# When a Heavy Tailed Service Minimizes Age of Information

Rajat Talak, Sertac Karaman, and Eytan Modiano

Abstract—Age-of-information (AoI) is a newly proposed performance metric of information freshness. It differs from the traditional delay metric, because it is destination centric and measures the time that elapsed since the last received fresh information update was generated at the source. We show that AoI and packet delay differ in a fundamental way in certain systems, i.e. minimizing one can imply maximizing the other.

We consider two queueing systems, namely a single server last come first serve queue with preemptive service (LCFSp) and  $G/G/\infty$  queue, and show that a heavy tailed service distribution, that results in the worst case packet delay or variance in packet delay, respectively, minimizes AoI. For the specific case of M/G/1 LCFSp and  $G/G/\infty$  queue, we also prove that deterministic service, that minimizes packet delay and variance in packet delay, respectively, results in the worst case AoI.

## I. INTRODUCTION

A source generates fresh information updates, encodes it into a time stamped update packet, and sends it across a network to the destination. The network, composed of several routers or nodes, forwards the update packets, along an appropriate path, to the destination. The routers may queue the update packets, and even drop them at times depending on the network conditions. The goal for the network designer is to ensure that the destination receives fresh information.

Age of information (AoI) is a newly proposed metric for information freshness [1], [2], that measures the time that elapses since the last received fresh update was generated at the source. A typical evolution of AoI for a single sourcedestination system is shown in Figure 1. The AoI increases linearly in time, until the destination receives a fresh packet. Upon reception of a fresh packet *i*, at time  $t'_i$ , the AoI drops to the time since packet *i* was generated, which is  $t'_i - t_i$ ; here  $t_i$  is the time of generation of packet *i*.

AoI was first studied for the first come first serve (FCFS) M/M/1, M/D/1, and D/M/1 queues in [2]. Since then, AoI has been analyzed for several queueing systems [2]–[15], with the goal to minimize AoI. Two time average metrics of AoI, namely, peak and average age are generally considered. Peak age for FCFS G/G/1, M/G/1 and multi-class M/G/1 queueing systems was analyzed in [3], while the discrete time FCFS queue was studied in [4]. Preemptive and non-preemptive last come first serve (LCFS) queue with Poisson arrival and Gamma distributed service was analyzed in [5].

Age for M/M/2 and M/M/ $\infty$  systems was studied in [7], [8] to demonstrate the advantage of having parallel servers, while [11], analyzed parallel LCFS queues with preemptive





Fig. 1. Age evolution in time. Update packets generated at times  $t_i$  and received, by the destination, at times  $t'_i$ . Packet 3 is received out of order, and thus, doesn't contribute to age.

service (LCFSp). Average age for a series of LCFSp queues in tandem was analysed in [13], [16]. Complexity of extending the traditional queuing theory analysis to analyzing multi-hop, multi-server systems has lead [12] to propose stochastic hybrid system method to compute average age, and its moments.

Packet delay, and its variants, have traditionally been considered as measures of communication latency. Optimizing for packet delay in a network, however, is known to be a hard problem. For a single server system, it is known that less variability in service time distribution usually improves packet delay [17], while a heavy tailed service worsens it. The same is true for age metrics, in the single server FCFS queue [9], [18]. In [19], it was proved that the minimum age and delay can be achieved under LCFSp service discipline, when the service times are exponentially distributed. LCFSp is also known to reduce the variance in packet delay [20]. In this work, however, we provide two instances of queueing systems, for which AoI and packet delay differ in a fundamental way, and minimizing one can imply maximizing the other.

We consider a single server G/G/1 LCFS queue with preemptive service and an infinite server G/G/ $\infty$  queue. The arrivals are modeled as a renewal process, while the service is independent and identically distributed across update packets. For both, G/G/1 LCFS queue with preemptive service and G/G/ $\infty$  queue, we show that three heavy tailed service time distributions, namely Pareto, log-normal, and Weibull, minimize AoI. For the specific case of M/G/1 LCFS queue with preemptive service, we show that deterministic service, which minimizes packet delay, results in the worst case age. Similarly, for the G/G/ $\infty$  queue, we show that deterministic service, that minimizes variability in packet delay, maximizes average age, across all service time distributions. In most of the literature, an age metric is analyzed and optimized over queue scheduling discipline, update generation and service rate. This, and its extended works [18], [21], is among the first work to consider age minimization over the space of inter-generation and service time distributions, and bring out a fundamental difference between age and delay.

*Organization:* A generic definition of AoI, peak age, and average age is provided in Section II. LCFS queues with preemptive service is considered in Section III and the infinite server queue  $G/G/\infty$  is considered in Section IV. We conclude in Section V.

# II. Age of Information

Let a source generate update packets at times  $t_1, t_2, \ldots$  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 for time prior to its generation  $t < t_i$ . The generated packet traverses a queueing system, to reach the destination. Let the 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.

Age of information at the destination node, at time t, is defined as the minimum age across all received packets up to time t:

$$A(t) = \min_{i \in \mathcal{P}(t)} A^{i}(t), \tag{1}$$

where  $\mathcal{P}(t) \subset \{1, 2, 3, ...\}$  denotes the set of packets received by the destination, up to time t. Age A(t) increases linearly, till the destination receives an *informative packet* [7], whose age is less than the age at the destination. The age A(t) is then set to the age of this informative packet, and then continues to increase linearly till the next reception.

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

$$A^{\text{ave}} = \limsup_{T \to \infty} \mathbb{E}\left[\frac{1}{T} \int_0^T A(t) dt\right],\tag{2}$$

where the expectation is over the packet generation and packet service processes. Notice that the age A(t) peaks just before reception of an informative packet. Peak age is defined to be the average of all such peaks:

$$A^{\mathsf{p}} = \limsup_{T \to \infty} \mathbb{E}\left[\frac{1}{N(T)} \sum_{k=1}^{N(T)} P^{k}\right],\tag{3}$$

where  $P^k$  denotes the *k*th peak and N(T) denotes the number of peaks till time *T*. The expectation is, again, over the packet generation and packet service processes. The systems we consider will bear the property that  $N(T) \to \infty$  almost surely as  $T \to \infty$ .

In the following sections, we analyze and optimize peak and average age for the single server, G/G/1 LCFS queue with preemptive service, and the average age for the G/G/ $\infty$  queue.



Fig. 2. 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. Packets are generated according to a Poisson process. The service rate  $\mu = 1$ , while the packet generation rate  $\lambda$  varies from 0.5 to 0.99.

## **III. LCFS QUEUES**

Consider a LCFS G/G/1 queue with preemptive service, in which a newly arrived packet gets priority for service immediately. Update packets are generated according to a renewal process, with inter-generation times distributed according to  $F_X$ . The service times are distributed according to  $F_S$ , i.i.d. across packets. Next, we derive explicit expressions for peak and average age for general inter-generation and service time distributions.

Lemma 1: For the LCFS G/G/1 queue, the peak and average age is given by

and

$$\begin{split} A^{\mathrm{p}}_{\mathrm{G/G/1}} &= \frac{\mathbb{E}\left[X\right]}{\mathbf{P}\left[S < X\right]} + \frac{\mathbb{E}\left[S\mathbb{I}_{S < X}\right]}{\mathbf{P}\left[S < X\right]},\\ A^{\mathrm{ave}}_{\mathrm{G/G/1}} &= \frac{1}{2}\frac{\mathbb{E}\left[X^{2}\right]}{\mathbb{E}\left[X\right]} + \frac{\mathbb{E}\left[\min\left(X,S\right)\right]}{\mathbf{P}\left[S < X\right]}, \end{split}$$

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

*Proof:* See our technical report [18].

We now prove that a heavy tailed continuous service time distribution minimizes both peak and average age. In Figure 2, we plot average age as a function of packet generation rates  $\lambda$ , 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

$$F_{S}(s) = \begin{cases} 1 - \left(\frac{\theta(\alpha)}{s}\right)^{\alpha} & \text{if } s \ge \theta(\alpha) \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

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

We observe in Figure 2 that the Pareto service yields better age than the exponential and deterministic service. Furthermore, observe that the heavier the tail of the Pareto distribution, i.e. the closer  $\alpha$  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.

Similar behavior is observed for two other heavy tailed service distributions: log-normal and Weibull. Log-normal service distribution, with mean  $1/\mu$ , is given by:

$$S = \exp\left\{-\log\mu - \frac{\sigma^2}{2} + \sigma N\right\},\tag{5}$$

where  $N \sim \mathcal{N}(0,1)$  is the standard normal distribution. Weibull service distribution, with mean  $1/\mu$ , is given by:

$$F_S(s) = 1 - e^{-(s/\beta)^{\kappa}},$$
 (6)

for all  $s \ge 0$ , 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. The parameters  $\sigma$  and  $\kappa$  determine the tail of the log-normal and Weibull service time distributions, respectively. Higher  $\sigma$  and lower  $\kappa$  imply heavier tails.

We now prove simple lower-bounds on the peak and average age, and show that the peak and average age approaches the lower-bound for the three heavy tailed service time distributions.

**Theorem** *1:* The peak and average age for the LCFS G/G/1 queue are lower bounded by

$$A^{\mathrm{p}}_{\mathrm{G/G/1}}(\lambda,\mu) \geq \mathbb{E}\left[X\right] \quad \text{and} \quad A^{\mathrm{ave}}_{\mathrm{G/G/1}}(\lambda,\mu) \geq \frac{1}{2} \frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]}.$$

Further, both the lower-bounds are simultaneously achieved for

- 1) Pareto distributed service (4) as  $\alpha \rightarrow 1$ ,
- 2) Log-normal distributed service (5) as  $\sigma \to +\infty$ ,
- 3) Weibull distributed service (6) as  $\kappa \to 0$ ,

for all packet generation and service rates,  $\lambda$  and  $\mu$ , respectively.

*Proof:* The lower-bounds follow directly from the age expressions obtained in Lemma 1, and noticing that  $\mathbf{P}[S < X] \leq 1$ . The distributions, namely the Pareto, lognormal, and Weibull, are all parametric distributions parameterized here by  $\alpha$ ,  $\sigma$ , and  $\kappa$ , respectively. We, therefore, prove the following generic result, which gives us a sufficient conditions for the optimality of peak and average age for a general, parametric continuous service time distribution  $F_S$ , parameterized by  $\eta$ .

**Lemma** 2: Let a parametric, continuous, service time distribution, with parameter  $\eta$ , satisfy

1) 
$$\mathbb{E}[S] = 1/\mu$$
,  
2)  $\mathbb{E}[\mathbb{I}_{\{S>x\}}] \to 0 \text{ as } \eta \to \eta^*$ , and  
3)  $\mathbb{E}[S\mathbb{I}_{\{S\leq x\}}] \to 0 \text{ as } \eta \to \eta^*$ ,  
or all  $x > x_0$ , and some  $x_0 > 0$  a

for all  $x \ge x_0$ , and some  $x_0 > 0$  and  $\eta^*$ . Then the peak and average age for LCFS queue, with preemptive service, is minimized by the service time distribution  $F_S$  as  $\eta \to \eta^*$ .

**Proof:** Let for a parametric, continuous, service time distribution  $F_S$  the stated properties hold. Notice that conditions 2 and 3 in the Lemma, along with bounded convergence theorem [22], imply  $\mathbf{P}[S < X] \rightarrow 1$  and  $\mathbb{E}[S\mathbb{I}_{S < X}] \rightarrow 0$  as  $\eta \rightarrow \eta^*$ . This proves that the peak age, given in Lemma 1, approaches its lower-bound:

$$A_{\mathsf{G}/\mathsf{G}/1}^{\mathsf{p}} = \frac{\mathbb{E}\left[X\right]}{\mathbf{P}\left[S < X\right]} + \frac{\mathbb{E}\left[S\mathbb{I}_{S < X}\right]}{\mathbf{P}\left[S < X\right]} \to \mathbb{E}\left[X\right],$$

as  $\eta \to \eta^*$ .

For the average age, notice that

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

Once again, using conditions 2 and 3 in the Lemma, and bounded convergence theorem, we have  $\mathbb{E}\left[X\mathbb{I}_{\{S \geq X\}}\right] \to 0$  and  $\mathbb{E}\left[S\mathbb{I}_{\{S < X\}}\right] \to 0$  as  $\eta \to \eta^*$ . We already know that  $\mathbf{P}\left[S < X\right] \to 1$  as  $\eta \to \eta^*$  from the arguments for peak age optimality. Substituting all this in the average age expression in Lemma 1, we obtain  $A_{G/G/1}^{\text{ave}} \to \frac{1}{2} \frac{\mathbb{E}[X]}{\mathbb{E}[X]}$  as  $\eta \to \eta^*$ . It, therefore, suffices to prove that the sufficient conditions

It, therefore, suffices to prove that the sufficient conditions in Lemma 2 are satisfied by Pareto, log-normal, and Weibull distributions. We know, by definition, that all these distributions are continuous and have mean  $\mathbb{E}[S] = 1/\mu$ . The other conditions are verified in our technical report [18].

# A. M/G/1 Queue

To highlight the contrast between packet delay and AoI metrics, we consider the special case of M/G/1 queue. Here, the update packets are generated according to a Poisson process. The inter-generation times X are exponentially distributed with rate  $\lambda$ . In [5], comparing the performance of LCFS queues M/M/1 and M/D/1 with preemptive service, it was shown numerically that deterministic service performed worse than exponential service. We now show that deterministic service yields the worst peak and average age, across all service time distributions.

Theorem 2: For the LCFS M/G/1 queue,

$$A^{\mathbf{p}}_{\mathbf{M}/\mathbf{G}/\mathbf{I}}(\lambda,\mu) \leq A^{\mathbf{p}}_{\mathbf{M}/\mathbf{D}/\mathbf{I}}(\lambda,\mu) \text{ and} A^{\mathrm{ave}}_{\mathbf{M}/\mathbf{G}/\mathbf{I}}(\lambda,\mu) \leq A^{\mathrm{ave}}_{\mathbf{M}/\mathbf{D}/\mathbf{I}}(\lambda,\mu),$$

for all packet generation and service rates,  $\lambda$  and  $\mu$ , respectively.

## Proof: See technical report [18].

It should be intuitive that if the packets in service are often preempted, then very few packets will complete service on time, and this will result 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  $\mathbf{P}[S > X] = 1 - \mathbb{E}[e^{-\lambda S}]$ , as X is exponentially distributed with rate  $\lambda$ . This can be upper-bounded by  $1 - e^{-\lambda \mathbb{E}[S]} = \mathbf{P}[\mathbb{E}[S] > X]$ , using Jensen's inequality, which is nothing but the probability of preemption under deterministic service:  $S = \mathbb{E}[S]$  almost surely.

Age of Information vs Packet Delay: Comparing age with packet delay for the LCFS queue with preemptive service results in a peculiar conclusion. The packet delay for a LCFS M/G/1 queue is given by [17]:

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

where  $\rho = \frac{\lambda}{\mu}$ . Note that this expression of packet delay  $\mathbb{E}[D]$  is minimized when the service time S is deterministic, namely  $S = \mathbb{E}[S]$  almost surely; follows from Jensen's inequality  $\mathbb{E}[S^2] \geq \mathbb{E}[S]^2$ . However, from Theorem 2 we know that deterministic service time maximizes age. This leads to the conclusion that, for the LCFS M/G/1 queue, *the service time distribution that minimizes delay, maximizes age of information*. It is also noteworthy that the three heavy tailed service time distributions, which minimize peak and average age, have  $\mathbb{E}[S^2] \to +\infty$ , and therefore, result in unbounded packet delay.

In the next section, we consider a queueing system with infinite servers, and show that a service time distribution that minimizes delay variance, maximizes age of information.

## **IV. INFINITE SERVERS**

Next, consider the G/G/ $\infty$  queue, where every newly generated packet is assigned a new server. Let  $F_X$  and  $F_S$  denote the inter-generation and service times, respectively. We focus only on the average age metric, and leave the optimization of peak age for future work. We first derive an expression for average age for the system.

**Lemma** 3: For the  $G/G/\infty$  queue, the average is given by

$$A_{\mathrm{G/G/\infty}}^{\mathrm{ave}} = \frac{1}{2} \frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \mathbb{E}\left[\min_{l \ge 0} \left\{\sum_{k=1}^{l} X_k + S_{l+1}\right\}\right],$$

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:* See our technical report [18].

We now prove that deterministic service yields the worst average age, across all service time distributions. **Theorem** 3: For the infinite server  $G/G/\infty$  system,

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

for all packet generation and service rates,  $\lambda$  and  $\mu,$  respectively.

*Proof:* When the service times are deterministic, i.e.  $S_l = 1/\mu$  a.e., we have  $\min_{l\geq 0} \left\{ \sum_{k=1}^{l} X_k + S_{l+1} \right\} = S_1$ . However, in general we have the inequality  $\min_{l\geq 0} \left\{ \sum_{k=1}^{l} X_k + S_{l+1} \right\} \leq S_1$ . Applying this in the average age expression of Lemma 3, we get the result. See technical report [18] for the detailed proof.

Intuitively, in the  $G/G/\infty$  queue, packets do not get serviced in the same order as they are generated. 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. The packet order is retained exactly under a deterministic service, and therefore, it maximizes age.

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 distributions.

**Theorem** 4: For the infinite server  $G/G/\infty$  system, the average age is lower-bounded by

$$A_{\mathsf{G/G}\infty}^{\mathsf{ave}}(\lambda,\mu) \geq \frac{1}{2} \frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]}.$$

Further, the lower-bound is achieved for

- 1) Pareto distributed service (4) as  $\alpha \rightarrow 1$ ,
- 2) Log-normal distributed service (5) as  $\sigma \to +\infty$ ,
- 3) Weibull distributed service (6) as  $\kappa \to 0$ ,

for all packet generation and service rates,  $\lambda$  and  $\mu,$  respectively.

*Proof:* The lower-bound immediately follows from the average age expression in Lemma 3. We use a similar approach to that followed in the LCFS queue case, we show that the same sufficient conditions as in Lemma 2 suffices for the average age optimality for the  $G/G/\infty$  queue.

**Lemma** 4: Let a parametric, continuous, service time distribution, with parameter  $\eta$ , satisfy

1)  $\mathbb{E}[S] = 1/\mu$ , 2)  $\mathbb{E}[\mathbb{I}_{\{S>x\}}] \to 0$  as  $\eta \to \eta^*$ , and 3)  $\mathbb{E}[S\mathbb{I}_{\{S\leq x\}}] \to 0$  as  $\eta \to \eta^*$ , for all  $x \ge x_0$ , and some  $x_0 > 0$  and  $\eta^*$ . Then the average are for the C/C/a groups is minimized by the equivalent

age for the G/G/ $\infty$  queue is minimized by the service time distribution  $F_S$  as  $\eta \to \eta^*$ .

*Proof:* From Lemma 3, we deduce the following upper-

Authorized licensed use limited to: MIT Libraries. Downloaded on December 17,2020 at 00:14:54 UTC from IEEE Xplore. Restrictions apply.

bound:

$$A_{\mathsf{G/G/\infty}}^{\mathrm{ave}} \leq \frac{1}{2} \frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[X\right]} + \mathbb{E}\left[\min\{S_1, X_1 + S_2\}\right]$$

It, therefore, suffices to argue that  $\mathbb{E}\left[\min\{S_1, X_1 + S_2\}\right] \to 0$  as  $\eta \to \eta^*$ .

Let  $S_1$  and  $S_2$  be independent copies of a parametric, continuously distributed service time random variable, with parameter  $\eta$ , that satisfies all the conditions in the Lemma. Then, by conditions 2 and 3, and the bounded convergence theorem [22], we have  $\mathbb{E}\left[S_1\mathbb{I}_{\{S_1 \leq X_1\}}\right] \to 0$ ,  $\mathbb{E}\left[X_1\mathbb{I}_{\{S_1 > X_1\}}\right] \to 0$ and  $\mathbf{P}\left[S_1 > X_1\right] \to 0$  as  $\eta \to 0$ . This is because  $S_1\mathbb{I}_{\{S_1 \leq X_1\}}, X_1\mathbb{I}_{\{S_1 > X_1\}}$ , and  $\mathbb{I}_{S_1 > X_1}$  are bounded by  $X_1$ ,  $X_1$ , and 1, respectively, all of which are independent of the parameter  $\eta$ . This implies

$$\begin{split} & \mathbb{E}\left[\min\{S_{1}, X_{1} + S_{2}\}\right] \\ &= \mathbb{E}\left[S_{1}\mathbb{I}_{\{S_{1} \leq X_{1}\}}\right] + \mathbb{E}\left[\left[X_{1} + \min\{S_{1} - X_{1}, S_{2}\}\right]\mathbb{I}_{\{S_{1} > X_{1}\}}\right] \\ &\leq \mathbb{E}\left[S_{1}\mathbb{I}_{\{S_{1} \leq X_{1}\}}\right] + \mathbb{E}\left[\left[X_{1} + S_{2}\right]\mathbb{I}_{\{S_{1} > X_{1}\}}\right], \\ &= \mathbb{E}\left[S_{1}\mathbb{I}_{\{S_{1} \leq X_{1}\}}\right] + \mathbb{E}\left[X_{1}\mathbb{I}_{\{S_{1} > X_{1}\}}\right] + \mathbb{E}\left[S_{2}\mathbb{I}_{\{S_{1} > X_{1}\}}\right], \\ &\to 0, \quad \text{as} \quad \eta \to \eta^{*}. \end{split}$$

This proves  $A_{G/G/\infty}^{\text{ave}} \to \frac{1}{2} \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$ , which is the lower-bound, as  $\eta \to \eta^*$ .

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 our technical report [18].

Age of Information vs Packet Delay Variance: For the  $G/G/\infty$  queue as well, a comparison of age with packet delay leads to an interesting conclusion. The packet delay for the  $G/G/\infty$  system, is nothing but the service time S. The variance of packet delay, therefore, is minimized, when S is deterministic. This observation, and Theorem 3, imply that for the  $G/G/\infty$  queue, the service time distribution that reduces packet delay variance, maximizes average age of information. Furthermore, the heavy tailed service time distributions, that minimize average age, results in the worst case, unbounded, variance in packet delay; as for all the age minimizing distributions we have  $\mathbb{E}[S^2] \to +\infty$ .

# V. CONCLUSION

Age of information (AoI) is a newly proposed, destination centric, measure of information freshness, and differs significantly from the traditional latency metric of packet delay. By considering two simple queueing systems, we exposed a fundamental difference between these two performance metrics. We showed that minimizing one can result in the worst case behavior for the other.

For the M/G/1 LCFS queue with preemptive service, we showed that a heavy tailed service minimizes both peak and average AoI. Whereas, deterministic service, which minimizes packet delay, results in the worst case peak and average AoI. For the G/G/ $\infty$  queue, we showed that a heavy tailed service minimizes average age. Whereas, deterministic service, which

minimizes variance in packet delay, yields the worst case average AoI.

### REFERENCES

- S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. SECON*, pp. 350–358, Jun. 2011.
- [2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. INFOCOM*, pp. 2731–2735, Mar. 2012.
- [3] L. Huang and E. Modiano, "Optimizing age-of-information in a multiclass queueing system," in *Proc. ISIT*, pp. 1681–1685, Jun. 2015.
- [4] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. Mobihoc*, Jun. 2018.
- [5] E. Najm and R. Nasser, "Age of information: The gamma awakening," ArXiv e-prints, Apr. 2016.
- [6] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of the information through queues," arXiv e-prints arXiv:1709.04956, Sep. 2017.
- [7] C. Kam, S. Kompella, and A. Ephremides, "Effect of message transmission diversity on status age," in *Proc. ISIT*, pp. 2411–2415, Jun. 2014.
- [8] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. ISIT*, pp. 1583–1587, Jun. 2014.
- [9] 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.
- [10] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems," arXiv e-prints arXiv:1805.12586, Jun. 2018.
- [11] R. D. Yates, "Status updates through networks of parallel servers," in *Proc. ISIT*, pp. 2281–2285, Jun. 2018.
- [12] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," arXiv e-prints arXiv:1806.03487, Jun. 2018.
- [13] R. D. Yates, "Age of information in a network of preemptive servers," arXiv e-prints arXiv:1803.07993, Mar. 2018.
- [14] 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.
- [15] 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.
- [16] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Proc. Allerton*, pp. 486–493, Oct. 2017.
- [17] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Prentice Hall, 2 ed., 1992.
- [18] R. Talak, S. Karaman, and E. Modiano, "Can determinacy minimize age of information?," arXiv e-prints arXiv:1810.04371, Oct. 2018.
- [19] 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.
- [20] 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. 163164, 1962.
- [21] R. Talak and E. Modiano, "Age-delay tradeoffs in single server systems," in *Submitted to ISIT*, Jul. 2019.
- [22] R. Durrett, Probability: Theory and Examples. Cambridge University Press, 4 ed., 2010.