Age of Information with On-Off Service

Ashirwad Sinha, Praful D. Mankar, Nikolaos Pappas, Harpreet S. Dhillon

***Abstract*—This paper considers a communication system where a source sends time-sensitive information to its destination. We assume that both arrival and service processes of the messages are memoryless and the source has a single server with no buffer. Besides, we consider that the service is interrupted by an independent random process, which we model using the On-Off process. For this setup, we study the age of information for two queueing disciplines: 1) *non-preemptive*, where the messages arriving while the server is occupied are discarded, and 2) *preemptive*, where the in-service messages are replaced with newly arriving messages in the Off states. For these disciplines, we derive closed-form expressions for the mean peak age and mean age.**

***Index Terms*—Age of information, Peak age, On-Off process, preemptive discipline, non-preemptive discipline.**

I. Introduction

The sixth generation (6G) of mobile communication system is envisioned to support diverse use cases requiring massive machine-type communication (MMTC) and/or ultra-reliable low-latency communication (URLLC) [1], [2]. Many of these use cases will include remote monitoring and/or actuation where the timeliness of the information may be crucial. For example, the sensors in MMTC may transmit time-sensitive updates, such as obstacle detection in autonomous car driving or fault detection in production chain, to a central processing unit. In such cases, maintaining freshness of updates received at the central unit is critical. For this, a recently introduced metric, called *age of information* (AoI), is useful for measuring the freshness of information received at the destination [3]. Because of its analytical tractability, the mean AoI has emerged as a key performance indicator for the real-time MMTC [1]. Besides, the distribution of AoI is also useful for characterizing the performance of URLLC [4]. However, in several scenarios, the update service process gets interrupted, causing the undesired increase in the age of updates observed at the destination. Such scenarios, where the service toggles between *On and Off states*, include 1) a mobile user going in and out of outage, 2) resources sharing/scheduling, and 3) energy harvesting communication. The AoI under service interruptions caused by an external process has not received much attention yet, which is the main theme of this paper.

Related works: The authors of [3] introduced the AoI metric and derived its average for M/M/1, M/D/1, and D/M/1 queues under the first-come-first-serve (FCFS) discipline. Subsequently, the same authors derived average age for M/M/1 queue in [5] with two types of last-come-first-serve (LCFS) disciplines: 1) LCFS without preemption, where a new arriving update replaces the stale update in queue, and 2) LCFS with preemption, where a new arriving update replaces the in-service update. In a majority of works so far, the consideration of memoryless inter-arrival and service times processes act as the main facilitator. Besides, the average age has also been analyzed for a general arrival/service process. For example, [6] derived average age for M/G/1/1 queue with HARQ, [7] derived the mean age and mean peak age of G/G/1, M/G/1, and G/M/1 queues with FCFS and LCFS in-service preemption disciplines, and [8] derived the mean age and its bounds for G/G/1/1 queue with and without preemption.

Some recent works have started focusing on the distribution of age. The authors of [9] studied the age distribution for D/G/1 queue with FCFS discipline, whereas [10] derived a general formula for the stationary distribution of age that is applicable for a wide class of update systems and demonstrated its use for various queues with FCFS and preemptive/non-preemptive LCFS policies. On similar lines, [11] derived a general formula for the age distribution under ergodic settings, which is then used to derive the generating function of age for discrete systems. The authors of [12] derived the MGF of age for M/M/1 queue with and without preemption. Besides, a new approach based on the idea of stochastic hybrid system is developed for characterizing the distributional properties of age in [13], [14]. Besides, significant work exists on the age characterization for a variety of system settings, such as mutiple source system [15], HARQ based update systems [6], energy harvesting based update system with a single source [12] and with multiple sources [13], [14], FCFS queues in tandem [16], and M/M/1/2 system with random packet deadlines [17], M/G/1 queue with vacation server [18], etc. However, the AoI under externally interrupted service process remains unexplored, despite its relevance in many situations as aforementioned.

Contributions: This paper considers a modified M/M/1/1 system whose server toggles between on and off states according to an independent random process that models the service interruptions. We model these interruptions using an On-Off process such that the On and Off states duration are independent and exponentially distributed. For this setup, we derive the mean age and mean peak age for two disciplines: 1) non-preemptive, where the updates arriving when server is busy are dropped, and 2) preemptive, where the in-service updates are replaced with new arrivals during Off states. For the limiting cases of these On and Off state parameters, the derived expressions for the mean of age and peak age approach to their mean values for the $M/M/1/1$ queue with continuous service.

II. System Model

We consider a communication system where a source sends time sensitive updates about some physical process to its desti-

A. Sinha and P. D. Mankar are with SPCRC, IIIT Hyderabad, India (Email: ashirwad.sinha@students.iiit.ac.in, praful.mankar@iiit.ac.in). N. Pappas is with the Linköping University, Sweden (Email: nikolaos.pappas@liu.se). H. S. Dhillon is with Wireless@VT, Department of ECE, Virginia Tech, Blacksburg, VA (Email: hdhillon@vt.edu). The work of H. S. Dhillon was supported by US NSF under grant CNS-1814477.

nation. It is assumed that the source has a single server with no packet storing facility and both the arrival and service processes follow exponential inter-arrival and service times with rates λ and μ, respectively. Further, we assume that the service process is interrupted by an On-Off process, wherein the server operates normally during the On states and stays idle, which we term as off, during the Off states. The On and Off times are also assumed to be exponentially distributed with parameter $\kappa_{\rm o}$ and $\kappa_{\rm f}$, respectively. We consider two queueing disciplines: 1) *non-preemptive*, where the updates arriving while the server is occupied are discarded, and 2) *preemptive*, where the in-service updates are replaced with newly arriving messages, if any, in the Off states. For such systems, we aim to analyze the AoI which is defined as $\Delta(t) = t - U(t)$, where $U(t)$ is the generation instance of the most fresh update received by the destination. Fig. 1 shows sample paths of the age $\Delta(t)$ for non-preemptive and preemptive disciplines where t_k and t'_k denote the arrival and departure instances of k-th delivered update. The service time of the k-th update is denoted as $T_k = t'_k - t_k$, and the i-th On and Off states' periods are denoted as $T_{{\rm o},i}$ and $T_{{\rm f},i}$, respectively. Let $Y_k = t'_k - t'_{k-1}$ be the time between departures of k-th and $(k-1)$-th updates, and let $B_k = t_k - t'_{k-1}$ denote the time required to arrive the k-th update since $(k-1)$-th delivery. We denote T_k and Y_k as $T^\star_k$ and $Y^\star_k$ for the preemptive case.

We focus on the analysis of the *mean age* and *mean peak age*. The mean age, denoted by Δ, is defined as the time mean of the age process $\Delta(t)$, whereas the mean peak age, denoted by $\bar{\mathcal{A}}$, is defined as the mean of age process $\Delta(t)$ observed just before the delivery of updates. For the non-preemptive discipline, these can be expressed as

$$\bar{\mathcal{A}} = \mathbb{E}[Y_k] + \mathbb{E}[T_{k-1}], \tag{1}$$

$$\text{and } \Delta = 0.5\lambda_e \mathbb{E}[Y_k^2] + \lambda_e \mathbb{E}[Y_k T_{k-1}], \tag{2}$$

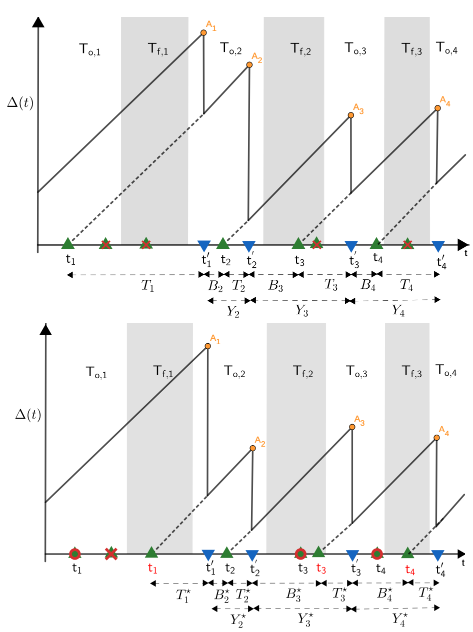

Figure 1. Typical sample path of age $\Delta(t)$ for non-preemption (top) and preemption (bottom) policies. The green up and blue down arrow markers on t-axis indicate update arrivals and departures, respectively. The red cross and circle markers on t-axis show discarded and preempted updates, respectively.

respectively, where λ_e is the effective arrival rate. Please refer to [19, Section III] for more details. For the preemptive discipline, we can determine the mean peak age $\bar{\mathcal{A}}^\star$ and the mean age $\Delta^\star$ using $T^\star_k$ and $Y^\star_k$ in the above expressions.

III. Age Analysis for On-Off service

We first present the age analysis for the non-preemptive case. From (1) and (2), and $Y_k = B_k + T_k$, it is clear that the mean age analysis requires the first two moments of service time T_k. As will be evident shortly, the key step in deriving the moments of T_k is conditioning it on the arrival of the delivered update in On and Off states. The probability of this conditional event is presented in the following lemma.

Lemma 1. *The probability that the delivered update arrives in the* On *state is*

$$\mathrm{P_{On}} = (\lambda + \kappa_{\rm f})(\lambda + \kappa_{\rm o} + \kappa_{\rm f})^{-1}. \tag{3}$$

Proof. Please refer to Appendix A for the proof. □

Lemma 2. *The first and second moments of service time T_k for non-preemptive policy are*

$$\bar{\mathrm{T}}^1_{\rm k} = \frac{1}{\mu} + \frac{\kappa_{\rm o}}{\kappa_{\rm f}}\left[\frac{1}{\mu} + \frac{1}{\lambda + \kappa_{\rm o} + \kappa_{\rm f}}\right], \tag{4}$$

$$\text{and } \bar{\mathrm{T}}^2_{\rm k} = \left(\frac{1}{\mu+\kappa_{\rm o}} + \frac{1}{\kappa_{\rm f}}\right)^2\left(1 + 3\frac{\kappa_{\rm o}}{\mu} + 2\frac{\kappa_{\rm o}^2}{\mu^2}\right) + \frac{1}{(\mu+\kappa_{\rm o})^2} + \frac{\mu+\kappa_{\rm o}}{\mu\kappa_{\rm f}^2}(1 - 2\mathrm{P_{On}}) - \frac{2}{\mu\kappa_{\rm f}}\mathrm{P_{On}}, \tag{5}$$

respectively, where $\mathrm{P_{On}}$ *is given in Lemma 1.*

Proof. Please refer to Appendix B for the proof. □

The effective arrival rate for any given queueing system can be calculated using the its inter-departure times distribution Y_k as $\lambda_{\rm eff} = \mathbb{E}[Y_k]^{-1}$. Finally, using the above lemmas along with (1) and (2), we obtain the mean age in the following theorem.

Theorem 1. *For $M/M/1/1$ queue with* On-Off *service under non-preemptive policy, the mean peak age and mean age are*

$$\bar{\mathcal{A}} = \frac{1}{\lambda} + \frac{2}{\mu} + \frac{2\kappa_{\rm o}}{\kappa_{\rm f}}\left[\frac{1}{\mu} + \frac{1}{\lambda + \kappa_{\rm o} + \kappa_{\rm f}}\right], \tag{6}$$

$$\text{and } \Delta = \frac{1}{1 + \lambda\bar{\mathrm{T}}^1_{\rm k}}\left[\frac{1}{\lambda} + \frac{\lambda}{2}\bar{\mathrm{T}}^2_{\rm k} + \bar{\mathrm{T}}^1_{\rm k}\right] + \bar{\mathrm{T}}^1_{\rm k}, \tag{7}$$

respectively, where $\bar{\mathrm{T}}^1_{\rm k}$ and $\bar{\mathrm{T}}^2_{\rm k}$ are given in Lemma 2.

Proof. The mean peak age given in (6) directly follows by substituting $\mathbb{E}[Y_k] = \mathbb{E}[B_k] + \mathbb{E}[T_k] = \frac{1}{\lambda} + \bar{\mathrm{T}}^1_{\rm k}$ in (1), where $\bar{\mathrm{T}}^1_{\rm k}$ is given in Lemma 2. Next, using $Y_k = B_k + T_k$, (2) and independence of T_k's, the mean age can be written as

$$\Delta = \frac{1}{\mathbb{E}[Y_k]}\left[\mathbb{E}[T_k^2] + \mathbb{E}[B_k^2] + 2\mathbb{E}[T_k]\mathbb{E}[B_k]\right] + \mathbb{E}[T_k].$$

Further, substituting $\mathbb{E}[Y_k] = \frac{1}{\lambda} + \bar{\mathrm{T}}^1_{\rm k}$, and $\bar{\mathrm{T}}^1_{\rm k}$ and $\mathbb{E}[T_k^2] = \bar{\mathrm{T}}^2_{\rm k}$ from Lemma 2, provides (7). □

Recall that in the preemptive policy the in-service update is replaced with a newly arriving update during Off states. For the analysis of this case, the crucial step is to derive the mean of the service time $T^\star_k$. Similar to the analysis of T_k, we derive the mean of $T^\star_k$ by conditioning on the arrival of delivered update in On/Off state. The probability of arrival of a delivered update in On state is smaller, compared to non-preemptive case, due to the replacement of older updates with new ones arrived in the Off states. This probability is derived in Lemma 3.

Lemma 3. *The probability that the successfully delivered update under preemption discipline arrives in the* On *state is*

$$\mathrm{P}^{\star}_{\mathrm{On}} = \mathrm{P}_{\mathrm{On}}(1-\beta)(1-\alpha\beta)^{-1}, \qquad (8)$$

where P_{On} *is given in* (3), $\beta = \frac{\kappa_{\mathrm{o}}}{\mu+\kappa_{\mathrm{o}}}$ *and* $\alpha = \frac{\kappa_{\mathrm{f}}}{\lambda+\kappa_{\mathrm{f}}}$.

Proof. Please refer to Appendix C for the proof. □

Lemma 4. *The mean service time* $T^{\star}_{k}$ *with preemption is*

$$\bar{\mathrm{T}}^{1,\star}_{\mathrm{k}} = \frac{1}{1-\gamma}\left[\frac{1}{\lambda+\kappa_{\mathrm{f}}} + \frac{1}{\mu+\kappa_{\mathrm{o}}}\frac{\lambda+\kappa_{\mathrm{o}}+\kappa_{\mathrm{f}}-\mu}{\lambda+\kappa_{\mathrm{o}}+\kappa_{\mathrm{f}}}\right], \qquad (9)$$

where $\gamma = \kappa_{\mathrm{o}}\kappa_{\mathrm{f}}(\lambda+\kappa_{\mathrm{f}})^{-1}(\mu+\kappa_{\mathrm{o}})^{-1}$.

Proof. Please refer to Appendix D for the proof. □

Finally, using results from Lemma 3 and 4, we obtain the mean age for the preemption case in the following theorem.

Theorem 2. *For* $M/M/1/1$ *queue with* On-Off *service and preemption policy, the mean peak age and mean age are*

$$\bar{\mathcal{A}}^{\star} = \frac{1}{\lambda} + \frac{1}{\mu} + \frac{\kappa_{\mathrm{o}}}{\kappa_{\mathrm{f}}}\left[\frac{1}{\mu} + \frac{1}{\lambda+\kappa_{\mathrm{o}}+\kappa_{\mathrm{f}}}\right] + \bar{\mathrm{T}}^{1,\star}_{\mathrm{k}}, \qquad (10)$$

$$\text{and} \quad \Delta^{\star} = \frac{1}{1+\lambda\bar{\mathrm{T}}^{1}_{\mathrm{k}}}\left[\frac{1}{\lambda} + \frac{\lambda}{2}\bar{\mathrm{T}}^{2}_{\mathrm{k}} + \bar{\mathrm{T}}^{1}_{\mathrm{k}}\right] + \bar{\mathrm{T}}^{1,\star}_{\mathrm{k}}, \qquad (11)$$

respectively, where $\bar{\mathrm{T}}^{1}_{\mathrm{k}}$ *and* $\bar{\mathrm{T}}^{2}_{\mathrm{k}}$ *are given in Lemma 2 and* $\bar{\mathrm{T}}^{1,\star}_{\mathrm{k}}$ *is given in Lemma 4.*

Proof. Due to preemption, the time of arrival of the update (that is getting delivered) since its last delivery is higher as compared to that under non-preemption case, i.e., $B^{\star}_{k} \geq B_{k}$. This increment is equal to the reduction in the service time due to preemption. This is because the preemption under memoryless service process essentially replaces the older packet with the new one without affecting the service/inter-departure time statistics. Hence, we have $Y_k = T_k + B_k = T^{\star}_{k} + B^{\star}_{k} = Y^{\star}_{k}$. This can also be verified from Fig. 1. This also implies that the effective arrival rates λ_e with the preemption and non-preemption disciplines are the same. Thus, using $\lambda_e = \frac{1}{\mathbb{E}[Y_k]}$,

$$\mathbb{E}[Y^{\star}_{k}] = \mathbb{E}[Y_k] = \lambda^{-1} + \bar{\mathrm{T}}^{1}_{\mathrm{k}},$$

$$\mathbb{E}[{Y^{\star}_{k}}^{2}] = \mathbb{E}[Y^{2}_{k}] = \bar{\mathrm{T}}^{2}_{\mathrm{k}} + 2\lambda^{-2} + 2\bar{\mathrm{T}}^{1}_{\mathrm{k}}\lambda^{-1},$$

and following the steps provided in the proof of Theorem 1, we obtain expressions given in (10) and (11). □

Remark 1. *The mean peak age and mean age derived for both non-preemptive (in Theorem 1) and primitive (in Theorem 2) disciplines approach to* $\frac{1}{\lambda} + \frac{2}{\mu}$ *and* $\frac{1}{\lambda} + \frac{2}{\mu} - \frac{1}{\lambda+\mu}$, *respectively, as* $\kappa_{\mathrm{o}} \to 0$ *and/or* $\kappa_{\mathrm{f}} \to \infty$. *These limiting values are equal to the mean peak age and mean age observed under conventional* $M/M/1/1$ *queue [19, Eq. (21) and Eq. (25)]. This is expected since the service tends to appear as uninteruppted as the* On *state duration becomes larger and/or the* Off *state duration becomes smaller, in which case the considered queue discipline will behave similar to the conventional* $M/M/1/1$ *queue.*

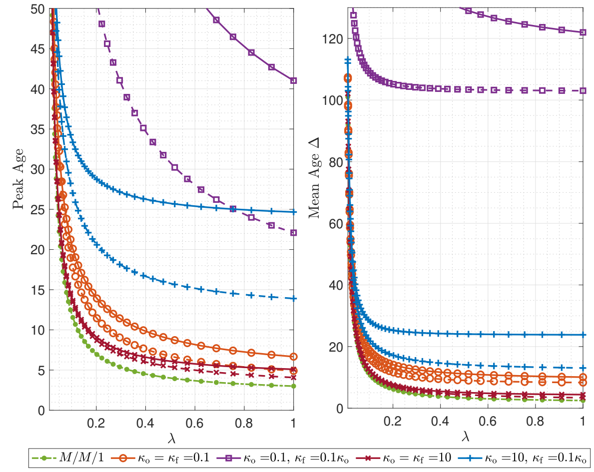

Figure 2. Age vs. λ for $\mu = 1$, $\frac{\mu}{\kappa_{\mathrm{o}}} \in \{1, 10\}$, and $\frac{\kappa_{\mathrm{o}}}{\kappa_{\mathrm{f}}} \in \{1, 10\}$. The solid and dotted lines show the non-preemptive and preemptive policies, respectively.

Fig. 2 shows the age as a function of λ for $\mu = 1$ and different scales of the On-Off state duration. The figure verifies that both peak and mean age are minimum for $M/M/1/1$ system without On-Off service, which is expected. Besides, the figure also illustrates that the preemption policy provides smaller peak as well as mean age as compared to non-preemption policy, for any given configuration of parameters. It can be seen that both the age metrics degrade for $\kappa_{\mathrm{o}} > \kappa_{\mathrm{f}}$. However, interestingly, it can be noted that the mean ages are smaller for $\kappa_{\mathrm{o}} \geq \mu$. This is because of the smaller On-Off cycles resulting in a frequent restart of the memoryless service such that it reduces the overall service time.

IV. Summary

This paper has analyzed the age of information for bufferless systems where both the update arrival and service processes are exponential and the service process is modulated with an independent On-Off process. We have derived the mean peak age and mean age for the discipline where the updates arriving when the server is busy are discarded. Next, we extended the analysis for the preemptive discipline where the source is assumed to be capable of replacing the stale updates with fresh ones arrived during Off states. The derived expressions for both the mean peak age and the mean age reduce to those for the simple $M/M/1$ queue as the On state parameter $\kappa_{\mathrm{o}} \to 0$ and/or the Off state parameter $\kappa_{\mathrm{f}} \to \infty$.

Appendix

A. Proof of Lemma 1

Let s_i and s'_i denote the start times of the i-th On and Off states, respectively. Without loss of generality, we assume that a successful delivery occurs at $t = 0$ and the corresponding On state is the 1-st On state. The probability that next update arrives in an On state is equal to the probability that the next arrival occurs in the interval $\bigcup_{i=1}^{\infty}(s_i, s'_i]$. Note that $s_1 = 0$, $s_i = \sum_{k=1}^{i-1}(T_{\mathrm{o},k} + T_{\mathrm{f},k})$ and $s'_i = s_i + T_{\mathrm{o},i}$, for $i = 2, 3, \ldots$, where $T_{\mathrm{o},k}$ and $T_{\mathrm{f},k}$ are the durations of k-th On and Off states, respectively. Since these intervals are disjoint, the probability that next arrival occurs in On state for given (s_i, s'_i)'s is

$$\mathrm{P}^{\mathrm{cond}}_{\mathrm{On}} = \sum_{i=1}^{\infty} F(s'_i) - F(s_i),$$

$$\stackrel{(a)}{=} \sum_{i=1}^{\infty} \exp(-\lambda s_i) - \exp(-\lambda(s_i + T_{\mathrm{o},i})),$$
$$= \sum_{i=1}^{\infty} [1 - \exp(-\lambda T_{\mathrm{o},i})] \prod_{k=1}^{i-1} \exp(-\lambda(T_{\mathrm{o},k} + T_{\mathrm{f},k})),$$

where $F(\cdot)$ is the CDF of arrival time. Step (a) follows using $F(s) = 1 - \exp(-\lambda s)$. Now, deconditioning $\mathrm{P}_{\mathrm{On}}^{\mathrm{cond}}$ gives

$$\mathrm{P}_{\mathrm{On}} \stackrel{(a)}{=} \sum_{i=1}^{\infty} \mathbb{E}_{T_{\mathrm{o}}}[1 - e^{-\lambda T_{\mathrm{o}}}] \left(\mathbb{E}_{T_{\mathrm{o}}}\left[e^{-\lambda T_{\mathrm{o}}}\right] \mathbb{E}_{T_{\mathrm{f}}}\left[e^{-\lambda T_{\mathrm{f}}}\right]\right)^{i-1},$$
$$= \sum_{l=0}^{\infty} \left[\frac{\kappa_{\mathrm{o}}}{\lambda + \kappa_{\mathrm{o}}} \frac{\kappa_{\mathrm{f}}}{\lambda + \kappa_{\mathrm{f}}}\right]^l \frac{\lambda}{\lambda + \kappa_{\mathrm{o}}},$$

where Step (a) follows directly as both $T_{\mathrm{o},k}$s and $T_{\mathrm{f},k}$s are i.i.d. sequences. Finally, by applying geometric series, we obtain (3).

B. Proof of Lemma 2

Let Z denote the remaining service time of the update and T_{o} denote the duration of an On state. Using the memoryless property of an exponential service, we can find the probability that the update gets served in an On state as

$$\mathrm{P}_{\mathrm{o}} = \mathbb{E}\left[\mathbb{P}[Z \leq t | T_{\mathrm{o}} = t]\right] = \frac{\mu}{\mu + \kappa_{\mathrm{o}}}. \tag{12}$$

Let Z' denote the service time spent in an On state in which the service gets completed. The distribution of Z' becomes

$$f_{Z'}(z) = f_Z(z | Z < T_{\mathrm{o}}) = \frac{\mathbb{P}(T_{\mathrm{o}} > z | Z = z) f_Z(z)}{\int_0^{\infty} \mathbb{P}(T_{\mathrm{o}} > Z | Z = z) f_Z(z) \mathrm{d}z},$$
$$= (\mu + \kappa_{\mathrm{o}}) \exp(-(\mu + \kappa_{\mathrm{o}}) z). \tag{13}$$

Let T_{o}' denote the On state duration in which the service does not get completed. We obtain the distribution of T_{o}' as

$$f_{T_{\mathrm{o}}'}(t) = f_{T_{\mathrm{o}}}(t | Z > T_{\mathrm{o}}) = (\mu + \kappa_{\mathrm{o}}) \exp(-(\mu + \kappa_{\mathrm{o}}) z). \tag{14}$$

Let $G_o = 0$ and $G_n = \sum_{l=1}^{n} T_{\mathrm{o},n}$. Consider $\mathcal{E}_n$ is an event where the service ends in the n-th On state after its arrival. We can determine the probability of $\mathcal{E}_n$ as

$$\mathbb{P}[\mathcal{E}_n] = \mathbb{P}[G_n \geq Z > G_{n-1}],$$
$$\stackrel{(a)}{=} \mathbb{P}[G_n \geq Z] \mathbb{P}[Z > G_l]^{n-1} \stackrel{(b)}{=} \mathrm{P}_{\mathrm{o}} \beta^{n-1}, \tag{15}$$

where Step (a) follows from the memoryless property of Z and the fact that $T_{\mathrm{o},n}$'s are i.i.d.s. Step (b) follows using (12) where $\beta = 1 - \mathrm{P}_{\mathrm{o}} = \frac{\kappa_{\mathrm{o}}}{\mu + \kappa_{\mathrm{o}}}$. Now, by the memoryless property, we can directly consider independently distributed exponential service times spent in different On states for deriving the expected service time by conditioning on the service times, On state durations, and $\mathcal{E}_n$. Given the arrival in On state, the mean of T_k becomes $\bar{\mathrm{T}}_{\mathrm{k}}^{1,\mathrm{o}} =$

$$\sum_{n=1}^{\infty} \left[\mathbb{E}[Z | Z < T_{\mathrm{o},n}] + \sum_{l=1}^{n-1} \mathbb{E}[T_{\mathrm{o},l} | Z > T_{\mathrm{o},l}] + \mathbb{E}[T_{\mathrm{f},l}]\right] \mathbb{P}[\mathcal{E}_n],$$
$$\stackrel{(a)}{=} \mathbb{E}[Z | Z < T_{\mathrm{o}}] + \sum_{n=1}^{\infty} (n-1)[\mathbb{E}[T_{\mathrm{o}} | Z > T_{\mathrm{o}}] + \mathbb{E}[T_{\mathrm{f}}]] \mathrm{P}_{\mathrm{o}} \beta^{n-1},$$
$$\stackrel{(b)}{=} (\mu + \kappa_{\mathrm{o}})^{-1} + \left[(\mu + \kappa_{\mathrm{o}})^{-1} + \kappa_{\mathrm{f}}^{-1}\right] \mathrm{P}_{\mathrm{o}} \mathcal{G}(n-1),$$
$$\stackrel{(c)}{=} \mu^{-1} + \kappa_{\mathrm{o}} (\mu \kappa_{\mathrm{f}})^{-1}, \tag{16}$$

where $\mathcal{G}(f(n)) = \sum_{n=1}^{\infty} f(n) \beta^{n-1}$, Step (a) follows from (15) and the fact that both $T_{\mathrm{o},l}$'s and $T_{\mathrm{f},l}$'s follow i.i.d. distributions. Step (b) follows using (13) and (14). Step (c) follows using $\mathcal{G}(n-1) = \beta(1-\beta)^{-2}$, $\beta = 1 - \mathrm{P}_{\mathrm{o}}$, and some algebric simplifications. Similarly, we obtain the mean of T_k conditioned on update arrived in an Off state as $\bar{\mathrm{T}}_{\mathrm{k}}^{1,\mathrm{f}} =$

$$\sum_{n=1}^{\infty} \left[\mathbb{E}[Z | Z < T_{\mathrm{o},n}] + (n-1) \mathbb{E}[T_{\mathrm{o}} | Z > T_{\mathrm{o}}] + n \mathbb{E}[T_{\mathrm{f}}]\right] \mathbb{P}[\mathcal{E}_n],$$
$$= \frac{1}{\mu + \kappa_o} + \frac{1}{\mu + \kappa_{\mathrm{o}}} \mathrm{P}_{\mathrm{o}} \mathcal{G}(n-1) + \frac{1}{\kappa_{\mathrm{f}}} \mathrm{P}_{\mathrm{o}} \mathcal{G}(n),$$
$$\stackrel{(a)}{=} \frac{1}{\mu + \kappa_o} + \frac{1}{\mu + \kappa_{\mathrm{o}}} \frac{\kappa_{\mathrm{o}}}{\mu} + \frac{1}{\kappa_{\mathrm{f}}} \mathrm{P}_{\mathrm{o}}{}^{-1},$$
$$= \mu^{-1} + (\kappa_{\mathrm{f}} \mu)^{-1} (\mu + \kappa_{\mathrm{o}}),$$

where Step (a) follows from the similar steps used in (16). Thus, the mean of T_k becomes $\bar{\mathrm{T}}_{\mathrm{k}}^{1} = \bar{\mathrm{T}}_{\mathrm{k}}^{1,\mathrm{o}} \mathrm{P}_{\mathrm{On}} + \bar{\mathrm{T}}_{\mathrm{k}}^{1,\mathrm{f}} (1 - \mathrm{P}_{\mathrm{On}})$.

Next, substituting P_{On} from (3) and further solving gives (4). Now, we derive the second moment of T_k using the similar approach presented above as

$$\bar{\mathrm{T}}_{\mathrm{k}}^{2,\mathrm{o}} = \sum_{n=1}^{\infty} \mathbb{E}\left[\left(Z' + \sum_{l=1}^{n-1} [T_{\mathrm{o},l}' + T_{\mathrm{f},l}]\right)^2\right] \mathbb{P}[\mathcal{E}_n].$$

Since Z' and $T_{\mathrm{o},l}'$ are equal in distribution [see (13) and (14)], we can write

$$\bar{\mathrm{T}}_{\mathrm{k}}^{2,\mathrm{o}} = \sum_{n=1}^{\infty} \mathbb{E}\left[\left(\sum_{l=1}^{n} T_{\mathrm{o},l}' + \sum_{l=1}^{n-1} T_{\mathrm{f},l}\right)^2\right] \mathbb{P}[\mathcal{E}_n].$$

Note that, since $T_{\mathrm{o},l}'$ follows exponential distribution with parameter $(\mu + \kappa_{\mathrm{o}})$ independently, we have $S_{\mathrm{o},n} = \sum_{l=1}^{n} T_{\mathrm{o},l}' \sim \mathrm{Gamma}(n, (\mu + \kappa_{\mathrm{o}})^{-1})$. Similarly, $W_{\mathrm{f},n-1} = \sum_{l=1}^{n-1} T_{\mathrm{f},l} \sim \mathrm{Gamma}(n-1, \kappa_{\mathrm{f}}^{-1})$. Using this, we can write $\bar{\mathrm{T}}_{\mathrm{k}}^{\mathrm{o},2}$

$$= \sum_{n=1}^{\infty} \left[\mathbb{E}[S_{\mathrm{o},n}^2] + \mathbb{E}[W_{\mathrm{f},n-1}^2] + 2\mathbb{E}[S_{\mathrm{o},n}] \mathbb{E}[W_{\mathrm{f},n-1}]\right] \mathbb{P}[\mathcal{E}_n],$$
$$= \sum_{n=1}^{\infty} \left[\frac{n(n+1)}{(\mu + \kappa_{\mathrm{o}})^2} + \frac{n(n-1)}{\kappa_{\mathrm{f}}^2} + 2 \frac{n}{\mu + \kappa_{\mathrm{o}}} \frac{n-1}{\kappa_{\mathrm{f}}}\right] \mathbb{P}[\mathcal{E}_n],$$
$$\stackrel{(a)}{=} \frac{\mathrm{P}_{\mathrm{o}} \mathcal{G}(n(n+1))}{(\mu + \kappa_{\mathrm{o}})^2} + \left[\frac{1}{\kappa_{\mathrm{f}}^2} + \frac{2}{\kappa_{\mathrm{f}} (\mu + \kappa_{\mathrm{o}})}\right] \mathrm{P}_{\mathrm{o}} \mathcal{G}(n(n-1)),$$
$$= \mathrm{P}_{\mathrm{o}} \left(\frac{1}{\mu + \kappa_{\mathrm{o}}} + \frac{1}{\kappa_{\mathrm{f}}}\right)^2 \mathcal{G}(n^2)$$
$$+ \mathrm{P}_{\mathrm{o}} \left(\frac{1}{(\mu + \kappa_{\mathrm{o}})^2} - \frac{1}{\kappa_{\mathrm{f}}^2} - \frac{2}{\kappa_{\mathrm{f}} (\mu + \kappa_{\mathrm{o}})}\right) \mathcal{G}(n), \tag{17}$$

where Step (a) follows using (15). With some algebraic calculations, we obtain $\mathcal{G}(n^2) = \mathcal{Z} P_o^{-1}$, where $\mathcal{Z} = 1 + 3\frac{\kappa_{\mathrm{o}}}{\mu} + 2\frac{\kappa_{\mathrm{o}}^2}{\mu^2}$. We also have $\mathcal{G}(n) = (1-\beta)^{-2} = \mathrm{P}_{\mathrm{o}}{}^{-2}$. Thus (17) becomes

$$\bar{\mathrm{T}}_{\mathrm{k}}^{2,\mathrm{o}} = \mathcal{Z} \left(\frac{1}{\mu + \kappa_{\mathrm{o}}} + \frac{1}{\kappa_{\mathrm{f}}}\right)^2 + \frac{\mathrm{P}_{\mathrm{o}}}{\mu^2} - \frac{1}{\mathrm{P}_{\mathrm{o}} \kappa_{\mathrm{f}}^2} - \frac{2}{\mu \kappa_{\mathrm{f}}}. \tag{18}$$

Similarly, we determine the second moment of T_k conditioned on the service started in Off state as

$$\bar{\mathrm{T}}_{\mathrm{k}}^{2,\mathrm{f}} \stackrel{(a)}{=} \sum_{n=1}^{\infty} \mathbb{E}\left[(S_{\mathrm{o},n} + W_{\mathrm{f},n}])^2\right] \mathbb{P}[\mathcal{E}_n],$$
$$= \sum_{n=1}^{\infty} \left[\mathbb{E}[S_{\mathrm{o},n}^2] + \mathbb{E}[W_{\mathrm{f},n}] + 2\mathbb{E}[S_{\mathrm{o},n}] \mathbb{E}[W_{\mathrm{f},n}]\right] \mathbb{P}[\mathcal{E}_n],$$
$$= \left[\frac{1}{(\mu + \kappa_{\mathrm{o}})^2} + \frac{1}{\kappa_{\mathrm{f}}^2}\right] \mathrm{P}_{\mathrm{o}} \mathcal{G}(n(n+1)) + \frac{2\mathrm{P}_{\mathrm{o}} \mathcal{G}(n^2)}{\kappa_{\mathrm{f}} (\mu + \kappa_{\mathrm{o}})},$$
$$= \mathcal{Z} \left(\frac{1}{\mu + \kappa_{\mathrm{o}}} + \frac{1}{\kappa_{\mathrm{f}}}\right)^2 + \frac{\mathrm{P}_{\mathrm{o}}}{\mu^2} + \frac{1}{\mathrm{P}_{\mathrm{o}} \kappa_{\mathrm{f}}^2}, \tag{19}$$

where Step (a) follows from $S_{\mathrm{o},n} = Z' + S_{\mathrm{o},n-1}$. Now, we can obtain the second moment of T_k as $\bar{\mathrm{T}}_{\mathrm{k}}^2 = \bar{\mathrm{T}}_{\mathrm{k}}^{2,\mathrm{o}} \mathrm{P}_{\mathrm{On}} + \bar{\mathrm{T}}_{\mathrm{k}}^{2,\mathrm{f}} (1 -$

$\mathrm{P_{On}}$). Finally, substituting (18) and (19) gives (5).

C. Proof of Lemma 3

The update arrived in an On state will be delivered if it does not get preempted during the Off states occurring in its service time. Thus, the probability that the update arrived in On state will be delivered is

$$\mathrm{P}_{\mathrm{On}}^{\star} = \mathrm{P}_{\mathrm{On}}\mathrm{P}_{\mathrm{pre}}^{\mathrm{c}}, \tag{20}$$

where $\mathrm{P_{On}}$ is given in Lemma 1 and $\mathrm{P}_{\mathrm{pre}}^{\mathrm{c}}$ is the probability that the update will not get preempted. By conditioning on the event $\mathcal{E}_n$ (defined in Appendix B), we obtain

$$\mathrm{P}_{\mathrm{pre}}^{\mathrm{c}} = \sum\nolimits_{n=1}^{\infty}\left[\prod\nolimits_{l=1}^{n-1}\mathbb{P}[\text{No arrival in } T_{\mathrm{f},l}]\right]\mathbb{P}[\mathcal{E}_n],$$
$$\stackrel{(a)}{=} \mathrm{P_o}\sum\nolimits_{n=1}^{\infty}\alpha^{n-1}\beta^{n-1} = (1-\beta)\frac{1}{1-\alpha\beta},$$

where Step (a) follows using (15), $\mathbb{P}[\text{No arrival in } T_{\mathrm{f},l}] = \alpha = \frac{\kappa_{\mathrm{f}}}{\lambda+\kappa_{\mathrm{f}}}$ and independence of $T_{\mathrm{f},l}$'s. Finally, substituting the above $\mathrm{P}_{\mathrm{pre}}^{\mathrm{c}}$ and $\mathrm{P_{On}}$ from Lemma 1 in (20) completes the proof.

D. Proof of Lemma 4

We construct the proof on the similar lines as in Appendix B. Recall, in Appendix B, the probability $\mathbb{P}[\mathcal{E}_n]$ that the service of an update ends in the n-th On state after its arrival is derived for the case of no preemption. However, under preemption, this probability depends on the probability that service takes n On states and there is no preemption during the Off periods occurring during the service time. Let $\mathcal{E}_n^{\star}$ denote the event where service of an update ends in the n-th On state after its arrival under preemption. Note that $\mathcal{E}_n^{\star}$ includes n i.i.d. Off periods in the service if it starts from an Off state, otherwise it includes $n-1$ i.i.d. Off periods. Thus,

$$\mathbb{P}[\mathcal{E}_n^{\star}] = \begin{cases} \frac{\mathbb{P}[\mathcal{E}_n]\mathrm{P}_{\mathrm{n-1}}^{\mathrm{no-pre}}}{\sum_{n=1}^{\infty}\mathbb{P}[\mathcal{E}_n]\mathrm{P}_{\mathrm{n-1}}^{\mathrm{no-pre}}}, & \text{if ser. starts in On st.,}\\ \frac{\mathbb{P}[\mathcal{E}_n]\mathrm{P}_{\mathrm{n}}^{\mathrm{no-pre}}}{\sum_{n=1}^{\infty}\mathbb{P}[\mathcal{E}_n]\mathrm{P}_{\mathrm{n}}^{\mathrm{no-pre}}}, & \text{otherwise,}\end{cases} \tag{21}$$

where $\mathrm{P}_{\mathrm{n}}^{\mathrm{no-pre}}$ is the probability that there is no preemption in the n Off states occurring during the service time. We can directly evaluate $\mathrm{P}_{\mathrm{n}}^{\mathrm{no-pre}}$ as $\mathrm{P}_{\mathrm{n}}^{\mathrm{no-pre}} = \alpha^n$, where α is the probability that there is no preemption in a typical Off state which is obtained in Appendix C as $\alpha = \frac{\kappa_{\mathrm{f}}}{\lambda+\kappa_{\mathrm{f}}}$. Thus, substituting $\mathbb{P}[\mathcal{E}_n] = \mathrm{P_o}\beta^{\mathrm{n}-1}$ and $\mathrm{P_{no-pre}} = \alpha^n$ in (21) and simplifying, we get

$$\mathbb{P}[\mathcal{E}_n^{\star}] = (1-\gamma)\gamma^{n-1}, \tag{22}$$

where $\gamma = \alpha\beta$, irrespective of whether the services started in the On or Off state. Note that the distribution of Off state duration conditioned on no preemption is

$$f_{T_{\mathrm{f}}'}(t) = (\lambda+\kappa_{\mathrm{f}})\exp(-(\lambda+\kappa_{\mathrm{f}})t).$$

Now, similar to (16), we obtain the mean of $T_k^{\star}$ given the update arrived in the On state as

$$\bar{\mathrm{T}}_{\mathrm{k}}^{1\star,\mathrm{o}} = \sum\nolimits_{n=1}^{\infty}\left[\mathbb{E}[Z'] + (n-1)\mathbb{E}[T_{\mathrm{o}}'] + (n-1)\mathbb{E}[T_{\mathrm{f}}']\right]\mathbb{P}[\mathcal{E}_n^{\star}],$$
$$\stackrel{(a)}{=} \mathbb{E}[Z'] + [\mathbb{E}[T_{\mathrm{o}}'] + \mathbb{E}[T_{\mathrm{f}}']](1-\gamma)\gamma\sum\nolimits_{n=1}^{\infty}(n-1)\gamma^{n-2},$$
$$= \frac{1}{\mu+\kappa_{\mathrm{o}}} + \left[\frac{1}{\mu+\kappa_{\mathrm{o}}} + \frac{1}{\lambda+\kappa_{\mathrm{f}}}\right](1-\gamma)\gamma\frac{\mathrm{d}}{\mathrm{d}\gamma}\sum\nolimits_{l=0}^{\infty}\gamma^l,$$
$$= \left[(\mu+\kappa_{\mathrm{o}})^{-1} + \gamma(\lambda+\kappa_{\mathrm{f}})^{-1}\right](1-\gamma)^{-1}, \tag{23}$$

where Step (a) follows using (22). Similarly, we obtain the mean of $T_k^{\star}$ given the update arrived in the Off state as

$$\bar{\mathrm{T}}_{\mathrm{k}}^{1\star,\mathrm{f}} = \sum\nolimits_{n=1}^{\infty}\left[\mathbb{E}[Z'] + (n-1)\mathbb{E}[T_{\mathrm{o}}'] + n\mathbb{E}[T_{\mathrm{f}}']\right]\mathbb{P}[\mathcal{E}_n^{\star}],$$
$$= \mathbb{E}[Z'] + \mathbb{E}[T_{\mathrm{o}}'](1-\gamma)\gamma\sum_{n=1}^{\infty}(n-1)\gamma^{n-2} + \mathbb{E}[T_{\mathrm{f}}']\sum_{n=1}^{\infty}n\gamma^{n-1},$$
$$= \left[\frac{1}{\mu+\kappa_{\mathrm{o}}} + \frac{1}{\lambda+\kappa_{\mathrm{f}}}\right]\frac{1}{1-\gamma}. \tag{24}$$

Next, we obtain mean $\bar{\mathrm{T}}_k^{\star}$ as

$$\bar{\mathrm{T}}_{\mathrm{k}}^{1\star} = \bar{\mathrm{T}}_{\mathrm{k}}^{1\star,\mathrm{o}}\mathrm{P}_{\mathrm{On}}^{\star} + \bar{\mathrm{T}}_{\mathrm{k}}^{1\star,\mathrm{f}}(1-\mathrm{P}_{\mathrm{On}}^{\star}).$$

Finally, by substituting $\bar{\mathrm{T}}_{\mathrm{k}}^{1*,\mathrm{o}}$ from (23), $\bar{\mathrm{T}}_{\mathrm{k}}^{1*,\mathrm{f}}$ from (24), and $\mathrm{P}_{\mathrm{On}}^{\star}$ from Lemma 3 in the above expression and further simplifying, we obtain (9). This completes the proof.

References

[1] W. Jiang, B. Han, M. Habibi, and H. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[2] P. Popovski, F. Chiariotti, K. Huang, A. E. Kalør, M. Kountouris, N. Pappas, and B. Soret, "A perspective on time toward wireless 6G," *Proceedings of the IEEE*, vol. 110, no. 8, pp. 1116–1146, 2022.

[3] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, 2012.

[4] M. K. Abdel-Aziz, C.-F. Liu, S. Samarakoon, M. Bennis, and W. Saad, "Ultra-reliable low-latency vehicular networks: Taming the age of information tail," in *IEEE GLOBECOM*, 2018, pp. 1–7.

[5] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Conf. Info. Sci. Sys.*, 2012, pp. 1–6.

[6] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *IEEE ISIT*, 2017, pp. 131–135.

[7] R. Talak, S. Karaman, and E. Modiano, "Can determinacy minimize age of information?" [Online]. Available: https://arxiv.org/abs/1810.04371

[8] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization," *IEEE Trans. Info. Theory*, vol. 67, no. 11, pp. 7477–7489, 2021.

[9] J. P. Champati, H. Al-Zubaidy, and J. Gross, "Statistical guarantee optimization for age of information for the D/G/1 queue," in *IEEE INFOCOM*, 2018, pp. 130–135.

[10] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," *IEEE Trans. Info. Theory*, vol. 65, no. 12, pp. 8305–8324, 2019.

[11] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The age of information in a discrete time queue: Stationary distribution and non-linear age mean analysis," *IEEE J. Sel. Areas Comm.*, vol. 39, no. 5, pp. 1352–1364, 2021.

[12] M. A. Abd-Elmagid and H. S. Dhillon, "Closed-form characterization of the MGF of AoI in energy harvesting status update systems," *IEEE Trans. Info. Theory*, vol. 68, no. 6, pp. 3896–3919, 2022.

[13] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," *IEEE Trans. Info. Theory*, vol. 66, no. 9, pp. 5712–5728, 2020.

[14] M. A. Abd-Elmagid and H. S. Dhillon, "Joint distribution of ages of information in networks," 2022. [Online]. Available: https://arxiv.org/abs/2205.07448

[15] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Info. Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.

[16] C. Kam and S. Kompella, "On the age of information for non-preemptive queues in tandem," *Frontiers in Communications and Networks*, p. 24.

[17] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "On the age of information with packet deadlines," *IEEE Trans. Info. Theory*, vol. 64, no. 9, pp. 6419–6428, 2018.

[18] J. Xu, I.-H. Hou, and N. Gautam, "Age of information for single buffer systems with vacation server," *IEEE Trans. Net. Sci. Engg.*, pp. 1–1, 2021.

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