Bits Through Bufferless Queues

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 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 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 achievable rates for the case of bufferless M/M/1 and M/G/1 queues. In particular, it is shown that a bufferless M/M/1 queue at worst suffers less than 10% reduction in capacity when compared to an M/M/1 work-conserving 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] and sensor networks [3], but can also provide a reasonable abstraction of interactions in biological systems [4]. 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 [5], [6].

The study of information theoretic timing channels began in the seminal paper [7], 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 [7] that the capacity of an SSTQ with iid exponential service times ($\cdot$/M/1 queue) is equal to e^{-1} nats per average service time.

In this paper, we are also interested in studying the capacity of a timing channel described by an SSTQ. However, in contrast to [7], our focus is on a *bufferless* SSTQ 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 [8]. 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.

To the best of our knowledge, however, the capacity of bufferless SSTQs in the context of timing channels has not been explored in prior work. And while the bufferless SSTQ shares some similarities with the buffered SSTQ in [7], 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 this paper, we make the following contributions to the capacity analysis of timing channels described by bufferless SSTQs with iid service times. First, we describe the maximum likelihood (ML) decoder for decoding timing messages transmitted through a bufferless queue. Second, 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. Finally, we provide achievability results for bufferless M/M/1 and M/G/1 queues using information density methods [9]. In particular, for the bufferless M/M/1 queue, achievable rates are shown to coincide with our outer bound. In addition, it is shown that a bufferless M/M/1 queue at worst suffers less than 10% reduction in achievable rate when compared to an M/M/1 queue with infinite buffer [7].

We conclude with a brief discussion of other related work on timing channels. The setup studied in [7] corresponds to a continuous timing channel. A discrete-time version of this timing channel is analyzed in [10], [11]. In [12], [13], the authors revisited the timing channel of [7] and provided capacity analysis from the viewpoint of point processes. Finally, extensions of [7] for the case when the distribution of service times has bounded support is investigated in [14] and for the case of a compound timing channel described by a tandem of queues is analyzed in [15]. In all these works, however, the fundamental assumption is that the queues are work conserving.

The rest of the paper is organized as follows. In Section II, we provide an overview of our system, describe the optimal receiver, 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. Section V investigates achievable rates in our system and compares them to the outer bounds obtained in Sections III and IV. Concluding remarks are in Section VI.

Note that we use $f_X(\cdot)$ to denote the probability density

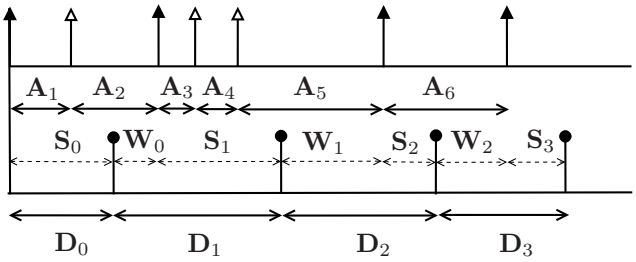

Fig. 1. One realization of the input and output sequence of the system is illustrated. The arrows with hollow arrowheads show the packets arriving at the server that are dropped, the arrows with solid arrowheads show the packets that enter the server, and the lines with circles on top show the packets departing the server.

function (PDF) of random variable X. Similarly $f_{X|Y}(\cdot|\cdot)$ is the conditional PDF of X given Y. In addition, $\exp(\cdot)$ denotes the inverse of $\log x$: $\exp(\log x) = x$.

II. System Model

The basic idea of the timing channel in [7] 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 contrast to [7], we consider a channel that consists 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.

In the following, we use S_i to denote the service time of the i^{th} packet admitted to service. As is customary in queuing systems, we assume that service times are iid 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.

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 into an infinite sequence of packet inter-arrival times $\overline{A_u} = (A_0 = 0, A_{u,1}, A_{u,2}, \cdots)$ where $A_{u,j}$ is the inter-arrival time between packets $j-1$ and j in codeword u. We refer to the packet submitted at time $A_0 = 0$ 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. We note that using a codeword with infinite length is not a new phenomenon and has been applied in [16] to ARQ systems where they design an infinite length codeword and transmit a part of it to the receiver or in [17] where an infinite length codeword is transmitted and in the receiver, after observing each packet, the decoding is performed.

B. Decoder

At the receiver, the decoder observes the inter-departure times $D_0, D_1, \cdots, D_n$ where D_0 is the departure time of packet 0 and D_i, $i > 0$, denotes the time between packet departures $i-1$ and i. These inter-departure times are used in estimating the index $V \in \{1, \cdots, M\}$ corresponding to the transmitted message. A decoding error occurs when $V \neq U$. 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

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

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

$$W_i(U, D_0^i) = \sum_{j=1}^{k_{i+1}} A_{U,j} - \sum_{j=0}^{i} D_j. \tag{2}$$

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

$$D_i = W_{i-1}(U, D_0^{i-1}) + S_i. \tag{3}$$

Equivalent to (2) and (3), we can explicitly represent W_i and D_i as functions of the arrival times A_1^∞ and past departures D_0^{i-1}:

$$W_i(A_1^\infty, D_0^i) = \sum_{j=1}^{k_{i+1}} A_j - \sum_{j=0}^{i} D_j, \tag{4}$$

$$D_i = W_{i-1}(A_1^\infty, D_0^{i-1}) + S_i. \tag{5}$$

After n codeword packets are received, the MAP decoder observes the departure times $D_0^n = d_0^n$ and finds the most probable codeword

$$u^*(d_0^n) = \arg\max_u P\left[U = u | D_0^n = d_0^n\right] \tag{6}$$

to have been transmitted. Since the codewords are equiprobable, we can rewrite (6) as the maximum likelihood problem

$$u^*(d_0^n) = \arg\max_u f_{D_0^n|U}\left[d_0^n | u\right] \tag{7}$$

$$= \arg\max_u f_{D_0}[d_0] \prod_{i=1}^{n} f_{D_i|D_0^{i-1},U}\left[d_i | d_0^{i-1}, u\right]. \tag{8}$$

$$= \arg\max_u \sum_{i=1}^{n} \log f_{D_i|D_0^{i-1},U}\left[d_i | d_0^{i-1}, u\right]. \tag{9}$$

Since $W_{i-1} = W_{i-1}(U, D_0^{i-1})$ is a deterministic function of

U, D_0^{i-1},

$$
\begin{aligned}
&f_{D_i|D_0^{i-1},U}\left[d_i|d_0^{i-1},u\right] \\
&\quad = f_{D_i|D_0^{i-1},U,W_{i-1}}\left[d_i|d_0^{i-1},u,w_{i-1}\right] && (10)\\
&\quad = f_{S_i|D_0^{i-1},U,W_{i-1}}\left[d_i-w_{i-1}|d_0^{i-1},u,w_{i-1}\right] && (11)\\
&\quad = f_{S_i}\left[d_i-w_{i-1}(u,d_0^{i-1})\right]. && (12)
\end{aligned}
$$

Note that (11) holds since $D_i = w_{i-1} + S_i$ given $W_{i-1} = w_{i-1}$ and that (12) follows since S_i is independent of $U, D_0^{i-1}, W_{i-1}(U, D_0^{i-1})$. Combining (9) and (12) and writing w_{i-1} explicitly as a function of u and d_0^{i-1}, we obtain

$$u^*(d_0^n) = \arg\max_u \sum_{i=1}^{n} \log f_{S_i}\left[d_i - w_{i-1}(u, d_0^{i-1})\right]. \quad (13)$$

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

$$T_n = \sum_{i=0}^{n} E\left[D_i\right] = E\left[S_0\right] + \sum_{i=1}^{n} E\left[W_{i-1} + S_i\right]. \quad (14)$$

Following [7], [18] 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 ϵ_n satisfying $\lim_{n\to\infty} \epsilon_n = 0$, 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 Theorems

We follow the approach of [7] in deriving a converse. Using P_e to denote the probability of a decoding error, we observe that Fano's inequality [19, sec. 2.10] and equiprobable U imply

$$
\begin{aligned}
H(U|V) &\le H(P_e) + P_e \log M_n && (15)\\
&\le H(P_e) + \epsilon_n \log M_n && (16)\\
&\le \log 2 + \epsilon_n \log M_n && (17)\\
&= \log 2 + \epsilon_n \log M_n + H(U) - \log M_n, && (18)
\end{aligned}
$$

where we assume that $H(P_e) \le \log 2$. We can conclude that

$$
\begin{aligned}
\log M_n &\le \frac{1}{1-\epsilon_n}\left[I(U;V) + \log 2\right] && (19)\\
&\le \frac{1}{1-\epsilon_n}\left[I(A_1^\infty; D_0^n) + \log 2\right], && (20)
\end{aligned}
$$

where (20) follows from the data processing lemma [19, sec. 2.8].

Before stating a converse for our system, we need the following lemmas.

Lemma 1. *The mutual information between the input codeword and the output departure times satisfies*

$$I(A_1^\infty; D_0^n) = \sum_{i=1}^{n} \left(h(W_{i-1} + S_i|D_0^{i-1}) - h(S_i)\right). \quad (21)$$

Proof: By the chain rule,

$$I(A_1^\infty; D_0^n) = I(A_1^\infty; D_0) + \sum_{i=1}^{n} I(A_1^\infty; D_i|D_0^{i-1}). \quad (22)$$

Since $D_0 = S_0$, which is independent of the code packet arrivals A_1^∞,

$$I(A_1^\infty; D_0) = I(A_1^\infty; S_0) = 0. \quad (23)$$

Moreover,

$$
\begin{aligned}
&I(A_1^\infty; D_i|D_0^{i-1}) \\
&= h(D_i|D_0^{i-1}) - h(D_i|A_1^\infty, D_0^{i-1}) && (24)\\
&= h(W_{i-1} + S_i|D_0^{i-1}) - h(W_{i-1} + S_i|A_1^\infty, D_0^{i-1}) && (25)\\
&= h(W_{i-1} + S_i|D_0^{i-1}) - h(S_i|A_1^\infty, D_0^{i-1}, W_{i-1}) && (26)\\
&= h(W_{i-1} + S_i|D_0^{i-1}) - h(S_i). && (27)
\end{aligned}
$$

Note that (26) holds since A_1^∞, D_0^{i-1} deterministically specify W_{i-1} using (4); (27) holds since S_i is independent of the arrivals A_1^∞, the prior departures D_0^{i-1} and the idle period W_{i-1}. The lemma follows from (22), (23) and (27). ■

Lemma 2. *The mutual information between the input codeword and the output departure times satisfies*

$$I(A_1^\infty; D_0^n) \le \sum_{i=1}^{n} I(W_{i-1}; W_{i-1} + S_i). \quad (28)$$

Proof: Based on Lemma 1,

$$
\begin{aligned}
I(A_1^\infty; D_0^n) &= \sum_{i=1}^{n} \left(h(W_{i-1} + S_i|D_0^{i-1}) - h(S_i)\right) \\
&\le \sum_{i=1}^{n} \left(h(W_{i-1} + S_i) - h(S_i)\right). && (29)
\end{aligned}
$$

Note that (29) holds since conditioning reduces entropy. ■

To develop universal bounds valid for all arrival and service processes, we follow the approach in [7] and define

$$c(a) \equiv \sup_{\substack{X\ge 0 \\ E[X]\le a}} I(X; X+S) \quad (30)$$

where X 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. We start with a relaxation of Lemma 2.

Lemma 3. *The mutual information between the input code-*

word and the output departure times satisfies

$$I(A_1^\infty; D_0^n) \le \sum_{i=1}^{n} c\left(\mathrm{E}\left[W_{i-1}\right]\right). \tag{31}$$

Proof: Lemma 2 and (30) imply

$$I(A_1^\infty; D_0^n) \le \sum_{i=1}^{n} I(W_{i-1}; W_{i-1}+S_i) \tag{32}$$

$$\le \sum_{i=1}^{n} \sup_{\substack{X_i \ge 0 \\ \mathrm{E}[X_i] \le \mathrm{E}[W_{i-1}]}} I(X_i; X_i+S_i) \tag{33}$$

$$= \sum_{i=1}^{n} c\left(\mathrm{E}\left[W_{i-1}\right]\right). \tag{34}$$

■

Now using Lemma 3 , we can define a general converse which is parallel to [7, Thm. 2].

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

$$C \le \overline{C}(S) \equiv \sup_{\lambda>0} \frac{c(\frac{1}{\lambda})}{\frac{1}{\lambda}+\frac{1}{\mu}}. \tag{35}$$

Proof: Let

$$R_n = \frac{(1-\epsilon_n)\log M_n}{T_n}. \tag{36}$$

Combining (14), (20), and Lemma 3 yields

$$R_n \le \frac{\frac{1}{n}\left[\sum_{i=1}^{n} c(E\left[W_{i-1}\right]) + \log 2\right]}{\frac{\mathrm{E}[S_0]}{n} + \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left[W_{i-1}\right] + \mathrm{E}\left[S_i\right]\right)}. \tag{37}$$

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

$$R_n \le \frac{c(\frac{1}{\lambda_n}) + \frac{\log 2}{n}}{\frac{1}{\lambda_n}+\frac{1}{\mu}} \le \sup_{\lambda>0} \frac{c(\frac{1}{\lambda})}{\frac{1}{\lambda}+\frac{1}{\mu}} + \frac{\log 2}{n/\mu}. \tag{38}$$

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

We can further loosen Theorem 1 by making use of the following lemma.

Lemma 4. *For a timing channel S, $c(a)$ defined in (30) satisfies*

$$c(a) \le \log(e) + \log(a + \mathrm{E}\left[S\right]) - h(S). \tag{39}$$

Proof: Based on (30), we have

$$c(a) = \sup_{\substack{E[X] \le a \\ X \ge 0}} h(X+S) - h(S). \tag{40}$$

Notice that $h(X+S)$ subject to the constraints $E\left[X\right] \le a$ and $X \ge 0$ and fixed service distribution will be maximized when $X+S$ has exponential distribution with rate $(a+\mathrm{E}\left[S\right])^{-1}$ [19]. The proof now follows from the entropy of an exponential distribution. ■

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

Theorem 2. *The bufferless timing queue S with $\mathrm{E}\left[S\right] = 1/\mu$ has capacity*

$$C \le \begin{cases} \frac{\log e + \log\left(\frac{1}{\mu}\right) - h(S)}{\mu^{-1}}, & h(S) < \log\left(1/\mu\right), \\ \frac{\log e}{\exp(h(S))}, & h(S) \ge \log\left(1/\mu\right). \end{cases} \tag{41}$$

Proof: Based on Theorem 1 and Lemma 4, R_n defined in (36) satisfies

$$R_n \le \sup_{\lambda>0} \frac{\log e + \log\left(\frac{1}{\lambda}+\frac{1}{\mu}\right) - h(S)}{\frac{1}{\lambda}+\frac{1}{\mu}}. \tag{42}$$

By taking the derivative of the upper bound in (42) with respect to λ^{-1}, the optimal λ will satisfy

$$h(S) = \log\left(\frac{1}{\lambda^*} + \frac{1}{\mu}\right). \tag{43}$$

Since λ is a nonnegative number, when $h(S) < \log(1/\mu)$, the supremum is approached as $\lambda^{-1} \to 0$ and the universal upper bound will be

$$R_n \le \mu\left[\log e + \log(1/\mu) - h(S)\right]. \tag{44}$$

Otherwise,

$$R_n \le \frac{\log e}{\exp\left(h(S)\right)}. \tag{45}$$

■

IV. Queue-specific Outer Bounds

We note that Lemmas 1 and 2 make no particular assumptions regarding the statistical structure of the arrivals. However, in the absence of such assumptions, memory in the arrivals can induce idling times W_i that are difficult to characterize. 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 ith 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. This observation yields the following outer bound for iid inter-arrivals.

Theorem 3. *With iid inter-arrival times identical to A, the bufferless timing channel S has capacity C satisfying*

$$C \le \overline{C}(A,S) \equiv \frac{I(W; W+S)}{E\left[W\right] + E\left[S\right]}, \tag{46}$$

where W is independent of S but has the idling time distribution induced by A and S.

Proof: Since each service initiation marks a renewal, Lemma 2 reduces to

$$I(A_1^\infty; D_0^n) \le nI(W_{i-1}; W_{i-1} + S_i). \tag{47}$$

In addition, (14) yields

$$T_n = E[S_0] + n\left(E[W_{i-1}] + E[S_i]\right). \tag{48}$$

Combining (20), (47) and (48) yields

$$R_n \le \frac{nI(W_{i-1}; W_{i-1} + S_i) + \log 2}{\mathrm{E}[S_0] + n\left(\mathrm{E}[W_{i-1}] + \mathrm{E}[S_i]\right)}. \tag{49}$$

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

In general, computing the PDF of W is nontrivial as it can involve n-fold convolutions of the PDF of A_i. Thus, the primary use of Theorem 3 is the case when the A_i form a rate λ Poisson arrival process. In this case, the idling times W_i are exponential (λ) random variables independent of S, and the queueing system is an M/G/1 single server bufferless queue. For Poisson arrivals, 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 M/M/1 queue in which the service time is exponential with rate μ. In this case, S will have entropy

$$h(S) = \log e + \log \frac{1}{\mu} \tag{50}$$

and $D = W + S$ will have the hypoexponential distribution

$$f_D(d) = \frac{\mu\lambda}{\mu - \lambda}\left(e^{-\lambda d} - e^{-\mu d}\right), \qquad d \ge 0, \tag{51}$$

and entropy $h(D) = h_{\text{hypo}}(\lambda, \mu)$. Since $I(W; W + S) = h(D) - h(S)$, Theorem 3 yields the outer bound

$$\overline{C}(A, S) = R(\lambda, \mu) \tag{52}$$

where

$$R(\lambda, \mu) \equiv \frac{h_{\text{hypo}}(\lambda, \mu) - \log e + \log \mu}{1/\lambda + 1/\mu}. \tag{53}$$

The entropy $h_{\text{hypo}}(\lambda, \mu)$ cannot be computed in a closed form. Using numerical integration methods, the upper bound in (52) is computed as a function of λ/μ as shown in Fig. 2 (see Appendix A for proof that (53) is a function of λ/μ for fixed μ). It can be seen from this figure that when λ/μ is close to zero, corresponding to a queue that is idle most of the time, the upper bound on 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 upper bound on capacity.

Fig. 2 compares the Theorem 2 universal upper bound for the ·/M/1 queue to the upper bound derived for M/M/1 queue in (52). It can be seen from this figure that although the Theorem 2 bound is looser than (52), the two upper bounds almost coincide for $0 < \lambda/\mu < 0.2$.

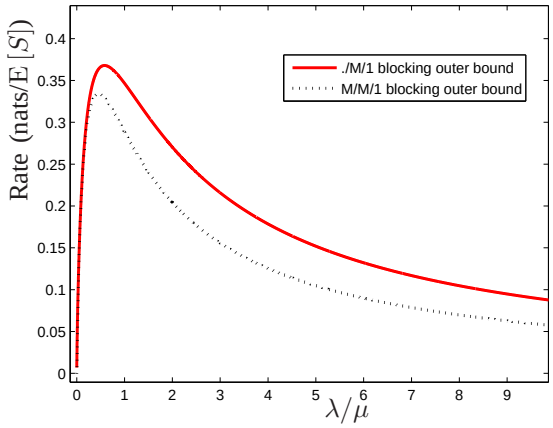

Fig. 2. Comparison between the bufferless ·/M/1 and M/M/1 queue upper bounds for $0 < \lambda/\mu < 10$ where λ is the arrival rate and μ is the service rate. The M/M/1 upper bound coincides with the achievable rate for M/M/1 as well.

V. Achievability

While the upper bounds in Sections III and IV make use of the maximization of the mutual information between idling time and inter-departure time, the only parameters in our control for coding purposes are the inter-arrival times. In order for our system to achieve the Theorem 2 upper bound, two conditions must be fulfilled: 1) The inter-departure times must be iid so $h(W_{k-1}+S_k|D_1^{k-1}) = h(W_{k-1}+S_k)$ which leads to equality in (28); 2) The inter-arrival times must be distributed such that asymptotically, the induced idling time maximizes (35). The first condition is satisfied only when the service time is exponential; Otherwise, the relationship between W_i and S_i would create dependency between consecutive inter-departure times D_i and D_{i+1}. When S has exponential distribution, [7, Theorem 3] shows that among the distributions with $\mathrm{E}[W + S] \le 1/\lambda + 1/\mu$, $I(W; W+S)$ is maximized when the distribution of W is a mixture of an exponential with expected value $\mu^{-1} + \lambda^{-1}$ and an impulse at the origin. The resulting distribution for inter-departure time will be exponential with expected value $\mu^{-1} + \lambda^{-1}$ which is the distribution used in Theorem 2. In our system, since W cannot have zero value, the above conditions cannot be satisfied simultaneously and the Theorem 2 upper bound is not achievable.

A. Achievability for the M/M/1 Queue

To derive achievability results, we use the information density method introduced in [9]. For the bufferless timing queue, the information density is given by

$$i_{A_1^\infty; D_0^n}(A_1^\infty; D_0^n) = \log \frac{f_{D_0^n|A_1^\infty}(D_0^n|A_1^\infty)}{f_{D_0^n}(D_0^n)}. \tag{54}$$

We will employ the following definition and theorem.

Definition 2. The *liminf in probability* of a sequence of random variables Q_n is

$$\begin{aligned}&\liminf_{n\to\infty} \text{-p } Q_n \\ &= \sup\left\{\alpha > 0 | \lim_{n\to\infty} P\left[Q_n \le \alpha - \gamma\right] = 0, \forall \gamma > 0\right\}.\end{aligned}$$

Lemma 5 ([9])**.** *A sufficient condition for rate R to be achievable is existence of some input process A_1^∞ for which*

$$\liminf_{n\to\infty} \text{-p } \left[\frac{1}{T_n} i_{A_1^\infty;D_0^n}(A_1^\infty; D_0^n)\right] \ge R.$$

We will use Lemma 5 to prove the following achievability result expressed in terms of $R(\lambda,\mu)$ given in (53).

Theorem 4. *The M/M/1 bufferless queue with service rate μ and arrival rate λ has capacity*

$$C(\lambda,\mu) \ge R(\lambda,\mu).$$

Proof: In the M/M/1 queue, the arrival process is Poisson with rate λ. As noted at the start of Section IV, the queue has a renewal each time a packet enters service. These inter-renewal times are of the form $S_i + W_i$ where S_i and W_i may be dependent, but S_i, W_i are independent of S_j, W_j for $j \ne i$. For Poisson arrivals, the memorylessness of the exponential distribution implies S_i and W_i are independent. As a result, the inter-departure times D_i are iid hypoexponential random variables with PDF given by (50). Hence we can write

$$f_{D_0^n}(d_0^n) = f_{S_0}(d_0)\prod_{i=1}^n f_{D_i}(d_i). \tag{55}$$

It follows from Lemma 1 that the expected value of the information density will be

$$\begin{aligned}\mathrm{E}\left[i_{A_1^\infty;D_0^n}(A_1^\infty; D_0^n)\right] &= I(A_1^\infty; D_0^n) && (56)\\ &= \sum_{k=0}^n \left[h(W_{k-1}+S_k) - h(S_k)\right] && (57)\\ &= n\left(h(W+S) - h(S)\right) && (58)\\ &= n\left(h_{\text{hypo}}(\lambda,\mu) - \log e + \log\mu\right). && (59)\end{aligned}$$

Furthermore,

$$\begin{aligned}&f_{D_0^n|A_1^\infty}(d_0^n|a_1^\infty)\\ &= f_{S_0}(d_0)\prod_{i=1}^n f_{D_i|A_1^\infty,D_0^{i-1}}\left(d_i|a_1^\infty, d_0^{i-1}\right) && (60)\\ &= f_{S_0}(d_0)\prod_{i=1}^n f_{D_i|A_1^\infty,D_0^{i-1},W_{i-1}}\left(d_i|a_1^\infty, d_0^{i-1}, w_{i-1}\right) && (61)\\ &= f_{S_0}(d_0)\prod_{i=1}^n f_{S_i|A_1^\infty,D_0^{i-1},W_{i-1}}\left(d_i - w_{i-1}|a_1^\infty, d_0^{i-1}, w_{i-1}\right) && (62)\\ &= f_{S_0}(d_0)\prod_{i=1}^n f_{S_i}\left(d_i - w_{i-1}(a_1^\infty, d_0^{i-1})\right), && (63)\end{aligned}$$

where (61) holds due to (4), and (62) follows due to (5). Since the server processes the packets independent of the arrival process, S_i is independent of $A_1^\infty, D_0^{i-1}, W_{i-1}$, and thus (63) holds. Using (55) and (63), (54) normalized by T_n can be written as

$$\begin{aligned}&\frac{1}{T_n} i_{A_1^\infty;D_0^n}(A_1^\infty; D_0^n)\\ &= \frac{1}{T_n}\left[\sum_{i=1}^n \log\left(f_{S_i}(D_i - W_{i-1})\right) - \sum_{i=1}^n \log\left(f_{D_i}(D_i)\right)\right]\\ &= \frac{n}{T_n}\frac{1}{n}\left[\sum_{i=1}^n \log\left(f_{S_i}(S_i)\right) - \sum_{i=1}^n \log\left(f_{D_i}(D_i)\right)\right], \qquad (64)\end{aligned}$$

since $S_i = D_i - W_{i-1}$. Since the W_i are iid exponential (λ) random variables, (48) implies

$$\lim_{n\to\infty}\frac{n}{T_n} = \frac{1}{\mathrm{E}[W] + \mathrm{E}[S]} = \frac{1}{1/\lambda + 1/\mu}. \tag{65}$$

By the strong law of large numbers [20], it follows from (64) and (65) that

$$\lim_{n\to\infty}\frac{i_{A_1^\infty;D_0^n}(A_1^\infty; D_0^n)}{T_n} = \frac{h(D) - h(S)}{1/\lambda + 1/\mu} = R(\lambda,\mu) \quad \text{wp 1.}$$

It follows that the liminf in probability of $i_{A_1^\infty;D_0^n}(A_1^\infty; D_0^n)/T_n$ equals $R(\lambda,\mu)$ and thus by Lemma 5, rate $R(\lambda,\mu)$ is achievable. ■

Comparing Theorem 4 and the upper bound (52), we see that the achievable rate $R(\lambda,\mu)$ matches the upper bound for the M/M/1 bufferless queue. Thus $R(\lambda,\mu)$ is the capacity of the bufferless M/M/1 timing channel with arrival rate λ and service rate μ. This M/M/1 capacity is illustrated in Fig. 2. The maximum achievable rate in (52) is 0.3340 nats per average server time, and the maximum of universal upper bound is 0.3679 which implies that a bufferless M/M/1 queue at worst suffers less than 10% reduction in achievable rate when compared to the universal upper bound.

Fig. 3 illustrates the achievable upper bound of M/M/1 (4) and the universal upper bound of ·/M/1 (52). The ·/M/1 upper bound of "Bits through queues" (BTQ) paper [7, eq. 2.17-2.18] is plotted for comparison. In these plots, $0 < \lambda/\mu < 1$ since the ·/M/1 BTQ requires $\lambda \le \mu$ for stability. From this plot, we can see that the maximum value of the upper bound of ·/M/1 is equal to the maximum value of ·/M/1 BTQ which is 0.3679 nats per average service time.

B. Achievability for the M/G/1 queue

Theorem 5. *The M/G/1 bufferless queue S with arrival rate λ and average service time $\mathrm{E}[S] = 1/\mu$ has capacity*

$$C(\lambda, S) \ge R(\lambda,\mu).$$

Proof: The procedure for this proof is along the lines of the proof of [7, Thm. 7]. We assume the inter-departure times under general service have PDF $g_{D_0^n}(d_0^n)$, and the arrivals are a rate λ Poisson process. We further assume that $f_{D_0^n}$ is the PDF of the inter-departure times of system with a memoryless server of rate μ (which would be hypoexponential). Now

similar to [7],

$$i_{A_1^\infty;D_0^n}(A_1^\infty;D_0^n) = \log\frac{g_{D_0^n|A_1^\infty}}{g_{D_0^n}} \tag{66}$$

$$= \log\frac{g_{D_0^n|A_1^\infty}}{f_{D_0^n|A_1^\infty}} - \log\frac{g_{D_0^n}}{f_{D_0^n}} + \log\frac{f_{D_0^n|A_1^\infty}}{f_{D_0^n}}. \tag{67}$$

In Theorem 4, we showed that

$$\liminf_{n\to\infty} \text{-p}\ \frac{1}{T_n}\log\frac{f_{D_0^n|A_1^\infty}}{f_{D_0^n}} = R(\lambda,\mu). \tag{68}$$

We need to prove that

$$\liminf_{n\to\infty} \text{-p}\ \frac{1}{T_n}\left[\log\frac{g_{D_0^n|A_1^\infty}}{f_{D_0^n|A_1^\infty}} - \log\frac{g_{D_0^n}}{f_{D_0^n}}\right] \geq 0.$$

Note that (48) implies it is sufficient to prove that for every $\zeta > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left[\frac{1}{n}\left(\log\frac{g_{D_0^n|A_1^\infty}}{f_{D_0^n|A_1^\infty}} - \log\frac{g_{D_0^n}}{f_{D_0^n}}\right) \leq -\zeta\right] = 0.$$

Using the same method as [7],

$$\mathrm{P}\left[\frac{1}{n}\left(\log\frac{g_{D_0^n|A_1^\infty}}{f_{D_0^n|A_1^\infty}} - \log\frac{g_{D_0^n}}{f_{D_0^n}}\right) \leq -\zeta\right]$$

$$= \mathrm{P}_{g_{A_1^\infty,D_0^n}}\left[\frac{1}{n}\log\frac{g_{D_0^n|A_1^\infty}}{f_{D_0^n|A_1^\infty}}\frac{f_{A_1^\infty}f_{D_0^n}}{f_{A_1^\infty}g_{D_0^n}} \leq -\zeta\right] \tag{69}$$

$$= \mathrm{P}_{g_{A_1^\infty,D_0^n}}\left[\frac{1}{n}\log\frac{g_{A_1^\infty|D_0^n}}{f_{A_1^\infty|D_0^n}} \leq -\zeta\right] \tag{70}$$

$$= \iint_{g_{A_1^\infty|D_0^n} \leq e^{-\zeta n} f_{A_1^\infty|D_0^n}} g_{A_1^\infty|D_0^n}(x_1^\infty|y_0^n) f_{D_0^n}(y_0^n) dx_1^\infty\, dy_0^n \tag{71}$$

$$\leq e^{-\zeta n}\iint f_{A_1^\infty|D_0^n}(x_1^\infty|y_0^n) f_{D_0^n}(y_0^n) dx_1^\infty\, dy_0^n \tag{72}$$

$$= e^{-\zeta n}. \tag{73}$$

It follows that

$$\liminf_{n\to\infty} -\mathrm{p}\frac{1}{T_n}\left(\log\frac{g_{D_0^n|A_1^\infty}}{f_{D_0^n|A_1^\infty}} - \log\frac{g_{D_0^n}}{f_{D_0^n}}\right) \geq 0. \tag{74}$$

Now using (68) and (74), we see that for every $\zeta' > 0$,

$$\lim_{n\to\infty} P\left[\frac{1}{T_n} i_{A_1^\infty;D_0^n}(A_1^\infty;D_0^n) \leq R(\lambda,\mu) - \zeta'\right] = 0.$$

Thus Theorem 5 holds. ■

It must be noted that this is not necessarily a tight lower bound similar to [7]. The result of Theorem 5 shows that the exponential server has the lowest capacity for a fixed service rate among servers with Poisson arrivals.

VI. CONCLUSION

This paper studied the capacity of 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 the maximum likelihood decoder for decoding timing messages transmitted through a bufferless queue, 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 and M/G/1 queues. 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.

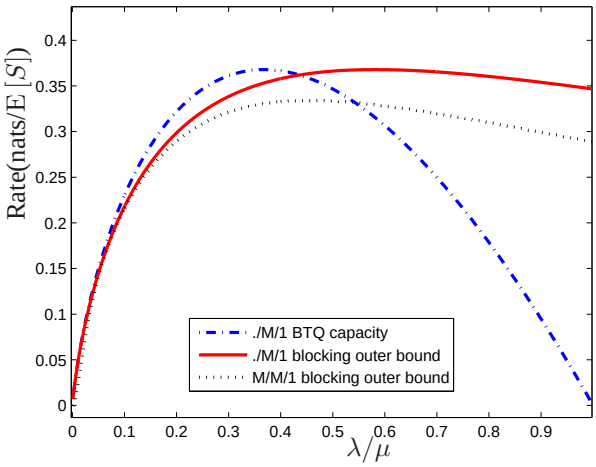

Fig. 3. Comparison of $C(\lambda)$ of the "bits through queues" (BTQ) paper [7, Theorem 4], the upper bound for bufferless ·/M/1 queue, and the capacity $R(\lambda,\mu)$ for the M/M/1 bufferless queue. All the systems have exponential service time of rate μ and arrival rate λ. Both BTQ and bufferless upper bound plots have maximum equal to 0.3679 nats per average service time whereas the maximum achievable rate is 0.3340 nats per average service time.

APPENDIX

In this part, the upper bound (52) for the M/M/1 queue is shown to be only a function of $\rho = \lambda/\mu$ for fixed μ. Initially,

the $h_{\text{hypo}}(\lambda,\mu)$ is rewritten using (51) as follows:

$$h_{\text{hypo}}(\lambda,\mu) \tag{75}$$

$$= -\int f_D(x)\log f_D(x)dx \tag{76}$$

$$= -\int f_D(x)\log\left[\mu e^{-\mu x}\frac{\rho}{1-\rho}\left(e^{-(\lambda-\mu)x}-1\right)\right]dx \tag{77}$$

$$= -\log\mu + \mu E\left[D\right]\log e - \log\left(\frac{\rho}{1-\rho}\right) - \int f_D(x)\log\left(e^{-(\lambda-\mu)x}-1\right)dx. \tag{78}$$

With the change of variable $y=(\lambda-\mu)\,x$,

$$h_{\text{hypo}}(\lambda,\mu) = -\log\mu + \left(1+\frac{1}{\rho}\right)\log e - \log\left(\frac{\rho}{1-\rho}\right) + \frac{\rho}{(1-\rho)^2}G\left(\rho\right), \tag{79}$$

where

$$G\left(\rho\right) = \int e^{-\frac{y}{\rho-1}}\left(e^{-y}-1\right)\log\left(e^{-y}-1\right)dy$$

is a function of ρ. Now substituting (79) in (53),

$$R(\mu,\lambda) = \mu\frac{\log e - \rho\log\left(\frac{\rho}{1-\rho}\right) + \left(\frac{\rho}{1-\rho}\right)^2 G\left(\rho\right)}{1+\rho}$$

which proves the claim.

REFERENCES

[1] B. Radosavljevic and B. Hajek, "Hiding traffic flow in communication networks," in *Proc. Military Communications Conference*, 1992, pp. 1096–1100.

[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] 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.

[4] 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," *plasmid*, vol. 30, no. 217.5, pp. 6–2.

[5] 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.

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

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

[8] 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.

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

[10] 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.

[11] J. A. Thomas, "On the shannon capacity of discrete time queues," in *IEEE International Symposium on Information Theory*. IEEE, 1997, p. 333.

[12] R. Sundaresan and S. Verdu, "Capacity of queues via point-process channels," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2697–2709, 2006.

[13] T. P. Coleman, "A simple memoryless proof of the capacity of the exponential server timing channel," in *Information Theory Workshop*. IEEE, 2009, pp. 101–105.

[14] S. H. Sellke, C.-C. Wang, N. Shroff, and S. Bagchi, "Capacity bounds on timing channels with bounded service times," in *IEEE International Symposium on Information Theory*. IEEE, 2007, pp. 981–985.

[15] P. Mimcilovic, "Mismatch decoding of a compound timing channel," in *Forty-Fourth Annual Allerton Conference on Communication, Control, and Computing*, 2006.

[16] E. Visotsky, V. Tripathi, and M. Honig, "Optimum ARQ design: A dynamic programming approach," in *Proceedings IEEE International Symposium on Information Theory, 2003*. IEEE, 2003, p. 451.

[17] Y. Polyanskiy, H. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, 2011.

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

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

[20] S. M. Ross, "Stochastic processes john wiley and sons," *New York*, 1983.