# Capacity Analysis of a Discrete-Time Bufferless Timing Channel

Mehrnaz Tavan, Roy D. Yates, and Waheed U. Bajwa
Department of Electrical and Computer Engineering
Rutgers University
Email: mt579@eden.rutgers.edu, ryates@winlab.rutgers.edu, waheed.bajwa@rutgers.edu

***Abstract*—This paper investigates the capacity of a discrete-time channel in which information is conveyed by the timing of consecutive packets passing through a queue with independent and identically distributed service times. Such timing channels are commonly studied under the assumption of a work-conserving queue. In contrast, this paper studies the case of a discrete-time bufferless queue that drops arriving packets while a packet is in service. Under this bufferless model, the paper provides upper bounds on the capacity of timing channels and establishes capacity for the case of bufferless M/M/1 queue.**

## I. Introduction

Timing channels convey information by the timing of consecutive packets – rather than by their contents. Such channels not only arise in many engineering contexts, such as covert communications [1], [2], energy harvesting communication systems [3], and sensor networks [4], but can also provide a reasonable abstraction of interactions in biological systems [5]. In addition, information theoretic understanding of timing channels can potentially help us attack the challenging problem of causal inference in systems where causal relationships are determined by timing information [6], [7].

Information theoretic study of timing channels began in the seminal paper [8], which characterizes the capacity of a timing channel described by a single-server timing queue (SSTQ) with independent and identically distributed (iid) service times. In particular, we have from [8] that the capacity of an SSTQ with iid exponential service times (·/M/1 queue) is equal to $e^{-1}$ nats per average service time. In [9], Bedekar and Azizoglu analyzed the discrete-time version of SSTQ discussed in [8] with iid service times. The maximum achievable rate for a discrete-time SSTQ with an arrival process of mean $1/\lambda$ time slots and geometrically distributed service time with mean $1/\mu$ time slots was shown to be

$$C(\lambda) = H_{\mathrm{Bin}}(\lambda) - \frac{\lambda}{\mu} H_{\mathrm{Bin}}(\mu), \tag{1}$$

which is the smallest among queues with the same average service time where $H_{Bin}(\cdot)$ is the binary entropy function. Moreover, [9] proved that for the capacity maximizing $\lambda$, the capacity is

$$C = \log\left[1 + \mu\,(1-\mu)^{(1-\mu)/\mu}\right]. \tag{2}$$

In [10] we presented the capacity analysis of continuous-time *bufferless* SSTQ with iid service times that discards incoming packets while a packet is in service. Bufferless SSTQs, despite their apparent simplicity, are effective in mathematically modeling some systems including protein synthesis networks [11]. Our interest in bufferless SSTQs is related to the mutual information in tweet sequences. Suppose Bob receives tweets from Alice and occasionally tweets in response. While formulating a response, Bob ignores subsequent tweets from Alice. In this model, we can view Alice's tweets as arrivals and Bob's tweets as departures from a queue. The time Bob spends formulating a response is the service time of a tweet admitted to the system. While the bufferless SSTQ is a simple model, it provides a starting point for characterizing how much information can be gleaned from tweet timing data.

In [10], we proved a single-letter upper bound on the channel capacity under arbitrary service distributions for the case of iid inter-arrival packet times. In addition, we presented a looser closed-form upper bound on the channel capacity under arbitrary service distributions. Finally, we provided achievability results for bufferless M/M/1 and M/G/1 queues using information density methods [12].

In this paper, we focus our attention on the capacity analysis of discrete-time bufferless SSTQs, which to the best of our knowledge has not been explored in prior work. While the bufferless SSTQ shares some similarities with the buffered SSTQ in [9], we will see that analyzing its capacity presents some new challenges in the absence of a one-to-one correspondence between incoming and departing packets. In contrast to continuous-time model in which mutual information is accumulated at the receiver with each departure from the queue, a discrete-time model enables a finer-grained characterization of mutual information. Specifically, with each discrete step in time, the departure or non-departure of a customer in service contributes to the mutual information accumulated by the receiver.

In this paper, we make the following contributions to the capacity analysis of discrete-time timing channels described by bufferless SSTQs with iid service times. First, we provide a single-letter upper bound on the channel capacity under arbitrary service distributions for the case of iid inter-arrival packet times. Next, we provide a single-letter upper bound and a looser closed-form upper bound on the channel capacity under arbitrary service distributions. For the case of ·/M/1 queue, this closed-form upper bound matches the one for the work conserving queue. Finally, we provide capacity results for bufferless M/M/1 queue.

978-1-4799-3001-2/14/$31.00 ©2014 IEEE

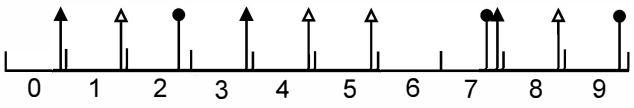


Fig. 1. One realization of the input and output sequence of the system is illustrated. The arrows with hollow arrowheads show the packets arriving at the server that are dropped, the arrows with solid arrowheads show the packets that enter the server, and the lines with circles on top show the packets departing the server. The corresponding input and output sequences are $X_0^9 = \{1,1,0,1,1,1,0,1,1,0\}$ and $Y_0^9 = \{0,0,1,0,0,0,0,1,0,1\}$ respectively.

The rest of the paper is organized as follows. In Section II, we provide an overview of our system, and provide a formal definition of capacity in our setup. Section III derives outer bounds on the timing capacity that hold for all arrival processes. Section IV provides outer bounds for specific arrival processes and service time distributions, and computes the capacity of bufferless M/M/1 queue. Concluding remarks are in Section V.

Note that we use $P_X(\cdot)$ to denote the probability mass function (PMF) of random variable $X$. Similarly $P_{X|Y}(\cdot|\cdot)$ is the conditional PMF of $X$ given $Y$.

## II. System Model

An important feature of the timing channel in [8] and [9] (the former in continuous-time and the latter in discrete-time) is to use packet inter-arrival times to the server to encode a message. The receiver, based on the departure times of packets from the server, decides which message has been transmitted. In this paper, we consider the discrete-time version of the model in [10] consisting of a single server bufferless queue with a zero packet waiting room. Upon arrival at an idle server, a packet immediately enters service; otherwise, if the server is busy with a previous packet, the incoming packet is blocked and discarded. To be consistent with the definition of discrete-time queues in [9], we assume that in each time slot, at most one arrival and one departure can occur.

We use $S_i$ to denote the service time of the $i$th packet admitted to service. As is customary in discrete-time queuing systems, we assume that service times are iid strictly positive integer-valued random variables, independent of packet arrival times. Thus we refer to the timing channel induced by the (bufferless) queue with service time $S$ as the (bufferless) timing channel $S$. In Fig. 1, one realization of the input and output sequences, including arriving, dropped and departing packets, is illustrated under our setup. In particular, a packet that remains in service at the end of slot $i-1$ will receive one unit of service in slot $i$, and if its service is completed, depart the instant before the end of slot $i$. Furthermore, an arrival in slot $i$ occurs at the end of the slot $i$, the instant after a possible departure. Thus, a packet that arrives in slot $i$ begin service in slot $i+1$ and departs no earlier than slot $i+1$ . As a result, the service time is $S \geq 1$.

### A. Encoder

The transmitted message is represented by the discrete uniform index $U \in \{1,\cdots,M\}$. At the transmitter, each message $U=u$ will be encoded in an infinite binary sequence codeword $\overline{X_u} = (X_0 = 1, X_{u,1}, X_{u,2}, \cdots)$. In this system, $X_{u,i} = 1$ means that a packet arrives at the server at the end of slot $i$. We refer to the packet submitted at time 0 as packet zero. This packet carries no timing information and serves only to initialize the system. Similarly, we refer to packets $1,2,\ldots$ as codeword packets as their inter-arrival times define the codewords.

### B. Decoder

At the receiver, the decoder observes the sequence $\overline{Y} = Y_0, Y_1, \cdots, Y_{K_n} \in \{0,1\}^{K_n}$ which is used in estimating the index $V \in \{1,\cdots,M\}$ corresponding to the transmitted message. A decoding error occurs when $V \neq U$. Observing a departure at the end of slot $i$ is represented by $Y_i = 1$.

In our system, the number of observed departure times is fixed at $n$. In this case the length of corresponding departure sequence $Y$ (the number of time slots until the $n$th departure is observed) will be random since it will equal

$$K_n = \min\left\{k | \sum_{i=1}^{k} Y_i = n\right\}. \tag{3}$$

To simplify the entropy computation of a sequence with random length, we first construct an infinite sequence through concatenation of infinite zeros to $Y_0^{K_n}$ resulting in $Y_0^\infty = \left[Y_0^{K_n} \quad 0 \quad 0 \quad \cdots\right]$.

After $n$ codeword packets are received, the maximum a posteriori probability (MAP) decoder observes the departure times $Y_0^\infty = y_0^\infty$ and finds the most probable codeword to have been transmitted. Since the codewords are equiprobable, the decoder solves the maximum likelihood (ML) problem

$$u^*(y_1^\infty) = \arg\max_u P_{Y_\bullet^\infty|U}\left[y_0^\infty|u\right]. \tag{4}$$

Description of the ML decoder is omitted as the implementation would parallel that in [10].

### C. Capacity

In this work, we aim to compute the capacity of the bufferless timing channel. While each decoded message conveys $\log_2 M$ bits of information, the time required by the receiver to decode a message depends on the packet departure times. In particular, we assume that the receiver decodes after observing the departures of $n$ codeword packets. The expected time required to observe these departures is $T_n = \mathrm{E}\left[K_n\right]$.

Following [8], [13] the achievable rate and the capacity for our system are defined as follows.

**Definition 1.** If for every $\gamma > 0$, a sequence of codewords from a codebook with $M_n$ entries exists with $(\log M_n)/T_n > R - \gamma$ for all sufficiently large $n$, and the corresponding maximum probability of error $\epsilon_n$ satisfies $\lim_{n\to\infty} \epsilon_n = 0$, then the rate $R$ is achievable. The maximum rate $R$ that satisfies this definition is called the capacity of the timing channel and is denoted by $C$.

## III. Converse Theorem

We follow the approach of [8] in deriving a converse. Using $P_e$ to denote the probability of a decoding error, we observe that Fano's inequality [14, sec. 2.10] and equiprobable $U$ imply

$$\begin{aligned} H(U|V) &\le H_{\mathrm{Bin}}(P_e) + P_e \log M_n && (5)\\ &\le H_{\mathrm{Bin}}(P_e) + \epsilon_n \log M_n && (6)\\ &\le \log 2 + \epsilon_n \log M_n && (7)\\ &= \log 2 + \epsilon_n \log M_n + H(U) - \log M_n, && (8)\end{aligned}$$

where we assume that $H_{\mathrm{Bin}}(P_e) \le \log 2$. We can conclude that

$$\begin{aligned} \log M_n &\le \frac{1}{1-\epsilon_n}\left[I(U;V) + \log 2\right] && (9)\\ &\le \frac{1}{1-\epsilon_n}\left[I(X_0^\infty; Y_0^\infty) + \log 2\right], && (10)\end{aligned}$$

where (10) follows from the data processing lemma [14, sec. 2.8]. Before stating a converse for our system, we need the following lemmas.

**Lemma 1.** *For any non-negative integer-valued random variable $D$, the following relationship holds*

$$H(D) = \sum_{l=0}^{\infty} H_{\mathrm{Bin}}(P\left[D = l+1 | D > l\right]) P\left[D > l\right]. \quad (11)$$

*Proof:* Starting from the right side of (11),

$$\begin{aligned} &\sum_{l=0}^{\infty} H_{\mathrm{Bin}}(P\left[D = l+1 | D > l\right]) P\left[D > l\right]\\ &= \sum_{l=0}^{\infty} H_{\mathrm{Bin}}\left(\frac{P\left[D = l+1\right]}{P\left[D > l\right]}\right) P\left[D > l\right] && (12)\\ &= -\sum_{l=0}^{\infty}\left[P\left[D = l+1\right]\log\left(\frac{P\left[D = l+1\right]}{P\left[D > l\right]}\right)\right.\\ &\quad \left. + P\left[D > l+1\right]\log\left(\frac{P\left[D > l+1\right]}{P\left[D > l\right]}\right)\right] && (13)\\ &= H(D) - \sum_{l=0}^{\infty}\left[P\left[D > l+1\right]\log\left(P\left[D > l+1\right]\right)\right.\\ &\quad \left. - P\left[D > l\right]\log\left(P\left[D > l\right]\right)\right] && (14)\\ &= H(D), && (15)\end{aligned}$$

where (14) transforms into (15) since the summation in (14) is a telescoping sum. ∎

**Definition 2.** Given a random sequence $L_0, \cdots, L_n$ with PMF $P_{L_j}(l)$, we define, for $j \in \mathbb{Z}$ the indicator random variables

$$I_l(j) = \begin{cases} 1 & L_j = l,\\ 0 & \text{otherwise.}\end{cases} \quad (16)$$

Note that $E\left[I_l(j)\right] = P\left[I_l(j) = 1\right] = P_{L_j}(l)$. Moreover, $\sum_{j=0}^{n} I_l(j)$ is the number of elements in $\{L_0, \cdots, L_n\}$ that equal $l$.

**Theorem 1.** *For a system with iid service process where the arrival and service process are independent, the following relationship holds:*

$$\sum_{i=1}^{\infty} H(Y_i | Y_0^{i-1}, X_0^\infty) = nH(S). \quad (17)$$

*Proof:* At time $i$, let $J_i \in \{0, 1, \cdots, n\}$ specify how many departures we have observed. In addition, let $B_i \in \{-1, 0, 1, \cdots\}$ specify the number of units of service received by a customer in service if any such that $B_i = -1$ indicates that either the server is idle at the end of slot $i$ or $n$ departures have been observed.

Since $J_{i-1}$ and $B_{i-1}$ are deterministic functions of $(Y_0^{i-1}, X_0^\infty)$,

$$H(Y_i | Y_0^{i-1}, X_0^\infty) = H(Y_i | Y_0^{i-1}, B_{i-1}, J_{i-1}). \quad (18)$$

Furthermore,

$$H(Y_i | Y_0^{i-1}, B_{i-1} = b, J_{i-1} = j) = 0 \quad (19)$$

if $b = -1$ or $j = n$. Otherwise, for $0 \le j < n$ and $b \ge 0$,

$$\begin{aligned} &H(Y_i | Y_0^{i-1} = y_0^{i-1}, X_0^\infty = x_0^\infty, B_{i-1} = b, J_{i-1} = j)\\ &= H_{\mathrm{Bin}}(P\left[S_{j+1} > b+1 | S_{j+1} > b\right]).\end{aligned}$$

Consequently,

$$\begin{aligned} &\sum_{i=1}^{\infty} H(Y_i | Y_0^{i-1}, X_0^\infty)\\ &= \sum_{i=1}^{\infty}\sum_{j=0}^{n-1}\sum_{b=0}^{\infty}\left(H_{\mathrm{Bin}}(P\left[S_{j+1} > b+1 | S_{j+1} > b\right])\right.\\ &\quad \left. P\left[B_{i-1} = b, J_{i-1} = j\right]\right) && (20)\\ &= \sum_{b=0}^{\infty} H_{\mathrm{Bin}}(P\left[S > b+1 | S > b\right])\\ &\quad \left(\sum_{i=1}^{\infty}\sum_{j=0}^{n-1} P\left[J_{i-1} = j | B_{i-1} = b\right] P_{B_{i-1}}(b)\right). && (21)\end{aligned}$$

Using $\sum_{j=0}^{n-1} P\left[J_{i-1} = j | B_{i-1} = b\right] = 1$, and following Definition 2, (21) can be written as

$$\begin{aligned} &\sum_{i=1}^{\infty} H(Y_i | Y_0^{i-1}, X_0^\infty)\\ &= \sum_{b=0}^{\infty} H_{\mathrm{Bin}}(P\left[S > b+1 | S > b\right]) \sum_{i=1}^{\infty} P_{B_{i-1}}(b) && (22)\\ &= \sum_{b=0}^{\infty} H_{\mathrm{Bin}}(P\left[S > b+1 | S > b\right]) E\left[\sum_{i=1}^{\infty} I_b(i-1)\right], && (23)\end{aligned}$$

where $\sum_{i=1}^{\infty} I_b(i-1)$ can be interpreted as the number of service times $S_j$ such that $S_j > b$ which can be represented by $\mathrm{Count}_{\mathbf{S}^n_\bullet}(b)$. Based on this definition of the $\mathrm{Count}(.)$ function, $E\left[\mathrm{Count}_{\mathbf{S}^n_\bullet}(b)\right] = nP\left[S > b\right]$. As a result, (23) can be written as

$$\sum_{i=1}^{\infty} H(Y_i|Y_0^{i-1}, X_0^{\infty})$$
$$= \sum_{b=0}^{\infty} H_{\mathrm{Bin}}(P[S > b+1|S > b]) nP[S > b] \quad (24)$$
$$= nH(S). \quad (25)$$

Using Lemma 1, (25) follows from (24). ∎

**Lemma 2.** *The mutual information between the input codeword and the observed output sequence satisfies*

$$I(X_0^{\infty}; Y_0^{\infty}) = H(Y_0^{\infty}) - \sum_{i=1}^{n} H(S_i). \quad (26)$$

*Proof:* By the chain rule,

$$I(X_0^{\infty}; Y_0^{\infty}) = I(X_0^{\infty}; Y_0) + \sum_{i=1}^{\infty} I(X_0^{\infty}; Y_i|Y_0^{i-1}). \quad (27)$$

Since we assume that $Y_0 = 0$, $I(X_0^{\infty}; Y_0) = 0$. Moreover,

$$\sum_{i=1}^{\infty} I(X_0^{\infty}; Y_i|Y_0^{i-1}) = H(Y_0^{\infty}) - \sum_{i=1}^{\infty} H(Y_i|X_0^{\infty}, Y_0^{i-1})$$
$$= H(Y_0^{\infty}) - \sum_{i=1}^{n} H(S_i). \quad (28)$$

Note that (28) holds using Theorem 1. ∎

**Definition 3.** The sequence of packet inter-arrival times are represented by $\overline{A_u} = (A_0 = 0, A_{u,1}, A_{u,2}, \cdots)$ where $A_{u,j} \in \mathbb{N}$ is the inter-arrival time between packets $j-1$ and $j$ in codeword $\overline{X_u}$ (in slots). A fundamental difference between the $\overline{A_u}$ sequence and $\overline{X_u}$ sequence is that even if the $\overline{A_u}$ sequence is iid, the $\overline{X_u}$ sequence is not iid since the elements of $\overline{X_u}$ that correspond to the same $A_i$ are correlated.

The number of time slots between the $(i-1)$th and the $i$th observed departures correspond to $D_i$ where $D_0$ is the departure time of packet 0 and $K_n = \sum_{i=1}^{n} D_i$. In the bufferless queue, the subset of arrivals that are admitted into service is denoted by the subsequence $k_0 = 0, k_1, \cdots$ such that

$$k_i = \min\left\{ m | \sum_{j=1}^{m} A_j - \sum_{j=0}^{i-1} D_j > 0 \right\} \quad (29)$$

denotes the index of the packet $i > 0$ admitted to service. The time that the server is idle between departure $i$ and the next arrival is represented by $W_i$. Since the queue in our system is blocking and has no buffer, the idling time $W_i$ can be represented as a deterministic function of the message index $U$ and prior departures $D_0^i$ as

$$W_i(U, D_0^i) = \sum_{j=1}^{k_{i+1}} A_{U,j} - \sum_{j=0}^{i} D_j. \quad (30)$$

Based on the assumption that in each time slot, at most one arrival and one departure can happen such that if the next arrival occurs in the same timeslot the instant after a departure, the packet would enter the server with zero idling time so the idling time is $W_i \geq 0$. For ease of notation, we use $W_i(U, D_0^i)$ and the shorthand $W_i$ interchangeably. The relationship between departure time $D_i$ and the corresponding idling time and service time is

$$D_i = W_{i-1}(U, D_0^{i-1}) + S_i. \quad (31)$$

Equivalent to (30) and (31), we can explicitly represent $W_i$ and $D_i$ as functions of the arrival times $A_1^{\infty}$ and past departures $D_0^{i-1}$:

$$W_i(A_1^{\infty}, D_0^i) = \sum_{j=1}^{k_{i+1}} A_j - \sum_{j=0}^{i} D_j, \quad (32)$$
$$D_i = W_{i-1}(A_1^{\infty}, D_0^{i-1}) + S_i. \quad (33)$$

Since we assumed that in each time slot, at most one arrival and departure can happen, $A_{u,i} \geq 1$ for $i \geq 1$and $D_{u,j} \geq 1$ for $1 \leq j \leq n$.

Moreover, $T_n$ can be rewritten as

$$T_n = \sum_{i=0}^{n} \mathrm{E}[D_i] = \mathrm{E}[S_0] + \sum_{i=1}^{n} \mathrm{E}[W_{i-1} + S_i]. \quad (34)$$

**Lemma 3.** *The mutual information between the input codeword and the output departure times satisfies*

$$I(X_0^{\infty}; Y_0^{\infty}) \leq \sum_{i=1}^{n} I(W_{i-1}; W_{i-1} + S_i). \quad (35)$$

*Proof:* Based on Lemma 2

$$I(X_0^{\infty}; Y_0^{\infty}) = H(Y_0^{\infty}) - \sum_{i=1}^{n} H(S_i). \quad (36)$$

Since there is a bijective map between $Y_0^{\infty}$and $D_0^n$, $H(Y_0^{\infty}) = H(D_0^n)$ [15]. Since conditioning reduces the entropy,

$$H(D_0^n) = \sum_{i=0}^{n} H(D_i|D_0^{i-1}) \leq \sum_{i=0}^{n} H(D_i).$$

As a result, (36) can be written as

$$I(X_0^{\infty}; Y_0^{\infty}) \leq \sum_{i=1}^{n} (H(W_{i-1} + S_i) - H(S_i)). \quad (37)$$

∎

To develop universal bounds valid for all arrival and service processes, we follow the approach in [8] and define

$$c(a) \equiv \sup_{\substack{Z \geq 0 \\ \mathrm{E}[Z] \leq a}} I(Z; Z + S), \quad (38)$$

where $Z$ is independent of $S$. We note that $c(a)$ is a monotone concave function in the argument $a$, and that this will provide a universal upper bound on the capacity of the timing channel.

**Theorem 2.** *The discrete-time timing channel $S$ with $\mathrm{E}[S] = 1/\mu$ has capacity*

$$C \leq \overline{C}(S) \equiv \sup_{\lambda > 0} \frac{c(\frac{1}{\lambda} - 1)}{\frac{1}{\lambda} + \frac{1}{\mu} - 1}. \quad (39)$$

*Proof:* Let

$$R_n = \frac{(1 - \epsilon_n) \log M_n}{T_n}. \quad (40)$$

Combining (34), (10), and

$$I(X_0^\infty; Y_0^\infty) \le \sum_{i=1}^{n} c\left(\mathrm{E}\left[W_{i-1}\right]\right), \tag{41}$$

(the proof of (41) follows the approach in [10, Lemma 3] using Lemma 3) yields

$$R_n \le \frac{\frac{1}{n}\left[\sum_{i=1}^{n} c(E\left[W_{i-1}\right]) + \log 2\right]}{\frac{\mathrm{E}[S_0]}{n} + \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left[W_{i-1}\right] + \mathrm{E}\left[S_i\right]\right)}. \tag{42}$$

Defining $\lambda_n^{-1} - 1 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[W_{i-1}\right]$, concavity of $c(a)$ implies

$$R_n \le \frac{c(\frac{1}{\lambda_n} - 1) + \frac{\log 2}{n}}{\frac{1}{\lambda_n} + \frac{1}{\mu} - 1} \le \sup_{\lambda > 0} \frac{c(\frac{1}{\lambda} - 1)}{\frac{1}{\lambda} + \frac{1}{\mu} - 1} + \frac{\log 2}{n/\mu}. \tag{43}$$

The claim follows as $n \to \infty$. ∎

We can further loosen Theorem 2 by making use of the following lemma.

**Lemma 4.** *For a timing channel $S$, $c(a)$ defined in (38) satisfies*

$$c(a) \le (a + \mathrm{E}\left[S\right]) H_{\mathrm{Bin}}((a + \mathrm{E}\left[S\right])^{-1}) - H(S). \tag{44}$$

*Proof:* Based on (38), we have

$$c(a) = \sup_{\substack{E[Z] \le a \\ Z \ge 0}} H(Z + S) - H(S). \tag{45}$$

Notice that $H(Z+S)$ subject to the constraints $E\left[Z\right] \le a$ and $Z \ge 0$ and fixed service distribution will be maximized when $Z+S$ has geometric distribution with success probability $(a + \mathrm{E}\left[S\right])^{-1}$ [14]. The proof now follows from the entropy of a geometric distribution. ∎

However, there is no guarantee that for any given service distribution, there exists a nonnegative discrete random variable with $\mathrm{E}\left[Z\right] \le a$ such that its summation with $S$ has geometric distribution. As a result, $(a + \mathrm{E}\left[S\right]) H_{\mathrm{Bin}}((a + \mathrm{E}\left[S\right])^{-1}) - H(S)$ is an upper bound on $c(a)$. A universal upper bound on the capacity of the system can now be stated.

**Theorem 3.** *The bufferless timing queue $S$ with $\mathrm{E}\left[S\right] = 1/\mu$ has capacity*

$$C \le \log\left(1 + e^{-H(S)}\right), \tag{46}$$

*when $H(S) \ge \log\left(\frac{1}{\mu} - 1\right)$.*

*Proof:* Based on Theorem 2 and Lemma 4, $R_n$ defined in (40) satisfies

$$R_n \le \sup_{\lambda > 0} \frac{\left(\frac{1}{\lambda} + \frac{1}{\mu} - 1\right) H_{\mathrm{Bin}}(\left(\frac{1}{\lambda} + \frac{1}{\mu} - 1\right)^{-1}) - H(S)}{\frac{1}{\lambda} + \frac{1}{\mu} - 1}. \tag{47}$$

By taking the derivative of the upper bound in (47) with respect to $\lambda^{-1}$, the optimal $\lambda$, $\lambda^*$ will satisfy

$$H(S) = \log\left(\frac{1}{\lambda^*} + \frac{1}{\mu} - 2\right). \tag{48}$$

Since $\lambda$ is a nonnegative number between 0 and 1, $H(S) \ge \log(\frac{1}{\mu} - 1)$ must hold, and the upper bound (46) follows. ∎

For the case of geometric service distribution, we obtain the same outer bound as [9] (see (2)).

## IV. Queue-specific Outer Bounds

Although Lemmas 2 and 3 hold for any arrival process, characterization of $W_i$ will be difficult in case of having memory in the arrivals. To go further, we focus on the special case of codebooks with iid inter-arrival times. With iid inter-arrivals, each time a packet enters service, the queue undergoes a renewal. In particular, the $i$th renewal point marks the beginning of a service time $S_i$ and a set of subsequent iid packet inter-arrival times $A_{k_i+1}, A_{k_i+2}, \ldots$ such that the distributions of $S_i$ and $\{A_{k_i+j}\}$ are sufficient to evaluate the distribution of the number of packet arrivals that are dropped during the service as well as the idling time $W_i$ that follows the service completion. Because service times and inter-arrival times are both iid, a renewal occurs at the end of the idling period when the next arrival is admitted. We note that $W_i$ depends on $S_i$; however the renewal implies that $(S_0, W_0), (S_1, W_1), \cdots, (S_n, W_n)$ constitute independent tuples. Based on this observation, the following outer bound yields for

$$C \le \overline{C}(A, S) \equiv \frac{I(W; W + S)}{\mathrm{E}\left[W\right] + \mathrm{E}\left[S\right]}, \tag{49}$$

where $W$ is independent of $S$ but has the idling time distribution induced by $A$ and $S$. The proof follows the approach in [10, Theorem 3].

In general, computing the PMF of $W$ is nontrivial as it can involve $n$-fold convolutions of the PMF of $A_i$. Thus, the primary use of (49) is for the case when the $X_i$ form a Bernoulli arrival process with probability of $X_i$ being 1 equal to $\lambda$. In this case, the idling times $W_i$ are Geometric $(\lambda)$ random variables, $P_W(w) = \lambda(1-\lambda)^w, w = 0, 1, 2, \cdots$ independent of $S$, and the queueing system is an M/G/1 single server bufferless queue. For Bernoulli arrival process, the outer bound $\overline{C}(A, S)$ reduces to a straightforward numerical evaluation of $I(W; W + S)$.

As a special case, we analyze the discrete-time M/M/1 queue in which the service time is geometric with success probability $\mu$ so $P_S(s) = \mu(1-\mu)^{s-1}, s = 1, 2, 3, \cdots$. In this case, $S$ will have entropy

$$H(S) = \frac{H_{\mathrm{Bin}}(\mu)}{\mu} \tag{50}$$

and $D = W + S$ will have the following distribution:

$$P_D(d) = \frac{\mu\lambda}{\mu - \lambda}\left((1-\lambda)^d - (1-\mu)^d\right), \qquad d = 1, 2, 3, \cdots, \tag{51}$$

which we call Hypogeometric distribution whose entropy is $H(D) = H_{\mathrm{hypo}}(\lambda, \mu)$.

**Proposition 4.** The capacity of the discrete timing channel with memoryless service time with mean $1/\mu$ and memoryless arrival time with mean $1/\lambda$ is

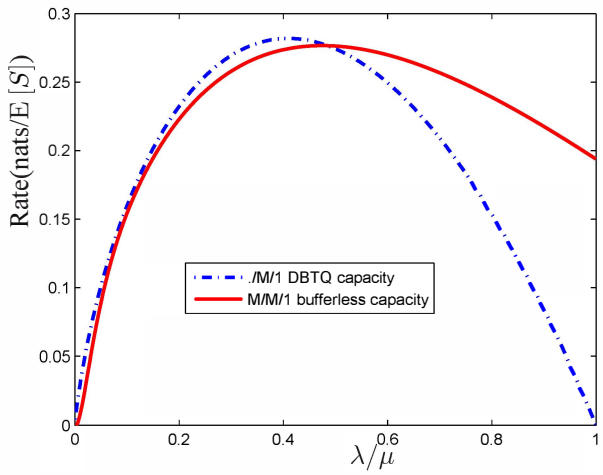


Fig. 2. Comparison of $C(\lambda)$ of the (DBTQ) paper [9, Proposition 4], and the capacity $R(\lambda,\mu)$ for the M/M/1 bufferless queue when $\mu = 0.6$. All the systems have geometric service time of success probability $\mu$ and arrival mean value of $1/\lambda$. The DBTQ plot has maximum equal to 0.2820 nats per average service time whereas the maximum achievable rate is 0.2767 nats per average service time.

$$\overline{C}(A,S) = \frac{H_{\text{hypo}}(\lambda,\mu) - \frac{H_{\text{bin}}(\mu)}{\mu}}{1/\lambda + 1/\mu - 1}. \tag{52}$$

*Proof:* Since $I(W;W+S) = H(D) - H(S)$, (49) yields the outer bound

$$\overline{C}(A,S) = R(\lambda,\mu) \tag{53}$$

where

$$R(\lambda,\mu) \equiv \frac{H_{\text{hypo}}(\lambda,\mu) - \frac{H_{\text{bin}}(\mu)}{\mu}}{1/\lambda + 1/\mu - 1}, \tag{54}$$

which completes the converse proof. The achievability part of the proposition can be proven using information density methods introduced in [12]. ■

The entropy $H_{\text{hypo}}(\lambda,\mu)$ cannot be computed in a closed form. Using numerical methods, the capacity in (53) is computed as a function of $\lambda/\mu$ as shown in Fig. 2 for $\mu = 0.6$. The ·/M/1 capacity of [9, eq. 7] is plotted for comparison (the supremum of which over $\lambda$ is equal to supremum of (46) over $\lambda$). In these plots, $0 < \lambda/\mu < 1$ since the ·/M/1 queue in [9, eq. 7] requires $\lambda \leq \mu$ for stability. It can be seen from this figure that when $\lambda/\mu$ is close to zero, corresponding to a queue that is idle most of the time, the capacity is also close to zero; this is to be expected since the time required to receive $n$ packets will be large in this case. On the other hand, when $\lambda \gg \mu$, the expected idling time reduces, but more and more packets are dropped, and it becomes difficult for the receiver to decode messages, resulting in a decreasing capacity.

The maximum achievable rate in (52) is 0.2767 nats per average server time, and the maximum of the universal upper bound in Theorem 3 is 0.282 when $\mu = 0.6$.

## V. Conclusion

This paper studied the capacity of discrete-time timing channels described by bufferless single-server timing queues with iid service times. One of the main challenges in the analysis of such timing channels is the lack of a one-to-one correspondence between packets arriving at and departing from the queue. This challenge was circumvented by resorting to codewords with infinite length, with the rate of the code defined using the average time it takes to observe the departure of $n$ codeword packets. In general, we believe that an information-theoretic understanding of the setup studied in here will help us address the challenge of causal inference in systems, such as (online) social networks, that lack a one-to-one correspondence between different actions (e.g., tweets versus retweets). In this regard, this paper discussed provided upper bounds on the channel capacity—including a single-letter upper bound and a looser universal upper bound, and computed achievable rates for bufferless M/M/1. Computing tighter upper bounds on the capacity and achievable rates for ·/M/1 and ·/G/1 queues that meet the upper bounds remain areas of future work.

## References


[1] S. Gianvecchio and H. Wang, "Detecting covert timing channels: an entropy-based approach," in *Proceedings of the 14th ACM conference on Computer and communications security*. ACM, 2007, pp. 307–316.

[2] J. Giles and B. Hajek, "An information-theoretic and game-theoretic study of timing channels," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2455–2477, 2002.

[3] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "Binary energy harvesting channel with finite energy storage," in *IEEE International Symposium on Information Theory*, 2013.

[4] G. Morabito, "Exploiting the timing channel to increase energy efficiency in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1711–1720, 2011.

[5] B. Krishnaswamy, C. M. Henegar, J. P. Bardill, D. Russakow, G. L. Holst, B. K. Hammer, C. R. Forest, and R. Sivakumar, "When bacteria talk: Time elapse communication for super-slow networks," in *IEEE International Conference on Communications*. IEEE, 2013, pp. 6348–6353.

[6] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of computational neuroscience*, vol. 30, no. 1, pp. 17–44, 2011.

[7] Y. Liu and S. Aviyente, "Information theoretic approach to quantify causal neural interactions from EEG," in *Asilomar Conference on Signals, Systems and Computers*, 2010, pp. 1380–1384.

[8] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[9] A. S. Bedekar and M. Azizoglu, "The information-theoretic capacity of discrete-time queues," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 446–461, 1998.

[10] M. Tavan, R. D. Yates, and W. U. Bajwa, "Bits through bufferless queues," in *51st Annual Allerton Conference on Communication, Control, and Computing*, 2013.

[11] C. Osorio and M. Bierlaire, "A tractable analytical model for large-scale congested protein synthesis networks," *European J. Operational Research*, vol. 219, no. 3, pp. 588–597, 2012.

[12] S. Verdu and T. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[13] R. Sundaresan and S. Verdu, "Robust decoding for timing channels," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 405–419, 2000.

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-interscience, 2012.

[15] B. Prabhakar and R. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 357–370, 2003.