Modeling and Performance Optimization of Slotted Aloha with Successive Transmission

Weilong Zhu, Wen Zhan, Minjie Liu, Jiyun Qiu
School of Electronics and Communication Engineering, Shenzhen Campus of Sun-Yat sen University, Shenzhen, China
Email:{zhuwlong5, liumj78}@mail2.sysu.edu.cn, {zhanwen6, qiujy8}@mail.sysu.edu.cn

2024 16th International Conference on Wireless Communications and Signal Processing (WCSP) | 979-8-3503-9064-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/WCSP62071.2024.10827702

***Abstract*—Enabling massive access and data transmission in Internet of Things (IoT) scenarios remains a persistent challenge. An efficient medium access control (MAC) scheme is essential for addressing this issue. This paper proposes a MAC scheme named Slotted Aloha with Successive Transmission (SAST), where upon successful transmission of the Head-of-Line (HoL) packet, the node delivers remaining packets with probability 1 until a collision occurs, thereby capitalizing on immediate channel availability. By formulating a vacation queuing model, the access/data throughput and access delay are explicitly characterized and optimized by properly choosing the HoL packet transmission probability. It reveals that SAST achieves a maximum data throughput of 0.5, 37% higher than e^{-1} in classic slotted Aloha. Practical insights are alos demonstrated through the example of 2-step small data transmission (SDT) random access in 5G. SAST can be seamlessly implemented into 5G, and compared to 2-step SDT scheme, SAST improves throughput performance while significantly reducing signaling overhead.**

***Index Terms*—Aloha, random access, successive transmission, vacation queuing model, 5G**

I. Introduction

With the emergence of the Internet of Things (IoT) concept, it has developed into an extensive networks including devices ranging from small sensors to smartphones, autonomous cars and even satellites. This expansive network has a profound impact on various application domains, including unmanned aerial vehicles (UAVs), industrial wireless sensor networks (IWSNs), environmental monitoring and others [1]–[3]. However, with the continuous evolution and progress of technology, in some complex IoT scenarios with diverse traffic such as smart factory [4], a key challenge is to ensure compatibility for the simultaneous access of a massive number of diverse IoT devices, while also accommodating the transmission of long data packets.

To accommodate the diverse traffic in IoT networks, random access schemes have emerged as a viable and adaptable solution due to simplicity and flexibility. In a random access framework, devices autonomously and independently decide when to initiate communication on the shared channel. Despite the proliferation of various random access methodologies, they can be fundamentally classified into two categories based on the establishment of a connection: connection-based scheme and the packet-based scheme.

For connection-based scheme, devices initiate by sending a request (typically much smaller than a data packet). It proceeds to the data transmission process in a collision-free manner only after receiving an acknowledgment from the receiver. This scheme is well-suited for scenarios where packets arrive frequently, and the data packet length significantly exceeds that of the request, e.g., LTE [5] and the RTS/CTS (Request To Send/Clear To Send) mechanism in WiFi [6]. Nevertheless, within the realm of IoT communications, the prevalent use of small data transmissions has made the connection-based random access method less effective, primarily due to the excessive overhead associated with establishing connections. Unlike the connection-based scheme, the packet-based scheme enables devices to send packets immediately without the need for connection setup, with overhead solely dependent on the packet's size. To facilitate Small Data Transmissions (SDT), the 3GPP has integrated data transmission into the random access process, introducing the 2/4-step SDT procedures in Release 17 [7], which have been shown to significantly reduce signaling overhead and energy efficiency [8].

Despite the significant advancements, the 2/4-step SDT random access schemes , inherently similar to the classic Aloha framework, often encounter performance constraints in response to access demands and the transmission of larger packets. This leads to reduced throughput and an inevitable onset of network congestion that is challenging to mitigate. The fundamental reason lies in the insufficient applicability of current random access mechanisms when faced with the complicated data traffic characteristics of IoT scenarios. Considerable efforts have been invested in enhancing and refining the classic slotted Aloha network. A prominent solution, grounded in an advanced receiver structure is developed, empowering devices with capability of Multi Packet Reception (MPR) [9]. To implement MPR, strategies based on Successive Interference Cancellation (SIC) [10]–[12] and Non-Orthogonal Multiple Access (NOMA) [13]–[15] have been suggested. However, these approaches necessitate significant alterations to the PHY layer, which may be impractical for cost-sensitive IoT deployments.

In this paper, a new and scalable random access scheme named Slotted Aloha with Successive Transmission (SAST) is proposed. In the SAST scheme, the nodes can transmit the Head-of-Line (HoL) packet according to the 2 step SDT

The work was supported in part by The Shenzhen Science and Technology Program (No.RCBS20210706092408010), in part by National Natural Science Foundation of China under Grant 62001524, in part by the Fundamental Research Funds for the Central Universities, Sun Yat-sen University, under Grant 24pnpy204, in part by the Open Fund of State Key Laboratory of Satellite Navigation System and Equipment Technology (No. CEPNT-2021KF-04).

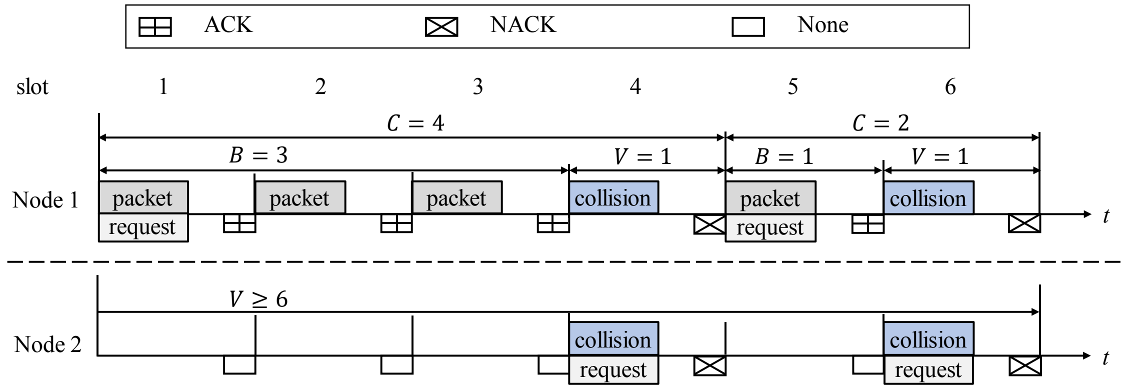

Fig. 1. Illustration of a two-node network with SAST scheme, where B, V, and C denote the busy period, vacation period, and cycle time of node, respectively.

standard. Once the HoL packet is successfully transmitted, the remaining packets in the node's buffer are transmitted successively with probability 1 until a collision occurs. The motivation for successive data transmission is to capitalize on immediate channel availability. The main contributions are summarized as follows:

- We propose and analyze Slotted Aloha with Successive Transmission (SAST). The maximum data throughput of SAST scheme is 0.5, obtaining 37% performance gain compared to classic slotted Aloha without any modification to the PHY layer.
- Based on 3GPP MAC specifications, we compare the signaling performance of SAST with 2-step SDT scheme from the perspective of the signaling-to-throughput Ratio (STR). When subjected to identical STR conditions, the SAST protocol exhibits a data throughput improvement of 23.65% over the 2-step SDT scheme.

II. System Model and Preliminary analysis

We consider a slotted Aloha network comprising n homogeneous nodes and a central server, where only the uplink from the nodes to the server is considered. A saturated system is assumed in which all nodes always have packets to transmit over a noiseless channel. At the beginning of each time slot, each node can independently decide whether to attempt channel access, and the nodes will be informed of their access's success (if attempted) by the end of that time slot. Each slot only allows to transmit at most one packet. The collision model is assumed, i.e., each packet can be transmitted successfully only when there is no concurrent transmission.

The Slotted Aloha with Successive Transmission (SAST) scheme can be elaborated as follows. Specifically, if a node attempt to access the channel with probability $q \in (0, 1]$, it directly transmits the Head-of-Line (HoL) packet to the server. If the server decodes the packet successfully, an acknowledgment (ACK) message will be sent to the node, indicating successful access. Otherwise, a negative acknowledgment (NACK) message is sent to indicate a failed transmission. Here, it is assumed that the ACK/NACK transmission is instantaneous and collision-free. With the successful transmission of the HoL packet, the node will deliver the remaining packets in the

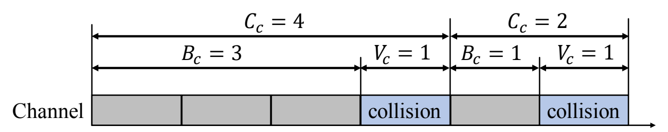

Fig. 2. Illustration of the work process of aggregate channel, where B_c, V_c, and C_c denote the busy period, vacation period, and cycle time of channel, respectively.

subsequent slots with probability 1 until a collision occurs. For the receiver, it replies with an ACK slot by slot, or a NACK when a collision occurs. The HoL packet can be regarded equivalent to the initial channel access request, and the HoL packet and access request are used interchangeably throughout the following discussion.

A. Vacation Queuing Model for SAST

Intuitively, nodes and the aggregate channel with SAST scheme switch between transmission and non-transmission states, which can be formulated using vacation queuing model. In this subsection, it will be characterized in detail from the perspective of node and that of channel, respectively.

From the perspective of node, Fig. 1 illustrates the contention process of a two-node case with SAST scheme. Considering Node 1 as an instance node, Node 1 sends its HoL packet (also an access request) at slot 1 with probability q. With the successful transmission, Node 1 delivers the subsequent packets one by one with probability 1 until a collision occurs at slot 4. This collision is caused by the concurrent transmission from Node 2. At slot 5, Node 1 re-initiates access request and successfully transmits the second HoL packet. However, another collision occurs at slot 6. Intuitively, the working process of node 1 keeps repeating two alternating periods: busy period, and vacation period. Denote B_i as the busy period of node i, during which it transmits its packets successfully, where $i \in \{1, 2, \ldots, n\}$. Denote V_i as the vacation period of node i, during which it is idle, or transmit a packet but failed. Denote C_i as a cycle period of node i, which is the duration between two consecutive time points that node i starts transmission, where $C_i = B_i + V_i$. As a homogeneous scenario is considered, we drop the subscript i for simplicity.

From the perspective of channel, as illustration in Fig. 2, the busy period and vacation period of channel can also be defined. The busy period of channel, denoted by B_c, is the time period in which packets are transmitted successfully. Both the channel busy period B_c and the node busy period B describe the transmission process of one node, which results in the same probability mass function. Thus, we use the symbol B to represent them both. The vacation period of channel, denoted by V_c, is the time period in which no packet is transmitted successfully, that is, either the channel is idle or a collision occurs. Accordingly, we define a cycle of the channel, denoted by C_c, as the duration between two consecutive time points that a batch of packets are beginning to be sent over channel, where $C_c = B_c + V_c$.

B. Performance metric

In this paper, the access/data throughput and access delay are considered. Define access throughput λ_{out}^a as the long-term average number of successful access requests (i.e., the HoL packet) per time slot. Define data throughput λ_{out}^d as the long-term average number of successful packets transmitted per time slot. Let $\overline{X}$ denote the mean value of the random variable X. On one hand, in a channel cycle C_c, only one access request can be accepted successfully, i.e., $1/\overline{C}_c$ is the frequency of the successful access requests. On the other hand, given a channel cycle, the packets can be transmitted successfully only in busy periods. Therefore, the proportion of time occupied by the busy period in a channel cycle is the frequency of successful transmission of packets. According to definition of throughput, the access throughput and data throughput can be obtained as

$$\lambda_{out}^a = \frac{1}{\overline{C}_c} = \frac{1}{\overline{B} + \overline{V}_c} \tag{1}$$

and

$$\lambda_{out}^d = \frac{\overline{B}}{\overline{C}_c} = \frac{\overline{B}}{\overline{B} + \overline{V}_c}. \tag{2}$$

When it comes to access delay, it is defined as the long-term mean time length of access requests (i.e., the HoL packet) from generation to acceptance, which is exactly equal to the mean length of node vacation period $\overline{V}$, i.e.,

$$D_A = \overline{V}. \tag{3}$$

As the above performance metrics are all related to the mean length of busy/vacation period, the derivation of the mean length of busy/vacation period will be analyzed in detail based on the system model.

III. Vacation queuing analysis for SAST

Denote Θ as the number of access quests at each slot, including the newly and retransmitted ones and θ is its mean value (θ is also called the attempt rate). In a idle slot, if the number of access requests is given i, then there must be i nodes that send requests with probability q and the other $n-i$ nodes that do not send requests with probability $1-q$. We can get

$$\Pr\{\Theta = i\} = C_n^i\,(1-q)^{n-i}\, q^i, i = 0, 1, \cdots, n. \tag{4}$$

where $C_n^i = \frac{n!}{i!(n-i)!}$. Here, Θ follows a binomial distribution with parameter $\{n, q\}$. Moreover, with a large number of node n and a small probability q, it can be approximately regarded as a Poisson random variable with parameter $\theta = nq$,i.e.,

$$\Pr\{\Theta = i\} = \frac{e^{-\theta}\theta^i}{i!}, i = 0, 1, \cdots, n. \tag{5}$$

Let us first derive the mean length of channel vacation period $\overline{V}_c$ and then the mean length of busy period $\overline{B}$ based on attempt rate θ in (5).

1) Mean length of channel busy period $\overline{B}$: Upon one node/channel enters busy period, then it will stay until a collision occurs. In each slot of the busy period, the probability that any other node sends request is $1 - e^{-\theta}$. If other $n-1$ nodes request transmission with probability $1 - e^{-\theta}$, then a collision occurs in the current time slot, i.e., the busy period ends and vacation period starts. Specifically, the length of busy period $B = 1$ if at least one of the other nodes sends a access request in the consecutive slot with probability $1-e^{-\theta}$, i.e., the the channel/node enters the next vacation period immediately. Otherwise, the node/channel remains in the busy period. By analogy, if $B = i$, the node/channel remains transmitting packets in the successive i slots, and then the node/channel encounters a collision at the $(i+1)$-th slot. The probability of busy period can be obtained as

$$\Pr\{B = i\} = \left(e^{-\theta}\right)^{i-1}(1 - e^{-\theta}), i = 1, 2, \ldots \tag{6}$$

It can be seen that B follows a geometric distribution with parameter $1 - e^{-\theta}$, which can give the mean length of busy period as

$$\overline{B} = \frac{1}{1 - e^{-\theta}}. \tag{7}$$

2) Mean length of channel vacation period $\overline{V}_c$: The channel shifts to vacation period due to a collision. And this collision slot is also the first slot for channel vacation period. With the attempt rate θ, the channel enters a busy period only when one request is transmitted with probability $\theta e^{-\theta}$. Specifically, the length of channel vacation period $V_c = 1$ if another node accesses the channel successfully in the consecutive slot with probability $\theta e^{-\theta}$, i.e., the channel enters the next busy period immediately. Otherwise, channel remains in the vacation period. By analogy, if $V_c = i$, the channel is either idle or encounters a collision in the previous i slots, and then the channel is occupied successfully by only one node at the $(i+1)$-th slot with probability $(1-\theta e^{-\theta})^i \theta e^{-\theta}$, i.e.,

$$\Pr\{V_c = i\} = \left(1 - \theta e^{-\theta}\right)^{i-1}\theta e^{-\theta}, i = 1, 2, \ldots \tag{8}$$

Similar to busy period, V_c also follows a geometric distribution with parameter $\theta e^{-\theta}$, which can give the mean length of channel vacation period as

$$\overline{V}_c = \frac{1}{\theta e^{-\theta}}. \tag{9}$$

3) Mean length of node vacation period $\overline{V}$: If a node shifts to its vacation due to a collision, it re-initiates a access request with probability q. At its first slot of vacation, three cases may occur:

C1: With probability $qe^{-\theta}$, the tagged node transmits packets to the receiver and other $n-1$ nodes do not request transmission. In this case, the node vacation period only holds one slot, i.e., $\overline{V}=1$.

C2: With probability $(1-q)\theta e^{-\theta}$, the tagged node does not transmit with probability $1-q$, and only one of the other $n-1$ nodes transmits successfully. The tagged node remains in vacation period. In this case, the tagged node will compete for the channel until it succeeds entering its next busy period. Before that, it will experience three stages in sequence: (1) one slot that it does not compete; (2) a busy period of another node; (3) a new node vacation period. Accordingly, this node vacation period hold: $\overline{V}=1+\overline{B}+\overline{V}$.

C3: With probability $1-(1-q)\theta e^{-\theta}-qe^{-\theta}$, no one succeeds in the first slot and the tagged node has to compete for the channel at the next time slot. Two stages will be experienced: (1) one free slot; (2) a new node vacation period. Thus, it holds: $\overline{V}=1+\overline{V}$.

Combining these three cases, the mean length of node vacation period can be written as

$$\overline{V}=\frac{e^{\theta}+\theta-1-q\theta}{q(1-e^{-\theta})}. \tag{10}$$

Substituting (5), (7) and (9) into (1) and (2), the access throughput and the data throughput can be rewritten as

$$\lambda_{out}^{a}=\frac{nq\left(1-e^{-nq}\right)}{e^{nq}+nq-1} \tag{11}$$

and

$$\lambda_{out}^{d}=\frac{nq}{e^{nq}+nq-1}. \tag{12}$$

Substituting (5), (10) into (3), the access delay can be rewritten as

$$D_A=\frac{e^{nq}+nq-1-n^2q^2}{q\left(1-e^{-nq}\right)}. \tag{13}$$

Theorem 1: The maximum access throughput of SAST scheme is given by

$$\lambda_{max}^{a}\approx 0.2384, \tag{14}$$

which is achieved if and only if

$$q^{*}\approx 1.2515/n. \tag{15}$$

Proof: Define $f\left(x\right)=\frac{x(e^x-1)}{e^x(e^x+x-1)}$. $f'\left(x\right)=\frac{(1-x)\left(e^{2x}-2e^x+1\right)+x^2}{e^x(e^x+x-1)^2}$. Since $e^x\left(e^x+x-1\right)^2>0$ is always true, here only the positive of $(1-x)\left(e^{2x}-2e^x+1\right)+x^2$ needs to be considered. Define $g(x)=(1-x)\left(e^{2x}-2e^x+1\right)+x^2$. It can be known by numerical calculation that there exists $x_0\approx 1.2515$ such that $g(x)>0$ when $0<x<x_0$ and $g(x)<0$ when $x>x_0$, i.e., x_0 is the maximum value point of function $f(x)$ at $(0,\infty)$. Thus, the maximum value of $f(x)$ at $(0,\infty)$ is $f\left(x_0\right)\approx 0.2384$. Using nq instead of x, the maximum access throughput λ_{max}^{a} in (14) and q^{*} in (15) can be obtained. ■

Theorem 2: The data throughput of SAST scheme is a monotonic decreasing function of nq, and the maximum data throughput is given by

$$\lambda_{max}^{d}=\lim_{nq\to 0}\lambda_{out}^{d}=\lim_{nq\to 0}\frac{nq}{e^{nq}+nq-1}=\frac{1}{2}. \tag{16}$$

Proof:
Define $f\left(x\right)=\frac{x}{e^x+x-1}$. $f'\left(x\right)=\frac{(1-x)e^x-1}{(e^x+x-1)^2}$. Define $g\left(x\right)=\left(1-x\right)e^x-1$. Since $g'\left(x\right)=-xe^x<0$ when $x\in(0,\infty)$, $g(x)$ is monotonically decreasing function at $(0,\infty)$, i.e., $g(x)<g(0)=0$ is always true when $x\in(0,\infty)$. Thus, $f\left(x\right)$ is monotonically decreasing function at $(0,\infty)$ as well. When x approaches 0, its function value approaches $1/2$. Using nq instead of x, the maximum data throughput in (16) can be obtained. ■

Theorem 3: Access delay and Access throughput can be approximated as

$$\lambda_{out}^{a}D_A\approx n. \tag{17}$$

Proof: Multiplying access delay and access throughput, we can get $\lambda_{out}^{a}D_A=n-\frac{(nq)^2}{e^{nq}+nq-1}$. Define $f\left(x\right)=\frac{x^2}{e^x+x-1}$. $f'\left(x\right)=\frac{x(e^x-1)(2-x)}{(e^x+x-1)^2}$. There exists $x_0=2$ such that $f'(x_0)>0$ when $0<x<x_0$ and $f'(x_0)<0$ when $x>x_0$, i.e., x_0 is the maximum value point of function $f(x)$ at $(0,\infty)$. Thus, the maximum value of $f(x)$ at $(0,\infty)$ is $f\left(x_0\right)=\frac{4}{e^2+1}\approx 0.4768$. Thus, given a large number of node $n\gg 0.4768$, the approximate expression in (17) can be obtained. ■

IV. Case Study: 5G Cellular Network

In this section, we explain how the proposed scheme can be used in current 5G cellular network and compared its performance with 2-step SDT scheme by leveraging the signaling-to-throughput Ratio (STR), i.e., the signaling overhead per successful data packet per slot.

A. 2-step SDT in 5G

According to 3GPP specifications [7], with 2-step SDT scheme, each node transmits its small packets in the random access procedure, such that the connection with base station is no longer needed. As shown in Fig. 4a, each backlogged node transmits one data packet along with a preamble in MsgA. If the receiver replies with the Random Access Response (RAR) and the Contention Resolution Response in MsgB, then the node knows whether its MsgA transmission is successful or not. Although the signaling overhead for connection establishment is avoided, the signaling overhead due to failed requests still exists and may increase when the number of nodes is large.

To characterize the signaling overhead in details, denote s (in a unit of bits) as the average size of signaling message exchanged between node and receiver. To be specific, one MsgA or MsgB consists of an average of s bits of signaling.

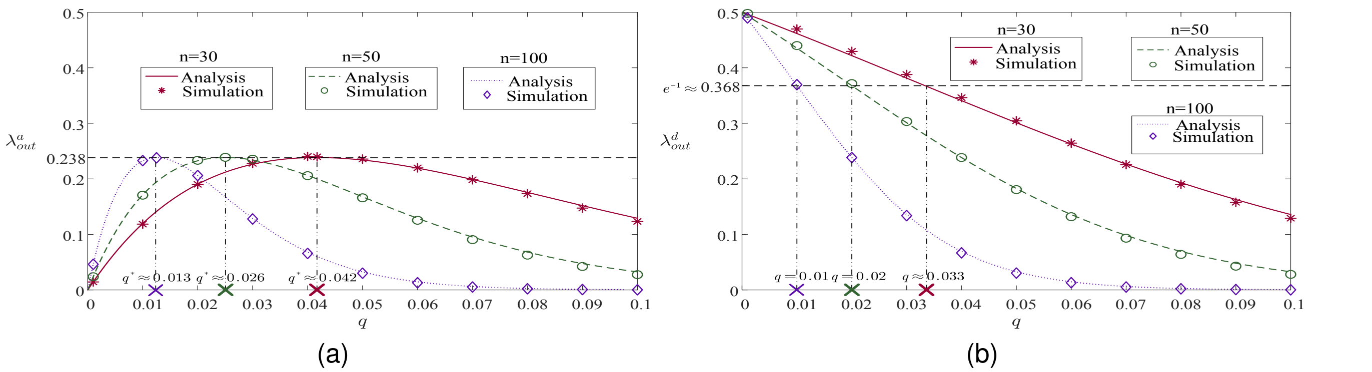

Fig. 3. (a) The access throughput λ_{out}^a, and (b) the data throughput of the SAST scheme λ_{out}^d versus the transmission probability q . $n \in \{30, 50, 100\}$.

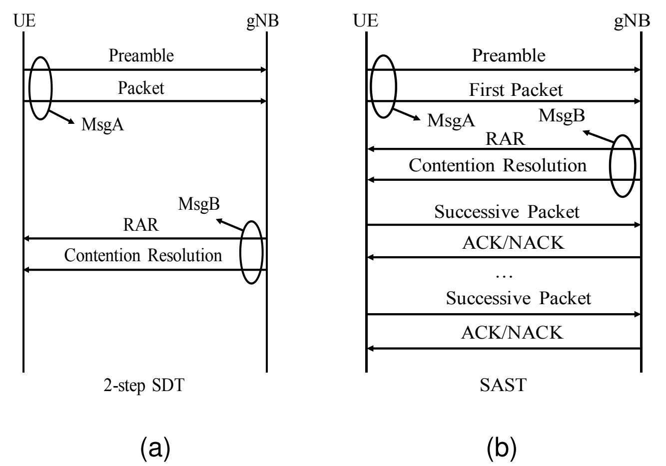

Fig. 4. Signaling change process between user and base station of (a) 2-step SDT scheme, (b)Slotted Aloha with successive transmission scheme.

Denote S as the time-average amount of signaling overhead per time slot. Denote F as the time-average amount of failed access per times slot. In fact, there exists a relationship, i.e., $F + \lambda_{out}^a = \theta = nq$. Denote STR as the ratio of average signaling overhead per time slot to the data throughput, i.e., signaling overhead per successful data packet per slot.

By observing Fig. 4a, we can see that no matter the access request is successful or not, two signaling messages are required. Thus, we have the signaling overhead per time slot in 2-step SDT scheme as $S_{SDT} = 2sF + 2s\lambda_{out}^a = 2s\frac{\lambda_{out}^a}{e^{-nq}}$. On the other hand, the data throughput is equal to the access throughput, i.e., $\lambda_{out}^d = \lambda_{out}^a$. The STR in 2-step SDT scheme can be obtained as

$$\text{STR}_{SDT} = \frac{S_{SDT}}{\lambda_{out}^d} = \frac{2s}{e^{-nq}}. \tag{18}$$

B. Slotted Aloha with Successive Transmission in 5G

As shown in Fig. 4b, similar to 2-step SDT scheme, each node with SAST will experience the same step for the transmission of the HoL packet. Upon the HoL packet is transmitted successfully, the node only needs to transmit the next packet without the preamble. If the receiver replies with an ACK message, then the node can transmit another packet successively. On the contrary, if a NACK message is replied, then the node has to repeat the transmission process. Here, ACK/NACK message or the subsequent failed packet are regarded as signaling overhead.

With one successful access in SAST scheme can transmit $\overline{B}$ packets , the subsequent fail packet that costs extra 2 signaling overheads is considered. Thus, we have the signaling overhead per time slot in SAST scheme as $S_{SAST} = 2sF + \lambda_{out}^a(\overline{B} + 3)s$. The STR in SAST scheme can be obtained as

$$STR_{SAST} = s\left(2e^{nq} + 2nq - e^{-nq}\right). \tag{19}$$

V. Simulation results

In this section, simulation results are presented to verify the above analysis. The simulation setting is the same as the system model described in SectionII, and each simulation is carried out for 10^7 time slots. In simulations, the access/data throughput is obtained by calculating the ratio of the sum of successful access requests/packets to the total time slot.

Fig. 3 depicts how the access throughput λ_{out}^a and data throughput λ_{out}^d vary with the transmission probability q in saturated case with the number of node $n \in \{30, 50, 100\}$. It can be seen in Fig. 3a that when the transmission probability q is small, the access throughput λ_{out}^a increases as q increases because more and more nodes can access to the network and the contention is not serious as well. But if q is large, the access throughput λ_{max}^a decreases because of the mounting channel contention. The maximum access throughput λ_{max}^a can be achieved when the transmission probability q is tuned properly, i.e., $q = q^* \approx 1.2515/n$. Meanwhile, the maximum access throughput $\lambda_{max}^a \approx 0.2384$ is not affected by the number of nodes. As for the data throughput λ_{out}^d in saturated case in Fig. 3b, the analysis and simulation both show that λ_{out}^d decreases as nq increases. Taking the maximum throughput in classic slotted Aloha $1/e$ as the threshold, (12) tells us that with $nq < 1$, the data throughput λ_{out}^d in the saturated case will always larger than $1/e$ and smaller than 0.5. Considering the access throughput λ_{out}^a and data throughput λ_{out}^d jointly, it can be found that the SAST scheme achieves high data throughput with a low access throughput. Intuitively, when the transmission probability q is small, it is difficult for the backlogged nodes to access the channel and transmit packets.

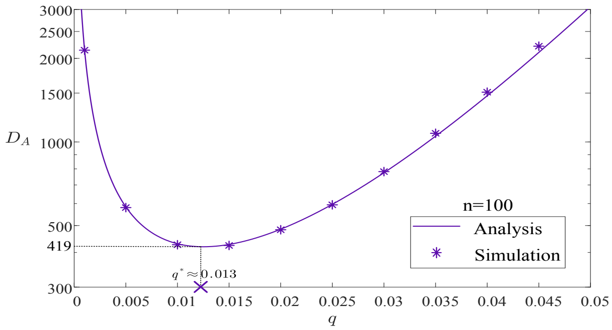

Fig. 5. Mean access delay D_A versus the transmission probability q. $n = 100$.

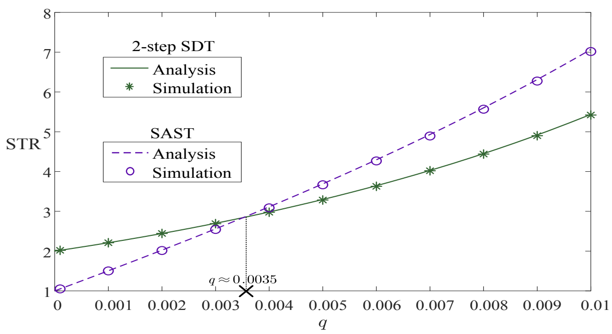

Fig. 6. (a) STR versus the transmission probability q of 2-step SDT and SAST. n=100. s=1.

Once an access request is successful, a large number of packets will be transmitted with a high successful probability, indicating that although the system access throughput is small, a large data throughput can be achieved with a small q.

Fig. 5 depicts how the mean access delay D_A varies with the transmission probability q. It can be observed from Fig. 5 and Fig. 3a that trends of the mean access delay and access throughput shown in (17) are exactly opposite. Therefore, increasing access throughput is equivalent to reducing access delay. Furthermore, by tuning q to maximize access throughput, access delay is also optimized to the minimum.

Fig. 6 depicts how the STR varies with the transmission probability q in saturated case. It can be seen that when the transmission probability $q < \frac{\mathbb{W}_0(0.5)}{n} = 0.0035$ [1], the SAST scheme outperforms 2-step SDT scheme from the perspective of STR (nearly half at most better in the case of $q \to 0$). In particular, when $q = 0.0035$, 2-step SDT and SAST have the same STR performance while according to (12), the data throughput of SAST is $\lambda_{out}^d = 0.4549$, still 23.69% higher than e^{-1}, i.e., the maximum throughput of 2-step SDT scheme.

VI. Conclusion

This paper proposes the SAST scheme for boosting the throughput and the signaling-to-Throughput Ratio (STR) performance of slotted Aloha. By establishing vacation queueing models for characterizing the behavior of both node and channel, the access/data throughput and access delay are derived and further optimized by properly tuning the transmission probability. To illustrate the practical insights of the SAST scheme, the 2-step SDT scheme in 5G cellular network is further considered as benchmark for comparison, where STR of both schemes are derived. The analysis demonstrates that the maximum data throughput of the SAST scheme can reach up to 0.5. Meanwhile, achieving such optimum throughput performance of SAST is simple: Reducing the transmission probability of each node. Moreover, the 5G case study reveals that compared with 2-step SDT scheme, the proposed SAST scheme can achieve a better throughput performance with much lower signaling overhead, indicating the SAST scheme is promising to be used in practical 5G for supporting a broad spectrum of IoT applications with stringent requirement on throughput and energy efficiency.

[1] $\mathbb{W}_0$ is one of the two banches of Lambert W function.

References

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, 2014.
[2] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.
[3] H.-M. Wang, Q. Yang, Z. Ding, and H. V. Poor, "Secure short-packet communications for mission-critical IoT applications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2565–2578, 2019.
[4] P. Castagno, V. Mancuso, M. Sereno, and M. A. Marsan, "A simple model of MTC flows applied to smart factories," *IEEE Trans. Mob. Comput.*, vol. 20, no. 10, pp. 2906–2923, 2020.
[5] W. Zhan and L. Dai, "Massive random access of machine-to-machine communications in lte networks: Modeling and throughput optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2771–2785, 2018.
[6] L. Dai and X. Sun, "A unified analysis of ieee 802.11 dcf networks: stability, throughput, and delay," *IEEE Trans. on Mobile Comput.*, vol. 12, no. 8, pp. 1558–1572, 2013.
[7] NR, *Medium Access Control (MAC) protocol specification (Release 17)*, document TS 38.321 V17.0.0, 3GPP, Mar. 2022.
[8] H. Zhou, Y. Deng, L. Feltrin, and A. Höglund, "Analyzing novel grant-based and grant-free access schemes for small data transmission," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2805–2819, 2022.
[9] J. Liu, J.-B. Seo, and H. Jin, "Online transmission control for random access with multipacket reception and reservation," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 23 160–23 175, Nov. 2022.
[10] E. Casini, R. De Gaudenzi, and O. Herrero, "Contention resolution diversity slotted Aloha (CRDSA): an enhanced random access scheme for satellite access packet networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.
[11] J.-B. Seo, Y. Hu, and H. Jin, "Time-offset Aloha with SIC," *IEEE Trans. on Mobile Comput.*, pp. 1–13, 2023.
[12] Y. Gu, Y. Xu, B. Zhang, Y. Wang, and Z. Yang, "Towards the random multi-access in SIoT: a generalized deduplication based CRDSA mechanism," *IEEE Internet Things J.*, 2024.
[13] J. Choi, "NOMA-based random access with multichannel Aloha," *IEEE J. Select. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.
[14] Y. Yuan, S. Wang, Y. Wu, H. V. Poor, Z. Ding, X. You, and L. Hanzo, "Noma for next-generation massive iot: Performance potential and technology directions," *IEEE Communications Magazine*, vol. 59, no. 7, pp. 115–121, 2021.
[15] C. G. Kang, A. T. Abebe, and J. Choi, "Noma-based grant-free massive access for latency-critical internet of things: A scalable and reliable framework," *IEEE Internet of Things Magazine*, vol. 6, no. 3, pp. 12–18, 2023.