On the Distribution of AoI

Jaya Prakash Champati
IMDEA Networks Institute, Spain

James Gross
KTH Royal Institute of Technology, Sweden

1 On the Distribution of AoI

Initial research on the analysis and minimization of AoI for different queuing systems focused on the average AoI statistics. However, several applications of interest, for example in safety-critical industrial systems, pose stringent requirements on the freshness of packets exchanged over a network. This circumstance, which we further elaborate below, together with related applications, motivates the research on studying the distribution of AoI.

Safety-critical control systems can be certified in two different ways (Leveson 2011). In the first case, *all* individual components of the system are subject to a stringent certification process. However, this requires to quantify failure probabilities of the components under operational conditions, which is not always possible. An alternative approach is to employ so called *safety layers* (Voss 2020), which is especially the case in distributed control systems that are closed over unreliable networks. Safety layers rest on the black channel principle, i.e., no internal state of the underlying communication system can be taken into consideration at run-time. Instead, the safety is validated periodically by a simple message exchange between a pair of safety layers, which are essentially software entities residing on top of layer 7 of the ISO/OSI stack. Example implementations comprise PROFIsafe, SafetyNetP as well as SafetyCAN, which operate according to the above principle (Voss 2020). In more detail, the entities exchange enumerated safety frames with message integrity codes generated from payload as well as shared secrets. If within a predefined safety deadline one of the safety layers involved cannot validate the integrity of the system through receiving a correctly encoded safety frame, then the corresponding safety layer, (a) raises a safety exception which can include switching the involved plant to a fail-safe mode (i.e. stopping the machine), and (b) stops sending safety messages to it's peers. As a consequence, safety can be guaranteed over an unreliable network at the price of (potentially frequent) events that shut down the plant, which do not reflect an immediate safety breach of the system (i.e. the trigger of a safety light barrier).

From this description it becomes clear why the freshness of the involved safety frames, and in particular the statistical distribution of the freshness is of interest. If we consider the messaging between two safety layers as an update process over a communication channel with random (unreliable) service and consider the stochastic process resembled by AoI of this system, then the tail mass of the

AoI distribution for this system with respect to the given deadline is a direct measure of the availability of the control system. The lower the tail mass, the higher will be the availability of the system. Thus, in a second step one might consider the optimization (minimization) of the tail mass, for instance, by tuning the frequency of safety messages. This emphasizes the relevance and potential applicability of AoI analysis with respect to its distribution.

This chapter concerns with fundamental analysis, closed-form expressions, and bounds on the distribution of AoI. We first provide a general formula for the distribution of AoI for a single-source single-server system, and focus primarily on two non-preemptive service systems: 1) GI/GI/1/1 which has no queue, and 2) GI/GI/1/2* which has a unit capacity queue and a new arrival always replaces a packet in the queue. The relevance of these systems in the context of AoI is due to the fact that, in both systems every packet served is an information update packet, i.e., every packet received at the destination has latest generation time. Furthermore, these systems potentially result in lower AoI statistics compared to that of the systems with larger capacity queue, because storing and transmitting old packets from the queue does not reduce AoI (Costa, Codreanu & Ephremides 2016, Kosta, Pappas, Ephremides & Angelakis 2019).

Most of the results of this chapter are from (Champati, Al-Zubaidy & Gross 2019*a*). A general formula for the distribution of AoI was also derived in (Inoue, Masuyama, Takine & Tanaka 2019), where the violation probability is characterized in terms of the distribution of peak AoI, and the distribution of system delay. The distribution of peak AoI for M/M/1/1, M/M/1/2, and M/M/1/2* systems was derived in (Costa et al. 2016) and that of PH/PH/1/1 and M/PH/1/2 systems[1] was derived in (Akar, Dogan & Atay 2020). The authors in (Kesidis, Konstantopoulos & Zazanis 2020) characterized the distribution of AoI for bufferless systems with arrival/service distributions with dependencies. The characterization of the distribution of AoI was used to compute average statistics of non-linear function of AoI for discrete-time queuing systems in (Kosta, Pappas, Ephremides & Angelakis 2020). Distribution of AoI in a tandem network with single source was studied in (Yates 2020), where moment generating function and stationary distribution were derived for Poisson arrivals and exponential service times under preemptive last-come-first-served policy, and in (Champati, Al-Zubaidy & Gross 2020), where periodic sampling rate is optimized to minimize the AoI violation probability.

Next, we present some notation that is used for the analysis in the sequel.

1.1 Notation, Definitions, and System Model

Consider a single source generating status updates or packets which are immediately dispatched to a single-server queueing system as shown in Figure 1.1. The

[1] PH stands for phase-type distribution.

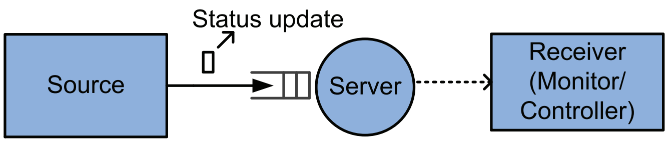

Figure 1.1 A single-source single-server system.

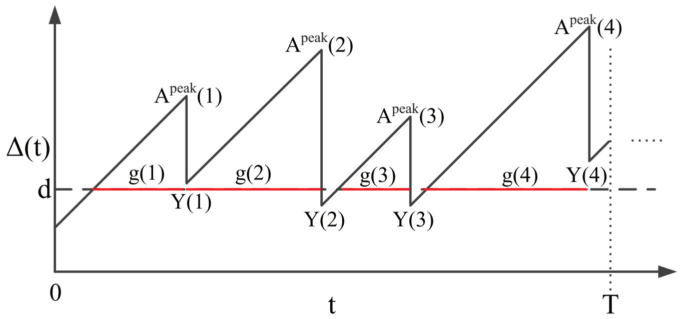

Figure 1.2 A sample path of AoI process.

inter-arrival time between the packets is denoted by the random variable Z with mean-arrival rate $\lambda = \frac{1}{\mathbb{E}[Z]}$. The arriving packets may be stored in a queue and are served by a server using some scheduling policy. We use the random variable X to denote the service time with mean-service rate $\mu = \frac{1}{\mathbb{E}[X]}$. We use packet n to refer to a packet that is nth in the sequence of departures at the server. Let $T_{\mathrm{D}}(n)$ denote the time instant of nth packet departure and $T_{\mathrm{A}}(n)$ denote the corresponding arrival instant.

The AoI metric, denoted by $\Delta(t)$, is defined as:

$$\Delta(t) \triangleq t - \max\{T_{\mathrm{A}}(n) : T_{\mathrm{D}}(n) \leq t\}. \tag{1.1}$$

For a given age limit $d \geq 0$, we are interested in computing the steady-state violation probability or simply violation probability given by

$$\mathbb{P}(\Delta > d) = \lim_{t \to \infty} \mathbb{P}(\Delta(t) > d).$$

The AoI process increases linearly in time with slope one until the departure of information an update packet at the server and it drops to a value equal to the system delay of that packet. Let $\{A^{\mathrm{peak}}(k), k \geq 1\}$ denote the peak AoI process, where $A^{\mathrm{peak}}(k)$ denotes the kth peak of $\Delta(t)$ as shown in Figure 1.2. Let $M(t)$ denote the number of peaks in the interval $(0, t]$. Also, in Figure 1.2 we plot $g(k)$, which is defined as the time duration for which AoI is greater than an age limit

d in the interval between $(k-1)$th peak and kth peak. As mentioned before, the characterization of the violation probability and the bounds presented in this chapter are obtained in terms of $g(k)$. Finally, we define $Y(k) = \Delta(T_{\mathrm{D}}(k))$, i.e. $Y(k)$ is the system delay of an update that departed at time $T_{\mathrm{D}}(k)$.

Note that the AoI peaks occur only at the departure instants of packets, but the converse might not be true as some packet departures might not result in a drop in the AoI. This may happen, for instance, in a GI/GI/1 queue under LCFS scheduling. If there is no new arrival during the service of a packet, the next packet from the queue does not reduce AoI upon its departure as its arrival time would be older than that of the previous departure. As noted before, we refer to packets that reduce AoI upon their departure as information update packets, and use k to index them as it uniquely identifies an information update packet that departs at kth AoI peak. We note that packet n and packet k may not refer to the same packet for $k = n$. We assume that the time average departure rate of information update packets, denoted by ν, is positive and finite. For stationary and ergodic AoI process, we have, almost surely, $\nu = 1/\mathbb{E}[T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)]$.

We study the GI/GI/1/1 and GI/GI/1/2* systems under non-preemptive scheduling. In both systems the inter-arrival times and the service times are i.i.d. As mentioned before, a packet being served always has arrival time later than that of the previous departure. Thus, AoI is reduced at each departure instant and all departures are information update packets. In these systems, packet n and packet k refer to the same packet for $n = k$, $T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)$ represents the inter-departure time, $M(t)$ is the number of departures till time t, and ν is simply the expected departure rate.

We use ω to denote a sample path of AoI, and Ω to denote the set of all sample paths. Let $\gamma(k)$ and $\Gamma(k)$ denote sample-path-wise lower and upper bounds for $g(k)$, i.e.,

$$\gamma(\omega, k) \leq g(\omega, k) \leq \Gamma(\omega, k), \; \forall k \text{ and } \forall \omega \in \Omega. \tag{1.2}$$

In the rest of the chapter, we explicitly drop ω if it is clear from the context.

The list of symbols used in this chapter are summarized in Table 1.1. We use $(x)^+$ for $\max(0, x)$, and $\mathbb{1}\{\cdot\}$ for the indicator function, where $\mathbb{1}\{E\}$ equals one if event E is true, and is zero, otherwise.

1.2 General Formula for the Distribution of AoI

Recall that AoI process increases linearly with slope one until the next information packet departure. Therefore, $A^{\mathrm{peak}}(k)$ can be determined from $\Delta(t)$ at departure of packet $(k-1)$, and the inter-departure time between kth and $(k-1)$th packets.

$$\begin{aligned} A^{\mathrm{peak}}(k) &= T_{\mathrm{D}}(k) - T_{\mathrm{A}}(k-1) \\ &= Y(k-1) + T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1). \end{aligned} \tag{1.3}$$

Table 1.1 List of Symbols

k	Index of an information update packet
$T_{\mathrm{A}}(k)$	Arrival time of packet k
$T_{\mathrm{D}}(k)$	Departure time of packet k
$\check{Z}_k$	Inter-arrival time between packet k and its *previous arrival*
$\hat{Z}_k$	Inter-arrival time between packet k and its *next arrival*
X_k	Service time of packet k
Y_k	System delay of packet k
ν	Expected departure rate of information update packets
I_k	Idle time of the server before starting service for packet k
W_k	Waiting time of packet k
$M(t)$	Number of AoI peaks in the interval $(0, t]$

Analysing $g(k)$ is central to the results presented in this chapter. In the following lemma, we express $g(k)$ in terms of kth AoI peak and inter-departure time between kth and $(k-1)$th packet.

LEMMA 1.1 *Given $d \geq 0$, for any sample path of $\Delta(t)$,*

$$g(k) = \min\{(A^{peak}(k) - d)^+, T_D(k) - T_D(k-1)\},\ \forall k. \tag{1.4}$$

Proof Consider the case where $A^{\mathrm{peak}}(k) \leq d$. For this case $g(k)$ is zero, by definition, which is satisfied by (1.4) as $T_{\mathrm{D}}(k) \geq T_{\mathrm{D}}(k-1)$. For $A^{\mathrm{peak}}(k) > d$ we further consider the following cases.

Case 1: $A^{\mathrm{peak}}(k) > d$ and $A^{\mathrm{peak}}(k) - d > T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)$. Using this in (1.3), we obtain $Y(k-1) > d$. This implies that $\Delta(t) > d$ during the entire interval $[T_{\mathrm{D}}(k-1), T_{\mathrm{D}}(k))$. Therefore, $g(k) = T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)$. This is the case for $g(2)$ in Figure 1.2.

Case 2: $A^{\mathrm{peak}}(k) > d$ and $A^{\mathrm{peak}}(k) - d \leq T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)$. Using this in (1.3), we obtain $Y(k-1) \leq d$. In this case the horizontal line, with y coordinate equal to d, intersects $\Delta(t)$ at some time $t' \in [T_{\mathrm{D}}(k-1), T_{\mathrm{D}}(k))$. Since $\Delta(t)$ increases linearly with slope one, by geometry we obtain $g(k) = T_{\mathrm{D}}(k) - t' = A^{\mathrm{peak}}(k) - d$.

From the above analysis, we conclude that $g(k)$ takes the minimum value of $(A^{\mathrm{peak}}(k) - d)^+$ and $T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)$, and the lemma follows. $\square$

Next, we characterize the violation probability in terms of $g(k)$ in the following theorem.

THEOREM 1.2 *If the AoI process is stationary and ergodic, the AoI violation probability, if exists, is given by*

$$\mathbb{P}(\Delta > d) = \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{M(T)} g(k), \ \ a.s., \tag{1.5}$$

where $g(k)$ is given in (1.4).

Proof Since $\Delta(t)$ is stationary and ergodic, by Birkhoff's ergodic theorem, we have

$$\mathbb{P}(\Delta > d) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{1}\{\Delta(\tau) > d\} d\tau, \ \text{a.s.} \tag{1.6}$$

The RHS above is the fraction of time $\Delta(t)$ is greater than d in a given sample path.

Consider a sample path of $\Delta(t)$ presented in Figure 1.2. Let $\delta(T)$ denote the duration for which $\Delta(t)$ is greater than d after the $M(T)$th peak and before time T. It is easy to see that

$$\begin{aligned} \int_0^T \mathbb{1}\{\Delta(\tau) > d\} d\tau &= \sum_{k=1}^{M(T)} g(k) + \delta(T) \\ \Rightarrow \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{1}\{\Delta(\tau) > d\} d\tau &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{M(T)} g(k). \end{aligned} \tag{1.7}$$

In the last step above, we use $\frac{\delta(T)}{T}$ goes to zero as T goes to infinity owing to the assumption that the departure rate is positive and finite. The result follows by substituting (1.7) in (1.6). □

Theorem 1.2 is quite general in the sense that it holds for any scheduling policy (e.g., FCFS/LCFS, preemptive/non-premptive etc.), general service times (possibly correlated), and general inter-arrival times (possibly correlated), as long as it is ensured that the resulting AoI process is stationary and ergodic. Note that even if the AoI process is stationary and ergodic, the violation probability may not exist. For example, for a D/G/1 system using FCFS the violation probability does not exist if $d < \frac{1}{\lambda}$ (Champati, Al-Zubaidy & Gross 2018).

The challenge in evaluating the infinite summation in the RHS of (1.5) is that the sequence $\{g(k), k \geq 1\}$ is not i.i.d., and we cannot directly use the Strong Law of Large Numbers (SLLN). However, we will later show that quantities involving $g(k)$ have structural independence property, defined below, which enables us to use SLLN.

DEFINITION 1.3 An infinite sequence of random variables $\{\Theta_n, n \geq 1\}$ is structurally independent and identically distributed (s.i.i.d.) iff Θ_n are identically distributed and have the following structural independence: for $1 \leq m < \infty$, Θ_{i+jm} is independent of Θ_{i+km}, for all $1 \leq i \leq m$, $j \geq 0$, $k \geq 0$, and $j \neq k$.

In the results that follow we make use of the following lemma, which extends SLLN for s.i.i.d. random variables.

LEMMA 1.4 *For any sequence $\{\Theta_n, n \geq 1\}$ that is s.i.i.d. according to Definition 1, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} \Theta_n = \mathbb{E}[\Theta], \text{ a.s.},$$

where $\mathbb{E}[\Theta] = \mathbb{E}[\Theta_n]$ for all n.

Proof The proof is based on partitioning the sum into multiple terms which themselves are infinite sums of i.i.d. random variables and then apply SLLN for these summations.

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} \Theta_n &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{\lfloor \frac{N-i+m}{m} \rfloor} \Theta_{i+(j-1)m} \\
&= \frac{1}{m} \sum_{i=1}^{m} \lim_{N\to\infty} \frac{\lfloor \frac{N-i+m}{m} \rfloor}{N/m} \sum_{j=1}^{\lfloor \frac{N-i+m}{m} \rfloor} \frac{\Theta_{i+(j-1)m}}{\lfloor \frac{N-i+m}{m} \rfloor} \\
&= \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}[\Theta] = \mathbb{E}[\Theta], \text{ a.s.}
\end{aligned}$$

In the third step above, we have used SLLN as $\{\Theta_{i+(j-1)m}, j \geq 1\}$ are i.i.d. (Definition 1.3), and $\lfloor \frac{N-i+m}{m} \rfloor$ differs from $\frac{N}{m}$ by utmost 1. ☐

THEOREM 1.5 *Given age limit $d \geq 0$, $\lambda > 0$, $0 < \mathbb{E}[X] = \frac{1}{\mu} < \infty$, $\{g(k), k \geq 1\}$ are s.i.i.d., and $\{T_D(k) - T_D(k-1), k \geq 1\}$ are s.i.i.d., then*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{M(T)} g(k) = \nu \mathbb{E}[g(k)], \text{ a.s.}$$

Proof We have

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{M(T)} g(k) = \lim_{T\to\infty} \frac{M(T)}{T} \cdot \frac{\sum_{k=1}^{M(T)} g(k)}{M(T)}. \tag{1.8}$$

Since $\lambda > 0$ and $\mathbb{E}[X] < \infty$, $M(T)$ approaches infinity, almost surely, as T approaches infinity, and we obtain,

$$\lim_{T\to\infty} \frac{T}{M(T)} = \lim_{T\to\infty} \sum_{k=1}^{M(T)} (T_\mathrm{D}(k) - T_\mathrm{D}(k-1))/M(T).$$

Since $\{T_\mathrm{D}(k) - T_\mathrm{D}(k-1), k \geq 1\}$ are s.i.i.d., from Lemma 1.4 we have

$$\lim_{T\to\infty} \frac{T}{M(T)} = \mathbb{E}[T_\mathrm{D}(k) - T_\mathrm{D}(k-1)], \text{ a.s.} \tag{1.9}$$

Similarly, we invoke Lemma 1.4 for $\{g(k), k \geq 1\}$ and obtain

$$\lim_{M(T)\to\infty} \frac{\sum_{k=1}^{M(T)} g(k)}{M(T)} = \mathbb{E}[g(k)], \text{ a.s.} \tag{1.10}$$

The result follows by substituting (1.9) and (1.10) in (1.8). □

Theorem 1.5 can be seen as an extension of renewal reward theorem for s.i.i.d. renewals and rewards. Later, we use the theorem to derive exact expressions for the violation probability for the D/GI/1/1 and M/GI/1/1 systems.

1.2.1 An Alternative Formula

In this section, we derive an alternative general formula (due to (Inoue et al. 2019)). It characterizes the distribution of AoI in terms of the distribution of the system delay and the distribution of the peak AoI and is presented in the following theorem.

THEOREM 1.6 *For $\nu \in (0, \infty)$, assume that $\lim_{t\to\infty} M(t)/t = \nu$, then for stationary and ergodic AoI process, the distribution of AoI, if exists, is given by*

$$\mathbb{P}(\Delta \leq d) = \nu \int_0^d (\mathbb{P}(Y \leq u) - \mathbb{P}(A^{peak} \leq u))du,$$

$\mathbb{P}(Y \leq u)$ and $\mathbb{P}(A^{peak} \leq u)$ are the steady-state distributions of the system delay and peak AoI processes, respectively.

Proof Let $h(k)$ denote the time duration for which AoI is *smaller* than the age limit d in the interval between $(k-1)$th peak and kth peak. Following the analysis in the proof of Theorem 1.2, we obtain

$$\begin{aligned}\mathbb{P}(\Delta \leq d) &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{M(T)} h(k), \text{ a.s.} \\ &= \lim_{K\to\infty} \frac{\nu}{K} \sum_{k=1}^{K} h(k), \text{ a.s.} \end{aligned} \tag{1.11}$$

We obtain the last equality above by using $M(T)$ goes to infinity as T goes to infinity (since ν is positive and finite), $\lim_{t\to\infty} M(t)/t = \nu$, and replacing $M(T)$ with K. Also, from Birkhoff's ergodic theorem, we obtain

$$\mathbb{P}(Y \leq u) = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\{Y(k) \leq u\}, \text{ a.s.} \tag{1.12}$$

$$\mathbb{P}(A^{\text{peak}} \leq u) = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\{A^{\text{peak}}(k) \leq u\}, \text{ a.s.} \tag{1.13}$$

By the definition of $h(k)$, we have

$$
\begin{aligned}
h(k) &= \int_{T_{\mathrm{D}}(k-1)}^{T_{\mathrm{D}}(k)} \mathbb{1}\{\Delta(t) \le d\} dt \\
&= \int_{T_{\mathrm{D}}(k-1)}^{T_{\mathrm{D}}(k)} \mathbb{1}\{Y(k-1) + t - T_{\mathrm{D}}(k-1) \le d\} dt \\
&= \int_{Y(k-1)}^{A^{\mathrm{peak}}(k)} \mathbb{1}\{u \le d\} du \\
&= \int_{0}^{\infty} \mathbb{1}\{u \le d\} \mathbb{1}\{Y(k-1) \le u\} \mathbb{1}\{A^{\mathrm{peak}}(k) > u\} du \\
&= \int_{0}^{d} \mathbb{1}\{Y(k-1) \le u\}(1 - \mathbb{1}\{A^{\mathrm{peak}}(k) \le u\}) du \\
&= \int_{0}^{d} (\mathbb{1}\{Y(k-1) \le u\} - \mathbb{1}\{A^{\mathrm{peak}}(k) \le u\}) du. \qquad (1.14)
\end{aligned}
$$

In the third step above, we have used change of variable and in the last step we have used $\mathbb{1}\{A^{\mathrm{peak}}(k) \le u\} = \mathbb{1}\{Y(k-1) \le u\}\mathbb{1}\{A^{\mathrm{peak}}(k) \le u\}$. We substitute (1.14) in (1.11), use bounded convergence theorem, and obtain the final result using (1.12) and (1.13). □

Using the above result, the authors in (Inoue et al. 2019) derived Laplace-Stieltjes transform of the distribution of AoI and thereby the moments of AoI for several systems including FCFS GI/GI/1, non-preemptive LCFS, preemptive LCFS, and the -/-/1/2* system with either exponential service or exponential inter-arrival times (cf. Table II (Inoue et al. 2019)).

In the rest of the chapter, we primarily use the formula in Theorem 1.5. This formula does not necessarily provide computational ease in deriving exact expressions when compared with that of the formula presented in Theorem 1.6. Nevertheless, as we will see later, it does provide an approach to derive worst-case performance guarantees for the upper bounds for the AoI violation probability. Deriving these performance guarantees using the formula in Theorem 1.6 is not known to us.

1.3 GI/GI/1/1

In a GI/GI/1/1 system, packet k is served upon its arrival, which implies $T_{\mathrm{D}}(k) = T_{\mathrm{A}}(k) + X_k$. Further, the inter-departure time is given by $T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1) = X_k + I_k$. We note that this relation is equally valid for the GI/GI/1/2* system. Therefore, for both systems

$$\nu = 1/(\mathbb{E}[X_k] + \mathbb{E}[I_k]). \qquad (1.15)$$

In the following we compute $A^{\text{peak}}(k)$ for a GI/GI/1/1 system.

$$
\begin{aligned}
A^{\text{peak}}(k) &= T_{\text{D}}(k) - T_{\text{A}}(k-1)\\
&= T_{\text{D}}(k) - T_{\text{A}}(k) + T_{\text{A}}(k) - T_{\text{D}}(k-1) + T_{\text{D}}(k-1) - T_{\text{A}}(k-1)\\
&= X_k + I_k + X_{k-1}. \qquad (1.16)
\end{aligned}
$$

The following lemma immediately follows from the above analysis and Lemma 1.1.

LEMMA 1.7 *In a GI/GI/1/1 system, given age limit d, for any sample path of $\Delta(t)$ the corresponding $g(k)$ is given by*

$$g(k) = \min\left\{(X_{k-1} + I_k + X_k - d)^+, X_k + I_k\right\}, \forall k \qquad (1.17)$$

We now provide a general expression for the violation probability in the following theorem.

THEOREM 1.8 *Consider a GI/GI/1/1 system, assuming the AoI process is stationary and ergodic, then for all $d \geq 0$, $\lambda > 0$, and $0 < \mathbb{E}[X] = \frac{1}{\mu} < \infty$, the violation probability, if exists, is given by:*

$$\mathbb{P}(\Delta > d) = \nu \mathbb{E}[g(k)], \quad a.s.,$$

where $g(k)$ is given by (1.17) *and ν is given by* (1.15).

Proof We note that the inter-arrival times $\{T_{\text{A}}(k) - T_{\text{A}}(k-1), k \geq 1\}$ in a GI/GI/1/1 system are i.i.d. To see this, the duration $T_{\text{A}}(k) - T_{\text{A}}(k-1)$ equals the sum of inter-arrival times of all dropped packets and the packet k starting from packet $k-1$, and only depends on the inter-arrival time and the service time of packet $k-1$. Therefore, the start of service of a packet is a renewal instant. This implies I_k are i.i.d. which further implies that $T_{\text{D}}(k) - T_{\text{D}}(k-1)$ are i.i.d. From (1.17) we infer that $g(k)$ are identically distributed random variables, and $g(k+2)$ is independent of the random variables $\{g(n), 1 \leq n \leq k\}$ for all k. Therefore, the sequence $\{g(k), k \geq 1\}$ is s.i.i.d. The result then follows from Theorems 1.2 and 1.5. □

Note that to compute the violation probability, we must compute $\mathbb{E}[g(k)]$. In the derivations that follow, we first compute the distribution of $g(k)$ toward this purpose. The following lemma presents a simplified expression for the distribution of $g(k)$.

LEMMA 1.9 *For a GI/GI/1/1 system,*

$$
\begin{aligned}
\mathbb{P}(g(k) > y) = &\int_0^d \mathbb{P}(X_k + I_k > y - x + d) f_X(x) dx\\
&+ \int_d^\infty \mathbb{P}(X_k + I_k > y) f_X(x) dx
\end{aligned}
$$

Proof From (1.17), we have

$$\mathbb{P}(g(k) > y) = \mathbb{P}\left(\min\{(X_{k-1} + X_k + I_k - d)^+, X_k + I_k\} > y\right)$$

$$
\begin{aligned}
&= \mathbb{P}\left(\max\{0, X_{k-1} + X_k + I_k - d\} > y, X_k + I_k > y\right) \\
&= \mathbb{P}\left((y < 0, X_k + I_k > y) \cup (X_{k-1} + X_k + I_k - d > y, X_k + I_k > y)\right) \\
&= \mathbb{P}\left(X_{k-1} + X_k + I_k - d > y, X_k + I_k > y\right) \\
&= \int_0^\infty \mathbb{P}\left(X_k + I_k > y + d - x, X_k + I_k > y\right) f_X(x)dx \\
&= \int_0^d \mathbb{P}\left(X_k + I_k > y - x + d\right) f_X(x)dx \\
&\qquad + \int_d^\infty \mathbb{P}(X_k + I_k > y) f_X(x)dx.
\end{aligned}
$$

□

Zero-wait policy

In a single-source-single-server queueing system using zero-wait policy, the source generates a packet only when there is a departure. It is easy to see that the statistics of the AoI process for this system will be same as that of GI/GI/1/1 when the input rate approaches infinity. Therefore, the following theorem immediately follows from Theorem 1.8, by substituting $I_k = 0$ as input rate is infinity.

THEOREM 1.10 *For the system using zero-wait policy, the violation probability is given by $\nu\mathbb{E}[g(k)]$, almost surely, where $g(k) = \min\{(X_{k-1} + X_k - d)^+, X_k\}$ and $\nu = \mu$.*

Since the AoI process is non-negative, the expected AoI for zero-wait policy is given by

$$
\mathbb{E}[\Delta(t)] = \int_0^\infty \nu\mathbb{E}[\min\{(X_{k-1} + X_k - y)^+, X_k\}]dy.
$$

Next, we derive exact expressions for AoI violation probability for the D/GI/1/1 and M/GI/1/1 systems.

1.3.1 D/GI/1/1: Exact Expressions

In a D/GI/1/1 system, the inter-arrival time is deterministic and is equal to $\frac{1}{\lambda}$. Intuitively, in a D/GI/1/1 system, we only need to consider the rate region $\lambda \geq \frac{1}{d}$ as AoI cannot be less than $\frac{1}{\lambda}$ when the samples are generated at rate λ. The following lemma asserts this intuition.

LEMMA 1.11 *For the D/GI/1/1 system, given $d \geq 0$ and $\lambda > 0$, the AoI violation probability only exists for $d \geq \frac{1}{\lambda}$.*

Proof We prove that $\mathbb{P}(\Delta > d)$ does not exist when $d < \frac{1}{\lambda}$. Consider the event $\{\Delta(t) > d\}$ at time t. If $d < \frac{1}{\lambda}$, there will be time instances, say $\hat{t}$, for which there is no arrival in the interval $[\hat{t} - d, \hat{t})$. This implies that at $\hat{t}$ the receiver cannot have a packet with arrival time greater than $\hat{t} - d$. Therefore, the event

$\{\Delta(\hat{t}) > d\}$ is true for all such $\hat{t}$. Let $\bar{t}$ denote any time instance $t \neq \hat{t}$, i.e., at $\bar{t}$ there exists an arrival in the interval $[\bar{t} - d, \bar{t})$. Since $d < \frac{1}{\lambda}$, there can be only one arrival in this interval. Therefore, for this case the event $\{\Delta(\hat{t}) > d\}$ is true if either the server is busy, in which case the packet is dropped, or the departure time of this packet exceeds $\bar{t}$.

From the above analysis, we conclude that $\mathbb{P}(\Delta(t) > d)$ depends on the value of t. Specifically, we infer that $\limsup_{t\to\infty} \mathbb{P}(\Delta(t) > d) = 1$, because the event $\{\Delta(\hat{t}) > d\}$ is true for all $\hat{t}$, which occur infinitely often as t goes to infinity. Similarly, we infer that $\liminf_{t\to\infty} \mathbb{P}(\Delta(t) > d) < 1$, because the time instances $\bar{t}$ also occur infinitely often and at these time instances the occurrence of the event $\{\Delta(\bar{t}) > d\}$ is uncertain. Since the limit supremum and limit infimum are not equal $\mathbb{P}(\Delta > d) = \lim_{t\to\infty} \mathbb{P}(\Delta(t) > d)$ does not exist for $d < \frac{1}{\lambda}$. □

We now present a closed form expression for the violation probability in the following theorem.

THEOREM 1.12 *For a D/GI/1/1 system, given* $d \geq \frac{1}{\lambda}$, $\lambda > 0$, *and* $0 < \mathbb{E}[X] = \frac{1}{\mu} < \infty$, *the violation probability is given by* $\nu\mathbb{E}[g(k)]$, *almost surely, where* $g(k)$ *is given by* (1.17), $\nu = \lambda/\mathbb{E}[\lceil \lambda X_k \rceil]$ *and* $I_k = \lceil \lambda X_{k-1} \rceil/\lambda - X_{k-1}$.

Proof Using the results from Lemma 1.7 and Theorem 1.8, it is sufficient to show that $I_k = \frac{\lceil \lambda X_{k-1} \rceil}{\lambda} - X_{k-1}$, which we argue to be true in the following. The time difference between the arrival of packet k and packet $(k-1)$ is given by $\frac{\lceil \lambda X_{k-1} \rceil}{\lambda}$. To see this, the service of packet $k-1$ starts upon its arrival, i.e., at $T_\mathrm{A}(k-1)$. During the service of packet $k-1$ the packets that arrived would be dropped and the packet that arrived immediately after $T_\mathrm{A}(k-1)+X_{k-1}$ is served. The number of arrivals since $T_\mathrm{A}(k-1)$ is given by $\lceil \lambda X_{k-1} \rceil$, and the time elapsed is $\frac{\lceil \lambda X_{k-1} \rceil}{\lambda}$. This implies that the idle time I_k is given by $\frac{\lceil \lambda X_{k-1} \rceil}{\lambda} - X_{k-1}$. □

In the following we compute the expression provided in Theorem 1.12 for exponential-service-time distribution.

THEOREM 1.13 *For a D/M/1/1 queue, given* $d \geq \frac{1}{\lambda}$, $\lambda > 0$, *and* $0 < \mathbb{E}[X] = \frac{1}{\mu} < \infty$, *the violation probability is given by* $\nu\mathbb{E}[g(k)]$, *almost surely, where* $\nu = \lambda(1 - e^{-\mu/\lambda})$ *and*

$$\mathbb{E}[g(k)] = \frac{e^{-\mu\frac{\lceil \lambda d \rceil}{\lambda}}}{\lambda(1 - e^{-\frac{\mu}{\lambda}})} + e^{-\mu\frac{\lfloor \lambda d \rfloor}{\lambda}} \left[\frac{\lceil \lambda d \rceil}{\lambda} - d + \frac{1}{\mu} \right] + \frac{e^{-\mu d}}{\mu} \left((e^{\frac{\mu}{\lambda}} - 1)\lfloor \lambda d \rfloor - 1 \right).$$

Proof In the following we first derive $\mathbb{E}[\lceil \lambda X \rceil]$.

$$\begin{aligned}
\mathbb{E}[\lceil \lambda X \rceil] &= \int_0^\infty \lceil \lambda x \rceil \mu e^{-\mu x} dx \\
&= \sum_{m=1}^{\infty} m \int_{\frac{m-1}{\lambda}}^{\frac{m}{\lambda}} \mu e^{-\mu x} dx \\
&= (e^{\mu/\lambda} - 1) \sum_{m=1}^{\infty} m (e^{-\mu/\lambda})^m \\
&= (e^{\mu/\lambda} - 1) e^{-\mu/\lambda} / (1 - e^{-\mu/\lambda})^2 = 1/(1 - e^{-\mu/\lambda}).
\end{aligned}$$

In the following we compute $\mathbb{P}(g(k) > y)$. Recall that $I_k = \frac{\lceil \lambda X_{k-1} \rceil}{\lambda} - X_{k-1}$ (Theorem 1.8). Using this and Lemma 1.9 we obtain

$$\begin{aligned}
\mathbb{P}(g(k) > y) &= \int_0^d \mathbb{P}\left(X_k + \frac{\lceil \lambda d \rceil}{\lambda} - x > y + d - x \right) f_X(x) dx \\
&\quad + \int_d^\infty \mathbb{P}\left(X_k + \frac{\lceil \lambda d \rceil}{\lambda} - x > y \right) f_X(x) dx \\
&= \int_0^d \mathbb{P}\left(X_k > y + d - \frac{\lceil \lambda d \rceil}{\lambda} \right) f_X(x) dx \\
&\quad + \int_d^\infty \mathbb{P}\left(X_k > y + x - \frac{\lceil \lambda d \rceil}{\lambda} \right) f_X(x) dx \\
&= A + B
\end{aligned}$$

We compute the terms A and B[2], and use $\mathbb{E}[g(k)] = \int_0^\infty \mathbb{P}(g(k) > y) dy$ to obtain the result. □

1.3.2 M/GI/1/1: Exact Expressions

For M/GI/1/1 system, the authors in (Najm, Yates & Soljanin 2017) derived expressions for the expected AoI and the expected peak AoI. For this system we provide an expression for the violation probability of AoI.

THEOREM 1.14 *For an M/GI/1/1 system, $\lambda > 0$, and $0 < \mathbb{E}[X] = \frac{1}{\mu} < \infty$, the violation probability, if exists, is given by $\nu \mathbb{E}[g(k)]$, almost surely, where $g(k)$ is given in* (1.17), $\frac{1}{\nu} = \frac{1}{\lambda} + \frac{1}{\mu}$, *and $I_k \sim Exp(\lambda)$.*

Proof The result follows from Theorem 1.8 and using the fact that in an M/G/1/1 system I_k and the inter-arrival times are identically distributed. □

For the special case of M/M/1/1, we have the following theorem.

THEOREM 1.15 *For the M/M/1/1 system, $\lambda > 0$, and $0 < \mathbb{E}[X] = \frac{1}{\mu} < \infty$,*

[2] The computation of A and B is not shown here as it involves lengthy expressions and can be referred from (Champati, Al-Zubaidy & Gross 2019*b*).

the violation probability, if exists, is given by $\nu\mathbb{E}[g(k)]$*, almost surely, where* $\frac{1}{\nu} = \frac{1}{\lambda} + \frac{1}{\mu}$*, and*

$$\mathbb{E}[g(k)] = \begin{cases} \frac{\mu^2(e^{-\lambda d} - e^{-\mu d})}{\lambda(\mu-\lambda)^2} + e^{-\mu d}\left(\frac{1}{\lambda} + \frac{1}{\mu} - \frac{\lambda d}{\mu-\lambda}\right) & \lambda \neq \mu, \\ \frac{\mu e^{-\mu d}}{2}\left(d + \frac{2}{\mu}\right)^2 & \lambda = \mu. \end{cases}$$

Proof Since $X_k \sim Exp(\mu)$ and $I_k \sim Exp(\lambda)$, we have

$$\mathbb{P}(X_k + I_k > y) = \begin{cases} \frac{\mu e^{-\lambda(y-x+d)} - \lambda e^{-\mu(y-x+d)}}{\mu-\lambda} & \lambda \neq \mu, \\ (1+\mu y)e^{-\mu y} & \lambda = \mu. \end{cases} \tag{1.18}$$

In the following we compute the distribution of $g(k)$ by substituting (1.18) in $\mathbb{P}(g(k) > y)$ given in Lemma 1.9.

Case 1: $\mu \neq \lambda$. For this case, we have

$$\begin{aligned} \mathbb{P}(g(k) > y) &= \int_0^d \frac{\mu e^{-\lambda(y-x+d)} - \lambda e^{-\mu(y-x+d)}}{\mu - \lambda} f_X(x)dx \\ &\quad + \frac{\mu e^{-\lambda y} - \lambda e^{-\mu y}}{\mu - \lambda} \int_d^\infty f_X(x)dx \\ &= \frac{\mu e^{-\mu(y+d)}}{\mu - \lambda}\left[\frac{\mu(e^{d(\lambda-\mu)} - 1)}{\lambda - \mu} - \lambda d\right] + \frac{(\mu e^{-\lambda y} - \lambda e^{-\mu y})e^{-\mu d}}{\mu - \lambda}. \end{aligned}$$

Integrating the above expression over y, we obtain the desired result.

Case 2: $\mu = \lambda$. For this case, we have

$$\begin{aligned} \mathbb{P}(g(k) > y) &= \int_0^d (1 + \mu(y - x + d))e^{-\mu(y-x+d)} f_X(x)dx \\ &\quad + \int_d^\infty (1 + \mu y)e^{-\mu y} f_X(x)dx \\ &= \mu e^{-\mu(y+d)} \int_0^d (1 + \mu(y - x + d))dx + (1 + \mu y)e^{-\mu(y+d)} \\ &= \mu e^{-\mu(y+d)}\left[(1 + \mu(y + d))d - \frac{\mu d^2}{2}\right] + (1 + \mu y)e^{-\mu(y+d)} \\ &= \mu d e^{-\mu(y+d)}\left[1 + \mu\left(y + \frac{d}{2}\right)\right] + (1 + \mu y)e^{-\mu(y+d)} \\ &= (\mu d + 1)(1 + \mu y)e^{-\mu(y+d)} + \frac{\mu^2 d^2}{2} e^{-\mu(y+d)}. \end{aligned}$$

Therefore, integrating the above expression over y, we obtain

$$\begin{aligned} \mathbb{E}[g(x)] &= e^{-\mu d}\left[(\mu d + 1)\frac{2}{\mu} + \frac{\mu d^2}{2}\right] \\ &= \frac{\mu e^{-\mu d}}{2}\left(d + \frac{2}{\mu}\right)^2. \end{aligned}$$

□

In the following, we illustrate the computation of the distribution for the case $\lambda = \mu$ using of the formula in Theorem 1.6. For this case, we have $\nu = \frac{\mu}{2}$. The system delay of packet k is equal to X_k, and therefore $Y_k = X_k$ for all k, and Y has the same distribution as the service time. Recall that, $A_k^{\text{peak}} = X_{k-1}+I_k+X_k$. Since $I_k \sim Exp(\lambda)$ and $\lambda = \mu$, peak AoI has Erlang distribution with scale parameter 3 and rate parameter μ. We have

$$\begin{aligned}
\mathbb{P}(\Delta \le d) &= \nu \int_0^d (\mathbb{P}(Y \le x) - \mathbb{P}(A^{\text{peak}} \le x))dx \\
&= \frac{\mu}{2} \int_0^d \left(1 - e^{-\mu x} - (1 - e^{-\mu x} - xe^{-\mu x} - \frac{x^2 e^{-\mu x}}{2})\right) dx \\
&= \frac{\mu}{2} \int_0^d \left(xe^{-\mu x} + \frac{x^2 e^{-\mu x}}{2}\right) dx \\
&= 1 - \frac{\mu^2 e^{-\mu d}}{4} \left(d + \frac{2}{\mu}\right)^2 .
\end{aligned}$$

In this case, the number of steps are less for computing the expression using the above formula because the distribution of peak AoI is readily available. However, in general, computing the distribution of peak AoI presents an additional step.

In the following theorem, we derive the violation probability for the system with zero-wait policy and exponentially distributed service times.

THEOREM 1.16 *For the system with zero-wait policy and exponentially distributed service times, the violation probability is given by*

$$\mathbb{P}(\Delta > d) = (1 + \mu d)e^{-\mu d}, \ \ a.s. \tag{1.19}$$

Proof The result can be obtained from Theorem 1.15 by utilizing the fact that the statistics of this system will be same as that for $M/M/1/1$ when λ approaches infinity. □

Interestingly, the distribution in (1.19) is gamma distribution with shape parameter 2 and scale parameter $\frac{1}{\mu}$. Further, the expected AoI in this case is $\frac{2}{\mu}$, a result reported in (Kaul, Yates & Gruteser 2012, Costa et al. 2016).

1.4 Upper Bounds

As one can expect $g(k)$ and $T_{\text{D}}(k) - T_{\text{D}}(k-1)$ depend on the idle time I_k and waiting time W_k in the queuing system. Therefore, computing $\mathbb{E}[g(k)]$ and ν is hard, in general, as the distributions of I_k and W_k become intractable for general inter-arrival time and service-time distributions. To this end, in the following theorem we present a result that is useful in deriving upper bounds for the violation probability and only requires the AoI process to be stationary.

THEOREM 1.17 *If the AoI process is stationary, then*

$$\mathbb{E}_\omega\left[\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{M(T)}\gamma(\omega,k)\right]\le\mathbb{P}(\Delta>d)\le\mathbb{E}_\omega\left[\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{M(T)}\Gamma(\omega,k)\right].$$

Proof Since $\Delta(t)$ is stationary, we have

$$\mathbb{P}(\Delta(t)>d)=\mathbb{E}_\omega[\mathbb{1}\{\Delta(\omega,t)>d\}],\ \forall t.$$

Therefore, for any t,

$$\begin{aligned}\mathbb{P}(\Delta(t)>d)&=\lim_{T\to\infty}\frac{1}{T}\int_0^T\mathbb{E}_\omega[\mathbb{1}\{\Delta(\omega,t)>d\}]dt\\&=\mathbb{E}_\omega\left[\lim_{T\to\infty}\frac{1}{T}\int_0^T\mathbb{1}\{\Delta(\omega,t)>d\}dt\right]\\&=\mathbb{E}_\omega\left[\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{M(T)}g(\omega,k)\right].\end{aligned}\tag{1.20}$$

Second step above is due to the fact that indicator function is non-negative. The third step is due to the fact that (1.7) is true for any ω. The result follows from (1.20) and (1.2). □

In terms of applicability, Theorem 1.17 is more general than Theorem 1.2 as it does not require ergodicity of the AoI process. Following Theorem 1.17, we strive to obtain upper bounds for the violation probability for GI/GI/1/1 and GI/GI/1/2* systems by finding bounds for $g(k)$.

In the following we establish a lower bound for $g(k)$ that is applicable to any single-source-single-server queueing system.

LEMMA 1.18 *For a single-source-single-server queuing system, it is true that $g(k)\ge\gamma^*(k)$, for all k, where*

$$\gamma^*(k)=\min\{(X_k+X_{k-1}+I_k-d)^+,X_k+I_k\}.$$

Proof For a single-server system it is easy to see that the inter-departure time between information update packets is at least the service time of a packet and idle time before its service started, i.e.,

$$T_{\mathrm{D}}(k)-T_{\mathrm{D}}(k-1)\ge X_k+I_k.\tag{1.21}$$

From (1.3) we have

$$\begin{aligned}A^{\mathrm{peak}}(k)&=T_{\mathrm{D}}(k)-T_{\mathrm{A}}(k-1)\\&\ge T_{\mathrm{D}}(k)-(T_{\mathrm{D}}(k-1)-X_{k-1})\\&\ge X_k+I_k+X_{k-1}\end{aligned}$$

The second step is due to the fact that a packet departure time is at least equal to its arrival time plus its service time. The last step is due to (1.21). □

We use the lower bound in Lemma 1.18 to analyse the performance of the upper bounds derived for the AoI violation probability for GI/GI/1/1 and GI/GI/1/2* systems. Nevertheless, this method is quite general and can be applied to other queueing systems.

1.4.1 Upper Bound for the GI/GI/1/1 system

In this section, we provide an upper bound for the violation probability for the GI/GI/1/1 system, and also analyse its performance. To this end, we first provide an upper bound for $g(k)$ in the following lemma.

LEMMA 1.19 *For a GI/GI/1/1 system, $g(k) \leq \Gamma_1(k)$ for all k, where*

$$\Gamma_1(k) = \min\left\{(X_{k-1} + \check{Z}_k + X_k - d)^+, X_k + \check{Z}_k\right\}. \tag{1.22}$$

Proof Recall that $\check{Z}_k$ is the inter-arrival time between packet k and its previous arrival. Therefore, we have $I_k \leq \check{Z}_k$. The result follows from using this in (1.17). □

Remark 1: In an M/G/1/1 system $\mathbb{E}[\Gamma_1(k)] = \mathbb{E}[g(k)]$, since for this system both I_k and $\check{Z}_k$ have the same distribution $\text{Exp}(\lambda)$. Thus, $\mathbb{E}[\Gamma_1(k)]$ is a tight upper bound for $\mathbb{E}[g(k)]$ for the GI/GI/1/1 system.

The following theorem presents an upper bound Φ_1 for the violation probability.

THEOREM 1.20 *For a GI/GI/1/1 system, given $d > 0$, assuming that the AoI process is stationary, the violation probability is bounded as follows:*

$$\mathbb{P}(\Delta > d) = \nu\mathbb{E}[\gamma^*(k)] \leq \Phi_1,$$

where γ^* is given by Lemma 1.18, and $\Phi_1 = \hat{\nu}\mathbb{E}[\Gamma_1(k)]$, for some $\hat{\nu} \geq \nu$, where ν is given in (1.15).

Proof The equality follows from the fact that $\gamma^*(k)$ is equal to $g(k)$ given in (1.17) for the GI/GI/1/1 system. It is easy to see that $\Gamma_1(k)$ are s.i.i.d., and as noted in the proof of Theorem 1.8, $T_{\mathrm{D}}(k) - T_{\mathrm{D}}(k-1)$ are i.i.d. Therefore, from Theorem 1.5 we infer that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{K(T)} \Gamma_1(k) = \nu\mathbb{E}[\Gamma_1(k)], \text{ a.s.}$$

Using the above equation in Theorem 1.17, we obtain $\mathbb{P}(\Delta > d) \leq \nu\mathbb{E}[\Gamma_1(k)]$. The result follows as $\hat{\nu} \geq \nu$. □

We define η below that will be used in describing the worst-case performance of Φ_1.

$$\eta \triangleq \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\nu}. \tag{1.23}$$

In the following theorem we present a worst-case-performance guarantee for Φ_1.

THEOREM 1.21 *For a GI/GI/1/1 system, for a given $\hat{\nu} \geq \nu$, Φ_1 has the following worst-case-performance guarantee .*

$$\Phi_1 \leq \frac{\hat{\nu}}{\nu} \cdot \mathbb{P}(\Delta > d) + \hat{\nu}\eta.$$

Proof Noting that $I_k \leq \check{Z}_k$, we have

$$\begin{aligned}\Gamma_1(k) &= \min\left\{(X_{k-1} + \check{Z}_k + X_k - d)^+, X_k + \check{Z}_k\right\} \\ &\leq \min\left\{(X_{k-1}+I_k+X_k - d)^+, X_k+I_k\right\} + (\check{Z}_k - I_k) \\ &\leq g(k) + (\check{Z}_k - I_k).\end{aligned}$$

Therefore, using Theorem 1.20, we obtain

$$\begin{aligned}\Phi_1 &\leq \hat{\nu}(\mathbb{E}[g(k)] + \mathbb{E}[\check{Z}_k] - \mathbb{E}[I_k]) \\ &= \frac{\hat{\nu}}{\nu} \cdot \mathbb{P}(\Delta(t) > d) + \hat{\nu}\left(\frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\nu}\right).\end{aligned}$$

In the last step above we have used Theorem 1.8 and (1.15). □

From Theorem 1.21, we infer that if $\hat{\nu} = \nu$, i.e., the departure rate is given, then Φ_1 overestimates the violation probability by at most η. We note that $\frac{1}{\lambda}+\frac{1}{\mu} \geq \frac{1}{\nu}$, and the relation holds with equality for an M/GI/1/1 system. Further, ν increases sub-linearly with λ in a GI/GI/1/1 system, in general. For example, $\nu = \lambda(1-e^{-\mu/\lambda})$ for the D/M/1/1 system (Theorem 1.13). Therefore, for a fixed μ, η decreases with λ, in general. In other words, the derived upper bound is tighter at higher utilization. Finally, the worst-case guarantee in Theorem 1.21 is provided for any $d \geq 0$. Therefore, we expect that Φ_1 may not be tight for larger d values for which the violation probability takes smaller values.

We require to compute the value of expected idle time to obtain ν. When ν is not tractable, we propose to use $\hat{\nu} = \min\{\lambda, \mu\}$, a trivial upper bound on the departure rate. We note however that the conclusion about tightness of the upper bound at higher utilization may no longer be valid in this case.

Note that, by replacing I_k using an appropriate quantity one can obtain upper bounds using the formula in Theorem 1.6, but obtaining performance guarantees for those bounds is not straightforward and requires further study.

1.5 The GI/GI/1/2* System

The analysis of a GI/GI/1/2* system follows similar steps to the analysis we have presented for the GI/GI/1/1 system. We first obtain expressions for I_k and $A^{\text{peak}}(k)$, and use them to obtain $g(k)$.

In Figure 1.3, we present a possible sequence of arrivals (in blue) and departures (in red) in a GI/GI/1/2* system. Note that there are no arrivals during the service of packet $(k-1)$. This happens only when $\hat{Z}_{k-1} > W_{k-1} + X_{k-1}$ and

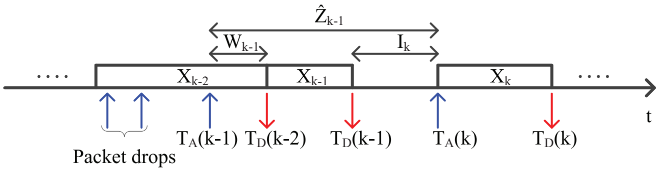

Figure 1.3 An example illustration of arrivals and departures in a GI/GI/1/2* system.

in this case, $I_k = \hat{Z}_{k-1} - W_{k-1} - X_{k-1}$. If $\hat{Z}_{k-1} \leq W_{k-1} + X_{k-1}$, then $I_k = 0$. Therefore, we have

$$I_k = (\hat{Z}_{k-1} - X_{k-1} - W_{k-1})^+. \tag{1.24}$$

Recall that $A^{\text{peak}}(k) = T_{\text{D}}(k) - T_{\text{A}}(k-1)$. From Figure 1.3, it is easy to infer that $A^{\text{peak}}(k) = X_k + X_{k-1} + I_k + W_k$. The following lemma immediately follows from the above analysis and Lemma 1.1.

LEMMA 1.22 *Given $d \geq 0$, for any sample path of $\Delta(t)$ in a GI/GI/1/2* system, we have for all k,*

$$g(k) = \min\left\{(X_k + X_{k-1} + I_k + W_{k-1} - d)^+, X_k + I_k\right\}. \tag{1.25}$$

Unlike the case of the GI/GI/1/1 system, for the GI/GI/1/2* system it is hard to derive a closed-form expression for the violation probability in terms of X_k, X_{k-1}, I_k and W_{k-1}, because $g(k)$, given in (1.25), does not satisfy the s.i.i.d. property. Further, computing the violation probability requires the distributions of both I_k and W_{k-1}. While these quantities can be computed for exponential service or exponential inter-arrival times (cf. (Inoue et al. 2019)), they become intractable for general inter-arrival and service-time distributions. To this end we present upper bounds in the next section.

1.5.1 Upper Bound for the GI/GI/1/2* system

In this subsection we propose an upper bound for the violation probability and analyse its worst-case performance.

LEMMA 1.23 *For a GI/GI/1/2* system, $g(k) \leq \Gamma_2(k)$ for all k, where*

$$\Gamma_2(k) = \min\{(X_k + X_{k-1} + \hat{Z}_{k-1} - d)^+, X_k + (\hat{Z}_{k-1} - X_{k-1})^+\}.$$

Proof Noting the expression for $g(k)$ given in (1.25), it is sufficient to show that $I_k + W_{k-1} \leq \hat{Z}_{k-1}$, and $I_k \leq (\hat{Z}_{k-1} - X_{k-1})^+$. The latter inequality follows from (1.24). The former inequality is obviously true if there are no arrivals during the service of packet $(k-1)$; see Figure 1.3. If there is an arrival during the service of packet $(k-1)$, then $I_k = 0$. In this case $I_k + W_{k-1} = W_{k-1} < \hat{Z}_{k-1}$, since by definition there should be no arrival after packet $(k-1)$ arrived and before its service started. □

In the following theorem we present an upper bound Φ_2 for the violation probability.

THEOREM 1.24 *For a GI/GI/1/2* system, assuming that the AoI process is stationary, the violation probability is bounded by,*

$$\nu\mathbb{E}[\gamma^*(k)] \leq \mathbb{P}(\Delta > d) \leq \Phi_2,$$

where $\Phi_2 = \hat{\nu}\mathbb{E}[\Gamma_2(k)]$, *for some* $\hat{\nu} \geq \nu$.

Proof The proof follows similar steps to the proof of Theorem 1.20 and is omitted. ☐

A worst-case-performance guarantee for Φ_2 is presented in the following theorem.

THEOREM 1.25 *For the GI/GI/1/2* system, for a given* $\hat{\nu} \geq \nu$, Φ_2 *has the following worst-case-performance guarantee.*

$$\Phi_2 \leq \frac{\hat{\nu}}{\nu} \cdot \mathbb{P}(\Delta > d) + \hat{\nu}\eta.$$

Proof It is easy to show that $\Gamma_2(k) \leq \gamma^*(k) + \hat{Z}_{k-1} - I_k$. The rest of the proof follows similar steps as in the proof of Theorem 1.21 and is omitted. ☐

Thus, Φ_2 also overestimates the violation probability by at most η, if ν is given. Therefore, given ν and for a fixed average service, Φ_2 is tighter at higher utilization. Since it is hard to compute ν, in general, in the numerical section we compute Φ_2 using $\hat{\nu} = \min\{\lambda, \mu\}$.

Remark 2: For both GI/GI/1/1 and GI/GI/1/2* systems $\nu = 1/(\mathbb{E}[X_k] + \mathbb{E}[I_k])$, and $g(k)$ for GI/GI/1/1 given by (1.17) seems to be closely related to $g(k)$ for GI/GI/1/2* given by (1.25). Also, one can expect that the idle time in GI/GI/1/2* will be lower compared to that of GI/GI/1/1. However, for a given d, a comparison between the violation probabilities in these systems is non-trivial because of the waiting time in GI/GI/1/2* and higher idle time in GI/GI/1/1.

Remark 3: When the input rate approaches infinity, the inter-arrival time, waiting time, and idle time approach zero. Therefore, the upper bounds Φ_1, Φ_2, and the respective violation probabilities in GI/GI/1/1 and GI/GI/1/2*, all converge to the violation probability in the system using zero-wait policy. Thus, both Φ_1 and Φ_2 are asymptotically tight.

1.6 Numerical Results

In this section, we validate the proposed upper bounds against the violation probability obtained through simulation for selected service-time and inter-arrival-time distributions. For all simulations we set $\mu = 1$ and thus the utilization increases with λ. We use $\lambda = .45$ and $d = 5$ as default values.

We first study the performance of Φ_1 in comparison with overestimation factor

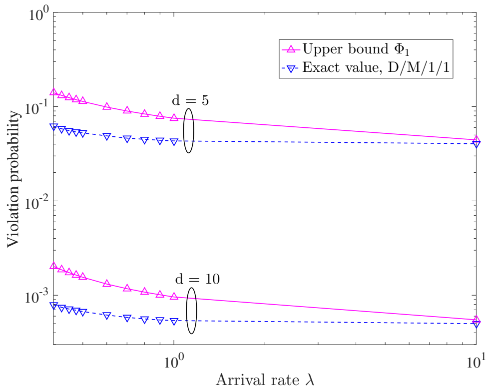

Figure 1.4 Performance of Φ_1 with varying λ, when ν is given, and $\mu = 1$

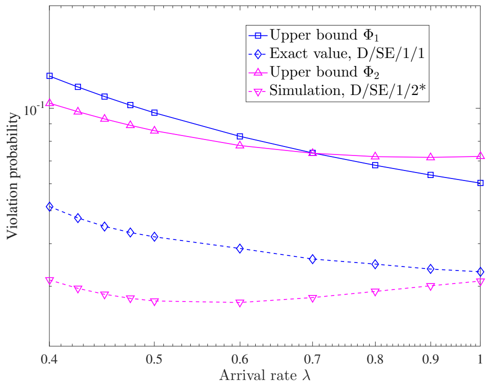

Figure 1.5 Performance of upper bounds with varying λ, $d = 5$, $\mu = 1$, and shift equal to 0.11 .

η, when ν is given. To this end we consider the D/M/1/1 system and compute Φ_1 by setting $\hat{\nu} = \nu = \lambda(1 - e^{-\mu/\lambda})$. In Figure 1.4, we plot Φ_1 against the exact value for the violation probability given in Theorem 1.13. Observe that the gap between Φ_1 and violation probability reduces as the arrival rate increases confirming our initial conclusion that the bound is tighter at higher utilization. Furthermore, Φ_1 approaches the simulated violation probability asymptotically. For $d = 5$ and $\lambda = 0.4$, we compute η to be 0.28, while the actual gap is 0.08. For the same setting, but for $d = 10$, η remains the same while the actual gap is 0.0012. This suggests that the proposed upper bound is much lower than the worst-case-performance guarantee.

Next, we consider two example systems where exact expressions for the distribution of AoI are hard to compute. For both systems, we use $\hat{\nu} = \min(\lambda, \mu)$ to compute Φ_1 and Φ_2. In the first example system, we choose deterministic arrivals and Shifted-Exponential (SE) service times, i.e., D/SE/1/1 and D/SE/1/2*. We set values of d and λ such that $d \geq \frac{1}{\lambda}$, $\mu = 1$ and shift parameter equal to 0.11. In Figures 1.5 and 1.6, we study the performance of the upper bounds,

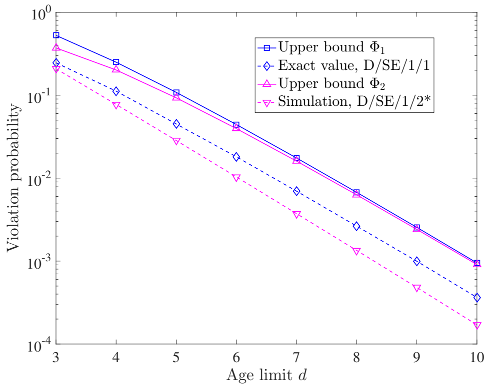

Figure 1.6 Performance of upper bounds with varying d, $\lambda = .45$, $\mu = 1$, and shift equal to 0.11.

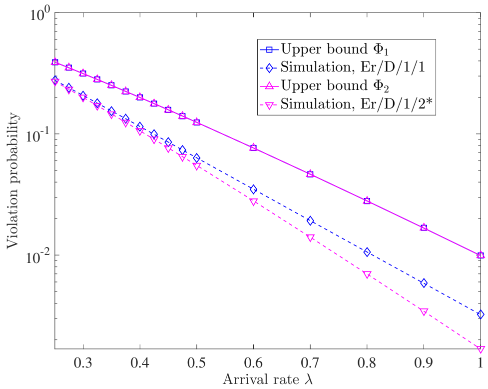

Figure 1.7 Performance of upper bounds with varying λ, Erlang shape parameter equal to 2, $d = 5$, and $\mu = 1$.

presented in Theorems 1.20 and 1.24, for varying arrival rate λ and varying age limit d, respectively. From Figure 1.5, we again observe that the upper bounds are tighter at higher utilization. For $\lambda > 1$ both upper bounds and the violation probabilities converge to 0.029. Interestingly, in contrast to D/SE/1/1 where the violation probability decreases with λ, D/SE/1/2* has minimum violation probability of 0.026 at around $\lambda = 0.6$. From Figure 1.6, we observe that both bounds are tighter at smaller d values. While the decay rate of Φ_1 matches with that of the simulated violation probability, Φ_2 becomes loose as d increases. We conjecture that this is due to the inequality $I_k + W_{k-1} \le \hat{Z}_{k-1}$ that we use to obtain this bound.

In Figures 1.7 and 1.8, we present a comparison for deterministic service and Erlang distributed inter-arrival times, i.e., Er/D/1/1 and Er/D/1/2*. We first note that for the parameter values chosen, Φ_1 and Φ_2 are equal in this case. From Figure 1.7, we observe that the bounds are not tight at larger arrival rate. This can be attributed to the use of $\hat{\nu} = \min(\lambda, \mu)$. From Figure 1.8, we observe that

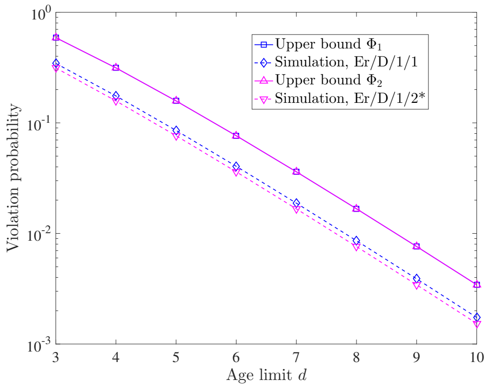

Figure 1.8 Performance of upper bounds with varying d, $\lambda = .45$, Erlang shape parameter equal to 2, and $\mu = 1$.

the decay rate of the bounds matches the decay rate of the violation probabilities. Finally, it is worth noting that, the violation probability in -/-/1/2* is lower than that in -/-/1/1 for the above example systems.

In conclusion, for the considered systems, the upper bounds are well within an order of magnitude from the violation probability. For most cases the decay rate of the proposed bounds follow the decay rate of the simulated violation probability as d increases. Also, the performance of these upper bounds can be improved further by finding non-trivial upper bounds for ν. Thus, we believe that the proposed upper bounds can be useful as first-hand metrics for measuring freshness of status updates in these systems.

References

Akar, N., Dogan, O. & Atay, E. U. (2020), 'Finding the exact distribution of (peak) age of information for queues of ph/ph/1/1 and m/ph/1/2 type', *IEEE Transactions on Communications* pp. 1–1.

Champati, J. P., Al-Zubaidy, H. & Gross, J. (2018), Statistical guarantee optimization for age of information for the D/G/1 queue, *in* 'Proc. IEEE INFOCOM Workshop', pp. 130–135.

Champati, J. P., Al-Zubaidy, H. & Gross, J. (2019*a*), On the distribution of AoI for the GI/GI/1/1 and GI/GI/1/2* systems: Exact expressions and bounds, *in* 'Proc. IEEE INFOCOM', pp. 37–45.

Champati, J. P., Al-Zubaidy, H. & Gross, J. (2019*b*), 'On the distribution of AoI for the GI/GI/1/1 and GI/GI/1/2* systems: Exact expressions and bounds', *CoRR* **abs/1905.04068**.

Champati, J. P., Al-Zubaidy, H. & Gross, J. (2020), 'Statistical guarantee optimization for AoI in single-hop and two-hop FCFS systems with periodic arrivals', *IEEE Transactions on Communications* pp. 1–1.

Costa, M., Codreanu, M. & Ephremides, A. (2016), 'On the age of information in status update systems with packet management', *IEEE Transactions on Information Theory* **62**(4), 1897–1910.

Inoue, Y., Masuyama, H., Takine, T. & Tanaka, T. (2019), 'A general formula for the stationary distribution of the age of information and its application to single-server queues', *IEEE Transactions on Information Theory* **65**(12), 8305–8324.

Kaul, S., Yates, R. & Gruteser, M. (2012), Status updates through queues, *in* 'Proc. Conference on Information Sciences and Systems (CISS)'.

Kesidis, G., Konstantopoulos, T. & Zazanis, M. A. (2020), 'The distribution of age-of-information performance measures for message processing systems', *Queueing Systems: Theory and Applications* **95**(3), 203–250.

Kosta, A., Pappas, N., Ephremides, A. & Angelakis, V. (2019), 'Age of information performance of multiaccess strategies with packet management', *Journal of Communications and Networks* **21**(3), 244–255.

Kosta, A., Pappas, N., Ephremides, A. & Angelakis, V. (2020), Non-linear age of information in a discrete time queue: Stationary distribution and average performance analysis, *in* 'Proc. IEEE ICC', pp. 1–6.

Leveson, N. (2011), *Engineering a Safer World - Systems Thinking Applied To Safety*, MIT Press.

Najm, E., Yates, R. D. & Soljanin, E. (2017), Status updates through M/G/1/1 queues with HARQ, *in* 'Proc. IEEE ISIT', pp. 131–135.

Voss, W. (2020), ‘A comprehensible guide to industrial ethernet’.

Yates, R. D. (2020), ‘The age of information in networks: Moments, distributions, and sampling’, *IEEE Transactions on Information Theory* **66**(9), 5712–5728.