On the Asymptotic Performance of Delay-Constrained Slotted ALOHA

Lei Deng*, Jing Deng†, Po-Ning Chen‡, and Yunghsiang S. Han*
*School of Electrical Engineering & Intelligentization, Dongguan University of Technology
†Department of Computer Science, University of North Carolina at Greensboro
‡Department of Electrical and Computer Engineering, National Chiao Tung University

Abstract—**Motivated by the proliferation of real-time applications in multimedia communication systems, tactile Internet, networked controlled systems, and cyber-physical systems, supporting delay-constrained traffic become critical for the communication system. In delay-constrained traffic, each packet has a hard deadline and if it cannot be delivered before its deadline, it becomes useless and will be removed from the system. In this work, we consider a slotted ALOHA system where multiple stations need to deliver delay-constrained traffic to a common receiver by accessing a shared channel. We prove that, under the frame-synchronized traffic pattern, the maximum system timely throughput converges to $1/e = 36.8\%$ as the number of stations goes to infinity, which is the same as the asymptotic maximum system throughput for delay-unconstrained slotted ALOHA system with saturate traffic. While this is not completely surprising, we further investigate the speed of such a maximum system throughput approaching $1/e$ under borderline traffic.**

I. Introduction

Since Abramson's invention in 1970 [1], ALOHA-type protocols have been widely used for multiple users to access a shared communication channel due to its extreme simplicity and decentralized nature. One popular type is *slotted ALOHA*, where users are synchronized and can only transmit at the beginning of a slot [2]. It is well-known that the optimal asymptotic system throughput of slotted ALOHA system is $1/e = 36.8\%$ [3]. In addition, there are many works to investigate the stability region of slotted ALOHA, i.e., to characterize the feasible input rates under which the system can be stabilized in the sense that all queues will not blow up [4], [5]. There are also many types of extension for slotted ALOHA protocol, including multi-packet reception [6], framed slotted ALOHA [7], coded slotted ALOHA [8], etc.

Most of existing work focus on delay-unconstrained case in that any packet can be delivered in whatever long time. However, nowadays with the proliferation of real-time applications, communication networks need to support more and more *delay-constrained* traffic. Typical examples include multimedia communication systems such as real-time streaming and video conferencing [9], tactile Internet [10], [11], networked controlled systems (NCSs) such as remote control of unmanned aerial vehicles (UAVs) [12], [13], and cyber-physical systems (CPSs) such as medical tele-operations, X-by-wire vehilces/avionics, factory automation, and robotic collaboration [14]. In such applications, each packet has a hard deadline: if it is not delivered before its deadline, its validity will expire and it will be removed from the system. Thus, it is important to examine the performance of slotted ALOHA protocol to deliver such delay-constrained traffic.

There are some works to study slotted ALOHA with delay-constrained traffic. Zhang et al. in [15] investigated the system throughput and the optimal retransmission probability with delay-constrained *saturate* traffic in the sense that each station always has a new packet arrival once its queue becomes empty (namely each station always has a packet to transmit). We should note that such a saturate traffic model is not necessarily practical. For example, in NCSs and CPSs, the control messages usually arrive periodically. Thus, in this paper we analyze the slotted ALOHA system with a non-saturate delay-constrained traffic model. It is called frame-synchronized traffic patten, which was widely investigated in the packet scheduling policy design in the delay-constrained wireless communication community [9], [16].

In particular, in this paper, we make the following contributions:

- We formulate the delay-constrained slotted ALOHA problem and we propose an algorithm to compute the system throughput for any number of stations N, and delay D, and any retransmission probability p;
- We prove that the maximum system throughput converges to $1/e = 36.8\%$ when the number of stations N goes to infinity, which is the same as the asymptotic maximum system throughput for delay-unconstrained slotted ALOHA system with saturate traffic [3, Chapter 5.3.2];
- We further investigate the speed of such a maximum system throughput approaching $1/e$ under borderline traffic.

II. System Model

Multiple (say in total N) stations (e.g., mobile devices) share a common channel (e.g., a wireless channel) and they need to send delay-constrained data packets to a common receiver via the shared channel, as shown in Fig. 1. All packets have the same size and time is slotted where the slot duration is the time to transmit a packet from a station to the receiver and receive the feedback from the receiver about whether the packet has been successfully delivered or not (ACK/NACK). A station can only send a packet at the beginning of a slot. If only one station transmits, the packet will be delivered successfully. Otherwise, if two or more stations transmit in the same slot, a collision happens and all packets will be lost. The station

978-1-5386-5156-8/18/$31.00 ©2018 IEEE

knows whether the packet is delivered successfully or not at the end of the transmitted slot.

We assume that each packet has a hard delay/deadline $D \geq 1$. Once arriving at a station, a packet will expire and be removed from the system if it cannot be delivered in D slots. Such feature is fundamentally different from the delay-unconstrained scenario. In delay-constrained scenario, the (arrival) traffic patten greatly influences the system performance. As an initial study, in this paper, we assume a *frame-synchronized traffic pattern*. It can find applications in NCSs and CPSs where the system generates the control packets/messsages periodically [17] and it was widely investigated in the packet scheduling policy design in the delay-constrained wireless communication community [9], [16]. In the frame-synchronized traffic pattern, starting from slot 1, all stations have a packet arrival every D slots (which is also the hard delay of the packet). We call every consecutive D slots (staring from slot 1) a *frame*. The first frame is from slot 1 to slot D; the second frame is from slot $D+1$ to slot $2D$; and so on. Every station has a packet arrival at the beginning of a frame and the packet expires at the end of the frame. An example of $D = 3$ of the frame-synchronized traffic pattern is shown in Fig. 1.

In the traditional slotted ALOHA protocol, a station transmits immediately when a new packet arrives. If a collision happens, the packet is backlogged in the station's queue and it will be retransmitted (as an old packet) in the next slot with probability $p \in (0, 1]$. Thus, at the beginning of each frame (e.g., slot 1), a collision must happen because all stations have a new packet arrival. To avoid this problem, we modify the traditional slotted ALOHA protocol such that all station will always transmit/retransmit its (new or old) packet with probability p. For simplicity, we call p the retransmission probability.

Given system parameters — the packet delay D, the number of stations N, and the retransmission probability p, we define the *system timely throughput* as the per-slot average number of delivered packets before expiration of the system, i.e.

$$R(D,N,p) \triangleq \lim_{t\to\infty} \frac{\mathbb{E}\left[\substack{\text{number of packets delivered before}\\ \text{expiration from slot 1 to slot } t}\right]}{t}. \quad (1)$$

Our system timely throughput only counts those packets that have been delivered before expiration and ignores those packets that expire and are removed from the system after their deadlines. For simplicity, we sometimes call it *system throughput*.

Since our traffic pattern is frame-synchronized and all stations's retransmission probabilities are the same at all slots, it is easy to see that the average number of delivered packets before expiration is the same from all frames. Thus, we only need to focus on the first frame and the system timely throughput becomes

$$R(D,N,p) = \frac{\mathbb{E}\left[\substack{\text{number of packets delivered before}\\ \text{expiration from slot 1 to slot } D}\right]}{D}. \quad (2)$$

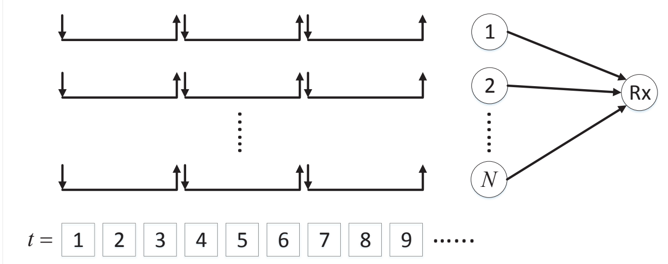

Fig. 1. System model with $D = 3$.

Note that (2) also indicates that the limit in (1) exists and thus (1) is well-defined.

We define the optimal retransmission probability as,

$$p^*(D,N) \triangleq \arg\max_{p\in(0,1]} R(D,N,p), \quad (3)$$

which maximizes the system throughput $R(D,N,p)$. We also denote the maximum/optimal system throughput as

$$R^*(D,N) \triangleq \max_{p\in(0,1]} R(D,N,p) = R(D,N,p^*(D,N)). \quad (4)$$

Next we will show how to compute $R(D,N,p)$ and analyze the asymptotic performance of $R^*(D,N)$ when the number of stations goes to infinity (i.e., $N \to \infty$). Before that, let us see a special case when $D = 1$.

When $D = 1$, then every station has a new packet arrival every slot. This is similar to the case in the *delay-unconstrained* slotted ALOHA with *saturate* traffic [3, Chapter 5.3.2]. Under this special case, the system throughput becomes

$$R(1,N,p) = \binom{N}{1} p(1-p)^{N-1} = Np(1-p)^{N-1}. \quad (5)$$

It is easy to see that the optimal retransmission probability is

$$p^*(1,N) = 1/N. \quad (6)$$

Further, the maximum system throughput is

$$R^*(1,N) = (1-1/N)^{N-1}. \quad (7)$$

Clearly, we have

$$\lim_{N\to\infty} R^*(1,N) = \frac{1}{e} = 36.8\%. \quad (8)$$

The results in (5), (6), (7), and (8) are the same with those in the delay-unconstrained slotted ALOHA with saturate traffic [3, Chapter 5.3.2] where all N stations will transmit/retransmit its packet with probability p and each station always has a new packet arrival once its packet has been delivered successfully.

Beyond this special case, next we propose an algorithm to compute $R(D,N,p)$ for general delay D in Sec. III. We further show that for any delay D, the maximum system throughput converges to $1/e = 36.8\%$ in Sec. IV.

III. An Algorithm to Compute $R(D, N, p)$

Let us show an algorithm to compute $R(D, N, p)$. From (2), we know that we only need to focus on the first frame from slot 1 to slot D. We denote an outcome of the first frame as $\boldsymbol{s} = (s_1, s_2, \cdots, s_D)$ where $s_t = 1$ if there is a packet delivered successfully in slot t (because there is only one station transmitting at this slot) and $s_t = 0$ otherwise (because there is no station transmitting at this slot or there are two or more stations transmitting simultaneously, which causes collision, at this slot). Denote the set of all possible outcomes as $\mathcal{S}$. Since we have in total N stations, we can at most transmit $\min\{D, N\}$ packets in the first frame. Thus, the set of all possible outcomes $\mathcal{S}$ is

$$\mathcal{S} = \left\{ (s_1, s_2, \cdots, s_D) \in \{0,1\}^D : \sum_{t=1}^{D} s_t \leq \min\{D, N\} \right\}. \tag{9}$$

Define

$$f(\boldsymbol{s}) \triangleq \sum_{t=1}^{D} s_t, \quad \boldsymbol{s} = (s_1, s_2, \cdots, s_D) \in \mathcal{S}, \tag{10}$$

as the total number of packets delivered successfully in the frame.

Let us use a random-variable vector $\boldsymbol{S} = (S_1, S_2, \cdots, S_D)$ to denote the outcome of the slotted ALOHA system. Then, if we can compute the probability of each outcome, i.e.

$$P(\boldsymbol{S} = \boldsymbol{s}) = P(S_1 = s_1, S_2 = s_2, \cdots, S_D = s_D), \tag{11}$$

based on (2), we can obtain the system throughput as

$$R(D, N, p) = \frac{\sum_{\boldsymbol{s} \in \mathcal{S}} f(\boldsymbol{s}) P(\boldsymbol{S} = \boldsymbol{s})}{D}. \tag{12}$$

The remaining problem is how to compute $P(\boldsymbol{S} = \boldsymbol{s})$ for any outcome $\boldsymbol{s} \in \mathcal{S}$. For an outcome $\boldsymbol{s} = (s_1, s_2, \cdots, s_D)$, we define function

$$g_t(\boldsymbol{s}) = N - \sum_{\tau=1}^{t-1} s_\tau, \quad \forall t = 1, 2, \cdots, D \tag{13}$$

and we set $g_1(\boldsymbol{s}) = N$ by convention. Here $g_t(\boldsymbol{s})$ is the number of remaining stations who have not delivered their packets before slot t. Then the probability of delivering a packet at slot t conditioning on historical outcome $(s_1, \cdots, s_{t-1})$ is

$$\begin{aligned} &P\left(S_t = 1 \,|(S_1, S_2, \cdots S_{t-1}) = (s_1, s_2, \cdots, s_{t-1})\right) \\ &= \begin{cases} 0, & \text{if } g_t(\boldsymbol{s}) = 0 \text{ ;} \\ g_t(\boldsymbol{s}) p (1-p)^{g_t(\boldsymbol{s})-1}, & \text{Otherwise.} \end{cases} \end{aligned} \tag{14}$$

and the probability of not delivering any packet at slot t conditioning on historical outcome $(s_1, \cdots, s_{t-1})$ is

$$\begin{aligned} &P\left(S_t = 0 \,|(S_1, S_2, \cdots S_{t-1}) = (s_1, s_2, \cdots, s_{t-1})\right) \\ &= \begin{cases} 1, & \text{if } g_t(\boldsymbol{s}) = 0 \text{ ;} \\ 1 - g_t(\boldsymbol{s}) p (1-p)^{g_t(\boldsymbol{s})-1}, & \text{Otherwise.} \end{cases} \end{aligned} \tag{15}$$

Then we can compute $P(\boldsymbol{S} = \boldsymbol{s})$ as follows,

$$\begin{aligned} &P(\boldsymbol{S} = \boldsymbol{s}) = P(S_1 = s_1) \times P(S_2 = s_2 | S_1 = s_1) \\ &\times P(S_3 = s_3 | (S_1, S_2) = (s_1, s_2)) \\ &\times \cdots \\ &\times P\left(S_D = s_D | (S_1, S_2, \cdots, S_{D-1}) = (s_1, s_2, \cdots, s_{D-1})\right). \end{aligned} \tag{16}$$

Remarks. The number of outcomes, i.e., $|\mathcal{S}|$, is at most 2^D. For each possible outcome $\boldsymbol{s}$, we need D steps to compute the probability $P(\boldsymbol{S} = \boldsymbol{s})$ according to (16). Thus, the total time complexity to get the system throughput $R(D, N, p)$ is $O(D2^D)$. Now we can use this algorithm to compute $R(D, N, p)$ for any system parameters D, N and p, and thus we can numerically obtain the optimal retransmission probability $p^*(D, N)$ and the maximum system throughput $R^*(D, N)$ (with certain step-size error).

IV. Asymptotic Performance

In (8), we have shown that the maximum system throughput converges to $1/e$ when $D = 1$. This result is the same as that in delay-unconstrained slotted ALOHA with saturate traffic. For general (fixed) $D > 1$, when N is large, in slot 1, it has N stations who have packets; in slot 2, there are at least $N - 1$ stations who have packets. Similarly, in any slot $t \in \{1, 2, \cdots, D\}$, there are at least $N - t + 1$ stations who have packets. Since $N - t + 1 \to \infty$ (as $N \to \infty$) for $1 \leq t \leq D$, it is reasonable to anticipate that a similar situation to the delay-unconstrained case with saturate traffic load occurs at each slot, i.e., the number of stations who have packets goes unbounded at all slots. According to this intuition, we may guess that the maximum system throughput converges to $1/e = 36.8\%$ for any delay D. We now present our main result.

***Theorem** 1:* For any $D \geq 1$, we have

$$\lim_{N \to \infty} R^*(D, N) = 1/e. \tag{17}$$

Next we prove Theorem 1 rigourously. Towards that end, we first prove some preliminarily lemmas.

A. Preliminarily Lemmas

The optimal retransmission probability $p^*(D, N)$ depends on the number of stations N and the packet delay D. In this section, we fix D but evaluate the asymptotical performance when $N \to \infty$. For notational convenience, we drop the dependence of the retransmission probability on D and denote $p(N)$ as the retransmission probability when the number of stations is N.

Again we consider the first frame from slot 1 to slot D. In any slot $t \in \{1, 2, \cdots, D\}$, we denote random variable $M_{N,t}$ as the number of stations who have already delivered their packets before slot t. Clearly $M_{N,t} \in \{0, 1, 2, \cdots, t-1\}$. Then by denoting the probability of $M_{N,t} = m$ by $P(M_{N,t} = m)$, we prove the following lemma.

***Lemma** 1:* If the retransmission probability $p(N)$ satisfies

$$\lim_{N \to \infty} p(N) = 0, \tag{18}$$

and

$$\lim_{N\to\infty} N \cdot p(N) = 1, \tag{19}$$

then the sequence $\{P(M_{N,t} = m)\}_{N=1}^{\infty}$ has a limit for any $t \in \{1, 2, \cdots, D\}$ and any $m \in \{0, 1, 2, \cdots, t-1\}$. Namely, there exists a non-negative real number γ_t^m such that

$$\lim_{N\to\infty} P(M_{N,t} = m) = \gamma_t^m,$$
$$\forall t \in \{1, 2, \cdots, D\}, m \in \{0, 1, \cdots, t-1\}. \tag{20}$$

In addition,

$$\sum_{m=0}^{t-1} \gamma_t^m = 1, \quad \forall t \in \{1, 2, \cdots, D\}. \tag{21}$$

Proof: We prove the existence of the limit (i.e., (20)) by induction with respect to t. Clearly, when $t = 1$,

$$P(M_{N,1} = 0) = 1, \quad \forall N. \tag{22}$$

Thus,

$$\lim_{N\to\infty} P(M_{N,1} = 0) = 1 \triangleq \gamma_1^0. \tag{23}$$

Suppose that when $t = k \in \{1, 2, \cdots, D-1\}$, the limit $\lim_{N\to\infty} P(M_{N,t} = m)$ exists for any m, i.e.,

$$\lim_{N\to\infty} P(M_{N,k} = m) = \gamma_k^m,$$
$$\forall k \in \{1, 2, \cdots, D-1\}, m \in \{0, 1, \cdots, k-1\}. \tag{24}$$

We then consider $t = k+1$. For $m = 0$,

$$\begin{aligned}
&P(M_{N,k+1} = m) = P(M_{N,k+1} = 0)\\
&= \sum_{m'=0}^{k-1} P(M_{N,k+1} = 0 | M_{N,k} = m') P(M_{N,k} = m')\\
&= P(M_{N,k+1} = 0 | M_{N,k} = 0) P(M_{N,k} = 0)\\
&= \left[1 - \binom{N}{1} p(N)(1 - p(N))^{N-1}\right] P(M_{N,k} = 0).
\end{aligned} \tag{25}$$

Due to (18) and (19), we have

$$\begin{aligned}
&\lim_{N\to\infty} (1 - p(N))^{N-1} = \lim_{N\to\infty} \exp\{(N-1)\ln[1 - p(N)]\}\\
&= \exp\left\{\lim_{N\to\infty} (N-1)\ln[1 - p(N)]\right\}\\
&= \exp\left\{\lim_{N\to\infty} N\ln[1 - p(N)]\right\}\\
&= \exp\left\{\lim_{N\to\infty} Np(N) \cdot \frac{\ln[1 - p(N)]}{p(N)}\right\}\\
&= \exp\left\{\lim_{N\to\infty} Np(N) \cdot \lim_{N\to\infty} \frac{\ln[1 - p(N)]}{p(N)}\right\}\\
&= \exp\left\{1 \cdot \lim_{u\to 0} \frac{\ln(1-u)}{u}\right\} \quad (\text{let } u = p(N))\\
&= \exp\left\{\lim_{u\to 0} \frac{-\frac{1}{1-u}}{1}\right\} \quad (\text{by L'Hospital's Rule})\\
&= \exp(-1) = 1/e.
\end{aligned} \tag{26}$$

Therefore, taking limit in (25), we have

$$\begin{aligned}
&\lim_{N\to\infty} P(M_{N,k+1} = 0)\\
&= \lim_{N\to\infty} \left[1 - Np(N)(1 - p(N))^{N-1}\right] \cdot \lim_{N\to\infty} P(M_{N,k} = 0)\\
&= \left[1 - \lim_{N\to\infty} [Np(N)] \cdot \lim_{N\to\infty} (1 - p(N))^{N-1}\right] \cdot \gamma_k^0\\
&= (1 - 1 \cdot 1/e) \cdot \gamma_k^0 = (1 - 1/e)\gamma_k^0 \triangleq \gamma_{k+1}^0.
\end{aligned} \tag{27}$$

Thus, the limit $\lim_{N\to\infty} P(M_{N,k+1} = 0)$ exists.

Similarly, for $m \in \{1, 2, \cdots, k-1\}$, we have[1]

$$\begin{aligned}
&P(M_{N,k+1} = m)\\
&= \sum_{m'=0}^{k-1} P(M_{N,k+1} = m | M_{N,k} = m') P(M_{N,k} = m')\\
&= P(M_{N,k+1} = m | M_{N,k} = m-1) P(M_{N,k} = m-1)\\
&+ P(M_{N,k+1} = m | M_{N,k} = m) P(M_{N,k} = m)\\
&= \tbinom{N-(m-1)}{1} p(N)(1 - p(N))^{N-(m-1)-1} P(M_{N,k} = m-1)\\
&+ \left[1 - \binom{N-m}{1} p(N)(1 - p(N))^{N-m-1}\right] P(M_{N,k} = m).
\end{aligned}$$

Due to (18) and (19), we can use the similar analysis in (26) and (27) to show that

$$\lim_{N\to\infty} P(M_{N,k+1} = m) = 1/e \cdot \gamma_k^{m-1} + (1 - 1/e)\gamma_k^m \triangleq \gamma_{k+1}^m.$$

Finally, for $m = k$, we have

$$\begin{aligned}
&P(M_{N,k+1} = k)\\
&= P(M_{N,k+1} = k | M_{N,k} = k-1) P(M_{N,k} = k-1)\\
&= \tbinom{N-(k-1)}{1} p(N)(1 - p(N))^{N-(k-1)-1} P(M_{N,k} = k-1)
\end{aligned}$$

and

$$\lim_{N\to\infty} P(M_{N,k+1} = k) = 1/e \cdot \gamma_k^{k-1}. \tag{28}$$

Therefore, for $t = k+1$, the limit $\lim_{N\to\infty} P(M_{N,k+1} = m)$ exists for any $m \in \{0, 1, 2, \cdots, k\}$, which completes the induction proof. Thus, (20) holds.

To prove (21), we can simply take limit in both sides of the following equality,

$$\sum_{m=0}^{t-1} P(M_{N,t} = m) = 1, \forall t \in \{1, 2, \cdots, D\}. \tag{29}$$

The proof is completed. ■

We further present another preliminary lemma.

***Lemma** 2:* Suppose that a_n, b_n are two bounded sequences where $b_n \geq 0, \forall n$. If

$$\limsup_{n\to\infty} a_n \leq 0,$$

then

$$\limsup_{n\to\infty} a_n b_n \leq 0.$$

[1] Since we focus on the limit as N goes to infinity, without loss of generality, we assume that N is large enough, e.g., $N \geq D+1$, which also applies to later proofs in this section whenever we prove issues on the limit.

Proof: We prove this lemma by contradiction. Suppose not. Then, since $\{a_n b_n\}$ is a bounded sequence, there exists a $\phi > 0$ such that

$$\limsup_{n\to\infty} a_n b_n = \phi,$$

which implies that there exists a subsequence $\{n_k\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} a_{n_k} b_{n_k} = \phi > 0. \tag{30}$$

Thus, there exists a $k_0 \geq 1$ such that

$$a_{n_k} b_{n_k} > 0, \quad \forall k \geq k_0. \tag{31}$$

Since $b_n \geq 0, \forall n$, we have

$$a_{n_k} > 0, \quad \forall k \geq k_0, \tag{32}$$

which implies that

$$\limsup_{k\to\infty} a_{n_k} > 0. \tag{33}$$

Then we have

$$\limsup_{n\to\infty} a_n \geq \limsup_{k\to\infty} a_{n_k} > 0, \tag{34}$$

which contradicts with $\limsup_{n\to\infty} a_n \leq 0$.

Therefore, we must have $\limsup_{n\to\infty} a_n b_n \leq 0$. ■

B. The System Throughput when $p(N) = 1/N$

In this subsection, we leverage Lemma 1 to show that the system throughput goes to $1/e$ when the retransmission probability is $p(N) = 1/N$.

Lemma 3: Consider the retransmission probability $p(N) = 1/N$ when the number of stations is N. Then the system throughput converges to $1/e$, i.e.,

$$\lim_{N\to\infty} R(D, N, 1/N) = 1/e. \tag{35}$$

Proof: Again, we consider the first frame from slot 1 to slot D. For any slot $t \in \{1, 2, \cdots, D\}$ in the first frame, recall that we use random variable $M_{N,t}$ to denote the number of stations who have already delivered their packets before slot t. We then denote random variable $S_{N,t}$ as the number of packet delivered at slot t. Then, the probability of delivering a packet in slot t is,

$$\begin{aligned}
P(S_{N,t} = 1) &= \sum_{m=0}^{t-1} P(S_{N,t} = 1 | M_{N,t} = m) P(M_{N,t} = m) \\
&= \sum_{m=0}^{t-1} \binom{N-m}{1} p(N) \left(1 - p(N)\right)^{N-m-1} P(M_{N,t} = m) \\
&= \sum_{m=0}^{t-1} \frac{N-m}{N} \left(1 - \frac{1}{N}\right)^{N-m-1} P(M_{N,t} = m).
\end{aligned} \tag{36}$$

Note that $p(N) = 1/N$ satisfies (18) and (19). Thus, (20) and (21) in Lemma 1 hold. Then, we take limit in (36) and get

$$\begin{aligned}
&\lim_{N\to\infty} P(S_{N,t} = 1) \\
&= \lim_{N\to\infty} \left[\sum_{m=0}^{t-1} \frac{N-m}{N} \left(1 - \frac{1}{N}\right)^{N-m-1} P(M_{N,t} = m) \right] \\
&= \sum_{m=0}^{t-1} \left\{ \lim_{N\to\infty} \frac{N-m}{N} \left(1 - \frac{1}{N}\right)^{N-m-1} P(M_{N,t} = m) \right\} \\
&= \sum_{m=0}^{t-1} \left\{ \lim_{N\to\infty} \tfrac{N-m}{N} \left(1 - \tfrac{1}{N}\right)^{N-m-1} \cdot \lim_{N\to\infty} P(M_{N,t} = m) \right\} \\
&= \sum_{m=0}^{t-1} \frac{1}{e} \cdot \gamma_t^m = \frac{1}{e} \cdot \sum_{m=0}^{t-1} \gamma_t^m = \frac{1}{e} \cdot 1 = \frac{1}{e}.
\end{aligned} \tag{37}$$

Therefore, when N goes to infinity, the probability of delivering a packet in any slot t is $1/e$. Thus, the system throughput also converges to $1/e$, i.e.,

$$\lim_{N\to\infty} R(D, N, 1/N) = \lim_{N\to\infty} \frac{\sum_{t=1}^{D} P(S_{N,t} = 1)}{D} = 1/e,$$

which completes the proof. ■

C. Proof of Main Result (Theorem 1)

We then proceed to prove our main result (Theorem 1) when we use the optimal retransmission probability $p(N) = p^*(D, N)$.

Since $p^*(D, N)$ is the optimal retransmission probability to maximize the system throughput, we have

$$R^*(D, N) = R(D, N, p^*(D, N)) \geq R(D, N, 1/N). \tag{38}$$

Thus, according to Lemma 3, we have

$$\liminf_{N\to\infty} R^*(D, N) \geq \liminf_{N\to\infty} R(D, N, 1/N) = 1/e. \tag{39}$$

Then Theorem 1 holds if we can verify

$$\limsup_{N\to\infty} R^*(D, N) \leq 1/e. \tag{40}$$

For any slot $t \in \{1, 2, \cdots, D\}$ in the first frame, suppose that we have $M_{N,t} = m$ stations who have already delivered their packets before slot t. Clearly, we have $m \in \{0, 1, \cdots, t-1\}$ and there are $N - m$ remaining stations who have a packet at slot t. Then the probability of delivering a packet in slot t is

$$\begin{aligned}
&\alpha(N, m) = (N-m) p^*(N, D) \left[1 - p^*(N, D)\right]^{N-m-1} \\
&\leq \max_{p\in(0,1]} \left[(N-m) p (1-p)^{N-m-1}\right] \\
&= \left(1 - \frac{1}{N-m}\right)^{N-m-1},
\end{aligned} \tag{41}$$

where the last equality follows from the fact that

$$\arg\max_{p\in(0,1]} p(1-p)^{N-m-1} = \frac{1}{N-m}. \tag{42}$$

Then, we have

$$\limsup_{N\to\infty} \alpha(N,m) \le \limsup_{N\to\infty} \left(1-\frac{1}{N-m}\right)^{N-m-1} = \frac{1}{e}, \tag{43}$$

implying that

$$\limsup_{N\to\infty} \left[\alpha(N,m)-\frac{1}{e}\right] \le 0. \tag{44}$$

Recall that we use random variable $S_{N,t}$ to denote the number of packet delivered at slot t. Then, the probability of delivering a packet in slot t is $P(S_{N,t}=1)$ and we have

$$\begin{aligned}
&P(S_{N,t}=1)-\frac{1}{e} \\
&=\sum_{m=0}^{t-1} P(S_{N,t}=1|M_{N,t}=m)P(M_{N,t}=m)-\frac{1}{e} \\
&=\sum_{m=0}^{t-1} \alpha(N,m)P(M_{N,t}=m)-\frac{1}{e}\cdot\sum_{m=0}^{t-1} P(M_{N,t}=m) \\
&=\sum_{m=0}^{t-1} \left[\alpha(N,m)-\frac{1}{e}\right] P(M_{N,t}=m).
\end{aligned}$$

If we set $a_N=\alpha(N,m)-\frac{1}{e}, b_N=P(M_{N,t}=m)$, we can see that $\{a_N\}$ and $\{b_N\}$ satisfy the conditions in Lemma 2, thus,

$$\limsup_{N\to\infty} a_N b_N = \limsup_{N\to\infty} \left[\alpha(N,m)-\frac{1}{e}\right] P(M_{N,t}=m) \le 0.$$

Then, we have

$$\begin{aligned}
&\limsup_{N\to\infty} P(S_{N,t}=1)-\frac{1}{e} = \limsup_{N\to\infty} \left[P(S_{N,t}=1)-\frac{1}{e}\right] \\
&=\limsup_{N\to\infty} \sum_{m=0}^{t-1} \left[\alpha(N,m)-\frac{1}{e}\right] P(M_{N,t}=m) \\
&\le \sum_{m=0}^{t-1} \left\{\limsup_{N\to\infty} \left[\alpha(N,m)-\frac{1}{e}\right] P(M_{N,t}=m)\right\} \\
&\le \sum_{m=0}^{t-1} 0 = 0.
\end{aligned} \tag{45}$$

Therefore, we have

$$\limsup_{N\to\infty} P(S_{N,t}=1) \le \frac{1}{e}. \tag{46}$$

Thus, the system throughput $R^*(D,N)$ satisfies,

$$\begin{aligned}
&\limsup_{N\to\infty} R^*(D,N) = \limsup_{N\to\infty} \frac{\sum_{t=1}^{D} P(S_{N,t}=1)}{D} \\
&\le \frac{\sum_{t=1}^{D} \limsup_{N\to\infty} P(S_{N,t}=1)}{D} \\
&\le \frac{\sum_{t=1}^{D} \frac{1}{e}}{D} = \frac{1}{e},
\end{aligned} \tag{47}$$

which proves (40). The proof for Theorem 1 is completed.

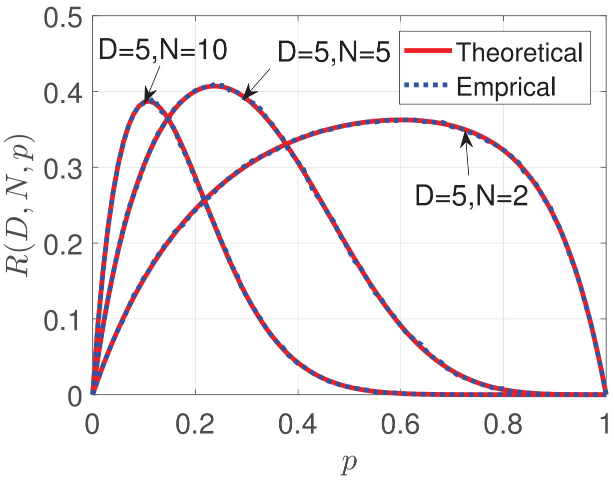

Fig. 2. The system throughput $R(D,N,p)$ with $D=5$.

D. A Conjecture on Optimal Retransmission Probability

Theorem 1 shows that the asymptotic maximum system throughput converges to $1/e = 36.8\%$ for any delay D. This is one key indicator for asymptotic performance that we hope to understand for delay-constrained slotted ALOHA. It suggests that the asymptotic maximum system for delay-constrained case is the same as that for delay-unconstrained case with saturate traffic. Another indicator for asymptotic performance is how the optimal retransmission probability $p^*(D,N)$ changes as N goes to infinity. Our Lemma 3 shows that when $p(N)=1/N$, the system throughput converges to $1/e$, which is exactly the asymptotic maximum system throughput (achieved by the optimal retransmission probability $p^*(D,N)$). It is thus reasonable to conjecture that $p^*(D,N)$ behaves as $1/N$ when N is large enough.

***Conjecture** 1:* For any $D \ge 1$, we have

$$\lim_{N\to\infty} N\cdot p^*(D,N) = 1. \tag{48}$$

Our numerical simulations support this conjecture.

V. Simulation

In this section, we verify our theoretical analysis by simulations and also show asymptotic performance results of the delay-constrained slotted ALOHA system.

A. Verification of Algorithmic Computation

We first verify our algorithm in Sec. III by computing $R(D,N,p)$. We show the theoretical system throughput computed by the algorithm in Sec. III and show the empirical system throughput by simulating a delay-constrained slotted ALOHA system for 50000 slots. The result is shown in Fig. 2. We consider $D=3$ and $N=2,5,10$ and we vary p from 0 to 1. For all cases, we can see that theoretical system throughput matches well with the empirical system throughput. This verifies the correctness of our algorithmic computation in Sec. III.

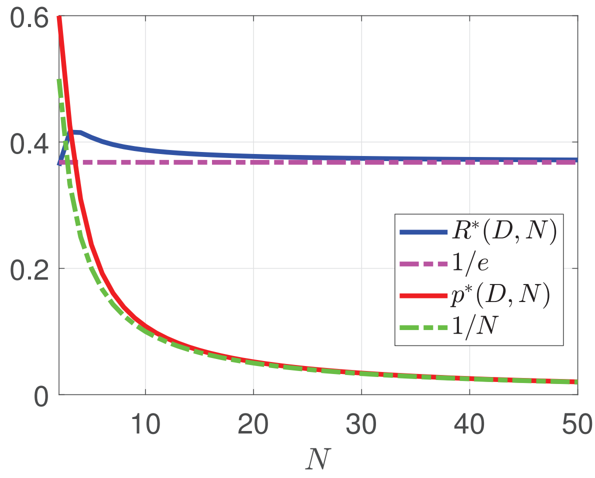

Fig. 3. The optimal retransmission probability and maximum system throughput when $D = 5$.

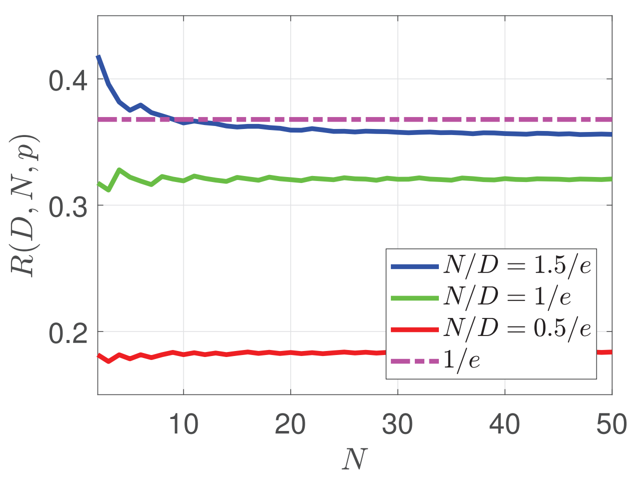

Fig. 4. The maximum system throughput with three different levels of system load.

B. Asymptotic Performance of $R^*(D,N)$ and $p^*(D,N)$

We fix $D = 5$ and change N from 2 to 50. We plot the maximum system throughput $R^*(D,N)$ and the optimal retransmission probability $p^*(D,N)$ in Fig. 3. First, we can see that indeed $R^*(D,N)$ converges to $1/e$, which verifies our main result (Theorem 1). We also note that indeed the $p^*(D,N)$ behaves asymptotically as $1/N$, which verifies Conjecture 1. (In fact, we can further plot the curve $N \cdot p^*(D,N)$ and we can observe that $N \cdot p^*(D,N)$ converges to 1).

C. Asymptotic Performance of Fixed System Load

If we define system load as the average number of packets injected into the system per slot, the system load becomes $\frac{N}{D}$ in the delay-constrained slotted ALOHA system. In our analysis, we fix D and let N goes to infinity. This means that we let the system load $\frac{N}{D}$ go to infinity, or saturated traffic. It would be also interesting to see the asymptotic performance when we fix the system load N/D but scale up the system (i.e., let N go to infinity). We consider three different levels of borderline (around $1/e$) system load $N/D = 1/2e$ (low), $N/D = 1/e$ (medium), and $N/D = 1.5/e$ (high). The results are shown in Fig. 4.

As we can see, the maximum system throughput $R^*(D,N)$ also converges. However, they converges to different values. When the system load is low $(N/D = 1/2e)$ $R^*(D,N)$ converges to $1/2e$, which means that almost all packets can be delivered successfully before expiration. This is because the system load is low and we have much more slots (supplies) than the number of packets (demands) and thus it is possible to serve almost all demands. When the system load is medium $(N/D = 1/e)$, $R^*(D,N)$ converges to $0.3208 = 1/e - 0.0471$. This shows that we cannot successfully deliver all packets and a small fraction of packets will expire. When the system load is high $(N/D = 1.5/e)$, $R^*(D,N)$ converges to $0.3561 = 1/e - 0.0118$. We can see that under borderline traffic, the maximum system throughput can approach $1/e$ at a relatively high speed (when the system load increases from $1/e$ to $1.5/e$, the maximum system throughput is almost close to $1/e$). Overall, we have seen some interesting asymptotic performance results when the system scales up with fixed system load, which deserves further research efforts.

VI. Conclusion

In this paper, we have analyzed the asymptotic performance of the delay-constrained slotted ALOHA system under the frame-synchronized traffic pattern. We have shown that the maximum system throughput converges to $1/e$ when the number of stations goes to infinity. Furthermore, we have investigated the speed of such a maximum system throughput approaching $1/e$ under borderline traffic.

In order to capture more practical scenarios, it may be interesting to investigate the asymptotic performance of the delay-constrained slotted ALOHA system with fixed system load and with non-frame-synchronized traffic pattern.

VII. Acknowledgment

The corresponding author is Yunghsiang S. Han. The first author would like to thank Yijin Zhang and Yuan-Hsun Lo for insightful discussions. This work was partially supported by NSFC of China (No. 61671007) and the Dongguan Science and Technology Program (No. 2015108101006).

References

[1] N. Abramson, "The ALOHA system: Another alternative for computer communications," in *Proc. AFIPS Fall Joint Computing Conference*, 1970, pp. 281–285.

[2] L. G. Roberts, "ALOHA packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, no. 2, pp. 28–42, 1975.

[3] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach, 6/E*. Addison-Wesley, 2013.

[4] W. Rosenkrantz and D. Towsley, "On the instability of slotted ALOHA multiaccess algorithm," *IEEE Transactions on Automatic Control*, vol. 28, no. 10, pp. 994–996, 1983.

[5] S. Ghez, S. Verdu, and S. C. Schwartz, "Stability properties of slotted Aloha with multipacket reception capability," *IEEE Transactions on Automatic Control*, vol. 33, no. 7, pp. 640–649, 1988.

[6] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2636–2656, 2005.

[7] J. Yu and L. Chen, "Stability analysis of frame slotted Aloha protocol," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1462–1474, 2017.

[8] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6815–6832, 2015.

[9] L. Deng, C.-C. Wang, M. Chen, and S. Zhao, "Timely wireless flows with general traffic patterns: Capacity region and scheduling algorithms," *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3473–3486, 2017.

[10] G. P. Fettweis, "The tactile Internet: Applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, 2014.

[11] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile Internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, 2016.

[12] J. Baillieul and P. J. Antsaklis, "Control and communication challenges in networked real-time systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 9–28, 2007.

[13] Z. Jiang, H. Cheng, Z. Zheng, X. Zhang, X. Nie, W. Li, Y. Zou, and W. S. Wong, "Autonomous formation flight of UAVs: Control algorithms and field experiments," in *Proc. CCC*, 2016, pp. 7585–7591.

[14] K. Kang, K.-J. Park, L. Sha, and Q. Wang, "Design of a crossbar VOQ real-time switch with clock-driven scheduling for a guaranteed delay bound," *Real-Time Systems*, vol. 49, no. 1, pp. 117–135, 2013.

[15] Y. Zhang, Y.-H. Lo, F. Shu, and J. Li, "Optimal slotted ALOHA under delivery deadline constraint for multiple-packet reception," *arXiv preprint arXiv:1706.00350*, pp. 1–11, 2017.

[16] I.-H. Hou, V. Borkar, and P. R. Kumar, "A theory of QoS for wireless," in *Proc. IEEE INFOCOM*, 2009, pp. 486–494.

[17] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.