Information Freshness and Timeliness Analysis in The Finite Blocklength Regime for Mission-Critical Applications

Di Zhang, *Senior Member, IEEE,* Mingxiao Sun, Lulu Song, Shaobo Jia, Aimin Li, Tingting Zhang, Anwer Al-Dulaimi, *Senior Member, IEEE,* and Shahid Mumtaz, *Senior Member, IEEE*

***Abstract*—Mission-critical applications are of significant importance to sixth generation (6G)'s massive and ubiquitous Internet of things (IoT) communications. The mission-critical applications mostly fall within the scope of finite blocklength (FBL), and in order to assess the information freshness, age of information (AoI) has been introduced. However, packet error is inevitable in the FBL regime, which exerts impacts on the time for successful packet transmission, and thus increases the AoI. To optimize the performance of AoI, the management of queue packets is an effective way. Motivated by optimizing the AoI performance in the FBL regime, we consider a system equipped with a single buffer, and propose two schemes of packet management in this article. We subsequently derive the closed-form expressions for the average AoI and the average peak AoI and we discuss the relationship between AoI and the factors, i.e. the blocklength, data generation rate and signal-to-noise ratio. Afterwards, we give the optimal blocklength expression associated with the optimal AoI. In order to examine the information timeliness in the network under the proposed schemes, the closed-form expressions of the average delay are deduced. The simulation results validate the theoretical analysis and demonstrate the advantage of the proposed scheme in terms of the performance of AoI, delay, and their trade-off.**

***Index Terms*—Packet management, age of information, delay, finite blocklength, mission-critical applications, 6G, Internet of things.**

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 62301502, the Henan Natural Science Foundation for Excellent Young Scholar under Grant 242300421169 and 252300421224, and the Fundamental Research Funds for the Central Universities under Grant B240203012. (Corresponding author: Shaobo Jia.)

Di Zhang is with the School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, 518107, China (E-mail: zhangd263@mail.sysu.edu.cn).

Mingxiao Sun, Lulu Song and Shaobo Jia are with the Department of Electrical and Information Engineering, Zhengzhou University, China (E-mail: dr.di.zhang@ieee.org, eiemxsun@gs.zzu.edu.cn, lulu_song@gs.zzu.edu.cn, ieshaobojia@zzu.edu.cn).

Aimin Li is with the School of Electronics Engineering Harbin Institute of Technology (Shenzhen), Shenzhen, China (E-mail: liaimin@stu.hit.edu.cn).

Tingting Zhang is with the Key Laboratory of Maritime Intelligent Cyberspace Technology (Hohai University), Ministry of Education, China (E-mail: 20151951@hhu.edu.cn).

Anwer Al-Dulaimi is with the College of Technological Innovation, Zayed University, Abu Dhabi, UAE, and also with the Veltris, CTO Office, Toronto. Canada (E-mail: anwer.aldulaimi@ieee.org).

Shahid Mumtaz is with the with the Department of Applied Informatics, Silesian University of Technology Akademicka 16 44-100 Gliwice, Poland and also with the Department of Computer Sciences, Nottingham Trent University, NG1 4FQ Nottingham, U.K. (E-mail: dr.shahid.mumtaz@ieee.org).

Part of this study has been accepted by IEEE ICCT-2024.

I. Introduction

Mission-critical applications, such as remote surgery and industrial automation control, play a significant role in sixth generation (6G) [1]–[3], in which the interactive information falls typically within the scope of finite blocklength (FBL) regime [4], [5]. In these mission-critical Internet of things (IoT) applications, devices are required to sense the surrounding physical environment and monitor the system status in real-time to provide timely and effective information for intelligent decision making and control [6]. For example, each driverless car is required to update the information on surrounding vehicles and the environment, as well as the high-precision real-time map information [7]. With the advancement of such technologies, the requirement for low-latency communications is becoming increasingly stringent. This trend has propelled FBL technology as a pivotal solution that reduces transmission delay by compressing packet length, thereby significantly enhancing the timeliness of information delivery.

In mission-critical applications, the information freshness constitutes a crucial metric as the utility of real-time information deteriorates over time. The inherent delay and potential obstruction of data transmission in the network will severely impact the freshness of the information at the receiving end [8]. Recent studies have revealed that the freshness of information has far-reaching implications, influencing various critical metrics in different applications, including but not limited to security protocols and the accuracy of data processing and decision-making processes [9], [10]. The correlation between information freshness and these metrics underscores the need for systems to prioritize the timely delivery and update of information, as failure to do so could compromise operational effectiveness and user safety. Thus, ensuring that the data remain fresh and timely is not just a technical challenge but also a fundamental requirement for improving user's satisfaction and maintaining robust security frameworks within diverse fields of mission-critical applications [11], [12].

In order to measure the freshness of information, the concept of age of information (AoI) has been introduced [13]. AoI is defined as the time elapsed since the generation of the last successfully received packet, specifically indicating the freshness of the received information. Assuming that the instant of generation of the latest received packet is $g(t)$, the instantaneous AoI is represented as $A(t) \triangleq t - g(t)$. It is crucial to emphasize that a lower AoI value is di-

rectly associated with obtaining fresher information at the destination, which constitutes a key prerequisite in numerous mission-critical applications. This forms a sharp contrast to conventional performance indicators such as throughput and delay, which have traditionally been used to assess the efficiency of communication systems. Moreover, AoI is not only affected by the transmission time of packets but also by several other factors, including the update rate and the efficacy of queue management utilized within the system [9]. The multi-dimensional characteristic of AoI highlights its significance in evaluating the performance of communication networks where the timely delivery of information is of utmost importance.

Compared to the infinite blocklength regime (IBL), the finite packet length reduces transmission time and improves update frequency by compressing data length, thus achieving lower latency and AoI. However, a short block increases the block error rate (BLER) [14], and packet loss or decoding failures can further cause a continuous rise in AoI. Therefore, blocklength has a two-sided effect on AoI and delay. With the escalating demands for information timeliness across diverse applications, AoI and delay analysis within the FBL regime are significant in mission-critical applications [15]. By addressing these factors, we can take significant strides toward enhancing the quality of service (QoS), ultimately ensuring that they meet the diverse and rigorous demands.

A. Related Work

To enhance the freshness of information in communication systems, an extensive range of queuing methods and packet management schemes have been profoundly investigated, as emphasized in [13], [16]–[18]. In their seminal work, [13] carried out an in-depth analysis of the average AoI by applying the first-come-first-serve (FCFS) queuing scheme. This investigation was conducted under three distinct queuing models corresponding to the IBL regime: M/M/1, M/D/1, and D/M/1 (where the Kendall symbols respectively specifies the inter-arrival time distribution, the service time distribution and the number of servers and can be extended with an extra entry to denote the number of waiting rooms for customers). On this basis, subsequent studies presented in [16] redirected its focus to the last-come-first-serve (LCFS) queuing scheme, which demonstrated that giving priority to the transmission of the most recent updates can result in substantial enhancements in AoI performance. Additionally, this prioritization strategy highlights the significance of timely information dissemination in various applications. Furthermore, the study of optimal status update policies was advanced in [18], where the methodology of Markov decision processes was utilized. The authors of [19] derived general bounds for peak AoI distribution by leveraging independence between information generation and transmission processes, validated through numerical case studies in M/M/1 and D/M/1 queuing systems. Moreover, the challenges of the average AoI in multi-class queuing systems were tackled in [20], providing valuable insights into how different classes of data can be managed and prioritized within a queuing framework, and considering both preemptive and non-preemptive schemes, study in [21] analyzed the average AoI expression in large-scale IoT scenario. These works collectively contributes to a deeper comprehension of how queuing theory and management strategies can effectively impact information freshness.

However, since AoI continues to accumulate until a packet is successfully received, packet error becomes another vital issue. This issue is particularly prominent given that the timely delivery of packets is essential for maintaining the freshness of information. In this context, the work presented in [22] specifically addressing the problem related to the blocklength that minimizes the AoI violation probability. Besides, [23] discussed the probability of SPC transmission success derived under the constraint of unreliable channel and RF power supply, and proposes static random and maximum weight scheduling strategies to optimize the expected weighted sum AoI. Moreover, the authors in [24] optimized the short packet length of two-hop communication to minimize the decoding error probability and reasonably predict the subsequent adjacent state information to shorten the transmission delay. Further, study in [25] addressed the optimization of both the average and instantaneous AoI through an innovative approach involving blocklength adaptation within a short packet sensor system.

Studies mentioned before assume an infinite queue capacity to guarantee a complete packet delivery, which is especially suitable for scenarios with unharsh latency demands. However, in high-frequency update scenarios, systems often prioritize processing newly generated packet while discarding outdated information to maintain efficiency and relevance. Such systems employ finite buffers that simultaneously mitigate queuing-induced latency and conserve resources by eliminating obsolete data transmissions.

To adapt to numerous scenarios and effectively reduce AoI, researchers have engaged in the analysis of various queuing theory and packet management schemes. For instance, [26] established that the preemptive LCFS policy simultaneously optimizes AoI, throughput, and delay in both infinite and finite buffer queuing systems without requiring packet dropping, with proven age-optimality under arbitrary arrival processes. Furthermore, M/G/1/1 queues with lower variance perform better than those with equivalent average service times. The blocklength adaptation method was also used to minimize the average AoI in M/G/1/1 systems [27], where the authors studied the performance of AoI under the traditional protocol where the sender discards the error status packets directly. For the multi-source M/G/1/1 queue model, the moment generating functions of AoI for each source are derived under self-preemption, globally preemptive, and non-preemptive strategies in [28]. Based on these results, the average AoI and peak AoI for two-source systems are calculated, thereby revealing the impact of service time distribution parameters. Additionally, in [29], authors derived the average AoI within the M/M/1 queuing framework by implementing a random mixing system (SHS) in a state update system characterized by two independent sources, and authors in [30] also analyzed multi-source single-buffer system and manage packet priorities using SHS approach to calculate the average AoI of different sources.

Considering the scenarios associated with the high update rate in mission-critical applications, it is of paramount importance to ensure that the most up-to-date packets are transmitted promptly to prevent queue congestion. To effectively tackle this issue, an analysis of AoI was carried out employing the M/G/1/1 queuing model. This analysis scrutinized three distinct packet management schemes: *non-preemption* (NP), *preemptive* (PR), and *retransmission* (RT) [31]. The findings therein demonstrated that although the PR scheme theoretically aims to improve the AoI performance, it does not consistently guarantee a superior outcome in all circumstances [32], which indicates that an M/G/1/1 non-preemptive queue substantially enhances performance of AoI, especially with diverse service processes. Moreover, the study disclosed that in both M/M/1/1 and M/G/1/1 queuing systems, when the server is engaged in an ongoing transmission, any newly arriving packets are instantly discarded under the NP scheme, while the PR scheme enables new arrivals to preempt the ongoing transmission. However, in scenarios with high update rate, frequent preemption can lead to failed transmission within the PR scheme. Simultaneously, the NP scheme's propensity to drop fresher packets also has adverse effects on the AoI performance. Nevertheless, these detrimental effects can be alleviated in systems that incorporate a waiting room capable of buffering at least one packet. The potential benefits of allowing the server to wait before processing packet transmissions are explored in [33], particularly in non-preemptive queues characterized by Poisson arrival processes.

On the other hand, [34] examined the probability that delay, as well as the average peak AoI, exceeding a predefined thresholds within a short packet communication framework. The statistical distribution of both delay and average peak AoI specific to short packet communication systems was derived in [35]. Additionally, the interaction between delay and AoI under the FBL regime, specifically within non-preemptive and re-transmission schemes—was analyzed in [36], [37], where the authors proposed an algorithm aimed at optimizing the balance between AoI and latency, and authors in [38] also established a fundamental AoI-delay tradeoff in multi-server systems, proving that minimizing average AoI asymptotically inevitably causes packet delay and its variance to diverge to infinity. In this case, a trade-off between AoI and delay serving as a valuable metric to assess the performance and efficiency of the system, highlight the interrelated nature of these two factors to ensure a timely and reliable packet delivery in advanced communication scenarios, is of vital importance.

B. Contributions

Motivated by the above issues, this article delves into the intricate function of packet management, specifically focusing on the effects of blocklength, data generation rate, signal-to-noise ratio (SNR) in relation to both AoI and delay. The system is supported by a buffer within the FBL mechanism, which differs from the infinite-buffer system and bufferless system examined in prior studies [31], [37]. The study aims to provide a comprehensive analysis and understanding of the interactions between these factors. The contributions of this research can be summarized as follows.

(1) We first conduct an in-depth investigation into the influence of packet management schemes, named NR and RiB scheme, on the information freshness of a communication system within the FBL regime. Based on this, we derive closed-form expressions for both the average AoI and the average peak AoI. These expressions are tailored for distinct packet management schemes while taking into account the BLER. We also reveal the relationship between blocklength and the AoI, and give the optimal blocklength with numerical method.
(2) In order to reveal the relationship between information timeliness and the packet management schemes, we afterwards give the average delay expression for two schemes, i.e., NR and RiB schemes in this article. We find that the two packet management schemes have distinctive effects to the information timeliness.
(3) To validate our theoretical analysis, we present simulation results that reinforce our findings and illustrate the effectiveness of the proposed packet management scheme that incorporates replacement mechanisms. The results indicate that the performance achieved by RiB scheme is better suited for applications with high update rate and an exclusive focus on the latest data, such as autonomous driving and tele-surgery. Furthermore, our simulations encompass an examination of the average delay, alongside a detailed exploration of the trade-off between AoI and delay.

II. SYSTEM MODEL

A. System Description

For the sake of generality, we consider a point-to-point communication system that equipped with a buffer, which is illustrated in Fig. 1. The source generates N-bit information according to a Poisson process with update rate λ, and the information will be encoded into a packet of m symbols. We also assume that the buffer can accommodate only one packet, which is in accordance with the process of data acquisition and updating by sensors in IoT systems. As update rate is frequently high in IoT-FBL systems, such as in intelligent driving or industrial control systems, implementing this buffer can not only prevent transmission interruptions caused by preemption but also mitigate the risk of overload due to infinite queuing. In addition, the duration of each symbol is T_s, and the transmission time of each packet is $M = mT_s$.

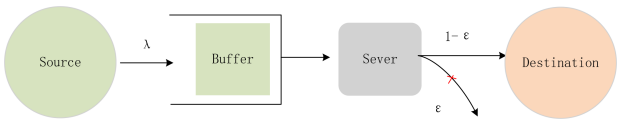


Fig. 1: System model of the packet delivery process, includes a source, a buffer, a server and a destination.

As shown in Fig. 1, the BLER during the transmission is ε, measuring the probability of packet errors in the transmission

process. Additionally, according to the prior study in [39], ε can be given as

$$\varepsilon \approx \mathbb{Q}\left(\frac{\frac{1}{2}\log_2(1+\gamma)-\frac{N}{m}}{\log_2(e)\sqrt{1-\frac{1}{(1+\gamma)^2}}}\right), \tag{1}$$

where γ is the received SNR, and the formula of the Q-function is $\mathbb{Q}(x)=\int_x^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$.

B. The Packet management Schemes

We propose two packet management schemes in this article, which are summarized as follows:

- *No Replacement* (NR): When the server is idle, an arriving packet is transmitted immediately. However, if the server is busy, the behavior depends on the buffer state: 1) if the buffer is empty, the arriving packet is stored in the buffer; 2) if the buffer is full, the arriving packet is discarded. This rule prioritizes direct transmission or conditional buffering, ensuring packets are only retained when there is available buffer space.
- *Replace in Buffer* (RiB): When the server is idle, an arriving packet is transmitted immediately. If the server is busy, the arriving packet is always stored in the buffer, regardless of whether the buffer is empty or full. This scheme emphasizes retaining all incoming packets during busy periods, requiring dynamic buffer management.

A fundamental premise supporting this mechanism requires the instantaneous packet clearance from the buffer upon transmission initiation, creating capacity for subsequent updates. Additionally, the system relies on real-time feedback from the destination to the source. Specifically, if a decoding failure is reported, the source persistently retransmits the current packet until either 1) successful reception is confirmed by the destination or 2) a new packet is generated. Notably, if a new packet generates, the source abandons unresolved retransmission attempts and prioritizes transmitting the latest packet.

For the sake of clear observation, the evolution of AoI is depicted in Fig. **??**, where on the horizontal axis, g_i denotes the generation time of the i-th update, and d_i denotes the successfully received time of the i-th update. The 'x' denotes the failure of a packet transmission. In addition to that, W_i denotes the waiting time in the buffer of the i-th successfully transmitted packet, T_i is the time staying in the system of the i-th successfully transmitted packet, Y_i denotes the interval between two consecutively successful transmission d_i and d_{i-1}, K_i is the interval form the start of the service immediately after a successful transmission to the next successful transmission, and X_i is the generation interval between g_i and g_{i+1}. Since the status updates are generated according to a Poisson process, X_i thus follows an exponential distribution with a probability density function (PDF) given by $f_{X_i}(x)=\lambda e^{-\lambda x}$.

As illustrated in Fig. **??** (a), two packets arrive sequentially in the network while the first packet is still being processed. In accordance with the NR scheme, the third packet is discarded. In contrast, under the RiB scheme, the third packet replaces the second packet and is ultimately transmitted, as depicted in Fig. **??** (b). In other words, while the NR scheme does not allow for buffered packets to be replaced, the RiB scheme permits buffered packets to be replaced. This highlights the key distinction between the two schemes.

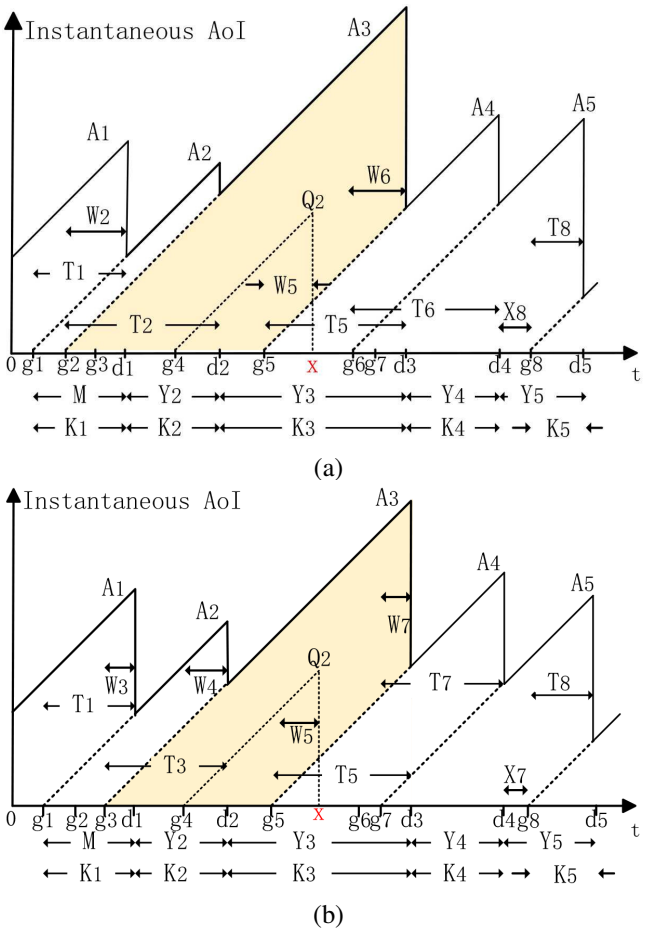


Fig. 2: The impact of SNR on the average peak AoI: λ is 1 packets/ms.

III. The Information Freshness Analysis

As mentioned before, we adopt AoI as the criterion of information freshness. In line with Fig. **??**, the average peak AoI can be given as $A_i=T_{i-1}+Y_i$. Let $N_t=\max\{\mathrm{i}|d_i<t\}$ be the number of successfully received packets during the interval $[0,t]$, the average AoI can be described as the sum of geometric areas under the instantaneous AoI curve, which is

$$\Delta=\lim_{t\to\infty}\frac{N_t}{t}\frac{1}{N_t}\sum_{i=1}^{N_t}Q_i=\lim_{t\to\infty}\frac{N_t}{t}\mathbb{E}\{Q_i\}, \tag{2}$$

where $\mathbb{E}$ represents the expectation operator. According to Fig. **??**, the area of Q_i can be expressed as

$$Q_i=\frac{(Y_i+T_{i-1})^2}{2}-\frac{{T_i}^2}{2}. \tag{3}$$

Since Y_i and T_i are mutually independent, we can obtain

$$\mathbb{E}\{Q\}=\frac{\mathbb{E}\{Y^2\}}{2}+\mathbb{E}\{Y\}\mathbb{E}\{T\}. \tag{4}$$

In addition, we have $\lim_{t\to\infty}\frac{N_t}{t}=\frac{1}{\mathbb{E}\{Y\}}$, then the average AoI can be simplified as

$$\Delta=\frac{\mathbb{E}\{Y^2\}}{2\mathbb{E}\{Y\}}+\mathbb{E}\{T\}, \tag{5}$$

and the average peak AoI can be rephrased as

$$\Delta^P=\mathbb{E}\{Y\}+\mathbb{E}\{T\}. \tag{6}$$

In the following section, we evaluate the average AoI and the average peak AoI of the proposed schemes by $\mathbb{E}\{Y\}$, $\mathbb{E}\{T\}$ and $\mathbb{E}\{Y^2\}$, which respectively are the expectations of Y_i, T_i and Y_i^2.

A. AoI Analysis for the NR Scheme

In the NR system, a packet may undergo several potential processes upon its arrival: 1) it may be transmitted immediately, 2) it may be stored in a buffer, or 3) it may be discarded outright. The determination of the specific process is contingent upon the server's current occupied status and the availability of buffer space. Assuming that the buffer has the capacity to accommodate only one packet, we can derive the probability of the server being idle by calculating the likelihood of no new arrivals occurring during the service period of the preceding packet, which is expressed as follows

$$p_i=\mathrm{e}^{-\lambda M}. \tag{7}$$

Otherwise, the probability of a busy sever after the last transmission is $p_b=1-p_i=1-e^{-\lambda M}$. Thus, $\mathbb{E}\{Y\}$ can be expressed by

$$\mathbb{E}\{Y\}=p_i(\mathbb{E}\{X\}+\mathbb{E}\{K\})+p_b\mathbb{E}\{K\}, \tag{8}$$

where the first item with probability p_i signifies that the system needs to wait for an update after a successful transmission. And the second item with probability p_b denotes that there has been a packet in the buffer after the last transmission. Concretely, if the transmission of the i-th packet is successful, we have $K_i=M$, otherwise $K_i=M+p_iX_{i+1}+K_i^{'}$. This is because there is a probability of packet errors in the FBL regime, then the server needs to wait for a new update with probability p_i if there is no packet in the buffer. Due to the memory-less nature of the Poisson process, the waiting time and the interval between the generation of state updates obey the same distribution, which can be represented by X_{i+1}. Then $\mathbb{E}\{K\}$ is

$$\mathbb{E}\{K\}=(1-\varepsilon)\,M+\varepsilon\mathbb{E}(M+p_iX+\hat{K}), \tag{9}$$

where $\mathbb{E}\{X\}=\frac{1}{\lambda}$ and $\hat{K}$ is the remaining process of transmitting a new packet when the previous transmission is unsuccessful, and $\hat{K}$ has the same distribution as K, which gives

$$\mathbb{E}\{K\}=\frac{M}{1-\varepsilon}+\frac{\varepsilon e^{-\lambda M}}{\lambda(1-\varepsilon)}. \tag{10}$$

Substituting the results of $\mathbb{E}\{K\}$, $\mathbb{E}\{X\}$ into (8), we have,

$$\mathbb{E}\{Y\}=\frac{M}{1-\varepsilon}+\frac{e^{-\lambda M}}{\lambda(1-\varepsilon)}. \tag{11}$$

According to Fig. **??**, the experienced time of the successfully transmitted packet is $T_i=p_iK_i+p_b(W_i+K_i)$, where the waiting time is $W_i=T_{i-1}-X_i$. Then the expectation of T can be given by

$$\mathbb{E}\{T\}=p_i\mathbb{E}\{K\}+p_b(\mathbb{E}\{W\}+\mathbb{E}\{K\}), \tag{12}$$

and the waiting time in the buffer, say, $\mathbb{E}\{W\}$, will be

$$\mathbb{E}\{W\}=\mathbb{E}\{T\}-\mathbb{E}\{X|X<M\}. \tag{13}$$

The conditional expectation $\mathbb{E}\{X|X<M\}$ is the generation interval under the condition that the source generates an update before the current transmission is completed, which is

$$\mathbb{E}\{X|X<M\}=\frac{1}{\lambda}-\frac{Me^{-\lambda M}}{1-e^{-\lambda M}}. \tag{14}$$

Combining (10), (12) with (13) will give $\mathbb{E}\{T\}$ as

$$\mathbb{E}\{T\}=\frac{Me^{\lambda M}}{(1-\varepsilon)}+\frac{\varepsilon}{\lambda(1-\varepsilon)}+M-\frac{1-e^{-\lambda M}}{\lambda e^{-\lambda M}}. \tag{15}$$

Substituting (11) and (15) into (6), which gives the expression of the average peak AoI as

$$\Delta_{NR}^P=\frac{M+Me^{\lambda M}}{1-\varepsilon}+\frac{e^{-\lambda M}+\varepsilon}{\lambda(1-\varepsilon)}+M-\frac{(1-e^{-\lambda M})}{\lambda e^{-\lambda M}}. \tag{16}$$

To obtain the closed-form expression of the average AoI, we need to derive the expectation of $\mathbb{E}\left\{Y^2\right\}$. Due to the independence between W_i and K_i, the following equation holds.

$$\mathbb{E}\{Y^2\}=p_i(\mathbb{E}\{X^2\}+\mathbb{E}\{K^2\}+2\mathbb{E}\{X\}\mathbb{E}\{K\})+p_b\mathbb{E}\{K^2\}, \tag{17}$$

where $\mathbb{E}\{X^2\}=\frac{2}{\lambda^2}$. Then according to (10), we have

$$\begin{aligned}\mathbb{E}\{K^2\}&=(1-\varepsilon)M^2+\varepsilon(M^2+{p_i}^2\mathbb{E}\{X^2\}+\mathbb{E}\{K^2\}\\&+2p_iM\mathbb{E}\{X\}+2M\mathbb{E}\{K\}+2p_i\mathbb{E}\{X\}\mathbb{E}\{K\}).\end{aligned} \tag{18}$$

Combining $\mathbb{E}\{K\}$, $\mathbb{E}\{X\}$, $\mathbb{E}\{X^2\}$ with $\mathbb{E}\{K^2\}$, we have

$$\begin{aligned}\mathbb{E}\{Y^2\}&=\frac{2e^{-\lambda M}}{\lambda^2}+\frac{M^2}{1-\varepsilon}+\frac{2\varepsilon e^{-2\lambda M}}{\lambda^2(1-\varepsilon)}+\\&\frac{2\varepsilon Me^{-\lambda M}}{\lambda(1-\varepsilon)}+\frac{2e^{-\lambda M}+2\varepsilon\lambda M}{\lambda(1-\varepsilon)}\{\frac{M}{(1-\varepsilon)}+\frac{\varepsilon e^{-\lambda M}}{\lambda(1-\varepsilon)}\}.\end{aligned} \tag{19}$$

Consequently, with the expressions of $\mathbb{E}\{Y\}$, $\mathbb{E}\{T\}$ and $\mathbb{E}\{Y^2\}$ in hand, the average AoI of the NR scheme will be

$$\begin{aligned}\Delta_{NR}&=\frac{Me^{\lambda M}}{(1-\varepsilon)}+\frac{\varepsilon}{\lambda(1-\varepsilon)}+M-\frac{(1-e^{-\lambda M})}{\lambda e^{-\lambda M}}\\&+\frac{1}{\lambda M+e^{-\lambda M}}\{\frac{(1-\varepsilon)e^{-\lambda M}+\varepsilon e^{-2\lambda M}}{\lambda}+\frac{\lambda M^2}{2}\\&+\varepsilon Me^{-\lambda M}+\frac{(e^{-\lambda M}+\varepsilon\lambda M)(\lambda M+\varepsilon e^{-\lambda M})}{\lambda(1-\varepsilon)}\}.\end{aligned} \tag{20}$$

B. AoI Analysis for the RiB Scheme

In the RiB scheme, the packet may undergo the following processes after arrival: 1) transmitted immediately, 2) wait in the buffer and transmitted ultimately, 3) replaced by a new arrival. Therefore, it is necessary to consider the probability of replacement. Note that replacement only impacts the duration

that packets within the buffer and does not affect the interval between two consecutive successful transmissions. In this case, the partial results obtained in subsection A are still feasible, i.e., $\mathbb{E}\{K\}$, $\mathbb{E}\{Y\}$ and $\mathbb{E}\{Y^2\}$. We thus only need to focus on $\mathbb{E}\{T\}$ in this subsection. In order to obtain $\mathbb{E}\{T\}$, we need to consider the probability of replaced packets. Assuming that there are at most two updates, or rather once replacement within time M. Leveraging the characteristics of the Poisson distribution, we can calculate the probabilities of generating one and two updates as follows,

$$\begin{aligned} p_1 &= e^{-\lambda M}\frac{(\lambda M)^1}{1!} = e^{-\lambda M}(\lambda M), \\ p_2 &= e^{-\lambda M}\frac{(\lambda M)^2}{2!} = \frac{e^{-\lambda M}(\lambda M)^2}{2}. \end{aligned} \tag{21}$$

We afterwards define the absence of replacement as event ϕ, and the opposite as event ψ. The waiting time of the successfully transmitted packet in the buffer under the RiB scheme can be given by

$$\mathbb{E}\{W\} = p_1(\mathbb{E}\{T\} - \mathbb{E}\{X|\phi\}) + p_2(\mathbb{E}\{T\} - \mathbb{E}\{\hat{X}|\psi\}), \tag{22}$$

where $\hat{X}$ represents the cumulative sum of two update intervals. Hence, we have

$$\begin{aligned} \mathbb{E}\{X|\phi\} &= \frac{\int_0^M x\lambda e^{-\lambda x}dx}{\lambda M e^{-\lambda M}} = \frac{1-e^{-\lambda M}}{\lambda^2 M e^{-\lambda M}} - \frac{1}{\lambda}, \\ \mathbb{E}\left\{\hat{X}|\psi\right\} &= \frac{\int_0^M \frac{\lambda}{2}xe^{-\frac{\lambda x}{2}}dx}{\frac{\lambda M}{2}e^{\frac{-\lambda M}{2}}} = \frac{4(e^{\frac{\lambda M}{2}}-1)}{\lambda^2 M} - \frac{2}{\lambda}. \end{aligned} \tag{23}$$

Then substituting (10), (22) into (12), the time of packet within the buffer can be expressed by

$$\begin{aligned} \mathbb{E}\{T\} = \frac{1}{(2-e^{-\lambda M})}\{&\frac{Me^{\lambda M}}{1-\varepsilon} + \frac{\varepsilon}{\lambda(1-\varepsilon)} + \\ &(1-e^{-\lambda M})(M+\lambda M^2 - 2M(e^{\frac{\lambda M}{2}}-1) - \frac{e^{\lambda M}-1}{\lambda})\}. \end{aligned} \tag{24}$$

Substituting (11) and (24) into (6) will give the average peak AoI as

$$\begin{aligned} \Delta_{RiB}^P = &\frac{M}{1-\varepsilon} + \frac{e^{-\lambda M}}{\lambda(1-\varepsilon)} + \frac{1}{(2-e^{-\lambda M})}\{\frac{\lambda M e^{\lambda M}+\varepsilon}{\lambda(1-\varepsilon)} \\ &+ (1-e^{-\lambda M})(M+\lambda M^2 - 2M(e^{\frac{\lambda M}{2}}-1) - \frac{e^{\lambda M}-1}{\lambda})\}. \end{aligned} \tag{25}$$

In line with the analysis in subsection A, the average AoI of the RiB scheme can be given as

$$\begin{aligned} \Delta_{RiB} = &\frac{1}{(2-e^{-\lambda M})}\{\frac{\lambda M e^{\lambda M}+\varepsilon}{\lambda(1-\varepsilon)} + \\ &(1-e^{-\lambda M})(M+\lambda M^2 - 2M(e^{\frac{\lambda M}{2}}-1) - \frac{e^{\lambda M}-1}{\lambda})\} \\ &+ \frac{1}{\lambda M + e^{-\lambda M}}\{\frac{(1-\varepsilon)e^{-\lambda M}+\varepsilon e^{-2\lambda M}}{\lambda} + \frac{\lambda M^2}{2} + \\ &\varepsilon M e^{-\lambda M} + \frac{(e^{-\lambda M}+\varepsilon\lambda M)(\lambda M + \varepsilon e^{-\lambda M})}{\lambda(1-\varepsilon)}\}. \end{aligned} \tag{26}$$

Remark 1: Based on the analysis, the difference between the RiB scheme and the NR scheme is rooted in $\mathbb{E}\{T\}$, as detailed in equations (15) and (24). This is also the main distinction between (16) and (25), (20) and (26), which emerges from the way packets are handled and the impact of replacement depicted in Fig. **??**. In scenarios where λ is elevated—indicating a higher replacement probability—the RiB scheme demonstrates more considerable advantage on reduction in $\mathbb{E}\{T\}$. We present and confirm this trend in the part of simulation results.

IV. The Optimal Blocklength and Information Timeliness Analyzes

A. The Optimal Blocklength Analysis

Determining the appropriate blocklength for various systems is crucial, especially for the freshness-sensitive applications. According to the closed-form expressions derived above, it is clear to observe the significant impacts of blocklength on the AoI. Specifically, a smaller packet blocklength can potentially reduce the average AoI as $M = mTs$ is diminished. Nevertheless, this reduction is countered by an increase of BLER augmented by reducing m. Hence, there exists an optimal blocklength that minimizes the value of the AoI with given update rate λ. Take the average peak AoI as example, We analyse the problem theoretically. For the NR scheme, the optimal packet blocklength is the solution to the formula $\frac{\mathrm{d}\Delta_{NR}^P}{\mathrm{d}m} = 0$, which can be calculated by

$$\begin{aligned} &\frac{T_s e^{-\lambda M} - T_s(1+(1+\lambda M)e^{\lambda M})}{1-\varepsilon} + T_s(1-\frac{1}{e^{-\lambda M}}) = \\ &\qquad \frac{M(1+e^{\lambda M}) + \frac{e^{-\lambda M}}{\lambda}}{(1-\varepsilon)^2}\frac{\partial\varepsilon}{\partial m}, \end{aligned} \tag{27}$$

The derivative of the Q-function can be expressed as follows:

$$\frac{\partial Q(x)}{\partial x} = -\frac{exp(-\frac{x^2}{2})}{\sqrt{2\pi}}, \tag{28}$$

then the term $\frac{\partial\varepsilon}{\partial m}$ in (27) is given by

$$\frac{\partial\varepsilon}{\partial m} = -\frac{\exp\left(-\frac{\Xi^2}{2}\right)}{\sqrt{2\pi}}\frac{\sqrt{2}\left(\frac{N}{m\sqrt{m}} + \frac{\log_2(1+\gamma)}{2\sqrt{m}}\right)}{2\log_2(e)\sqrt{1-\frac{1}{(1+\gamma)^2}}}, \tag{29}$$

where the letter Ξ is that

$$\Xi = \frac{\frac{1}{2}\log_2(1+\gamma) - \frac{N}{m}}{\log_2 e\sqrt{\frac{1}{2m}(1-\frac{1}{(1+\gamma)^2})}}. \tag{30}$$

Substituting (1), (28), (29), and (30) into (27) produces a complete equation involving the parameters λ, γ, m, N, T_s. If all parameter values except m are known, they can be substituted into (27) to yield a numerical solution for the optimal blocklength.

For the RiB scheme, the optimal blocklength is the solution to the equation $\frac{\mathrm{d}\Delta_{RiB}^{P}}{\mathrm{d}m}=0$, which can be given by

$$\frac{T_s(e^{-\lambda M}-1)}{1-\varepsilon}+\frac{\lambda T_s e^{-\lambda M}\cdot\Omega-(2-e^{-\lambda M})\frac{\partial\Omega}{\partial m}}{(2-e^{-\lambda M})^2}= \frac{\lambda M+e^{-\lambda M}}{\lambda(1-\varepsilon)^2}\frac{\partial\varepsilon}{\partial m}, \tag{31}$$

where Ω is in (25) and the term $\frac{\partial\Omega}{\partial m}$ is given by

$$\begin{aligned}\frac{\partial\Omega}{\partial m}=&\frac{T_s e^{\lambda M}}{1-\varepsilon}+\left(\frac{Me^{\lambda M}}{(1-\varepsilon)^2}+\frac{1}{\lambda(1-\varepsilon)^2}\right)\frac{\partial\varepsilon}{\partial m}+\\ &\lambda T_s e^{-\lambda M}\left(3M+\lambda M^2-2Me^{\frac{\lambda M}{2}}-\frac{e^{\lambda M}-1}{\lambda}\right)+\\ &(1-e^{-\lambda M})(T_s+2\lambda T_s M-2T_s(e^{\frac{\lambda M}{2}}-1)\\ &-\lambda T_s M e^{\frac{\lambda M}{2}}-T_s e^{\lambda M})\end{aligned} \tag{32}$$

The numerical solution for the optimal block length can be obtained by substituting parameter values into (31) using the same method as in the NR scheme. This method enables the determination of the optimal blocklength while maintaining low complexity.

B. The Information Timeliness Analysis

We adopted the delay as the criterion of information timeliness. The average delay is used in this article. To assess the performance of transmission delay in the considered system, we analysis the expressions of average delay incurred by the RiB and NR schemes during data packet transmission.

The delay of each packet in the network includes transmission delay, propagation delay, processing delay and queuing delay. We focus on transmission delay and queuing delay in this article, which play a dominant role in the total delay [40]. As shown in Fig. 3, the delay of each packet solely pertains to its actions, rather than the successive packets.

The re-transmission policy is adopted to increase the transmission reliability if the buffer is empty in the proposed schemes. The introduction of this policy will not affect AoI, which is analyzed from the perspective of the receiving end. Note that only the successfully received packets are taken into consideration in the analysis. As described before in the analysis of AoI, assuming the number of successfully received packets during the interval $[0,t]$ as N_t, and then the average delay can be given by

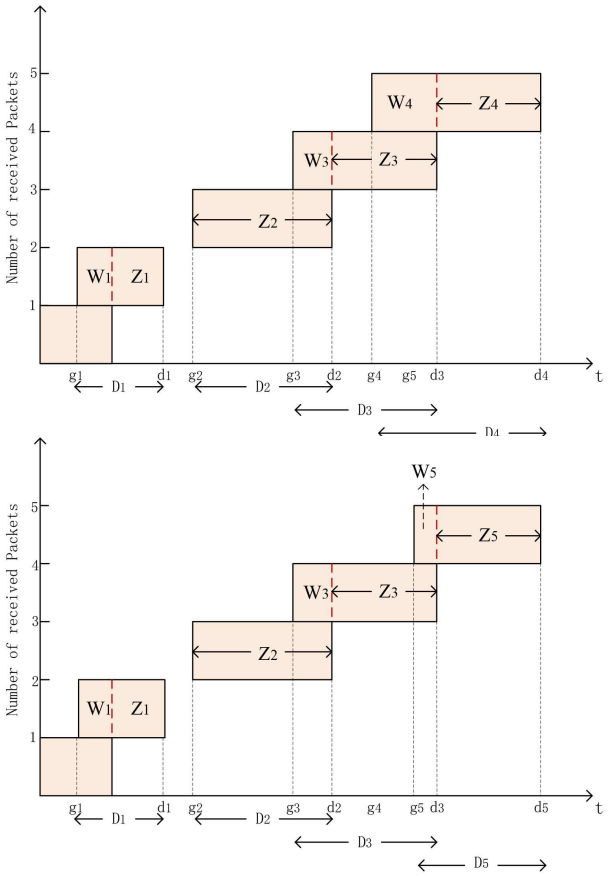
$$\bar{D}=\lim_{t\to\infty}\frac{1}{N_t}\sum_{i=1}^{N_t}D_i=\mathbb{E}\{W\}+\mathbb{E}\{Z\}, \tag{33}$$

where $\mathbb{E}\{W\}$is the waiting time in the queue and $Z_i=(1-\varepsilon)\mathbb{E}\{S\}+\varepsilon(\mathbb{E}\{S\}+p_i Z_{i+1})$ means the transmission delay with the RN policy. The term Z_i and Z_{i+1} obey the same distribution. Then the $\mathbb{E}\{Z\}$ can be easily expressed by

$$\mathbb{E}\{Z\}=\frac{M}{1-\varepsilon e^{-\lambda M}}. \tag{34}$$

According to the similar steps in (13), the average waiting time of packets of the NR scheme will be

$$\mathbb{E}\{W\}=\frac{\lambda M e^{\lambda M}+\varepsilon}{\lambda(1-\varepsilon)}+M-\frac{1}{\lambda e^{-\lambda M}}+\frac{Me^{-\lambda M}}{1-e^{-\lambda M}}. \tag{35}$$



(b)

Fig. 3: The evolution of delay: (a) sample path of the NR scheme; (b) sample path of the RiB scheme.

Subsequently, the average delay of the NR scheme will be

$$\begin{aligned}\bar{D}_{NR}=&\frac{M}{1-\varepsilon e^{-\lambda M}}+\frac{\lambda M e^{\lambda M}+\varepsilon}{\lambda(1-\varepsilon)}+M-\frac{1}{\lambda e^{-\lambda M}}\\ &+\frac{Me^{-\lambda M}}{1-e^{-\lambda M}}.\end{aligned} \tag{36}$$

The average delay of the RiB scheme can be obtained while following a similar method. We omit the detailed derivations and give the final result of it here, which is

$$\begin{aligned}\bar{D}_{RiB}=&\frac{M}{1-\varepsilon e^{-\lambda M}}+\mathbb{E}\{T\}-\frac{1-e^{-\lambda M}}{\lambda}+Me^{-\lambda M}\\ &-2M(e^{-\frac{\lambda M}{2}}-e^{-\lambda M})+\lambda M^2 e^{-\lambda M},\end{aligned} \tag{37}$$

where the term $\mathbb{E}\{T\}$ is given by (24).

Remark 2: In addition to the relationship with blocklength, delay also exhibits a significant correlation with update rate described in (36) and (37). As previously analyzed, both excessively low and high update rate can adversely affect the AoI. Therefore, further investigation is warranted to elucidate the interaction between delay and update rate. By approximating formula (36) using the first-order Taylor approximation of

$e^{-\lambda M}$ around the point 0, we can observe that

$$\tilde{D}_{NR} = \frac{M}{1-\varepsilon} + \frac{M(1+\lambda M)}{1-\varepsilon} + \frac{\varepsilon}{\lambda(1-\varepsilon)} + \frac{1}{\lambda} - \frac{1}{\lambda(1-\lambda M)}. \quad (38)$$

Then the derivative about λ of (38) is that

$$\frac{\partial \tilde{D}_{NR}}{\partial \lambda} = \frac{M^2}{1-\varepsilon} + \frac{1-2\lambda M}{\lambda^2(1-\lambda M)^2} - \frac{1}{\lambda^2(1-\varepsilon)}. \quad (39)$$

By inspecting (39), it can be deduced that it reaches a value of 0 at a certain point. Hence, we can infer that the delay of the NR scheme initially decreases and then increases as λ rises. Conversely, exploring the delay of the RiB scheme in the same manner, we find that

$$\tilde{D}_{RiB} = \frac{2M}{1-\varepsilon} + \frac{\varepsilon}{\lambda(1-\varepsilon)(1+\lambda M)} - \lambda M^2 - \lambda^2 M^3, \quad (40)$$

and

$$\frac{\partial \tilde{D}_{RiB}}{\partial \lambda} = \frac{\varepsilon}{1-\varepsilon} \frac{-1-2\lambda M}{\lambda^2(1+\lambda M)^2} - M^2 - 2\lambda M^2. \quad (41)$$

Formulas (40) and (41) reveals that regardless of the value of λ, $\frac{\partial \tilde{D}_{RiB}}{\partial \lambda}$ is always negative ($\lambda > 0$). As λ increases, the delay of RiB scheme consistently diminishes. In other words, the RiB scheme is particularly suitable for mission-critical scenarios characterized by higher update rate when considering delay performance. This proposition is supported by the simulation results.

V. Numerical Results

In the simulations, we set the information of each status update as $N = 160$ bits, the SNR as 3 dB. The duration time for each symbol is that $T_s = 0.006$ ms. We respectively simulate the AoI versus the blocklenth m and update rate λ under different schemes and system settings. The AoI curve of the infinite queue system is incomplete here due to the congestion in the queue when $\lambda > \frac{1-\varepsilon}{M}$.

Fig. 4 depicts the average peak AoI and the average AoI in relation to the blocklength m from 160 to 240 channel use (cu) with different update rate: 0.5 packets/ms, 1 packets/ms and 2 packets/ms. We can see that the RiB scheme outperforms the others, especially when λ is 2 packets/ms, showcasing the superiority of replacing with newly arrived packets. Nevertheless, the infinite queue system achieves a lower AoI when the blocklength is small. This is because packets waiting in the queue can be transferred quickly, while systems with no packets in the buffer need to wait for the next update. Moreover, the AoI of the NR scheme is greatly large when λ is 1 and 2 packets/ms, because the packet in the buffer is out of date in this scheme. The bufferless system employing non-preemptive scheme exhibits intermediate AoI performance between the NR and RiB schemes, as it inherently eliminates queuing delays.

The AoI trajectory observed across all schemes-decreasing to a minimum near 180 cu before increasing-results from the fundamental tradeoff between transmission reliability and transmission time. When $m < 180$ cu, AoI minimization is

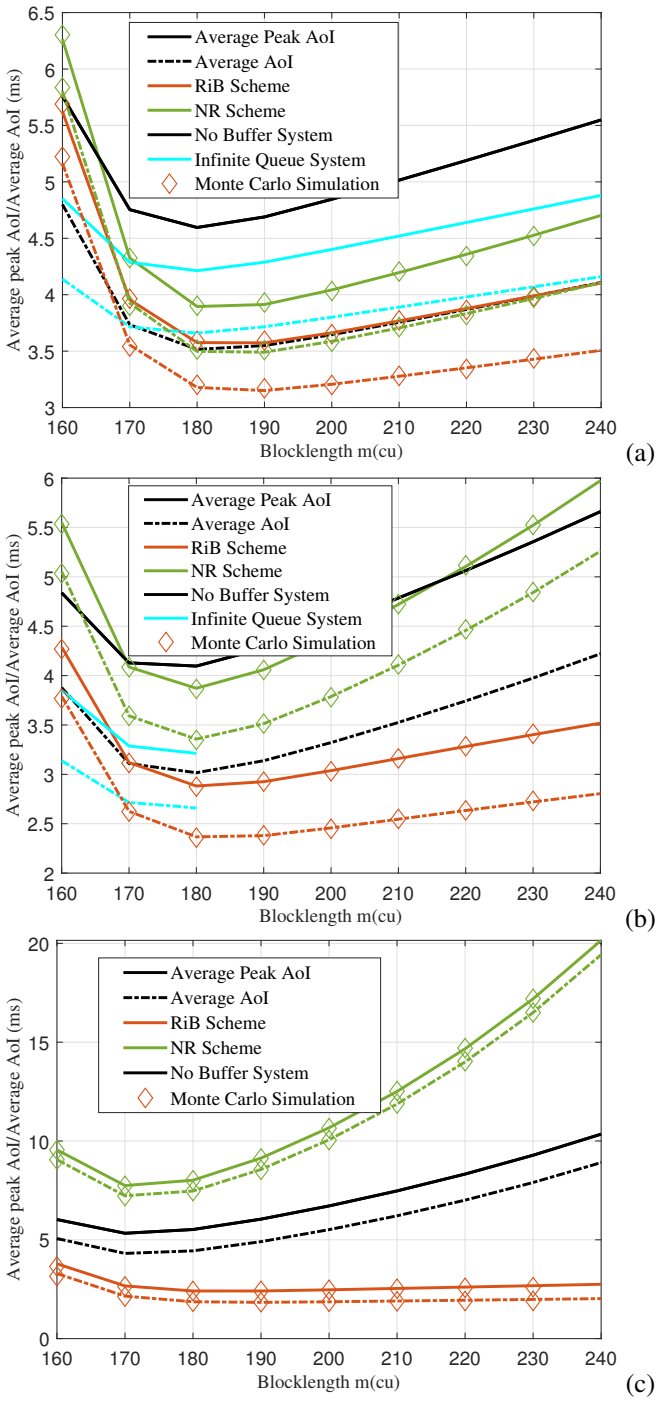


Fig. 4: The average peak AoI and the average AoI versus blocklength m: (a) $\lambda = 0.5$ packets/ms; (b) $\lambda = 1$ packets/ms; (c) $\lambda = 2$ packets/ms.

dominated by the exponential decay of BLER with increasing blocklength. Conversely, when $m > 180$ cu, the linear scaling of service time with blocklength emerges as the dominant constraint, and the marginal improvement in BLER cannot compensate for the increased transmission time. This dual optimization leads to the observed minimum AoI at a critical blocklength of approximately 180 cu.

Fig. 5 displays the average peak AoI and the average AoI

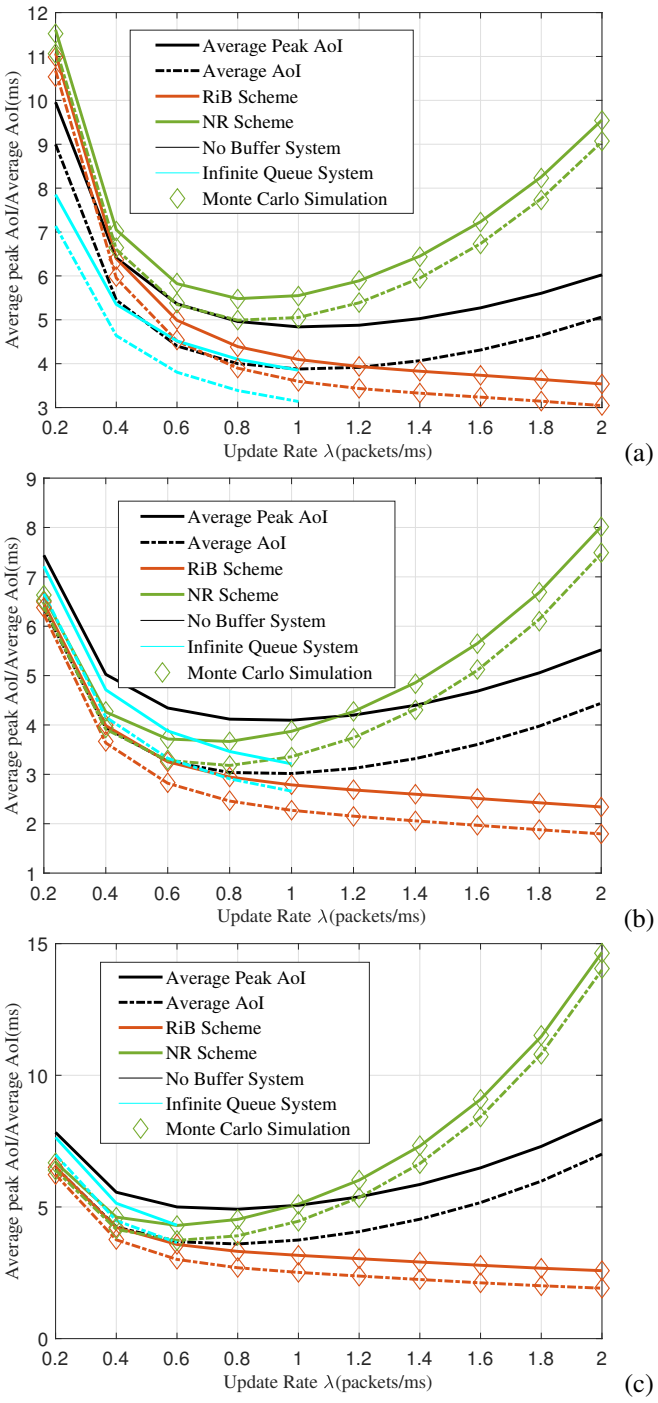


Fig. 5: The average peak AoI and the average AoI versus update rate λ: (a) m=160 cu; (b) m=180 cu; (c) m=220 cu.

to the update rate λ from 0.2 to 2 packets/ms with different blocklength: 160 cu, 180 cu, 220 cu. As displayed here, when m is 160 cu and λ is less than 1 packets/ms, the AoI of the infinite queue system is the lowest. In the RiB scheme, the value of AoI continuously decreases as λ increases and remains lower than that of the NR scheme and the no buffer scheme, because the replacement improves the freshness of the buffer packet and enables the system to achieve a low AoI at higher λ. The NR scheme exhibits similar performance of AoI to the RiB scheme when λ is low, which complies with the system settings, and the AoI increases significantly at moderately higher λ as the buffer is prone to retain the previous packets.

The scheme selection mechanism operates under these framework. For instance, in safety-critical vehicular networks requiring update rates >1 packet/ms (e.g., collision avoidance systems), the RiB strategy proves advantageous for maintaining real-time control.

The effects of varying SNR on the average peak AoI are highlighted in Fig. 6. It is apparent that, all else being equal, a lower AoI is observed at an SNR of 5 dB. This is attributed to the fact that a higher SNR reduces the probability of packet errors during transmission. Notably, when the SNR is sufficiently high, it allows for a smaller blocklength to be utilized. This reduction in blocklength contributes to decreased transmission delays, which in turn significantly improves the freshness of the information received at the destination. Thus, optimizing SNR is crucial for enhancing the efficiency of information delivery and minimizing AoI.

Fig. 7 illustrates the optimal average peak AoI versus the update rate with the settings of the optimal packet blocklength for the two proposed schemes. The optimal values of the packet blocklength are set to the numerical results analyzed by (27) and (31), respectively. Specifically, the optimal packet blocklength of the system without buffer and packet management can be obtained by a one-dimensional exhaustive search. From the comparison in Fig. 7, the influence of SNR on the optimal average peak AoI are evident, as SNR and blocklength concurrently affect BLER. The figure demonstrates that increasing the transmission SNR enables more aggressive blocklength compression while maintaining reliability constraints, thereby achieving lower AoI. The SNR-blocklength-AoI correlation plays a pivotal role in optimizing communication system performance.

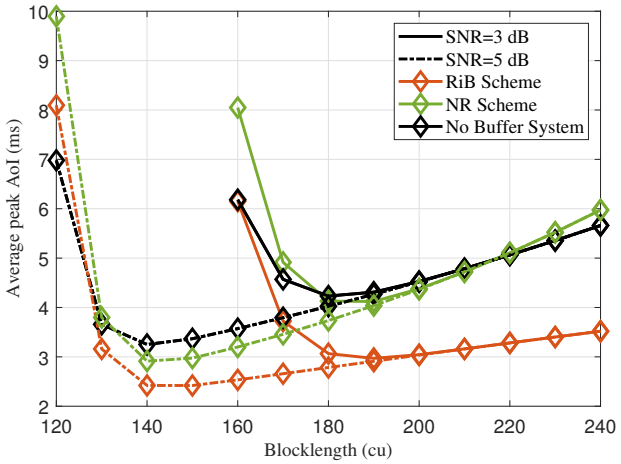


Fig. 6: The impact of SNR on the average peak AoI: λ is 1 packets/ms.

Fig. 8 illustrates the relationship between average delay, blocklength, and update rate. Examining the graph, the average delay is significantly affected by both the update rate and the blocklength. It reveals that extreme values for blocklength

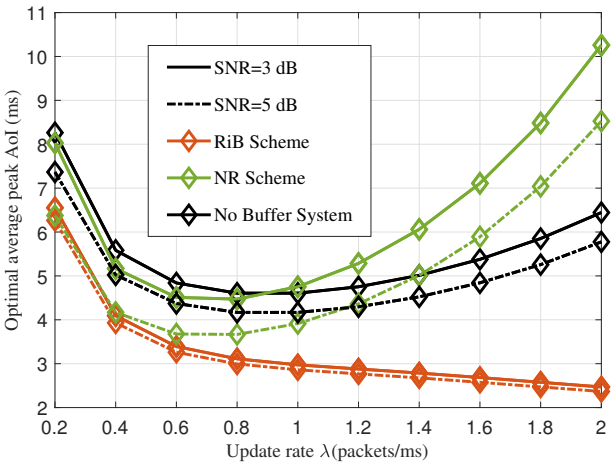


Fig. 7: The optimal average peak AoI versus the update rate λ with optimal packet blocklength.

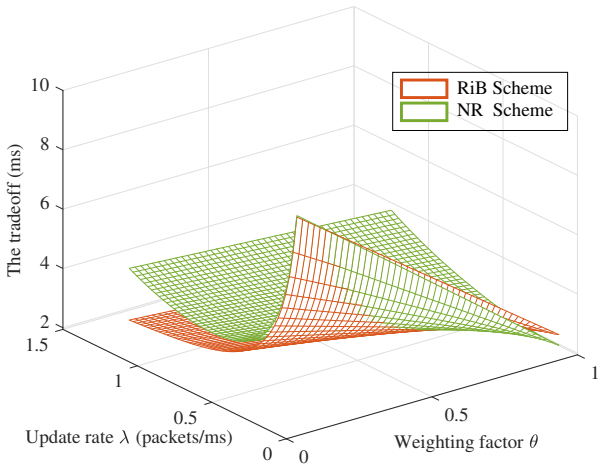


Fig. 9: The tradeoff between the average peak AoI and delay is $(1-\theta)\Delta^P + \theta\bar{D}$, where θ is the weighting factor and the blocklength m is 180 cu.

result in an increase in average delay. Specifically, either a large or a small blocklength will augment the average delay, which is consistent with the analysis of AoI. In particular, the increase in λ tends to have a much more substantial effect on the NR scheme compared to other scenarios. The underlying reason stems from the no-replace principle inherent in the NR scheme, which leads to extended waiting times for packets in the queue. As λ rises, the congestion in the queue becomes more pronounced, thereby exacerbating the delays experienced by packets. Consequently, this interplay between λ, blocklength, and average delay highlights the complexities involved in optimizing communication systems for timely information delivery.

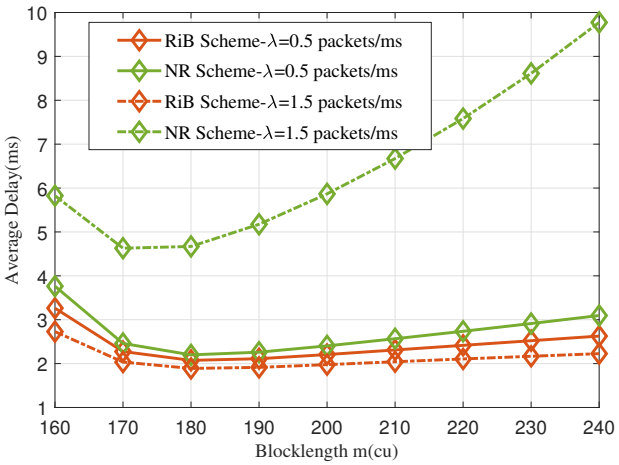


Fig. 8: The average delay versus the blocklength m and the update rate λ.

In addition, we investigate the tradeoff between the AoI and delay that $(1-\theta)\Delta^P + \theta\bar{D}$, where θ is the weighting factor from 0.1 to 1. The tradeoff can be used in a variety of scenarios, considering both the performance of AoI and delay. A larger value of θ indicates that the system places greater emphasis on achieving low delay. Fig. 9 reveals the causality between the tradeoff and λ, θ. It can be observed from Fig. 9 that the overall tradeoff of the RiB scheme is smaller than that of the NR scheme with the identical system settings. This indicates that the RiB scheme demonstrates superior performance on AoI and delay. As θ approaches 1, which implies that the system increasingly prioritizes delay over AoI, the NR scheme might achieve a lower tradeoff than the RiB scheme, when λ is less than 0.5 packets/ms. In this scenario, the RiB scheme does not have distinct advantages over the NR scheme. Moreover, it is crucial to recognize that the tradeoff does not directly imply the system performance about delay and AoI, as the weight coefficient is tailored to meet the specific requirements of the applications.

The joint minimization of communication delay and AoI constitutes a critical design requirement in mission-critical applications where outdated or delayed data jeopardize operational integrity and safety. In autonomous systems, e.g., self-driving vehicles and drone swarms, low latency is critical for enabling real-time obstacle detection and path planning, while data freshness ensures the reliability of positional information in dynamic and unpredictable environments. In certain instances, selecting an appropriate weight factor becomes essential for systems aimed at effectively balancing delay and AoI metrics.

VI. Conclusion

This article investigates the impact of packet management schemes on the performance of information freshness and timeliness in the system under the FBL regime. We find that the RiB scheme outperforms the NR scheme, primarily due to its effective strategy for replacing the outdated packets in the buffer with newly arrived ones. This advantage allows for a more adaptive response to scenarios necessitating higher update rate, thereby enhancing the information freshness and timeliness performances. The distinction of NR and RiB scheme lies in their implementation: the NR scheme does not involve packet replacement, leading lower implementation complexity but poorer performance under high update rate, making it suitable for environmental and status monitoring

applications. In contrast, the RiB scheme employs replacement to always transmit the latest data, which is ideal for contexts like the Industrial IoT for the cost of data obsolescence is significant. Furthermore, our study gives a mathematical framework to obtain the optimal blocklength. The simulation results clarify the intricate relationship between AoI, average delay, blocklength, and update rate, revealing a trade-off between maintaining low average delay and achieving minimal average peak AoI. Building on the conclusions drawn, future studies could focus on enabling the adaptive adjustment of the weight factor in dynamic scenarios, as well as analyzing and optimizing parameters such as blocklength, update rate, and SNR to achieve an optimal trade-off between AoI and delay.

REFERENCES

[1] D. Zhang, J. J. P. C. Rodrigues, Y. Zhai, and T. Sato, "Design and implementation of 5G e-health systems: Technologies, use cases, and future challenges," *IEEE Communications Magazine*, vol. 59, no. 9, pp. 80–85, 2021.

[2] M. Tang, S. Cai, and V. K. N. Lau, "Online system identification and optimal control for mission-critical IoT systems over MIMO fading channels," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 157–21 173, 2022.

[3] S. Jia, R. Wang, Y. Lou, N. Wang, D. Zhang, K. Singh, and S. Mumtaz, "Secrecy performance analysis of UAV-assisted ambient backscatter communications with jamming," *IEEE Transactions on Wireless Communications*, 2024.

[4] L. D. Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

[5] L. Song, D. Zhang, S. Jia, P. Zhu, and Y. Li, "STAR-RIS-aided NOMA for secured xURLLC," *IEEE Transactions on Vehicular Technology*, 2025.

[6] S. Kim and J.-H. Na, "Improving latency using codes in mission-critical communication," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, 2017, pp. 730–734.

[7] N. Jiang, S. Yan, Z. Liu, C. Hu, and M. Peng, "Communication and computation assisted sensing information freshness performance analysis in vehicular networks," in *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2022, pp. 969–974.

[8] F. Mohammed, A. S. M. Kayes, E. Pardede, and W. Rahayu, "A framework for measuring iot data quality based on freshness metrics," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 1242–1249.

[9] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends® in Networking*, vol. 12, pp. 162–259, 11 2017.

[10] R. Liu, M. Hua, K. Guan, X. Wang, L. Zhang, T. Mao, D. Zhang, Q. Wu, and A. Jamalipour, "6G enabled advanced transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 10 564–10 580, 2024.

[11] X. Lan, Z. Xu, J. He, L. Liu, Q. Chen, and T. Li, "Information-freshness-aware wireless multiuser uplink physical-layer security communication," *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 15 940–15 956, 2024.

[12] H. Xue, D. Zhang, C. Wu, Y. Ji, S. Long, C. Wang, and T. Sato, "Parameter estimation-aided edge server selection mechanism for edge task offloading," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 2, pp. 2506–2519, 2024.

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

[14] J. Guo, D. Zhang, I. Lee, Y. Li, and M. Shirvanimoghaddam, "Partitioned analog fountain codes for short packet communications," *IEEE Communications Letters*, vol. 28, no. 6, pp. 1248–1252, 2024.

[15] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[16] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:332244

[17] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 2569–2573.

[18] G. Stamatakis, N. Pappas, and A. Traganitis, "Optimal policies for status update generation in an iot device with heterogeneous traffic," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5315–5328, 2020.

[19] Z. Li, A. Zhong, Y. Jiang, T. Tang, and R. Wang, "A bound on peak age of information distribution," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 88–93.

[20] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 1681–1685.

[21] B. Gabr, H. ElSawy, K. G. Seddik, and W. Mesbah, "Age of information for preemptive/non-preemptive transmissions in large-scale iot networks," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Dec 2022, pp. 329–334.

[22] M. K. Abdel-Aziz, S. Samarakoon, C.-F. Liu, M. Bennis, and W. Saad, "Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1911–1924, 2020.

[23] X. Ma, X. Jia, Y. Li, and K. Xue, "Minimizing age of information in wireless powered communication network based on short packet communication," in *2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*, 2024, pp. 1–5.

[24] Y. Zhang, Y. Chen, B. Yu, X. Diao, and Y. Cai, "Minimizing age of information based on predictions and short packet communications in uav relay systems," in *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2021, pp. 1–5.

[25] B. Han, Y. Zhu, Z. Jiang, Y. Hu, and H. D. Schotten, "Optimal blocklength allocation towards reduced age of information in wireless sensor networks," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[26] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 2569–2573.

[27] B. Yu, Y. Cai, D. Wu, and Z. Xiang, "Average age of information in short packet based machine type communication," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10 306–10 319, 2020.

[28] M. Moltafet, M. Leinonen, and M. Codreanu, "Moment generating function of age of information in multisource m/g/1/1 queueing systems," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6503–6516, 2022.

[29] ——, "Average age of information for a multi-source m/m/1 queueing model with packet management," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1765–1769.

[30] X. Zhang, X. Jia, H. Chang, and H. Tian, "Average age of information for multi-source single buffer preemption queueing model with packet dropping in service," in *2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*, 2024, pp. 1–5.

[31] R. Wang, Y. Gu, H. Chen, Y. Li, and B. Vucetic, "On the age of information of short-packet communications with packet management," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[32] D. Deng, Z. Chen, Y. Jia, L. Liang, S. Fang, and M. Wang, "Age of information in a multiple stream M/G/1/1 non-preemptive queue," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[33] P. Zou, O. Ozel, and S. Subramaniam, "Waiting before serving: A companion to packet management in status update systems," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3864–3877, 2020.

[34] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal, "Reliable transmission of short packets through queues and noisy channels under latency and peak-age violation guarantees," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 721–734, 2019.

[35] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal-Biyikoglu, "Delay and peak-age violation probability in short-packet transmissions," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2471–2475.

[36] J. Cao, X. Zhu, Y. Jiang, and Z. Wei, "Can AoI and delay be minimized simultaneously with short-packet transmission?" in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6.

[37] J. Cao, X. Zhu, Y. Jiang, Z. Wei, S. Sun, and Y. Wang, "Peak AoI vs. delay: Closed-form analysis in the finite block length regime," in *2021*

13th International Conference on Wireless Communications and Signal Processing (WCSP), 2021, pp. 1–6.

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

[39] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[40] D. Bertsekas and R. Gallager, *Data networks (2nd ed.)*. USA: Prentice-Hall, Inc., 1992.