Grant Free Age-Optimal Random Access Protocol for Satellite-based Internet of Things

Tao Yang, Jian Jiao, *Member, IEEE,* Shaohua Wu, *Member, IEEE,* Rongxing Lu, *Fellow, IEEE* and Qinyu Zhang, *Senior Member, IEEE*

***Abstract*—In satellite-based Internet of Things (S-IoT) system, the timely status updating of terrestrial sensing user equipments (UEs) to satellite could be hampered by the long propagation delay, especially in massive machine type communications (mMTC). To guarantee the information freshness in S-IoT, a new performance indicator called age of information (AoI) is exploited to analyze the average AoI (AAoI) in the overload case of mMTC, and a grant free age-optimal (GFAO) random access protocol is proposed to lower the AAoI. Specifically, the closed-form expression of AAoI is derived by tracing the instantaneous AoI evolution of each UE through Markov analysis. Then, the proposed GFAO random access protocol is proved to achieve a minimum AAoI and a maximum throughput in S-IoT, by adjusting the number of access time slots in each transmission frame in the overload case of mMTC. Extensive simulations are conducted to validate the theoretical analysis, and show that there exists different optimal value of access time slots in system load region from 0.2 to 3, which can minimize AAoI and maximize throughput in the proposed GFAO random access protocol.**

***Index Terms*—Satellite-based Internet of Things, grant free random access, age of information, Markov analysis, status update.**

I. Introduction

IN the upcoming of Internet of Everything, the ability for broadband multiple access *anywhere and anytime* is urgently needed by lots of application domains, such as smart agriculture, intelligent manufacturing, smart grid and airplane entertainment [1]. Due to the dramatic increasing of the IoT devices, the massive machine type communications (mMTC) may induce traffic jamming and decline on quality of services (QoS), the design of a pragmatic random access protocol is decisive. However, the existing terrestrial networks are unlikely to fully support these communication scenarios due to the extreme topographies or the cost of deploying the terrestrial stations in a rural environment. Considering the ubiquitous coverage inherited from the satellites, Internet of Things (IoT) over satellite, referred to as satellite-based IoT (S-IoT), has become an attractive system architecture to provide a cost-efficient solution [2].

Moreover, the next generation of millimeter wave (mmWave) band high-throughput satellite (HTS) is viewed as a promising solution for the S-IoT, which has the capability of providing broadband access and agile deployment for the above mentioned massive access services [3], [4]. Besides, massive user equipments (UEs) are settled for the above mentioned applications, which keep constant surveillance, sensing and updating the status information to HTS [5], and the timeliness of the massive access data updates to HTS is also of paramount importance, since the obsolete information may lead to unpredictable or even disaster result. A simple example is that a set of sensors in the environment monitoring system to detect the disastrous phenomenons, such as forest fires and earthquakes, which should feedback to the control center for effective reaction as soon as possible [6], [7]. However, the geosynchronous orbit (GEO) HTS may lead to a 500 ms round trip time, and the low earth orbit (LEO) HTS has a lower propagation delay as about 4 to 10 ms but implies the Doppler shifts [8]. Consider the long propagation delay caused by huge distances in S-IoT [9], it is an open challenging to devise a practical random access protocol that can realize the massive connectivity of UEs over the shadowed-Rician fading channel in S-IoT, and guarantee the information freshness simultaneously [10].

A. Motivations and Related Works

The conventional indicators such as delay or throughput are insufficient to characterize the timeliness of the status updates and information freshness. To quantitatively depict the timeliness and freshness, a new performance indicator called age of information (AoI) is proposed in [11] and [12], and the definition of AoI is the elapsed time since the latest resolved status update generated by UE. Some related works have focused on the AoI in multiple access [13]–[15], [17], [18]. Minimizing the AoI through the grant-based UEs access scheduling has been analyzed in [13], [14]. The authors in [15] study the AoI of the carrier sense multiple access (CSMA) scheme. The AoI of UE side and related optimization in slotted Aloha random access schemes are considered in [17], [18]. However, there is still lack of work on the joint design of AoI and mMTC in S-IoT system.

The practical implementation of massive access wireless communication is one of the open challenges in the upcoming

Manuscript received xxx, 2021. This work was supported in part by the National Natural Sciences Foundation of China (NSFC) under Grant 62071141, Grant 61871147, Grant 61831008, and Grant 62027802. in part by the Natural Science Foundation of Guangdong Province under Grant 2020A1515010505, in part by the Guangdong Science and Technology Planning Project under Grant 2018B030322004, in part by the Shenzhen Science and Technology Program under Grant ZDSYS20210623091808025 and Grant GXWD20201230155427003-20200822165138001, and in part by the Major Key Project of PCL under Grant PCL2021A03-1. (*Corresponding author: Jian Jiao.*)

T. Yang, J. Jiao, S. Wu, and Q. Zhang are with the Communication Engineering Research Centre, Harbin Institute of Technology (Shenzhen), Guangdong 518055, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: 19S152106@stu.hit.edu.cn; jiaojian@hit.edu.cn; hitwush@hit.edu.cn; zqy@hit.edu.cn).

R. Lu is with the Faculty of Computer Science, University of New Brunswick, Fredericton, NB E3B 5A3, Canada, (e-mail: rlu1@unb.ca).

S-IoT system [19]. The mMTC is dominated by the uplink-oriented short packet transmissions, where a HTS should provide massive access in uplink mMTC [20]. Consider the tremendous growth UEs in future mMTC applications with S-IoT, the dedicated pilot allocation in conventional grant-based scheme is impractical. Moreover, the HTS is hundreds or thousands of kilometers away from the UEs, the grant free random access can reduce the propagation delay in the total access delay, and also the signaling overhead that induced by traditional grant-based schemes. Besides, for grant free random access system, each activated UE randomly chooses a pilot sequence from a predefined pilot sequence set, and then updates to HTS together with the data payload, which could mitigate the system overload and pilot collision [21], [22]. Thus, it is a natural choice for S-IoT to prefer decentralized pilot allocation and exploit grant free random access [23]–[25].

The most widely used grant free random access protocols in satellite communications are Aloha protocols, and many ameliorated versions of Aloha protocols are proposed to improve the system performance [26], [27], which can partially alleviate the conflict of massive access [28]. A coded slotted Aloha (CSA) scheme for non-collaborative random access is proposed in [29], where the bipartite graph with the erasure coding is utilized for the asymptotical analysis. Further, the frameless Aloha protocol for time-varying channels is introduced in [30] by leveraging on rateless coding, where the theoretical performance analysis and sub-optimal solution between the access failure probability (AFP) and activate probability of UEs are presented under limited access time slots.

Although the above grant free random access can partially alleviate the pilot collision in massive access, the available pilot sequences cannot satisfy the explosive growth of IoT UEs, and the pilot collision becomes the bottleneck in massive access. To alleviate this issue, some relevant works have increased the access time slots to provide secondary access opportunity for collision UEs. A strongest-user collision resolution (SUCR) random access protocol is proposed in [31], where the access point (AP) allows the strongest UE to perform the second access. Moreover, the extension versions named SUCR combined idle pilots access (SUCR-IPA) and SUCR combined graph-based pilots access (SUCR-GBPA) are proposed in [32], [33]. In SUCR-IPA, the strongest UE is performed as the same in the SUCR, and if the collision UE in the first access can pass an access class barring (ACB) check, they can contend for the idle pilots to perform the second access. The steps of SUCR-GBPA are similarly to SUCR-IPA, but the data payload is transmitted in the pilot repeating and reselecting phase rather than after the pilots allocating phase. However, the SUCR and its improved versions still could not completely solve the congestion problem from a massive number of burst UEs transmission. Since the idle pilots is significant declining in an overload case, the AFP that an activated UE fails to access to HTS would increase as well as the system load (i.e., ratio of the number of activated UEs to that of available pilot sequences) in SUCR, which leads to the increasing of AoI. It is worth noting that the characteristics of delay tolerant but reliability-critical in typical space communications, and a tradeoff between the minimum average AoI (AAoI) and an optimal value of access time slots under the restricted pilots can be achieved.

B. Contributions

The main contributions in this paper are concluded as follows.

1) *Grant Free Age-Optimal Random Access Protocol:* We design a grant free age-optimal (GFAO) random access protocol to lower the AAoI for the timeliness services in S-IoT, where each activated UE randomly selects a pilot, and transmits it along with its data payload to a HTS in the first access time slot. Then, the collision UEs can independently retransmit their data payloads after randomly reselecting a pilot in subsequent access time slots. We show that our GFAO random access protocol can achieve lower AAoI and higher throughput by adjusting the number of access time slots in each transmission frame.
2) *Markovain Analysis for GFAO Random Access Protocol:* We first analyze the instantaneous AoI evolution via Markovain analysis, and obtain the relationship between the AAoI and number of UEs, activated probability and available resources (i.e., number of the available pilot sequences and access time slots). Then, the closed-form expressions of AAoI for the proposed GFAO random access protocol are derived. Moreover, we derive the expressions of AFP and throughput, and analyze the relationship between throughput and AAoI.
3) *Parameters Optimization and Validation of GFAO Protocol:* We prove that there exists an optimal number of access time slots to achieve minimum AAoI and maximum throughput under any given system load as shown in Fig. 8 and Fig. 9, respectively. Furthermore, extensive simulations are conducted to validate our theoretical derivations, and also show that our GFAO protocol outperforms the related state-of-the-art schemes by optimizing the access time slots to the system load to simultaneously enable the high throughput and low AAoI.

The rest of this paper is presented as follows. Section II exhibits system model and the principles of the GFAO random access protocol. Section III describes the Markovain analysis for the proposed GFAO random access protocol. The parametrical analysis for GFAO random access protocol is depicted in section IV. Section V shows the simulation results and Section VI exhibits the conclusion.

II. System Model and GFAO Random Access Protocol

A. System Model

Our GFAO random access protocol for a time-critical mMTC S-IoT system is depicted in Fig. 1, where K UEs update their status to a HTS. Let $\mathbf{P} = \{P_1, P_2, \cdots, P_\tau\}$ denote a predetermined pilot set, which has τ available pilot sequences with the same length. Assume that each access frame is divided

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

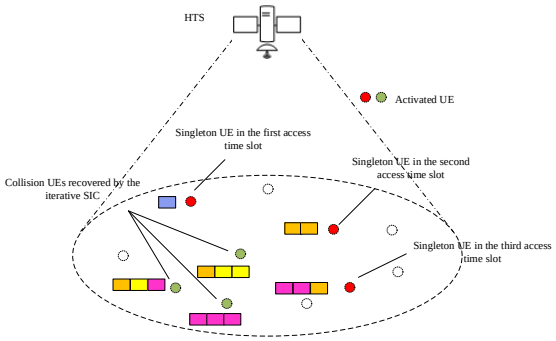

Fig. 1. System model of uplink transmission in a random access scenario for a time-critical status update S-IoT system in a transmission frame.

TABLE I
LIST OF ABBREVIATIONS

Acronyms	Definition
IoT	Internet of Things
S-IoT	Satellite-based IoT
UEs	User Equipments
mMTC	massive Machine Type Communication
AoI	Age of Information
AAoI	Average AoI
GFAO	Grant Free Age-Optimal
QoS	Quality of Services
HTS	High Throughput Satellite
LEO	Low Earth Orbit
GEO	Geosynchronous Earth Orbit
CSMA	Carrier Sense Multiple Access
CSA	Coded Slotted Aloha
SUCR	Strongest User Collision Resolution
SUCR-IPA	SUCR combined Idle Pilots Access
SUCR-GBPA	SUCR combined Graph-Based Pilots Access
AP	Access Point
AFP	Access Failure Probability
SIC	Successive Interference Cancellation

into N equal length access time slots, and each access time slot can update one status packet (i.e., a selected pilot and update data payload). Note that one access time slot is consisted of the packet transmission delay, processing delay at HTS and UE, and propagation delay [34]. Considering that the altitude of LEO HTS is 765 km in this paper, thus the two-way propagation delay is 5.1 ms, and we assume that transmission delay takes 1 ms since the short packet transmissions are dominated in mMTC [35]. Moreover, the processing delay are negligible compare to the long propagation delay. Thus, the length of one access time slot is set as 7.1 ms [36]. Note that a contention timer in satellite link can be set up to 50 ms [37], [38], thus a transmission frame could be able to support 6 access time slots in this S-IoT system. The acronyms throughout the paper are summarized in Table I.

At the beginning of each transmission frame, each of K UEs generates a status packet with probability λ, and those UEs generating status packets are called activated UEs, denoted by $\mathbf{U} = \{U_1, U_2, ..., U_{K_a}\}$. Moreover, a buffer in each UE can only cache single status packet, and UE will discard the status packet at the end of each transmission whether it successfully access to the HTS or not [39].

Moreover, the channel from UEs to HTS is modeled as the widely-used shadowed-Rician fading channel in this paper [42]. Specially, the channel gain ξ from U_i to HTS is given by

$$\xi_i(t) = \xi_i^{\text{sca}}(t)\exp(j\varpi_i^{\text{sca}}(t)) + \xi_i^{\text{los}}(t)\exp(j\varpi_i^{\text{los}}), \quad (1)$$

where $\xi_i^{\text{sca}}(t)\exp(j\varpi_i^{\text{sca}}(t))$ represents the scattering component, and $\xi_i^{\text{los}}(t)\exp(j\varpi_i^{\text{los}})$ is the light-of-sight (LoS) component between U_i and HTS.

Note that the relative high altitude of LEO HTS, the Doppler shifts in the LoS component $\xi_i^{\text{los}}(t)\exp(j\varpi_i^{\text{los}})$ are identical for diversity propagation paths of UEs [43]. Moreover, we assume the terrestrial UEs are quasi-static, and $(j\varpi_i^{\text{sca}}(t))$ in the scattering component is the stationary random phase process with uniform distribution over $[0, 2\pi)$.

Therefore, the Doppler effects in our system can be modeled as a constant multiplicator factor. In addition, the bandwidth of mmWave band LEO HTS is about 800 MHz to 2 GHz, a guard band that double than the Doppler shifts is utilized to relieve the influence of Doppler shifts on the system [44]. Besides, the amplitudes of scattering and LoS components are following Rayleigh and Nakagami-m distributions, respectively, and the probability density functions (PDFs) are as follows

$$\begin{cases} p_{\xi_i^{\text{sca}}}(a) = \frac{a}{b}\exp\left(\frac{-a^2}{2b}\right), & a \geq 0, \\ p_{\xi_i^{\text{los}}}(z) = \frac{2\imath^{\imath}}{\Upsilon(\imath)\Omega^{\imath}} z^{2\imath-1}\exp\left(\frac{-\imath z^2}{\Omega}\right), & z \geq 0, \end{cases} \quad (2)$$

where $2b$ and Ω represent average power of the scattering and LoS components, respectively, $\imath$ is the Nakagami-m parameter, and $\Upsilon(.)$ is the Gamma function. We assume that the transmission power of all UEs is P_T. Hence, under the shadowed-Rician fading model, the distribution of received power for a single UE, denoted by $f_{sR}(p)$, can be expressed as [42]

$$\begin{aligned} f_{sR}(p) = &\left(\frac{2b\imath}{2b\imath+\Omega}\right)^{\imath}\frac{1}{2b}\exp\left(-\frac{p}{2bP_T}\right) \\ &\cdot_1 F_1\left(\imath, 1, \frac{\Omega p}{2b(2b\imath+\Omega)P_T}\right), \end{aligned} \quad (3)$$

where ${}_1F_1(a, b, c)$ represents the confluent hypergeometric function.

B. Procedure of GFAO Random Access Protocol

Fig. 2 illustrates the procedure of the proposed GFAO random access protocol, and four steps are executed for the GFAO random access protocol. Prior to the four steps, a control signal is broadcasted from the HTS to all UEs for estimating the average channel gain, and allowing each UE to synchronize with the HTS for the following grant free random access.

Step 1 Let K_a $(K_a \leq K)$ denote the number of activated UEs in a transmission frame, in the first access time slot, a pilot is randomly selected by each of K_a UEs from the predetermined pilot set $\mathbf{P}$, and sent to the HTS followed by the metadata (e.g., the selected pilot index, timestamp of the status

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

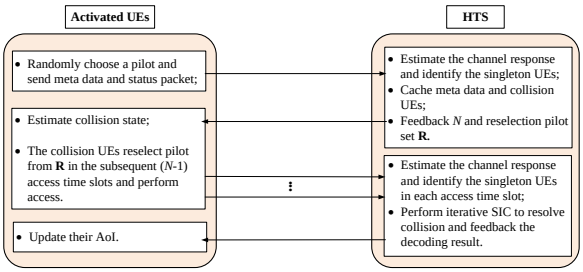

Fig. 2. The procedure of the proposed GFAO random access protocol.

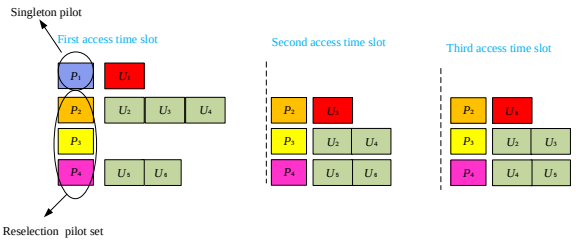

Fig. 3. An example of multi-slot accessing procedure in GFAO random access protocol.

packet, etc.) and data payload together. It is worth noting that the probability that each UE chooses a certain pilot is equal to $\frac{1}{\tau}$, and if more than one UEs choose a same pilot would incur the pilot collision.

Step 2 The HTS first estimates the received pilots in the first access time slot, then it utilizes the channel response to identify the pilot collision and perform decoding procedure. Specifically, in the first access time slot, the HTS identifies the state of a pilot sequence by exploiting received signal power along with statistical channel information [40], where the detail analysis is provided in the following Section II.C. There are three states for each pilot sequence:

Case 1: If a pilot is not selected by any UE, then the pilot is called idle pilot;

Case 2: If a pilot is selected by only one UE, then the pilot is called singleton pilot, and the HTS immediately identifies the status packets with the singleton pilots and decodes them at the end of the current transmission frame;

Case 3: If a pilot is selected by more than one UEs, then the pilot is named collision pilot.

We define $\mathbf{R}$ as the reselection pilot set, which consists both of idle pilots and collision pilots. In the end of the first access time slot, HTS broadcasts a feedback of N and $\mathbf{R}$ to the UEs.

Step 3 According to the feedback $\mathbf{R}$ and N, each UE can detect independently they are in collision at the first access time slot or not, where the UEs select a singleton pilot will not perform access in the rest of $N-1$ multi-slot. Then, each collision UE chooses a random pilot sequence from $\mathbf{R}$ in each of subsequent $N-1$ multi-slot, and directly sends it with their meta data and data payload to the HTS, where the HTS will not feedback until the end of the last access time slot.

Step 4 Similar to the Step 2, the HTS estimates the received pilot in each $N-1$ access time slot, and identifies the singleton pilot and caches the collision UEs for the later decoding. At the end of the N-th access time slot, the HTS immediately recovers the status packets with the singleton pilot in the transmission frame. If the HTS still has collision UEs in its buffer, the HTS performs successive interference cancellation (SIC) to recover the collision UEs via the recovered singleton UEs [41].

To facilitate the analysis, we assume that there are six activated UEs denoted by $\mathbf{U} = \{U_1, U_2, ..., U_6\}$ and four available pilot sequences denoted by $\mathbf{P} = \{P_1, P_2, ..., P_4\}$ as shown in Fig. 3, and there are three access time slots in this transmission frame and each UE has an orthogonal codebook to encode their status packets [19]. In the first access time slot, only U_1 chooses a singleton pilot P_1, and P_2, P_4 are collision pilots which are selected by $\{U_2, U_3, U_4\}$ and $\{U_5, U_6\}$, respectively, and P_3 is idle pilot. Thus, the reselection pilot set $\mathbf{R} = \{P_2, P_3, P_4\}$. After receiving the feedback and the reselection pilot set $\mathbf{R}$, U_1 will keep silent, and each of the collided UEs $\{U_2, U_3, ..., U_6\}$ reselects a pilot from $\mathbf{R}$ at random and performs access in the next two access time slots. Since there is no prior knowledge of random activity and pilot selections of K UEs for the HTS, the HTS needs to identify the pilot contamination via the statistic channel information and their codebooks, then it performs the iterative SIC, which is described in detail as follows.

Firstly, the HTS recovers the status packets with singleton pilots in the three access time slots, i.e., P_1 chosen by U_1 in the first access time slot, P_2 chosen by U_3 in the second access time slot and P_2 chosen by U_6 in the third access time slot. Secondly, the HTS removes all the copies of the status packets of U_3 and U_6 in the transmission frame, and P_4 chosen by U_5 in the first access time slot, P_4 chosen by U_5 in the second access time slot and P_3 chosen by U_2 in the third access time slot become singleton pilots. Then, the HTS recovers the status packets of U_5 and U_2, and removes all the copies of the status packets of U_5 and U_2 in the transmission frame. Thus, pilots chosen by U_4 in each access time slot become singleton pilots, and the status packet of U_4 can be recovered. Finally, the HTS broadcasts a feedback of the decoding states to all activated UEs.

C. Total Interfering Power in GFAO Random Access Protocol

Since the signal power of the collision pilots hinders the decoding process of singleton UEs, we need to analyze the distribution of total number of collision UEs and their total interfering power in a transmission frame. Thus, the expressions for total number of collision UEs in the first access time slot and total interfering power in the GFAO random access system are derived in the following.

In each transmission frame, since each UE actives independently with a certain probability λ, we assume that the distribution of the number of activated UEs in a transmission frame is Poisson distribution. Recall that the number of pilots in $\mathbf{P}$ is τ, let $\gamma = E[K_a]$ represent the average number of activated UEs. The probability that an activated UE fails to access in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

the first access time slot is denoted by $\eta = 1 - \exp(-\gamma/\tau)$. Thus, the average number of collision UEs in a same pilot satisfies a Poisson distribution with parameter γ/τ.

We denote Θ_s as the event that there exist s singleton pilots in the first access time slot of a transmission frame, i.e., there exist s singleton UEs with different pilots from $\mathbf{P}$, which means that there are $\tau - s$ pilots in reselection pilot set $\mathbf{R}$. For clearer analysis, let $H_{\tau-s}$ denote the event that the number of UEs in each of these reselection pilots is not 1, and Φ_ν denotes the event that there are ν collision UEs in all collision pilots. Thus, the conditional probability we need to derive is $\Pr(\Phi_\nu | H_{\tau-s})$. Moreover, we can express the conditional probability as piecewise functions. When $\nu = 0$, the result of $\Pr(\Phi_0 | H_{\tau-s})$ can be directly obtained based on the exponential function of the Poisson distribution as follows:

$$\Pr(\Phi_0 | H_{\tau-s}) = \left(\frac{\exp\left(-\frac{\gamma}{\tau}\right)}{\Pr(\Theta_1)} \right)^{\tau-s}, \tag{4}$$

where $\Pr(\Theta_1) = 1 - \left(\frac{\gamma}{\tau}\right)\exp\left(-\frac{\gamma}{\tau}\right)$. When $\nu = 1$, i.e., in $\tau - s$ reselection pilots the total number of UEs is 1, which is an impossible event, thus

$$\Pr(\Phi_1 | H_{\tau-s}) = 0. \tag{5}$$

When $\nu \geq 2, \tau - s = 1$, i.e., the total number of UEs in one of reselection pilots is ν, which can be directly computed by Poisson distribution as follows:

$$\Pr(\Phi_\nu | H_1) = \frac{\left(\frac{\gamma}{\tau}\right)^\nu \exp\left(-\frac{\gamma}{\tau}\right)}{\nu! \Pr(\Theta_1)}. \tag{6}$$

In general circumstance, i.e., $\nu \geq 2$ and $\tau - s \geq 2$, the conditional probability $\Pr(\Phi_\nu | H_{\tau-s})$ can be derived by exploiting the recursion formula, thus we can obtain the expression as follows:

$$\begin{aligned} \Pr(\Phi_\nu \mid H_{\tau-s}) &= \frac{\Pr(H_{\tau-s} \mid \Phi_\nu) \Pr_{\tau-s}(\Phi_\nu)}{\Pr(H_{\tau-s})} \\ &= \frac{\left(1 - \Pr\left(\bar{H}_{\tau-s} \mid \Phi_\nu\right)\right) \Pr_{\tau-s}(\Phi_\nu)}{\Pr(H_{\tau-s})} \\ &= \frac{\Pr_{\tau-s}(\Phi_\nu) - \Pr\left(\Phi_\nu, \bar{H}_{\tau-s}\right)}{\Pr(H_{\tau-s})}, \end{aligned} \tag{7}$$

where $\bar{H}_{\tau-s}$ represents the event that there exists a pilot in $\mathbf{R}$ in which the number of UEs is 1, which is the inverse event of $H_{\tau-s}$.

By replacing the event $H_{\tau-s}$ into event Θ_s, the corresponding piecewise functions are given as Eq. (8) at the top of next page, where $\Pr(\Theta_s)$, $\Pr_s(\Phi_\nu)$ and $\Pr(\overline{\Theta}_1) = 1 - \Pr(\Theta_1)$ are given by

$$\begin{cases} \Pr(\Theta_s) = \left(1 - \frac{\gamma}{\tau}\exp\left(-\frac{\gamma}{\tau}\right)\right)^s, \\ \Pr_s(\Phi_\nu) = \frac{\left((\tau-s)\frac{\gamma}{\tau}\right)^\nu \exp\left(-(\tau-s)\frac{\gamma}{\tau}\right)}{\nu!}, \\ \Pr(\overline{\Theta}_1) = \left(\frac{\gamma}{\tau}\right)\exp\left(-\frac{\gamma}{\tau}\right), \end{cases} \tag{9}$$

where $\sigma = \min\{\tau - s, \nu\}$ and $\Pr(\Phi_{\nu-i}, \Theta_{s-i}) = \Pr(\Phi_{\nu-i} | \Theta_{s-i}) \Pr(\Theta_{s-i})$.

The total number of collision UEs and the distribution of received power can be exploited to calculate the total interfering power. Hence, under the shadowed-Rician fading channel, the distribution of total interfering power $g(I|\Theta_s)$ can be derived as follows,

$$g(I|\Theta_s) = \sum_{\nu=0}^{\infty} f_{sR}(I) \underset{\nu-1}{\otimes} f_{sR}(I) \times \Pr(\Phi_\nu | \Theta_s), \tag{10}$$

where $\underset{\nu-1}{\otimes}$ denotes $\nu - 1$ times convolution.

We choose the parameters in [42] as $b = 0.251$, $\imath = 5.21$, $\Omega = 0.278$, the transmit distance between UEs and HTS is 765 km, the transmit power is 100 W, the transmit frequency is 30 GHz, and the bandwidth and thermal noise density are 200 MHz and -174 dBm/Hz, respectively, and the Monte Carlo simulations are shown in Fig. 4 and Fig. 5, which validate the accuracy of the derived expressions that agree well with the simulation results under different parameters.

Fig. 4 shows the relationship between collision probability η and total number of collision UEs ν. It can be observed that η is fluctuating over ν. Specifically, the collision probability η is decreasing when ν is odd, and is increasing when ν is even, since the probability of a collision pilot chosen by two UEs simultaneously (i.e., $\Pr(\Phi_2|\Theta_{s-1})$) is higher than that of a collision pilot chosen by three UEs simultaneously (i.e., $\Pr(\Phi_3|\Theta_{s-1})$). Besides, the probability of $\nu = 4$ (i.e., $2\Pr(\Phi_2|\Theta_{s-2}) + \Pr(\Phi_4|\Theta_{s-1})$), which means two collision pilots are selected by four UEs (i.e., every two UEs choose one of the two collision pilots) or one pilot is selected by four UEs simultaneously, is higher than that of $\nu = 3$.

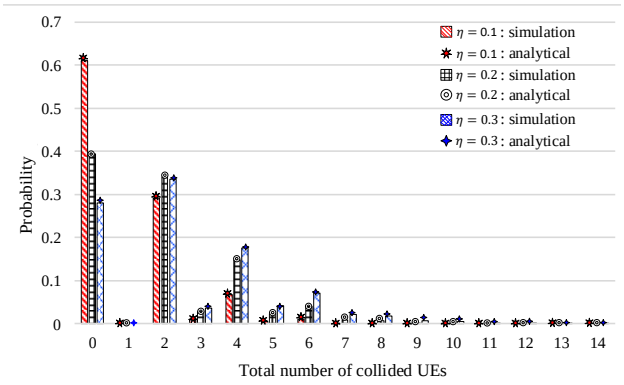

Fig. 4. The collision probability.

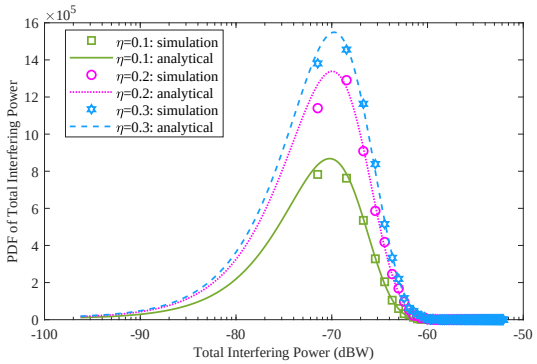

Fig. 5. Total interfering power and collision probability under the shadowed-Rician fading channel.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

$$\begin{cases} \Pr\left(\Phi_0\left|\Theta_s\right.\right) = \left(\frac{\exp\left(-\frac{\gamma}{\tau}\right)}{\Pr(\Theta_1)}\right)^{\tau-s}, \\ \Pr\left(\Phi_1\left|\Theta_s\right.\right) = 0, \\ \Pr\left(\Phi_\nu\left|\Theta_s\right.\right) = \frac{\Pr_s(\Phi_\nu) - \sum\limits_{i=1}^{\sigma}\binom{s}{i}\Pr\left(\overline{\Theta}_1\right)^i \Pr(\Phi_{\nu-i},\Theta_{s-i})}{\Pr(\Theta_{\tau-s})},\ \nu \geq 2, s \leq \tau-2, \\ \Pr\left(\Phi_\nu\left|\Theta_s\right.\right) = \frac{\left(\frac{\gamma}{\tau}\right)^\nu \exp\left(-\frac{\gamma}{\tau}\right)}{\nu!\Pr(\Theta_1)},\ \nu \geq 2, s = \tau-1, \end{cases} \tag{8}$$

The PDF of total interfering power for different η is shown in Fig. 5. It can be observed that the peak values of these PDF are increasing with η, and the total interfering power are mostly lower than a quite small value -60 dBW as shown in Fig. 5. Therefore, we can assume that the SIC decoding process of singleton UEs is not affected by the interfering power in our paper.

III. Analysis of AoI for GFAO Random Access Protocol

We present the analysis of the AAoI of each UE for the GFAO random access protocol in this section. First, we define the fundamental notations of AoI for the following analysis, and model the evolution of instantaneous AoI for an individual UE by exploiting the Markov analysis. Then, we derive the expressions of the AAoI and further analyze the relationship of AAoI to the system throughput.

A. Fundamental Notations of AoI

Recall that the AoI is the elapsed time since the latest status update packet is generated at the UE until it is resolved by the HTS. Let $t = 0, 1, 2, 3...$ denote the time, which is normalized to one access time slot. We define $G_i(t)$ as the timestamp of the latest status packet generated from the i-th UE as of time t. Thus, the instantaneous AoI of the i-th UE observed by itself is given by

$$\delta_i(t) = t - G_i(t). \tag{11}$$

If the i-th UE sends a status update to the HTS, $\delta_i(t)$ grows linearly over t. Besides, $\delta_i(t)$ drops to zero if the packet successfully accesses to the HTS or the transmission frame is finished, as the UE discards the status packet at the end of each transmission frame whether it successfully accesses to the HTS or not.

We denote the instantaneous AoI of the i-th UE observed by the HTS as $\Delta_i(t)$. Note that if a packet is resolved, the HTS updates the state of the i-th UE, and $\Delta_i(t)$ drops to the gap between the generation time of the resolved packet and the current time. Let Δ^i denote the AAoI of the i-th UE at the HTS, according to the aforementioned features, the definition of Δ^i can be given by

$$\Delta^i = \lim_{t\to\infty} \frac{1}{t}\sum_0^t \Delta_i(t). \tag{12}$$

TABLE II
The definition of notations

Notation	Definition
K	Total number of UEs
K_a	Number of activated UEs
N	Number of slots in each transmission frame
$\delta_i(t)$	Instantaneous AoI of the i-th UE
$\Delta_i(t)$	Instantaneous AoI of the i-th UE at HTS
Δ^i	AAoI of the i-th UE, which is equal to the system AAoI at HTS
τ	Number of pilot sequences
λ	Activated probability of the i-th UE
P_{err}	AFP of the GFAO random access protocol
P_{nc}	Non-collision probability at the first access time slot
g	System load
T	Throughput of the GFAO random access protocol

Also, the operation of each UE is independent identically distributed (i.i.d.) in the random access system, the system AAoI can be defined as

$$\Delta = \frac{1}{K}\sum_{i=1}^{K}\Delta^i, \tag{13}$$

and we can estimate Δ by exploiting Δ^i.

B. Derivation of System AAoI in GFAO Random Access Protocol

We analyze the AAoI of single UE to model the AAoI of the GFAO random access protocol in this subsection, and the related notations are summarized in Table II. In our GFAO random access protocol, the index of transmission frames are denoted by $\chi = 1, 2, 3, ...$, and for the i-th UE, there exist four different cases in the χ-th transmission frame as follows.

Case I: The i-th UE is activated and selects a singleton pilot at the first access time slot in the χ-th transmission frame;

Case II: The i-th UE is activated and selects a collision pilot and its packet transmitted during the χ-th transmission frame is resolved by the HTS via the iterative SIC;

Case III: The i-th UE is activated and selects a collision pilot, but the packet is failed to be resolved by the HTS;

Case IV: The i-th UE is not activated at the χ-th transmission frame.

Then, we define S_χ^i as the service time of the i-th UE in the χ-th transmission frame, i.e., the gap between the generation time to the reception time of the packet. Thus, we have $S_\chi^i = 1$ in case I, $S_\chi^i = N$ in case II, and $S_\chi^i = 0$ in case III and case IV. Accordingly, we can denote Γ_χ^i as the elapsed time since

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

the i-th UE successfully accesses to the HTS at the first access time slot until the end of χ-th transmission frame. In case I, $\Gamma_{\chi}^{i} = N-1$, and under the other three cases, $\Gamma_{\chi}^{i} = 0$.

Moreover, consider the i-th UE in case I, II, III, i.e., the activated states of i-th UE, we define g_j^i as the generation time of the j-th packet. Note that not every packet can be successfully received by the HTS, and we define d_m^i as the time of the m-th packet resolved by HTS from the i-th UE, where $m \leq j$. Therefore, we define W_m^i as the consumed time between the m-th packet and the $(m-1)$-th packet, which is the sum of the transmission frames that the i-th UE is not activated or the packets are failed to access. Thus, W_m^i is an integer multiples of N. For simplicity, let S_m^i denote the service time of the m-th packet in the χ-th transmission frame, and Γ_{m-1}^i denotes the elapsed time of the $(m-1)$-th packet in the χ'-th transmission frame. Then, let Y_m^i denote the time interval from the reception time of $(m-1)$-th packet to that of the m-th packet. Therefore, we have $Y_m^i = S_m^i + \Gamma_{m-1}^i + W_m^i$.

In our GFAO random access protocol, $\Delta_i(t)$ only can drop under case I and II. To simplify the analysis, we denote $A_i(\chi)$ as the state of the i-th UE at the current transmission frame, where $A_i(\chi) = 1$ means the i-th UE is activated during the χ-th transmission frame, and $A_i(\chi) = 0$ otherwise. Then, when $A_i(\chi) = 1$, we denote $B_i(\chi)$ as the access state of status packet when the χ-th transmission frame is finished, where $B_i(\chi) = 1$ means the packet is successfully recovered at the HTS, otherwise $B_i(\chi) = 0$. The indicator $C_i(\chi)$ is used to represent the states of $B_i(\chi) = 1$, where $C_i(\chi) = 1$ represents the packet is resolved at the first access time slot, and $C_i(\chi) = 0$ means successful decoded by HTS through the iterative SIC when the χ-th transmission frame is finished.

According to our GFAO random access protocol, $\Delta_i(t)$ drops to N only when $B_i(\chi) = 1$, i.e., the i-th UE is activated at the current transmission frame and the HTS successfully resolved the packet. Therefore, the evolution of $\Delta_i^{\chi}(t)$ at the end of the χ-th transmission frame is given by

$$\Delta_i^{\chi}(t) = \begin{cases} N, \text{ if } B_i(\chi) = 1 \\ \Delta_i^{\chi-1}(t) + N, \text{ if } A_i(\chi) = 0 \text{ or } B_i(\chi) = 0. \end{cases} \tag{14}$$

An example of the evolution of $\delta_i(t)$ and $\Delta_i^{\chi}(t)$ are shown in Fig. 6. The blue line represents the instantaneous AoI $\Delta_i^{\chi}(t)$ of the i-th UE observed by the HTS. The red solid and dotted lines and the green line are depicted the instantaneous AoI $\delta_i(t)$, where the red solid line represents the status packets are successful decoded when the transmission frame is finished, otherwise is the red dotted line. Note that the packet is discarded when the transmission frame is finished whether it is successfully resolved by the HTS or not, which means $\delta_i(t) = 0$ when each transmission frame is finished. Moreover, the green line represents that it is a singleton packet at the first access time slot in this transmission frame.

As shown in Fig. 6, at g_1, g_2, g_3, g_4 and g_5, the i-th UE becomes activated since a packet is generated at the buffer in each transmission frame. The packets generate at g_1, g_3 and g_5 are recovered through the iterative SIC decoding, resulting in the blue line dropping to N as $\Delta_i^1(t) = \Delta_i^3(t) = \Delta_i^5(t) = N$. The packet generates at g_2 is singleton packet in the first

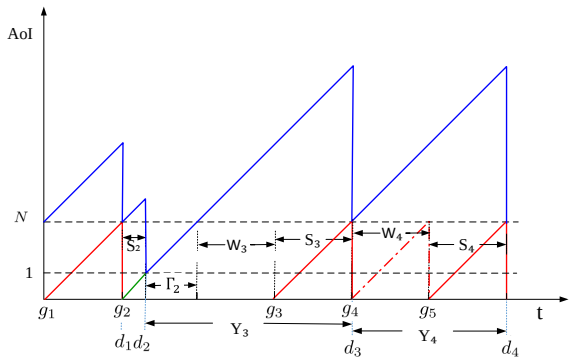

Fig. 6. The evolution of the instantaneous AoI δ_i at the i-th UE and Δ_i at HTS, where the red and green lines are δ_i at the i-th UE, and the blue line represents Δ_i of the i-th UE at the HTS.

access time slot, resulting in $\Delta_i^{\chi}(t)$ dropping to 1, and the i-th UE keeps silent in the rest $\Gamma_{\chi} = N-1$ access time slots, and $\Delta_i^2(t) = N$. Besides, The packet generates at g_4 fails to access, resulting the $\Delta_i^5(t) = 2N$ at g_5.

Since the AAoI Δ^i can be influenced by $\Delta_i^{\chi}(t)$, and successfully receptions can significantly decrease $\Delta_i^{\chi}(t)$, thus we mainly analyze the AAoI over multiple receptions. Assume that the HTS has resolved M status packets from the i-th UE in total c transmission frames, which includes l access time slots, i.e. $\frac{l}{N} = c$. To analyze the AAoI of M receptions, the AoI of the m-th resolved status packet ($0 \leq m \leq M$) is derived as follows.

Note that the time interval between the m-th reception in the χ-th transmission frame and the $(m-1)$-th reception in the χ'-th transmission frame is $Y_m^i = \Gamma_{m-1}^i + W_m^i + S_m^i$, where $\Gamma_{m-1}^i = N-1$ if the last reception is a singleton packet, and the expression of Γ_{m-1}^i can be given as follows,

$$\Gamma_{m-1}^i = \begin{cases} N-1, \text{ if } C_i(\chi') = 1, \\ 0, \text{ otherwise.} \end{cases} \tag{15}$$

Then, note that W_m^i denotes the consumed time gap between the m-th reception and the $(m-1)$-th reception, and there may exist $|\chi - \chi'| \geq 1$ transmission frames between two successful receptions. Thus, let $1 \leq k \leq |\chi - \chi'|$, we have $W_m^i = (k-1)N$, and W_m^i follows the geometric distribution.

Moreover, S_m^i follows a binomial distribution, which is related to two conditions of $C_i(\chi)$. Therefore, $S_2 = 1$ if $C_i(2) = 1$, and $S_3 = N$ as $C_i(4) = 0$ as shown in Fig. 6, and we can define S_m^i as follows,

$$S_m^i = \begin{cases} 1, \text{ if } C_i(\chi) = 1, \\ N, \text{ if } C_i(\chi) = 0. \end{cases} \tag{16}$$

We denote X_m^i as the sum of W_m^i and S_m^i. As shown in Fig. 6, the AoI consumptions of a status update packet at HTS can be split into four segments: 1) A rectangle whose base is Γ_m^i and height is 1; 2) A triangle whose base is Γ_m^i and height is $N-1$; 3) A rectangle whose base is X_m^i and height is N; 4) An isosceles triangle whose waist is X_m^i.

$$\Delta^i = \lim_{M\to\infty} \frac{M}{\sum_{m=1}^{M} Y_m^i} \times \left[\sum_{m=1}^{M} \frac{N X_m^i}{M} + \sum_{m=1}^{M} \frac{(X_m^i)^2}{2M} + \sum_{m=1}^{M} \frac{\Gamma_m^i}{M} + \sum_{m=1}^{M} \frac{(\Gamma_m^i)^2}{2M} \right]. \tag{17}$$

Therefore, when the number of resolved status packets M grows to infinity, the AAoI Δ^i is the sum of AoI consumptions of all M status packets divided by the sum of Y_m^i, which is given as Eq. (17) at the top of this page.

Since the i.i.d. processes of all UEs follow the law of large numbers, we can further simplify Eq. (17) as follows,

$$\begin{aligned}\Delta^i &= \frac{\left[N\mathrm{E}\left[X_m^i\right] + \frac{1}{2}\mathrm{E}\left[(X_m^i)^2\right] + \mathrm{E}[\Gamma_m^i] + \frac{1}{2}\mathrm{E}\left[(\Gamma_m^i)^2\right]\right]}{\mathrm{E}\left[Y_m^i\right]} \\ &= \frac{\left[(N-1)\mathrm{E}\left[X_m^i\right] + \frac{1}{2}\mathrm{E}\left[(X_m^i)^2\right] + \frac{1}{2}\mathrm{E}\left[(\Gamma_m^i)^2\right]\right]}{\mathrm{E}\left[Y_m^i\right]} + 1.\end{aligned} \tag{18}$$

Since W_m^i, S_m^i and Γ_m^i are i.i.d., we have $\mathrm{E}\left[X_m^i\right] = \mathrm{E}\left[W_m^i + S_m^i\right] = \mathrm{E}\left[W_m^i\right] + \mathrm{E}\left[S_m^i\right]$ and $\mathrm{E}\left[Y_m^i\right] = \mathrm{E}\left[W_m^i + S_m^i + \Gamma_m^i\right] = \mathrm{E}\left[W_m^i\right] + \mathrm{E}\left[S_m^i\right] + \mathrm{E}[\Gamma_m^i]$.

By exploiting the above analysis of Γ_m^i, W_m^i and S_m^i, $\mathrm{E}[\Gamma_m^i]$, $\mathrm{E}\left[(\Gamma_m^i)^2\right]$, $\mathrm{E}\left[W_m^i\right]$, $\mathrm{E}\left[(W_m^i)^2\right]$, $\mathrm{E}\left[S_m^i\right]$ and $\mathrm{E}\left[(S_m^i)^2\right]$ can be expressed as,

$$\begin{cases} \mathrm{E}[\Gamma_m^i] = (N-1)P_a^i P_{nc}, \\ \mathrm{E}\left[(\Gamma_m^i)^2\right] = (N-1)^2\lambda P_{nc}, \\ \mathrm{E}\left[W_m^i\right] = \frac{N}{\lambda(1-P_{err})}, \\ \mathrm{E}\left[(W_m^i)^2\right] = N^2\frac{2-\lambda(1-P_{err})}{(\lambda(1-P_{err}))^2}, \\ \mathrm{E}\left[S_m^i\right] = \frac{P_{nc}}{1-P_{err}} + N\left(1-\frac{P_{nc}}{1-P_{err}}\right), \\ \mathrm{E}\left[(S_m^i)^2\right] = \frac{P_{nc}}{1-P_{err}} + N^2\left(1-\frac{P_{nc}}{1-P_{err}}\right), \end{cases} \tag{19}$$

where P_{nc} represents the probability that the i-th UE chooses a singleton pilot from $\mathbf{P}$ at the first access time slot. Recall that the number of activated UEs is K_a, the expression of P_{nc} is given by

$$P_{nc} = C_\tau^1 \left(\frac{1}{\tau}\right)\left(1-\frac{1}{\tau}\right)^{K_a-1} = \left(1-\frac{1}{\tau}\right)^{K_a-1}, \tag{20}$$

where $K_a = \lambda K$.

Moreover, the AFP in the GFAO random access protocol consists of three parts: 1) The probability that a collision UE is not recovered through SIC; 2) The probability that a singleton pilot is identified as a collision pilot, which can be solved after subsequent $N-1$ multi-slot reselection phase; 3) The probability that a collision pilot is identified as a singleton pilot. Let P_{err} denote the AFP of the GFAO random access protocol, which is given by

$$\begin{aligned} P_{err} = &[1-(1-\frac{1}{\tau})^{K_a-1}] \\ &\times \sigma Q(\frac{\sqrt{N\tau}(G^* - \beta_0(N\tau)^{-2/3} - G)}{\sqrt{\alpha_0^2 + G}}), \end{aligned} \tag{21}$$

where the first term in Eq. (21) is the probability that the i-th UE chooses a collision pilot at the first access time slot, which would perform pilot reselection from $\mathbf{R}$ and access to HTS. The second term in Eq. (21) is the probability that a UE fails to be resolved by HTS in this transmission frame, denoted by ε. We define $G = \frac{K_a}{\tau N}$ as the average number of UEs that choose a pilot per access time slot, G^* is the related decoding threshold, and α_0, β_0, σ are suitable scaling parameters of ε. A brief derivation of ε is introduced as follows.

Note that the SIC in GFAO random access protocol can be regarded as the BP decoding in erasure code [45], and the authors in [46] propose a generalized finite length analysis method for the BP decoding in erasure code. Let P_{BEP} denote the block error probability after BP decoding in (n, m) erasure code with erasure probability δ, which can be expressed as

$$P_{\text{BEP}} = Q(\frac{\sqrt{n}(\delta_r^* - \beta_r n^{-2/3} - \delta)}{\alpha}) + O(n^{-1/3}), \tag{22}$$

where $Q(\cdot)$ is the tail probability of standard normal distribution, $r = (n-m)/n$ represents the nominal rate. α_r and β_r are constant values that related to r, and $\alpha = \sqrt{\alpha_r^2 + \delta(1-\delta)}$. Moreover, let δ^* denote the threshold of BP decoding, and $P_{\text{BEP}} \to 0$ if $\delta < \delta^*$ and else $P_{\text{BEP}} \to 1$.

Similarly, let G^* denote the SIC decoding threshold, which leads to $\varepsilon \to 0$ if $G < G^*$ and else $\varepsilon \to 1$. Considering the mapping relationship between the SIC in our GFAO random access protocol and BP decoding in erasure code, and G can be expressed as a function of δ and m as

$$G = \frac{K_a}{N\tau} = \frac{\delta n}{m}. \tag{23}$$

Further, the SIC decoding threshold G^* can be obtained by substituting δ^* into Eq. (23) as follows:

$$G^* = \frac{\delta^* n}{m} = \frac{\delta^*}{1-r}. \tag{24}$$

Then, by substituting Eq. (23) and Eq. (24) into Eq. (22), we can finish the derivation of ε as follows,

$$\varepsilon = \sigma Q(\frac{\sqrt{N\tau}(G^* - \beta_0(N\tau)^{-2/3} - G)}{\sqrt{\alpha_0^2 + G}}), \tag{25}$$

where α_0, β_0 and σ are computed by the density evolution method for $r = 0$.

Finally, recall that we can estimate the system AAoI Δ by exploiting the AAoI of the i-th UE Δ^i as shown in Eq. (18), and we have

$$\Delta = \frac{\Lambda}{\Psi} + 1, \tag{26}$$

where $\Lambda = \left[(N-1)\mathrm{E}\left[X_m^i\right] + \frac{1}{2}\mathrm{E}\left[(X_m^i)^2\right] + \frac{1}{2}\mathrm{E}\left[(\Gamma_m^i)^2\right]\right]$, and $\Psi = \mathrm{E}\left[Y_m^i\right]$. By substituting Eq. (19) into Λ and Ψ, we have

$$\begin{aligned} \Lambda = &\frac{N(3N-2)}{2} + \frac{N(3N-2)}{2}\frac{1}{\lambda(1-P_{err})} \\ &+ N^2\frac{1}{(\lambda(1-P_{err}))^2} - \frac{N(N-1)}{\lambda(1-P_{err})}\frac{P_{nc}}{1-P_{err}} \\ &- \frac{(3N-1)(N-1)}{2}\frac{P_{nc}}{1-P_{err}} + \frac{(N-1)^2}{2}\lambda P_{nc}, \end{aligned} \tag{27}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

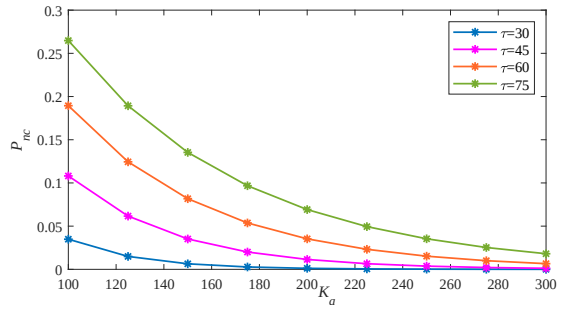

Fig. 7. The impact of number of activated UEs K_a for the probability of non-collision at the first access time slot with different value of τ.

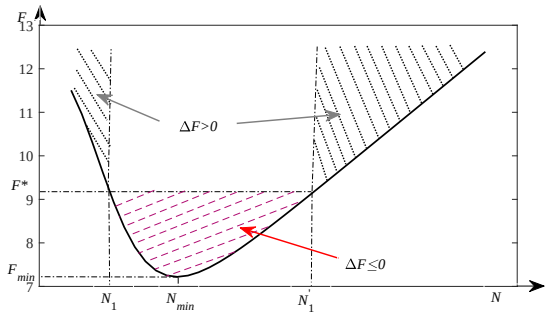

Fig. 8. The diversification of the system AAoI with N access time slots.

and

$$\Psi = N + \frac{N}{\lambda(1-P_{err})} - \frac{(N-1)P_{nc}}{1-P_{err}} + (N-1)\lambda P_{nc}. \quad (28)$$

C. Relationship between System AAoI and Throughput

Let T denote the system throughput, i.e., average sum of recovered UEs in each access time slot of one transmission frame for our GFAO random access protocol, and the relationship between the AFP and system throughput in our GFAO random access protocol can be given by

$$T = \frac{K_a}{N}\left(1 - P_{err}\right). \quad (29)$$

By substituting Eq. (29) into Eqs. (26)-(28), we can obtain an expression of the system AAoI Δ related to the throughput, which can be expressed as Eq. (30) at the top of next page.

IV. Simplification of System AAoI and Optimization of GFAO Random Access Protocol

In this section, we first simplify the expression of system AAoI Δ. Then, to figure out the effects of number of access time slots N and available pilot sequences τ in each transmission frame on Δ, we perform an optimal analysis on N and τ. In addition, the impact on the throughput T is also considered.

A. Simplification of the System AAoI

To further analyze the system AAoI Δ, we conduct some reasonable operations to simplify Eq. (26). First, as shown in Fig. 7, when K_a increases, the non-collision probability P_{nc} in Eq. (20) at the first access time slot in each transmission frame is $P_{nc} \to 0$. Hereby, we assume that $P_{nc} = 0$, and the simplified system AAoI F is given as follows,

$$F = \frac{N}{\lambda\left(1-P_{err}\right)\left(1+\lambda\left(1-P_{err}\right)\right)} + \frac{3N}{2}. \quad (31)$$

Furthermore, we assume that the activated probability of the i-th UE $\lambda = 1$, i.e., the number of activated UEs $K_a = K$. Thus, we can further simplified F as follows,

$$\begin{aligned} F &= \frac{N}{\left(1-P_{err}\right)\left(2-P_{err}\right)} + \frac{3N}{2} \\ &= N\left(\frac{1}{\left(1-P_{err}\right)\left(2-P_{err}\right)} + \frac{3}{2}\right), \end{aligned} \quad (32)$$

where $N/(1-P_{err})$ is the average end-to-end delay.

Similarly, we can also simplify P_{err} and T as:

$$P_{err} = \sigma Q\left(\frac{\sqrt{N\tau}(G^* - \beta_0(N\tau)^{-2/3} - g/N)}{\sqrt{\alpha_0^2 + g/N}}\right), \quad (33)$$

and

$$T = \frac{K}{N}\{1 - \sigma Q(\frac{\sqrt{N\tau}(G^* - \beta_0(N\tau)^{-2/3} - g/N)}{\sqrt{\alpha_0^2 + g/N}})\}. \quad (34)$$

Meanwhile, Eq. (30) can also be simplified as follows,

$$F = \frac{\frac{N(3N-2)}{2} + \frac{(3N-2)K}{2T} + \left(\frac{K}{T}\right)^2}{N + \frac{K}{T}} + 1. \quad (35)$$

B. Minimization of System AAoI

It is straightforward that N and P_{err} are key factors that influence F according to Eq. (32), N and τ are key parameters of P_{err} according to Eq. (33). Recall F in Eq. (32), let $f_1(N) = N$, $f_2(N) = \frac{1}{(1-P_{err})(2-P_{err})} + \frac{3}{2}$, and F can be regarded as a function of N as follows,

$$F(N) = f_1(N) \cdot f_2(N). \quad (36)$$

Thus, we can compute the partial derivation of $F(N)$ to N as Eq. (37) at the top of next page, where $\frac{\partial P_{err}(N)}{\partial N}$ is the partial derivation of $P_{err}(N)$ related to N as in Eq. (33) and given as Eq. (38) at the top of next page, which indicates that $P_{err}(N)$ is a monotonically non-increasing function of N. Thus, by substituting $\frac{\partial P_{err}(N)}{\partial N}$ into $\frac{\partial F(N)}{\partial N}$, we can obtain that $F(N)$ is a non-monotonic function of N, and we have the following Theorem 1.

Theorem 1: There exists an optimal N, denoted by N_{min} to achieve a minimum F_{min} under a given τ, and two conditions are given as follows,

- Supposing $N \geq N_{min}$, then $F(N) \geq F_{min}$;
- Supposing $N < N_{min}$, then $F(N) > F_{min}$.

Proof. There exists any value in N denotes as $N_1 < N_1'$ that $F(N_1) = F(N_1') = F^*$ as shown in Fig. 8. Moreover, we can observe that when $0 < N < N_1$ or $N_1' < N$, we have $\Delta F = F(N) - F^* > 0$, and when $N_1 \leq N \leq N_1'$, we have $\Delta F = F(N) - F^* \leq 0$. Thus, the expression of ΔF is given as Eq. (39) at the top of next page.

Let Δf_A denote the numerator in the first item of Eq. (39). Note that the denominator in the first item of Eq. (39) is larger than 0 as $0 \leq P_{err} < 1$, Thus, two cases can be differentiated as follows.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

$$\Delta = \frac{\frac{N(3N-2)}{2}+\frac{(3N-2)K_a}{2\lambda T}+\left(\frac{K_a}{\lambda T}\right)^2-\frac{(N-1)P_{nc}K_a^2}{N\lambda T^2}+\frac{(3N-1)(N-1)P_{nc}K_a}{2NT}+\frac{(N-1)^2}{2}\lambda P_{nc}}{N+\frac{K_a}{\lambda T}-\frac{N-1P_{nc}K_a}{NT}+(N-1)\lambda P_{nc}}+1. \tag{30}$$

$$\frac{\partial F(N)}{\partial N}=\frac{\partial f_1(N)}{\partial N}f_2(N)+f_1(N)\frac{\partial f_2(N)}{\partial N}=\frac{1}{(P_{err}-1)(P_{err}-2)}-\frac{N}{(P_{err}-1)^2(P_{err}-2)}\frac{\partial P_{err}(N)}{\partial N}-\frac{N}{(P_{err}-1)(P_{err}-2)^2}\frac{\partial P_{err}(N)}{\partial N}+\frac{3}{2}. \tag{37}$$

$$\begin{aligned}\frac{\partial P_{err}(N)}{\partial N}=-\frac{1}{\sqrt{\pi}}\Bigg(\sigma e^{-\frac{N\tau\left(G^*-\frac{\beta_0}{(N\tau)^{\frac{2}{3}}-g/N}\right)^2}{2\left(\alpha_0^2+\frac{g}{N}\right)}}\Bigg(\frac{\left(G^*-\frac{\beta_0}{(N\tau)^{\frac{2}{3}}}-g/N\right)\sqrt{2}\tau}{4\sqrt{N\tau}\sqrt{\alpha_0^2+\frac{g}{N}}}+\frac{\sqrt{N\tau}\left(\frac{2\beta_0\tau}{3(N\tau)^{\frac{5}{3}}}+\frac{g}{N^2}\right)\sqrt{2}}{2\sqrt{\alpha_0^2+\frac{g}{N}}}\\+\frac{\sqrt{N\tau}\left(G^*-\frac{\beta_0}{(N\tau)^{\frac{2}{3}}}-\frac{g}{N}\right)\sqrt{2}g}{4\left(\alpha_0^2+\frac{g}{N}\right)^{3/2}N^2}\Bigg)\Bigg)\le 0.\end{aligned} \tag{38}$$

$$\Delta F=f_1(N)\cdot f_2(N)-f_1(N_1)\cdot f_2(N_1)=\frac{N(1-P_{err}(N_1))(2-P_{err}(N_1))-N_1(1-P_{err}(N))(2-P_{err}(N))}{(1-P_{err}(N))(2-P_{err}(N))(1-P_{err}(N_1))(2-P_{err}(N_1))}+\frac{3}{2}(N-N_1). \tag{39}$$

Case a $\Delta f_A > 0$: The condition of $\Delta f_A > 0$ can be exhibited as,

$$N > N_1\frac{(1-P_{err}(N))(2-P_{err}(N))}{(1-P_{err}(N_1))(2-P_{err}(N_1))}, \tag{40}$$

and Eq. (41) at the top of next page, where $|.|$ represents absolute value.

Thus, ${N_1}' = N_1\frac{(1-P_{err}(N))(2-P_{err}(N))}{(1-P_{err}(N_1))(2-P_{err}(N_1))}$. From Eq. (40) we can find out that $F(N) > F^*$ if $N > N_1'$ or $N < N_1$, as the area of $\Delta F > 0$ shown in Fig. 8. Besides, when $N = N_1 = {N_1}'$, F approximately approaches to F^*.

Case b $\Delta f_A \le 0$: Similarly, the conditions of $\Delta f_A \le 0$ can be exhibited as,

$$N_1 \le N \le {N_1}', \tag{42}$$

and Eq. (43) at the top of next page.

As shown in Fig. 8, there exists an optimal value of access time slots from N_1 to N_1' that can achieve the lowest system AAoI F_{min} in the region of $\Delta F \le 0$. Therefore, the minimum system AAoI F_{min} for our GFAO random access protocol can be easily found as the number of access time slots is an integer. □

C. Maximization of System Throughput

Similar to the analysis in Section IV. B, we can achieve a maximum system throughput T_{max} under a certain number of access time slots N_{max} for a given τ according to Eq. (34), and we have the following Theorem 2.

Theorem 2: There exists an optimal N, denoted by N_{max} to achieve a maximum T_{max} under a given τ, and two conditions are given as follows,

- Supposing $N \ge N_{max}$, then $T(N) \le T_{max}$;
- Supposing $N < N_{max}$, then $T(N) < T_{max}$.

Proof. Define T^* as the system throughput under the given number of access time slots N_2 and N_2' ($N_2 \le N_{max} \le N_2'$), and we have $T^* = T(N_2) = T(N_2')$ as shown in Fig. 9. Then, we define the interval between T^* and T as ΔT, and the expression of ΔT is given by,

$$\begin{aligned}\Delta T &= \frac{K(1-P_{err}(N))}{N}-\frac{K(1-P_{err}(N_2))}{N_2}\\&=[N_2-N+NP_{err}(N_2)-N_2P_{err}(N)]\frac{K}{NN_2}.\end{aligned} \tag{44}$$

Let the first item in Eq. (44) is defined as $\Delta f_T = N_2 - N + N\cdot P_{err}(N_2) - N_2\cdot P_{err}(N)$. Consider that $N\cdot N_2 > 0$ and $K > 0$, therefore if and only if $\Delta f_T > 0$, $\Delta T > 0$. Thus, two cases can be discussed in the following.

Case c $\Delta f_T < 0$: To satisfy $\Delta f_T < 0$, we have

$$N > N_2(1-P_{err}(N))/(1-P_{err}(N_2)). \tag{45}$$

Hence, ${N_2}' = N_2(1-P_{err}(N))/(1-P_{err}(N_2))$. According to Eq. (45), we can obtain that when $N > {N_2}'$, the system throughput would degenerate than T^*, corresponding to the

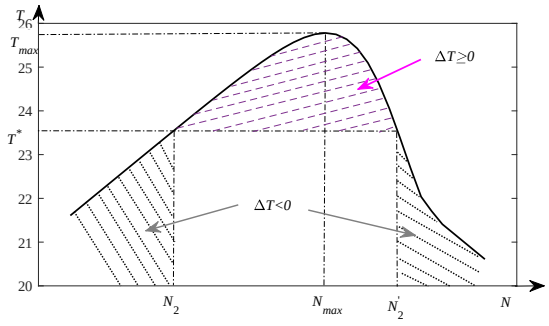

Fig. 9. The diversification of system throughput T with N access time slots.

$$\frac{N\left(1-P_{err}\left(N_1\right)\right)\left(2-P_{err}\left(N_1\right)\right)-N_1\left(1-P_{err}\left(N\right)\right)\left(2-P_{err}\left(N\right)\right)}{\left(1-P_{err}\left(N\right)\right)\left(2-P_{err}\left(N\right)\right)\left(1-P_{err}\left(N_1\right)\right)\left(2-P_{err}\left(N_1\right)\right)}>\frac{3}{2}\left|N-N_1\right|. \tag{41}$$

$$\left|\frac{N\left(1-P_{err}\left(N_1\right)\right)\left(2-P_{err}\left(N_1\right)\right)-N_1\left(1-P_{err}\left(N\right)\right)\left(2-P_{err}\left(N\right)\right)}{\left(1-P_{err}\left(N\right)\right)\left(2-P_{err}\left(N\right)\right)\left(1-P_{err}\left(N_1\right)\right)\left(2-P_{err}\left(N_1\right)\right)}\right|\geq\frac{3}{2}\left(N-N_1\right). \tag{43}$$

region of $\Delta T<0$ in Fig. 9. Besides, when $N=N_2{}'$, $T(N)$ approximately approaches to T^*.

Case d $\Delta f_T\geq 0$: Similarly, the condition to guarantee $\Delta f_T\geq 0$ can be exhibited as

$$N_2\leq N\leq N_2{}'. \tag{46}$$

It is worth noting that the system throughput T can be improved by increasing the number of access time slots from N_2 to N_2', which is corresponding to the region of $\Delta T\geq 0$ in Fig. 9. Thus, the maximum system throughput T_{max} of our GFAO random access protocol can be obtained, and we can express the partial derivative of ΔT as

$$\frac{\partial\Delta T}{\partial N}=\frac{-K(N\frac{\partial P_{err}(N)}{\partial N}+1-P_{err}(N))}{N^2}. \tag{47}$$

Recall that P_{err} is monotonically non-increasing according to Eq. (38), then T_{max} can be achieved at $\frac{\partial\Delta T}{\partial N}|_{N=N_{max}}=0$. Thus, a numerical solution of N in practical communication environment can be easily obtained as N is an integer. □

D. Optimal Number of Activated UEs and Number of Access Time Slots under Required AFP

In this subsection, we first investigate the variable number of activated UE K_a and N under a given AFP P_{err}^* with $0<\lambda<1$. Without loss of generality, we assume that $P_{nc}=0$ since $P_{nc}\to 0$ with large K_a. Thus, the asymptotic AAoI Δ_S can be expressed as Eq. (48) at the top of next page.

Let $x=\frac{K}{K_a(1-P_{err}^*)}$, the above Eq. (48) can be expressed as follows,

$$\Delta_S=\frac{\frac{3N-2}{2}+\frac{3N-2}{2}x+Nx^2}{1+x}+1. \tag{49}$$

Thus, we can obtain a quadratic function with x as follows,

$$\Delta_S=N\left((1+x)+\frac{1}{1+x}\right)-\frac{1}{2}N. \tag{50}$$

Therefore, we can find a lower bound of Δ_S as $\frac{3}{2}N$ since $x>0$. Assume that there exists a minimum AAoI $\Delta_S^*>\frac{3}{2}N$ under P_{err}^* and K_a, then, we can further derive a maximum K_a^* under Δ_S^*. Thus, Eq. (49) with $x^*=\frac{K}{K_a^*(1-P_{err}^*)}$ under Δ_S^* can be rewritten as follows,

$$(x^*)^2+\left(\frac{3}{2}-\frac{\Delta_S^*}{N}\right)x+\left(\frac{3}{2}-\frac{\Delta_S^*}{N}\right)=0. \tag{51}$$

According to the quadratic formula and $\Delta_S^*>\frac{3}{2}N$, the solution of Eq. (51) can be expressed as

$$x^*=\frac{-\left(\frac{3}{2}-\frac{\Delta_S^*}{N}\right)^2+\sqrt{\left(\frac{1}{2}+\frac{\Delta_S^*}{N}\right)^2-4}}{2}. \tag{52}$$

TABLE III
SIMULATION PARAMETERS

Description	Values
System load	$0.2\leq g\leq 3.0$
Altitude from HTS to UE	765 km
Number of predetermined pilots in $\mathbf{P}$	$20<\tau<120$
Number of UEs	$100<K<1000$
Number of access time slots	$1\leq N\leq 6$

Consider that $\sqrt{\left(\frac{1}{2}+\frac{\Delta_S^*}{N}\right)^2-4}>\left(\frac{3}{2}-\frac{\Delta_S^*}{N}\right)^2$ since $x>0$, we have

$$K_a^*=\frac{2K^*}{\left(-\left(\frac{3}{2}-\frac{\Delta_S^*}{N}\right)^2+\sqrt{\left(\frac{1}{2}+\frac{\Delta_S^*}{N}\right)^2-4}\right)\left(1-P_{err}^*\right)}, \tag{53}$$

and the partial derivation of K_a^* related to N can be obtain as Eq. (54) at the top of next page.

Note that $\frac{\partial K_a^*}{\partial N}\geq 0$ since the first and second items are no large than 0 in Eq. (54), which means that under given P_{err}^* and K, K_a^* is a monotonically non-decreasing function of N. With the increasing of N, the number of activated UEs that can access in our S-IoT system is increasing under Δ_S^* with given P_{err}^* and K.

Moreover, similar to the analysis of K_a^*, we can obtain the minimum number of access time slots N^* under Δ_S^*, and the expression of $N^*(K_a)$ is given as follows,

$$N^*(K_a)=\frac{\Delta_S^*\left(1+\frac{K}{K_a(1-P_{err}^*)}\right)}{\left(\frac{K}{K_a(1-P_{err}^*)}\right)^2+\frac{3}{2}\frac{K}{K_a(1-P_{err}^*)}+\frac{3}{2}}. \tag{55}$$

Besides, the partial derivation of $N^*(K_a)$ is related to K_a can be obtained as Eq. (56) at the top of next page, which indicates that $N^*(K_a)$ is a non-monotonic function of K_a. Thus, there exists an optimal K_a to achieve a minimum access delay N^* under the minimum AAoI Δ_S^*.

V. SIMULATION RESULTS AND DISCUSSION

In this section, we present extensive Monte Carlo simulations to verify our theoretical analysis, and the simulation results are average values over 10,000 iterations. Since a transmission frame could be able to support 6 access time slots in this S-IoT system, the simulation range of N is from 1 to 6, the main simulation parameters are summarized in Table III. In addition, the possible operations to optimize the system AAoI based on simulation results are discussed.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

$$\Delta_S = \frac{\frac{N(3N-2)}{2} + \frac{N(3N-2)}{2}\frac{K}{K_a(1-P^*_{err})} + N^2\left(\frac{K}{K_a(1-P^*_{err})}\right)^2}{N + N\frac{K}{K_a(1-P^*_{err})}} + 1 = \frac{\frac{3N-2}{2} + \frac{3N-2}{2}\frac{K}{K_a(1-P^*_{err})} + N\left(\frac{K}{K_a(1-P^*_{err})}\right)^2}{1 + \frac{K}{K_a(1-P^*_{err})}} + 1. \tag{48}$$

$$\frac{\partial K^*_a}{\partial N} = \frac{-2K\left(-\left(\frac{3}{2} - \frac{\Delta^*_S}{N}\right)^2 + \sqrt{\left(\frac{1}{2} + \frac{\Delta^*_S}{N}\right)^2 - 4}\right)^{-2}}{(1 - P^*_{err})}\left(-2\left(\frac{3}{2} - \frac{\Delta^*_S}{N}\right)^{-3}\frac{\Delta^*_S}{N^2} - \sqrt{\left(\frac{1}{2} + \frac{\Delta^*_S}{N}\right)^2 - 4}\cdot\left(\frac{1}{2} + \frac{\Delta^*_S}{N}\right)\frac{\Delta^*_S}{N^2}\right). \tag{54}$$

$$\frac{\partial N^*(K_a)}{K_a} = \frac{\Delta^*_S\left(-\frac{K}{K^2_a(1-P^*_{err})}\right)}{\left(\frac{K}{K_a(1-P^*_{err})}\right)^2 + \frac{3}{2}\frac{K}{K_a(1-P^*_{err})} + \frac{3}{2}} - \frac{\Delta^*_S\left(1 + \frac{K}{K_a(1-P^*_{err})}\right)\left(-2\frac{K^2}{K^3_a(1-P^*_{err})^2} - \frac{3}{2}\frac{K}{K^2_a(1-P^*_{err})}\right)}{\left(\left(\frac{K}{K_a(1-P^*_{err})}\right)^2 + \frac{3}{2}\frac{K}{K_a(1-P^*_{err})} + \frac{3}{2}\right)^2}. \tag{56}$$

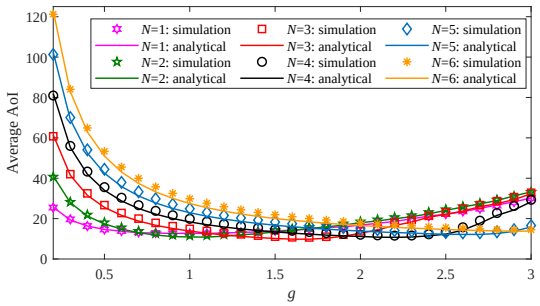

Fig. 10. The variation of AAoI derived in section IV over the system load g in GFAO random access protocol for different number of slots N with $\tau = 50$.

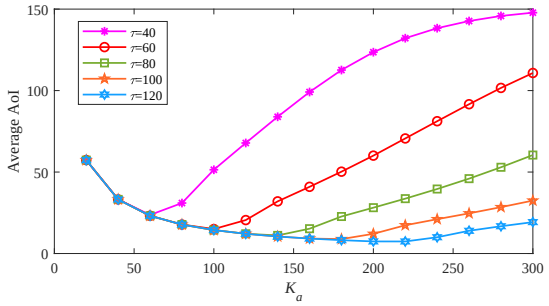

Fig. 11. The AAoI of GFAO random access protocol versus the number of activated UEs K_a for different number of available pilots τ with $N = 3$.

TABLE IV
THE SEGMENTATION AND CORRESPONDING N FOR $K = 300$ TO ACHIEVE THE SMALLEST AAOI

Group	System load g	Number of slots N	System AAoI Δ
1	$0.2 \leq g \leq 0.7$	1	13.1816
2	$0.7 < g \leq 1.2$	2	10.8473
3	$1.2 < g \leq 1.9$	3	9.8104
4	$1.9 < g \leq 2.5$	4	10.7253
5	$2.5 < g \leq 2.9$	5	12.1824
6	$g > 2.9$	6	14.4634

Fig. 10 illustrates the variation of system AAoI over the system load g in our GFAO random access protocol, it can be observed that the simulation results agree well with their theoretical results in Eq. (26) under different parameters. Moreover, we can conclude from Fig. 10 that: First, when the system load g increases, all curves of system AAoI under different N is convex, where Δ is high in the low system load region, then obviously decreases to a lowest Δ, which are almost the same under different N. Further, the system AAoI keeps increasing when their system load increases. Second, we can observe that under a certain system load, there exists a minimum AAoI and its corresponding number of access time slots. For example, under $g = 1.5$, the minimum $\Delta = 9.8104$ and its corresponding $N = 3$. Third, it can be observed that when $0 < g < 0.7$, Δ grows when N increases, and the smallest Δ appears under $N = 1$, which indicates that the conventional random access schemes can work out under low system load. When $g = 0.7$, the overlap among the curve $N = 1$ and the curve $N = 2$ appears, and the overlap among $N = 2$ and $N = 3$ appears at about $g = 1.2$. Also, when $0.7 < g \leq 1.2$, the minimum Δ appears under $N = 2$. Besides, the overlap among $N = 3$ and $N = 4$ appears at $g = 1.9$, and when $1.2 < g \leq 1.9$, the minimum AAoI appears under $N = 3$, and so on. Therefore, the system load $g \in (0, 3.0]$ can be divided into five regions as shown in Table IV. Thus, our GFAO random access protocol can achieve a minimum Δ under different g by selecting an optimal N.

Fig. 11 shows the AAoI Δ versus number of activated UEs for different number of available pilots with $N = 3$. First, we can find that different K_a and τ achieves their minimum Δ at the range of $1.2 < g = K_a/\tau \leq 1.9$, which also agrees well with the result of Fig. 10 as $N = 3$. Second, the minimum Δ decreases with the increasing of τ and K_a, which validates the efficiency of our GFAO random access protocol in massive

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2022.3171450, IEEE Transactions on Communications

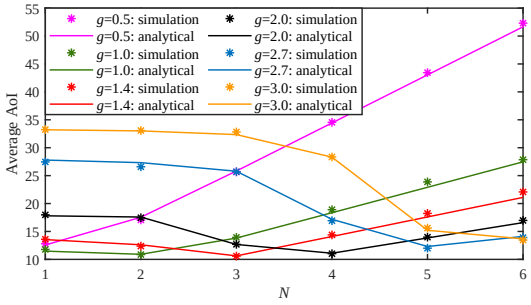

Fig. 12. The AAoI of GFAO random access protocol versus the number of slots N for different system load g.

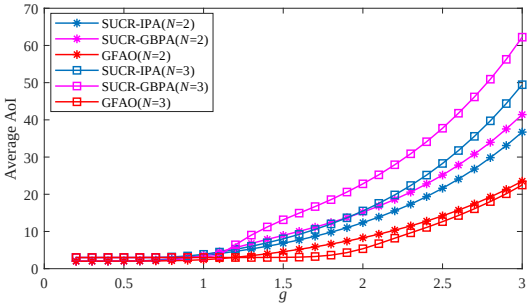

Fig. 14. The AAoI comparison between GFAO, SUCR-GBPA and SUCR-IPA with $\lambda = 1$.

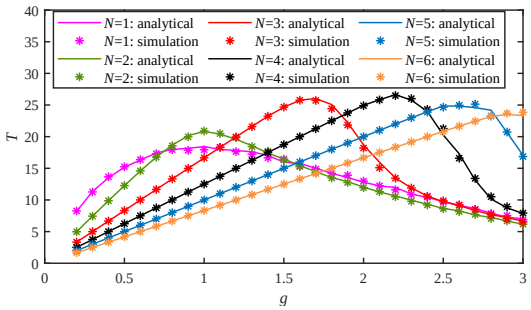

Fig. 13. The variation of throughput derived in section IV over system load g in GFAO random access protocol for different number of slots N with $\tau = 50$.

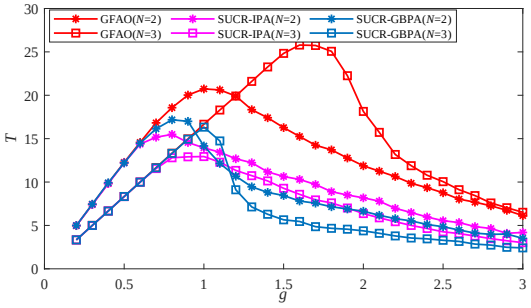

Fig. 15. The throughput comparison between GFAO, SUCR-GBPA and SUCR-IPA.

access.

Furthermore, some optimal operations can be conducted according to Table IV to lower the AAoI via adjusting the number of access time slots in different region of g as shown in Fig. 12. We can observe that the curves of Δ corresponding to different g, a minimum AAoI can be achieved under different N, which matches the result in Table IV.

Fig. 13 shows that the throughput T versus system load g under different number of slots N. As shown in Fig. 13, we can obtain that under different region of g, there exists a certain N, which can make the T achieve a maximum value. The phenomenon is similar to that in Fig. 10 because of the conclusion that by exploiting the GFAO random access protocol, the AAoI and the throughput can achieve their optimal value in different region of g since $g = \frac{K_a}{\tau}$ with a given τ.

Fig. 14 and Fig. 15 show the AAoI Δ and throughput T comparison between the GFAO, SUCR-IPA [32] and SUCR-GBPA [33] random access protocols versus the system load g under $N = 2$ and $N = 3$, where $\tau = 60$. Note that to be fairly comparison, we have revised the SUCR-IPA and SUCR-GBPA as the grant free protocol, which can upload their status packets in the first access time slot. It can be observed that our GFAO random access protocol outperforms the SUCR-IPA and SUCR-GBPA random access protocols under $N = 2$, i.e., the original setup of access time slots in these grant-free random access protocols. In the conventional system load region as $0.2 \leq g \leq 0.7$, our GFAO has similar Δ and T with the compared random access protocols, because the system has enough idle pilots to support mMTC. When $g > 0.7$, our GFAO has higher T than the compared random access protocols as shown in Fig. 15, which validates that our GFAO is more suitable in mMTC with large g. In the high system load region, since the idle pilots is significant declining when $g > 1$ and leads to the increasing of AFP P_{err}, and thus deteriorates the performance of Δ and T.

Finally, Fig. 16 shows that the maximum number of activated UEs K_a^* versus minimum number of access time slot N^* under minimum AAoI Δ_S^* with given AFP P_{err}^*, thus the expected number of accessed UEs is $K_a^*(1 - P_{err}^*)$. Compare with state-of-art schemes, our GFAO random access protocol has almost double higher $K_a^*(1 - P_{err}^*)$ than the SUCR-IPA and SUCR-GBPA random access protocols in the overload region with the increasing of N^*, which validates that the GFAO random access protocol has significant improvement for massive access in S-IoT system.

VI. Conclusions

In this paper, we proposed a GFAO random access protocol to guarantee the timely status updating for massive access in S-IoT system. By exploiting the new performance indicator named AoI, we evaluated the instantaneous AoI and derived the theoretical expressions of system AAoI and throughput to the GFAO random access protocol via the Markov analysis. Further, we investigated the relationship between the AAoI and number of access time slots in each transmission frame, and showed that the AAoI was a non-monotonic function of the number of access time slots, pointing out the existence of a minimum number of access time slots which can optimize

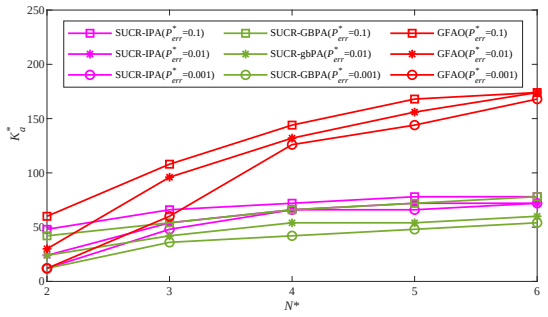

Fig. 16. The number of activated UEs that can access the system under minimum AAoI Δ_S^* with given AFP P_{err}^* and $\tau = 60$.

the system AAoI. Extensive Monte Carlo simulation results were conducted to validate that the minimum AAoI and the maximum throughput can be obtained by selecting the optimal access time slots in diversity system load regions. Our future work would extend our GFAO random access protocol to the unequal AoI requirement coexistence services and other more complex emerging systems.

REFERENCES

[1] H. Yao, L. Wang, X. Wang, Z. Lu and Y. Liu, "The space-terrestrial integrated network: An overview", *IEEE Commun. Mag.*, vol. 56, pp. 178–185, Sep. 2018.

[2] J. Liu, Y. Shi, Z. M. Fadlullah and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 4, pp. 2714–2741, 2018.

[3] Z. Lin, et al., "Supporting IoT with rate-splitting multiple access in satellite and aerial integrated networks," *IEEE Internet of Things Journal*, early access, doi: 10.1109/JIOT.2021.3051603.

[4] I. Akyildiz, A. Kak. "The internet of space things/cubesats," *IEEE Network*, vol. 33, no. 5, pp. 212–218, 2019.

[5] J. Jiao, Y. Sun, S. Wu, Y. Wang and Q. Zhang, "Network utility maximization resource allocation for NOMA in satellite-based internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3230–3242, 2020.

[6] Space Exploration Technologies. SpaceX non-geostationary satellite system Attachment A: "technical information to supplement Schedule S. https://licensing.fcc.gov/myibfs/download.do?attachment key= 1158350," Nov. 2016.

[7] T. Ho, T. Tran, T. Nguyen, et al., "Next-generation wireless solutions for the smart factory, smart vehicles, the smart grid and smart cities," [Online]. Available: https://arxiv.org/abs/1907. 10102.

[8] S. Cioni, R. Gaudenzi, O. Herrero, et al., "On the satellite role in the era of 5G massive machine type communications," *IEEE Network*, vol. 33, no. 5, pp. 54–61, 2018.

[9] D. Li, S. Wu, Y. Wang, J. Jiao, and Q. Zhang,"Age-optimal HARQ design for freshness-critical satellite-IoT systems," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2066–2076, 2020.

[10] O. Ayan, H. Murat Grsu, A. Papa and W. Kellerer, "Probability analysis of age of information in multi-hop networks," *IEEE Networking Letters*, vol. 2, no. 2, pp. 76-80, June 2020.

[11] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proceedings of 8th Annual IEEE Commun. Society Conf. on Sensor, Mesh and Ad Hoc Commun. and Networks*, pp. 350–358, June. 2011.

[12] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proceedings of IEEE Conf. on Computer Commun. (INFOCOM)*, pp. 2731–2735, Mar. 2012.

[13] B. Zhou, W. Saad, "Minimum age of information in the internet of things with non-uniform status packet sizes," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1933–1947, March 2020.

[14] A. Gong, T. Zhang, H. Chen and Y. Zhang, "Age-of-information-based scheduling in multiuser uplinks with stochastic arrivals: A POMDP approach,"*GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, 2020.

[15] A. Maatouk, M. Assaad and A.Ephremides, "On the age of information in a CSMA environment," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 818–831, April 2020.

[16] A. Baiocchi and I. Turcanu, "Age of information of one-hop broadcast communications in a CSMA network," *IEEE Communications Letters*, vol. 25, no. 1, pp. 294–298, Jan. 2021.

[17] A. Munari, A. Frolov, "Average age of information of irregular repetition slotted ALOHA," [Online]. Available: http://arxiv.org/abs/2004.01998.

[18] O. T. Yavascan and E. Uysal, "Analysis of slotted ALOHA with an age threshold," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1456–1470, 2021.

[19] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 148–156, 2020.

[20] R. Giuliano, F. Mazzenga and A. Vizzarri, "Satellite-based capillary 5G mMTC networks for environmental applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 10, pp. 40–48, Oct. 2019.

[21] Z. Zhang, et al., "User activity detection and channel estimation for grant-free random access in LEO satellite-enabled Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8811-8825, 2020.

[22] Z. Zhang, Y. Li, C. Huang et al., "DNN-aided block sparse Bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12000-12012, 2019.

[23] Y. Yuan, Z. Yuan, G. Yu, C. Hwang, P. k. Liao, A. Li, and K. Takeda, "Non-orthogonal transmission technology in LTE evolution," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 68–74, 2016.

[24] B. Zhao, et. al, "Spatial group based optimal uplink power control for random access in satellite networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7354–7365, 2020.

[25] B. Di, H. Zhang, L. Song, Y. Li, and G. Y. Li, "Ultra-dense LEO: Integrating terrestrial-satellite networks into 5G and beyond for data offloading," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 47–62, Jan. 2019.

[26] S. Wang, Y. P. Hong, "Transmission control with imperfect CSI in channel-aware slotted ALOHA networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 5214–5224, October 2009.

[27] R. Ma, V. Misra and D. Rubenstein, "An analysis of generalized slotted-aloha protocols," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 936–949, June 2009.

[28] E. Paolini, C. Stefanovic, G. Liva, and P. Popovski, "Coded random access: Applying codes on graphs to design random access protocols," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 144–150, 2015.

[29] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 477–487, 2011.

[30] C. Stefanovic, P. Popovski, and D. Vukobratovic, "Frameless ALOHA protocol for wireless networks," *IEEE Communications Letters*, vol. 16, no. 12, pp. 2087–2090, 2012.

[31] E. Björnson, E. De Carvalho, J. H. Sørensen, E. G. Larsson, and P. Popovski, "A random access protocol for pilot allocation in crowded massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2220–2234, 2017.

[32] H. Han, X. Guo, and Y. Li, "A high throughput pilot allocation for M2M communication in crowded massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9572–9576, 2017.

[33] H. Han, Y. Li, and X. Guo, "A graph-based random access protocol for crowded massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7348–7361, 2017.

[34] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 4, pp. 3098–3130, 2018.

[35] O. Kodheli, A. Guidotti and A. Vanelli-Coralli, "Integration of satellites in 5G through LEO constellations," in *2017 IEEE Global Communications Conference (GLOBECOM 2017)*, pp. 1–6, 2017.

[36] 3GPP, "System architecture for the 5G system," TS 23.501 V17.2.0, 2021.

[37] N. -N. Dao et al., "Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1193–1225, 2021.

[38] 3GPP, "Study on using satellite access in 5G," TR 22.822 V16.0.0, 2018.

[39] T. Yang, J. Jiao, L. Xu, S. Wu and Q. Zhang, "Age-optimal multi-slot pilot allocation random access protocol for S-IoT," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2021.

[40] J. Jiao, J. Zhou, S. Wu and Q. Zhang, "Superimposed pilot code-domain NOMA scheme for satellite-based internet of things," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2732–2743, June 2021.

[41] E. Casini, R. De Gaudenzi, and O. del Rio Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1408–1419, April 2007.

[42] A. Abdi, et. al, "A new simple model for land mobile satellite channels: first-and second-order statistics," *IEEE Transactions on Wireless Communications*, vol. 2, no. 3, pp. 519–528, 2003.

[43] L. You, K. -X. Li, J. Wang, X. Gao, X. -G. Xia and B. Ottersten, "Massive MIMO transmission for LEO satellite communications," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1851–1865, Aug. 2020.

[44] D. Goto, H. Shibayama, F. Yamashita and T. Yamazato, "LEO-MIMO satellite systems for high capacity transmission," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[45] A. Amraoui, A. Montanari, and R. Urbanke, "Analytic determination of scaling parameters," in *2006 IEEE International Symposium on Information Theory*, 2006, pp. 562–566.

[46] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 473–498, 2009.

[47] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted aloha," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 477–487, 2010.