Impact of NOMA on Age of Information: A Grant-Free Transmission Perspective

Zhiguo Ding, *Fellow, IEEE*, Robert Schober, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

Abstract

The aim of this paper is to characterize the impact of non-orthogonal multiple access (NOMA) on the age of information (AoI) of grant-free transmission. In particular, a low-complexity form of NOMA, termed NOMA-assisted random access, is applied to grant-free transmission in order to illustrate the two benefits of NOMA for AoI reduction, namely increasing channel access and reducing user collisions. Closed-form analytical expressions for the AoI achieved by NOMA assisted grant-free transmission are obtained, and asymptotic studies are carried out to demonstrate that the use of the simplest form of NOMA is already sufficient to reduce the AoI of orthogonal multiple access (OMA) by more than 40%. In addition, the developed analytical expressions are also shown to be useful for optimizing the users' transmission attempt probabilities, which are key parameters for grant-free transmission.

I. Introduction

Grant-free transmission is an important enabling technique to support the sixth-generation (6G) services, including ultra massive machine type communications (umMTC) and enhanced ultra-reliable low latency communications (euRLLC) [1], [2]. Unlike conventional grant-based transmission, grant-free transmission enables users to avoid multi-step signalling and directly transmit data signals together with access control information, which can reduce the system overhead significantly, particularly for scenarios with massive users requiring short-package transmission. Grant-free transmission can be realized by applying the random access protocols developed for conventional computer networks, such as ALOHA random access [3], [4].

Z. Ding and H. V. Poor are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA. Z. Ding is also with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, UK (email: zhiguo.ding@manchester.ac.uk, poor@princeton.edu). R. Schober is with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg (FAU), Germany (email: robert.schober@fau.de).

Alternatively, massive multi-input multi-output (MIMO) can also be applied to support grant-free transmission by using the excessive spatial degrees of freedom offered by massive MIMO [5]–[7].

Recently, the application of non-orthogonal multiple access (NOMA) to grant-free transmission has received significant attention due to the following two reasons. First, the NOMA principle is highly compatible, and the use of NOMA can significantly improve the reliability and spectral efficiency of random access and massive MIMO based grant-free protocols [8]–[11]. Second, more importantly, the use of NOMA alone is sufficient to support grant-free transmission. For example, NOMA-based grant-free transmission has been proposed in [12], where a Bayesian learning based scheme has been designed to ensure successful multi-user detection, even if the number of active grant-free users is unknown. The principle of NOMA can also be used to develop so-called semi-grant-free transmission protocols, where the bandwidth resources which would be solely occupied by grant-based users are released for supporting multiple grant-free users in a distributed manner [13]. In addition, NOMA-based grant-free transmission has also been shown to be robust and efficient in various communication scenarios, such as satellite communication networks, secure Internet of Things (IoT), intelligent reflecting surface (IRS) networks, marine communication systems, etc., see [14]–[17].

The aim of this paper is to characterize the impact of NOMA on the performance of grant-free transmission with respect to a recently developed new performance metric, termed the age of information (AoI) [18]–[20]. In particular, the AoI describes the freshness of data updates collected in the network, and is an important metric to measure the success of the 6G services, including umMTC and euRLLC. We note that the impact of NOMA on the AoI has been previously studied mainly for grant-based networks. For example, for two-user grant-based networks, the capability of NOMA to reduce the AoI has been shown to be related to the spectral efficiency gain of NOMA over orthogonal multiple access (OMA) [21]. For multi-user grant-based networks, cognitive-radio (CR) NOMA has been used to illustrate that NOMA provides two benefits for AoI reduction [22]. One is that, with NOMA, users have more opportunities to update their base station than with OMA, and the other is that, with NOMA, users can be scheduled to transmit earlier than with OMA. These benefits of NOMA for improving the AoI of grant-based networks have also been confirmed in [23]–[26].

In this paper, the impact of NOMA on the AoI of grant-free transmission is investigated from the perspective of performance analysis, which is different from the existing work focusing on resource allocation [29]. In particular, a low-complexity form of NOMA, termed NOMA-assisted random access, is adopted in order to illustrate the two benefits of NOMA for AoI reduction, namely increasing channel access and reducing user collisions [27]. The key element of the proposed performance analysis is the modelling of the channel competition among the grant-free users as a Markov chain, which is different from the performance analysis for grant-based NOMA networks [21]–[24]. The main contributions of this paper are two-fold:

- Analytical expressions for the AoI achieved by NOMA assisted grant-free transmission are obtained, by rigorously characterizing the state transition probabilities of the considered Markov chain. We note that by using NOMA-assisted random access, the base station creates multiple preconfigured receive signal-to-noise ratio (SNR) levels, which makes NOMA-assisted grant-free transmission similar to multi-channel ALOHA. As a result, the calculation of the state transition probabilities for the NOMA case is more challenging than that for the OMA case, which can be viewed as single-channel ALOHA. By exploiting the properties of the considered Markov chain and also the characteristics of successive interference cancellation (SIC), closed-form expressions for the state transition probabilities are developed for NOMA assisted grant-free transmission.
- Valuable insights regarding the relative performance of NOMA and OMA assisted grant-free transmission are also obtained. For example, for the case where users always have updates to deliver, asymptotic expressions are developed to demonstrate that the use of NOMA can almost halve the AoI achieved by OMA, even if the simplest form of NOMA is implemented. In addition, the optimal choices of the users' transmission probabilities for random access with NOMA and OMA are obtained and compared. This study reveals that NOMA-assisted grant-free transmission is fundamentally different from multi-channel ALOHA due to the use of SIC. Furthermore, simulation results are provided to verify the developed analytical expressions and also demonstrate that the use of NOMA can significantly reduce the AoI compared to OMA, particularly for the case of low transmit SNR and a massive number of grant-free users.

The remainder of this paper is organized as follows. In Section II, the system model for grant-free transmission is introduced. In Section III, the AoI achieved by OMA and NOMA assisted grant-free transmission is analyzed, and numerical results are presented in Section IV. Finally, the paper is concluded in Section V, and all proofs are collected in the appendix.

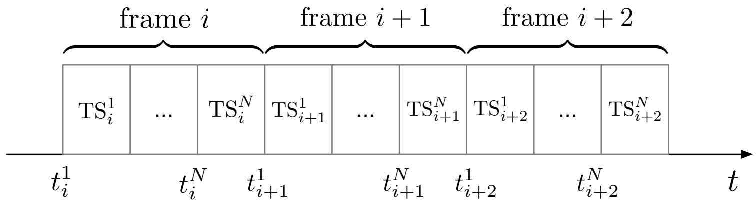


Fig. 1. Considered slotted time frame structure.

II. NOMA Assisted Grant-Free Transmission

Consider a grant-free communication network with M users, denoted by U_m, $1 \leq m \leq M$, communicating with the same base station. Assume that each time frame comprises N time slots, each of duration of T seconds, where the n-th time slot of the i-th frame is denoted by TS_i^n, and the starting time of TS_i^n is denoted by t_i^n, $1 \leq n \leq N$ and $i \geq 1$, as shown in Fig. 1.

A. Data Generation Models

For the considered grant-free transmission scenario, each user tries to deliver one update to the base station in each time frame. When the users' updates are generated depends on which of the following two data generation models is used [22].

1) Generate-at-request (GAR): For GAR, the base station requests all users to simultaneously generate their updates at the beginning of each time frame. GAR is applicable to many important IoT applications, such as structural health monitoring and autonomous driving.

2) Generate-at-will (GAW): For GAW, a user's update is generated right before its transmit time slot. GAW has been commonly used in the AoI literature, since it ensures that the delivered updates are freshly generated.

In this paper, GAR will be focused on due to the following two reasons. First, the AoI expression for grant-free transmission for GAW can be straightforwardly obtained from that for GAR, as shown in the next section. Second, if there are retransmissions within one time frame, GAW requires a user to repeatedly generate updates, and hence, causes a higher energy consumption than GAW. For grant-based transmission, this increase in energy consumption is not severe since the number of retransmissions in one frame is small [22]. However, in the grant-free case, a user might have to carry out a large number of retransmissions due to potential collisions, which means that GAW can cause a significantly higher energy consumption than GAR.

B. Channel Access Modes

1) Orthogonal Multiple Access (OMA): A simple example of OMA based grant-free transmission is slotted ALOHA, as described in the following. Prior to TS_i^n, assume that j users have successfully delivered their updates to the base station. For grant-free transmission, each of the remaining $M-j$ users will independently make a transmission attempt with the same transmit power, denoted by P, and the same transmission probability, denoted by $\mathbb{P}_{\mathrm{TX}}$. $\mathbb{P}_{\mathrm{TX}}$ can be based on a fixed choice, or be state-dependent, i.e., $\mathbb{P}_j = \frac{1}{M-j}$ [28].

There are three possible events which cause an update failure for a user: i) the user does not make a transmission attempt; ii) a collision occurs, i.e., there are more than one concurrent transmissions; iii) an outage occurs due to the user's weak channel condition, i.e., $\log(1 + P|h_m^{i,n}|^2) \leq R$, where R denotes the user's target data rate, and $h_m^{i,n}$ denotes U_m's channel gain in TS_i^n. In this paper, all users are assumed to have the identical target data rates, and their channel gains are assumed to be independent and identically complex Gaussian distributed with zero mean and unit variance.

2) Non-orthogonal Multiple Access (NOMA): Recall that the principle of NOMA can be implemented in different forms. For the purpose of illustration, a particular form of NOMA, termed NOMA-assisted random access, is adopted to reduce the AoI of grant-free transmission [27]. In particular, prior to transmission, the base station configures K receive SNR levels, denoted by $P_1 \geq \cdots \geq P_K$. If U_m chooses P_k during TS_i^n, it will scale its transmitted signal by $\sqrt{\frac{P_k}{|h_m^{i,n}|^2}}$. The base station carries out SIC by decoding the signal delivered at SNR level, P_k, before decoding the one at P_{k+1}, $1 \leq k \leq K-1$. The SNR levels are preconfigured to guarantee the success of SIC, i.e., the following conditions need to be satisfied:

$$\log\left(1 + \frac{P_k}{1 + \sum_{i=k+1}^{K} P_i}\right) = R, \quad 1 \leq k \leq K-1, \tag{1}$$

and $\log(1 + P_K) = R$, which means $P_K = 2^R - 1$ and $P_k = \left(2^R - 1\right)\left(1 + \sum_{i=k+1}^{K} P_i\right)$, where the noise power is assumed to be normalized to one.

Again assume that there are j users which have successfully sent their updates to the base station prior to TS_i^n. Each of the remaining $M-j$ users will first randomly choose an SNR level with equal probability, denoted by $\mathbb{P}_K = \frac{1}{K}$, and independently make a transmission attempt with probability $\mathbb{P}_{\mathrm{TX}}$. For illustrative purposes, assume that U_m is among the $M-j$ remaining users, and chooses P_k. The possible events which cause U_m's update to fail are listed as follows:

- The user does not make an attempt for transmission;
- The receive SNR level chosen by the user is not feasible due to the user's transmit power budget, i.e., P_k is not feasible for U_m in TS_i^n if $\frac{P_k}{|h_m^{i,n}|^2} > P$;
- Another user also chooses P_k, which leads to a collision at P_k and hence a failure at the k-th stage of SIC;
- Prior to the k-th stage of SIC, SIC has already been terminated due to one or more failures in the previous SIC stages.

C. AoI Model

AoI is an important performance metric for quantifying the freshness of the updates delivered to the base station. We note that for the considered grant-free scenario, all the users experience the same AoI. Therefore, without loss of generality, U_1's instantaneous AoI at time t is focused on and defined as follows [18]:

$$\Delta(t) = t - T(t), \tag{2}$$

where $T(t)$ denotes the generation time of U_1's freshest update successfully delivered to the base station. U_1's average AoI of the considered network is given by

$$\bar{\Delta} = \lim_{T_\Delta \to \infty} \frac{1}{T_\Delta} \int_0^{T_\Delta} \Delta(t) dt. \tag{3}$$

The AoI achieved by OMA and NOMA assisted grant-free transmission will be analyzed in the following section.

III. AoI of Grant-Free Transmission

As discussed in the previous section, U_1 is treated as the tagged user and its AoI will be focused on in this section, without loss of generality. For the example shown in Fig. 2, U_1 successfully sends its updates to the base station in TS_n^i of frame i and TS_l^{i+2} of frame $i+2$, but fails in frame $i+1$. For the AoI analysis, the following metrics are required:

- S_j: The time duration between the generation time and the receive time of the j-th successful update. For the example shown in Fig. 2, $S_{j-1} = nT$ and $S_j = lT$.
- Y_j: The time duration between the $(j-1)$-th and the j-th successful updates. For the example shown in Fig. 2, $Y_j = (N-n)T + NT + lT$.
- X_j: The number of frames between the $(j-1)$-th and the j-th successful updates. An example of $X_j = 2$ is shown in Fig. 2.

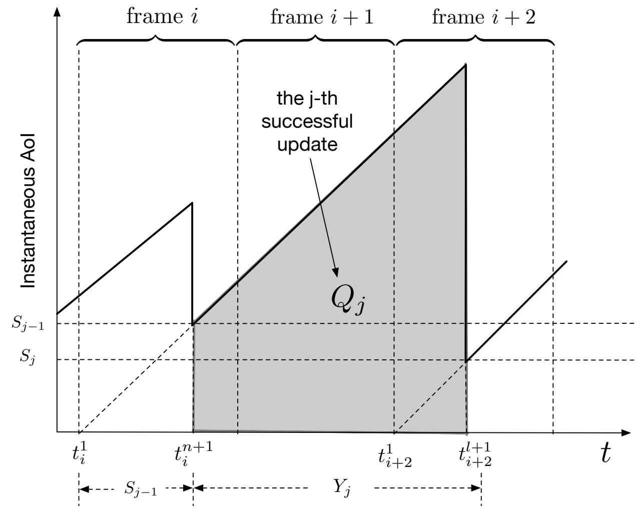


Fig. 2. Illustration of AoI evolution with grant-free transmission for GAR.

We note that in the literature of random access, S_j is termed the service delay and Y_j is termed the inter-departure time [28].

By using the aforementioned metrics, for GAR, U_1's averaged AoI is given by

$$\begin{aligned}\bar{\Delta} = &\lim_{J\to\infty} \frac{\sum_{j=1}^{J} Q_j}{\sum_{j=1}^{J} Y_j} = \lim_{J\to\infty} \frac{\sum_{j=1}^{J} \left(S_{j-1}Y_j + \frac{1}{2}Y_j^2\right)}{\sum_{j=1}^{J} Y_j} \\ = &\frac{\mathcal{E}\{S_{j-1}Y_j\}}{\mathcal{E}\{Y_j\}} + \frac{\mathcal{E}\{Y_j^2\}}{2\mathcal{E}\{Y_j\}}, \end{aligned} \tag{4}$$

where J denotes the total number of successful updates, Q_j denotes the area of the shaded region shown in Fig. 2, $\mathcal{E}\{Y_j^2\} = \lim_{J\to\infty} \frac{\sum_{j=1}^{J} Y_j^2}{J}$, $\mathcal{E}\{Y_j\} = \lim_{J\to\infty} \frac{\sum_{j=1}^{J} Y_j}{J}$, and $\mathcal{E}\{\cdot\}$ denotes the expectation.

Remark 1: We note that the AoI expressions in (4) of this paper and [28, Eq. (3)] are consistent. The reason why there is an extra factor of $\frac{1}{2}$ in [28, Eq. (3)] is that the users' instantaneous AoI was assumed to be discrete-valued in [28], instead of continuous-valued as in this paper.

Remark 2: For GAW, the user's averaged AoI is given by

$$\bar{\Delta}^{\mathrm{GAW}} = \lim_{J\to\infty} \frac{\sum_{j=1}^{J} \left(TY_j + \frac{1}{2}Y_j^2\right)}{\sum_{j=1}^{J} Y_j} = T + \frac{\mathcal{E}\{Y_j^2\}}{2\mathcal{E}\{Y_j\}}, \tag{5}$$

which is simpler than the AoI expression in (4). Therefore, the analytical results developed for GAR are straightforwardly applicable to the case for GAW.

A. Generic Expressions for $\mathcal{E}\{S_{j-1}Y_j\}$, $\mathcal{E}\{Y_j\}$, and $\mathcal{E}\{Y_j^2\}$

As shown in (4), the AoI is a function of $\mathcal{E}\{S_{j-1}Y_j\}$, $\mathcal{E}\{Y_j\}$, and $\mathcal{E}\{Y_j^2\}$, and generic expressions for these metrics have been derived in [28], and will be briefly introduced in this subsection. In particular, the considered grant-free transmission can be modelled by a Markov chain with $M+1$ states, denoted by s_k, $0 \leq k \leq M$. In particular, s_k, $0 \leq k \leq M-1$, denotes the transient state, where k users, other than U_1, have successfully delivered their updates to the base station. s_M means that U_1 has successfully delivered its update to the base station. Define the state transition probability from s_j to s_i by $P_{j,i}$, $0 \leq i, j \leq M$. Build an $M \times M$ matrix, denoted by $\mathbf{P}$, whose element in the $(i+1)$-th column and the $(j+1)$-th row is $P_{j,i}$, $0 \leq i, j \leq M-1$. Furthermore, build an $M \times 1$ vector, denoted by $\mathbf{p}$, whose $(j+1)$-th element is $P_{j,M}$. Once $\mathbf{P}$ and $\mathbf{p}$ are available, $\mathcal{E}\{S_{j-1}Y_j\}$, $\mathcal{E}\{Y_j\}$, and $\mathcal{E}\{Y_j^2\}$ can be obtained as follows.

Denote by Z the number of time slots required by U_1 to successfully deliver its update to its base station. Then, the probability mass function (pmf) of Z is given by

$$\mathbb{P}\left(Z=n\right) = \mathbf{s}_0^T \mathbf{P}_M^{n-1} \mathbf{p}, \quad n = 1, 2, \cdots, \tag{6}$$

where $\mathrm{s}_0 = \begin{bmatrix} 1 & \mathbf{0}_{1\times(M-1)} \end{bmatrix}^T$ denotes the initial probability vector and $\mathbf{0}_{m\times n}$ denotes an all-zero $m \times n$ matrix. Therefore, the probability that U_1 cannot complete an update within one frame is given by $P_{\text{fail}} = \mathbb{P}\left(Z > N\right) = \mathbf{s}_0^T \mathbf{P}_M^N \mathbf{1}$, where $\mathbf{1}$ denotes an $M \times 1$ all-one vector. Therefore, the pmf of access delay, S_j, can be written as follows:

$$\mathbb{P}(S_j = nT) = \frac{\mathbb{P}\left(Z=n\right)}{1-P_{\text{fail}}} = \frac{\mathbf{s}_0^T \mathbf{P}_M^{n-1} \mathbf{p}}{1 - \mathbf{s}_0^T \mathbf{P}_M^N \mathbf{1}}, \tag{7}$$

for $1 \leq n \leq N$, which means $\mathcal{E}\{S_j\} = T\sum_{n=1}^N n \frac{\mathbf{s}_0^T \mathbf{P}_M^{n-1} \mathbf{p}}{1 - \mathbf{s}_0^T \mathbf{P}_M^N \mathbf{1}}$ and $\mathcal{E}\{S_j^2\} = T^2\sum_{n=1}^N n^2 \frac{\mathbf{s}_0^T \mathbf{P}_M^{n-1} \mathbf{p}}{1 - \mathbf{s}_0^T \mathbf{P}_M^N \mathbf{1}}$.

Similarly, the pmf of X_j is given by

$$\mathbb{P}(X_j = n) = P_{\text{fail}}^{n-1}(1 - P_{\text{fail}}), \tag{8}$$

which means that $\mathcal{E}\{X_j\} = \frac{1}{1-P_{\text{fail}}}$ and $\mathcal{E}\{X_j^2\} = \frac{1+P_{\text{fail}}}{(1-P_{\text{fail}})^2}$. The expressions of $\mathcal{E}\{X_j\}$ and $\mathcal{E}\{X_j^2\}$ can be used to evaluate $\mathcal{E}\{Y_j\}$ and $\mathcal{E}\{Y_j^2\}$, since $\mathcal{E}\{Y_j\} = TN\mathcal{E}\{X_j\}$ and $\mathcal{E}\{Y_j^2\} = N^2T^2\mathcal{E}\left\{X_j^2\right\} + 2\mathcal{E}\left\{S_j^2\right\} - 2\mathcal{E}\left\{S_j\right\}^2$. Furthermore, $\mathcal{E}\{S_j\}$, $\mathcal{E}\{S_j^2\}$, and $\mathcal{E}\{Y_j\}$ can be used to evaluate $\mathcal{E}\{S_{j-1}Y_j\}$ which can be expressed as follows:

$$\mathcal{E}\{S_{j-1}Y_j\} = \mathcal{E}\{S_j\}\mathcal{E}\{Y_j\} - \mathcal{E}\{S_j^2\} + \mathcal{E}\{S_j\}^2,$$

where the last step follows by the fact that S_{j-1} and $Y_j - (NT - S_{j-1})$ are independent.

As discussed above, the crucial step to evaluate the AoI is to find $\mathbf{P}$ and $\mathbf{p}$, which depends on the used multiple access schemes.

B. OMA-Based Grant-Free Transmission

The state transition probabilities for the OMA case can be straightforwardly obtained, as shown in the following. With OMA, a single user can be served in each time slot, which means that the number of successful users after one time slot can be increased by one at most. Therefore, most of the state transition probabilities in matrix $\mathbf{P}$ are zero, except for $P_{j,j}$, $P_{j,j+1}$, and $P_{j,M}$, $0 \leq j \leq M-1$. In particular, $P_{j,j}$ denotes the probability of the event that no user succeeds, and is given by [28]

$$P_{j,j} = 1-(M-j)\mathbb{P}_{\rm TX}e^{-\frac{\epsilon}{P}}\left(1-\mathbb{P}_{\rm TX}\right)^{M-j-1}, \tag{9}$$

where $\epsilon = 2^R-1$. $P_{j,j}$ denotes the probability of the event that a single user, but not U_1, succeeds and is given by

$$P_{j,j+1} = (M-j-1)\mathbb{P}_{\rm TX}e^{-\frac{\epsilon}{P}}\left(1-\mathbb{P}_{\rm TX}\right)^{M-j-1}. \tag{10}$$

Furthermore, the j-th element of $\mathbf{p}$, denoted by $P_{j,M}$, is given by

$$P_{j,M} = \mathbb{P}_{\rm TX}e^{-\frac{\epsilon}{P}}\left(1-\mathbb{P}_{\rm TX}\right)^{M-j-1}. \tag{11}$$

C. NOMA-Based Grant-Free Transmission

The benefit of using NOMA is that more users can be admitted simultaneously than for OMA. In particular, with NOMA, the number of successful users after one time slot can be increased by K at most, whereas the number of successful users was no more than 1 for OMA. This means that the non-zero state transition probabilities in matrix $\mathbf{P}$ include $P_{j,j}$, $P_{j,j+i}$, and $P_{j,M}$, $0 \leq j \leq M-1$, $1 \leq i \leq K$ and $j+i \leq M-1$.

The analysis of the state transition probabilities for the NOMA case is more challenging than that for the OMA case, mainly due to the application of SIC. For example, a collision at SNR level P_k can prevent all those users, which choose SNR level P_i, $i > k$, from being successful. The following lemma provides a high-SNR approximation for the state transition probabilities.

Lemma 1. *At high SNR, the state transition probability, $P_{j,j}$, $0 \leq j \leq M-1$, can be approximated as follows:*

$$\begin{aligned} P_{j,j} \approx & 1-\sum_{m=1}^{M-j}\binom{M-j}{m}\mathbb{P}_{\rm TX}^m(1-\mathbb{P}_{\rm TX})^{M-j-m} \\ & \times \sum_{k=1}^{K} m\mathbb{P}_K\left(1-k\mathbb{P}_K\right)^{m-1}, \end{aligned} \tag{12}$$

the state transition probability, $P_{j,j+1}$, $0 \leq j \leq M-2$, can be approximated as follows:

$$\begin{aligned} P_{j,j+1} \approx & (M-j)\mathbb{P}_{\mathrm{TX}}(1-\mathbb{P}_{\mathrm{TX}})^{M-j-1}\frac{M-j-1}{M-j}K\mathbb{P}_K \\ & + \sum_{m=2}^{M-j}\binom{M-j}{m}\mathbb{P}_{\mathrm{TX}}^m(1-\mathbb{P}_{\mathrm{TX}})^{M-j-m}\sum_{k=1}^{K-1}\frac{M-j-1}{M-j}m\mathbb{P}_K \\ & \left[(1-k\mathbb{P}_K)^{m-1} - \sum_{\kappa=k+1}^{K}(m-1)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-2}\right], \end{aligned} \tag{13}$$

and the state transition probability, $P_{j,j+i}$, $0 \leq j \leq M-3$ and $2 \leq i \leq \min\{M-1-j, K\}$, can be approximated as follows:

$$\begin{aligned} P_{j,j+i} \approx & \binom{M-j}{i}\mathbb{P}_{\mathrm{TX}}^i(1-\mathbb{P}_{\mathrm{TX}})^{M-j-i}\sum_{k_1=1}^{K-i+1}\sum_{k_2=k_1+i-1}^{K}\frac{M-j-i}{M-j}\mathbb{P}_K^i\binom{k_2-k_1-1}{i-2}\prod_{p=1}^{i}p \\ & + \sum_{m=i+1}^{M-j}\binom{M-j}{m}\mathbb{P}_{\mathrm{TX}}^m(1-\mathbb{P}_{\mathrm{TX}})^{M-j-m} \\ & \times \sum_{k_1=1}^{K-i}\sum_{k_2=k_1+i-1}^{K-1}\frac{M-j-i}{M-j}\mathbb{P}_K^i\binom{k_2-k_1-1}{i-2}\prod_{p=0}^{i-1}(m-p) \\ & \times \left[(1-k_2\mathbb{P}_K)^{m-i} - \sum_{\kappa=k_2+1}^{K}(m-i)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-i-1}\right]. \end{aligned} \tag{14}$$

Proof. See Appendix A. □

Once the transition probability matrix $\mathbf{P}$ is obtained, the elements of $\mathbf{p}$ can be obtained straightforwardly by applying $\mathbf{P}\mathbf{1}+\mathbf{p} = \mathbf{1}$, where recall that $\mathbf{1}$ denotes an $M \times 1$ all-one vector.

The closed-form analytical expressions shown in Lemma 1 allow the evaluation of the impact of NOMA on the AoI without carrying out intensive Monte Carlo simulations. However, the expressions of the state transition probabilities shown in Lemma 1 are quite involved, which makes it difficult to obtain insights about the performance difference between OMA and NOMA. For this reason, the special case of $K = 2$ and $N = 1$ is focused on in the remainder of this section. $K = 2$ means that there are two SNR levels, i.e., the base station needs to carry out two-stage SIC only, which is an important case in practice due to its low system complexity. $N = 1$ implies that there is one time slot in each frame, i.e., in each time slot, all users have updates to deliver and hence always participate in contention.

For this case, the following lemma provides the optimal choice for the transmission probability $\mathbb{P}_{\mathrm{TX}}$.

Lemma 2. *For the special case of $K = 2$ and $N = 1$, the optimal choice for the transmission probability $\mathbb{P}_{\mathrm{TX}}$ is given by*

$$\mathbb{P}^*_{\mathrm{TX}} = \frac{\eta}{M}, \tag{15}$$

for $M \to \infty$ and $P \to \infty$, where η is the root of the following equation: $\left(1 - \frac{\eta}{2}\right) e^{-\frac{\eta}{2}} + \left(1 - \frac{\eta^2}{2}\right) e^{-\eta} = 0$.

Proof. See Appendix B. □

Remark 3: Note that $\left(1 - \frac{\eta}{2}\right) e^{-\frac{\eta}{2}} + \left(1 - \frac{\eta^2}{2}\right) e^{-\eta} = 0$ is not related to M, which means that η is not a function of M. By applying off-shelf root solvers, the exact value of η can be straightforwardly obtained as follows: $\eta \approx 1.6646$.

By using Lemma 2, the AoI performance difference between NOMA and OMA is analyzed in the following proposition.

Proposition 1. *For the special case of $K = 2$ and $N = 1$, for $M \to \infty$ and $P \to \infty$, the ratio between the AoI achieved by NOMA and OMA is given by*

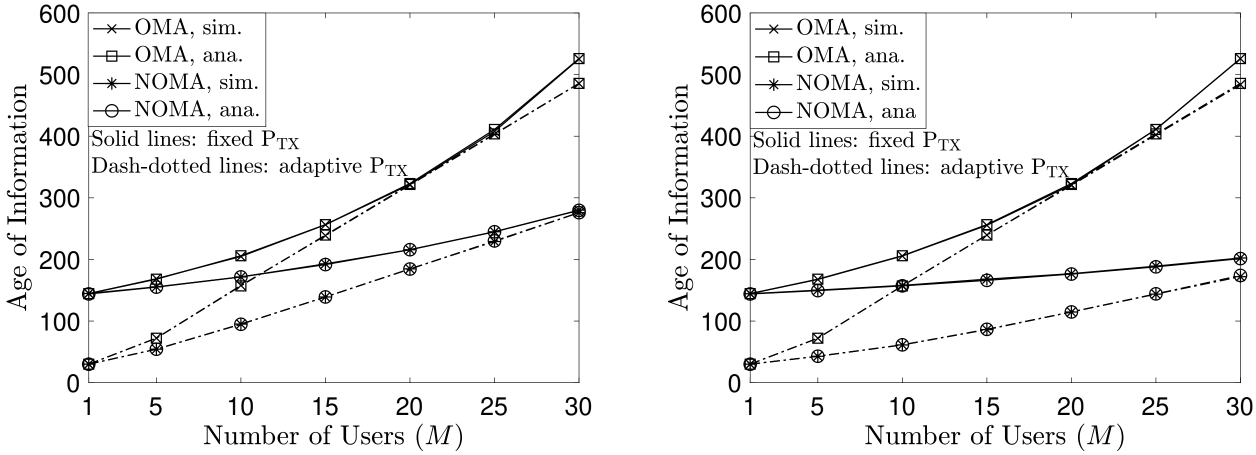
$$\frac{\bar{\Delta}^N}{\bar{\Delta}^O} \approx \frac{2e^{\eta - 1}}{\eta\left(e^{\frac{\eta}{2}} + 1 + \frac{\eta}{2}\right)}, \tag{16}$$

where $\bar{\Delta}^N$ and $\bar{\Delta}^O$ denote the AoI achieved by NOMA and OMA, respectively.

Proof. See Appendix C. □

Remark 4: Recall that $\eta \approx 1.6646$, which means that the ratio in Proposition 1 is $\frac{\bar{\Delta}^N}{\bar{\Delta}^O} \approx 0.5653$, i.e., the use of NOMA can almost halve the AoI achieved by OMA. We note that this significant performance gain is achieved by using NOMA with two SNR levels only, i.e., the simplest form of NOMA. By implementing NOMA with more than two SNR levels, the performance gain of NOMA over OMA can be further increased, as shown in the next section.

Remark 5: As shown in the proof for Proposition 1, the optimal choice of $\mathbb{P}_{\mathrm{TX}}$ for OMA is $\frac{1}{M}$. This is expected as explained in the following. With M users competing for access in a single channel, the use of a transmission probability of $\frac{1}{M}$ is reasonable. By using the same rationale, one might expect that $\frac{2}{M}$ should be optimal for the NOMA transmission probability with two SNR levels. However, Lemma 2 shows that the optimal value of $\mathbb{P}_{\mathrm{TX}}$ is $\frac{1.6646}{M}$, which is a more conservative choice for transmission than $\frac{2}{M}$. The reason for this are potential SIC errors. In particular, although there are two channels (or two SNR levels, P_1 and P_2), a collision



(a) With Two SNR Levels ($K = 2$)

(b) With Four SNR Levels ($K = 4$)

Fig. 3. Impact of the number of users on the average AoI achieved with OMA and NOMA assisted grant-free transmission for GAR. $T = 6$, $P = 20$ dB, $R = 0.5$ BPCU, and $N = 8$. For the fixed choice of $\mathbb{P}_{\rm TX}$, $\mathbb{P}_{\rm TX} = 0.05$, and for the adaptive choice of $\mathbb{P}_{\rm TX}$, $\mathbb{P}_{\rm TX} = \min\left\{1, \frac{K}{M}\right\}$ for NOMA, and $\mathbb{P}_{\rm TX} = \frac{1}{M-j}$ for OMA, i.e., a state-dependent choice is used for OMA as discussed in Section II-B.

at P_1 causes SIC to immediately terminate, which means that P_2 can no longer be used to serve any users, i.e., the number of the effective channels is less than 2.

Remark 6: Motivated by the results shown in Lemma 2 and Proposition 1, for the general case of $K > 2$, a simple choice of $\mathbb{P}_{\rm TX} = \min\left\{1, \frac{K}{M}\right\}$ can be used for NOMA. In fact, the simulation results presented in the next section show that this choice is sufficient to realize a significant performance gain of NOMA over OMA. We note that this choice of $\mathbb{P}_{\rm TX}$ depends on M only, unlike the the state-dependent choice of $\mathbb{P}_{\rm TX}$ used for OMA, which is $\mathbb{P}_{\rm TX} = \frac{1}{M-j}$ [28]. Therefore, an important direction for future research is to find a more sophisticated state-dependent choice of $\mathbb{P}_{\rm TX}$ for NOMA-assisted grant-free transmission.

Remark 7: Proposition 1 implies that the application of NOMA is particularly beneficial for grant-free transmission, compared to its application to grant-based transmission. Recall that in grant-based networks, one important benefit of using NOMA for AoI reduction is that a user can be scheduled to transmit earlier than with OMA [22]. However, for a user which has already been scheduled to transmit early in OMA, the impact of NOMA on the user' AoI can be insignificant, particularly at high SNR. Unlike grant-based networks, Proposition 1 shows that in grant-free networks, the use of NOMA can reduce the AoI of OMA by more than 40%, and this significant performance gain applies for all users in the network.

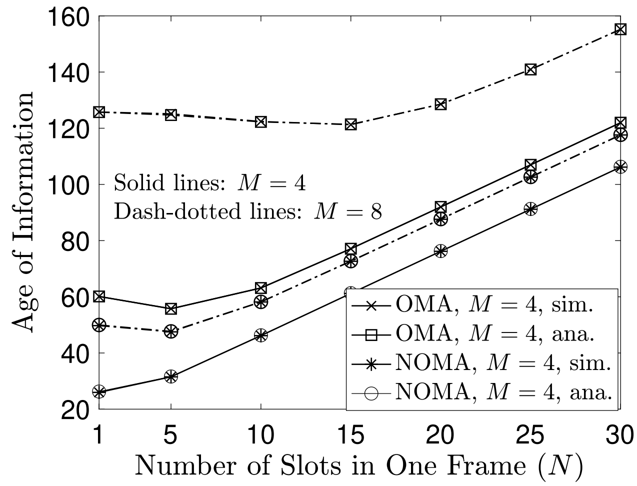


Fig. 4. Impact of the number of time slots in each frame on the average AoI achieved by the OMA and NOMA assisted grant-free transmission schemes for GAR. $K = 4$, $T = 6$, $P = 20$ dB, $R = 0.5$ BPCU, $M = 8$, and the adaptive choices for $\mathbb{P}_{\rm TX}$ are used.

IV. Simulation Results

In this section, simulation results are presented to demonstrate the AoI achieved by the considered grant-free transmission schemes and to also verify the developed analytical results.

In Fig. 3, the impact of the number of users on the average AoI achieved by the considered grant-free transmission schemes is investigated. As can be seen from the figure, the AoI achieved with NOMA is significantly lower than that of OMA. In addition, Fig. 3 demonstrates that the performance gain of NOMA over OMA increases as the number of users, M, grows. This observation can be explained by using Proposition 1 which states that, for $K = 2$, $N = 1$, $M \rightarrow \infty$ and $P \rightarrow \infty$, $\bar{\Delta}^N \approx 0.5653\bar{\Delta}^O$, or equivalently $\bar{\Delta}^N - \bar{\Delta}^O \approx 0.4347\bar{\Delta}^O$. Because increasing M increases $\bar{\Delta}^O$, the performance gain of NOMA over OMA also increases as the number of users grows. Therefore, the use of NOMA is particularly important for grant-free transmission with a massive number of users, an important use case for 6G. Between the two choices of $\mathbb{P}_{\rm TX}$, the adaptive choice yields a better AoI than the fixed choice. For the two subfigures in Fig. 3, different numbers of SNR levels, K, are used. By comparing the two subfigures, one can observe that the AoI achieved by the NOMA scheme can be reduced by increasing the number of SNR levels. This is because the use of more SNR levels makes user collisions less likely to happen. Fig. 3 also demonstrates the accuracy of the AoI expressions developed in Lemma 1.

In Fig. 4, the impact of the number of time slots in each frame, N, on the average AoI achieved by the two considered grant-free transmission schemes is studied. As can be seen from

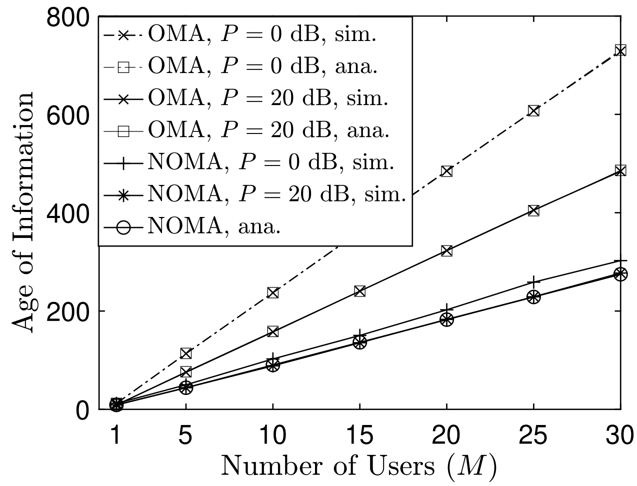


Fig. 5. AoI performance achieved by the two grant-free transmission schemes for the special case with $K = 2$ and $N = 1$, where $T = 6$, $R = 0.5$ BPCU, and the optimal choices of $\mathbb{P}_{\rm TX}$ are used.

the figure, the use of NOMA can always realize lower AoI than OMA, regardless of the choices of N. An interesting observation from Fig. 4 is that a small increase of N, e.g., from 1 to 5, can reduce the AoI. This is because the likelihood for users to deliver their updates to the base station is improved if there are more time slots in each frame. However, after $N \geq 10$, further adding more time slots in each frame increases the AoI, which can be explained with the following example. Assume that U_1 can always successfully update its base station in the first time slot of each frame. For this example, U_1's AoI is simply the length of one time frame, and hence its AoI is increased if there are more time slots in one frame.

As discussed in the previous section, the special case with $K = 2$ and $N = 1$ is important in practice, and hence the AoI realized by the OMA and NOMA assisted grant-free transmission schemes is investigated in Fig. 5. In particular, Fig. 5 demonstrates that when the transmit SNR is large, i.e., $P = 20$ dB, the developed analytical results match perfectly with the simulation results. For the case with low transmit SNR, i.e., $P = 0$ dB, there is a small gap between the simulation and analytical results, which is due to the fact that the analytical results were developed based on the high SNR assumption. It is also interesting to observe that OMA based grant-free transmission is sensitive to a change of the transmit SNR. For example, the AoI achieved by OMA is around 480 seconds for $P = 20$ dB and increased to 730 seconds for $P = 0$ dB, whereas the performance loss for NOMA due to the change of P is insignificant. In other words, the performance gain of NOMA over OMA is particularly large at low SNR. This is a valuable property in practice since most AoI sensitive applications, such as IoT and sensor

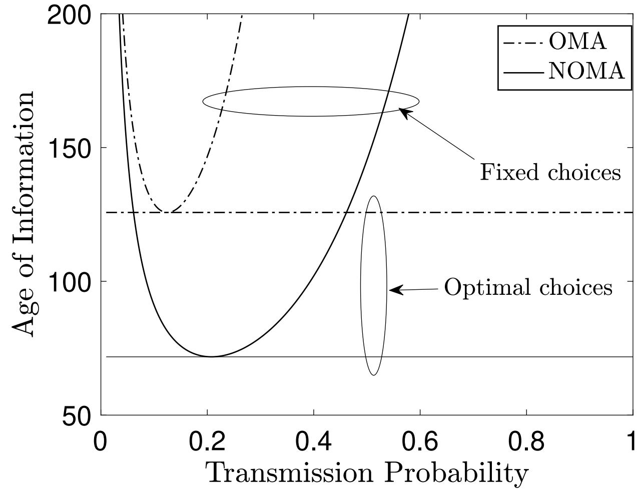


Fig. 6. Illustration of the impact of $\mathbb{P}_{\rm TX}$ on the AoI achieved by the considered grant-free schemes for GAR. $T = 6$, $P = 20$ dB, $R = 0.5$ BPCU, $K = 2$, $N = 1$, and $M = 8$.

networks, are energy constrained and operate in the low SNR regime.

In Fig. 6, the AoI of grant-free transmission is shown as a function of the transmission probability, $\mathbb{P}_{\rm TX}$, and the figure demonstrates that the choices of $\mathbb{P}_{\rm TX}$ are crucial to the AoI performance of grant-free transmission. Furthermore, Fig. 6 shows that NOMA assisted grant-free transmission always yields a smaller AoI than the OMA case, if the value for the transmission probability are used for both of the schemes. As shown in Lemma 2 and the proof for Proposition 1, $\mathbb{P}_{\rm TX} = \frac{1}{M}$ is optimal for OMA, and $\mathbb{P}_{\rm TX} = \frac{\eta}{M}$ is optimal for NOMA in the case with $K = 2$ and $N = 1$. Fig. 6 verifies the optimality of these choices of $\mathbb{P}_{\rm TX}$, since the minimal AoIs achieved by the fixed choices of $\mathbb{P}_{\rm TX}$ match perfectly with the AoIs realized by the optimal choices of $\mathbb{P}_{\rm TX}$.

In Fig. 7, the performance of the considered OMA and NOMA grant-free schemes is compared by using the following AoI ratio, $\frac{\bar{\Delta}^N}{\bar{\Delta}^O}$. For the special case of $N = 1$ and $K = 2$, Proposition 1 predicts that this ratio is 0.5653 for large M, which is confirmed by Fig. 7. If K is fixed, i.e., $K = 2$, an increase of N does not change the ratio significantly, particularly in the case of large M. By introducing more SNR levels, i.e., increasing K, the AoI ratio can be further reduced, which means that the performance gain of NOMA over OMA can be increased by introducing more SNR levels. This is expected since increasing K reduces the likelihood of user collisions and ensures that users can update the base station earlier.

For all the previous simulation results, GAR has been considered, which means that a user's update is generated at the beginning of a time frame, instead of at the beginning of a time slot as

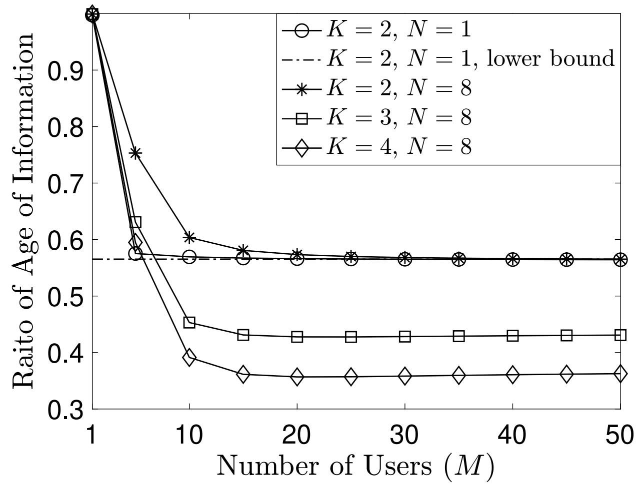


Fig. 7. Impact of the number of users on the ratio between the AoI achieved with the NOMA and OMA, i.e., $\frac{\bar{\Delta}^N}{\bar{\Delta}^O}$, for GAR. $T = 6$, $P = 20$ dB, $R = 0.5$ BPCU, and the optimal choices of $\mathbb{P}_{\mathrm{TX}}$ are used.

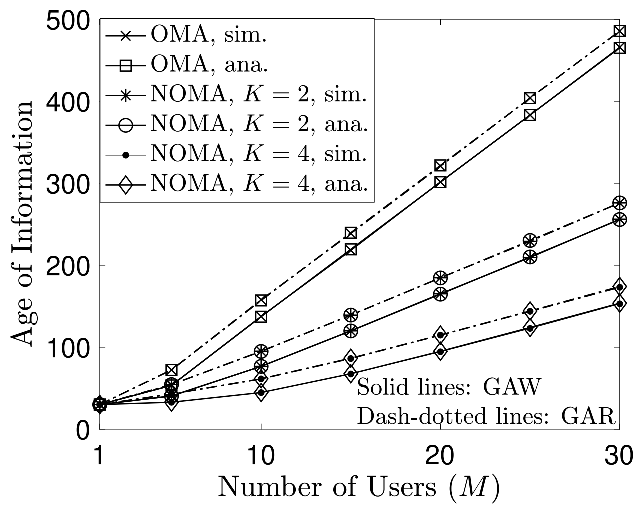


Fig. 8. AoI achieved by the considered grant-free transmission schemes for different data generation models. $N = 8$, $T = 6$, $R = 0.5$ BPCU, and the adaptive choices of $\mathbb{P}_{\mathrm{TX}}$ are used.

for GAW. In Fig. 8, the AoI achieved by the considered grant-free transmission schemes for the two different data generation models is illustrated. As can be seen from the figure, for both GAR and GAW, NOMA assisted grant-free transmission always outperforms the OMA based scheme. In addition, the figure shows that the AoI realized by the considered schemes for GAW is smaller than that for GAR, because, for GAW, each update is generated right before its delivery time, i.e., there is no service delay S_j. We also note that the difference between the AoI for GAW and GAR is not significant, but the use of GAW can cause a higher energy consumption than GAR, since GAW requires a user to re-generate an update for each retransmission.

V. Conclusions

In this paper, the impact of NOMA on the AoI of grant-free transmission has been investigated by applying a particular form of NOMA, namely NOMA-assisted random access. By modelling grant-free transmission as a Markov chain and accounting for SIC, closed-form analytical expressions for the AoI achieved by NOMA assisted grant-free transmission have been obtained, and asymptotic studies have been carried out to demonstrate that the use of the simplest form of NOMA is already sufficient to reduce the AoI of OMA by more than 40%. In addition, the developed analytical results have also been shown useful for optimizing the users' transmission probabilities, $\mathbb{P}_{\mathrm{TX}}$, which is crucial for performance maximization of grant-free transmission.

In this paper, concise and insightful analytical results have been developed for the special case of $N = 1$. An important direction for future research is the development of similar insightful results for the general case of $N \geq 1$. We also note that the principle of NOMA can be implemented in different forms, such as power-domain and code-domain NOMA. Hence, an interesting direction for future research is to investigate whether other forms of NOMA can be applied to further reduce the AoI of grant-free transmission and improve the performance gain of NOMA over OMA.

Appendix A
Proof for Lemma 1

The proof is divided into three parts to evaluate $P_{j,j}$, $P_{j,j+1}$, and $P_{j,j+i}$, $i \geq 2$, respectively. Throughout the proof, the high SNR assumption is made, which ensures that all the SNR levels, P_k, $1 \leq k \leq K$, are feasible for each user, i.e., transmission failures are due to user collisions only.

A. Evaluating $P_{j,j}$

To find the expression for $P_{j,j}$, assume that j users have successfully delivered their updates to the base station. Therefore, each of the remaining $M - j$ users independently makes an attempt to transmit with the probability, $\mathbb{P}_{\mathrm{TX}}$, at a randomly chosen SNR level. $P_{j,j}$ is the probability of the event that none of the $M - j$ users succeeds.

Define $E_{P_k|j}$ as the event that given $M - j$ remaining users, a user successfully updates the base station by using the k-th SNR level, P_k, and no user succeeds at P_i, $i < k$. The reason to include the constraint that no user succeeds at P_i, $i < k$, in the definition of $E_{P_k|j}$ is to ensure that $E_{P_k|j}$ and $E_{P_p|j}$, $k \leq p$, are uncorrelated. For example, the event that U_i succeeds by using P_2 and U_j succeeds by using P_3 belongs to $E_{P_2|j}$ only, and is not included in $E_{P_3|j}$.

Therefore, the probability $P_{j,j}$ can be expressed as follows:

$$\begin{aligned} P_{j,j} =& 1 - \mathbb{P}\left(E_{P_1|j} \cup \cdots \cup E_{P_K|j}\right) \\ =& 1 - \sum_{k=1}^{K} \mathbb{P}\left(E_{P_k|j}\right). \end{aligned} \tag{17}$$

Further define $E_{m|j}$ as the event that among the $M-j$ remaining users, there are m active users which make the transmission attempts, and define $E_{P_k|m}$ as the event that among the m active users, a single user successfully updates the base station by using the k-th SNR level, P_k, and no user chooses P_i, $i<k$, which means

$$P_{j,j} = 1 - \sum_{k=1}^{K} \sum_{m=1}^{M} \binom{M-j}{m} \mathbb{P}\left(E_{m|j}\right) \mathbb{P}\left(E_{P_k|m}\right). \tag{18}$$

By using the transmission attempt probability, $\mathbb{P}_{\mathrm{TX}}$, the probability, $E_{m|j}$, can be obtained as follows:

$$\mathbb{P}(E_{m|j}) = \mathbb{P}_{\mathrm{TX}}^{m}(1-\mathbb{P}_{\mathrm{TX}})^{M-j-m}. \tag{19}$$

Without loss of generality, assume that U_i is one of the m active users which make transmission attempts. The probability for the event that U_i chooses P_k, and no user chooses P_i, $i<k$, is given by

$$\mathbb{P}_K \left(1 - k\mathbb{P}_K\right)^{m-1}, \tag{20}$$

where $(1-k\mathbb{P}_K)$ is the probability of the event that a user which is not U_i cannot choose P_p, $1 \leq p \leq k$. Therefore, $\mathbb{P}\left(E_{P_k|j,m}\right)$ can be approximated as follows:

$$\mathbb{P}\left(E_{P_k|m}\right) \approx m\mathbb{P}_K \left(1 - k\mathbb{P}_K\right)^{m-1}, \tag{21}$$

since each of the m active users can be the successful user with the equal probability, where the high SNR assumption is used, i.e., all SNR levels are feasible to each user and only the errors caused by user collisions are considered. By combining (18), (19), and (21), $P_{j,j}$ can be expressed as follows:

$$\begin{aligned} P_{j,j} \approx & 1 - \sum_{m=1}^{M-j} \binom{M-j}{m} \mathbb{P}_{\mathrm{TX}}^{m}(1-\mathbb{P}_{\mathrm{TX}})^{M-j-m} \\ & \times \sum_{k=1}^{K} m\mathbb{P}_K \left(1 - k\mathbb{P}_K\right)^{m-1}. \end{aligned} \tag{22}$$

B. Evaluating $P_{j,j+1}$

Recall that $P_{j,j+1}$ is also conditioned on the assumption that j users have successfully updated the base station, and $P_{j,j+1}$ is the probability of the event that there is a single successful update from a user which cannot be U_1.

Define $E^1_{P_k|j}$ as the event that given $M-j$ remaining users, a single user, other than U_1, successfully updates the base station by using the k-th SNR level. At high SNR, the following two conclusions can be made regarding $E^1_{P_k|j}$. On the one hand, due to the feature of SIC, $E^1_{P_k|j}$ implies that no user chooses P_i, $i<k$, which can be shown by contradiction. Assume that U_p chooses P_i. If U_p is the only user choosing P_i, U_p becomes an additional successful user, which contradicts the assumption that there is a single successful user. If multiple users choose P_i, a collision occurs and SIC needs to be terminated at P_i, which contradicts the assumption that a successful update happens at P_k. On the other hand, $E^1_{P_k|j}$ does not exclude the event that an SNR level, P_i, $i>k$, is chosen by a user; however, $E^1_{P_k|j}$ does imply that if P_i, $i>k$, is chosen, a collision must happen at this SNR level, otherwise there will be an additional successful user. We also note that $E^1_{P_k|j}$ is different from $E_{P_k|j}$ for the following two reasons. First, $E_{P_k|j}$ does not exclude the event that the successful user is U_1. Second, for $E_{P_k|j}$, it is still possible for a user to succeed at SNR levels, P_i, $i>k$.

By using $E^1_{P_k|j}$, the probability $P_{j,j+1}$ can be expressed as follows:

$$\begin{aligned} P_{j,j+1} =& \mathbb{P}(E^1_{L_1|j} \cup \cdots \cup E^1_{L_K|j}) \\ =& \sum_{k=1}^{K} \mathbb{P}(E^1_{L_k|j}), \end{aligned} \tag{23}$$

where the last step follows by the fact that the $E^1_{P_k|j}$, $1 \leq k \leq K$, are uncorrelated.

Similar to $E_{P_k|m}$, define $E^1_{P_k|m}$ as the event that among the m active users, a single user, other than U_1, successfully updates the base station by using the k-th SNR level. By using $E_{m|j}$ and $E^1_{P_k|m}$, $P_{j,j+1}$ can be expressed as follows:

$$P_{j,j+1} = \sum_{k=1}^{K} \sum_{m=1}^{M} \binom{M-j}{m} \mathbb{P}\left(E_{m|j}\right) \mathbb{P}\left(E_{P^1_k|m}\right). \tag{24}$$

The analysis of $\mathbb{P}\left(E_{P^1_k|m}\right)$ is challenging. First, we assume that $m \geq 2$, i.e., there are more than one active users. For illustrative purposes, assume that U_2 is an active user. Denote by E^+_{2k} the event that U_2 succeeds by using P_k, and $\mathbb{P}\left(E^+_{2k}\right)$ is given by

$$\mathbb{P}\left(E^+_{2k}\right) \approx \mathbb{P}_K(1-k\mathbb{P}_K)^{m-1}, \tag{25}$$

where the high SNR approximation is used, and the reason to have $(1-k\mathbb{P}_K)^{m-1}$ is that the k highest SNR levels are no longer available to the other $m-1$ active users. Because $m \geq 2$, $k \leq K-1$, i.e., U_2 cannot succeed by using P_K, which can be explained by using a simple example with U_2 and U_3 being the active users ($m=2$). As discussed previously, if U_2 chooses P_K, U_3 has to choose P_p, $p>K$, which is not possible.

Furthermore, denote by E_{2k}^{-} the event that U_2 succeeds by using P_k, and there is at least one additional user which succeeds by using P_i, $i \geq k+1$. $\mathbb{P}\left(E_{2k}^{-}\right)$ is given by

$$\mathbb{P}\left(E_{2k}^{-}\right) \approx \mathbb{P}_K \sum_{\kappa=k+1}^{K} (m-1)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-2}, \tag{26}$$

which is obtained in a similar manner as $\mathbb{P}\left(E_{P_k|m}\right)$ in (21).

By using E_{2k}^{+} and E_{2k}^{-}, the probability for the event that among the m active users, only U_2 succeeds by using P_k, denoted by E_{2k}, is given by

$$\begin{aligned}\mathbb{P}\left(E_{2k}\right) =& \mathbb{P}\left(E_{2k}^{+}\right) - \mathbb{P}\left(E_{2k}^{-}\right) \\ \approx& \mathbb{P}_K\left[(1-k\mathbb{P}_K)^{m-1} - \sum_{\kappa=k+1}^{K}(m-1)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-2}\right].\end{aligned} \tag{27}$$

Intuitively, $\mathbb{P}\left(E_{P_k^1|m}\right)$ should be simply $\mathbb{P}\left(E_{P_k^1|m}\right) = (m-1)\mathbb{P}\left(E_{2k}\right)$, i.e., including $(m-1)$ cases corresponding to U_i, $2 \leq i \leq m$. If U_1 is one of the active users, indeed, $\mathbb{P}\left(E_{P_k^1|m}\right) = (m-1)\mathbb{P}\left(E_{2k}\right)$. However, if U_1 is not an active user, $\mathbb{P}\left(E_{P_k^1|m}\right) = m\mathbb{P}\left(E_{2k}\right)$. Therefore, by taking into account the fact that U_1 might not be an active user, $\mathbb{P}\left(E_{P_k^1|m}\right)$ can be expressed as follows:

$$\begin{aligned}\mathbb{P}\left(E_{P_k^1|m}\right) \approx& \left[\frac{m}{M-j}(m-1) + \frac{M-j-m}{M-j}(m)\right]\mathbb{P}\left(E_{2k}\right) \\ =& \frac{M-j-1}{M-j} m\mathbb{P}_K\left[(1-k\mathbb{P}_K)^{m-1} - \sum_{\kappa=k+1}^{K}(m-1)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-2}\right],\end{aligned} \tag{28}$$

for the case $m \geq 2$.

For the case $m=1$, i.e., there is a single active user, $\mathbb{P}\left(E_{P_k^1|m}\right)$ is simply zero if this user is U_1. Otherwise, $\mathbb{P}\left(E_{P_k^1|m}\right) = \mathbb{P}_K$, i.e., P_k is selected by the active user. Therefore, for $m=1$, $\mathbb{P}\left(E_{P_k^1|m}\right)$ is given by

$$\mathbb{P}\left(E_{P_k^1|m}\right) \approx \frac{M-j-1}{M-j}\mathbb{P}_K. \tag{29}$$

By combining (24), (28) and (29), the probability, $P_{j,j+1}$, can be obtained as shown in the lemma.

C. Evaluating $P_{j,j+i}$, $i \geq 2$

We note that for $P_{j,j+1}$, there is a single successful update, and $P_{j,j+1}$ has been analyzed by first specifying which SNR level, i.e., P_k, is used for this successful update. The same method could be applied to analyze $P_{j,j+i}$, $i \geq 2$; however, the resulting expression can be extremely complicated if the number of the used SNR levels is large.

Instead, the feature of SIC can be used to simplify the analysis of $P_{j,j+i}$. Consider the following example with $i = 3$ and $K = 8$. Assume that the following three SNR levels, P_2, P_4, and P_5, are used by the successful users. The key observation for simplifying the performance analysis is that those SNR levels prior to P_2 and between the chosen ones are not selected by any users, e.g., no user chooses P_1 and P_3. This observation can be explained by taking P_3 as an example. If this SNR level has been selected, a collision must happen at this level, otherwise there should be an additional successful user. However, if a collision does happen at P_3, SIC needs to be terminated in the third SIC stage, and hence no successful update can happen at P_4, which contradicts to the assumption that P_4 is used by a successful user. As a result, no one chooses P_3.

By using this observation, we note that only two SNR levels are significant to the analysis of $P_{j,j+i}$, namely the highest and the lowest SNR levels, which are denoted by P_{k_1} and P_{k_2}, respectively. For the aforementioned example, $P_{k_1} = P_2$ and $P_{k_2} = P_5$. Define $E^i_{P_k|j}$ as the event that given $M - j$ remaining users, i users which are not U_1 successfully update the base station, where P_{k_1} and P_{k_2} are the highest and lowest used SNR levels. By using $E^i_{P_k|j}$, the probability $P_{j,j+i}$ can be expressed as follows:

$$
\begin{aligned}
P_{j,j+i} =& \sum_{k_1=1}^{K-i+1} \sum_{k_2=k_1+i-1}^{K} \mathbb{P}(E^i_{L_k|j}), \\
=& \sum_{k_1=1}^{K-i+1} \sum_{k_2=k_1+i-1}^{K} \sum_{m=1}^{M} \binom{M-j}{m} \mathbb{P}\left(E_{m|j}\right) \mathbb{P}\left(E_{P^i_k|m}\right),
\end{aligned}
\tag{30}
$$

where $E_{P^i_k|m}$ is defined similar to $E^i_{P_k|j}$ with the assumption that there are m active users.

In order to find $\mathbb{P}\left(E_{P^i_k|m}\right)$, again, we first focus on the case $m \geq i + 1$, i.e., there are more than i active user. Define E_{ik} as the probability for the particular event that among the m active users, U_2 succeeds by using P_{k_1}, U_j succeeds by using P_{k_1+j-2}, $3 \leq j \leq i$, and U_{i+1} succeeds by using P_{k_2}. Similar to $\mathbb{P}\left(E_{2k}\right)$, $\mathbb{P}\left(E_{ik}\right)$ can be approximated at high SNR as follows:

$$\mathbb{P}\left(E_{ik}\right) \approx \mathbb{P}_K^i \left[(1-k_2\mathbb{P}_K)^{m-i} - \sum_{\kappa=k_2+1}^{K} (m-i)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-i-1}\right], \tag{31}$$

where the term, $(1-k_2\mathbb{P}_K)^{m-i}$, is due to the fact that the remaining $m-i$ active users can choose P_i, $i > k_2$, only, and the second term in the bracket is obtained similar to $\mathbb{P}\left(E_{P_k|m}\right)$ in (21).

Following steps similar to those to obtain $\mathbb{P}\left(E_{P_k^1|m}\right)$ in (28), $\mathbb{P}\left(E_{P_k^i|m}\right)$ can be obtained from $\mathbb{P}\left(E_{ik}\right)$ as follows:

$$\begin{aligned}\mathbb{P}\left(E_{P_k^i|m}\right) \approx & \mathbb{P}\left(E_{ik}\right)\binom{k_2-k_1-1}{i-2}\underbrace{\left[\frac{m}{M-j}(m-1)+\frac{M-j-m}{M-j}m\right]}_{\text{repetitions in the highest SNR level}} \times \cdots \times \\ & \underbrace{\left[\frac{m-i+1}{M-j-i+1}(m-i)+\frac{M-j-m}{M-j-i+1}(m-i+1)\right]}_{\text{repetitions in the lowest SNR level}} \\ = & \mathbb{P}\left(E_{ik}\right)\binom{k_2-k_1-1}{i-2}\frac{M-j-i}{M-j}m\cdots(m-i+1),\end{aligned} \tag{32}$$

for the case $m \geq i+1$, where $\binom{k_2-k_1-1}{i-2}$ is the number of the possible choices for the $(i-2)$ SNR levels which are between P_{k_1} and P_{k_2}.

The special case $m = i$ means that there are m active users and each of the active users is a successful user. Therefore, the probability in (31), $\mathbb{P}\left(E_{ik}\right)$, is simply $\mathbb{P}_K^i$, and hence for the case $m = i$, $\mathbb{P}\left(E_{P_k^i|m}\right)$ can be expressed as follows:

$$\mathbb{P}\left(E_{P_k^i|m}\right) \approx \mathbb{P}_K^i \binom{k_2-k_1-1}{i-2}\frac{M-j-i}{M-j} i \cdots 1. \tag{33}$$

By combining (32) and (33), the expression of $P_{j,j+i}$ can be obtained as shown in the lemma. Therefore, the proof for the lemma is complete.

Appendix B
Proof for Lemma 2

The proof comprises first simplifying the state transition probabilities, then developing an asymptotic expression of the AoI, and finally finding the optimal choice of $\mathbb{P}_{\text{TX}}$.

A. Simplifying the State Transition Probabilities

For the case of $N = 1$ and $K = 2$, only the following transition probabilities, $P_{0,0}$, $P_{0,1}$, $P_{0,2}$, need to be focused on.

In particular, the expression of $P_{0,0}$ can be simplified as follows:

$$\begin{aligned} P_{0,0} \approx & 1 - \sum_{m=1}^{M} \binom{M}{m} \mathbb{P}_{\mathrm{TX}}^{m} (1-\mathbb{P}_{\mathrm{TX}})^{M-m} \sum_{k=1}^{K} m\mathbb{P}_K \left(1-k\mathbb{P}_K\right)^{m-1} \\ \overset{(1)}{=} & 1 - \sum_{m=1}^{M} \binom{M}{m} \mathbb{P}_{\mathrm{TX}}^{m} (1-\mathbb{P}_{\mathrm{TX}})^{M-m} m\mathbb{P}_K \left(1-\mathbb{P}_K\right)^{m-1} - M\mathbb{P}_{\mathrm{TX}}(1-\mathbb{P}_{\mathrm{TX}})^{M-1}\mathbb{P}_K \\ \overset{(b)}{=} & 1 - \sum_{m=1}^{M} \binom{M}{m} \mathbb{P}_{\mathrm{TX}}^{m} (1-\mathbb{P}_{\mathrm{TX}})^{M-m} m\mathbb{P}_K^{m} - M\mathbb{P}_{\mathrm{TX}}(1-\mathbb{P}_{\mathrm{TX}})^{M-1}\mathbb{P}_K, \end{aligned}$$

where step (a) follows by the fact that $(1 - K\mathbb{P}_K)^{m-1} \neq 0$ only if $m = 1$, and step (b) follows by $\mathbb{P}_K = 1 - \mathbb{P}_K$ for $K = 2$. By using the properties of the binomial coefficients, $P_{0,0}$ can be simplified as follows:

$$\begin{aligned} P_{0,0} \approx & 1 - M\mathbb{P}_K\mathbb{P}_{\mathrm{TX}} \sum_{i=0}^{M-1} \binom{M-1}{i} (\mathbb{P}_K\mathbb{P}_{\mathrm{TX}})^{i} \\ & \times (1-\mathbb{P}_{\mathrm{TX}})^{M-1-i} - M\mathbb{P}_{\mathrm{TX}}(1-\mathbb{P}_{\mathrm{TX}})^{M-1}\mathbb{P}_K \\ = & 1 - M\mathbb{P}_K\mathbb{P}_{\mathrm{TX}}(\mathbb{P}_K\mathbb{P}_{\mathrm{TX}} + 1 - \mathbb{P}_{\mathrm{TX}})^{M-1} - M\mathbb{P}_{\mathrm{TX}}(1-\mathbb{P}_{\mathrm{TX}})^{M-1}\mathbb{P}_K. \end{aligned} \tag{34}$$

Similarly, for the case of $K = 2$, $P_{0,1}$ can be first written as follows:

$$\begin{aligned} P_{0,1} \approx & M\mathbb{P}_{\mathrm{TX}}(1-\mathbb{P}_{\mathrm{TX}})^{M-1}\frac{M-1}{M}K\mathbb{P}_K \\ & + \sum_{m=2}^{M} \binom{M}{m} \mathbb{P}_{\mathrm{TX}}^{m}(1-\mathbb{P}_{\mathrm{TX}})^{M-m} \sum_{k=1}^{K-1} \frac{M-1}{M} m\mathbb{P}_K \\ & \left[(1-k\mathbb{P}_K)^{m-1} - \sum_{\kappa=k+1}^{K} (m-1)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-2} \right]. \end{aligned} \tag{35}$$

Define $\tau(m) = \sum_{k=1}^{K-1} \frac{M-1}{M} m\mathbb{P}_K[(1-k\mathbb{P}_K)^{m-1} - \sum_{\kappa=k+1}^{K}(m-1)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-2}]$. Note that in the expression of $\tau(m)$, $k \geq 1$, and hence $\kappa \geq 2$. For the case of $K = 2$, $1 - 2\mathbb{P}_K = 0$. By using this observation, $\tau(m)$ can be simplified as follows:

$$\tau(m) = \frac{m(M-1)}{M}\mathbb{P}_K(1-\mathbb{P}_K)^{m-1}, \tag{36}$$

for $m > 2$, and

$$\tau(2) = \left(\frac{m}{M}(m-1) + \frac{M-m}{M}m\right)\mathbb{P}_K\left[(1-\mathbb{P}_K) - \mathbb{P}_K\right] = 0. \tag{37}$$

By using the simplified expression of $\tau(m)$, $P_{0,1}$ can be simplified as follows:

$$\begin{aligned} P_{0,1} \approx & M\mathbb{P}_{\text{TX}}(1-\mathbb{P}_{\text{TX}})^{M-1}\frac{M-1}{M}K\mathbb{P}_K \\ & + \sum_{m=3}^{M}\binom{M}{m}\mathbb{P}_{\text{TX}}^{m}(1-\mathbb{P}_{\text{TX}})^{M-m}\frac{m(M-1)}{M}\mathbb{P}_K(1-\mathbb{P}_K)^{m-1} \\ = & (M-1)\mathbb{P}_{\text{TX}}(1-\mathbb{P}_{\text{TX}})^{M-1} \\ & + (M-1)\sum_{m=3}^{M}\binom{M-1}{m-1}\mathbb{P}_{\text{TX}}^{m}(1-\mathbb{P}_{\text{TX}})^{M-m}\mathbb{P}_K^{m}, \end{aligned} \tag{38}$$

where the last step follows by the fact that $1-\mathbb{P}_K=\mathbb{P}_K$.

Finally, $P_{0,2}$ can rewritten as follows:

$$\begin{aligned} P_{0,2} \approx & \binom{M}{2}\mathbb{P}_{\text{TX}}^{i}(1-\mathbb{P}_{\text{TX}})^{M-2}\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K}\frac{M-i}{M}\mathbb{P}_K^2\binom{k_2-k_1-1}{0}\prod_{p=0}^{1}(2-p) \\ & + \sum_{m=3}^{M}\binom{M}{m}\mathbb{P}_{\text{TX}}^{m}(1-\mathbb{P}_{\text{TX}})^{M-m}\sum_{k_1=1}^{K-2}\sum_{k_2=k_1+1}^{K-1}\frac{M-i}{M}\mathbb{P}_K^2\binom{k_2-k_1-1}{0}\prod_{p=0}^{1}(m-p) \\ & \times\left[(1-k_2\mathbb{P}_K)^{m-2}-\sum_{\kappa=k_2+1}^{K}(m-2)\mathbb{P}_K(1-\kappa\mathbb{P}_K)^{m-3}\right] \\ & \overset{(a)}{=}\binom{M}{2}\mathbb{P}_{\text{TX}}^{2}(1-\mathbb{P}_{\text{TX}})^{M-2}2\frac{M-2}{M}\mathbb{P}_K^2=\frac{(M-1)!}{(M-3)!}\mathbb{P}_{\text{TX}}^{2}(1-\mathbb{P}_{\text{TX}})^{M-2}\mathbb{P}_K^2, \end{aligned} \tag{39}$$

where step (a) follows by employing the properties of the binomial coefficients.

B. Asymptotic Studies of AoI

By using the above transition probabilities, the probability for the event that U_1 cannot complete an update within one frame, P_{fail}, can be simplified as follows:

$$\begin{aligned} P_{\text{fail}} \approx & \mathbb{P}(Z>N)=\mathbf{s}_0^T\mathbf{P}_M^N\mathbf{1}=\sum_{j=0}^{2}P_{0,j} \\ = & 1-M(x+1-\mathbb{P}_{\text{TX}})^{M-1}x-M(1-\mathbb{P}_{\text{TX}})^{M-1}x+2(M-1)(1-\mathbb{P}_{\text{TX}})^{M-1}x \\ & \underbrace{+(M-1)\sum_{i=2}^{M}\binom{M-1}{i}(1-\mathbb{P}_{\text{TX}})^{M-i-1}x^{i+1}}_{\tau_0}+\frac{(M-1)!}{(M-3)!}(1-\mathbb{P}_{\text{TX}})^{M-2}x^2, \end{aligned} \tag{40}$$

where $x = \mathbb{P}_{\text{TX}}\mathbb{P}_K$. By using the properties of binomial coefficients, τ_0 can be rewritten as follows:

$$\begin{aligned}\tau_0 =& x(M-1)\sum_{i=2}^{M}\binom{M-1}{i}(1-\mathbb{P}_{\text{TX}})^{M-i-1}x^i \\ =&(1-\mathbb{P}_{\text{TX}}+x)^{M-1}(M-1)x-(1-\mathbb{P}_{\text{TX}})^{M-1}(M-1)x \\ &-(M-1)(1-\mathbb{P}_{\text{TX}})^{M-2}(M-1)x^2.\end{aligned} \tag{41}$$

By using the simplified expression of τ_0, P_{fail} can be rewritten as follows:

$$\begin{aligned}P_{\text{fail}} \approx& 1-M(x+1-\mathbb{P}_{\text{TX}})^{M-1}x-M(1-\mathbb{P}_{\text{TX}})^{M-1}x \\ &+2(M-1)(1-\mathbb{P}_{\text{TX}})^{M-1}x \\ &+(1-\mathbb{P}_{\text{TX}}+x)^{M-1}(M-1)x-(1-\mathbb{P}_{\text{TX}})^{M-1}(M-1)x \\ &-(M-1)(1-\mathbb{P}_{\text{TX}})^{M-2}(M-1)x^2+(M-1)(M-2)(1-\mathbb{P}_{\text{TX}})^{M-2}x^2 \\ =&1-(\mathbb{P}_{\text{TX}}\mathbb{P}_K+1-\mathbb{P}_{\text{TX}})^{M-1}\mathbb{P}_{\text{TX}}\mathbb{P}_K \\ &-(1-\mathbb{P}_{\text{TX}}+(M-1)\mathbb{P}_{\text{TX}}\mathbb{P}_K)(1-\mathbb{P}_{\text{TX}})^{M-2}\mathbb{P}_{\text{TX}}\mathbb{P}_K.\end{aligned} \tag{42}$$

The simplified expression of P_{fail} can be used to facilitate the asymptotic studies of the AoI. For the case of $N=1$, the pmf of the access delay can be obtained from P_{fail} as follows:

$$\mathbb{P}(S_j=T)=\frac{\mathbf{s}_0^T\mathbf{P}_M^0\mathbf{p}}{1-\mathbf{s}_0^T\mathbf{P}_M\mathbf{1}}=\frac{p_{0,M}}{p_{0,M}}=1, \tag{43}$$

where $p_{0,M}$ is the first element of $\mathbf{p}$. As a result, $\mathcal{E}\{S_j\}$ and $\mathcal{E}\{S_j^2\}$ can be obtained as follows:

$$\mathcal{E}\{S_j\}=T\sum_{n=1}^{1}n\frac{\mathbf{s}_0^T\mathbf{P}_M^{n-1}\mathbf{p}}{1-\mathbf{s}_0^T\mathbf{P}_M^N\mathbf{1}}=T, \tag{44}$$

and

$$\mathcal{E}\{S_j^2\}=T^2\sum_{n=1}^{1}n^2\frac{\mathbf{s}_0^T\mathbf{P}_M^{n-1}\mathbf{p}}{1-\mathbf{s}_0^T\mathbf{P}_M^N\mathbf{1}}=T^2. \tag{45}$$

Furthermore, recall that the inter-departure time Y_j can be rewritten as follows:

$$Y_j=(NT-S_{j-1})+(X_j-1)NT+S_j. \tag{46}$$

Therefore, for the case of $N=1$, the expectation of Y_j can be simplified as follows:

$$\begin{aligned}\mathcal{E}\{Y_j\} =& \mathcal{E}\{(NT-S_{j-1})+(X_j-1)NT+S_j\} \\ =& TN\mathcal{E}\{X_j\}=TN\frac{1}{1-P_{\text{fail}}},\end{aligned} \tag{47}$$

which is obtained by using the fact that X_j follows the geometric distribution. Similarly, the expectation of Y_j^2 can be simplified as follows:

$$\begin{aligned}\mathcal{E}\{Y_j^2\} =& N^2T^2\mathcal{E}\left\{X_j^2\right\} + 2\mathcal{E}\left\{S_j^2\right\} - 2\mathcal{E}\left\{S_j\right\}^2 \\ =& N^2T^2\frac{1+P_{\text{fail}}}{\left(1-P_{\text{fail}}\right)^2} + 2T^2 - 2T^2.\end{aligned} \tag{48}$$

Finally, for the case of $N=1$, the averaged AoI can be expressed as follows:

$$\begin{aligned}\bar{\Delta}^N =& \frac{\mathcal{E}\{S_j\}\mathcal{E}\{Y_j\} - \mathcal{E}\{S_j^2\} + \mathcal{E}\{S_j\}^2}{\mathcal{E}\{Y_j\}} + \frac{\mathcal{E}\{Y_j^2\}}{2\mathcal{E}\{Y_j\}} \\ =& \frac{T\mathcal{E}\{Y_j\} - T^2 + T^2}{\mathcal{E}\{Y_j\}} + \frac{\mathcal{E}\{Y_j^2\}}{2\mathcal{E}\{Y_j\}} \\ =& T + \frac{N^2T^2\frac{1+P_{\text{fail}}}{(1-P_{\text{fail}})^2}}{2TN\frac{1}{1-P_{\text{fail}}}} \approx T + \frac{NT(2-f(\mathbb{P}_{\text{TX}}))}{2f(\mathbb{P}_{\text{TX}})},\end{aligned} \tag{49}$$

where $f(\mathbb{P}_{\text{TX}}) = (\mathbb{P}_{\text{TX}}\mathbb{P}_K + 1 - \mathbb{P}_{\text{TX}})^{M-1}\mathbb{P}_{\text{TX}}\mathbb{P}_K + (1 - \mathbb{P}_{\text{TX}} + (M-1)\mathbb{P}_{\text{TX}}\mathbb{P}_K)(1 - \mathbb{P}_{\text{TX}})^{M-2}\mathbb{P}_{\text{TX}}\mathbb{P}_K$.

C. Finding the Optimal Choice of $\mathbb{P}_{\text{TX}}$

The considered AoI minimization problem can be expressed as follows: $\min\limits_{0\leq\mathbb{P}_{\text{TX}}\leq 1} \bar{\Delta}^N$. It is challenging to show whether $\bar{\Delta}^N$ is a convex function of $\mathbb{P}_{\text{TX}}$, given the complex expression of $\bar{\Delta}^N$. However, it can be shown that $\bar{\Delta}^N$ first decreases and then increases as $\mathbb{P}_{\text{TX}}$ grows, as shown in the following.

The first order derivative of $\bar{\Delta}^N$ with respect to $\mathbb{P}_{\text{TX}}$ is given by

$$\begin{aligned}\left(\bar{\Delta}^N\right)' \approx& \frac{NT}{2}\left(\frac{-f'(\mathbb{P}_{\text{TX}})}{f(\mathbb{P}_{\text{TX}})} - \frac{2f'(\mathbb{P}_{\text{TX}})' - f(\mathbb{P}_{\text{TX}})f'(\mathbb{P}_{\text{TX}})}{y^2}\right) \\ =& -\frac{NT}{2}\frac{2f'(\mathbb{P}_{\text{TX}})}{f^2(\mathbb{P}_{\text{TX}})},\end{aligned} \tag{50}$$

which shows that the monotonicity of $\bar{\Delta}^N$ is decided by the sign of $f'(\mathbb{P}_{\text{TX}})$ only. In the following, we will first show that $f'(\mathbb{P}_{\text{TX}}) = 0$ has a single root for $0 \leq \mathbb{P}_{\text{TX}} \leq 1$.

With some straightforward algebraic manipulations, $f'(\mathbb{P}_{\text{TX}})$ can be expressed as follows:

$$\begin{aligned}f'(\mathbb{P}_{\text{TX}}) =& \left[1 - M\frac{\mathbb{P}_{\text{TX}}}{2}\right]\left(1 - \frac{\mathbb{P}_{\text{TX}}}{2}\right)^{M-2} \\ &+ \left[1 - 2\mathbb{P}_{\text{TX}} - M(M-3)\frac{\mathbb{P}_{\text{TX}}^2}{2}\right](1-\mathbb{P}_{\text{TX}})^{M-3}.\end{aligned} \tag{51}$$

Because $0 \leq \mathbb{P}_{\text{TX}} \leq 1$, both $\left(1 - \frac{\mathbb{P}_{\text{TX}}}{2}\right)$ and $(1 - \mathbb{P}_{\text{TX}})$ are positive.

On the one hand, if $\mathbb{P}_{\text{TX}} \geq \frac{2}{M}$, $1 - M\frac{\mathbb{P}_{\text{TX}}}{2} \leq 0$, otherwise $1 - M\frac{\mathbb{P}_{\text{TX}}}{2} \geq 0$. On the other hand, there are two roots for $1 - 2\mathbb{P}_{\text{TX}} - M(M-3)\frac{\mathbb{P}_{\text{TX}}^2}{2} = 0$, namely $\frac{\sqrt{4+2M(M-1)}-4}{M(M-3)}$ and $\frac{-\sqrt{4+2M(M-1)}-4}{M(M-3)}$. When $M \to \infty$, the two roots can be approximated as follows:

$$\frac{\pm\sqrt{4+2M(M-1)}-4}{M(M-3)} \approx \pm\frac{\sqrt{2}}{M}. \tag{52}$$

Therefore, $1-2\mathbb{P}_{\text{TX}}-M(M-3)\frac{\mathbb{P}_{\text{TX}}^2}{2} \leq 0$ if $\mathbb{P}_{\text{TX}} \geq \frac{\sqrt{2}}{M}$, otherwise $1-2\mathbb{P}_{\text{TX}}-M(M-3)\frac{\mathbb{P}_{\text{TX}}^2}{2} > 0$. Denote the root(s) of $f'(\mathbb{P}_{\text{TX}}) = 0$ by $\mathbb{P}_{\text{TX}}^*$, which is bounded as follows:

$$\frac{\sqrt{2}}{M} \leq \mathbb{P}_{\text{TX}}^* \leq \frac{2}{M}. \tag{53}$$

A key observation from (53) is that the upper and lower bounds on $\mathbb{P}_{\text{TX}}^*$ are of the same order of $\frac{1}{M}$. Therefore, $\mathbb{P}_{\text{TX}}^*$ can be expressed as $\mathbb{P}_{\text{TX}}^* = \frac{\eta}{M}$, where $1 \leq \eta \leq 2$. By using this expression for $\mathbb{P}_{\text{TX}}^*$, $f'(\mathbb{P}_{\text{TX}}^*) = 0$ can be expressed as follows:

$$\begin{aligned} 0 = &\left(1-\frac{\eta}{2}\right)\left(1-\frac{\eta}{2M}\right)^{M-2} \\ &+ \left(1-2\frac{\eta}{M}-(M-3)\frac{\eta^2}{2M}\right)\left(1-\frac{\eta}{M}\right)^{M-3}. \end{aligned} \tag{54}$$

In order to find an explicit expression of $\mathbb{P}_{\text{TX}}^*$, we note that $f_x(x) = \left(1-\frac{a}{x}\right)^x$ can be approximated at $x \to \infty$ as follows:

$$\begin{aligned} \ln f_x(x) =& \lim_{x\to\infty} x \ln\left(1-\frac{a}{x}\right) = \lim_{z\to 0}\frac{\ln(1-az)}{z} \\ =& -\lim_{z\to 0}\frac{a}{1-az} = -a, \end{aligned} \tag{55}$$

which implies $\left(1-\frac{a}{M}\right)^M = e^{-a}$ for large M.

Therefore, for $M \to \infty$, (54) can be approximated as follows:

$$0 = \left(1-\frac{\eta}{2}\right)e^{-\frac{\eta}{2}} + \left(1-\frac{\eta^2}{2}\right)e^{-\eta}, \tag{56}$$

where the value of η can be straightforwardly obtained by applying off-the-shelf root solvers. It is important to point out that (56) has a single root, which means that $\mathbb{P}_{\text{TX}}^* = \frac{\eta}{M}$ is the single root of $f'(\mathbb{P}_{\text{TX}}) = 0$. Therefore, $\mathbb{P}_{\text{TX}}^* = \frac{\eta}{M}$ is the optimal choice of $\mathbb{P}_{\text{TX}}$ to minimize $\bar{\Delta}^N$, and the proof is complete.

Appendix C
Proof for Proposition 1

Based on Lemma 2, the optimal choice for the transmission probability is given by $\mathbb{P}_{\rm TX}^{*}=\frac{\eta}{M}$. By using this optimal choice of $\mathbb{P}_{\rm TX}$, $P_{\rm fail}$ can be expressed as follows:

$$\begin{aligned} P_{\rm fail} \approx & 1-\left(1-\frac{\eta}{2M}\right)^{M-1}\frac{\eta}{2M} \\ & -\left(1-\frac{\eta}{M}+(M-1)\frac{\eta}{2M}\right)\left(1-\frac{\eta}{M}\right)^{M-2}\frac{\eta}{2M}. \end{aligned} \tag{57}$$

Again applying the following limit: $\lim\limits_{M\to\infty}\left(1-\frac{a}{M}\right)^{M}=e^{-a}$, $P_{\rm fail}$ can be approximated for large M as follows:

$$\begin{aligned} P_{\rm fail} \approx & 1-e^{-\frac{\eta}{2}}\frac{\eta}{2M}-\left(1-\frac{\eta}{M}+(M-1)\frac{\eta}{2M}\right)e^{-\eta}\frac{\eta}{2M} \\ \approx & 1-e^{-\frac{\eta}{2}}\frac{\eta}{2M}-\left(1+\frac{\eta}{2}\right)e^{-\eta}\frac{\eta}{2M}. \end{aligned} \tag{58}$$

By using this approximation of $P_{\rm fail}$, the AoI achieved by NOMA can be approximated as follows:

$$\begin{aligned} \bar{\Delta}^{N} \approx & T+\frac{NT}{2}\frac{2-e^{-\frac{\eta}{2}}\frac{\eta}{2M}-\left(1+\frac{\eta}{2}\right)e^{-\eta}\frac{\eta}{2M}}{e^{-\frac{\eta}{2}}\frac{\eta}{2M}+\left(1+\frac{\eta}{2}\right)e^{-\eta}\frac{\eta}{2M}} \\ \approx & \eta+\frac{NT}{2}\left(\frac{2}{e^{-\frac{\eta}{2}}\frac{\eta}{2M}+\left(1+\frac{\eta}{2}\right)e^{-\eta}\frac{\eta}{2M}}-1\right) \\ \approx & NT\frac{2Me^{\eta}}{\eta\left(e^{\frac{\eta}{2}}+1+\frac{\eta}{2}\right)}. \end{aligned} \tag{59}$$

In order to find an explicit expression for the AoI achieved by OMA for the case of $N=1$, the transition probabilities can be expressed as follows:

$$P_{0,0}=1-M\mathbb{P}_{\rm TX}\left(1-\mathbb{P}_{\rm TX}\right)^{M-1}, \tag{60}$$

and

$$P_{0,1}=(M-1)\mathbb{P}_{\rm TX}\left(1-\mathbb{P}_{\rm TX}\right)^{M-1}. \tag{61}$$

Therefore, for OMA, the failure probability, $P_{\rm fail}$, is given by

$$P_{\rm fail}=P_{0,0}+P_{0,1}=1-\mathbb{P}_{\rm TX}\left(1-\mathbb{P}_{\rm TX}\right)^{M-1}. \tag{62}$$

By using the expression for P_{fail}, the AoI achieved by OMA can be expressed as follows:

$$\begin{aligned}\bar{\Delta}^{O} =&1+\frac{NT(1+P_{\text{fail}})}{2(1-P_{\text{fail}})} \\ =&1+\frac{NT\left(2-\mathbb{P}_{\text{TX}}\left(1-\mathbb{P}_{\text{TX}}\right)^{M-1}\right)}{2\mathbb{P}_{\text{TX}}\left(1-\mathbb{P}_{\text{TX}}\right)^{M-1}}.\end{aligned} \tag{63}$$

Define $f_O(\mathbb{P}_{\text{TX}}) = \mathbb{P}_{\text{TX}}\left(1-\mathbb{P}_{\text{TX}}\right)^{M-1}$. Similar to the NOMA case, it is straightforward to show that the optimal choice of $\mathbb{P}_{\text{TX}}$ for OMA is the root of $f_O'(\mathbb{P}_{\text{TX}})=0$. Note that $f_O'(\mathbb{P}_{\text{TX}})$ can be expressed as follows:

$$f_O'(\mathbb{P}_{\text{TX}}) =(1-M\mathbb{P}_{\text{TX}})(1-\mathbb{P}_{\text{TX}})^{M-2}, \tag{64}$$

which is the reason why $\mathbb{P}_{\text{TX}}^{*}=\frac{1}{M}$ is optimal for the OMA case. By using $\mathbb{P}_{\text{TX}}^{*}=\frac{1}{M}$, P_{fail} can be expressed as follows:

$$P_{\text{fail}} =1-\frac{1}{M}\left(1-\frac{1}{M}\right)^{M-1}, \tag{65}$$

and hence the AoI achieved by OMA is given by

$$\begin{aligned}\bar{\Delta}^{O} =&1+\frac{NT}{2}\frac{2-\mathbb{P}_{\text{TX}}\left(1-\mathbb{P}_{\text{TX}}\right)^{M-1}}{\mathbb{P}_{\text{TX}}\left(1-\mathbb{P}_{\text{TX}}\right)^{M-1}} \\ =&1+\frac{NT}{2}\left[\frac{2M}{\left(1-\frac{1}{M}\right)^{M-1}}-1\right].\end{aligned} \tag{66}$$

For $M\rightarrow\infty$, the AoI achieved by OMA can be approximated as follows:

$$\bar{\Delta}^{O} \approx\frac{NT}{2}\frac{2M}{\left(1-\frac{1}{M}\right)^{M-1}}\approx NTMe. \tag{67}$$

Therefore, for $M\rightarrow\infty$, the ratio between the AoI realized by NOMA and OMA is given by

$$\frac{\bar{\Delta}^{N}}{\bar{\Delta}^{O}}\approx\frac{NT\frac{2Me^{\eta}}{\eta\left(e^{\frac{\eta}{2}}+1+\frac{\eta}{2}\right)}}{NTMe}=\frac{2e^{\eta-1}}{\eta\left(e^{\frac{\eta}{2}}+1+\frac{\eta}{2}\right)}, \tag{68}$$

which completes the proof.

References

[1] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the standardization of grant-free operation and the associated NOMA strategies in 3GPP," *IEEE Commun. Standards Mag.*, vol. 3, no. 4, pp. 60–66, Dec. 2019.

[2] X. You, C. Wang, J. Huang *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 110301, pp. 1–74, Feb. 2021.

[3] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Random access for M2M communications with QoS guarantees," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2889–2903, Jul. 2017.

[4] J. Choi, J. Ding, N.-P. Le, and Z. Ding, "Grant-free random access in machine-type communication: Approaches and challenges," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 151–158, Feb. 2022.

[5] L. Liu and W. Yu, "Massive connectivity with massive MIMO - Part I: Device activity detection and channel estimation," *IEEE Trans. on Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[6] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sept. 2018.

[7] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Transa. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.

[8] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A novel analytical framework for massive grant-free NOMA," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2436–2449, Mar. 2019.

[9] Y. Ma, G. Ma, N. Wang, Z. Zhong, and B. Ai, "OTFS-TSMA for massive internet of things in high-speed railway," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 519–531, Jan. 2022.

[10] J. Ding, D. Qu, and J. Choi, "Analysis of non-orthogonal sequences for grant-free RA with massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 150–160, Jan. 2020.

[11] M. Elbayoumi, M. Kamel, W. Hamouda, and A. Youssef, "NOMA-assisted machine-type communications in UDN: State-of-the-art and challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1276–1304, 2020.

[12] X. Zhang, P. Fan, J. Liu, and L. Hao, "Bayesian learning-based multiuser detection for grant-free NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6317–6328, Aug. 2022.

[13] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, Jun. 2019.

[14] Q. Hu, J. Jiao, Y. Wang, S. Wu, R. Lu, and Q. Zhang, "Multi-type services coexistence in uplink NOMA for dual-layer LEO satellite constellation," *IEEE Internet of Things Journal*, to appear in 2022.

[15] K. Cao, H. Ding, B. Wang, L. Lv, J. Tian, Q. Wei, and F. Gong, "Enhancing physical layer security for IoT with non-orthogonal multiple access assisted semi-grant-free transmission," *IEEE Internet of Things Journal*, to appear in 2022.

[16] J. Chen, L. Guo, J. Jia, J. Shang, and X. Wang, "Resource allocation for irs assisted SGF NOMA transmission: A MADRL approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1302–1316, Apr. 2022.

[17] Y. Liu, M. Zhao, L. Xiao, and S. Zhou, "Pilot domain NOMA for grant-free massive random access in massive MIMO marine communication system," *China Commun.*, vol. 17, no. 6, pp. 131–144, Jun. 2020.

[18] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inform. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

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

[20] A. A. Al-Habob, O. A. Dobre, and H. V. Poor, "Age- and correlation-aware information gathering," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 273–277, Feb. 2022.

[21] A. Maatouk, M. Assaad, and A. Ephremides, "Minimizing the age of information: NOMA or OMA?" in *Proc. IEEE INFOCOM WKSHPS)*, Paris, France, 2019.

[22] Z. Ding, R. Schober, and H. V. Poor, "Age of information: Can CR-NOMA help?" *IEEE Trans. Wireless Commun.*, Available on-line at arXiv:2209.09562, 2022.

[23] H. Pan, J. Liang, S. C. Liew, V. C. M. Leung, and J. Li, "Timely information update with nonorthogonal multiple access," *IEEE Trans. Industrial Informatics*, vol. 17, no. 6, pp. 4096–4106, Jun. 2021.

[24] Q. Ren, T.-T. Chan, J. Liang, and H. Pan, "Age of information in sic-based non-orthogonal multiple access," in *Proc. IEEE WCNC*, Austin, TX, USA, 2022.

[25] X. Feng, S. Fu, F. Fang, and F. R. Yu, "Optimizing age of information in RIS-assisted NOMA networks: A deep reinforcement learning approach," *IEEE Wireless Commun. Lett.*, to appear in 2022.

[26] Z. Gao, A. Liu, C. Han, and X. Liang, "Non-orthogonal multiple access based average age of information minimization in LEO satellite-terrestrial integrated networks," *IEEE Trans. Green Commun. and Net.*, 2022.

[27] J. Choi, "NOMA based random access with multichannel ALOHA," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, 2017.

[28] Y. H. Bae and J. W. Baek, "Age of information and throughput in random access-based IoT systems with periodic updating," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 821–825, Apr. 2022.

[29] H. Zhang, Y. Kang, L. Song, Z. Han, and H. V. Poor, "Age of information minimization for grant-free non-orthogonal massive access using mean-field games," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7806–7820, Nov. 2021.