Peak Age of Information Analysis for Real-Time Status Update Systems Based on Martingale Theory

Hangyu Yan, Xuefen Chi, Wanting Yang, Zehui Xiong, *Senior Member, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*—In this paper, we propose a martingale-based method to analyze the peak age of information (PAoI) in real-time status update systems. Specifically, we establish the relationship between the PAoI and the inter-arrival time as well as the service time of status update packets. Martingales are then constructed for the inter-arrival time and service time to capture their stochastic characteristics. Utilizing the constructed martingales, we derive a closed-form expression for the upper bound of the PAoI violation probability under the First-Come-First-Served scheduling policy. It is worth emphasizing that the proposed method is applicable where the service time of status update packets is independently and identically distributed or follows a Markov process. Numerical results verify the accuracy and generality of the proposed method in analyzing PAoI.**

***Index Terms*—Peak age of information, martingale theory, status update systems, Markov process, scheduling policy.**

I. Introduction

As 6G networks become widely deployed, a variety of applications with strict real-time status update requirements have emerged, such as autonomous driving, disaster monitoring, and the industrial Internet of Things (IIoT) [1]–[3]. In these applications, destination decisions and control actions rely on timely received status information [4], [5]. To quantify information freshness at the destination, the age of information (AoI) was introduced in [6], defined as the time elapsed since the generation of the most recently updated information. Subsequently, various AoI-related metrics have been proposed, such as average AoI and peak age of information (PAoI) [7], [8]. This work focuses on PAoI, i.e., the maximum AoI within each update, which directly captures the worst-case status staleness relevant to time-sensitive decision/control.

Due to the inherent randomness of wireless channels and the uncertainty in the arrival times of status update packets (SUPs), achieving deterministic PAoI is both costly and challenging [9]. As a practical alternative, real-time requirements can be expressed by an allowable staleness limit together with an acceptable violation level, and the upper bound of peak age of information violation probability (UB-PAoIVP) provides a tractable bound on the risk that the PAoI exceeds this limit under stochastic conditions. To this end, various studies have focused on deriving closed-form UB-PAoIVP expressions under different assumptions. For instance, [10] derived closed-form expressions for the UB-PAoIVP under the First-Come-First-Served (FCFS) scheduling strategy in M/M/1 and M/D/1 queues. Building on this strategy, [11] proposed an integrated satellite-terrestrial network architecture model and derived the UB-PAoIVP using the Mellin transform and stochastic network calculus. Similarly, [12] investigated the impact of computation and communication delays on the PAoI in mobile edge computing systems, also deriving a closed-form UB-PAoIVP. However, these methods suffer from several limitations. The use of Boolean inequalities in [11] and [12] leads to relatively loose bounds, which may not accurately capture the system's worst-case PAoI performance. Subsequently, [13] conducted a more precise analysis of the relationship between inter-arrival times and service times, and derived a tighter closed-form expression for the UB-PAoIVP. Nonetheless, their method is limited to scenarios where both inter-arrival and service times are independently and identically distributed (i.i.d.) and follow continuous distributions. Moreover, when the service times exhibit Markov properties [14], the methods proposed in [10]–[13] are no longer applicable. These limitations highlight the current lack of a general analytical framework that can accommodate diverse service time models. Therefore, developing a unified method for analyzing the UB-PAoIVP remains an important and unsolved research problem.

Martingale theory is a powerful tool for probabilistic analysis in communication systems. With appropriate transformations, non-martingale processes can be converted into martingales, enabling the use of martingale theorems and Doob's inequality. These tools support the derivation of tight and tractable performance bounds, and have been widely used in delay and backlog analysis [15]. Motivated by these advantages, we propose a martingale-based method for analyzing the PAoI in real-time status update systems. The main contributions are summarized as follows.

- We construct corresponding martingales for the inter-arrival time and service time of SUPs. These martingales effectively capture the stochastic characteristics of the inter-arrival time and service time, providing theoretical support for PAoI analysis.
- Based on the constructed martingales, we derive a closed-form expression for the UB-PAoIVP under the FCFS scheduling policy. The expression is applicable when the service times of SUPs are either i.i.d. (following discrete or continuous distributions) or follow a Markov process.

This work was supported in part by Natural Science Foundation of Jilin Province under Grant 20230101063JC. (Corresponding author: Xuefen Chi.)

H. Yan is with the School of Electrical and Information Engineering, Yunnan Key Laboratory of Unmanned Autonomous Systems, Yunnan Minzu University, Kunming 650500, China (e-mail: yanhy21@mails.jlu.edu.cn).

X. Chi is with the Department of Communications Engineering, Jilin University, Changchun 130012, China (e-mail: chixf@jlu.edu.cn).

W. Yang is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore (e-mail: wanting_yang@sutd.edu.sg).

Z. Xiong is with the School of Electronics, Electrical Engineering and Computer Science (EEECS), Queen's University Belfast, Belfast, BT7 1NN, U.K. (e-mail: z.xiong@qub.ac.uk)

Z. Han is with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea, 446-701 (e-mail: hanzhu22@gmail.com).

TABLE I
SUMMARY OF IMPORTANT NOTATIONS.

Notation	Description
$\mathbb{E}[\cdot]$	Expectation operator
$\Pr(\cdot)$	Probability of event $(\cdot)$
$T^{\mathrm{A}}(n)$, $T^{\mathrm{S}}(n)$, $T^{\mathrm{D}}(n)$	the arrival time, service time, and departure time of the n-th SUP
$T^{\mathrm{I}}(m,n)$	The inter-arrival time between the m-th and the n-th SUPs
$T^{\mathrm{S}}(m,n)$	The cumulative service time of the system from the m-th to the n-th SUPs
$P^{\mathrm{AoI}}(n)$	The PAoI for the n-th SUP
Δ_{th}	PAoI threshold
$f_{T^{\mathrm{I}}(n-1,n)}(x)$	The probability density function of inter-arrival time

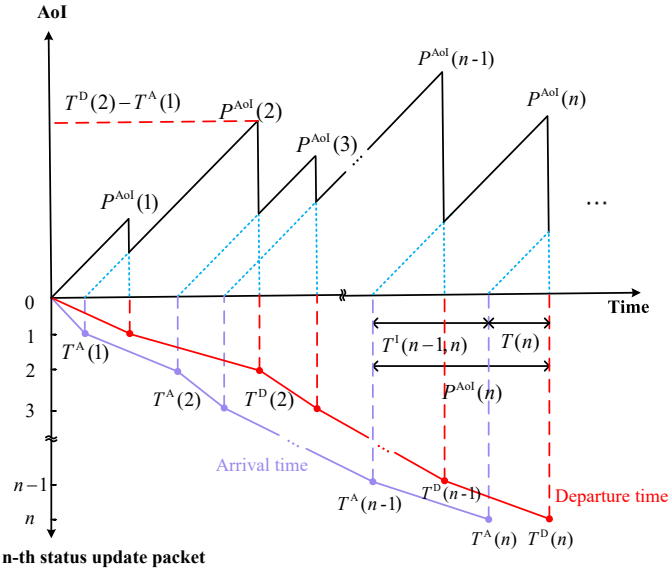


Fig. 1. AoI evolution for status update packets.

- We validate the derived closed-form expression through three simulation cases, confirming its accuracy and demonstrating its potential value in guiding parameter design and performance optimization for real-time systems.

Important notations used in this paper are listed in Table I.

II. System Model

We consider a real-time status update system in which a transmitter node continuously monitors a physical process (e.g., temperature, humidity, or pressure) and transmits the measurement results to a destination node in the form of SUPs. Let $T^{\mathrm{A}}(n)$, $T^{\mathrm{S}}(n)$ and $T^{\mathrm{D}}(n)$ represent the arrival time, service time, and departure time of the n-th SUP, respectively. The inter-arrival time between the $(n-1)$-th and the n-th SUPs can be defined as $T^{\mathrm{I}}(n-1,n) = T^{\mathrm{A}}(n) - T^{\mathrm{A}}(n-1)$ for $n = 1, 2, 3, \ldots$. Then, the inter-arrival time between the m-th and the n-th SUPs, with $1 \le m \le n$, can be expressed as

$$T^{\mathrm{I}}(m,n) = \textstyle\sum_{j=m+1}^{n} T^{\mathrm{I}}(j-1,j). \tag{1}$$

To maintain generality, we set $T^{\mathrm{A}}(0) = 0$ and $T^{\mathrm{I}}(n,n) = 0$. Similarly, the cumulative service time of the system from the m-th to the n-th SUPs can be expressed as

$$T^{\mathrm{S}}(m,n) = \textstyle\sum_{j=m}^{n} T^{\mathrm{S}}(j). \tag{2}$$

Fig. 1 demonstrates the impact of SUPs on the system's AoI under the FCFS strategy. It can be observed from the figure that the system's AoI reaches its peak immediately just before each SUP is received by the destination node. The PAoI, denoted by $P^{\mathrm{AoI}}(n)$, for the n-th SUP can be expressed as [16]

$$P^{\mathrm{AoI}}(n) = T^{\mathrm{D}}(n) - T^{\mathrm{A}}(n-1). \tag{3}$$

The departure time of the n-th SUP in the system is given by

$$T^{\mathrm{D}}(n) = \max_{1\le m\le n} \left\{ T^{\mathrm{A}}(m) + T^{\mathrm{S}}(m,n) \right\}. \tag{4}$$

In addition, to ensure the stable operation of the system, the following condition must be satisfied:

$$\mathbb{E}[T^{\mathrm{I}}(n-1,n)] \ge \mathbb{E}[T^{\mathrm{S}}(n)], \tag{5}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. By leveraging (1) and (4), we can derive the sojourn time of the n-th SUP as follows:

$$T(n) = T^{\mathrm{D}}(n) - T^{\mathrm{A}}(n) = \max_{1\le m\le n} \left\{ T^{\mathrm{S}}(m,n) - T^{\mathrm{I}}(m,n) \right\}, \tag{6}$$

where $T(n) \ge 0$. We only need to set m equal to n to prove this conclusion. Based on (3) and (6), the PAoI of n-th SUP can be further expressed as

$$P^{\mathrm{AoI}}(n) = T^{\mathrm{I}}(n-1,n) + \max_{1\le m\le n} \left\{ T^{\mathrm{S}}(m,n) - T^{\mathrm{I}}(m,n) \right\}. \tag{7}$$

III. Upper Bound Analysis for PAoI Violation Probability Using Martingale Theory

In this section, we utilize martingale theory to analyze the UB-PAoIVP of the system described in Section II, deriving the closed-form expression in Theorem 3.1. In practical systems, the inter-arrival time in many traffic flows typically exhibits independence and memorylessness [13]. Accordingly, this paper models the inter-arrival time as i.i.d. Two service-time models are considered: i.i.d. and Markov process, which respectively characterize the system's independent per-update service and the temporally correlated service induced by wireless channel fading.

A. Martingales for Inter-arrival Time and Service Time

In this subsection, we first review important aspects of martingale theory through the following definition.

Definition 3.1 (Martingale): Let $\mathcal{G}_n$ be a filtration on the given probability space, where $\mathcal{G}_n \subseteq \mathcal{G}_{n+m}$, for all $m, n \in \mathbb{N}$. A discrete-time stochastic sequence $\{\mathcal{Y}(n), n \ge 0\}$ is called a martingale if it satisfies the following conditions [17]:

$$\begin{aligned} &\mathbb{E}\left[|\mathcal{Y}(n)|\right] < \infty, \\ &\mathbb{E}\left[\mathcal{Y}(n+1) \mid \mathcal{G}_n\right] = \mathcal{Y}(n), \end{aligned} \tag{8}$$

where $\{\mathcal{Y}(n), n \ge 0\}$ is $\mathcal{G}_n$ measurable. Further, if $\mathbb{E}[\mathcal{Y}(n+1) \mid \mathcal{G}_n] \le \mathcal{Y}(n)$, the stochastic sequence $\{\mathcal{Y}(n); n \ge 0\}$ is a super-martingale. Similarly, if $\mathbb{E}[\mathcal{Y}(n+1) \mid \mathcal{G}_n] \ge \mathcal{Y}(n)$, the stochastic sequence $\{\mathcal{Y}(n); n \ge 0\}$ is a sub-martingale. For a given martingale or super-martingale, the probability that its maximum value exceeds a positive threshold x can be estimated using Doob's inequality, which is given by

$$\Pr\left\{\sup_{0\le t\le n} |\mathcal{Y}(t)| > x\right\} \le \mathbb{E}\left[|\mathcal{Y}(0)|\right]/x. \tag{9}$$

Based on the definition of a martingale, we construct corresponding martingales for the inter-arrival time and service time of the SUPs, respectively.

Lemma 3.1 (Martingale for Inter-Arrival Time): In a real-time status update system, the inter-arrival time between the m-th and n-th SUPs can be represented as $T^{\mathrm{I}}(m,n) = \sum_{j=m+1}^{n} T^{\mathrm{I}}(j-1,j)$, where $T^{\mathrm{I}}(j-1,j)$ is i.i.d. and $1 \leq m \leq n$. For every $\vartheta > 0$, the stochastic process:

$$\mathrm{M}_{T^{\mathrm{I}}}(m,n;\vartheta) := e^{\vartheta((n-m)\mathcal{K}_{T^{\mathrm{I}}} - T^{\mathrm{I}}(m,n))} \tag{10}$$

is a martingale. Here, the martingale parameter $\mathcal{K}_{T^{\mathrm{I}}} = -\ln\left(\mathbb{E}\left[\exp(-\vartheta T^{\mathrm{I}}(j-1,j))\right]\right)/\vartheta$.

Proof: Let us prove it using the definition of a martingale. For practical systems, $T^{\mathrm{I}}(j-1,j)$ is finite. Therefore, $\mathbb{E}\left[|\mathrm{M}_{T^{\mathrm{I}}}(m,n;\vartheta)|\right] < \infty$ can be obtained. Next, we can derive

$$\begin{aligned}
&\mathbb{E}[\mathrm{M}_{T^{\mathrm{I}}}(m,n+1;\vartheta)|T^{\mathrm{I}}(1,2),\ldots,T^{\mathrm{I}}(n-1,n)] \\
&= e^{\vartheta((n-m)\mathcal{K}_{T^{\mathrm{I}}} - T^{\mathrm{I}}(m,n))}\mathbb{E}[e^{\vartheta(-T^{\mathrm{I}}(n,n+1))}]e^{\vartheta\mathcal{K}_{T^{\mathrm{I}}}} \\
&= \mathrm{M}_{T^{\mathrm{I}}}(m,n;\vartheta).
\end{aligned} \tag{11}$$

■

Lemma 3.2 (Martingale for Service Time): For a real-time status update system, the cumulative service time from the m-th to the n-th SUP is denoted as $T^{\mathrm{S}}(m,n) = \sum_{j=m}^{n} T^{\mathrm{S}}(j)$, $1 \leq m \leq n$. For any $\vartheta > 0$, the following stochastic process:

$$\mathrm{M}_{T^{\mathrm{S}}}(m,n,\vartheta) := h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n))e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}})} \tag{12}$$

is a martingale. if $T^{\mathrm{S}}(j)$ is i.i.d., $h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n)) = 1$ and $\mathcal{K}_{T^{\mathrm{S}}} = \ln\left(\mathbb{E}\left[\exp(\vartheta T^{\mathrm{S}}(j))\right]\right)/\vartheta$. On the other hand, when $T^{\mathrm{S}}(j)$ is a Markov process with a finite state space $\mathcal{S} = \{s_1, s_2, \cdots, s_N\}$, where N denotes the number of states. Let p and q be the indices of the corresponding states with $p, q \in \{1,2,3,\cdots,N\}$. $\mathbf{P}$ is the transition probability matrix of $T^{\mathrm{S}}(j)$. $\mathbf{P}(\vartheta)$ is the exponential column-transformed matrix of $\mathbf{P}$, which is expressed as $\mathbf{P}_{p,q}(\vartheta) = \Pr(j+1 = q | j = p)e^{\vartheta T^{\mathrm{S}}(j+1)} = \mathbf{P}_{p,q}e^{\vartheta T^{\mathrm{S}}(j+1)}$. We can obtain $\mathcal{K}_{T^{\mathrm{S}}} \geq \ln(sp(\mathbf{P}(\vartheta)))/\vartheta$ and $sp(\mathbf{P}(\vartheta))$ is the spectral radius of the matrix $\mathbf{P}(\vartheta)$. $h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n))$ is the state parameter of $T^{\mathrm{S}}(n)$ in the right eigenvector corresponding to the spectral radius.

Proof: For a practical system, $T^{\mathrm{S}}(j)$ is finite, which directly implies that $\mathbb{E}\left[|\mathrm{M}_{T^{\mathrm{S}}}(m,n,\vartheta)|\right] < \infty$. When $T^{\mathrm{S}}(j)$ is a Markov process, we can derive

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{M}_{T^{\mathrm{S}}}(m,n+1,\vartheta)|T^{\mathrm{S}}(1),T^{\mathrm{S}}(2),...,T^{\mathrm{S}}(n)\right] \\
&\overset{(a)}{=} e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}})} \\
&\qquad \times \mathbb{E}\left[h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n+1))e^{\vartheta T^{\mathrm{S}}(n+1)}|T^{\mathrm{S}}(n)\right]e^{-\vartheta\mathcal{K}_{T^{\mathrm{S}}}} \\
&\overset{(b)}{=} e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}})}(\mathbf{P}(\vartheta)h_{T^{\mathrm{S}}})(T^{\mathrm{S}}(n))e^{-\vartheta\mathcal{K}_{T^{\mathrm{S}}}} \\
&\overset{(c)}{=} h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n))e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}})}sp(\mathbf{P}(\vartheta))e^{-\vartheta\mathcal{K}_{T^{\mathrm{S}}}} \\
&\overset{(d)}{\leq} h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n))e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}})},
\end{aligned} \tag{13}$$

where step (c) utilizes the Perron-Frobenius Theorem. When $T^{\mathrm{S}}(j)$ is i.i.d., we can derive

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{M}_{T^{\mathrm{S}}}(m,n+1;\vartheta)|T^{\mathrm{S}}(1),T^{\mathrm{S}}(2),...,T^{\mathrm{S}}(n)\right] \\
&= e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}})}\mathbb{E}\left[e^{\vartheta T^{\mathrm{S}}(n+1)}\right]e^{-\vartheta\mathcal{K}_{T^{\mathrm{S}}}} \\
&= \mathbb{E}\left[\mathrm{M}_{T^{\mathrm{S}}}(m,n+1;\vartheta)\right].
\end{aligned} \tag{14}$$

■

Lemma 3.3: Let $\mathcal{X}$ and $\mathcal{Y}$ be two independent non-negative random variables. Suppose that $\overline{F}_{\mathcal{X}}(x) \leq f(x)$ and $\overline{F}_{\mathcal{Y}}(x) \leq g(x)$, where $f, g \in \overline{\mathcal{F}}$. Then, for all $x \geq 0$,

$$\Pr\{\mathcal{X} + \mathcal{Y} > x\} \leq 1 - (\bar{f} * \bar{g})(x), \tag{15}$$

where $\bar{f}(x) = 1 - f(x)$ and $\bar{g}(x) = 1 - g(x)$. Here $\overline{F}_{\mathcal{X}}(x)$ and $\overline{F}_{\mathcal{Y}}(x)$ denote the complementary cumulative distribution functions of $\mathcal{X}$ and $\mathcal{Y}$, respectively. $\overline{\mathcal{F}}$ is the set of non-negative, wide-sense decreasing functions.
The proof of Lemma 3.3 can be found in [18].

B. Upper Bound of PAoI Violation Probability

Based on the martingales constructed for the inter-arrival time and service time, we obtain the following theorem.

Theorem 3.1: Consider a stable real-time status update system, where $\mathrm{M}_{T^{\mathrm{I}}}(m,n;\vartheta)$ and $\mathrm{M}_{T^{\mathrm{S}}}(m,n,\vartheta)$ are the martingales representing the inter-arrival time and the cumulative service time between the m-th and the n-th SUPs, respectively. For any PAoI threshold $\Delta_{\mathrm{th}} \geq 0$, the corresponding UB-PAoIVP is given by (20), where $f_{T^{\mathrm{I}}(n-1,n)}(\cdot)$ is the probability density function of the distribution that the inter-arrival time $T^{\mathrm{I}}(n-1,n)$ follows. The parameter Θ can be expressed as

$$\begin{aligned}
\Theta := \min_{1\leq m\leq j\leq n, 1\leq m\leq k\leq n} \big\{&h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(j))| \\
&T^{\mathrm{S}}(j) > T^{\mathrm{I}}(k-1,k)\big\}.
\end{aligned} \tag{16}$$

The parameter ϑ^* is set to

$$\vartheta^* := \max\{\vartheta > 0 : \mathcal{K}_{T^{\mathrm{S}}} \leq \mathcal{K}_{T^{\mathrm{I}}}\}. \tag{17}$$

Proof: Based on Lemmas 3.1 and 3.2, we first construct a martingale to describe the PAoI, as shown below:

$$\begin{aligned}
\mathrm{M}(m,n;\vartheta) := &h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n)) \\
&\times e^{\vartheta(T^{\mathrm{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\mathrm{S}}} + (n-m)\mathcal{K}_{T^{\mathrm{I}}} - T^{\mathrm{I}}(m,n))},
\end{aligned} \tag{18}$$

where we set $\mathcal{K}_{T^{\mathrm{S}}} \leq \mathcal{K}_{T^{\mathrm{I}}}$ to ensure that (18) is a super-martingale. We can derive the UB-PAoIVP as follows:

$$\begin{aligned}
&\Pr\left\{P^{\mathrm{AoI}}(n) > \Delta_{\mathrm{th}}\right\} \overset{(a)}{=} \Pr\Big\{T^{\mathrm{I}}(n-1,n) \\
&\qquad + \max_{1\leq m\leq n}\left\{T^{\mathrm{S}}(m,n) - T^{\mathrm{I}}(m,n)\right\} > \Delta_{\mathrm{th}}\Big\} \\
&\overset{(b)}{=} 1 - \Pr\Big\{T^{\mathrm{I}}(n-1,n) \\
&\qquad + \max_{1\leq m\leq n}\left\{T^{\mathrm{S}}(m,n) - T^{\mathrm{I}}(m,n)\right\} \leq \Delta_{\mathrm{th}}\Big\} \\
&\overset{(c)}{\leq} 1 - \int_0^{\Delta_{\mathrm{th}}} \Pr\left\{\max_{1\leq m\leq n}\left\{T^{\mathrm{S}}(m,n) - T^{\mathrm{I}}(m,n)\right\} \leq x\right\} \\
&\qquad \times f_{T^{\mathrm{I}}(n-1,n)}(\Delta_{\mathrm{th}} - x)dx \\
&\overset{(d)}{\leq} 1 - \int_{\mathcal{K}_{T^{\mathrm{I}}}}^{\Delta_{\mathrm{th}}} \left(1 - \frac{\mathbb{E}[h_{T^{\mathrm{S}}}(T^{\mathrm{S}}(n))]}{\Theta}e^{-\vartheta^*(x-\mathcal{K}_{T^{\mathrm{I}}})}\right) \\
&\qquad \times f_{T^{\mathrm{I}}(n-1,n)}(\Delta_{\mathrm{th}} - x)dx,
\end{aligned} \tag{19}$$

where we apply (7) to obtain step (a). By applying Lemma 3.3, step (c) can be derived from step (b). For the $\Pr\left\{\max_{1\leq m\leq n}\left\{T^{\mathrm{S}}(m,n) - T^{\mathrm{I}}(m,n)\right\} \leq x\right\}$ in step (c), we further derive it as shown in (21). Applying the conclusion

$$\Pr\left\{P^{\text{AoI}}(n) > \Delta_{\text{th}}\right\} \leq \begin{cases} \min\left\{1, 1 - \int_{\mathcal{K}_{T^{\text{I}}}}^{\Delta_{\text{th}}} \left(1 - \frac{\mathbb{E}[h_{T^{\text{S}}}(T^{\text{S}}(n))]}{\Theta} e^{-\vartheta^*(x-\mathcal{K}_{T^{\text{I}}})}\right) \cdot f_{T^{\text{I}}(n-1,n)}(\Delta_{\text{th}} - x)dx\right\}, & \Delta_{\text{th}} \geq \mathcal{K}_{T^{\text{I}}}, \\ 1, & \text{Otherwise.} \end{cases} \tag{20}$$

from (21), step (d) can be derived from step (c) in (19). By substituting specific values, it is easy to demonstrate that the right-hand side of the inequality in step (d) of (19) is a decreasing function of ϑ. Therefore, to achieve the tightest bound, we define ϑ^*, which is expressed as shown in (17). Additionally, as probabilities cannot exceed 1 in practical scenarios, we modify (19) to obtain (20).

$$\begin{aligned} &\Pr\left\{\max_{1\leq m\leq n}\left\{T^{\text{S}}(m,n) - T^{\text{I}}(m,n)\right\} \leq x\right\} \\ &\overset{(a)}{=} 1 - \Pr\left\{\max_{1\leq m\leq n}\left\{T^{\text{S}}(m,n) - T^{\text{I}}(m,n)\right\} > x\right\} \\ &\overset{(b)}{\geq} 1 - \Pr\left\{\max_{1\leq m\leq n}\left\{T^{\text{S}}(m,n) - (n-m+1)\mathcal{K}_{T^{\text{S}}}\right.\right. \\ &\qquad \left.\left. +(n-m)\mathcal{K}_{T^{\text{I}}} - T^{\text{I}}(m,n)\right\} > x - \mathcal{K}_{T^{\text{I}}}\right\} \\ &\overset{(c)}{\geq} 1 - \Pr\left\{\max_{1\leq m\leq n}\left\{h_{T^{\text{S}}}(T^{\text{S}}(n))e^{\vartheta^*(T^{\text{S}}(m,n) - T^{\text{I}}(m,n))}\right\}\right. \\ &\qquad\qquad \left. > \Theta e^{\vartheta^*(x-\mathcal{K}_{T^{\text{I}}})}\right\} \\ &\overset{(d)}{\geq} 1 - \frac{\mathbb{E}[h_{T^{\text{S}}}(T^{\text{S}}(n))]}{\Theta} e^{-\vartheta^*(x-\mathcal{K}_{T^{\text{I}}})}. \end{aligned} \tag{21}$$

Based on (16) and (18), steps (b) and (c) in (21) can be derived. step (d) is obtained from step (c) in (21) using (9). ■

C. Case Study

In this subsection, we conduct a detailed case study and derive the corresponding model parameters. Assuming that the arrival process of SUPs follows a Poisson process, the inter-arrival time between two adjacent SUPs follows an exponential distribution with a mean of λ. The probability density function (PDF) of inter-arrival time is given by

$$f_{T^{\text{I}}(n-1,n)}(x) = \lambda^{-1} e^{-\lambda^{-1}x}. \tag{22}$$

Based on (22) and Lemma 3.1, we can derive $\mathcal{K}_{T^{\text{I}}} = -\ln\left(\frac{1}{1+\lambda\vartheta}\right)/\vartheta$.

For the service time of SUPs, three cases are considered.

Case 1: The service time follows a Markov process with a state space $\{s_1, s_2\}$. l and v denote the transition probabilities between the two states. The state transition probability matrix can be expressed as $\mathbf{P} = \begin{bmatrix} 1-l & v \\ l & 1-v \end{bmatrix}$. The corresponding steady-state probabilities can be given as

$$\pi_1 = v/(l+v), \pi_2 = l/(l+v). \tag{23}$$

Based on Lemma 3.2, we can get

$$\mathbf{P}(\vartheta) = \begin{bmatrix} (1-l)e^{\vartheta s_1} & ve^{\vartheta s_2} \\ le^{\vartheta s_1} & (1-v)e^{\vartheta s_2} \end{bmatrix}, \tag{24}$$

where the spectral radius of (24) is $sp(\mathbf{P}(\vartheta))$, and its corresponding right eigenvector is $V = [h_{T^{\text{S}}}(s_1), h_{T^{\text{S}}}(s_2)]$. We can further derive $\mathbb{E}\left[h_{T^{\text{S}}}(T^{\text{S}}(n))\right] = \pi_1 \cdot h_{T^{\text{S}}}(s_1) + \pi_2 \cdot h_{T^{\text{S}}}(s_2)$.

Case 2: The service time is specified to be i.i.d. and follows a continuous distribution, modeled as an exponential distribution with mean μ, expressed as

$$f_{T^{\text{S}}(n)}(y) = \mu^{-1} e^{-\mu^{-1}y}. \tag{25}$$

Based on (25) and Lemma 3.2, we can derive $\mathcal{K}_{T^{\text{S}}} = \ln\left(\frac{1}{1-\mu\vartheta}\right)/\vartheta$.

When the service time is i.i.d., (20) can be further simplified. Taking *Case 2* as an example, by substituting its parameters into (20), we obtain

$$\Pr\left\{P^{\text{AoI}}(n) > \Delta_{\text{th}}\right\} \leq \begin{cases} \frac{(1+\lambda\vartheta^*)e^{-\vartheta^*\Delta_{\text{th}}} - \lambda\vartheta^*(1+\lambda\vartheta^*)\frac{1}{\lambda\vartheta^*}e^{-\frac{\Delta_{\text{th}}}{\lambda}}}{1-\lambda\vartheta^*}, & \Delta_{\text{th}} \geq \frac{\lambda\mu\ln(1+\frac{\lambda-\mu}{\mu})}{\lambda-\mu} \\ 1, & \text{Otherwise,} \end{cases} \tag{26}$$

where $\vartheta^* = (\lambda-\mu)/\lambda\mu$.

Case 3: The service time is specified to be i.i.d. and follows a discrete distribution, modeled as a generalized Bernoulli distribution, which can be expressed as

$$T^{\text{S}}(n) = \begin{cases} t_1, & p_s, \\ t_2, & 1-p_s, \end{cases} \tag{27}$$

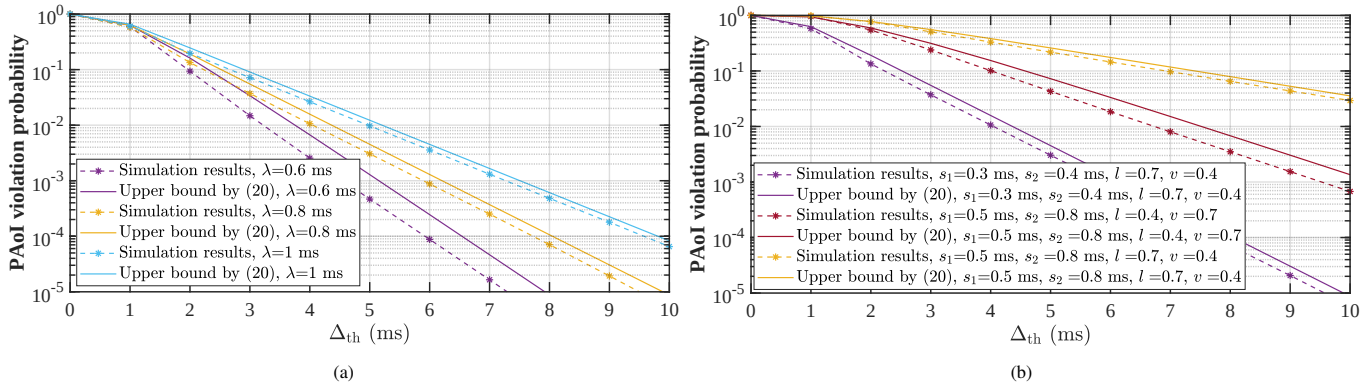
where t_1 and t_2 are the durations of the service, and p_s is the probability that the service time is t_1. Based on (27) and Lemma 3.2, we can derive

$$\mathcal{K}_{T^{\text{S}}} = \ln\left(p_s e^{\vartheta t_1} + (1-p_s)e^{\vartheta t_2}\right)/\vartheta. \tag{28}$$

IV. Numerical Results

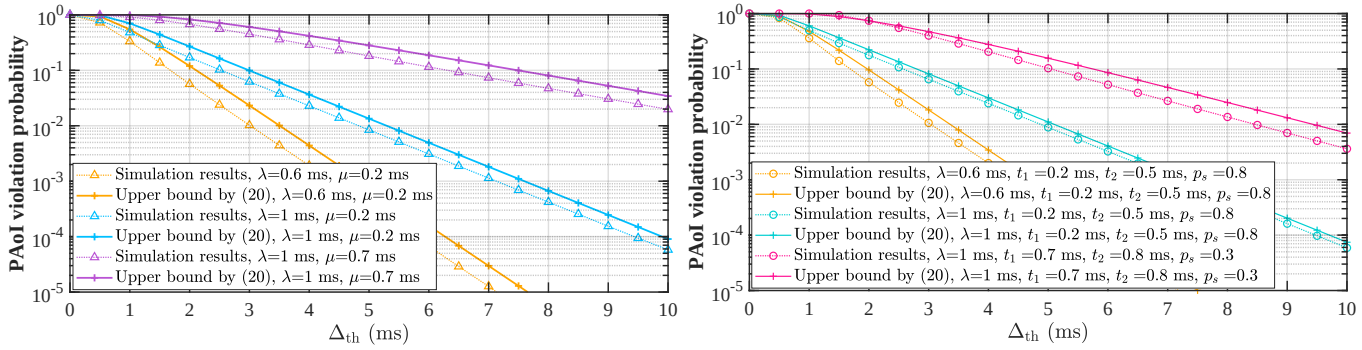
In this section, we verify the accuracy and generality of the derived UB-PAoIVP through numerical results. In our simulation, the number of SUPs is set to 10^8. Subsequently, the Monte Carlo method is used to generate the inter-arrival times of adjacent SUPs and the service times of each SUP for the three cases previously proposed, respectively. The relevant parameters are provided in the corresponding figures. For each PAoI threshold, 100 simulation runs are conducted. The time unit is set to milliseconds (ms).

Figs. 2 (a) and (b) illustrate the relationship between the PAoI threshold and the PAoI violation probability under various inter-arrival time and service time parameters, where the service time is modeled as a Markov process (*Case 1*). Additionally, the theoretical upper bounds obtained by (20) are compared with the simulation results. In Fig. 2 (a), the service time parameters $s_1 = 0.3\,\text{ms}, s_2 = 0.4\,\text{ms}, l = 0.7, v = 0.4$ are fixed, while the inter-arrival time parameter is varied. In Fig. 2 (b), the inter-arrival time parameter $\lambda = 0.8\,\text{ms}$ is fixed, while the service time parameters are varied. It can be observed from Figs. 2 (a) and (b) that the theoretical UB-PAoIVP closely align with the simulation results, exhibiting consistent decay trends. This demonstrates the accuracy of the derived (20) in analyzing the AoI performance under the given conditions. Furthermore, the results in Figs. 2 (a) and (b) indicate that



(a) (b)

Fig. 2. PAoI violation probability versus PAoI threshold when the service time is modeled as a Markov process (*Case 1*). (a) Fixing service time while varying inter-arrival time parameters. (b) Fixing inter-arrival time while varying service time parameters.



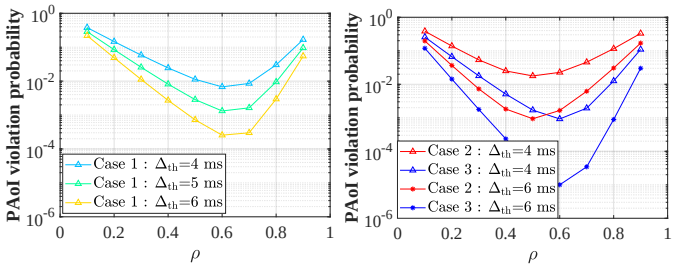
(a) (b)

Fig. 3. PAoI violation probability versus PAoI threshold when the service time is modeled as i.i.d. (a) *Case 2*-The service time follows an exponential distribution. (b) *Case 3*-The service time follows a generalized Bernoulli distribution.

different inter-arrival time and service time parameter settings can influence the prediction accuracy of the theoretical upper bound obtained from (20). For instance, in Fig. 2 (a), as the inter-arrival time parameter λ increases, the accuracy of the upper bound obtained by (20) improves. Therefore, in PAoI optimization, identifying the impact of specific parameters on accuracy can facilitate more efficient allocation and utilization of available resources.

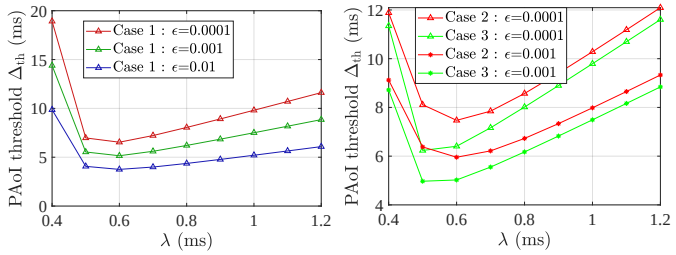
Figs. 3 (a) and (b) further reveal the relationship between the PAoI threshold and the PAoI violation probability under different inter-arrival time and service time parameters when the service time is modeled as i.i.d. Specifically, Fig. 3 (a) corresponds to *Case 2*, where the service time follows an exponential distribution, while Fig. 3 (b) corresponds to *Case 3*, where the service time follows a generalized Bernoulli distribution. By comparing the simulation results with the theoretical upper bounds derived from (20) across different parameter settings and distributions, it is evident that when service times are i.i.d., the proposed (20) continues to accurately estimate the PAoI violation probability. In addition, the results in Figs. 2 and 3 demonstrate that the derived theoretical method has practical value in real time systems, as it can rapidly evaluate information freshness and replace complex simulations, thereby assisting system designers in efficiently configuring update frequencies and resources to ensure real-time transmission requirements.

The system utilization is an important performance metric, defined as $\rho = \frac{\mathbb{E}[T^{\text{S}}(n)]}{\mathbb{E}[T^{\text{I}}(n-1,n)]}, (0 < \rho < 1)$. Figs. 4 (a) and (b) illustrate the variation of the PAoI violation probability with respect to the system utilization ρ under three different service time models. Fig. 4 (a) shows the results for *Case 1*, where the service time parameters are fixed as $s_1 = 0.3\,\text{ms}, s_2 = 0.4\,\text{ms}, l = 0.7$, and $v = 0.4$. Fig. 4 (b) presents the results for *Case 2* and *Case 3*. For *Case 2*, we set $\mu = 0.3$, and for *Case 3*, the service time parameters are set as $t_1 = 0.2\,\text{ms}, t_2 = 0.5$ ms, $p_s = 0.8$. The system utilization is adjusted by varying the inter-arrival time parameters. Figs. 4 (a) and (b) show that, under three service time models, the PAoI violation probability first decreases and then increases as the ρ increases. This non-monotonic behavior arises because the PAoI depends on the inter-arrival time, queuing delay, and service time of status updates. When ρ is low, status updates are too infrequent to maintain freshness, whereas when ρ is high, excessive arrivals cause congestion and long delays. Therefore, there exists an optimal utilization level that minimizes the PAoI violation probability. Notably, this optimal point remains relatively stable across different PAoI thresholds, suggesting that it is mainly determined by system dynamics rather than threshold settings. This finding provides valuable insights for real-time applications such as disaster monitoring, IIoT, and vehicular networks. It implies that system design should avoid extreme utilization levels and instead operate near the optimal region to ensure both the timeliness and reliability of information delivery, thereby improving decision-making



(a) (b)

Fig. 4. PAoI violation probability versus system utilization. (a) Service time follows a Markov process *(Case 1)*. (b) Service time follows exponential and generalized Bernoulli distributions, respectively *(Case 2 and Case 3)*.



(a) (b)

Fig. 5. PAoI threshold versus average inter-arrival time. (a) Service time follows a Markov process *(Case 1)*. (b) Service time follows exponential and generalized Bernoulli distributions, respectively *(Case 2 and Case 3)*.

efficiency and maintaining system stability.

The PAoI violation probability is denoted by ϵ. Figs. 5 (a) and (b) further illustrate the relationship between the PAoI threshold Δ_{th} and the average inter-arrival time λ under different ϵ constraints for three cases. The service time parameters are kept consistent with those in Fig. 4. As shown in Figs. 5 (a) and (b), with the increase of λ, the PAoI threshold first decreases and then increases, indicating the existence of an optimal λ that minimizes the PAoI threshold Δ_{th}. Furthermore, The results show that variations in the violation probability ϵ have little effect on the value of the optimal λ. This suggests that, in practical systems, when only the violation probability requirement of the PAoI is adjusted, the update rate can remain unchanged, which simplifies system configuration and reduces operational complexity.

V. Conclusions

In this work, the probabilistic distribution of the UB-PAoIVP in a real-time status update system has been investigated. We have constructed martingales for the inter-arrival time and service time, establishing their relationship with the PAoI in the martingale domain. Based on this martingale framework, we have derived a closed-form expression for the UB-PAoIVP. Notably, the derived expression is applicable to the service time, whether it is modeled as i.i.d. or as a Markov process. We have conducted simulations in three different cases, and the results have shown that the derived UB-PAoIVP closely matches the simulation results, demonstrating the accuracy and generality of the proposed method. These findings can further serve as design guidelines for time-sensitive systems (e.g., autonomous driving, IIoT, and disaster monitoring), where the obtained UB-PAoIVP provides a conservative worst-case timeliness guarantee and thereby supports system-parameter selection to satisfy prescribed reliability requirements. In future work, we will further investigate the PAoI performance under more complex system settings, and also extend the analysis to average AoI.

References

[1] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.

[2] Y. Emami, H. Gao, K. Li, L. Almeida, E. Tovar, and Z. Han, "Age of information minimization using multi-agent uavs based on ai-enhanced mean field resource allocation," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 13 368–13 380, Sep. 2024.

[3] H. Tong, S. Wang, Z. Yang, J. Zhao, M. Bennis, and C. Yin, "Semantic-aware remote state estimation in digital twin with minimizing age of incorrect information," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2023, pp. 4534–4539.

[4] Z. Fang, J. Wang, Y. Ren, Z. Han, H. V. Poor, and L. Hanzo, "Age of information in energy harvesting aided massive multiple access networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1441–1456, May 2022.

[5] Y. Zhang, Z. Xiong, D. Niyato, P. Wang, and Z. Han, "Information trading in internet of things for smart cities: A market-oriented analysis," *IEEE Network*, vol. 34, no. 1, pp. 122–129, Feb. 2020.

[6] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. on Sensor, Mesh and Ad Hoc Commun. Netw.*, Jun. 2011, pp. 350–358.

[7] W. Lin, L. Li, J. Yuan, Z. Han, M. Juntti, and T. Matsumoto, "Age-of-information in first-come-first-served wireless communications: Upper bound and performance optimization," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9501–9515, Sep. 2022.

[8] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, Apr. 2016.

[9] X. Zhou, D. Niyato, and C. Yuen, "Age of information driven task allocation for periodic updating crowdsensing: A contract theory based approach," *IEEE Internet Things J.*, Nov. 2024, early access.

[10] L. Hu, Z. Chen, Y. Dong, Y. Jia, L. Liang, and M. Wang, "Status update in iot networks: Age-of-information violation probability and optimal update rate," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11 329–11 344, Jul. 2021.

[11] J. Wang, W. Cheng, and H. Vincent Poor, "Statistical delay and error-rate bounded qos provisioning for aoi-driven 6g satellite- terrestrial integrated networks using fbc," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 15 540–15 554, Oct. 2024.

[12] Q. Meng, H. Lu, and L. Qin, "Energy optimization in statistical aoi-aware mec systems," *IEEE Commun. Lett.*, vol. 28, no. 10, pp. 2263–2267, Oct. 2024.

[13] A. Zhong, Z. Li, D. Wu, T. Tang, and R. Wang, "Stochastic peak age of information guarantee for cooperative sensing in internet of everything," *IEEE Internet Things J.*, vol. 10, no. 17, pp. 15 186–15 196, Sep. 2023.

[14] M. Fidler, J. P. Champati, J. Widmer, and M. Noroozi, "Statistical age-of-information bounds for parallel systems: When do independent channels make a difference?" *IEEE J. Sel. Areas Inf. Theory.*, vol. 4, pp. 591–606, Oct. 2023.

[15] Q. Zhang, M. Xiong, D. Zhou, Y. Dong, Q. Liu, W. Lu, and Z. Han, "Resonant beam information and power transfer: Multiple access modeling and delay analysis," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 17 651–17 665, Nov. 2024.

[16] X. Zhang, J. Wang, and H. V. Poor, "Aoi-driven statistical delay and error-rate bounded qos provisioning for murllc over uav-multimedia 6g mobile networks using fbc," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3425–3443, Nov. 2021.

[17] R. S. Liptser and A. N. Shiryayev, *Theory of Martingales*. Norwell, MA: Kluwer Academic Publishers, 1989.

[18] Y. Jiang and Y. Liu, *Stochastic network calculus*. Springer-Verlag, 2008.