2020 IEEE/CIC International Conference on Communications in China (ICCC)

Maximum Sum Rate of Slotted Aloha for mMTC with Short Packet

Weihua Liu*‡, Xinghua Sun*‡, Wen Zhan*, Xijun Wang†
*School of Electronics and Communication Engineering, Sun Yat-sen University, Guangdong, China
Email: liuwh9@mail2.sysu.edu.cn, sunxinghua@mail.sysu.edu.cn, zhanw6@mail.sysu.edu.cn
†School of Electronics and Information Technology, Sun Yat-sen University, Guangdong, China
Email: wangxijun@mail.sysu.edu.cn
‡Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, 865 Changning Road, Shanghai 200050 China

***Abstract*—Massive machine type communications (mMTC) is one of the key use cases of the fifth-generation (5G). Grant-free access has emerged as a promising technique to reduce the access delay. As one of the representative grant-free schemes, slotted Aloha has recently attracted much attention. This paper focuses on the optimization of sum rate of a slotted Aloha network to achieve the low latency. By deriving the probability of successful transmissions of head-of-line packets, the network sum rate is obtained as explicit function of key system parameters. Based on this analytical expression, the maximum sum rate is derived by optimizing the transmission probabilities of nodes and the blocklength of packets. The effect of the number of information bits per packet and the retry limit on the optimal sum rate performance is characterized. The analysis shows that the retry limit does not affect the maximum sum rate, while a larger number of information bits per packet ameliorates the maximum sum rate in the finite blocklength region.**

***Index Terms*—Aloha, finite blocklength region, network sum rate.**

I. Introduction

Along with the arrival of the era of Internet of Things (IoT), billions of everyday objects (e.g., watches, doors and vehicles) are expected to be connected and join the Internet via radio communication link. For enabling the vision of IoT, Machine Type communications (MTC) [1] has been considered as the key foundation, which can provide pervasive wireless connectivity for autonomous devices with minimum or no human intervention. The large volume of traffic generated by massive MTC devices will have a major influence on the next-generation cellular networks. As such, massive MTC (mMTC) has been identified as one of the main use cases for 5G.

The mMTC traffic has distinctive characteristics compared to the traditional human-type communications traffic [2]. For instance, the number of MTC devices can be very large, e.g., over 10^4 per cell, while each MTC device (e.g., smart meter) usually transmits short packets, which can be only several bits [3]. Coordinating the massive short packet transmissions in a centralized manner is very inefficient and may also induce a great deal of signalling cost that overwhelms the communication system. Accordingly, supporting the massive access from mMTC based on random access protocols is a consensus that gradually comes to being [4]. As one of the representative random access protocols, Aloha and its variations have gained renewed interests recently and been applied in many emerging low power and long range technologies, such as Sigfox and Long Range (LoRa) Radio, for supporting mMTC applications [5]. With Aloha, each node independently decides when to transmit and backs off if its transmission fails.

Despite its simplicity in concept, how to analyze and optimize the performance of Aloha has long been known as notoriously difficult. There has been a long line of research on Aloha, which dates back to Abramson's landmark paper [6] in 70's, in which by modeling the aggregate traffic as a Poisson random variable with parameter G, the network throughput (i.e., the average number of successfully decoded packets per time slot) is obtained as Ge^{-G}, and maximized at e^{-1} when $G = 1$. To achieve the maximum network throughput, various strategies have been developed to adjust the transmission probability of nodes by periodically estimating the channel load [7], [8]. Similar parameter tuning approaches were recently developed for the massive random access of MTC devices in cellular networks based on the realtime/statistical information of the traffic on the random access channel [9]–[13].

In above studies, it is usually assumed that the packet can be successfully decoded at the receiver if and only if there is no concurrent transmission. This assumption can greatly simplify the analysis while implicitly indicates an arbitrarily small packet error probability that can only be achieved with infinite blocklength [14]. However, for short-packet transmissions in mMTC, particularly those low-latency mMTC services, the blocklength is finite and sometimes, small, leading to an nonzero probability that transmissions fail due to decoding error [15].

Note that given number of information bits to be transmitted, each node can choose its blocklength of packets based on the channel conditions, that would affect the sum rate

This work was supported in part by Guangdong Basic and Applied Basic Research Foundation (2019A1515011906), by Sensor Network & Communication (Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Science) under grant 20190901 and 20190912, and by Fundamental Research Funds for the Central Universities under 19lgpy79 and 19lgpy77.

(i.e., the average successfully transmitted information bits per channel use) performance. With a small blocklength, the sum rate performance may degrade due to a large packet error probability. Yet, if the blocklength is too large, each node's information encoding rate becomes small, which also degrades the sum rate performance. Therefore, besides the transmission probability of each node, the blocklength of each packet is also an important system parameter that should be carefully tuned. This leads to new questions: in finite blocklength region, what is maximum sum rate that the Aloha networks can achieve and how to jointly tune the transmission probability of each node and the blocklength of each packet to achieve it?

This paper aims to contribute to the understanding of the above open issues by considering an n-node slotted Aloha network where all the nodes transmit to a single receiver via an Additive white Gaussian noise (AWGN) channel. Each node encodes k bits of information into one packet of blocklength N, and one packet would be dropped after Mth transmission failure. By extending the analytical framework in [16], the network sum rate is obtained as explicit functions of the system parameters, based on which the optimal sum rate performance is further characterized by jointly tuning the transmission probabilities of nodes and the blocklength of packets. The analysis reveals that both the maximum sum rate and the corresponding optimal blocklength are independent of the retry limit M, while the optimal transmission probability of nodes varies with M. To improve the maximum sum rate, the number of information bits per packet k should be enlarged, while the performance gain becomes marginal if k is large.

The remainder of this paper is organized as follows. Section II presents the system model and preliminary analysis. In Section III, the network sum rate is derived and maximized. Conclusions are summarized in Section IV.

II. System Model and Preliminary Analysis

Consider an n-node Aloha network where each node transmits to one common receiver. Assume that each node always has information bits to send. Each node encodes k information bits to a codeword, i.e., a sequence of symbols, as one packet to be transmitted over the channel. We will refer to the number of symbols N as the blocklength (which is also the packet length). Assume the time axis is slotted, and each packet transmission lasts for one slot. The information encoding rate of each node is then equal to the ratio of the number of information bits per packet k to the blocklength N, i.e.,

$$R = \frac{k}{N}. \tag{1}$$

An AWGN channel is assumed between each node and the receiver. To ensure fairness, each node performs power control to compensate for the large-scale fading, which results in an identical received signal to noise ratio (SNR) ρ at the receiver. The channel capacity is defined as the largest information encoding rate k/N for which the packet error probability can be made arbitrarily small by choosing the blocklength N sufficiently large. With the AWGN channel, as N goes to infinity, the information encoding rate of each node can approach the channel capacity $\log_2(1+\rho)$ by random coding. In the finite blocklength region, nevertheless, we have rate loss from the channel capacity, and the information encoding rate of each node can be approximately written as [17]

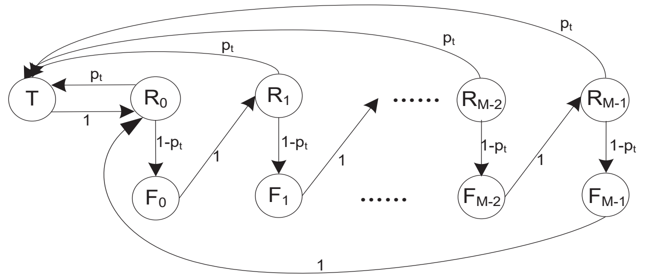


Fig. 1. Embedded markov chain of the state transition process of an individual HOL packet with retry limit M.

$$R = \log_2(1+\rho) - \sqrt{\frac{V}{N}}Q^{-1}(\epsilon) + \frac{1}{2N}\log_2 N, \tag{2}$$

where $Q^{-1}(\cdot)$ denotes the inverse of the Gaussian Q function and the channel dispersion V is given by $V = \rho\frac{2+\rho}{1+\rho^2}(\log_2 e)^2$, and ϵ denotes the packet error probability, which can be obtained as

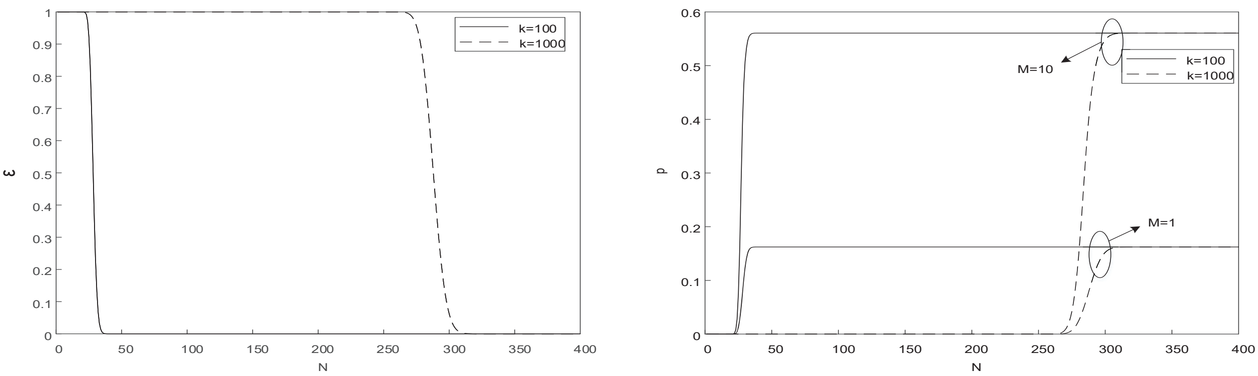
$$\epsilon = Q\left(\frac{N\log_2(1+\rho) - k + (\log_2 N)/2}{\sqrt{NV}}\right), \tag{3}$$

by combining (1) and (2).

The classic collision model is assumed, i.e., one packet transmission would fail if more than one node attempt to transmit at the same time. Therefore, a packet transmission is successful as long as there are no concurrent transmissions and no decoding error. Assume perfect and instant feedback of the transmission outcome from the receiver. For each node, if the head-of-line (HOL) packet in its buffer has experienced ith transmission failure, it would attempt to retransmit the packet with probability of q_i. To avoid excessive access delay, consider that one packet would be dropped after M-th transmission failure, and M is refered to as the retry limit. Without loss of generality, let $q_i = q_0\mathcal{Q}(i)$, where q_0 is the initial transmission probability and $\mathcal{Q}(i)$ is an arbitrary monotonic non-increasing function of i with $\mathcal{Q}(0) = 1$ and $\mathcal{Q}(i) \leq \mathcal{Q}(i-1)$, $i = 1, \cdots, M$.

A. State Characterization of HOL Packets

The behavior of the HOL packet in each node's queue can be characterized by a discrete-time Markov renewal process $(\boldsymbol{X}, \boldsymbol{V}) = \{(X_j, V_j), j = 0, 1, \ldots\}$, where X_j denotes the state of the HOL packet at the j-th transition and V_j denotes the epoch at which the j-th transition occurs. Fig. 1 shows the embedded markov chain $\boldsymbol{X} = \{X_j\}$. The states of $\{X_j\}$ can be divided into three categories: 1) waiting to request (State R_i, $i = 0, \ldots, M-1$) 2) collision (State F_i,$i = 0, \ldots, M-1$) and 3) successful transmission (State T). A HOL packet moves from State R_i to State T if the transmission succeeds;



(a) (b)

Fig. 2. Packet error probability ϵ and the probability of successful transmissions of HOL packets p versus the blocklength N. $\rho = 10$. $q_0 = 0.1$. $n = 20$. $\mathcal{Q}(i) = 2^{-i}$. (a) ϵ versus N. (b) p versus N.

otherwise, it will stay in State F_i until the end of failure and then shifts to State R_{i+1}. A HOL packet will be dropped after the transmission fails for M times, i.e., when it leaves State F_{M-1}, and a new HOL packet is initially at State R_0.

Let p_t denote the probability of successful transmissions of HOL packets at time slot $t = 1, 2, \ldots$. The markov chain in Fig. 1 is uniformly strongly ergodic if and only if the limit

$$\lim_{t\to\infty} p_t = p \tag{4}$$

exists. The steady-state probability distribution of the embeded markov chain can then be obtained as

$$\pi_{R_i} = \frac{(1-p)^i}{1-(1-p)^M} \cdot \pi_T, \text{ for } i = 0, \ldots, M-1, \tag{5}$$

and

$$\pi_{F_i} = \frac{(1-p)^{i+1}}{1-(1-p)^M} \cdot \pi_T, \text{ for } i = 0, \ldots, M-1. \tag{6}$$

The interval between successive transitions, i.e., $V_{j+1} - V_j$, is called the holding time in State $X_j, j = 0, 1, \ldots$. In particular, the mean holding time τ_T in State T and the mean holding time τ_{F_i} in State $F_i, i = 0, \ldots, M-1$, depend on the transmission time of each packet, which equals one time slot. The mean holding time τ_{R_i} in State $R_i, i = 0, \ldots, M-1$, on the other hand, is determined by the backoff protocol. Specially, it is given by the expected time interval when the HOL packet stays at State R_i before it is transmitted. Recall that q_i denotes the probability of accessing the channel of a State-R_i HOL packet in each time slot. The mean holding time τ_{R_i} in State R_i, $i = 0, \ldots, M-1$, is thus given by

$$\tau_{R_i} = \frac{1}{q_i} - 1. \tag{7}$$

Finally, the limiting state probabilities of the Markov renewal process $(\boldsymbol{X}, \boldsymbol{V})$ are given by

$$\tilde{\pi}_j = \frac{\pi_j \cdot \tau_j}{\pi_T \cdot \tau_T + \sum_{i=0}^{M-1} \pi_{F_i} \cdot \tau_F + \sum_{i=0}^{M-1} \pi_{R_i} \cdot \tau_{R_i}}, \tag{8}$$

$j \in S$, where S is the state space of $\boldsymbol{X}$. Specifically, by combining (5)-(8), the probability of one HOL packet being in State T can be obtained as

$$\tilde{\pi}_T = \frac{1}{\frac{1}{1-(1-p)^M} \sum_{i=0}^{M-1} \frac{(1-p)^i}{q_i}}. \tag{9}$$

B. Probability of Successful Transmissions of HOL packets

One HOL packet is successfully received when all the other nodes do not attempt to access the channel and this HOL packet is decoded by the receiver successfully. Therefore, we have

$$p = \Pr\{\text{Other HOL packets are in State } R_i \text{ and do not attempt to access the channel}\} \cdot (1-\epsilon), \tag{10}$$

where ϵ is the packet error probability. According to the embedded Markov chain shown in Fig. 1, the steady-state probability of successful transmission of a HOL packet p can be obtained as

$$p = \left\{1 - \sum_{i=1}^{M-1} \tilde{\pi}_{R_i} q_i^{'}\right\}^{n-1} \cdot (1-\epsilon). \tag{11}$$

As the mean sojourn time of each HOL packet in State R_i is given by $\tau_{R_i} = \frac{1}{q_i} - 1$, the equivalent transmission probability is given by $q_i^{'} = \frac{1}{\tau_{R_i}} = \frac{q_i}{1-q_i}$. We then have

$$p \overset{\text{with a large } n}{\approx} \exp\left\{-n \sum_{i=0}^{M-1} \tilde{\pi}_{R_i} \frac{q_i}{1-q_i}\right\} \cdot (1-\epsilon). \tag{12}$$

By combining (8), (9) and (11), the probability of successful transmissions of HOL packets p can be further obtained as

$$\begin{aligned} p &= \exp\left\{-\frac{n\tilde{\pi}_T}{p}\right\}(1-\epsilon) \\ &= \exp\left\{-\frac{n}{\frac{p}{1-(1-p)^M}\sum_{i=0}^{M-1}\frac{(1-p)^i}{q_i}}\right\}(1-\epsilon). \end{aligned} \tag{13}$$

Recall that one HOL packet would be discarded if and only if its transmission fails for M times. The probability that a HOL packet is dropped, p_d, is then given by

$$p_d = (1-p)^M. \tag{14}$$

The reliability η is defined as the probability that a HOL packet is not dropped, i.e.,

$$\eta = 1-(1-p)^M. \tag{15}$$

Fig. 2 demonstrates how the packet error probability ϵ and the probability of successful transmissions of HOL packets p vary with the blocklength N with the number of information bits per packet $k = 100$ or 1000. Specifically, it can be observed from Fig. 2(a) that the packet error probability ϵ decreases as the number of information bits k decreases or the blocklength N increases. In particular, it would decrease sharply within a small range of N. For example, with $k = 100$, ϵ quickly drops from 1 to 0 when N increases from 20 to 40. Generally, the point $\frac{k}{\log_2(1+\rho)}$ is included in this small range of N where ϵ has a significant change. For instance, in the case of $k = 100$ and $\rho = 10$, ϵ drops from 1 to 0 when N increases from 20 to 40 and the point $k/\log_2(1+\rho) \approx 29 \in (20, 40)$. Moreover, for the probability of successful transmissions of HOL packets p, it can be observed from Fig. 2(b) that as the blocklength N increases or the number of information bits k decreases, p increases since the packet error probability is improved as shown in Fig. 2(a). p can also be improved with a larger retry limit M.

III. Network Sum Rate

In this section, we will characterize the network sum rate and study how to properly select the initial transmission probability of each node q_0 and the blocklength N to maximize the network sum rate.

A. *Maximum Sum Rate*

The network sum rate is defined as the average successfully transmitted information bits per channel use, which can be obtained as

$$C = \hat{\lambda}_{out} \cdot R, \tag{16}$$

where $\hat{\lambda}_{out}$ is the network throughput, which equals the fraction of time slots that have successful packet transmissions, and the information encoding rate of each node R is given by (2).

In saturated conditions that each node always has information bits to transmit, the node throughput equals the service rate of its queue. According to Fig. 1, each node has a successful packet transmission if and only if its HOL packet is in State T. Therefore, the node throughput equals the probability of the HOL packet being in State T, $\tilde{\pi}_T$, and the network throughput is thus given by

$$\hat{\lambda}_{out} = n\tilde{\pi}_T = -p\ln\frac{p}{1-\epsilon}, \tag{17}$$

by combining (9) and (13).

According to (1), (16) and (17), the network sum rate can be derived as

$$C = -\frac{k}{N}\cdot p\ln\frac{p}{1-\epsilon}. \tag{18}$$

It can be seen from (3), (13) and (18) that for given signal to noise ratio ρ and the number of information bits per packet k, C is determined by the initial transmission probability of each node q_0 and the blocklength N. Intuitively, with a larger blocklength N, the network throughput $\hat{\lambda}_{out}$ can be improved due to a lower packet error probability ϵ, which, however, might be achieved at the cost of the information encoding rate of each node R according to (2). Thus, to maximize the network sum rate, the initial transmission probability of each node q_0 and the blocklength N should be jointly tuned. The following theorem presents the maximal sum rate $C_{\max} = \max_{\{N,q_0\}} C$ and the corresponding optimal settings of initial transmission probability of each node q_0^* and the blocklength N^*.

Theorem 1. *The maximum sum rate $C_{\max}$ is given by*

$$C_{\max} = \frac{k}{N^*}\frac{1-Q\left(\frac{N^*\log_2(1+\rho)-k+(\log_2 N^*)/2}{\sqrt{N^*V}}\right)}{e}, \tag{19}$$

which is achieved when the blocklength N is set to N^* and the initial transmission probability q_0 is set to q_0^*, where N^* is the single non-zero root of the following equation

$$\begin{aligned}1 =& Q\left(\frac{N\log_2(1+\rho)-k+(\log_2 N)/2}{\sqrt{NV}}\right)+\\ &\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(N\log_2(1+\rho)-k+(\log_2 N)/2)^2}{2NV}\right\}\\ &\cdot\frac{N\log_2(1+\rho)+k-\frac{1}{\ln 2}-(\log_2 N)/2}{2\sqrt{NV}},\end{aligned}$$

and q_0^* is given by

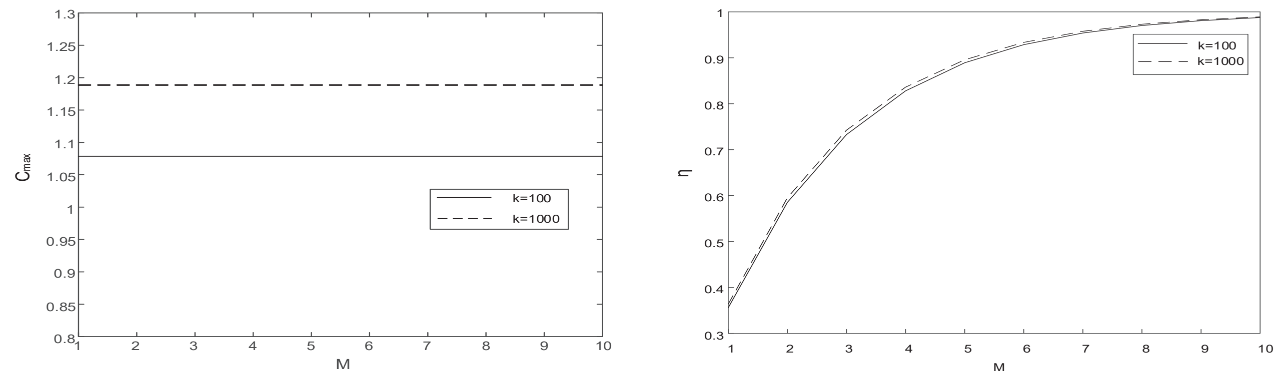
$$q_0^* = \frac{p^*}{n[1-(1-p^*)^M]}\sum_{i=0}^{M-1}\frac{(1-p^*)^i}{\mathcal{Q}(i)}, \tag{20}$$

where $p^* = \left(1-Q\left(\frac{N^*\log_2(1+\rho)-k+(\log_2 N^*)/2}{\sqrt{N^*V}}\right)\right)e^{-1}$.

Proof: See Appendix A. ■

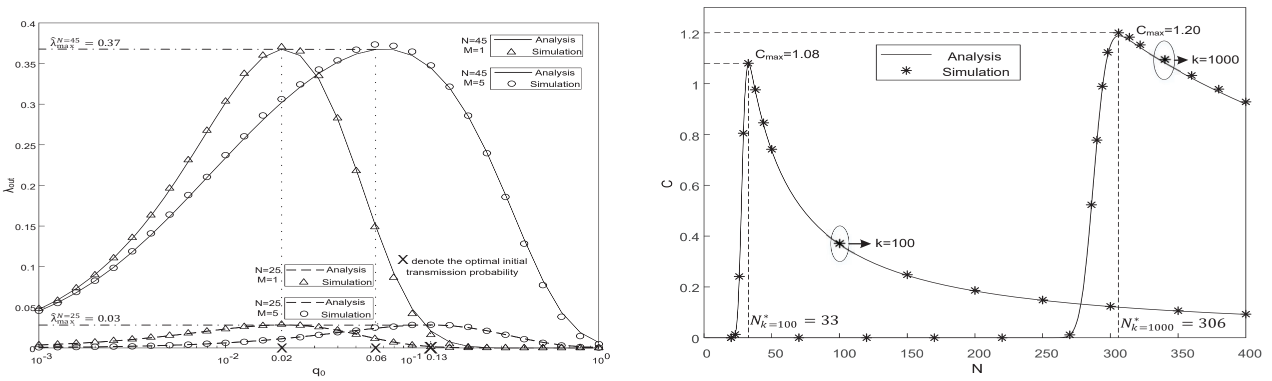
It is revealed in Theorem 1 that both the maximum sum rate $C_{\max}$ and the corresponding optimal blocklength N^* depend on the number of information bits per packet k and the signal-to-noise ratio ρ, while is independent of the retry limit M.

Fig. 3(a) illustrates how the maximum sum rate $C_{\max}$ varies with the retry limit M with the number of information bits per packet $k = 100$ or 1000. It can be observed from Fig. 3(a) that $C_{\max}$ is independent of the retry limit M, and increases as k increases. As $k \to \infty$, we have $\lim_{k\to\infty} C_{\max} = e^{-1}\log_2(1+\rho)$ according to (19), implying that the performance gain achieved by increasing k would be limited if k is large. In this case, as the optimal blocklength also goes to infinity according to (20), the information encoding rate of each node R approaches the channel capacity $\log_2(1+\rho)$. With a small number of



(a) (b)

Fig. 3. Maximum sum rate $C_{\max}$ and the reliability η versus the retry limit M. $\rho = 10$, $n = 20$. $N = N^*$. $q_0 = q_0^*$. (a) $C_{\max}$ versus M. (b) η versus M.



(a) (b)

Fig. 4. (a) Network throughput $\hat{\lambda}_{out}$ versus the initial backoff factor q_0. $\rho = 10$. $n = 50$. $k = 100$. (b) Network sum rate C versus the blocklength N. $\rho = 10$. $n = 50$. $M = 1$. $q_0 = q_0^*$. $\mathcal{Q}(i) = 2^{-i}$.

information bits per packet k, nevertheless, a loss in the network sum rate would incur in the short blocklength region.

Although the maximum sum rate $C_{\max}$ does not depend on M, the reliability η can be significantly improved as M increases, as Fig. 3(b) illustrates. Here we can see that from the perspective of the sum rate performance, a larger retry limit M is preferable, as the maximum sum rate is insensitive to M yet the reliability performance can be significantly improved. In contrast to the sum rate performance, the reliability is insensitive to the number of information bits per packet k when sum rate C is maximized according to Theorem 1.

B. Simulation Results

In this subsection, we present simulation results to verify the proceeding analysis. Note that event-driven simulations are conducted, and each simulation is carried out for 10^6 time slots. The simulation setting is consistent with the system model and the details are omit here.

Fig. 4(a) shows how the network throughput $\hat{\lambda}_{out}$ varies with initial transmission probability q_0 under various values of the blocklength N and the retry limit M. In simulations, we count the total number of successful access requests in each simulation run, i.e., 10^6 time slots. The network throughput is then obtained by calculating the ratio of the number of successful access requests to the number of time slots. It can be clearly seen from Fig. 4(a) that the network throughput $\hat{\lambda}_{out}$ is sensitive to the variation of q_0, implying that to maximize $\hat{\lambda}_{out}$, the initial transmission probability q_0 should be carefully selected. We can observe that by optimally tuning q_0 according to (20), the maximum network throughput $\hat{\lambda}_{\max}$ can be achieved. Note that the network sum rate C is proportional to the network throughput $\hat{\lambda}_{out}$, as shown in (18). Thus, the optimal initial transmission probability q_0^* for maximizing C is equivalent to that for maximizing $\hat{\lambda}_{out}$. It is also demonstrated in Fig. 4(a) that the maximum network throughput $\hat{\lambda}_{\max}$ is independent of the retry limit M, and increases as the blocklength N increases. The simulation results match with the analysis well.

Fig. 4(b) further presents how the network sum rate C varies with the blocklength N with the network throughput is maximized. As Theorem 1 indicates, in this case, N needs to be further tuned to optimize the network sum rate C, and the optimal blocklength N can be obtained according to (20). The simulation results in Fig. 4(b) indicates that the network sum rate is very sensitive to N. When N is too small, C is degraded due to a large packet error probability ϵ. When N is large, although ϵ is significantly improved, each node's information encoding rate becomes small, leading to a low network sum rate.

IV. Conclusion

This paper characterizes the sum rate performance of Aloha networks with packet dropping in the finite blocklength region. By deriving the probability of successful transmissions of HOL packets, an explicit expression of the network sum rate is obtained and further optimized by jointly choosing the transmission probabilities of nodes and the blocklength of packets. The effect of the number of information bits per packet and the retry limit on the network optimal performance is characterized. It is found that increasing the retry limit does not affect the maximum sum rate while can improve the reliability. To enhance the sum rate performance, on the other hand, the network should enlarge the number of information bits per packet.

Appendix A
Proof of Theorem 1

According to (17), $\hat{\lambda}_{out}$ depends on the probability of successful transmissions of HOL packets p, and in turn depends on the initial transmission probability of each node q_0. By carefully turning q_0, $\hat{\lambda}_{out}$ can be maximized. By combining (1) and (16), we then have

$$C_{\max} = \max_{\{N,q_0\}} \frac{k}{N} \cdot \hat{\lambda}_{out} = \max_{N} \frac{k}{N} \max_{\{q_0\}} \hat{\lambda}_{out} = \max_{N} \frac{k}{N} \hat{\lambda}_{\max}, \tag{21}$$

where the maximum network throughput $\hat{\lambda}_{\max}$ can be written as

$$\hat{\lambda}_{\max} = \max_{\{q_0\}} \hat{\lambda}_{out} = \frac{1-\epsilon}{e}, \tag{22}$$

which is achieved when the initial transmission probability is set to be

$$q_0 = \frac{e^{-1}(1-\epsilon)}{n[1-(1-e^{-1}(1-\epsilon))^M]} \sum_{i=0}^{M-1} \frac{(1-e^{-1}(1-\epsilon))^i}{\mathcal{Q}(i)}. \tag{23}$$

Therefore, we have

$$C_{\max} = \max_{\{N\}} \frac{k}{N} \cdot \frac{1-\epsilon}{e}, \tag{24}$$

by combining (21) and (22).

Let $C(N) = \frac{k}{N} \cdot \frac{1-\epsilon}{e}$. Then the derivative of C with respective to N is given by

$$C'(N) = -\frac{k}{eN^2}[1-\epsilon(N) - N\epsilon'(N)]. \tag{25}$$

Letting $C'(N) = 0$ yields the optimal blocklength N^*, which is given by (20). The maximun sum rate $C_{\max}$ is then obtained by combining (24) and $N = N^*$. The optimal transmission probability can then be obtained by combining (20) and (23).

References

[1] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the internet of things: Communication technologies and challenges," *IEEE Access*, vol. 6, pp. 3619–3647, 2018.

[2] A. Rajandekar and B. Sikdar, "A survey of MAC layer issues and protocols for machine-to-machine communications," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 175–186, Apr. 2015.

[3] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives," *IEEE Commun. Surv. Tutor*, vol. 16, no. 1, pp. 4–16, 2014.

[4] A. Hoglund, D. P. Van, T. Tirronen, O. Liberg, Y. Sui, and E. A. Yavuz, "3GPP release 15 early data transmission," *IEEE Commun. Stand. Mag.*, vol. 2, no. 2, pp. 90–96, Jun. 2018.

[5] H. Wang and A. O. Fapojuwo, "A survey of enabling technologies of low power and long range machine-to-machine communications," *IEEE Commun. Surv. Tuts.*, vol. 19, no. 4, pp. 2621–2639, 2017.

[6] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. 25, no. 1, pp. 117–128, Jan. 1977.

[7] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Autom. Control*, vol. 27, no. 3, pp. 559–569, Jun. 1982.

[8] R. Rivest, "Network control by bayesian broadcast," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 323–328, May 1987.

[9] C. Oh, D. Hwang, and T. Lee, "Joint access control and resource allocation for concurrent and massive access of M2M devices," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 8, pp. 4182–4192, Aug. 2015.

[10] Y. Wang, T. Wang, Z. Yang, D. Wang, and J. Cheng, "Throughput-oriented non-orthogonal random access scheme for massive MTC networks," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1777–1793, Mar. 2020.

[11] W. Zhan and L. Dai, "Massive random access of machine-to-machine communications in LTE networks: Modeling and throughput optimization," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 4, pp. 2771–2785, Apr. 2018.

[12] H. Wu, C. Zhu, R. J. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA using access history for event-driven M2M communications," *IEEE-ACM Trans. Netw.*, vol. 21, no. 6, pp. 1904–1917, Dec. 2013.

[13] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. S. Wong, "D-ACB: Adaptive congestion control algorithm for bursty M2M traffic in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9847–9861, Dec. 2016.

[14] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge Univ. Press, 2005.

[15] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[16] L. Dai, "Stability and delay analysis of buffered aloha networks," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 8, pp. 2707–2719, Aug. 2012.

[17] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.