for the LCFS with preemptive service, we assume that a preempted packet only has to serve its residual service time to complete its service, and not begin the service again when the queue empties. This assumption does not affect the average age computation. However, if the system is such that a preempted packet has to re-start its service, the plotted delay is only an underestimate.

A. Contributions

We consider an m server system in which each server can service at most one update packet at any given time. Update packets are generated according to a renewal process at a fixed rate. The system is not allowed to drop packets and has an infinite buffer. The system designer decides the queue *scheduling discipline*, i.e. the order in which the packets get serviced, the *routing*, which determines the server for each arriving update packet, and the *service time distribution*.

In order to observe the age-delay tradeoff, we consider the problem of minimizing packet delay (and packet delay variance), subject to an average age constraint, over the space of all queue scheduling disciplines, routing, and service time distributions, with a fixed mean service time budget of $1/\mu$ for each queue. When the updates are generated according to a Poisson process we show that there is a strong age-delay tradeoff, namely, as the average age approaches its minimum, the delay approaches infinity. When the updates are generated according to a general renewal process, we show that there is a strong age-delay variance tradeoff, i.e. as the average age approaches its minimum, the variance in packet delay approaches infinity.

The proof of this result involves first proving the same result in two special cases of the m server system: (1) A single server system, i.e. $m = 1$, in which the queue scheduling discipline is fixed to LCFSp, and (2) An infinite server system, i.e. $m = \infty$. In both these cases, we derive sufficient conditions on the average age minimizing service time distribution. We then show that these sufficient conditions are satisfied by the three heavy tailed service time distributions, namely Pareto, log-normal, and Weibull, asymptotically in its tail parameter. This helps us establish the age-delay tradeoff results in the two special cases of the m server system. We also observe a certain age-delay disparity in these two cases in which the delay (or delay variance) minimizing service time distributions result in the worst case average age.

The results derived for the two special cases are then used to prove the strong age-delay and age-delay variance tradeoffs for the m server system. To the best of our knowledge, this is the first work to establish an age-delay tradeoff result. A preliminary version of the this work was available on arXiv [1] and appeared in ISIT 2019 [2], [3]. This work builds upon the results in [1]–[3].

B. Organization

In Section II, we describe the system model and provide a general definition of AoI. In Section III, we formulate the age-delay and age-delay variance problems for the m server system. The age-delay tradeoff result for the m server system is also stated and discussed here. In Sections IV and V, we prove the age-delay tradeoff result in the two special cases of the single server LCFSp and infinite server systems. The paper culminates in Section VI with a proof of the age-delay tradeoff for the m server system. We conclude in Section VII.

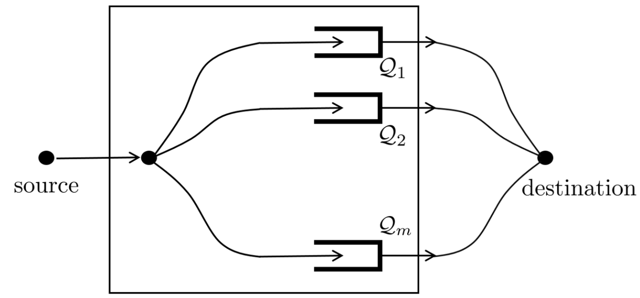

Fig. 3. Illustration of the m server queueing system.

II. System Model

A source generates update packets according to a renewal process, at a given rate λ. The update packets enter a queueing system, which consists of m servers shown in Figure 3. Each server has a rate μ, and can service at most one update packet at any given time. The service times are independent and identically distributed across update packets. A scheduler determines routing and scheduling of update packets, which upon service reach the destination. Our goal is to ensure minimum age and/or minimum delay at the destination.

The system designer has control over three things:

1) *service*: it can decide the service time distribution, given the mean service time of $1/\mu$;
2) *routing*: it can determine the server that an update packet connects for service; and
3) *scheduling*: it can decide the order in which the update packets get serviced at each server.

A *scheduler* implements the routing and scheduling policy.The scheduler is also not allowed to drop any packets, and has access to an infinite buffer. The scheduler can preempt packets, i.e., it can pause a packet under service to serve a newly arrived packet. In the case of a preemption, the preempted packet only has to compete its residual service left, after it is put back in service.[2] Further, in determining the order of service of generated packets, we assume that the scheduler is not privy to the service times of the individual packets. We also assume that only the service time distribution can be set before hand by the system designer, and not the service times of individual packets.

We use X to denote the inter-generation time of update packets with distribution F_X, and S to denote the service time random variable, with distribution F_S. Note that $\mathbb{E}[X] = 1/\lambda$ and $\mathbb{E}[S] = 1/\mu$ is fixed. We assume that $\lambda < \mu$, i.e. there is enough serving capacity in the network to service the generated updates.We use Minimize or $\min$, instead of the technically correct $\inf$, for ease of presentation. We now define the latency measures of packet delay, delay variance, and average age.

[2]The modeling assumption here that a preempted packet only has to complete its residual service left, after it is put back in service, does not affect the age metric. It only underestimates the packet delay in situations where a preempted packet may have some additional incurred service time.

A. Delay and Age of Information

Let the update packets be generated at times $t_1, t_2, \ldots$, and let the update packet i reach the destination at time $t_i^{'}$. The update packets may not reach the destination in the same order as they were generated. In Figure 1, packet 3 reaches the destination before packet 2, i.e. $t_3^{'} < t_2^{'}$. Delay for the ith packet is $D_i = t_i^{'} - t_i$, and the packet delay for the system is given by

$$D = \limsup_{N \to \infty} \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} D_i\right], \tag{1}$$

where the expectation is over the update generation, service times, and scheduling discipline. We skip a formal definition, but will use the notation VarD to denote variance in packet delay. For the m server queueing system considered, we note that VarD is lower-bounded by the variance in service time distribution F_S.

Age of a packet i is defined as the time since it was generated: $A^i(t) = (t - t_i)\mathbb{I}_{\{t > t_i\}}$, which is 0 by definition for time prior to its generation $t < t_i$. Age of information at the destination node, at time t, is defined as the minimum age across all received packets up to time t:

$$A(t) = \min_{i \in \mathcal{P}(t)} A^i(t), \tag{2}$$

where $\mathcal{P}(t) \subset \{1, 2, 3, \ldots\}$ denotes the set of packets received by the destination, up to time t. Notice that AoI increases linearly, and drops only at the times of certain packet receptions: $t_1^{'}, t_3^{'}, t_4^{'}, \ldots$, but not $t_2^{'}$ in Figure 1. Such an age drop occurs only when an update packet with a lower age, than all packets received thus far, is received by the destination. We refer to such packets, that result in age drops, as the *informative packets* [13].

We consider a time averaged metrics of age of information, namely, the average age. The average age is defined to be the time averaged area under the age curve:

$$A^{\text{ave}} = \limsup_{T \to \infty} \mathbb{E}\left[\frac{1}{T}\int_0^T A(t)dt\right], \tag{3}$$

where the expectation is over the packet generation and packet service processes.

It is important to note that the age $A(t)$, and therefore the average age, is defined from the view of the destination, and not a packet. $A(t)$ is the time since the last received informative packet was generated at the source. It, therefore, does not matter how long the non-informative packets take to reach the destination. This is unlike packet delay, which accounts for every packet in the system equally.

We use the notation $D(F_S, \pi_Q)$, $\text{VarD}(F_S, \pi_Q)$, and $A^{\text{ave}}(F_S, \pi_Q)$ to make explicit the dependency of delay, its variance, and average age on the service time distribution F_S and the queue scheduling policy π_Q.

In the next section, we pose the age-delay tradeoff problems. Age-delay tradeoff results are then proved for two special cases in Section IV and Section V, before arriving at the result for the m server system in Section VI.

III. Age-Delay Tradeoff

Motivated by the example in Figure 2, we define two age-delay tradeoff problems. One, minimizes delay while the other minimizes delay variance, both over an average age constraint. The age-delay tradeoff is defined as:

$$\begin{aligned} T(\text{AoI}) = \underset{F_S, \pi_Q}{\text{Minimize}} \quad & D(F_S, \pi_Q) \\ \text{subject to} \quad & A^{\text{ave}}(F_S, \pi_Q) \leq \text{AoI}, \\ & \mathbb{E}[S] = 1/\mu. \end{aligned} \tag{4}$$

Here, the function $T(\text{AoI})$ denotes the minimum delay that can be achieved for the m server queueing system, with an average age constraint of $A^{\text{ave}}(F_S, \pi_Q) \leq \text{AoI}$. It might seem that both minimum age and delay could be attained simultaneously. We will show that, $T(\text{AoI}) \to \infty$ as AoI approaches the minimum average age

$$\begin{aligned} A_{\min} = \underset{F_S, \pi_Q}{\text{Minimize}} \quad & A^{\text{ave}}(F_S, \pi_Q), \\ \text{subject to} \quad & \mathbb{E}[S] = 1/\mu. \end{aligned} \tag{5}$$

Variability in packet delay is also an important metric in system performance. We define the age-delay variance tradeoff problem to be:

$$\begin{aligned} V(\text{AoI}) = \underset{F_S, \pi_Q}{\text{Minimize}} \quad & \text{VarD}(F_S, \pi_Q) \\ \text{subject to} \quad & A^{\text{ave}}(F_S, \pi_Q) \leq \text{AoI}, \\ & \mathbb{E}[S] = 1/\mu. \end{aligned} \tag{6}$$

Here, the function $V(\text{AoI})$ denotes the minimum delay variance that can be achieved for the m server queueing system, with an average age constraint of $A^{\text{ave}}(F_S, \pi_Q) \leq \text{AoI}$. Counter to our intuition, we show that $V(\text{AoI}) \to +\infty$ as AoI approaches its minimum value $A_{\min}$.

Ideally, we would like to obtain every point on the tradeoff curves, i.e., completely characterize the functions $T(\text{AoI})$ and $V(\text{AoI})$. The following result shows that the tradeoff curves can be done by optimizing a linear combination of average age and packet delay.

Theorem *1:* The points on the age-delay tradeoff curve $(\text{AoI}, T(\text{AoI}))$ can be obtained by solving

$$\begin{aligned} \underset{F_S, \pi_Q}{\text{Minimize}} \quad & D(F_S, \pi_Q) + \nu A^{\text{ave}}(F_S, \pi_Q) \\ \text{subject to} \quad & \mathbb{E}[S] = 1/\mu, \end{aligned} \tag{7}$$

for all $\nu \geq 0$. Similarly, the points on the age-delay variance tradeoff curve $(\text{AoI}, V(\text{AoI}))$ can be obtained by solving (7), by replacing $D(F_S, \pi_Q)$ with $\text{VarD}(F_S, \pi_Q)$.

Proof: This follows from Theorem II.2 in [54]. ∎

Theorem 1 motivates optimization of a latency metric that is a linear combination of average age and packet delay (or packet delay variance). This problem, however, is not easy to solve. For instance, in the case of a single server, i.e. $m = 1$, with Poisson arrivals, the delay is minimized with deterministic service times and the variance in delay is

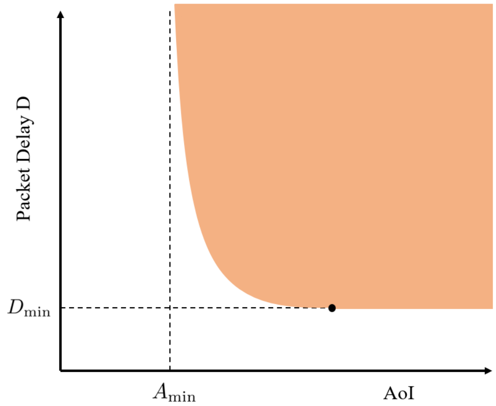

Fig. 4. Illustration of strong age-delay tradeoff.

minimized under the FCFS service discipline [38]. Exactly the opposite is true for the metric of average age. We will show in Section IV that the LCFS queue scheduling policy with heavy tailed service minimizes average age. It, therefore, appears that the delay term and the average age term in (7) are pulling the decision variables in opposite directions.

We say that a *strong age-delay tradeoff* exists for $T(\text{AoI})$ if $T(\text{AoI}) \to +\infty$ and AoI approaches $A_{\min}$. Conversely, *no age-delay tradeoff* exists for $T(\text{AoI})$ if the minimum average age and the minimum packet delay can be achieved simultaneously. Similar definition apply for age-delay variance tradeoff $V(\text{AoI})$. Figure 4 illustrates a strong age-delay tradeoff. Note that this matches with our numerical results in Figure 2.

We show that for the m server system defined above there is a strong age-delay tradeoff when the update generation is Poisson, and a strong age-delay variance tradeoff for the general update generation process.

Theorem *2:* For the m server system, the following statements are true:

1) The minimum achievable average age is $A_{\min} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$.
2) When update generation is Poisson, there is a strong age-delay tradeoff.
3) When the update generation is a general renewal process, there is a strong age-delay variance tradeoff.

In Theorem 2, the update generation process is kept fixed. Thus, the strong age-delay tradeoffs hold even if we could control the inter-generation time distribution F_X, with a mean budget constraint of $\mathbb{E}[X] = 1/\lambda$. The optimal update generation, with the mean constraint $\mathbb{E}[X] = 1/\lambda$, that minimizes the $A_{\min}$ is the periodic update generation.

It seems counterintuitive at first that a strong tradeoff should exist between delay, or delay variance, and average age. However, a close examination reveals the following insight:

For age minimization it becomes necessary that the informative packets, the packets that reduce age, get serviced as soon as they arrive, while the non-informative packets, may incur as long a service time and queueing delay, as they do not matter in the age calculations. As we do this, the packet delay gets dominated by the delay of the non-informative packets, resulting in the two age-delay tradeoffs.

In what follows, we prove this strong tradeoff between age-delay and age-delay variance. The proof of Theorem 2 is given in Section VI. It relies on age-delay tradeoff results in two special cases, which are first studied in Sections IV and V.

In Section IV, we consider the special case of a single server system, i.e. $m = 1$, in which the scheduling policy π_Q is fixed to the last come first server with preemptive service (LCFSp). We prove the statements of Theorem 2 for this special case. Namely, we show that a age-delay tradeoff exists when the update generation is Poisson, and a age-delay variance tradeoff exists when the updates are generated according to a general renewal process.

In Section V, we consider the special case of an infinite server queue, i.e. $m = \infty$, in which the scheduling policy π_Q assigns every newly generated update to a new server. We again prove the statements of Theorem 2 for this case.

Our approach in both these sections is as follows: We first derive an expression for the average age, and use it to obtain the minimum average age $A_{\min}$. We then use this to prove the two strong age-delay tradeoffs. Using these two special cases, in Section VI, we finally prove Theorem 2.

IV. LCFSp Queue

In this section, we consider a special case of the m server system. We consider a single server system, i.e. $m = 1$, in which the queue scheduling discipline π_Q is fixed to the LCFSp. The age and delay metrics, therefore, are just a function of the service time distribution F_S. The age-delay and the age-delay variance problem, for this case, reduces to

$$\begin{aligned} T(\text{AoI}) = \underset{F_S}{\text{Minimize}} \quad & D(F_S) \\ \text{subject to} \quad & A^{\text{ave}}_{\text{G/G/1}}(F_S) \leq \text{AoI}, \\ & \mathbb{E}[S] = 1/\mu, \end{aligned} \tag{8}$$

and

$$\begin{aligned} V(\text{AoI}) = \underset{F_S}{\text{Minimize}} \quad & \text{VarD}(F_S) \\ \text{subject to} \quad & A^{\text{ave}}_{\text{G/G/1}}(F_S) \leq \text{AoI}, \\ & \mathbb{E}[S] = 1/\mu, \end{aligned} \tag{9}$$

where we use the notation $A^{\text{ave}}_{\text{G/G/1}}(F_S)$ to denote the average age for the LCFSp queue. The optimization is only over the service time distribution. We omit the dependence on F_S, whenever convenient.

The rest of this section is organized as follows. In Section IV-A, we derive an expression for average age, and characterize the minimum average age. We also show that heavy tailed service time distributions achieve the minimum average age. In Section IV-B, we then use these results to prove that there is a strong age-delay and age-delay variance tradeoff.

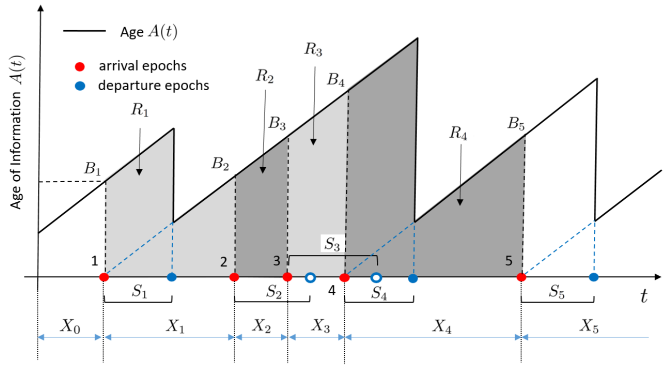

Fig. 5. Age $A(t)$ evolution in time t for the LCFS queue with preemptive service.

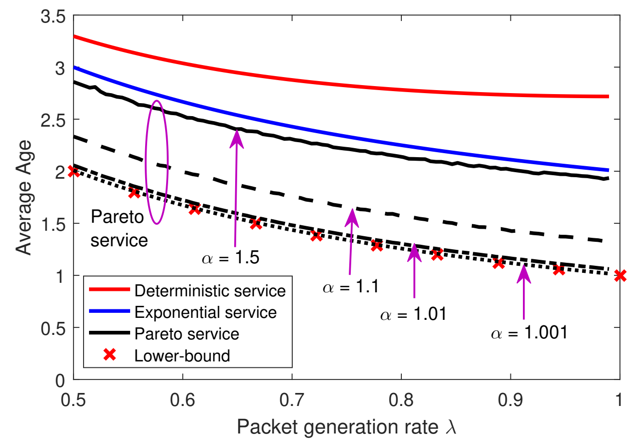

Fig. 6. Plotted is the average age under deterministic, exponential, and Pareto ($\alpha = 1.5, 1.1, 1.01$, and 1.001) distributed service times distributions for the LCFS queue with preemptive service. Service rate $\mu = 1$, while the packet generation rate λ varies from 0.5 to 0.99.

In Section IV-C, we prove an age-delay disparity when the update generation is Poisson.

A. Minimizing Age

We first derive explicit expression average age for general inter-generation and service time distributions. We assume at least one of the distributions F_X and F_S to be continuous.

Lemma *1:* The average age $A^{\text{ave}}_{\text{G/G/1}}(F_S)$ is given by

$$A^{\text{ave}}_{\text{G/G/1}}(F_S) = \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \frac{\mathbb{E}\left[\min\left(X, S\right)\right]}{\mathbf{P}\left[S < X\right]},$$

where X and S denotes the independent inter-generation and service time distributed random variables, respectively.

Proof: Let X_i denote the inter-generation time between the ith and $(i+1)$th update packet. Due to preemption, not all packets get serviced on time to contribute to age reduction. We illustrate this in Figure 5. Observe that packets 2 and 3 arrive before packet 4. However, packet 2 is preempted by packet 3, which is subsequently preempted by packet 4. Thus, packet 4 is serviced before 2 and 3. Service of packet 2 and 3 (not shown in figure) does not contribute to age curve $A(t)$ because they contain stale information.

In order to analyze this, define S_i to be the virtual service time for packet i, such that $\{S_i\}_{i\geq 1}$ are i.i.d., and distributed according to the service time distribution F_S. If $S_i < X_i$, then the packet i is serviced, and the age $A(t)$ drops to S_i, which is the time since generation of the packet i. In Figure 5, we observe this for packets 1, 4 and 5. However, if $S_i > X_i$, the service of packet i is preempted, and the server starts serving the newly arrived packet $(i+1)$. In Figure 5, observe that $S_2 > X_2$ and $S_3 > X_3$, while $S_4 < X_4$, and thus, packet 4 gets serviced before 2 and 3.

For computing the average age, which is nothing but the time averaged area under the age curve $A(t)$, we compute the sum $\sum_{i=1}^{m} R_i$, where R_i is the area under $A(t)$ between the ith and $(i+1)$th generation of update packets; see Figure 5. To do so, we obtain a recursion for B_i, the age $A(t)$ at the time of generation of the ith update packet: define $Z_i \triangleq \sum_{k=0}^{i-1} X_k$ and $B_i = A(Z_i)$, and show that

$$R_i = \begin{cases} B_i X_i + \frac{1}{2}X_i^2 & \text{if } X_i < S_i \\ B_i S_i + \frac{1}{2}X_i^2 & \text{if } X_i \geq S_i \end{cases}. \tag{10}$$

The detailed proof is given in Appendix A. ■

We now prove that a heavy tailed continuous service time distribution minimizes the average age. In Figure 6, we plot average age as a function of packet generation rates λ, for three different service time distributions: deterministic service, exponential service, and Pareto service. The cumulative distribution function for a Pareto service distribution, with mean $1/\mu$, is given by

$$F_S(s) = \begin{cases} 1 - \left(\frac{\theta(\alpha)}{s}\right)^{\alpha} & \text{if } s \geq \theta(\alpha) \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

where $\theta(\alpha) = \frac{1}{\mu}\left(1 - \frac{1}{\alpha}\right)$ and $\alpha > 1$ is the shape parameter. The shape parameter α determines the tail of the distribution. The closer the shape parameter is to 1, the heavier is the tail.

We observe in Figure 6 that the Pareto service yields better age than the exponential service. Furthermore, observe that the heavier the tail of the Pareto distribution, i.e. the closer α is to 1, the lower is the age. Also plotted is the age lower-bound $1/\lambda$, as no matter what the service, the age cannot decrease below the inverse rate at which packets are generated.

We observe similar behavior not just for Pareto distributed service, but also for other heavy tailed distributions. In Figure 7, we plot average age for log-normal service distribution, another heavy-tail distribution, with mean $1/\mu$ given by:

$$S = \exp\left\{-\log\mu - \frac{\sigma^2}{2} + \sigma N\right\}, \tag{12}$$

where $N \sim \mathcal{N}(0, 1)$ is the standard normal distribution and σ is a parameter that determines the tail of the distribution F_S.

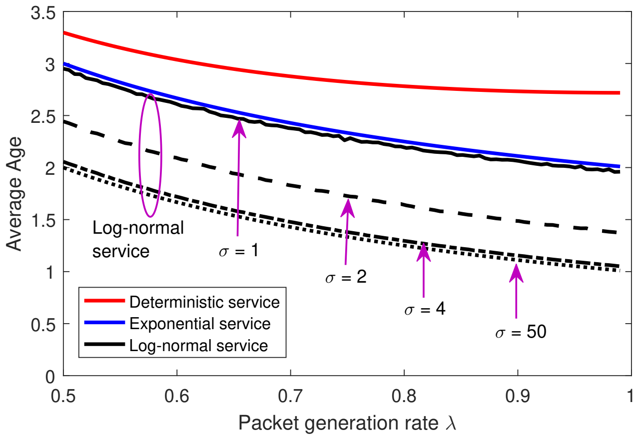

Fig. 7. Plotted is the average age under deterministic, exponential, and log-normal ($\sigma = 1, 2, 4,$ and 50) distributed service times distributions for the LCFS queue with preemptive service. Service rate $\mu = 1$, while the packet generation rate λ varies from 0.5 to 0.99.

Higher σ implies heavier tail, and in Figure 7 we observe that it results in smaller age, that approaches the age lower-bound of $1/\lambda$ as $\sigma \to +\infty$. We observe similar behavior for Weibull distributed service, with mean $1/\mu$:

$$F_S(s) = 1 - e^{-(s/\beta)^{\kappa}}, \tag{13}$$

for all $s \geq 0$, where $\beta = [\mu\Gamma(1 + 1/\kappa)]^{-1}$, as $\kappa \downarrow 0$; here $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ is the gamma function.

We now show that the minimum average age is given by $A_{\text{min}} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$. We first prove that the average age is lower-bounded by $\frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ for any service time distribution, and then show that this lower-bound is in fact achievable for the three heavy tailed service time distributions.

Theorem *3:* The average age is lower bounded by

$$A^{\text{ave}}_{\text{G/G/1}}(F_S) \geq \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]}.$$

Further, this lower-bound is achieved for

1) Pareto distributed service (11) as $\alpha \to 1$,
2) Log-normal distributed service (12) as $\sigma \to +\infty$, and
3) Weibull distributed service (13) as $\kappa \to 0$.

Proof: The lower-bound on the average age follows directly from the age expressions obtained in Lemma 1, and noticing that $\frac{\mathbb{E}[\min\{X,S\}]}{\mathbf{P}[S<X]} \geq 0$. The distributions, namely the Pareto, log-normal, and Weibull, are all parametric distributions parameterized here by α, σ, and κ, respectively. We, therefore, prove the following generic result, which gives us a sufficient conditions for the optimality of the average age for a general, parametric continuous service time distribution F_S, parameterized by η. We hide the dependence of the parameter η on S and F_S for notational convenience.

Lemma *2:* Let a parametric, continuous service time S, with parameter η, satisfy

1) $\mathbb{E}\left[S\right] = 1/\mu$ for all η,
2) $\mathbb{E}\left[\mathbb{I}_{\{S>x\}}\right] \to 0$ as $\eta \to \eta^*$, and
3) $\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] \to 0$ as $\eta \to \eta^*$,

for all $x > 0$ and for some η^*. Then,

$$\lim_{\eta\to\eta^*} A^{\text{ave}}_{\text{G/G/1}}(F_S) = \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]}. \tag{14}$$

Proof: Let for a parametric, continuous service time distribution F_S the stated properties hold. Note that

$$\mathbb{E}\left[\min\{X,S\}\right] = \mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] + \mathbb{E}\left[S\mathbb{I}_{\{S\leq X\}}\right],$$

Using conditions 2 and 3 in the Lemma, and the dominated convergence theorem [55], we have $\mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] \to 0$, $\mathbb{E}\left[S\mathbb{I}_{\{S\leq X\}}\right] \to 0$, and $\mathbf{P}\left[S < X\right] \to 1$ as $\eta \to \eta^*$. In order to illustrate how this is proved, we show the steps to prove $\mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] \to 0$ using condition 2. The remaining claims follow the same line of argument. Let $Y_\eta = \mathbb{E}\left[X\mathbb{I}_{\{S>X\}}|X\right]$. Here, Y_η is a function of the parameter η only via S, as X does not depend on η. By the towering property of conditional expectations we have $\mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] = \mathbb{E}\left[Y_\eta\right]$. Further, we have $Y_\eta \leq X$ for all η and, by condition 2, $Y_\eta \to 0$ as $\eta \to \eta^*$. Since X is integrable, i.e., it has finite expectation, we can apply the dominated convergence theorem [55], which yields

$$\mathbb{E}\left[Y_\eta\right] = \mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] \to 0. \tag{15}$$

Substituting $\mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] \to 0$, $\mathbb{E}\left[S\mathbb{I}_{\{S\leq X\}}\right] \to 0$, and $\mathbf{P}\left[S < X\right] \to 1$ as $\eta \to \eta^*$ in the average age expression in Lemma 1, we obtain $A^{\text{ave}}_{\text{G/G/1}}(F_S) \to \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ as $\eta \to \eta^*$. ■

It, therefore, suffices to prove that the sufficient conditions in Lemma 2 are satisfied by the Pareto, log-normal, and Weibull distributions. We know, by definition, that all these distributions are continuous and have mean $\mathbb{E}\left[S\right] = 1/\mu$ for all parameter values. The other conditions are verified in Appendix B. ■

We showed that the minimum average age $A_{\text{min}} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ is achieved under the three heavy tailed service time distributions. Such heavy tailed services' induce maximum variation in the service times, and thus, will yield a worse delay and delay variance. However, it is not clear whether these are the only distributions that can achieve minimum age. Perhaps, we may be able to find a distribution F_S, that minimizes age and well as delay and delay variance. In the next section, we prove that this is not so, and that there is a strong age-delay and age-delay variance tradeoff.

B. Age-Delay Tradeoff

We now prove that there exists a strong age-delay and age-delay variance tradeoff for the single server system, when the queue scheduling is fixed to LCFSp.

Theorem *4:* For a single server system under LCFSp scheduling policy, the following statements are true:

1) When the update generation is Poisson, there is a strong age-delay tradeoff.
2) When the update generation is a general renewal process, there is a strong age-delay variance tradeoff.

Proof: Let $A_{\min} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ denote the minimum average age. We have to show that as AoI $\to A_{\min}$ in (8), $T(\text{AoI}) \to +\infty$, when the updates are generated according to a Poisson process. We also have to show that as AoI $\to A_{\min}$ in (9), $V(\text{AoI}) \to +\infty$ for general update generation process.

When the update generation is Poisson, the queue is a M/G/1 LCFSp queue. The packet delay for this queue is given by [56]:

$$D(F_S) = \frac{\lambda}{2}\frac{\mathbb{E}\left[S^2\right]}{1-\rho} + \mathbb{E}\left[S\right]. \quad (16)$$

Furthermore, the delay variance is lower-bounded by the variance in service time, namely, $\text{VarD}(F_S) \geq \mathbb{E}\left[S^2\right] - \mathbb{E}\left[S\right]^2$. Therefore, it suffices to show that as AoI $\to A_{\min}$ in (8) and (9) we have $\mathbb{E}\left[S^2\right] \to +\infty$.

In the following, we prove the strong age-delay tradeoff. The arguments are exactly the same for establishing the strong age-delay variance tradeoff, as we only have to show that $\mathbb{E}\left[S^2\right] \to +\infty$.

To establish the strong age-delay tradeoff, we use the expressions for average age derived in Lemma 1. Let S_{AoI} denote the service time, and $F_{S_{\text{AoI}}}$ the corresponding service time distribution, that solves (8). Now, as AoI $\to A_{\min}$ in (8) we must have

$$\begin{aligned} A^{\text{ave}}_{\text{G/G/1}}(F_{S_{\text{AoI}}}) &= \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \frac{\mathbb{E}\left[\min\left(X, S_{\text{AoI}}\right)\right]}{\mathbf{P}\left[S_{\text{AoI}} < X\right]} \\ &\to \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} = A_{\min}, \quad (17) \end{aligned}$$

which implies

$$\lim_{\text{AoI}\to A_{\min}} \frac{\mathbb{E}\left[\min\left(X, S_{\text{AoI}}\right)\right]}{\mathbf{P}\left[S_{\text{AoI}} < X\right]} = 0. \quad (18)$$

Now, notice that $\mathbf{P}\left[S_{\text{AoI}} < X\right]$, being probability, is bounded by 1. Therefore, for (18) to hold, it must be the case that

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[\min\left(X, S_{\text{AoI}}\right)\right] = 0. \quad (19)$$

Substituting the fact $\mathbb{E}\left[\min\left(X, S_{\text{AoI}}\right)\right] = \mathbb{E}\left[X\mathbb{I}_{\{X<S_{\text{AoI}}\}}\right] + \mathbb{E}\left[S_{\text{AoI}}\mathbb{I}_{\{X\geq S_{\text{AoI}}\}}\right]$ in (19) we get

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[X\mathbb{I}_{\{X<S_{\text{AoI}}\}}\right] = 0, \quad (20)$$

and

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[S_{\text{AoI}}\mathbb{I}_{\{X\geq S_{\text{AoI}}\}}\right] = 0. \quad (21)$$

Now, (20) and (21) implies that there exists a $x_0 > 0$ such that

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[\mathbb{I}_{\{x_0<S_{\text{AoI}}\}}\right] = 0, \quad (22)$$

and

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[S_{\text{AoI}}\mathbb{I}_{\{x_0\geq S_{\text{AoI}}\}}\right] = 0. \quad (23)$$

This can be established by a short proof by contradiction. Using Lemma 8 in Appendix E, along with (22) and (23), we obtain

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[\mathbb{I}_{\{x<S_{\text{AoI}}\}}\right] = 0, \quad (24)$$

and

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[S_{\text{AoI}}\mathbb{I}_{\{x\geq S_{\text{AoI}}\}}\right] = 0, \quad (25)$$

for all $x \geq x_0$. Lemma 10 in Appendix E shows that these two conditions in (24) and (25) imply

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[S^2_{\text{AoI}}\right] = +\infty, \quad (26)$$

which proves our result. ∎

In the next subsection, we bring out an even stronger contrast between age and delay, than the strong age-delay tradeoff. We show that the delay minimizing service time distribution results in the worst case average age, and that the average age minimizing service time distribution results in the worst case delay.

C. Age-Delay Disparity under Poisson Update Generation

To bring out the contrast between packet delay and AoI metrics, we consider the special case in which the update packets are generated according to a Poisson process. In this case, the single server system is nothing but a M/G/1 LCFSp queue. We now show that deterministic service yields the worst average age, across all service time distributions. We use the notation $A^{\text{ave}}_{G/G/1}$ for average age for the G/G/1 LCFSp queue, and omit the dependence on service time distribution F_S for convenience.

Theorem *5:* For a single server system under LCFSp scheduling policy and Poisson update generation, the deterministic service yields the worst case average age:

$$A^{\text{ave}}_{\text{M/G/1}} \leq A^{\text{ave}}_{\text{M/D/1}}.$$

Proof: See Appendix C. ∎

It should be intuitive that if the packets in service are often preempted, then very few packets complete service on time, and this results in a very high AoI. It turns out that deterministic service maximizes the probability of preemption. For the LCFS M/G/1 queue, the probability of preemption is given by $\mathbf{P}\left[S > X\right] = 1 - \mathbb{E}\left[e^{-\lambda S}\right]$, as X is exponentially distributed with rate λ. This can be upper-bounded by $1 - e^{-\lambda\mathbb{E}[S]} = \mathbf{P}\left[\mathbb{E}\left[S\right] > X\right]$, using Jensen's inequality, which is nothing but the probability of preemption under deterministic service: $S = \mathbb{E}\left[S\right]$ almost surely.

Comparing age with packet delay for the LCFSp queue results in a peculiar conclusion. The packet delay for a M/G/1 LCFSp queue is given by [56]:

$$D = \frac{\lambda}{2}\frac{\mathbb{E}\left[S^2\right]}{1-\rho} + \mathbb{E}\left[S\right].$$

Note that this expression of packet delay D is minimized when the service time S is deterministic, namely $S = \mathbb{E}\left[S\right]$ almost surely; follows from Jensen's inequality $\mathbb{E}\left[S^2\right] \geq \mathbb{E}\left[S\right]^2$. However, from Theorem 5 we know that deterministic service time maximizes age. This leads to the conclusion that, for the M/G/1 LCFSp queue, *the service time distribution that minimizes delay, maximizes age of information*.

V. Infinite Servers

In this section, we consider the case when $m = \infty$, i.e. there are infinite servers in the system. The queue scheduling policy π_Q is also fixed, and it assigns a new server to every arriving update packet. We call this the *work conserving* scheduling policy. The infinite server system, under the work conserving policy, is nothing but the G/G/∞ queue. With the scheduling policy π_Q fixed, the age and delay metrics are just a function of the service time distribution F_S.

Note that under the above scheduling policy, the packet delay incurred equals the service time, and thus, $D = \mathbb{E}\left[S\right] = 1/\mu$. This implies that the minimum age and minimum delay, which is $1/\mu$, can be achieved simultaneously, and the service time distribution that achieves this can be obtained by solving (5).

The age-delay variance tradeoff problem, on the other hand, is not so trivial. This can be written as

$$\begin{aligned} V(\text{AoI}) = \underset{F_S}{\text{Minimize}} \quad & \text{VarD}(F_S) \\ \text{subject to} \quad & A^{\text{ave}}_{\text{G/G/}\infty}(F_S) \leq \text{AoI}, \\ & \mathbb{E}\left[S\right] = 1/\mu, \end{aligned} \tag{27}$$

where the notation $A^{\text{ave}}_{\text{G/G/}\infty}(F_S)$ denotes the average age for the G/G/∞ queue. The optimization is only over the service time distribution. We omit the dependence on F_S, whenever convenient.

The rest of this section is organized as follows. In Section V-A, we derive an expressions for the average age, and characterize its minimum. We also show that heavy tailed service time distributions achieve the minimum average age. In Section V-B, we use these results to prove that there is a strong age-delay variance tradeoff. In Section V-C, we obtain an average age maximizing service time distribution, and point to the disparity between the average age metric and delay variance.

A. Minimizing Age

We first derive an expression for average age for the system.

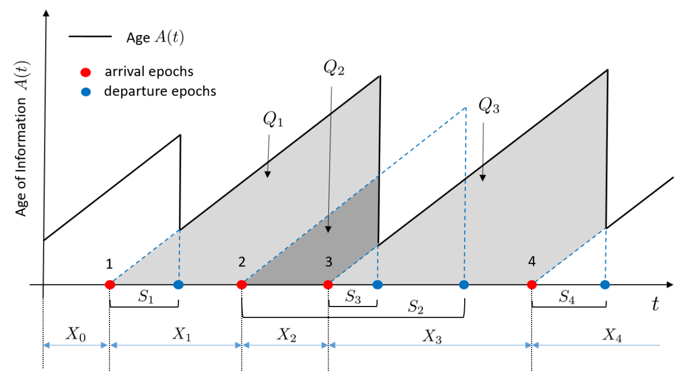

Fig. 8. Age $A(t)$ evolution over time t for G/G/∞ queue.

Lemma *3:* For the G/G/∞ queue, the average is given by

$$A^{\text{ave}}_{\text{G/G/}\infty}(F_S) = \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \mathbb{E}\left[\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right],$$

where X and $\{X_k\}_{k\geq 1}$ are i.i.d. distributed according to F_X, while $\{S_k\}_{k\geq 1}$ are i.i.d. distributed according to F_S.

Proof: For the G/G/∞ queue, each arriving packet is serviced by a different server. As a result, the packets may get serviced in an out of order fashion. Figure 8, which plots age evolution for the G/G/∞ queue, illustrates this. In Figure 8, observe that packet 3 completes service before packet 2. As a result, the age doesn't drop at the service of packet 2, as it now contains stale information. To analyze average age, it is important to characterize these events of out of order service.

Let X_i denote the inter-generation time between the ith and $(i+1)$th packet, and S_i denote the service time for the ith packet. In Figure 8, $X_2 + S_3 < S_2$, and therefore, packet 3 completes service before packet 2. To completely characterize this, define $Z_i \triangleq \sum_{k=0}^{i-1} X_k$ to be the time of generation of the ith packet. Note that the ith packet gets serviced at time $Z_i + S_i$, the $(i+1)$th packet gets serviced at time $Z_i + X_i + S_{i+1}$, and similarly, the $(i+l)$th packet gets serviced at time $Z_i + \sum_{k=1}^{l} X_{i+k-1} + S_{i+l}$, for all $l \geq 1$.

Let J_i denote the time at which the first amongst the packets $\{i, i+1, i+2, \ldots\}$ get served. Namely,

$$\begin{aligned} J_i &= \min\{Z_i + S_i, Z_i + X_i + S_{i+1}, X_i + X_{i+1} + S_{i+2}, \ldots\}. \\ &= Z_i + \min\{S_i, X_i + S_{i+1}, X_i + X_{i+1} + S_{i+2}, \ldots\}. \end{aligned}$$

Let D_i denote the time from the generation of the ith packet, Z_i, to the time at which the first amongst the packets $\{i, i+1, i+2, \ldots\}$ get served, J_i:

$$\begin{aligned} D_i = J_i - Z_i, &= \min\{S_i, X_i + S_{i+1}, X_i + X_{i+1} + S_{i+2}, \ldots\} \\ &= \min_{l\geq 0}\left\{\sum_{k=1}^{l} X_{i+k-1} + S_{i+l}\right\}. \end{aligned} \tag{28}$$

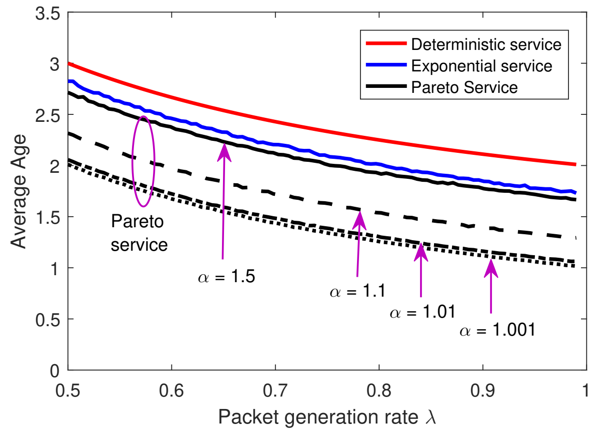

Fig. 9. Plotted is the average age under deterministic, exponential, and Pareto ($\alpha = 1.5, 1.1, 1.01$, and 1.001) distributed service times distributions for the infinite server M/G/∞ queue. Service rate $\mu = 1$, while the packet generation rate λ varies from 0.5 to 0.99.

In Figure 8, note that $D_1 = S_1$, $D_2 = X_2 + S_3$, $D_3 = S_3$, and $D_4 = S_4$.

The area under the age curve $A(t)$ is nothing but the sum of the areas of the trapezoids Q_i (see Figure 8). Applying the renewal reward theorem [57], by letting the reward for the ith renewal, namely $[Z_i, Z_i + X_i)$, be the area Q_i, we get the average age to be:

$$A^{\text{ave}}_{\text{G/G/}\infty}(F_S) = \frac{\mathbb{E}[Q_i]}{\mathbb{E}[X_i]}. \tag{29}$$

It is easy to see that

$$Q_i = \frac{1}{2}(X_i + D_{i+1})^2 - \frac{1}{2}D_{i+1}^2, \tag{30}$$

as the trapezoid Q_i extends from the time of the ith packet generation to the time at which the $(i+1)$th, or a packet that arrives after the $(i+1)$th packet, is served; which is nothing but $X_i + D_{i+1}$. For illustration, note that $Q_1 = \frac{1}{2}(X_1 + X_2 + S_3)^2 - \frac{1}{2}(X_2 + S_3)^2$, which is same as (30), for $i = 1$, since $D_2 = X_2 + S_3$. Substituting (30) in (29), we obtain

$$A^{\text{ave}}_{\text{G/G/}\infty}(F_S) = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]} + \frac{\mathbb{E}[X_i D_{i+1}]}{\mathbb{E}[X_i]}. \tag{31}$$

We obtain the result by noting that X_i and D_{i+1} are independent. ■

In Figure 9, we plot the average age for the M/G/∞ queue under three service distributions: deterministic, exponential, and Pareto distribution (given in (11)), with mean $1/\mu$. We observe that the heavy tail Pareto distributed service performs better than the exponential service. Also, heavier tail or decreasing α results in improvement in age. It appears, like in the LCFSp queue, that as $\alpha \downarrow 1$ the average age approaches the lower bound $1/\lambda$. Similar observations are made for the log-normal distributed service (12) and Weibull distributed service (13), we $\sigma \to +\infty$ and $\kappa \to 0$, respectively.

We now prove a simple lower bound on the average age, and show that the average age converges to this lower bound for the three heavy tailed service time distribution.

Theorem *6:* For the infinite server G/G/∞ queue, the average age is lower-bounded by

$$A^{\text{ave}}_{\text{G/G/}\infty}(F_S) \geq \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}.$$

Further, the lower-bound is achieved for

1) Pareto distributed service (11) as $\alpha \to 1$,
2) Log-normal distributed service (12) as $\sigma \to +\infty$, and
3) Weibull distributed service (13) as $\kappa \to 0$.

Proof: The lower-bound immediately follows from the average age expression in Lemma 3. We use a similar approach to that followed in the proof of Theorem 3. We show that the same sufficient conditions as in Lemma 2 suffices for the average age optimality for the G/G/∞ queue.

Lemma *4:* Let a parametric, continuous, service time distribution, with parameter η, satisfy

1) $\mathbb{E}[S] = 1/\mu$,
2) $\mathbb{E}\left[\mathbb{I}_{\{S>x\}}\right] \to 0$ as $\eta \to \eta^*$, and
3) $\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] \to 0$ as $\eta \to \eta^*$,

for some η^*, and all $x > 0$. Then

$$\lim_{\eta\to\eta^*} A^{\text{ave}}_{\text{G/G/}\infty}(F_S) = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}. \tag{32}$$

Proof: It suffices to argue that when the above conditions hold for a parametric random variable S, we have

$$\lim_{\eta\to\eta^*} \mathbb{E}\left[\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right] = 0. \tag{33}$$

This is proved in Lemma 9 in Appendix E. ■

It, now, suffices to argue that the three heavy tailed service time distributions satisfy the conditions in Lemma 4. All the three heavy tailed distributions are continuous, and have mean $\mathbb{E}[S] = 1/\mu$, by definition. The other two properties are verified in Appendix B. ■

Thus, the minimum age can be achieved by the three heavy tailed service time distribution. For these three distributions, the second moment approaches infinity, as their tails get heavier; namely as $\alpha \to 1$, $\sigma \to +\infty$, and $\kappa \to 0$. This implies that the delay variance, which is lower-bounded by the variance in service time, also approaches infinity. However, it is not known that there is no other distribution that can simultaneously minimize average age and delay variance. In the next sub-section, we prove just that, and show that there is a strong age-delay variance tradeoff.

B. Age-Delay Tradeoff

We now prove that there is a strong age-delay variance tradeoff.

Theorem *7:* For the infinite server system, under the work conserving scheduling policy, there is a strong age-delay variance tradeoff.

Proof: Let $A_{\min} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ denote the minimum average age. We have to show that as AoI $\to A_{\min}$ in (27), $V(\text{AoI}) \to +\infty$. Note that the delay variance is lower-bounded by the variance in service time, namely, $\text{VarD}(F_S) \geq \mathbb{E}\left[S^2\right] - \mathbb{E}\left[S\right]^2$. Therefore, it suffices to show that as AoI $\to A_{\min}$ in (27) we have $\mathbb{E}\left[S^2\right] \to +\infty$.

To establish this, we use the average age expression derived in Lemma 3. Let S_{AoI} denote the service time, and $F_{S_{\text{AoI}}}$ the corresponding service time distribution, that solves (27). Now, as AoI $\to A_{\min}$ in (27) we must have

$$A^{\text{ave}}_{\text{G/G/}\infty}(F_{S_{\text{AoI}}}) = \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \mathbb{E}\left[\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right] \to \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} = A_{\min}, \quad (34)$$

where S_l and X_k are independent and distributed according to $F_{S_{\text{AoI}}}$ and F_X, respectively. This implies

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right] = 0. \quad (35)$$

From Lemma 9 and Lemma 10, in Appendix E, (35) implies that

$$\lim_{\text{AoI}\to A_{\min}} \mathbb{E}\left[S^2_{\text{AoI}}\right] = +\infty, \quad (36)$$

which proves the result. ■

In the next sub-section, we prove an even stronger disparity between average age and delay. We show that the service time distribution that minimizes delay variance, i.e. deterministic service, yields the worst case age.

C. Age-Delay Disparity

We first prove that deterministic service yields the worst average age, across all service time distributions.

Theorem *8:* For the infinite server G/G/∞ queue,

$$A^{\text{ave}}_{\text{G/G/}\infty}(\lambda, \mu) \leq A^{\text{ave}}_{\text{G/D/}\infty}(\lambda, \mu),$$

for all packet generation and service rates, λ and μ, respectively.

Proof: See Appendix D. ■

The intuition is as follows: In the G/G/∞ queue, packets do not get serviced in the same order as they are generated. However, a swap in order helps improve age, because it means that a packet that arrived later was served earlier. Therefore, the service that swaps the packet order the least maximizes age. Under deterministic service, the packet order is retained exactly, with probability 1, and therefore, yields the maximum age.

Notice that, for the G/G/∞ queue, packet delay equals the service time, and therefore, deterministic service minimizes delay variance. This observation, along with Theorem 8, imply that for the G/G/∞ queue, the *service time distribution that reduces packet delay variance, maximizes average age of information*.

The next section considers the general m server system, and proves the age-delay tradeoff result of Theorem 2.

VI. m Server System

In this section, we consider the m server system and prove Theorem 2, which asserts a strong tradeoff between age-delay and age-delay variance. Recall that $A^{\text{ave}}(F_S, \pi_Q)$, $D(F_S, \pi_Q)$, and $\text{VarD}(F_S, \pi_Q)$ denote the average age, delay, and delay variance, respectively, under the scheduling policy π_Q and the service time distribution F_S.

We first derive the minimum average age $A_{\min}$, over the space of all scheduling policies π_Q and service time distributions F_S.

Lemma *5:* The minimum average age $A_{\min} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$.

Proof: The fact that $\frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ is a lower-bound on the average age, can be proved by pretending that each update packet spends zero time in the system, i.e. $t_i = t_i^{'}$. This provides a sample path lower bound for the age process. In this sample path, the age drops to 0 at every t_i, and increases to $t_{i+1} - t_i$, just before dropping to 0 again at t_{i+1}. The average age of this artificially constructed age process is $\frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$, and since it is a sample path wise lower-bound, we have $A^{\text{ave}}(F_S, \pi_Q) \geq \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$. This lower-bound $\frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ is independent of the scheduling policy π_Q and the service time distribution F_S. Therefore, we have $A_{\min} \geq \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$.

In Section IV, we showed that this lower-bound can be achieved by a single server system, i.e. $m = 1$, under the LCFSp scheduling policy with heavy tailed service. Therefore, choosing to route update packets only through a single sever, scheduling packets in that server with LCFSp scheduling policy with heavy tailed service, we can achieve this lower bound average age. Thus, $A_{\min} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$. ■

We now prove the strong age-delay tradeoff and age-delay variance tradeoff.

Lemma *6:* For the m server system, the following statements are true:

1) When the update generation is Poisson, there is a strong age-delay tradeoff.
2) When the update generation is a general renewal process, there is a strong age-delay variance tradeoff.

Proof:

1. Age-delay tradeoff: Consider Poisson update generation. It is known that the minimum delay can be attained by any work conserving scheduling policy [22], [23]. We define the following work conserving scheduling policy π_Q^*:

1) Generated updates are queued in a single FCFS queue.
2) Whenever a server is free, an update packet at the head of the FCFS queue, is assigned to that server.

This scheduling policy, begin work conserving, attains minimum average delay for a given service time distribution, i.e.

$$D(F_S, \pi_Q) \geq D(F_S, \pi_Q^*), \tag{37}$$

for all scheduling policies π_Q. The m server system under Poisson update generation and scheduling policy π_Q^* is nothing but the M/G/m queue. For the M/G/m queue, the average delay, namely $D(F_S, \pi_Q^*)$, is lower-bounded by a constant times the variance in service time [58]:

$$D(F_S, \pi_Q^*) \geq c\left(\mathbb{E}\left[S^2\right] - \mathbb{E}\left[S\right]^2\right), \tag{38}$$

where in [58] the constant relates to the delay in the FCFS M/M/m queue. From (38) and (37), we have

$$D(F_S, \pi_Q) \geq c\left(\mathbb{E}\left[S^2\right] - \mathbb{E}\left[S\right]^2\right), \tag{39}$$

for any scheduling policy π_Q.

From (39), it is clear, that in order to prove a strong age-delay tradeoff, for Poisson update generation, it suffices to argue that $\mathbb{E}\left[S^2\right] \rightarrow +\infty$ as AoI $\rightarrow A_{\min}$ in (4); where $A_{\min} = \frac{1}{2}\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ denotes the minimum average age. In the rest of the proof, we prove just this.

Let S_{AoI} and $F_{S_{\text{AoI}}}$ denote the service time random variable and its distribution, respectively, that solves (4). We argue that as AoI $\rightarrow A_{\min}$ in (4) we must have $\mathbb{E}\left[S_{\text{AoI}}^2\right] \rightarrow +\infty$.

We first note that the average age $A^{\text{ave}}(F_S, \pi_Q)$, under any queue scheduling policy π_Q, is lower-bounded by the average age for the G/G/∞ queue:

$$A^{\text{ave}}(F_S, \pi_Q) \geq A^{\text{ave}}_{\text{G/G/}\infty}(F_S). \tag{40}$$

This is because, in G/G/∞ queue, an arriving packet is immediately put to service, and therefore, incurs no queueing delay. Due to this the average age for the G/G/∞ queue serves as a lower-bound for any m server queue, in a stochastic ordering sense. Taking expected value yields (40).

We know the average age for the G/G/∞ queue to be:

$$A^{\text{ave}}_{\text{G/G/}\infty} = \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right], \tag{41}$$

where S_l and X_k are independent random variables with distributions $F_{S_{\text{AoI}}}$ and F_X, respectively. Notice that the first term in (41) is nothing but $A_{\min}$. Therefore, as AoI $\rightarrow$ $A_{\min}$ in (4), it must be the case that $\mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right] \rightarrow 0$. Lemmas 9 and 10, in Appendix E, prove that $\mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right] \rightarrow 0$ implies $\mathbb{E}\left[S_{\text{AoI}}^2\right] \rightarrow +\infty$.

2. Age-delay variance tradeoff: Note that the variance in packet delay is lower-bounded by the variance in service time, under any scheduling policy π_Q. Therefore,

$$\text{VarD}(F_S, \pi_Q) \geq \mathbb{E}\left[S^2\right] - \mathbb{E}\left[S\right]^2. \tag{42}$$

Therefore, in order to establish the strong age-delay variance tradeoff, we need to argue that $\mathbb{E}\left[S^2\right] \rightarrow +\infty$ and AoI $\rightarrow$ $A_{\min}$ in (6).

Let S_{AoI} and $F_{S_{\text{AoI}}}$ denote the service time random variable and its distribution that solves (6). We have to argue that as AoI $\rightarrow A_{\min}$ in (6) we have $\mathbb{E}\left[S_{\text{AoI}}^2\right] \rightarrow +\infty$. Note that, we just proved this in establishing the age-delay tradeoff. ∎

VII. Conclusion and Open Problems

We considered an m server system in which each server serves at most one update packet at any given time. Updates are generated according to a renewal process and the system designer controls the scheduling discipline, routing, and the service time distribution. When the updates are generated according to a Poisson process, we show that there is a strong age-delay tradeoff, i.e. as the average age approaches its minimum the packet delay tends to infinity. However, for a general update generation process, we prove a strong age-delay variance tradeoff. The proof involves first establishing similar age-delay tradeoff results for two special cases of the m server system, namely, the single server system with LCFSp service and the infinite server system. For the two cases, we also show that heavy tailed service time distributions asymptotically minimize average age, as their tail gets heavier.

Though seemingly counterintuitive, the age-delay tradeoff is natural and occurs due to the delays incurred by the non-informative packets. When the system attempts to minimize age, it minimizes waiting and service times for the informative packets. This results in very high waiting and service times for the non-informative packets, which dominate the packet delay, and causes it to increase unboundedly. We therefore expect similar age-delay tradeoffs to exist in other communication systems as well, and investigating them is an open question for future research.

In a real communication system, the space of service times, routing, and scheduling policies that one can choose from may be more constrained than assumed here. For instance, a lot of factors affect and constraint variability in service in real systems. For a link layer packet, the service time variability may be due to channel fading, sudden drops in SNR, packet collisions due to contention etc. For a transport layer packet, on the other hand, the service time variability is affected by the underlying routing mechanisms, packet management, packet fragmentation used, ARQ protocols used by all the underlying links, exogenous traffic etc. Understanding age-delay tradeoffs under these realistic constraints is important and interesting direction for future work.

APPENDIX

A. Proof of Lemma 1

Let $A(t)$ denote the age at time t. Let B_i denote the age at the generation of the ith update packet, i.e. $Z_i = \sum_{k=0}^{i-1} X_k$:

$$B_i = A(Z_i). \tag{43}$$

Then, we have the following recursion for B_i:

$$B_{i+1} = \begin{cases} X_i & \text{if } S_i < X_i \\ B_i + X_i & \text{if } S_i \geq X_i \end{cases}, \tag{44}$$

for all $i \geq 0$. This can be written as

$$B_{i+1} = X_i + B_i \left(1 - \mathbb{I}_{S_i < X_i}\right). \tag{45}$$

Note that B_i is independent of S_i and X_i. Further, $\{B_i\}_{i\geq 1}$ is a Markov process, and can be shown to be positive recurrent using the drift criteria [59]; using the fact that X_i and S_i are continuous random variables and $\mathbf{P}[S_i < X_i] < 1$. Taking expected value, and noting that at stationarity $\mathbb{E}[B_i] = \mathbb{E}[B_{i+1}]$, we get

$$\mathbb{E}[B] = \frac{\mathbb{E}[X]}{\mathbf{P}[S < X]}. \tag{46}$$

We now compute the average age. Let R_i denote the area under the age curve $A(t)$ between the generation of packet i and packet $i+1$:

$$R_i \triangleq \int_{Z_i}^{Z_i + X_i} A(t) dt, \tag{47}$$

where $Z_i = \sum_{k=0}^{i-1} X_k$ is the time of generation of the ith update packet. This R_i can be computed explicitly to be

$$R_i = \begin{cases} B_i X_i + \frac{1}{2} X_i^2 & \text{if } X_i < S_i \\ B_i S_i + \frac{1}{2} X_i^2 & \text{if } X_i \geq S_i \end{cases}, \tag{48}$$

which can be written compactly as

$$R_i = \frac{1}{2} X_i^2 + B_i \min\left(X_i, S_i\right). \tag{49}$$

Since, B_i is independent of X_i and S_i, taking expected value at stationarity we obtain

$$\mathbb{E}[R] = \frac{1}{2}\mathbb{E}\left[X^2\right] + \mathbb{E}[B]\,\mathbb{E}\left[\min\left(X, S\right)\right]. \tag{50}$$

Using renewal theory, the average age can be obtained to be

$$A^{\text{ave}}_{\text{G/G/1}} = \frac{\mathbb{E}[R]}{\mathbb{E}[X]}, \tag{51}$$

$$= \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}[X]} + \frac{\mathbb{E}[B]}{\mathbb{E}[X]}\mathbb{E}\left[\min\left(X, S\right)\right]. \tag{52}$$

Substituting (46) we get the result.

B. Properties of the Heavy Tailed Distributions

Lemma *7:* For any $x > 0$, we have $\mathbf{P}[S > x] \to 0$ and $\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right] \to 0$ for:

1) Pareto distributed service S, as $\alpha \to 1$; see (11).
2) Log-normal distributed service S, as $\sigma \to +\infty$; see (12).
3) Weibull distributed service S, as $\kappa \to 0$; see (13).

Proof:

1. Pareto Service: Choose a $x > 0$. Then there exists a $\overline{\alpha}_x > 1$ such that $\theta(\alpha) = \frac{1}{\mu}\frac{\alpha-1}{\alpha} < x$ for all $\alpha < \overline{\alpha}_x$. For such any $\alpha < \overline{\alpha}_x$, we have $\mathbf{P}[S > x] = \left(\frac{\theta(\alpha)}{x}\right)^{\alpha} \to 0$ as $\alpha \downarrow 1$, since $\theta(\alpha) \to 0$ as $\alpha \downarrow 1$.

For the second part, we first compute $\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right]$ for $\alpha < \overline{\alpha}_x$:

$$\mathbb{E}\left[S\mathbb{I}_{S \leq x}\right] = \int_{\frac{1}{\mu}\left(1-\frac{1}{\alpha}\right)}^{x} s f_S(s) ds = \frac{\alpha}{\mu^{\alpha}} \int_{\frac{1}{\mu}\left(1-\frac{1}{\alpha}\right)}^{x} \frac{\left(1-\frac{1}{\alpha}\right)^{\alpha}}{s^{\alpha}} ds.$$

Substituting $y = \alpha s/(\alpha - 1)$, and solving the definite integral, we get

$$\mathbb{E}\left[S\mathbb{I}_{S \leq x}\right] = \frac{1}{\mu} - \frac{1}{\mu}\frac{(\alpha/\mu)^{\alpha-1}}{(\alpha-1)^{\alpha-1}} x^{\alpha-1}. \tag{53}$$

From the above expression, it can be deduced that $\mathbb{E}\left[S\mathbb{I}_{S \leq x}\right] \to 0$ as $\alpha \downarrow 1$.

2. Log-normal Service: Choose a $x > 0$. From (12) notice that

$$\mathbf{P}[S > x] = \mathbf{P}\left[N > \frac{\log(x\mu)}{\sigma} + \frac{\sigma}{2}\right] \to 0,$$

as $\sigma \to +\infty$.

For the second part, using the relation (12) between the log-normal service time and normal random variable N, we can compute the expectation $\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right]$ to be

$$\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right] = \frac{1}{\mu} - \frac{1}{\mu}\Phi\left(-\frac{\log(x\mu)}{\sigma} + \frac{\sigma}{2}\right),$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2} dt$. Taking the limit $\sigma \to +\infty$ we get $\Phi\left(-\frac{\log(x\mu)}{\sigma} + \frac{\sigma}{2}\right) \to 1$, and therefore, $\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right] \to 0$.

3. Weibull Service: Choose a $x > 0$. Using the distribution function (13), we can conclude $\mathbf{P}[S > x] = e^{-(x\mu)^{\kappa}} e^{-[\Gamma(1+1/\kappa)]^{\kappa}}$. Using Sterling's formula, $[\Gamma(1+1/\kappa)]^{\kappa} \geq 1/\kappa$, and therefore $[\Gamma(1+1/\kappa)]^{\kappa} \to +\infty$ as $\kappa \to 0$. Therefore, we have $\mathbf{P}[S > x] \to 0$ as $\kappa \to 0$.

For the second part, we can explicitly derive the conditional expectation $\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right]$ using the distribution (13):

$$\mathbb{E}\left[S\mathbb{I}_{\{S \leq x\}}\right] = \int_0^x \frac{\kappa}{\beta}\left(\frac{t}{\beta}\right)^{\kappa-1} e^{-(t/\beta)^{\kappa}} t dt,$$

$$= \frac{1}{\mu\Gamma(1+1/\kappa)} \int_0^{(x\mu\Gamma(1+1/\kappa))^{\kappa}} y^{1/\kappa} e^{-y} dy, \tag{54}$$

which is obtained by substituting $\beta = [\mu\Gamma(1+1/\kappa)]^{-1}$ and changing variables $y = (t/\beta)^{\kappa}$. Using lower-bounds given by Sterling approximation on Gamma function, we can deduce that (54), and therefore $\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right]$, approaches 0 as $\kappa \to 0$. ■

C. Proof of Theorem 5

We first show that the average age for Poisson update generation is given by

$$A^{\text{ave}}_{\text{M/G/1}} = \frac{\mathbb{E}[S]}{\mathbf{P}[S < X]}. \tag{55}$$

Let $A(t)$ be the age at time t, and B_i be the age at the time of generation of the ith update packet $Z_i = \sum_{k=0}^{i-1} X_k$:

$$B_i = A(Z_i). \tag{56}$$

Let B denote the distribution of B_i at stationarity. By PASTA property and ergodicity of the age process $A(t)$ we have $A^{\text{ave}}_{\text{M/G/1}} = \mathbb{E}[B]$, as update generation process is a Poisson process. Substituting the expression for $\mathbb{E}[B]$ in (46), from Appendix A, we obtain (55).

Now, substituting $S = \mathbb{E}[S]$ almost surely we get the average age expression for the $M/D/1$ LCFSp queue to be

$$A^{\text{ave}}_{\text{M/D/1}} = \frac{\mathbb{E}[S]}{\mathbf{P}[\mathbb{E}[S] < X]} = \frac{\mathbb{E}[S]}{e^{-\lambda\mathbb{E}[S]}}, \tag{57}$$

where we have used the fact that the packet inter-generation time X is exponentially distributed. We obtain $A^{\text{ave}}_{\text{M/G/1}} \leq A^{\text{ave}}_{\text{M/D/1}}$ by noting that

$$\mathbf{P}[S < X] = \mathbb{E}\left[e^{-\lambda S}\right] \geq e^{-\lambda\mathbb{E}[S]}, \tag{58}$$

by Jensen's inequality.

D. Proof of Theorem 8

From Lemma 3, it is clear that the average age depend on service time through the term:

$$\mathbb{E}\left[\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right]. \tag{59}$$

We show that this quantity is maximized when service times are deterministic, i.e. $S = \mathbb{E}[S]$ almost surely.

First, notice that

$$\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\} = S_1, \tag{60}$$

if S_k are all equal and deterministic. This is because $X_k \geq 0$ almost surely. Thus, the peak and average age for the G/D/∞ queue is given by

$$A^{\text{p}}_{\text{G/D/}\infty} = \mathbb{E}[X] + \mathbb{E}[S], \tag{61}$$

and

$$A^{\text{ave}}_{\text{G/D/}\infty} = \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}[X]} + \mathbb{E}[S]. \tag{62}$$

Furthermore, we must have

$$\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\} \leq S_1, \tag{63}$$

since S_1 is the first term in the minimization. Therefore,

$$\mathbb{E}\left[\min_{l\geq 0}\left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right] \leq \mathbb{E}[S_1] = \mathbb{E}[S]. \tag{64}$$

Applying this to the peak and average age expression from Lemma 3, we get

$$A^{\text{p}}_{\text{G/G/}\infty} \leq \mathbb{E}[X] + \mathbb{E}[S], \tag{65}$$

and

$$A^{\text{ave}}_{\text{G/G/}\infty} \leq \frac{1}{2}\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}[X]} + \mathbb{E}[S]. \tag{66}$$

The result follows from (61), (62), (65), and (66).

E. Properties of Service Time Random Variable S

Here, we derive several asymptotic properties of the service time distributions and their implications. These properties are used throughout the paper.

Let S be a continuous random variable with distribution F_S, with parameter η, such that $\mathbb{E}[S] = 1/\mu$ for all η. For notational convenience, we hide the dependence of S and F_S on η. We are interested in the S and F_S as $\eta \to \eta^*$, for some specific η^*.

> **Lemma** *8:* If $\exists$ a $x_0 > 0$ such that $\mathbf{P}[S > x_0] \to 0$ and $\mathbb{E}\left[S\mathbb{I}_{\{S\leq x_0\}}\right] \to 0$ as $\eta \to \eta^*$ then
>
> $$\lim_{\eta\to\eta^*} \mathbf{P}[S > x] = 0 \text{ and } \lim_{\eta\to\eta^*} \mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] = 0, \tag{67}$$
>
> for all $x \geq x_0$.

Proof: Let there be a $x_0 > 0$ such that

$$\lim_{\eta\to\eta^*} \mathbf{P}[S > x_0] = 0 \text{ and } \lim_{\eta\to\eta^*} \mathbb{E}\left[S\mathbb{I}_{\{S<x_0\}}\right] = 0. \tag{68}$$

Take a $x > x_0$. Then, $\mathbb{I}_{\{S>x\}} \leq \mathbb{I}_{\{S>x_0\}}$, and therefore, $\mathbf{P}[S > x] \leq \mathbf{P}[S > x_0]$. This and (68) implies

$$\lim_{\eta\to\eta^*} \mathbf{P}[S > x] = 0. \tag{69}$$

For a $x > x_0$, we can re-write $\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right]$ as

$$\begin{aligned}
\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] &= \mathbb{E}\left[S\mathbb{I}_{\{S\leq x_0\}}\right] + \mathbb{E}\left[S\mathbb{I}_{\{x_0<S\leq x\}}\right], && (70)\\
&\leq \mathbb{E}\left[S\mathbb{I}_{\{S\leq x_0\}}\right] + x\mathbb{E}\left[\mathbb{I}_{\{x_0<S\leq x\}}\right], && (71)\\
&\leq \mathbb{E}\left[S\mathbb{I}_{\{S\leq x_0\}}\right] + x\mathbf{P}[S > x_0]. && (72)
\end{aligned}$$

Using (68), which states that both the terms in (72) tend to 0 as $\eta \to \eta^*$, we get

$$\lim_{\eta\to\eta^*} \mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] = 0. \tag{73}$$

Since (69) and (73) hold for any $x > x_0$, we have the result. ■

In the infinite server case, the average age expression in Lemma 3 has a term

$$\mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right], \tag{74}$$

where S_l and X_k are independent, distributed according to F_S and F_X, respectively. We would like to derive conditions on S such that

$$\mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right] \to 0,$$

as η approaches certain η^*, for a given distribution F_X. The following result, derives an equivalent condition that only requires verifying certain properties of F_S.

Lemma *9:* For S_l and X_k that are i.i.d. distributed according to F_S and F_X, respectively, we have

$$\lim_{\eta\to\eta^*}\mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right] = 0, \tag{75}$$

if and only if, for all $x > 0$, we have

$$\lim_{\eta\to\eta^*}\mathbf{P}\left[S > x\right] = 0, \text{ and } \lim_{\eta\to\eta^*}\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] = 0. \tag{76}$$

Proof: **(a)** We first prove that (75) implies (76). Let $Z = \min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)$. We first lower-bound Z as follow:

$$Z = \min\{S_1, X_1{+}S_2, X_1{+}X_2{+}S_3, \ldots\} = \min\{S_1, X_1{+}Z'\},$$

where $Z' = \min\{S_2, X_2{+}S_3, X_2{+}X_3{+}S_4, \ldots\}$. Since $Z' \geq 0$, we must have $Z \geq \min\{S_1, X_1\}$. Without loss of generality, we can loose the subscripts and write $Z \geq \min\{S, X\}$, where $S \sim F_S$ and $X \sim F_X$.

If $\mathbb{E}\left[Z\right] \to 0$ as $\eta \to \eta^*$ then clearly $\mathbb{E}\left[\min\{S, X\}\right] \to 0$ as $\eta \to \eta^*$. Pick a $x_0 > 0$ such that $\mathbf{P}\left[X \geq x_0\right] > 0$. Note that such an $x_0 > 0$ always exists since $\mathbb{E}\left[X\right] = 1/\lambda > 0$. Now construct $\hat{X}$ such that:

$$\hat{X} = \begin{cases} 0 & \text{if } X < x_0 \\ x_0 & \text{if } X \geq x_0 \end{cases}.$$

Clearly, $\hat{X} \leq X$, and thus, $\min\{S, \hat{X}\} \leq \min\{S, X\}$, which implies $\mathbb{E}\left[\min\{S, \hat{X}\}\right] \to 0$. Since $\hat{X}$ takes only two values, namely 0 and x_0, we have $\mathbb{E}\left[\min\{S, \hat{X}\}\right] = \mathbb{E}\left[\min\{S, x_0\}\right]\mathbf{P}\left[X \geq x_0\right]$. Further, $\mathbf{P}\left[X \geq x_0\right]$ does not depend on S, and therefore, is independent of the parameter η. Therefore, $\mathbb{E}\left[\min\{S, \hat{X}\}\right] \to 0$ implies

$$\lim_{\eta\to\eta^*}\mathbb{E}\left[\min\{S, x_0\}\right] = 0. \tag{77}$$

Now, notice that

$$\mathbb{E}\left[\min\{S, x_0\}\right] = \mathbb{E}\left[S\mathbb{I}_{\{S\leq x_0\}}\right] + x_0\mathbb{E}\left[\mathbb{I}_{\{S>x_0\}}\right], \tag{78}$$

Substituting (78) in (77) we get

$$\lim_{\eta\to\eta^*}\mathbb{E}\left[S\mathbb{I}_{\{S\leq x_0\}}\right] = 0 \text{ and } \lim_{\eta\to\eta^*}\mathbb{E}\left[\mathbb{I}_{\{S>x_0\}}\right] = 0. \tag{79}$$

Using Lemma 8, (79) implies

$$\lim_{\eta\to\eta^*}\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] = 0 \text{ and } \lim_{\eta\to\eta^*}\mathbb{E}\left[\mathbb{I}_{\{S>x\}}\right] = 0, \tag{80}$$

for all $x \geq x_0$.

Now, we had chosen x_0 to be such that $\mathbf{P}\left[X \geq x_0\right] > 0$. Since $\mathbb{E}\left[X\right] = 1/\lambda > 0$, the choice of x_0 could be as small, and close to 0, as possible. This and (80) yield the result in (76).

(b) We now prove that (76) implies (75). First, note that (76) along with the bounded convergence theorem [55] imply

$$\lim_{\eta\to\eta^*}\mathbf{P}\left[S > X\right] = 0 \text{ and } \lim_{\eta\to\eta^*}\mathbb{E}\left[S\mathbb{I}_{\{S\leq X\}}\right] = 0. \tag{81}$$

Using the same arguments we also have

$$\lim_{\eta\to\eta^*}\mathbb{E}\left[X\mathbb{I}_{\{S>X\}}\right] = 0. \tag{82}$$

Secondly, note that

$$\mathbb{E}\left[\min_{l\geq 0}\left(\sum_{k=1}^{l} X_k + S_{l+1}\right)\right] \leq \mathbb{E}\left[\min\{S_1, X_1 + S_2\}\right]. \tag{83}$$

It suffices to show that $\mathbb{E}\left[\min\{S_1, X_1 + S_2\}\right] \to 0$ as $\eta \to \eta^*$. To see this, we write $\mathbb{E}\left[\min\{S_1, X_1 + S_2\}\right]$ as:

$$\begin{aligned}
&\mathbb{E}\left[\min\{S_1, X_1 + S_2\}\right] \\
&= \mathbb{E}\left[S_1\mathbb{I}_{\{S_1\leq X_1\}}\right] + \mathbb{E}\left[\left[X_1 + \min\{S_1 - X_1, S_2\}\right]\mathbb{I}_{\{S_1>X_1\}}\right], \\
&\leq \mathbb{E}\left[S_1\mathbb{I}_{\{S_1\leq X_1\}}\right] + \mathbb{E}\left[\left[X_1 + S_2\right]\mathbb{I}_{\{S_1>X_1\}}\right], \\
&= \mathbb{E}\left[S_1\mathbb{I}_{\{S_1\leq X_1\}}\right] + \mathbb{E}\left[X_1\mathbb{I}_{\{S_1>X_1\}}\right] + \mathbb{E}\left[S_2\mathbb{I}_{\{S_1>X_1\}}\right], \\
&\to 0, \quad \text{as} \quad \eta \to \eta^*,
\end{aligned}$$

where the last equation follows from (81) and (82). ■

We now give a sufficient condition on the service time distributions F_S, parameterized by η, to have its second moment approach infinity. This result is used in proving the strong age-delay and age-delay variance tradeoffs.

Lemma *10:* For the parameterized, service time random variable S, we have $\lim_{\eta\to\eta^*}\mathbb{E}\left[S^2\right] = +\infty$ if

$$\lim_{\eta\to\eta^*}\mathbf{P}\left[S > x\right] = 0, \text{ and } \lim_{\eta\to\eta^*}\mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] = 0, \tag{84}$$

for all $x \geq x_0$, and some $x_0 > 0$.

Proof: Let the two conditions (84) hold for S. First, note that $\mathbb{E}\left[S^2\right] \geq \mathbb{E}\left[S^2\mathbb{I}_{\{S>x\}}\right] \geq x\mathbb{E}\left[S\mathbb{I}_{\{S>x\}}\right]$, for all $x > 0$. We can write $\mathbb{E}\left[S\mathbb{I}_{\{S>x\}}\right]$ as $\mathbb{E}\left[S\right] - \mathbb{E}\left[S\mathbb{I}_{\{S\leq x\}}\right] \to 1/\mu$ as $\eta \to \eta^*$ by (84) and the fact that $\mathbb{E}\left[S\right] = 1/\mu$. Therefore, we have $\liminf_{\eta\to\eta^*}\mathbb{E}\left[S^2\right] \geq x/\mu$ for all $x \geq x_0$. This can only be true if $\lim_{\eta\to\eta^*}\mathbb{E}\left[S^2\right] = +\infty$. ■

References

[1] R. Talak, S. Karaman, and E. Modiano, "Can determinacy minimize age of information?," *arXiv e-prints arXiv:1810.04371*, Oct. 2018.

[2] R. Talak and E. Modiano, "Age-delay tradeoffs in single server systems," in *Proc. ISIT*, pp. 340–344, Jul. 2019.

[3] R. Talak, S. Karaman, and E. Modiano, "When a heavy tailed service minimizes age of information," in *Proc. ISIT*, pp. 345–349, Jul. 2019.

[4] X. Jiang, H. Shokri-Ghadikolaei, G. Fodor, E. Modiano, Z. Pang, M. Zorzi, and C. Fischione, "Low-latency networking: Where latency lurks and how to tame it," *Proc. of the IEEE*, pp. 1–27, Aug. 2018.

[5] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile internet," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 460–473, Mar. 2016.

[6] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, and M. Dohler, "Toward haptic communications over the 5G tactile internet," *Commun. Surveys Tuts.*, vol. 20, pp. 3034–3059, Fourthquarter 2018.

[7] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, "The tactile internet: vision, recent progress, and open challenges," *IEEE Commun. Mag.*, vol. 54, pp. 138–145, May 2016.

[8] R. Talak, S. Karaman, and E. Modiano, "Speed limits in autonomous vehicular networks due to communication constraints," in *Proc. CDC*, pp. 4998–5003, Dec. 2016.

[9] I. Bekmezci, O. K. Sahingoz, and S. Temel, "Flying ad-hoc networks (FANETs): A survey," *Ad Hoc Networks*, vol. 11, pp. 1254–1270, May 2013.

[10] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.

[11] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. SECON*, pp. 350–358, Jun. 2011.

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

[13] C. Kam, S. Kompella, and A. Ephremides, "Effect of message transmission diversity on status age," in *Proc. ISIT*, pp. 2411–2415, Jun. 2014.

[14] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Proc. CISS*, pp. 1–6, Mar. 2012.

[15] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, pp. 1807–1827, Mar. 2019.

[16] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. ISIT*, pp. 1681–1685, Jun. 2015.

[17] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Proc. Allerton*, pp. 486–493, Oct. 2017.

[18] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. Mobihoc*, Jun. 2018.

[19] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *To appear in IEEE/ACM Trans. Netw.*, Dec. 2019.

[20] V. Tripathi, R. Talak, and E. Modiano, "Age of information for discrete time queues," *arXiv e-prints arXiv:1901.10463*, Jan. 2019.

[21] E. Najm and R. Nasser, "Age of information: The gamma awakening," *ArXiv e-prints*, Apr. 2016.

[22] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *Proc. ISIT*, pp. 2569–2573, Jul. 2016.

[23] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of the information through queues," *IEEE Trans. Inf. Theory*, Aug. 2019.

[24] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. ISIT*, pp. 576–580, Jun. 2017.

[25] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," *IEEE/ACM Trans. Netw.*, vol. 27, Jun. 2019.

[26] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. ISIT*, pp. 1583–1587, Jun. 2014.

[27] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "The stationary distribution of the age of information in FCFS single-server queues," in *Proc. ISIT*, pp. 571–575, Jun. 2017.

[28] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems," *arXiv e-prints arXiv:1805.12586*, Jun. 2018.

[29] R. D. Yates, "Status updates through networks of parallel servers," in *Proc. ISIT*, pp. 2281–2285, Jun. 2018.

[30] R. D. Yates, "Age of information in a network of preemptive servers," *arXiv e-prints arXiv:1803.07993*, Mar. 2018.

[31] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Age of information with a packet deadline," in *Proc. ISIT*, pp. 2564–2568, Jul. 2016.

[32] Y. Inoue, "Analysis of the age of information with packet deadline and infinite buffer capacity," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2639–2643, Jun. 2018.

[33] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Controlling the age of information: Buffer size, deadline, and packet replacement," in *Proc. MILCOM*, pp. 301–306, Nov. 2016.

[34] Y. Sun, I. Kadota, R. Talak, and E. Modiano, *Age of Information: A New Metric for Measuring Information Freshness*. Morgan & Claypool, 2019.

[35] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," *IEEE Trans. Inf. Theory*.

[36] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," *arXiv e-prints arXiv:1806.03487*, Jun. 2018.

[37] 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 Trans. Inf. Theory*, vol. 63, pp. 7492–7508, Nov. 2017.

[38] J. F. C. Kingman, "The effect of queue discipline on waiting time variance," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 58, no. 1, p. 163–164, 1962.

[39] T. Rolski and D. Stoyan, "On the comparison of waiting times in GI/G/1 queues," *Operations Research*, vol. 24, no. 1, pp. 197–200, 1976.

[40] R. W. Wolff, "The effect of service time regularity on system performance," Mar. 1977.

[41] W. Whitt, "The effect of variability in the GI/G/s queue," *J. Applied Prob.*, vol. 17, Dec. 1980.

[42] W. Whitt, "Minimizing delays in the GI/G/1 queue," *Operations Research*, vol. 32, Jan.-Feb. 1984.

[43] B. Hajek, "The proof of a folk theorem on queueing delay with applications to routing in networks," *J. Association for Computing Machinery*, vol. 30, Oct. 1983.

[44] W. Whitt, "Comparison conjectures about the M/G/s queue," *Operation Research Letters*, vol. 2, Oct. 1983.

[45] X. Zhao, W. Chen, J. Lee, and N. B. Shroff, "Delay-optimal and energy-efficient communications with markovian arrivals," *arXiv e-prints arXiv:1908.11797*, Aug.

[46] P.-C. Hsieh, I.-H. Hou, and X. Liu, "Delay-optimal scheduling for queueing systems with switching overhead," *arXiv e-prints arXiv:1701.03831*, Jan. 2017.

[47] S. Wang and N. Shroff, "Towards fast-convergence, low-delay and low-complexity network optimization," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 1, Dec. 2017.

[48] Y. Sun, C. E. Koksal, and N. B. Shroff, "On delay-optimal scheduling in queueng systems with replications," *arXiv e-prints arXiv:1603.07322*, Feb. 2017.

[49] J. Liu, A. Eryilmaz, N. B. Shroff, and E. S. Bentley, "Heavy-ball: A new approach to tame delay and convergence in wireless network optimization," in *Proc. INFOCOM*, pp. 1–9, Apr. 2016.

[50] M. J. Neely, "Delay-based network utility maximization," *IEEE/ACM Trans. Netw.*, vol. 21, pp. 41–54, Feb. 2013.

[51] B. Ji, C. Joo, and N. B. Shroff, "Delay-based back-pressure scheduling in multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 21, pp. 1539–1552, Oct. 2013.

[52] L. Hunag, S. Moeller, M. J. Neely, and B. Krishnamachari, "LIFO-backpressure achieves near-optimal utility-delay tradeoff," *IEEE/ACM Trans. Netw.*, vol. 21, Jun. 2013.

[53] G. R. Gupta and N. B. Shroff, "Delay analysis and optimality of scheduling policies for multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 19, Feb. 2011.

[54] M. J. Neely, *Network Optimization: Notes and Exercies*. http://www-bcf.usc.edu/ mjneely/, Updates Fall 2018.

[55] R. Durrett, *Probability: Theory and Examples*. Cambridge University Press, 4 ed., 2010.

[56] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Prentice Hall, 2 ed., 1992.

[57] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*. Prentice Hall, 1 ed., 1989.

[58] V. Gupta, M. Harchol-Balter, J. G. Dai, and B. Zwart, "On the inapproximability of m/g/k: why two moments of job size distribution are not enough," *Queueing Systems*, vol. 64, pp. 5–48, Jan. 2010.
[59] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

Rajat Talak is a Postdoctoral Associate at SPARK-lab in the Department of Aeronautics and Astronautics, at the Massachusetts Institute of Technology. He recently obtained a Ph.D. degree in networked autonomy from the Massachusetts Institute of Technology, where he worked with Prof. Eytan Modiano and Prof. Sertac Karaman. Prior to this, he has obtained a B.Tech. degree in electronics and communication engineering from the National Institute of Technology Karnataka, India, in 2010; and a M.Sc. degree from the department of electrical communication engineering at the Indian Institute of Science, Bengaluru, India in 2013, where he was awarded the prestigious Prof. F M Mowdawalla medal for his thesis. His research interests are in the field of multi-agent systems, communication networks, information freshness, machine learning, and autonomy. His recent focus has been towards studying age of information and guaranteeing information freshness in wireless networks. His other notable work includes developing a new theoretical framework of uncertainty variables for machine learning and robotic perception. He is the recipient of ACM MobiHoc 2018 best paper award.

Eytan Modiano is Professor in the Department of Aeronautics and Astronautics and Associate Director of the Laboratory for Information and Decision Systems (LIDS) at MIT. Prior to Joining the faculty at MIT in 1999, he was a Naval Research Laboratory Fellow between 1987 and 1992, a National Research Council Post Doctoral Fellow during 1992-1993, and a member of the technical staff at MIT Lincoln Laboratory between 1993 and 1999. Eytan Modiano received his B.S. degree in Electrical Engineering and Computer Science from the University of Connecticut at Storrs in 1986 and his M.S. and PhD degrees, both in Electrical Engineering, from the University of Maryland, College Park, MD, in 1989 and 1992 respectively.

His research is on modeling, analysis and design of communication networks and protocols. He is the co-recipient of the Infocom 2018 Best paper award, the MobiHoc 2018 best paper award, the MobiHoc 2016 best paper award, the Wiopt 2013 best paper award, and the Sigmetrics 2006 best paper award. He is the Editor-in-Chief for IEEE/ACM Transactions on Networking, and served as Associate Editor for IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking. He was the Technical Program co-chair for IEEE Wiopt 2006, IEEE Infocom 2007, ACM MobiHoc 2007, and DRCN 2015. He had served on the IEEE Fellows committee in 2014 and 2015, and is a Fellow of the IEEE and an Associate Fellow of the AIAA.