Improved Scaling Law for Status Update Timeliness in Massive IoT by Elastic Spatial Multiplexing

Zhiyuan Jiang, Sheng Zhou, Zhisheng Niu, *Fellow, IEEE*
{zhiyuan, sheng.zhou, niuzhs}@tsinghua.edu.cn,
Beijing National Research Center for Information Science and Technology (BNRist),
Tsinghua University, Beijing 100084, China

***Abstract*—In this paper, the wireless uplink is considered for status update with a large number of terminals. Thy key problem we address is that whether spatial multiplexing of multiple terminals, enabled by the massive multiple-input multiple-output technology, can help to improve the scaling law of age-of-information versus the number of terminals, on account of the mandatory pilot overhead. Based on a queuing theory analysis, we show that the proposed elastic spatial multiplexing scheme, which assigns an optimized pilot length that is smaller than the number of transmitting terminals on account of random packet arrivals, can indeed improve the scaling law compared with the optimal scaling law without spatial multiplexing, by a factor that is related to the packet lengths and arrival rates. Simulation results are provided to validate our findings.**

***Index Terms*—Internet-of-Things, age-of-information, massive multiple-input multiple-output, queuing theory**

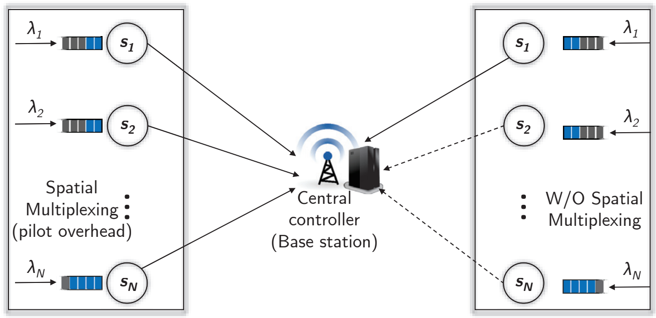

Fig. 1. System architecture. The left part depicts transmissions with spatial multiplexing of all terminals; the right part is concerned with transmissions without spatial multiplexing.

I. Introduction

Internet-of-Things (IoT) is emerging to be one of the revolutionary technologies of modern societies. In the foreseeable future, machine-type communication (MTC) will replace human-based applications (videos, texts) and dominant the data traffic in wireless communication systems. In many scenarios, the MTC involves terminals reporting status update of sensory data, e.g., temperature, pressure and etc., to destination nodes. One of the major challenges in designing such systems is to achieve *timely status update* from terminals and, meanwhile, still *scalable* in the massive IoT regime with a large number of terminals.

In order to quantify the status update timeliness, age-of-information (AoI) is recently proposed which denotes the time elapsed since the last status observation [1], [2]. The AoI jointly accounts for the information rate of status variation, which is related to sampling rate at the information source, and data communication delay, and hence distinguishes itself from the conventional end-to-end (e2e) communication delay metric, making it suitable to characterize the status update timeliness in IoT systems. The study of AoI optimization receives broad attention in the literature, including e.g., single-queue analysis [1], multiple sources sharing one queue [3], general service distributions [4] and scheduling problems with multiple queues [5].

The considered scenario in this paper is shown in Fig. 1, where a central node is receiving status update from terminals. Our previous work [6] has shown that a round-robin policy can achieve the optimal scaling law in the massive IoT regime, assuming that only one terminal can be scheduled at one time; several other works also consider this problem and propose near-optimal solutions based on, e.g., Whittle's index policies [7]–[9]. However, as far as we know, no existing work has considered spatial multiplexing enabled by massive multiple-input multiple-output (MIMO) technology [10] which will play a pivotal part in future systems. Based on spatial multiplexing, multiple terminals can transmit simultaneously leveraging spatially orthogonal channels such that the system scalability is significantly improved. However, on the flip side, spatial multiplexing entails a mandatory pilot overhead which is proportional to the number of terminals assuming time-division-duplexing (TDD) operations. The question of particular interest is that whether spatial multiplexing can benefit the AoI performance, considering the pilot overhead.

In this paper, we show through theoretical analysis and appropriate approximations that compared with the optimal scaling result without spatial multiplexing, i.e., $\frac{DN}{2}$ where D is the number of time slots consumed to transmit one packet [6], the proposed elastic spatial multiplexing scheme (ESM) can achieve an improved scaling law of $\frac{\delta N}{2\beta}$ where $\delta < 1$ is related to status packet arrival rates and $\beta \geq 1$ is related to the physical-layer multi-terminal detection performance (both defined formally in later sections). Considering the pilot overhead, ESM allows all terminals to transmit pilot symbols while only some of them may have status packets due to random packet arrivals, and hence ESM can benefit from spatial multiplexing by carefully designing the pilot length.

Notations: We use $\mathcal{O}(M^N)$ to denote polynomials of M with order lower than N. The expectation operator is denoted by $\mathbb{E}[\cdot]$. The k-th binomial coefficient is denoted by $\binom{N}{K}$. The modulo operator is denoted by mod. The complementary error function is defined by $\mathrm{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} \mathrm{d}t$.

II. System Model and Preliminaries

We consider a single-cell cellular system wherein a base station (BS), or central controller, equipped with M antennas is responsible for collecting status updates from N terminals, e.g., sensors and monitors. A time slotted system is considered. The status updates are conveyed by randomly generated packets at each terminal, reflecting the current status information sensed by terminals and stored at terminal queues. The packet arrivals are modeled by independently identically distributed (i.i.d.) Bernoulli processes with mean rates $\lambda_n \in [0, 1]$, $\forall n = \{1, ..., N\}$. The status update packets are transmitted in the wireless uplink, where each packet is assumed to consume a constant of D ($D \geq 1$) time slots to be transmitted; this corresponds to the scenario that the status packets are with constant lengths and retransmissions due to channel failure are not considered. We adopt the AoI metric as our main optimization target in this paper. Concretely, the τ-horizon time-average AoI of the system is defined by

$$\Delta_{\pi}^{(\tau,N)} \triangleq \frac{1}{\tau N} \sum_{t=1}^{\tau} \sum_{n=1}^{N} \mathbb{E}[h_{n,\pi}(t)], \tag{1}$$

where π denotes an admissible policy, τ is the time horizon length, $h_{n,\pi}(t)$ denotes the AoI reported by terminal-n at the t-th time slot under policy π and

$$h_{n,\pi}(t) \triangleq t - \mu_{n,\pi}(t), \tag{2}$$

where $\mu_{n,\pi}(t)$ denotes the generation time of the status from terminal-n maintained at destination at time t. In particular, the long-time-average AoI of the system is concerned, which is defined by

$$\bar{\Delta}_{\pi,N} \triangleq \limsup_{\tau \to \infty} \Delta_{\pi}^{(\tau,N)}. \tag{3}$$

A. Optimal Scaling Law in Massive IoT without Spatial Multiplexing

Without spatial multiplexing, the uplink channel allows one terminal to transmit at one time, i.e., the protocol interference model is adopted which means that a collision happens whenever multiple terminals transmit simultaneously. The following corollary based on our previous work in [6] states the optimal scaling result with a massive number of IoT terminals and its attainability by the round-robin scheduling policy with one-packet buffers at terminals (only retain the most up-to-date packets), referred to as RR-ONE and denoted by RR in the subscript.

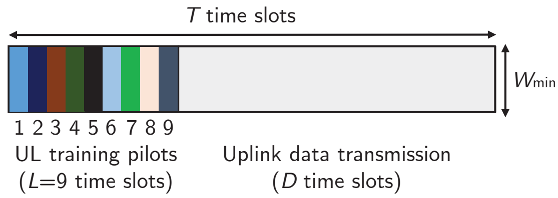

Fig. 2. Pilot-aided spatial multiplexing in massive MIMO enabled uplinks.

Corollary 1: Assuming the protocol interference model and that each transmission consumes D time slots, RR-ONE achieves a long-time average AoI of

$$\bar{\Delta}_{\mathsf{RR},N} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\lambda_n} + \frac{ND-1}{2} + D - 1. \tag{4}$$

The optimal scaling law between the long-time average AoI and the number of terminals in this scenario is

$$\bar{\Delta}_{\mathsf{opt},N} = \frac{DN}{2} + \mathcal{O}(N). \qquad \square \tag{5}$$

Proof: The performance of RR-ONE can be derived by a generalized Poisson-arrival-see-time-average (PASTA) theorem, and the optimality of the scaling factor can be proved by comparing RR-ONE to a AoI lower bound as in [6, Lemma 3], with a straightforward extension to incorporate the transmission time D. ■

Given the optimal scaling result without spatial multiplexing in Corollary 1, the main problem to be addressed is: **Can spatial multiplexing achieve an improved scaling law in the massive IoT regime?** In the following subsection, we will first describe the system model of spatial multiplexing, and then present our main result in the subsequent sections.

B. Uplink Spatial Multiplexing

Thanks to the massive MIMO technology, uplink spatial multiplexing of multiple terminals can be enabled, i.e, multiple terminals can transmit simultaneously and successfully (collision-free) by performing joint spatial multi-terminal detection at the BS side [10]. Since different terminals have distinct channel state information (CSI) vectors which become asymptotically orthogonal when $M \to \infty$, their packets can be efficiently decoded by estimating and exploiting the CSI vectors as spatial signatures. Such a process can be materialized by a pilot-aided uplink transmission scheme as depicted in Figure 2.

A transmission frame consists of T time slots,[1] among which L time slots are dedicated to uplink pilots transmission for channel training and the remaining D time slots are used for data transmission; in this paper, we assume that by spatial multiplexing, multiple terminals transmit in parallel and the

[1]In this paper, we assume that the channel coherence time is larger than T and that the channel stays constant throughout one transmission frame.

resulting transmission time, i.e., D time slots, is the same as the one occupied by a single terminal.[2]

However, there is a limit in terms of the number of spatial-multiplexed terminals given the pilot length L; exceeding the limit will result in unsuccessful channel estimations and hence failure in the signal decoding. The characterization of the limit is related to several key parameters and has been extensively investigated in the massive MIMO literature. In summary, two aspects play pivotal roles in determining this limitation:

- In a pure pilot-based scheme, the number of spatial-multiplexed terminals, denoted by K, cannot be more than L due to the fact that the CSI estimation process has to solve for K variables given L linear equations [11]. Furthermore, it often requires $K < L$ considering the uplink channel noise and requirement of CSI estimation accuracy for demodulation.
- In a data-aided pilot-based scheme (cf. [12], [13]), the constraint of $K < L$ can be relaxed by leveraging received data to facilitate the CSI estimations. It is shown that K can be larger than L, with low transmission failure probability.

On account of these existing works, and without loss of generality, we define the transmission success probability given K and L as $p_{\rm s}(K, L)$, that is, in the event that a transmission is successful, all spatial-multiplexed terminals update a packet; otherwise all updates fail since the signals from terminals cannot be distinguished. Whereas the specific physical-layer uplink CSI estimation and data transmission scheme is of critical importance, this paper focuses on the AoI performance analysis and hence the essence and net effect of uplink transmissions are abstracted by $p_{\rm s}(k, l)$. An arbitrary function of $p_{\rm s}(K, L)$ can be plugged into our result; an exemplary $p_{\rm s}(K, L)$ can be found in the recent work considering massive connectivity in massive MIMO [13].

C. Status Update and AoI Evolution

It is crucial to clarify the status update procedure, which includes deciding the following:

- *Scheduled terminal set*: Decide which set of terminals is scheduled in the transmission frame.
- *Packet management*: Once scheduled, each terminal transmits its associated pilot sequence and a status update packet (if its queue is not empty) based on a packet management scheme; the terminal stays silent if its queue is empty.

Due to the fact that a newer status always renders an older status meaningless, it is obvious that the optimal packet management scheme is to transmit the most up-to-date packet; this is equivalent to maintaining a one-packet buffer at each terminal and only retains the newest packet. Secondly, the number of scheduled terminals, denoted by S, is generally larger than the number of transmitting terminals K, on account of the fact that some terminals may have empty queues due to random packet arrivals and these terminals would remain silent even if scheduled. Consequently, to fully utilize the channel and achieve the best performance, one should schedule more terminals than the number of terminals the adopted pilot length (L) allows considering random packet arrivals; this is referred to as *elastic spatial multiplexing*. In particular, in this work, we let all terminals be scheduled in each transmission frame and design the optimal pilot length thereof.

The evolution of the AoI of terminal-n can be written as

$$h_{n,\pi}(t+1) = \begin{cases} h_{n,\pi}(t) + 1, & \text{if } t \bmod T \neq 0; \\ h_{n,\pi}(t) + 1 - u_{n,\pi}(t) p_{{\rm s},\pi}(t) g_{n,\pi}(t), & \text{otherwise}, \end{cases} \tag{6}$$

where $u_{n,\pi}(t) = 1$ denotes the terminal-n is scheduled in this transmission frame and zero otherwise, the transmission success probability is denoted by $p_{{\rm s},\pi}(t)$, and at the end of the frame the AoI reduction is denoted by $g_{n,\pi}(t)$ which equals the time duration (time slots) between the generation of the last received packet from terminal-n and the updated packet's generation time. Note that $g_{n,\pi}(t)$ equals zero if terminal-n has no packet to update.

III. Achievable Long-Time Average AoI by Elastic Spatial Multiplexing

In order to show that ESM can achieve better scaling result, we only need to find one status update strategy that achieves a scaling factor smaller than $\frac{D}{2}$ as in Corollary 1.[3] Towards this end, we consider the case that all N terminals are scheduled in each transmission frame and try to find the appropriate pilot length L; this scheme is denote by ESM.

In what follows, the long-time average AoI of ESM will be derived.

Theorem 1: The long-time average AoI of ESM is

$$\bar{\Delta}_{\mathsf{ESM},N} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\lambda_n} + \frac{1}{N}\sum_{n=1}^{N}\frac{2 - c_{\mathsf{s},n}(L,S)}{2c_{\mathsf{s},n}(L,S)}T + D - \frac{3}{2}, \tag{7}$$

where

$$\begin{aligned} c_{\mathsf{s},n}(L,S) &\triangleq \int_{-\infty}^{+\infty} q_n(x,S,m) p_s(x,L) \mathsf{d}x, \\ q_n(K,N,T) &\triangleq \binom{N}{K}(1-\delta_n)^{N-K}\delta_n^K, \\ \delta_n &\triangleq 1 - (1-\lambda_n)^T. \qquad \square \end{aligned} \tag{8}$$

Proof: See Appendix A. ■

[2]Existing work [10] has shown that such an assumption is approximately satisfied in the massive MIMO uplink as long as the number of BS antennas is larger than N.

[3]Clearly, this means that the proposed strategy in this paper is not necessarily achieving the optimal scaling law which is left for future work.

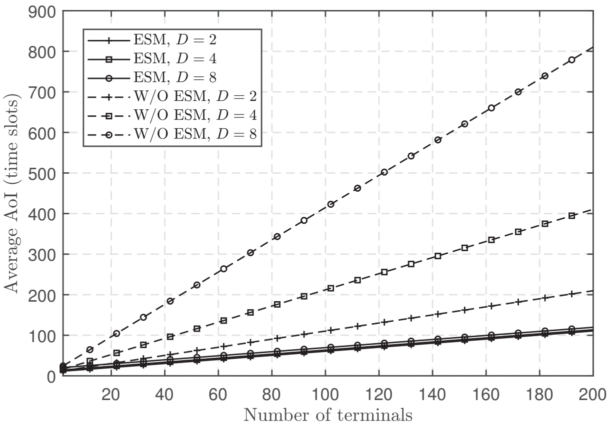

Fig. 3. Long-time average AoI for RR-ONE [6] and ESM with $\lambda_n = 1/10$, $\forall n$.

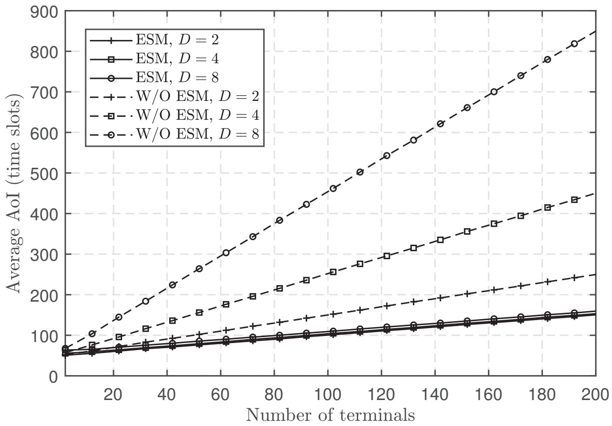

Fig. 4. Long-time average AoI for RR-ONE [6] and ESM with $\lambda_n = 1/50$, $\forall n$.

IV. Closed-Form Scaling Result by Elastic Spatial Multiplexing

Based on Theorem 1, the long-time average AoI can be analyzed numerically given an arbitrary physical layer transmission success probability function of $p_s(K, L)$. However, this formula cannot provide a closed-form scaling result that we desire. Hence, we will use an exemplary $p_s(K, L)$ to derive an explicit scaling result. Specifically, we assume that the physical layer transmission success probability is of the following threshold-type form:

$$p_{\mathrm{s}}(K, L) = \begin{cases} 1, & \text{if } K \leq \beta L; \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

where $\beta \in (0, \infty)$ is a constant irrespective with K and L. That is, the success probability of physical layer multi-terminal detection is one when the number of concurrent terminals is smaller than β times the pilot length, and zero otherwise. This threshold-based model is in fact validated by many existing works (cf. [13]). Based on this model we obtain the following theorem.

Theorem 2: The following scaling result is attainable with ESM:

$$\bar{\Delta}_{\mathsf{ESM},N} = \frac{\delta_{n_{\mathsf{max}}}}{2\beta} N + o(N), \tag{10}$$

where n_{max} is given in (25). □

Proof: See Appendix B. ■

Remark 1: Comparing the scaling results of Theorem 2 and Corollary 1, it is found that the scaling factor is $\frac{\delta_{n_{\mathsf{max}}}}{2\beta}$ by ESM and $\frac{D}{2}$ without spatial multiplexing. Due to the fact that $\delta < 1$, $\beta \geq 1$ and $D \geq 1$, the scaling factor of ESM is strictly smaller than that without spatial multiplexing. □

V. Simulation Results

In this section, we run Monte-Carlo simulations to validate the derived AoI performance and scaling results. The arrival packets are generated based on Bernoulli distributions with

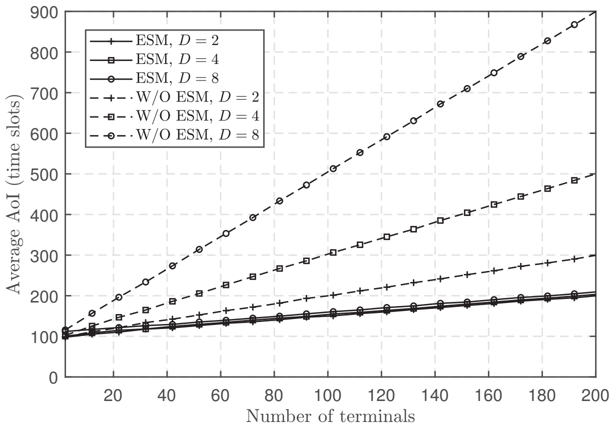

Fig. 5. Long-time average AoI for RR-ONE [6] and ESM with $\lambda_n = 1/100$, $\forall n$.

arrival rates given in the figures. The packet lengths for status update are 2, 4 and 8 time slots in the figures as shown by the legends. The pilot length in ESM is given by (25) and the physical-layer channel estimation is abstracted by (9) with $\beta = 1$. The simulation runs for 10^6 time slots to obtain the long-time average AoI. In comparison, we show the performance of RR-ONE in [6] which is proved to achieve the optimal scaling law without spatial multiplexing. It is shown by Fig. 3, 4 and 5 that the derived scaling law in Theorem 2 is well observed; moreover, the improved scaling factor, i.e., the slope of the curve, demonstrates that ESM benefit from spatial multiplexing such that the AoI is irrelevant to the packet lengths, i.e., D.

VI. Conclusions and Future Work

In this paper, it is found that spatial multiplexing or multi-terminal detection on the wireless uplink, on account of the mandatory pilot overhead and channel estimation error, can improve the scalability of status update with a large number of terminals in terms of AoI. In particular, the scaling factor of AoI with the number of terminals can be improved by a multiplicative factor proportional to the packet transmission time. While the paper shows an improved scaling result, the optimal scaling result with spatial multiplexing is still unclear, which is left for future work.

Appendix A
Proof of Theorem 1

First, the transmission success probability of a transmission frame is calculated. We adopt $p_{\mathsf{s}}(K, L)$ as an abstraction of physical-layer transmission success probability by spatial multiplexing, with a pilot length of L and the number of concurrent terminals of K. Given that the scheduling interval for each terminal is $T = D + L$, the distribution of the age of the packet at the terminal side when scheduled, i.e., a_n for terminal-n, is (in the case of empty queues, a_n equals the current AoI of terminal-n, which is denoted by h_n, for ease of exposition; note that $a_n < h_n$ if the queue is not empty and thus there is no ambiguity below)

$$\begin{aligned}\Pr\{a_n = a|\tau_n = T\} &= \lambda_n(1-\lambda_n)^a,\ a = 0, ..., T-1,\\ \Pr\{a_n = h_n|\tau_n = T\} &= (1-\lambda_n)^T,\end{aligned} \tag{11}$$

where τ_n denotes the time since last scheduling of terminal-n. Also note that we assume after a terminal transmits in a transmission frame, the transmitted packet leaves its queue even if the transmission fails since no feedback information from the BS is assumed. Based on ESM, all N terminals are scheduled in each transmission frame, the probability that there are K terminals that have non-empty queues is therefore following the binomial distribution, i.e.,

$$q_n(K, N, T) \triangleq \binom{N}{K}(1-\delta_n)^{N-K}\delta_n^K, \tag{12}$$

where $\delta_n \triangleq 1 - (1-\lambda_n)^T$. The transmission success probability in the s-th transmission frame is therefore

$$c_{\mathsf{s},n}(L, S) \triangleq \int_{-\infty}^{+\infty} q_n(x, S, m)p_s(x, L)\mathrm{d}x. \tag{13}$$

Armed with the above results, we are ready to calculate the long-time-average AoI. The following derivation is based on a generalization of the PASTA theorem, i.e., the ASTA property [14, Theorem 3.14]. Basically, each scheduling is seen as a calibration for the AoI of the outside observer (in this case is the BS) based on the observed age of the packet at the terminal side; due to possible transmission failures deriving from ROSE by elastic spatial multiplexing, the calibrations are also subject to failures. Figure 6 illustrates the evolution processes of AoI of terminal-n and the age of the packet at terminal-n and their interplay.

Observing Figure 6, although the scheduling interval of every terminal is fixed to m, the transmission is subject to failure which is denoted by the dotted line with $c_{\mathsf{s}} = 0$ in (22). Therefore, denoting the successful transmission interval of terminal-n as $X_{n,\pi}(l)$ where l denotes the interval index, it follows a geometric distribution, i.e.,

$$\begin{aligned}\Pr\{X_{n,\pi}(l) = T\gamma\} = (1 - c_{\mathsf{s},n}(L, S))^{\gamma-1}c_{\mathsf{s},n}(L, S),\\ \forall n \in \{1, ..., N\},\ l = 1, 2, ...\end{aligned} \tag{14}$$

Note that $c_{\mathsf{s},n}(L, S)$ is irrespective with l such that $X_{n,\pi}(l)$ is i.i.d. over time. The first and second moments of $X_{n,\pi}(l)$ then follow:

$$\begin{aligned}\mathbb{E}\left[X_{n,\pi}(l)\right] &= \mathbb{E}[\gamma]T = \frac{T}{c_{\mathsf{s},n}(L, S)},\\ \mathbb{E}\left[X^2_{n,\pi}(l)\right] &= \mathbb{E}[\gamma^2]T^2 = \frac{2 - c_{\mathsf{s},n}(L, S)}{c_{\mathsf{s},n}(L, S)^2}T^2.\end{aligned} \tag{15}$$

To proceed to apply the ASTA property and calculate the AoI, we first need to verify two crucial conditions. Denote $R_{n,\pi}(t)$ as the counting process of successful transmission times before the t-th time slot of terminal n, i.e., $R_{n,\pi}(t) \triangleq \sup\{r : \sum_{k=0}^{r} X_{n,\pi}(l) \le t\}$.

(i) The successful transmission interval processes $\{X_{n,\pi}(l),\ n = 1, ..., N\}$ are independent with the packet age processes $a_n(t),\ n = 1, ..., N\}$ at terminals, with finite first and second raw moments.
(ii) The counting processes $R_{n,\pi}(t),\ n = 1, ..., N$ are renewal processes. □

It is clear that both conditions are met based on (15) and the i.i.d. property of $X_{n,\pi}(l)$ illustrated before. Thereby, following the same arguments in, e.g., [15], the long-time-average AoI can be readily calculated by the sum of the highlighted geometric areas, denoted by $Q_{l,n}$ in Figure 6:

$$\begin{aligned}\bar{\Delta}_{n,\pi} &= \limsup_{\tau\to\infty} \frac{Y}{\tau}\frac{1}{Y}\sum_{l=1}^{Y} Q_{l,n}\\ &= \limsup_{\tau\to\infty} \frac{Y}{\tau} \lim_{Y\to\infty} \frac{1}{Y}\sum_{k=1}^{Y} Q_{l,n}\\ &= \frac{\mathbb{E}[Q_{l,n}]}{T}.\end{aligned} \tag{16}$$

The last equality is based on the elementary renewal theorem [16]. It then follows that

$$\begin{aligned}\bar{\Delta}_{n,\pi} \;&=\; \frac{1}{T}\mathbb{E}\Big[X_{n,\pi}(l)(a_n(s_l) + D - 1)\\ &\quad + (X_{n,\pi}(l) - 1)\frac{X_{n,\pi}(l)}{2}\Big]\\ &\overset{(a)}{=} \frac{1}{T}\Big(\mathbb{E}\left[X_{n,\pi}(l)\right]\mathbb{E}\left[a_n(s_l)\right] + D - 1\\ &\quad + \frac{1}{2}\left(\mathbb{E}[X^2_{n,\pi}(l)] - \mathbb{E}[X_{n,\pi}(l)]\right)\Big)\\ &\overset{(b)}{=} \mathbb{E}\left[A_n(s_k)\right] + D + \frac{2 - c_{\mathsf{s},n}(L, S)}{2c_{\mathsf{s},n}(L, S)}T - \frac{3}{2},\end{aligned} \tag{17}$$

where the equality (a) is based on the condition (i), i.e., $X_{n,\pi}(l)$ and $a_n(s_l)$ are independent, and the equality (b) follows from (15).

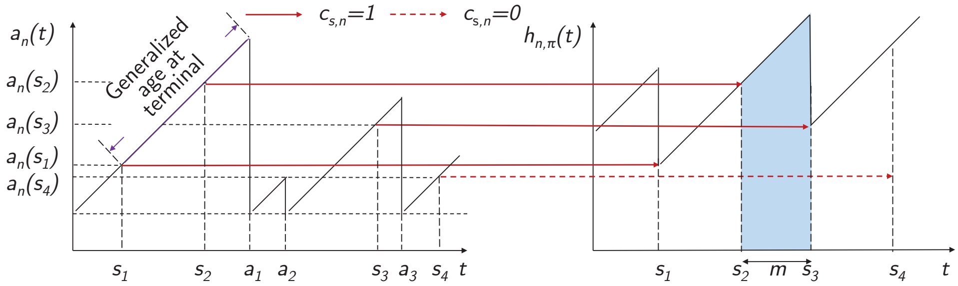

Fig. 6. Age of the packet at terminal-n assuming one packet buffer (left) and AoI at the BS (right).

Then we apply the ASTA property to calculate $\mathbb{E}\left[a_n(s_l)\right]$; the ASTA property is stated here for the completeness of the paper.

Lemma 1: [14, Theorem 3.14] Let U be a Markov state process and N be a counting process. Then ASTA holds for the pair (U, N) if U is left-continuous and the pair (U, N) is forward-pointwise independent, i.e., for all $t > 0$, $U(t)$ and $\{N(t+s) - N(s) : s \geq 1\}$ are independent.

Based on Lemma 1, let U be $\{a_n(t), t = 1, 2...\}$, and N be the counting process of the number of successful transmission times before time t. It follows from condition (i) that U and N are independent. The requirement for U to be left-continuous is also satisfied by prescribing that the age of the packet at the time of new packet arrival equals to the age of the old packet. Therefore, the conditions of Lemma 1 is upheld, expect that U, i.e., $\{a_n(t), t = 1, 2...\}$ is a Markov state process. This is obvious since the packet arrival process is Bernoulli. In addition, the steady-state stationary distribution is derived in the following lemma, facilitating the following calculation of AoI.

Lemma 2: $\{a_n(t), t = 1, 2...\}$ is a Markov state process with the steady-state stationary distribution given as

$$\alpha_n(j) = \alpha_n(1 - \lambda_n)^{j-1}, \tag{18}$$

where $\alpha_n(j)$ denotes the probability of the age of packet at terminal-n equals j time slots in the steady-state.

Proof: At each time slot, the probability of a packet arrival is λ_n and thereby the age decreases to one; otherwise, the age increases by one with probability of $1-\lambda_n$. The steady-state stationary distribution is hence a geometric distribution with parameter λ_n. ■

Then it follows that

$$\mathbb{E}\left[a_n(s_l)\right] = \lim_{Y\to\infty} \frac{1}{Y}\sum_{l=1}^{Y} a_n(s_l) = \mathbb{E}\left[a_n(t)\right] = \frac{1}{\lambda_n}, \tag{19}$$

which essentially states that the random and independent observation of the age process equals the statistical average in a long time duration. The proof of Theorem 1 is hereby concluded.

Appendix B
Proof of Theorem 2

The binomial distribution in (12) can be approximated by a continuous Gaussian distribution as stated in the De Moivre–Laplace theorem

$$q_n(K, S, m) \simeq \frac{1}{\sqrt{2\pi N\delta_n(1-\delta_n)}} e^{\frac{-(K-N\delta_n)^2}{2N\delta_n(1-\delta_n)}}, \tag{20}$$

The general rule-of-thumb of the approximation to be accurate is N is sufficiently large; specifically, the following condition should be satisfied: [17]

$$\frac{|1-2\delta_n|}{\sqrt{N\delta_n(1-\delta_n)}} = \frac{1}{\sqrt{N}}\left|\sqrt{\frac{1-\delta_n}{\delta_n}} - \sqrt{\frac{\delta_n}{1-\delta_n}}\right| < \frac{1}{3}. \tag{21}$$

It follows that

$$\begin{aligned} c_{\mathsf{s},n}(L, N) &\triangleq \int_{-\infty}^{+\infty} q_n(x, N, m) p_s(x, L)\mathrm{d}x \\ &\overset{(a)}{=} \int_{-\infty}^{\beta L} q_n(x, N, m)\mathrm{d}x \\ &= 1 - \frac{1}{2}\operatorname{erfc}\left(\frac{\beta L - N\delta_n}{\sqrt{2N\delta_n(1-\delta_n)}}\right), \end{aligned} \tag{22}$$

where the equality (a) stems from (9). Then, it becomes clear that

$$\begin{aligned} \bar{\Delta}_{\mathsf{ESM},N} &= \frac{1}{N}\sum_{n=1}^{N} \bar{\Delta}_{n,\pi} \\ &= \frac{1}{N}\sum_{n=1}^{N}\left(\mathbb{E}\left[a_n(s_l)\right] + \frac{2 - c_{\mathsf{s},n}(L, S)}{2c_{\mathsf{s},n}(L, N)}T + D - \frac{3}{2}\right) \\ &= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\lambda_n} + D - \frac{3}{2} \\ &\quad + \frac{1}{N}\sum_{n=1}^{N}\frac{1 + \frac{1}{2}\operatorname{erfc}\left(\frac{\beta L - N\delta_n}{\sqrt{2N\delta_n(1-\delta_n)}}\right)}{1 - \frac{1}{2}\operatorname{erfc}\left(\frac{\beta L - N\delta_n}{\sqrt{2N\delta_n(1-\delta_n)}}\right)}\frac{L+D}{2}. \end{aligned} \tag{23}$$

It is well-known that the complementary error function can be approximated by [18]

$$\operatorname{erfc}(x) \approx \sqrt{\frac{2e}{\pi}} \frac{\sqrt{\alpha-1}}{\alpha} e^{-\alpha x^2},\, x \geq 0,\, \alpha > 1, \tag{24}$$

and α can be adjusted to minimize the approximation error. Therefore, we can set L to be

$$L_{\mathsf{app}} = \max_{n\in\{1,\ldots,N\}} \left\{ \frac{N\delta_n + \omega\sqrt{2N\delta_n(1-\delta_n)}}{\beta} \right\}, \tag{25}$$

where $\omega > 0$ and $n_{\mathsf{max}} = \arg\max \left\{ \frac{N\delta_n+\omega\sqrt{2N\delta_n(1-\delta_n)}}{\beta} \right\}$, such that $\forall n \in \{1, ..., N\}$,

$$\operatorname{erfc}\left(\frac{\beta L_{\mathsf{app}} - N\delta_n}{\sqrt{2N\delta_n(1-\delta_n)}} \right) \lessapprox \sqrt{\frac{2e}{\pi}} \frac{\sqrt{\alpha-1}}{\alpha} e^{-\alpha\omega^2}. \tag{26}$$

It follows that the long-time average AoI with pilot length of L_{app} is

$$\begin{aligned} \bar{\Delta}_{\mathsf{ESM},N} \lessapprox & \frac{1}{N}\sum_{n=1}^{N} \frac{1}{\lambda_n} + D - \frac{3}{2} \\ & + \frac{1+\frac{1}{2}\sqrt{\frac{2e}{\pi}}\frac{\sqrt{\alpha-1}}{\alpha}e^{-\alpha\omega^2}}{1-\frac{1}{2}\sqrt{\frac{2e}{\pi}}\frac{\sqrt{\alpha-1}}{\alpha}e^{-\alpha\omega^2}} \\ & \times \left(\frac{N\delta_{n_{\mathsf{max}}} + \omega\sqrt{2N\delta_{n_{\mathsf{max}}}(1-\delta_{n_{\mathsf{max}}})}}{2\beta} + \frac{D}{2} \right). \end{aligned} \tag{27}$$

The scaling result can be obtained by

$$\begin{aligned} \lim_{N\to\infty} \frac{\bar{\Delta}_{\mathsf{ESM},N}}{N} &= \lim_{N\to\infty} \underbrace{\frac{1+\frac{1}{2}\sqrt{\frac{2e}{\pi}}\frac{\sqrt{\alpha-1}}{\alpha}e^{-\alpha\omega^2}}{1-\frac{1}{2}\sqrt{\frac{2e}{\pi}}\frac{\sqrt{\alpha-1}}{\alpha}e^{-\alpha\omega^2}}}_{\mathcal{M}_1} \\ & \quad \times \left(\frac{\delta_{n_{\mathsf{max}}}}{2\beta} + \underbrace{\frac{\omega\sqrt{2\delta_{n_{\mathsf{max}}}(1-\delta_{n_{\mathsf{max}}})}}{2\sqrt{N}\beta}}_{\mathcal{M}_2} \right) \\ &\overset{(a)}{\approx} \frac{\delta_{n_{\mathsf{max}}}}{2\beta}. \end{aligned} \tag{28}$$

The last approximation (a) is obtained by observing that a moderately large ω (e.g., $\omega > 3/\sqrt{\alpha}$) would lead to $\mathcal{M}_1 \approx 1$ and $\lim_{N\to\infty} \mathcal{M}_2 = 0$. With this, we conclude the proof.

References

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

[2] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *IEEE Int'l Symp. Info. Theory*, Jul 2012, pp. 2666–2670.

[3] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *arXiv preprint arXiv:1608.08622*, 2016.

[4] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *IEEE Int'l Symp. Info. Theory*, Jun 2015, pp. 1681–1685.

[5] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *IEEE Int'l Symp. Info. Theory*, Jul 2016, pp. 2569–2573.

[6] Z. Jiang, B. Krishnamachari, X. Zheng, S. Zhou, and Z. Niu, "Decentralized status update for age-of-information optimization in wireless multiaccess channels," in *IEEE Int'l Symp. Info. Theory*, Jun 2018.

[7] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Annu. Allerton Conf. Commun., Control, Comput.*, Sep 2016, pp. 844–851.

[8] Y.-P. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," *arXiv preprint arXiv:1801.03422*, 2018.

[9] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Can decentralized status update achieve universally near-optimal age-of-information in wireless multiaccess channels?" *arXiv preprint arXiv:1803.08189*, 2018.

[10] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov 2010.

[11] Z. Jiang, A. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[12] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2012, pp. 3249–3252.

[13] L. Liu and W. Yu, "Massive connectivity with massive MIMO-Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, pp. 1–1, 2018.

[14] B. Melamed and D. Yao, "The ASTA property," *Advances in Queueing: Theory, Methods and Open Problems*, pp. 195–224, 1995.

[15] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *IEEE Int'l Symp. Info. Theory*, Jun 2017, pp. 131–135.

[16] D. R. Cox, *Renewal theory*. Methuen London, 1967, vol. 1.

[17] G. E. Box, J. S. Hunter, and W. G. Hunter, *Statistics for experimenters: design, innovation, and discovery*. Wiley-Interscience New York, 2005, vol. 2.

[18] S. H. Chang, P. C. Cosman, and L. B. Milstein, "Chernoff-type bounds for the Gaussian error function," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2939–2944, Nov 2011.