

Capacity of Queues Via Point-Process Channels

Rajesh Sundaresan, *Senior Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

Abstract—A conceptually simple proof for the capacity formula of an exponential server timing channel is provided. The proof links the timing channel to the point-process channel with instantaneous noiseless feedback. This point-process approach enables a study of timing channels that arise in multiserver queues, queues in tandem, and other simple configurations. Although the capacities of such channels remain to be found, the paper provides some analytical bounds and highlights a method to find achievable rates via simulations.

Index Terms—Channels with feedback, direct-detection photon channel, intensity, nonlinear filtering, point process, Poisson channel, queue, rate, timing channel.

I. INTRODUCTION

THE capacity of the exponential server timing channel (ESTC) with service rate μ packets per second is $e^{-1}\mu$ nats per second [1]. The capacity of the point-process channel with maximum input intensity μ points per second, and no background intensity, is also $e^{-1}\mu$ nats per second (cf. [2], [3]). Furthermore, in both channels, the capacity does not increase in the presence of instantaneous noiseless feedback. In [1], the connection between the two channels in the presence of instantaneous noiseless feedback was discussed briefly. In [4], this connection was further explored—any strategy on the ESTC can be mapped to an equivalent strategy that uses feedback on the point-process channel. This observation implies that the capacity of the ESTC is upper-bounded by the capacity of the point-process channel with instantaneous noiseless feedback, i.e., $e^{-1}\mu$ nats per second.

From [1] and [4], we know that $e^{-1}\mu$ nats per second is indeed achievable on the ESTC. The route taken in [4] was to show the achievability result in full generality for a stationary and ergodic sequence of service times. Considerable attention has been focused on the ESTC (cf. [1], [5], [6]). However, not much is known about the capacity of other queueing systems such as multiserver queues, queues in tandem, or even the single-server queue with a finite buffer. For such systems, the approaches of [1], [5], [6], and [4] via Lindley's equation [7, p. 431] do not seem to extend easily to more general queueing systems.

In this paper, we take a point-process approach to study timing channels. We restrict our attention to servers whose ser-

vice times have the exponential distribution. In Section II, we first show the converse and the direct parts for the ESTC, using point-process techniques. We also find the capacity region when two or more users access this channel. This approach is then used in Sections III and IV to study timing channels in a multiserver queueing system and in some other simple networks of servers with exponential service times. In particular, we obtain bounds, either analytically or from simulations, on the capacities of multiserver queues in Section III, the single-server queue with spurious departures, and a pair of queues connected in tandem in Section IV. In all these examples, we identify the channel model, find upper bounds on the capacity, and then find achievable rates through simulations. Section V is a summary and discussion of the results.

Since the inclusion of the main results of this paper in [8], several papers have focused on the ESTC. We now summarize their results. Reference [9] identifies the error exponent of this channel for rates between $(\mu/4) \log 2$ and $e^{-1}\mu$ nats per second, and gives upper and lower bounds for rates below $(\mu/4) \log 2$ nats per second. These upper and lower bounds on the reliability function are based on the sphere-packing and random-coding bounds, respectively. The sphere-packing bound is identical to the sphere-packing bound for point-process channels without dark current. Similarly, the two channels' random-coding bounds also coincide. However, [10] shows that the reliability functions of the two channels are different; the zero-rate error exponent of the ESTC is $\mu/2$ while that of the point-process channel without feedback is $\mu/4$. Incidentally, the zero-rate error exponent of the point-process channel with feedback is μ . Thus, the restrictive use of feedback on the point-process channel, as suggested by the emulation of the ESTC on the point-process channel, does not achieve the best reliability with feedback in the low-rate regime. The point-process technique used in [10] is similar to that used in [8], and in this paper. While the focus of [10] is on the ESTC, the goal of the present paper is not only to highlight the connections between point-process channels and the ESTC, but also to apply the technique to study other queueing systems.

II. SINGLE QUEUEING STATION

In this section, we rederive the capacity formula for the ESTC. In showing the achievability result, we directly see the role played by Burke's output theorem, a well-known result for the $M/M/1$ queue ([11, Ch. V, Theorem T1]). We also find the capacity region when we allow more than one transmitter. We first give some mathematical preliminaries before we describe the channel in Section II-B1. In the interest of brevity, several facts and technical details are stated without proof. Where the arguments may not be obvious, we supply pointers to sections

Manuscript received February 15, 2005; revised September 28, 2005. This work was supported in part by the National Science Foundation under Grant NCR 9523805, in part by Qualcomm Inc., and in part by the Ministry of Human Resources and Development (MHRD, India) under Grant Part(2A) Tenth Plan (338/ECE). The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000.

R. Sundaresan is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

Communicated by R. Srikant, Guest Editor.

Digital Object Identifier 10.1109/TIT.2005.862079

and theorems in standard texts to help the reader complete the technical arguments.

A. Preliminaries

Let \mathbb{Z}_+ denote the set of nonnegative integers, $\overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$, and \mathbb{N} the set of natural numbers. Let (Ω, \mathcal{F}) be a measurable space. Fix finite $T \in (0, \infty)$. Let $(\mathcal{F}_t : t \in [0, T])$ be an increasing family of sub- σ -algebras that is right-continuous. An uppercase letter $X = (X_t : t \in [0, T])$ will denote a stochastic process on $[0, T]$. We say X is *adapted* to the family $(\mathcal{F}_t : t \in [0, T])$ if X_t is \mathcal{F}_t -measurable for each $t \in [0, T]$. Let

$$\mathcal{F}_t^X = \sigma\{X_s : s \in [0, t]\}, \quad \text{for } t \in [0, T]$$

and

$$\mathcal{F}_{t-}^X = \sigma\{X_s : s \in [0, t)\}, \quad \text{for } t \in (0, T].$$

We say that a stochastic process X is *predictable* with respect to $(\mathcal{F}_t : t \in [0, T])$ if X is measurable with respect to $([0, T] \times \Omega, \mathcal{L})$, where \mathcal{L} is the σ -algebra generated by all left-continuous processes adapted to $(\mathcal{F}_t : t \in [0, T])$. (In particular, any adapted left-continuous process is predictable.)

The input space is $(\mathcal{X}, \mathcal{F}_T^X)$, where the alphabet \mathcal{X} is the set of functions $x : [0, T] \rightarrow \mathbb{Z}_+$ that are nondecreasing and right-continuous. \mathcal{X} represents the set of arrival processes on $[0, T]$ with possible multiple arrivals at the same instant. Let $x \in \mathcal{X}$. This function's value at t , denoted by x_t , represents the number of arrivals in $[0, t]$. Similarly, the output space is $(\mathcal{Y}, \mathcal{F}_T^Y)$, where \mathcal{Y} is the set of functions $y : [0, T] \rightarrow \overline{\mathbb{Z}}_+$ that are nondecreasing, right-continuous, have unit jumps, and satisfy $y_0 = 0$. \mathcal{Y} represents the set of counting processes (or point processes) on $[0, T]$.

Let $(\Omega, \mathcal{F}_T, P)$ be a probability space. (In all settings considered in this paper, the space (Ω, \mathcal{F}_T) will be such that $\mathcal{F}_T^Y \subset \mathcal{F}_T$; this will be clear from the context). We say that the point process Y adapted to $(\mathcal{F}_t : t \in [0, T])$ has the *rate* (or *intensity*) process λ with respect to the family $(\mathcal{F}_t : t \in [0, T])$, if the following three conditions hold:

- (i) the nonnegative process $\lambda = (\lambda_t : t \in [0, T])$ is predictable with respect to the family $(\mathcal{F}_t : t \in [0, T])$;
- (ii) $\int_0^t \lambda_s ds < \infty$, P -almost surely (a.s), for each $t \in [0, T]$; and
- (iii) for every nonnegative process C that is predictable with respect to $(\mathcal{F}_t : t \in [0, T])$, we can write

$$E \left[\int_0^T C_s dY_s \right] = E \left[\int_0^T C_s \lambda_s ds \right]$$

where $\int_0^T C_s dY_s$ denotes Lebesgue–Stieltjes integration for a fixed $\omega \in \Omega$.

Let P_0 be the measure on $(\mathcal{Y}, \mathcal{F}_T^Y)$ such that Y is a point process having constant and unit-rate intensity with respect to $(\mathcal{F}_t^Y : t \in [0, T])$. This is the projection of the standard Poisson point process on \mathcal{F}_T^Y , and will be our reference measure on the space $(\mathcal{Y}, \mathcal{F}_T^Y)$. The dependence of P_0 on T is understood.

Even though we have allowed ∞ in the range of the observable y , our attention in this paper will be restricted to (P_0 -a.s.) nonexplosive point processes.

B. The $\cdot/M/1$ Queue

1) *Channel Model*: $P(x, dy)$ is a transition probability function [12, p. 315] from the space $(\mathcal{X}, \mathcal{F}_T^X)$ to the space $(\mathcal{Y}, \mathcal{F}_T^Y)$, if it satisfies the following measurability properties:

- for each $x \in \mathcal{X}$, the mapping $B \mapsto P(x, B)$ from \mathcal{F}_T^Y to $[0, 1]$ is a probability measure on $(\mathcal{Y}, \mathcal{F}_T^Y)$, and
- for each $B \in \mathcal{F}_T^Y$, the mapping $x \mapsto P(x, B)$ is \mathcal{F}_T^X -measurable.

Yet again, the dependence of $P(x, dy)$ on T is understood.

For each T , we now define a transition probability function that models the rate- μ ESTC with information encoded in the arrival times of packets. (See also the definition in [10, Sec. III]). We motivate our definition as follows. Consider an $M/M/1$ queue in equilibrium at $t = 0$. Define the state process $Q = (Q_t = X_t - Y_t : t \in [0, T])$ (right-continuous with left limits), where Q_t denotes the number of packets that remain in the system at time t . It is shown in [11, Example 2.6] and [11, Example 1.3] that for the $M/M/1$ queue, the departure process Y admits the rate process $\lambda = (\mu 1\{Q_{t-} > 0\} : t \in [0, T])$ with respect to $(\mathcal{F}_t^Q : t \in [0, T])$.

Fix T and $x \in \mathcal{X}$. Let

$$Q_t = x_t - Y_t, \quad t \in [0, T]. \quad (1)$$

Fix arbitrary $t \in [0, T]$. If $Q_{t-} = 0$, there is no packet in the system at time $t-$, and therefore no packet can depart at time t ; the rate of the point process of departures is 0 at time t . If $Q_{t-} > 0$, there is at least one packet in the system at $t-$. Due to the memoryless property of exponential service times, the residual time for the next departure is exponentially distributed with mean $1/\mu$ seconds, independent of the past. In other words, the rate of the point process of departures is μ at time t .

For a fixed $x \in \mathcal{X}$, therefore, the probability measure $P(x, \cdot)$ on $(\mathcal{Y}, \mathcal{F}_T^Y)$ is such that Y , the point process of departures in $[0, T]$, admits the rate process

$$\lambda = (\mu 1\{Q_{t-} > 0\} : t \in [0, T]) \quad (2)$$

with respect to $(\mathcal{F}_t^Y : t \in [0, T])$, where Q_t is as defined in (1).

We therefore model this channel by setting the Radon–Nikodym derivative of $P(x, \cdot)$ with respect to P_0 as

$$\frac{dP(x, \cdot)}{dP_0}(y) = p(x, y) \quad (3)$$

where

$$p(x, y) \triangleq \exp \left\{ \int_0^T [\log(\lambda_t) dy_t + (1 - \lambda_t) dt] \right\}. \quad (4)$$

The results [11, Ch. VI, Theorems T2–T4] ensure that the point process of departures Y admits the rate process (2) with respect to $(\mathcal{F}_t^Y : t \in [0, T])$ under the probability measure $P(x, \cdot)$. The function $p(x, y)$ is measurable with respect to $\mathcal{F}_T^X \vee \mathcal{F}_T^Y$, which implies that $P(x, dy)$ is a transition probability function. Thus, (3) and (4) suitably model the ESTC. Note that (3) and (4) are related to the sample function density for the self-exciting point process with rate λ [13, Theorem 5.2.2] with the difference being that (3) is written as a Radon–Nikodym derivative with respect to P_0 .

We adopt the following definitions for achievability and capacity. Each of M equiprobable messages is mapped to an el-

ement in \mathcal{X} . The decoder observes the departures in the time interval $[0, T]$ and declares one of the M messages as transmitted. An error occurs if the decoded message is different from the transmitted message. For a fixed T , a codebook with M codewords, and a decoder, let ε be the probability of error. We call this a (T, M, ε) -code. Rate R is *achievable* if, for every $\gamma > 0$, there is a sequence of $(T_n, M_n, \varepsilon_n)$ -codes that satisfies $\lim_{n \rightarrow \infty} T_n = \infty$, $(\log M_n)/T_n \geq R - \gamma$ for all sufficiently large n , and $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. The *capacity* is the supremum of all achievable rates.

The codes considered in [4] require all codewords to have the same number of packets. The codes considered in [1] require that packets exit before T on the average. The above definition does not impose such restrictions. These relaxations however do not increase the capacity of the ESTC.

$P(x, dy)$, in addition to modeling the ESTC, is also the resulting transition probability function under the coding technique (2) on the point-process channel. This coding technique utilizes feedback on the channel where background noise is absent and a peak constraint μ is placed on the rate. From [2] we know that $\lambda_t \in \{0, \mu\}$ is optimal on the point-process channel. The coding technique described in the above paragraphs also results in intensity values on the set $\{0, \mu\}$.

2) *Mutual Information*: Let ν be a probability measure on $(\mathcal{X}, \mathcal{F}_T^X)$. Then ν and $P(x, dy)$ define a joint probability measure π on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_T^X \vee \mathcal{F}_T^Y)$, denoted by

$$\pi(dx, dy) = \nu(dx)P_0(dy)p(x, y). \quad (5)$$

We can verify from Fubini's theorem [14, Theorem 18.3, p. 238] and some measurability arguments that, under π , the stochastic process Y has rate λ with respect to the information pattern $(\mathcal{F}_T^X \vee \mathcal{F}_t^Y : t \in [0, T])$.

Let π^X and π^Y be the restrictions of π to \mathcal{F}_T^X and \mathcal{F}_T^Y , respectively. Let $\pi^X \times \pi^Y$ be the product probability measure on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_T^X \vee \mathcal{F}_T^Y)$. Let the symbol “ \ll ” denote that the measure on the left side of the symbol is absolutely continuous with respect to the measure on the right side. From (5), we get that $\pi \ll \pi^X \times \pi^Y \ll \nu \times P_0$ [15, Corollary 5.3.1, p. 112], and that $\pi^Y \ll P_0$. Furthermore, [11, Ch. VI, Result R8] gives

$$\frac{d\pi^Y}{dP_0}(y) = \exp \left\{ \int_0^T \left[(\log \hat{\lambda}_t) dy_t + (1 - \hat{\lambda}_t) dt \right] \right\} \quad (6)$$

where Y has rate $\hat{\lambda}$ with respect to $(\mathcal{F}_t^Y : t \in [0, T])$ under the probability measure π^Y . More specifically, we can take $\hat{\lambda}_t = E[\lambda_t | \mathcal{F}_{t-}^Y]$, for each $t \in [0, T]$ [16, Theorem 18.3].

Equation (6) was proved for the special case of an intensity driven by a Markov process in [17], and is analogous to the well-known “estimator–correlator” formula for the likelihood ratio of a random signal in Gaussian noise [18]. Equation (6) tells us how to identify the likelihood ratio in two separate steps when the intensity is random. In the first step, for each $t \in [0, T]$, we perform an “estimation” to obtain $\hat{\lambda}_t$, the conditional expectation of the intensity λ_t at time t , given the past departures \mathcal{F}_{t-}^Y . In the second step, we observe that the point process admits this filtered signal as an \mathcal{F}_t^Y -intensity. Substitution of $\hat{\lambda}_t$ in place of λ_t in (4) yields the likelihood ratio in (6).

The normalized information density $(1/T) i_T(x, y)$ is therefore

$$\begin{aligned} & \frac{1}{T} \log \left(\frac{d\pi}{d(\pi^X \times \pi^Y)}(x, y) \right) \\ &= \frac{1}{T} \int_0^T [(\log \lambda_t) dy_t + (1 - \lambda_t) dt \\ & \quad - (\log \hat{\lambda}_t) dy_t - (1 - \hat{\lambda}_t) dt]. \end{aligned} \quad (7)$$

Finally, as a consequence of property (iii) in the definition of a rate process, we can write the normalized mutual information as

$$\frac{1}{T} I_T(X; Y) = \frac{1}{T} E \int_0^T dt [\phi(\lambda_t) - \phi(\hat{\lambda}_t)]$$

where $\phi(u) = u \log u$ (see [2], [16], [3], [11]). We define $\phi(0) \triangleq 0$. The function ϕ is strictly convex on $[0, \infty)$.

Related formulas for the derivative of mutual information with respect to the channel parameters have been derived recently in [19].

3) *Optimal Decoding*: Suppose that there are M equiprobable codewords. Each codeword is mapped to a sequence of arrivals $x \in \mathcal{X}$, in the time interval $[0, T]$. Suppose $y \in \mathcal{Y}$ is received. It is well known that the optimal decoder that minimizes the probability of error works as follows. For each codeword, assign a score of $\int_0^T [\log(\lambda_t) dy_t + (1 - \lambda_t) dt]$, where λ_t is obtained from (2) and (1). Then choose the codeword that maximizes this score, and in case of a tie choose the one with the least index.

Note that for a codeword under consideration, if $q_{t-} = 0$ and a departure is observed at time t , i.e., $dy_t = 1$, then the score is $-\infty$. This codeword therefore does not explain the received sequence of departures. If the codeword is indeed compatible, then λ_t at instants of departures is μ . Thus, the decision is based on maximizing $\int_0^T (1 - \lambda_t) dt$, or equivalently, is based on maximizing the net idling time of the server.

4) *Converse*: The following converse was proved in [3], [2]. Observe that it works for any predictable process λ on the point-process channel that satisfies $\lambda_t \in [0, \mu]$. The coding that arises in the queueing system (cf. (2) and (1)) is only a specific case.

Proposition 1: [3], [2] The capacity of the point-process channel with maximum intensity μ cannot exceed $e^{-1}\mu$ nats per second, even in the presence of instantaneous noiseless feedback.

Proof: This proof is taken from [3], [2]. We present it here because, when looking at achievability, it is instructive to see the conditions when the inequalities in the converse become equalities. Let λ be an arbitrary predictable process with $\lambda_t \in [0, \mu]$. We interpret λ as an encoding strategy of the input message in the presence of (instantaneous noiseless) feedback. Then the following sequence of inequalities holds:

$$\begin{aligned} & \frac{1}{T} I_T(X; Y) \\ &= \frac{1}{T} E \int_0^T dt [\phi(\lambda_t) - \phi(\hat{\lambda}_t)] \\ & \stackrel{(a)}{\leq} \frac{1}{T} E \int_0^T dt \phi(\lambda_t) - \phi \left(\frac{1}{T} E \int_0^T dt \hat{\lambda}_t \right) \end{aligned}$$

$$\begin{aligned}
&\stackrel{(b)}{=} \frac{1}{T} E \int_0^T dt \phi(\lambda_t) - \phi \left(\frac{1}{T} \int_0^T dt E \hat{\lambda}_t \right) \\
&\stackrel{(c)}{=} \frac{1}{T} E \int_0^T dt \phi(\lambda_t) - \phi \left(\frac{1}{T} \int_0^T dt E \lambda_t \right) \\
&\stackrel{(d)}{\leq} \max_{0 \leq a \leq \mu} \max_{F: \int_0^\mu zF(dz) = a} \int_0^\mu \phi(z) F(dz) - \phi(a) \\
&\stackrel{(e)}{=} \max_{0 \leq a \leq \mu} \max_{F: \int_0^\mu zF(dz) = a} \left(h(a) - \int_0^\mu h(z) F(dz) \right). \quad (8)
\end{aligned}$$

Inequality (a) follows from Jensen's inequality applied to the strictly convex function ϕ . Equality (b) follows from Fubini's theorem which allows us to move expectation inside the integral. Equality (c) follows from $E \hat{\lambda}_t = E \lambda_t$. Inequality (d) comes from fixing $(1/T) \int_0^T dt E \lambda_t = a$, maximizing over all possible distributions F on $[0, \mu]$ with mean a , followed by a maximization over a . Equality (e) represents an equivalent maximization with $h(z) = (z/\mu)\phi(\mu) - \phi(z) = z \log(\mu/z)$.

Observe that, because $h(0) = h(\mu) = 0$, and $h(z) > 0$ for every $z \in (0, \mu)$, the maximizing F puts mass only on the set $\{0, \mu\}$. Furthermore, because the mean is a , F puts masses $(1 - a/\mu)$ and a/μ on 0 and μ , respectively. Moreover, $h(a) = a \log(\mu/a)$ has maximum value $e^{-1}\mu$ at $a = e^{-1}\mu$. We therefore have that $(1/T)I_T(X; Y) \leq e^{-1}\mu$. \square

Remarks: This proves the converse for the timing channel without feedback. The converse is valid for any predictable λ that lies within $[0, \mu]$. Feedback therefore cannot increase the capacity of the timing channel.

The proof also indicates that if we restrict attention to those processes with $(1/T) \int_0^T E \lambda_t dt = a$, the normalized mutual information satisfies $(1/T)I_T(X; Y) \leq a \log(\mu/a)$.

5) *Direct Part:* The \liminf in probability of a sequence of random variables $(Z_n : n \in \mathbb{N})$ is the supremum of all reals α such that $\lim_{n \rightarrow \infty} P\{Z_n \leq \alpha\} = 0$. If the limit does not exist for any real α , we take the \liminf in probability to be $-\infty$. It was shown in [20, Theorem 2] that the \liminf in probability of the normalized information density is an achievable rate.

Proposition 2: Fix finite $T > 0$. For the ESTC given by $P(x, dy)$, there is an input probability measure ν such that $(1/T)I_T(X; Y) = e^{-1}\mu$. Furthermore, for any sequence $(T_n : n \in \mathbb{N})$ with $\lim_{n \rightarrow \infty} T_n = \infty$, the \liminf in probability of the sequence $((1/T_n) i_{T_n}(X; Y) : n \in \mathbb{N})$ is $e^{-1}\mu$.

Proof: From [1] and [4], we already know that the maximum mutual information $e^{-1}\mu$ nats per second is attained using Poisson input following equilibrium at time $t = 0$. The purpose of this proof is to show this directly from the converse (8).

Let ν be such that X_0 (i.e., Q_0), the initial number in the queue, has the equilibrium state distribution associated with an $M/M/1$ queue with Poisson arrivals of rate $a = e^{-1}\mu$, i.e.,

$$\nu\{X_0 = k\} = (1 - a/\mu)(a/\mu)^k, \quad \text{for } k \in \mathbb{Z}_+.$$

Furthermore, let the arrivals $X_t - X_0$ on $(0, T]$ form a Poisson process of rate $a = e^{-1}\mu$. Let λ_t and Q_t be defined as in (2) and (1). Let π be as defined in (5). Under π , Q is the state process of an $M/M/1$ queue starting from equilibrium at $t = 0$. We

can therefore apply a result due to Burke (see, e.g., [11, Ch. V, Theorem T1]), which states that

$$\pi\{Q_t = k | \mathcal{F}_t^Y\} = \nu\{Q_0 = k\} \quad (9)$$

for every $t \in [0, T]$, π -a.s. This means that Q_t is independent of \mathcal{F}_t^Y for every $t \in [0, T]$.

We now verify that all inequalities in (8) are equalities. Observe that, $\lambda_t \in \{0, \mu\}$. The system remains in equilibrium throughout $[0, T]$; λ_t is therefore either 0 or μ with probabilities $1 - a/\mu$ and a/μ , respectively, for every $t \in [0, T]$. Hence,

$$\frac{1}{T} \int_0^T dt E \lambda_t = a = e^{-1}\mu.$$

This implies that inequality (d) in (8) is an equality.

For inequality (a) to be an equality, we need

$$\hat{\lambda}_t = E[\lambda_t | \mathcal{F}_t^Y] = a, \quad \text{for each } t \in [0, T].$$

Note that λ_t is a function of Q_{t-} . From Burke's theorem, Q_s is independent of \mathcal{F}_s^Y for each $s \in [0, T]$. Consequently, Q_{t-} is independent of \mathcal{F}_{t-}^Y [21, Theorem 1.6], and therefore $\hat{\lambda}_t = E \lambda_t = a$. Hence, $(1/T)I_T(X; Y) = e^{-1}\mu$, i.e., the input probability measure ν maximizes mutual information.

We now consider the normalized information density. Consider an increasing unbounded sequence $(T_n : n \in \mathbb{N})$ taking values in $(0, \infty)$. Fix T_n . Consider the same input measure ν as in the first part of this proof. Since $\hat{\lambda}_t = a$ for every $t \in [0, T_n]$, we get

$$\frac{1}{T_n} i_{T_n}(x; y) = \frac{1}{T_n} \int_0^{T_n} [\log(\lambda_t/a) dy_t + (a - \lambda_t) dt].$$

Observe that a departure can occur only when there is a packet in the system. At all times, we know exactly how many packets are in the system from (1). If a packet exits at time t , we must have that $\lambda_t = \mu \mathbf{1}\{Q_{t-} > 0\} = \mu$. The first integral is therefore $(Y_{T_n}/T_n) \log(\mu/a)$; it converges by the renewal theorem to $a \log(\mu/a)$ almost surely, and therefore in probability. The second integral converges to 0 in probability because the time average of the quantity $\mu \mathbf{1}\{Q_{t-} > 0\}$ converges in probability to a [22, Theorem 6.1]. By setting $a = e^{-1}\mu$, we get that the normalized information density converges in probability to $e^{-1}\mu$. \square

Remarks: From this result and the converse, the capacity of the ESTC is $e^{-1}\mu$ nats per second. Moreover, the coding scheme utilizing feedback described in (2) and (1) achieves capacity on the point-process channel.

As a final remark, observe that for all $a \in (0, \mu]$, we have $(1/T)I_T(X; Y) = a \log(\mu/a)$, the maximum achievable under the constraint $(1/T) \int_0^T E \lambda_t dt = a$. (See remark at the end of Section II-B4).

C. Multiuser Capacity Region

Suppose now that two users input packets to the ESTC. The capacity region is the triangle given by $R_1 \geq 0$, $R_2 \geq 0$, and $R_1 + R_2 \leq e^{-1}\mu$ nats per second [23]. Indeed, by time sharing and by the single-user result with output constraint [1], this region is clearly achievable. To show the converse, note that even if the two users cooperate, they have to transmit information

through a queueing system whose maximum service rate is μ . Consequently, joint coding cannot achieve a (sum) rate larger than $e^{-1}\mu$ nats per second. In contrast, in the Gaussian case, joint encoding leads to an increase in available power. An analogous argument holds when there are more than two users.

III. THE $\cdot/M/m$ QUEUE

In this section, we show how the point-process approach can be extended to study a multiserver queue. We show an upper bound based on the capacity of the ESTC, and an analytically obtained lower bound. A discussion on the results can be found at the end of this section.

The model for the $\cdot/M/m$ queue timing channel is given by (1) and (3) with the rate $\lambda = (\lambda_t : t \in [0, T])$ being

$$\lambda_t = \mu \min\{m, Q_{t-}\} \quad (10)$$

i.e., the rate is μ times the number of servers among the m that are busy. The optimal decoding strategy is as before with λ_t given by (10), i.e., choose the codeword that maximizes $\int_0^T [\log(\lambda_t) dy_t + (1 - \lambda_t) dt]$. In case of a tie, choose the codeword with the least index.

Note that since the maximum possible rate is $m\mu$, the capacity is upper-bounded by $m\mu/e$ nats per second (this follows straightforwardly as in (8)).

For the direct part, although inequality (a) in (8) can be made an equality by choosing Poisson input, inequality (d) in (8) is clearly not an equality. This is because, unlike the single-server system, the encoder is unable to keep the rate at the extreme values 0 and $m\mu$; the intermediate values are unavoidable. The following result gives achievable rates on the $\cdot/M/m$ queue timing channel.

Proposition 3: For the $\cdot/M/m$ queue timing channel

$$g(m) = \sup_{0 < a \leq m\mu} \left(h_m(a) - \sum_{i=1}^{m-1} p_{m,a}(i) h_m(i\mu) \right)$$

is an achievable rate, where $h_m(x) = x \log(m\mu/x)$ for $x \in (0, m\mu]$ and $p_{m,a}(\cdot)$ is the equilibrium state distribution for the $M/M/m$ queue with input rate a .

Proof: The input distribution ν is as in the proof of Proposition 2, with input rate a . The initial number of packets in queue at time $t = 0$ is as per the equilibrium distribution for an $M/M/m$ queueing system. We therefore have an $M/M/m$ queueing system starting at equilibrium. We now apply Burke's result for the $M/M/m$ queue [11, Ch. V, Theorem T1] to get $\hat{\lambda}_t = a$ for every $t \in [0, T]$. Consequently, inequality (a) of (8) is an equality, and we get

$$\frac{1}{T} I_T(X; Y) = h_m(a) - E \left[\frac{1}{T} \int_0^T dt h_m(\lambda_t) \right]$$

with $\lambda_t \in \{0, \mu, 2\mu, \dots, m\mu\}$. Upon simplification, we get

$$\frac{1}{T} I_T(X; Y) = h_m(a) - E \left[\sum_{i=0}^m f_T(i) h_m(i\mu) \right]$$

where $f_T(i)$ is the fraction of the time the system is in state i ; its expected value is $p_{m,a}(i)$, independent of T . Thus, for each a , the normalized mutual information is independent of T . Upon

optimization over a , we obtain that $g(m)$ is the supremum of normalized mutual information under *Poisson arrivals*.

To show achievability, however, we need to show that the liminf in probability of the normalized information density is at least as large as $g(m)$. Fix an arbitrary $\varepsilon > 0$ and choose a so that the normalized mutual information under Poisson arrivals is $g(m) - \varepsilon$. Fix any strictly increasing sequence of $(T_n : n \in \mathbb{N})$ such that $T_n \in (0, \infty)$ and $\lim_{n \rightarrow \infty} T_n = \infty$. Substitution of $\hat{\lambda}_t = a$ in (7) simplifies the normalized information density to

$$\begin{aligned} & \frac{1}{T_n} i_{T_n}(x; y) \\ &= \frac{1}{T_n} \left[\int_0^{T_n} \log(\lambda_t/a) dy_t + \int_0^{T_n} (a - \lambda_t) dt \right]. \end{aligned} \quad (11)$$

The term

$$\frac{1}{T_n} \int_0^{T_n} (a - \lambda_t) dt \rightarrow 0, \text{ in probability} \quad (12)$$

as in the proof of Proposition 2.

To analyze the first term, the following lemma turns out to be useful.

Lemma 1: The sequence of random variables $(W_n : n \in \mathbb{N})$ given by

$$W_n \triangleq \frac{1}{T_n} \int_0^{T_n} \log(\lambda_t/a) dY_t \quad (13)$$

converges almost surely and in L^1 to a constant $c \in \mathbb{R}$.

Proof: See Appendix I. □

This lemma implies, in particular, that W_n converges in probability to c . Moreover, convergence in L^1 implies that

$$\lim_{n \rightarrow \infty} E \left[\frac{1}{T_n} \int_0^{T_n} \log(\lambda_t/a) dY_t \right] = c.$$

However, we saw earlier that the normalized mutual information did not depend on T_n . In light of this fact, and (12), we get

$$g(m) - \varepsilon = E \left[\frac{1}{T_n} i_{T_n}(X; Y) \right] = c$$

which establishes that $g(m)$ is an achievable rate. □

Table I gives the achievable rate $g(m)$ in Proposition 3 as a function of m and compares it to the known upper bound m/e nats per second. We take $\mu = 1$ packet per second. The values were obtained numerically by varying the input load factor a/m in steps of 0.01. The load factor that achieves $g(m)$ is reported in the next row. The last row indicates the upper bound.

Discussion: Setting $m = 1$ we get Proposition 2. Proposition 3 is thus a generalization. Note that the lower bound on capacity for $m = 2$ and $m = 3$ give significant improvements over the ESTC. For higher values of m , our lower bound saturates at about 0.5804 μ nats per second. The capacity of the $\cdot/M/\infty$ queue is however infinite. See Appendix II for a coding technique with batch arrivals that achieves reliable communication at any arbitrary bit rate. The limitation of Poisson inputs is that the transmitter is unable to explicitly control the rate of the departure process and keep it at extremes (i.e., at 0 or $m\mu$), given the constraints of the channel. Multiple arrivals at the same instant (with sufficiently far apart batches of arrivals)

TABLE I
ACHIEVABLE RATES FOR MULTISERVER QUEUES; $\mu = 1$ PACKET PER SECOND

m	1	2	3	4	5	6	7	8
a/m	0.3679	0.34	0.31	0.28	0.25	0.22	0.19	0.17
$g(m)$	0.3679	0.5014	0.5524	0.5716	0.5780	0.5798	0.5801	0.5802
m/e	0.3679	0.7358	1.1036	1.4715	1.8394	2.2073	2.5752	2.9430

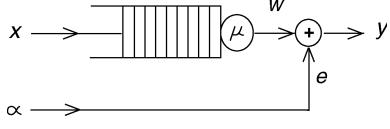


Fig. 1. Single-server queue with spurious departures.

is expected to give higher achievability rates. Our upper and the lower bounds are therefore quite loose and the capacity of the multiserver queue remains an interesting open problem.

IV. OTHER EXAMPLES

In this section, we analyze two simple configurations of exponential servers and get bounds on the timing channel capacity. These two simple examples demonstrate how to extend the point-process approach beyond the ESTC.

A. Single-Server Queue With Spurious Departures

Suppose that the output of a single-server queue is merged with a stream of Poisson arrivals having rate α (cf. Fig. 1). The departures from the two streams are indistinguishable to the receiver. The resulting model is similar to a point-process channel with background intensity α .

To simplify the recursion formulas in Proposition 4, we assume single arrivals at the input. Fix $T \in (0, \infty)$. The input space is $(\mathcal{X}, \mathcal{F}_T^X)$, where \mathcal{X} is the set of counting functions $x : [0, T] \rightarrow \mathbb{Z}_+$ (nondecreasing and right-continuous) with unit jumps and $x_0 \in \mathbb{Z}_+$. Let $(\mathcal{Y}, \mathcal{F}_T^Y)$ be the output space where \mathcal{Y} is the set of counting functions $y : [0, T] \rightarrow \mathbb{Z}_+$ with unit jumps and $y_0 = 0$.

Our first goal is to find the transition probability function that models this channel. We now outline the steps to do this.

Step 1: Fix $x \in \mathcal{X}$. We first identify an information pattern $(\mathcal{G}_t : t \in [0, T])$, where $\mathcal{F}_t^Y \subset \mathcal{G}_t$ for $t \in [0, T]$, so that Y has a known rate $\lambda = (\lambda_t : t \in [0, T])$ with respect to $(\mathcal{G}_t : t \in [0, T])$.

Let $(w_t : t \in [0, T])$ denote the departures from the queue, and let $(e_t : t \in [0, T])$ be the spurious departures. The output $y \in \mathcal{Y}$ observed by the receiver is $y = (y_t : t \in [0, T])$, where $y_t = e_t + w_t$ for $t \in [0, T]$. The service times are independent and exponentially distributed with mean $1/\mu$ seconds, and E is a Poisson process having rate α arrivals per second.

Fix $x \in \mathcal{X}$. Let $Q_t = x_t - W_t$ for $t \in [0, T]$. Clearly, with

$$\begin{aligned} \lambda_0 &= \alpha \\ \lambda_t &= \alpha + \mu 1\{Q_{t-} > 0\}, \quad t \in (0, T] \end{aligned}$$

Y has rate $\lambda = (\lambda_t : t \in [0, T])$ with respect to the information pattern $(\mathcal{G}_t : t \in [0, T])$, where $\mathcal{G}_t = \mathcal{F}_t^Y \vee \mathcal{F}_t^W$ for $t \in [0, T]$. Note that $\mathcal{F}_t^E \subset \mathcal{G}_t$ for $t \in [0, T]$ because $e_t = y_t - w_t$ for $t \in [0, T]$.

Step 2: The measure $P(x, \cdot)$ on $(\mathcal{Y}, \mathcal{F}_T^Y)$ that models the channel is represented by (cf. [11, Ch. VI, Result R8])

$$\frac{dP(x, \cdot)}{dP_0}(y) = \exp \left\{ \int_0^T \left[(\log \hat{\lambda}_t) dy_t + (1 - \hat{\lambda}_t) dt \right] \right\} \quad (14)$$

where Y has rate $\hat{\lambda}$ with respect to $(\mathcal{F}_t^Y : t \in [0, T])$ under $P(x, \cdot)$. Furthermore, we may assume that $\hat{\lambda}$ satisfies (cf. [16, Theorem 18.3], [21, Theorem 1.6])

$$\begin{aligned} \hat{\lambda}_0 &= \alpha, \\ \hat{\lambda}_t &= \lim_{s \uparrow t} E[\alpha + \mu 1\{Q_s > 0\} | \mathcal{F}_s^Y], \quad t \in (0, T]. \end{aligned} \quad (15)$$

Step 3: Fix $x \in \mathcal{X}$. Given the observed $y \in \mathcal{Y}$, we obtain the estimates for the queue states so that we can evaluate $E[1\{Q_t > 0\} | \mathcal{F}_t^Y]$ for $t \in [0, T]$. Substitution of this evaluation in (15) and (14) yields the transition probability function $P(x, dy)$ from $(\mathcal{X}, \mathcal{F}_T^X)$ to $(\mathcal{Y}, \mathcal{F}_T^Y)$. Given (14), the maximum-likelihood criterion for decoding is then straightforward. Equation (14) also tells us how to evaluate the information density, as we will see later in this subsection.

For the single-server queue with spurious departures, the following proposition shows how to calculate the estimates of queue sizes given the observations. Let $Z_t(n) \triangleq 1\{Q_t = n\}$ and $\hat{Z}_t(n) \triangleq E[1\{Q_t = n\} | \mathcal{F}_t^Y]$ for $n \in \mathbb{Z}_+$.

Proposition 4: Consider the single-server queue with spurious departures. Fix $x \in \mathcal{X}$. The process $(\hat{Z}_t(n) : t \in [0, T])$ for $n \in \mathbb{Z}_+$ can be recursively evaluated using the following update rules.

- (a) Initialize $\hat{Z}_0(n) = 1\{x_0 = n\}$ for $n \in \mathbb{Z}_+$.
- (b) If an arrival occurs at time τ , i.e., $dx_\tau = x_\tau - x_{\tau-} = 1$, then

$$\hat{Z}_\tau(n) = \hat{Z}_{\tau-}(n-1)1\{n > 0\}, \quad n \in \mathbb{Z}_+.$$

- (c) If a departure occurs at time τ , i.e., $dy_\tau = 1$, then

$$\hat{Z}_\tau(n) = \frac{\alpha \hat{Z}_{\tau-}(n) + \mu \hat{Z}_{\tau-}(n+1)}{\alpha + \mu (1 - \hat{Z}_{\tau-}(0))}, \quad n \in \mathbb{Z}_+.$$

- (d) Let τ_k and τ_{k+1} be two successive instants of discontinuity of $x + y$. Let $t \in (\tau_k, \tau_{k+1})$. Then

$$\hat{Z}_t(0) = \frac{\hat{Z}_{\tau_k}(0) \exp\{\mu(t - \tau_k)\}}{\hat{Z}_{\tau_k}(0) \exp\{\mu(t - \tau_k)\} + (1 - \hat{Z}_{\tau_k}(0))}$$

and for $n \in \mathbb{N}$

$$\hat{Z}_t(n) = \frac{\hat{Z}_{\tau_k}(n)}{\hat{Z}_{\tau_k}(0) \exp\{\mu(t - \tau_k)\} + (1 - \hat{Z}_{\tau_k}(0))}.$$

Proof: See Section IV-C. \square

Proposition 4 solves $(\hat{Z}_t(n) : t \in [0, T])$ explicitly for $n \in \mathbb{Z}_+$. Such explicit solutions are however hard to obtain in most

cases because of the difficulty in solving a system of nonlinear differential equations. In most cases, we can only write an integral equation for the updates. The single-server queue with spurious departures in this subsection and a pair of queues connected in tandem considered in Section IV-B are two exceptions where an explicit solution can indeed be found.

Step 4: We use [11, Ch. VI, Result R8] to obtain an expression for the mutual information. We proceed as in Section II-B2. Let ν be any probability measure on $(\mathcal{X}, \mathcal{F}_T^X)$. The input measure ν and the transition probability function $P(x, dy)$ in (14) define the joint probability measure π on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_T^X \vee \mathcal{F}_T^Y)$, denoted by $\pi(dx, dy) = \nu(dx)P(x, dy)$. We can again verify that under π , the stochastic process Y has rate $\hat{\lambda}$ with respect to the information pattern $(\mathcal{F}_t^X \vee \mathcal{F}_t^Y : t \in [0, T])$.

With π^X and π^Y denoting the restrictions of π to \mathcal{F}_T^X and \mathcal{F}_T^Y , respectively, we get that $\pi \ll \pi^X \times \pi^Y \ll \nu \times P_0$, and that $\pi^Y \ll P_0$. Furthermore, [11, Ch. VI, Result R8] gives

$$\frac{d\pi^Y}{dP_0}(y) = \exp \left\{ \int_0^T \left[(\log \hat{\lambda}_t) dy_t + (1 - \hat{\lambda}_t) dt \right] \right\}$$

where Y has rate $\hat{\lambda}$ with respect to $(\mathcal{F}_t^Y : t \in [0, T])$ under the probability measure π^Y . We may take $\hat{\lambda}_t = E[\hat{\lambda}_t | \mathcal{F}_{t-}^Y]$, for each $t \in [0, T]$ [16, Theorem 18.3]. The normalized information density $(1/T) i_T(x, y)$ is therefore given by

$$\begin{aligned} \frac{1}{T} \log \frac{d\pi}{d(\pi^X \times \pi^Y)}(x, y) \\ = \frac{1}{T} \int_0^T \left[(\log(\hat{\lambda}_t / \hat{\lambda}_t)) dy_t + (\hat{\lambda}_t - \hat{\lambda}_t) dt \right]; \end{aligned} \quad (16)$$

the normalized mutual information can be written as

$$\frac{1}{T} I_T(X; Y) = \frac{1}{T} E \int_0^T \left[\log(\hat{\lambda}_t / \hat{\lambda}_t) \right] dy_t \quad (17)$$

$$= \frac{1}{T} E \int_0^T dt \left[\phi(\hat{\lambda}_t) - \phi(\hat{\lambda}_t) \right] \quad (18)$$

where $\phi(u) = u \log u$.

Step 5: To find achievable information rates, we need to find the liminf in probability of the normalized information density, under a suitable input distribution.

For the single-server queue with spurious departures, let ν be such that

$$\nu \{X_0 = k\} = \left(1 - \frac{a}{\mu}\right) \left(\frac{a}{\mu}\right)^k$$

for $k \in \mathbb{Z}_+$, and $(X_t - X_0 : t \in (0, T])$ is a Poisson process having rate a . For each $t \in [0, T]$, we then have

$$\begin{aligned} \hat{\lambda}_t &= E \left[\hat{\lambda}_t | \mathcal{F}_{t-}^Y \right] \\ &= E \left[E \left[\lambda_t | \mathcal{F}_T^X \vee \mathcal{F}_{t-}^Y \right] | \mathcal{F}_{t-}^Y \right] \\ &= E \left[\lambda_t | \mathcal{F}_{t-}^Y \right] \quad (19) \\ &= \alpha + \mu E \left[1 \{Q_{t-} > 0\} | \mathcal{F}_{t-}^Y \right] \\ &= \alpha + a \end{aligned}$$

where (19) follows from [14, Theorem 34.4] and (20) follows from the fact that Q_s and (E_s, D_s) are independent, which im-

plies that Q_s and Y_s are independent. Substitution of (20) in (16) and (17) yields

$$\begin{aligned} \frac{1}{T} i_T(X, Y) &= \frac{1}{T} \sum_{k=1}^{Y_T} \log \left(\frac{\hat{\lambda}_{\tau_k}}{\alpha + a} \right) \\ &\quad + \frac{1}{T} \int_0^T \left((\alpha + a) - \hat{\lambda}_t \right) dt \end{aligned} \quad (21)$$

$$\frac{1}{T} I_T(X; Y) = E \left[\frac{1}{T} \sum_{k=1}^{Y_T} \log \left(\frac{\hat{\lambda}_{\tau_k}}{\alpha + a} \right) \right]. \quad (22)$$

For an increasing sequence $(T_n : n \in \mathbb{N})$, with $\lim_{n \rightarrow \infty} T_n = \infty$, we conjecture that the sequence $(1/T_n) i_{T_n}(X, Y)$ for the single-server queue with spurious departures converges in L^1 to a constant real number. Then this constant has to be the limiting normalized mutual information. Furthermore, the conjecture would imply convergence in probability to that constant. Some weaker results are known to hold—convergence in L^1 to a random variable, convergence almost surely to a finite value. They are consequences of uniform integrability and L^1 boundedness of the sequence of normalized information densities. For brevity, we omit the proofs of these weaker facts. As seen earlier in this paper, the conjecture holds true for the $M/M/m$ cases.

Under the conjecture, (21) and (22) are amenable to numerical evaluation through simulations. The simulation results are plotted in Fig. 2 and discussed at the end of this subsection.

Step 6: We can find an upper bound on the capacity of the single-server queue with spurious departures. Recall that $P(x, dy)$ (cf. (14)) is the transition probability function for the self-exciting point process having rate $\hat{\lambda}$ (cf. (15)), i.e., the rate $\hat{\lambda}_t$ at any instant of time instant t is determined by the input and by the past departures. The capacity of the point-process channel with instantaneous noiseless feedback, and where $\hat{\lambda}_t \in [\alpha, \alpha + \mu]$, is (cf. [2], [3])

$$C_\alpha = \alpha \left[e^{-1} (1 + \mu/\alpha)^{1+\alpha/\mu} - (1 + \alpha/\mu) \log(1 + \mu/\alpha) \right].$$

The capacity of the single-server queue with spurious departures is therefore upperbounded by C_α .

Simulation Results: The simulation results are plotted in Fig. 2 for $\mu = 1$ packet per second. The simulations are for $\alpha = 0.1$ and $\alpha = 0.5$ packets per second. The abscissa is the input rate a to the queue. The ordinate is the estimate of $(1/T) I_T(X; Y)$ in (22). The reported value for each a is an average of 500 values of $(1/T) i_T(x, y)$, where each realization of the process y has $y_T = 1000$. These simulations indicate that 0.21 nats per second is achievable when $\alpha = 0.1$, and that 0.09 nats per second is achievable when $\alpha = 0.5$; the corresponding upper bounds C_α are 0.251 and 0.132 nats per second, respectively, achieved at input intensity a/μ of 0.41 and 0.46, respectively. As can be observed from the figure, the maximizing a/μ values for the queueing case and the chosen α 's are close to the parameters of the point-process channel.

B. Two Queues in Tandem

In Section IV-A, we outlined a method to study the single-server queue with spurious departures. We now consider two exponential server queues in tandem with identical service rates of μ packets per second (cf. Fig. 3).

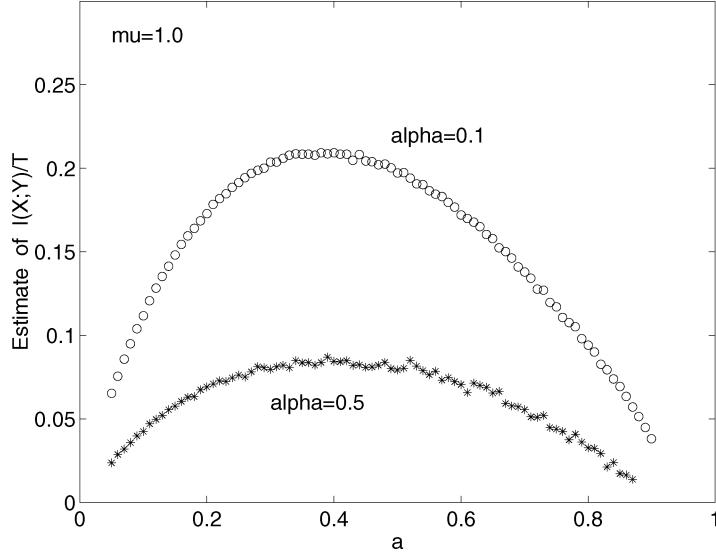


Fig. 2. Estimate of normalized mutual information on the single-server queue with spurious departures as a function of input rate, for $\alpha = 0.1$ and $\alpha = 0.5$.

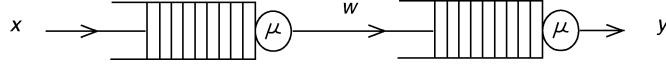


Fig. 3. Two exponential-server queues in tandem.

Step 1: Let $(\mathcal{X}, \mathcal{F}_T^X)$ and $(\mathcal{Y}, \mathcal{F}_T^Y)$ be as in Section IV-A. Let $x = (x_t : t \in [0, T])$ be the input, $w = (w_t : t \in [0, T])$ the departures from the first queue (i.e., the arrivals to the second queue), and $y = (y_t : t \in [0, T])$ the departures from the second queue, which are observed by the receiver. The transmitter can control the input process x and the initial state w_0 in the second queue. (For $t > 0$, it has no direct control or exact knowledge of w_t .) The state of the queue at any time $t \in [0, T]$ is $(Q_t^{(1)}, Q_t^{(2)})$, where $Q_t^{(1)} = x_t - W_t$ and $Q_t^{(2)} = W_t - Y_t$ for $t \in [0, T]$.

We choose $\mathcal{G}_t = \mathcal{F}_t^Y \vee \mathcal{F}_t^W$ for $t \in [0, T]$. Then for each fixed $x \in \mathcal{X}$, the process of departures Y is a point process having rate $\lambda = (\mu 1\{Q_t^{(2)} > 0\} : t \in [0, T])$ with respect to $(\mathcal{G}_t : t \in [0, T])$.

Step 2: The transition probability function $P(x, dy)$ that models the tandem-queue timing channel is represented by (14), where we take $\hat{\lambda}_0 = 0$ and for $t \in (0, T]$

$$\begin{aligned} \hat{\lambda}_t &= E[\lambda_t | \mathcal{F}_{t-}^Y] \\ &= \mu \lim_{s \uparrow t} E[1\{Q_s^{(2)} > 0\} | \mathcal{F}_s^Y]. \end{aligned} \quad (23)$$

Step 3: The estimates for the queue states can be evaluated from Proposition 5 below. Let

$$Z_t(n_1, n_2) \triangleq 1\{Q_t^{(1)} = n_1, Q_t^{(2)} = n_2\}$$

and

$$\hat{Z}_t(n_1, n_2) \triangleq E[1\{Q_t^{(1)} = n_1, Q_t^{(2)} = n_2\} | \mathcal{F}_t^Y]$$

for $n \in \mathbb{Z}_+$.

Proposition 5: Consider two queues in tandem with identical service rates of μ packets per second. Fix $x \in \mathcal{X}$. The process $(\hat{Z}_t(n_1, n_2) : t \in [0, T])$ can be recursively evaluated using the following update rules.

- (a) Initialize $\hat{Z}_0(n_1, n_2) = 1\{x_0 - w_0 = n_1, w_0 = n_2\}$ for $(n_1, n_2) \in \mathbb{Z}_+^2$.
- (b) If an arrival occurs at time τ , i.e., $dx_\tau = 1$, then

$$\hat{Z}_\tau(n_1, n_2) = \hat{Z}_{\tau-}(n_1 - 1, n_2) 1\{n_1 > 0\}$$

for $(n_1, n_2) \in \mathbb{Z}_+^2$.

- (c) If a departure occurs at time τ , i.e., $dy_\tau = 1$, then

$$\hat{Z}_\tau(n_1, n_2) = \frac{\hat{Z}_{\tau-}(n_1, n_2 + 1)}{1 - \hat{Z}_{\tau-}(n_1 + n_2 + 1, 0)}$$

for $(n_1, n_2) \in \mathbb{Z}_+^2$.

- (d) Let τ_k and τ_{k+1} be two successive instants of discontinuity of $x + y$. Let $t \in (\tau_k, \tau_{k+1})$, $s = t - \tau_k$, and $n = x_{\tau_k} - y_{\tau_k}$. There are exactly n packets in the system at time t . Furthermore

$$\hat{Z}_t(n, 0) = \frac{\hat{Z}_{\tau_k}(n, 0)}{1 + \mu s \hat{Z}_{\tau_k}(n, 0)};$$

for $1 \leq i \leq n - 1$, $\hat{Z}_t(n - i, i)$ is given by the expression

$$\begin{aligned} &\frac{\hat{Z}_{\tau_k}(n, 0)}{1 + \mu s \hat{Z}_{\tau_k}(n, 0)} \\ &+ \frac{e^{-\mu s} \sum_{j=1}^i \frac{(\mu s)^{i-j}}{(i-j)!} [\hat{Z}_{\tau_k}(n - j, j) - \hat{Z}_{\tau_k}(n, 0)]}{1 + \mu s \hat{Z}_{\tau_k}(n, 0)} \end{aligned}$$

and

$$\hat{Z}_t(0, n) = 1 - \sum_{i=0}^{n-1} \hat{Z}_t(n - i, i).$$

Proof: See Section IV-C. \square

Step 4: An input measure ν on $(\mathcal{X}, \mathcal{F}_T^X)$ and the transition probability function $P(x, dy)$ from $(\mathcal{X}, \mathcal{F}_T^X)$ to $(\mathcal{Y}, \mathcal{F}_T^Y)$ define the joint measure π on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_T^X \vee \mathcal{F}_T^Y)$ as before. With other analogous definitions, the normalized mutual information $(1/T)I(X; Y)$ is given by (17) and (18).

Step 5: As expected, we can evaluate $\hat{\lambda}_t$ when ν is such that the arrivals are Poisson with rate $a < \mu$ in $(0, T]$, and the queue

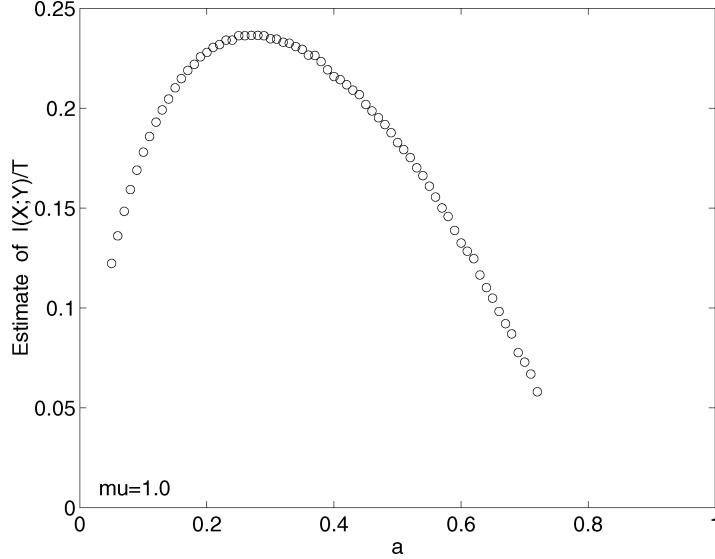


Fig. 4. Estimate of normalized mutual information for two queues connected in tandem as a function of input rate.

size is in equilibrium at time $t = 0$. Under these circumstances, $\hat{\lambda}_t = a$ for each $t \in [0, T]$, which leads to

$$\frac{1}{T} i_T(X; Y) = \frac{1}{T} \sum_{k=1}^{Y_T} \log \left(\frac{\hat{\lambda}_{\tau_k}}{a} \right) + \frac{1}{T} \int_0^T (a - \hat{\lambda}_t) dt \quad (24)$$

$$\frac{1}{T} I_T(X; Y) = E \left[\frac{1}{T} \sum_{k=1}^{Y_T} \log \left(\frac{\hat{\lambda}_{\tau_k}}{a} \right) \right]. \quad (25)$$

For this case too, for an increasing sequence $(T_n : n \in \mathbb{N})$ such that $\lim_{n \rightarrow \infty} T_n = \infty$, we conjecture that the sequence of normalized information density $(1/T_n) i_{T_n}(X; Y)$ converges in L^1 to a constant real number. Yet again, (24) and (25) are amenable to simulation. The simulation results are shown in Fig. 4. The capacity of this system is an open problem.

Simulation Results: In Fig. 4, the service rate is $\mu = 1$ packet per second. The abscissa is the input rate a to the first queue. The ordinate is the estimate of $(1/T) I_T(X; Y)$ in (25). The reported value for each a is an average of 200 values of $(1/T) i_T(x; y)$, where each realization of the second queue's departures y has $y_T = 1000$. The simulation indicates that 0.23 nats per second is achievable. The capacity of this system cannot exceed the capacity of the $\cdot/M/1$ queue, which is 0.36 nats per second.

C. Proofs of Proposition 4 and 5

Proof Outline: We first discuss the key ideas in the proofs with reference to Proposition 4. We also highlight the necessary modifications to the statements and proofs of [11, Ch. IV, Theorems T1–T2] for application in our setting. The same ideas are employed in the proof of Proposition 5.

Proposition 4 tells us how to evaluate the projections $\hat{Z}_t(n) = E[Z_t(n) | \mathcal{F}_t^Y]$ of the indicator function $Z_t(n) = 1 \{Q_t = n\}$, on the information pattern \mathcal{F}_t^Y . To prove Proposition 4, we first identify a representation for $Z_t(n)$ of the following form:

$$Z_t = Z_0 + \int_0^t h_s dx_s + \int_0^t f_s ds + m_t \quad (26)$$

where we have temporarily discarded the argument n in $Z_t(n)$. In (26), $m = (m_t : t \in [0, T])$ is a zero-mean \mathcal{G}_t -martingale which is almost surely of bounded variation, f and h are predictable with respect to $(\mathcal{G}_t : t \in [0, T])$. We assume that x and Y do not have common jumps. (Bremaud [11, Ch. IV, eq. 1.1] calls (26) a *semi-martingale* when the term involving the Lebesgue–Stieltjes integral is absent). We then follow the proofs of [11, Ch. IV, Theorems T1–T2], specialized to the case when f and h are predictable with respect to $(\mathcal{G}_t : t \in [0, T])$, to show that

$$\hat{Z}_t = \hat{Z}_0 + \int_0^t \hat{h}_s dx_s + \int_0^t \hat{f}_s ds + \hat{m}_t \quad (27)$$

where \hat{m} is a zero-mean \mathcal{F}_t^Y -martingale, and \hat{h} and \hat{f} are given by

$$\hat{h}_t = E [h_t | \mathcal{F}_{t-}^Y] = \lim_{s \uparrow t} E [h_s | \mathcal{F}_s^Y] \quad (28)$$

$$\hat{f}_t = E [f_t | \mathcal{F}_{t-}^Y] = \lim_{s \uparrow t} E [f_s | \mathcal{F}_s^Y]. \quad (29)$$

That \hat{f} and \hat{h} are predictable with respect to $(\mathcal{F}_t^Y : t \in [0, T])$ follows from the predictability of f and h with respect to $(\mathcal{G}_t : t \in [0, T])$, and from [16, Lemma 19.11]. Additionally, the Lebesgue–Stieltjes integral $\int_0^t h_s dx_s$ in Z_t is handled in exactly the same way as $\int_0^t f_s ds$ is handled in the proof of [11, Ch. IV, Theorems T1–T2]; the details are therefore omitted.

The *innovations process* \hat{m}_t can be represented as

$$\int_0^t K_s (dy_s - \hat{\lambda}_s ds) \quad (30)$$

where the process $(K_t : t \in [0, T])$ is predictable with respect to $(\mathcal{F}_t^Y : t \in [0, T])$ and plays the role of *innovations gain*. This gain is a sum of components that have to satisfy certain conditions; we indicate these in the proofs below. For more details, the reader is referred to [11, Ch. IV, Theorems T1–T2, T8].

As a final step, we explicitly solve the set of integral equations for $\hat{Z}_t(n)$. We are now ready to look at the proof of Proposition 4 in detail.

Proof of Proposition 4: Recall that the process $w = (w_t : t \in [0, T])$ of departures from the queue has rate $(\mu(1 - Z_{t-}(0)) : t \in [0, T])$ with respect to $(\mathcal{G}_t : t \in [0, T])$. Moreover, we can write

$$\begin{aligned} Z_t(n) &= Z_0(n) + \int_0^t [Z_{s-}(n-1)1\{n > 0\} - Z_{s-}(n)] dx_s \\ &\quad + \int_0^t [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] dw_s \\ &= Z_0(n) + \int_0^t [Z_{s-}(n-1)1\{n > 0\} - Z_{s-}(n)] dx_s \\ &\quad + \int_0^t [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] \\ &\quad \cdot \mu(1 - Z_{s-}(0)) ds + u_t(n) \end{aligned} \quad (31)$$

where

$$u_t(n) \triangleq \int_0^t [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] \cdot [dw_s - \mu(1 - Z_{s-}(0)) ds]. \quad (32)$$

Observe that for every $n \in \mathbb{Z}_+$ we have

$$\begin{aligned} [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] (1 - Z_{s-}(0)) \\ = [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}]. \end{aligned} \quad (33)$$

Furthermore, the process $Z(n) = (Z_t(n) : t \in [0, T])$ is bounded (in fact, either 0 or 1). From [11, Ch. II, Theorem T8(β)], the process $u(n) = (u_t(n) : t \in [0, T])$ is a $(\mathcal{G}_t : t \in [0, T])$ -martingale because

$$E \left[\int_0^t (Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}) \mu(1 - Z_{s-}(0)) ds \right]$$

is bounded within $[-2\mu t, 2\mu t]$ for every $t \in [0, T]$. Also, $u(n)$ is almost surely of bounded variation on $[0, T]$. Equation (31) is of the form (26). After substitution of (33) in (31), and from (27) and (30), we get

$$\begin{aligned} \hat{Z}_t(n) &= \hat{Z}_0(n) + \int_0^t [\hat{Z}_{s-}(n-1)1\{n > 0\} - \hat{Z}_{s-}(n)] dx_s \\ &\quad + \int_0^t [\hat{Z}_{s-}(n+1) - \hat{Z}_{s-}(n)1\{n > 0\}] \mu ds \\ &\quad + \int_0^t [\Psi_{1,s}(n) + \Psi_{2,s}(n) - \hat{Z}_{s-}(n)] [dy_s - \hat{\lambda}_s ds] \end{aligned} \quad (34)$$

where

$$\hat{\lambda}_s = \alpha + \mu(1 - \hat{Z}_{s-}(0)), \quad \text{for } s \in [0, T]$$

and $(\Psi_{i,t} : t \in [0, T])$, $i = 1, 2$, are predictable with respect to $(\mathcal{F}_t^Y : t \in [0, T])$, and satisfy

$$E \left[\int_0^t C_s Z_s(n) \lambda_s ds \right] = E \left[\int_0^t C_s \Psi_{1,s}(n) \hat{\lambda}_s ds \right] \quad (35)$$

$$E \left[\int_0^t C_s \Delta u_s(n) dy_s \right] = E \left[\int_0^t C_s \Psi_{2,s}(n) \hat{\lambda}_s ds \right] \quad (36)$$

for all nonnegative bounded processes $(C_t : t \in [0, T])$ that are predictable with respect to $(\mathcal{F}_t^Y : t \in [0, T])$.

To get $\Psi_{1,s}(n)$, observe that

$$\begin{aligned} Z_{s-}(n) \lambda_s &= Z_{s-}(n) (\alpha + \mu(1 - Z_{s-}(0))) \\ &= \alpha Z_{s-}(n) + \mu Z_{s-}(n) 1\{n > 0\} \\ &= Z_{s-}(n) (\alpha + \mu 1\{n > 0\}). \end{aligned}$$

Moreover, $\hat{\lambda}_s = \alpha + \mu(1 - \hat{Z}_{s-}(0))$. After substitution of these in (35), we can check that

$$\Psi_{1,s}(n) = \frac{\hat{Z}_{s-}(n)(\alpha + \mu 1\{n > 0\})}{\alpha + \mu(1 - \hat{Z}_{s-}(0))}. \quad (37)$$

To get $\Psi_{2,s}(n)$, we write

$$\Delta u_s(n) dy_s = C_s [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] dw_s.$$

This leads to

$$\begin{aligned} E \left[\int_0^t C_s \Delta u_s(n) dy_s \right] &= E \left[\int_0^t C_s [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] dw_s \right] \\ &= E \left[\int_0^t C_s [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] \right. \\ &\quad \cdot \mu(1 - Z_{s-}(0)) ds \left. \right] \\ &= E \left[\int_0^t C_s \cdot \mu \cdot (Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}) ds \right] \end{aligned} \quad (38)$$

where we have used (33) to get the last equality. A comparison of (38) with (36) shows that

$$\Psi_{2,s}(n) = \frac{\mu(\hat{Z}_{s-}(n+1) - \hat{Z}_{s-}(n)1\{n > 0\})}{\alpha + \mu(1 - \hat{Z}_{s-}(0))}. \quad (39)$$

Substitution of (37) and (39) in (34) yields

$$\begin{aligned} \hat{Z}_t(n) &= \hat{Z}_0(n) \\ &\quad + \int_0^t [\hat{Z}_{s-}(n-1)1\{n > 0\} - \hat{Z}_{s-}(n)] dx_s \\ &\quad + \int_0^t \hat{Z}_{s-}(n) (1\{n = 0\} - \hat{Z}_{s-}(0)) \mu ds \\ &\quad + \int_0^t \mu \left[\frac{\hat{Z}_{s-}(n+1) - \hat{Z}_{s-}(n)(1 - \hat{Z}_{s-}(0))}{\alpha + \mu(1 - \hat{Z}_{s-}(0))} \right] dy_s. \end{aligned} \quad (40)$$

The update rules (a), (b), and (c) follow straightforwardly from (40). For $t \in (\tau_k, \tau_{k+1})$ where τ_k and τ_{k+1} are two consecutive points of discontinuity of $x + y$, observe that

$$\frac{d\hat{Z}_t(n)}{dt} = \mu \hat{Z}_t(n) (1\{n = 0\} - \hat{Z}_t(0))$$

for $n \in \mathbb{Z}_+$, a system of differential equations with initial conditions $\hat{Z}_{\tau_k}(n)$ for $n \in \mathbb{Z}_+$. Using standard techniques to solve differential equations, we get (d). \square

Proof of Proposition 5: We give only a brief outline of the proof; most of the steps are analogous to the proof of Proposition 4. The state $Z_t(n_1, n_2)$ can be represented by

$$\begin{aligned} & Z_t(n_1, n_2) \\ &= Z_0(n_1, n_2) \\ &+ \int_0^t [Z_{s-}(n_1 - 1, n_2)1\{n_1 > 0\} - Z_{s-}(n_1, n_2)] dx_s \\ &+ \int_0^t [-Z_{s-}(n_1, n_2)1\{n_2 > 0\} + Z_{s-}(n_1, n_2 + 1)] dy_s \\ &+ \int_0^t [-Z_{s-}(n_1, n_2)1\{n_1 > 0\} \\ &+ Z_{s-}(n_1 + 1, n_2 - 1)1\{n_2 > 0\}] dw_s. \end{aligned}$$

This, after evaluation and substitution of the innovations gain, leads to the following equation for the estimates:

$$\begin{aligned} & \hat{Z}_t(n_1, n_2) \\ &= \hat{Z}_0(n_1, n_2) \\ &+ \int_0^t [\hat{Z}_{s-}(n_1 - 1, n_2)1\{n_1 > 0\} - \hat{Z}_{s-}(n_1, n_2)] dx_s \\ &+ \int_0^t \left[-\hat{Z}_{s-}(n_1, n_2) + \frac{\hat{Z}_{s-}(n_1, n_2 + 1)}{1 - \hat{Z}_{s-}(\cdot, 0)} \right] dy_s \\ &+ \int_0^t \left[\hat{Z}_{s-}(n_1, n_2) (1 - \hat{Z}_{s-}(\cdot, 0)) \right. \\ &\quad \left. - \hat{Z}_{s-}(n_1, n_2)1\{n_2 > 0\} - \hat{Z}_{s-}(n_1, n_2)1\{n_1 > 0\} \right. \\ &\quad \left. + \hat{Z}_{s-}(n_1 + 1, n_2 - 1)1\{n_2 > 0\} \right] \mu ds, \end{aligned} \quad (41)$$

where

$$\hat{Z}_s(\cdot, 0) = \sum_{n \in \mathbb{Z}_+} \hat{Z}_s(n, 0).$$

The update rules (a), (b), and (c) then follow straightforwardly. Let $n = x_{\tau_k} - y_{\tau_k}$. To get the update rule (d) for $t \in (\tau_k, \tau_{k+1})$, we need to solve the system

$$\begin{aligned} & \frac{1}{\mu} \frac{d\hat{Z}_t(n - i, i)}{dt} \\ &= \hat{Z}_t(n - i, i) (1 - 1\{n - i > 0\} - 1\{i > 0\}) \\ &\quad - \hat{Z}_t(n - i, i) \hat{Z}_t(n, 0) + \hat{Z}_t(n - i + 1, i - 1)1\{i > 0\} \end{aligned}$$

for $0 \leq i \leq n$. Let $n > 0$. When $i = 0$, we solve the differential equation

$$\frac{d\hat{Z}_t(n, 0)}{dt} = -\mu (\hat{Z}_t(n, 0))^2$$

to get

$$\hat{Z}_t(n, 0) = \frac{\hat{Z}_{\tau_k}(n, 0)}{1 + \mu(t - \tau_k) \hat{Z}_{\tau_k}(n, 0)}.$$

Suppose now that $1 \leq i \leq n - 1$. Then

$$\frac{1}{\mu} \frac{d\hat{Z}_t(n - i, i)}{dt}$$

$$= -\hat{Z}_t(n - i, i) (1 + \hat{Z}_t(n, 0)) + \hat{Z}_t(n - i + 1, i - 1).$$

We search for solutions of the form $\hat{Z}_t(n - i, i) = h_t(i)g_t$. We can always take

$$g_t = \frac{\exp\{-\mu(t - \tau_k)\}}{1 + \mu(t - \tau_k) \hat{Z}_{\tau_k}(n, 0)}$$

which implies that

$$\begin{aligned} \frac{dh_t(i)}{dt} \frac{\exp\{-\mu(t - \tau_k)\}}{1 + \mu(t - \tau_k) \hat{Z}_{\tau_k}(n, 0)} \\ = \mu \hat{Z}_t(n - i + 1, i - 1), \quad 1 \leq i \leq n - 1. \end{aligned}$$

We now proceed by induction. Solving for $h_t(i)$, substituting it in $\hat{Z}_t(n - i, i) = h_t(i)g_t$, we obtain the update rule (d) for $1 \leq i \leq n - 1$. The expression for $\hat{Z}_t(0, n)$ now follows because

$$\sum_{i=0}^n \hat{Z}_t(n - i, i) = 1, \quad \text{for every } t \in (\tau_k, \tau_{k+1}). \quad \square$$

V. SUMMARY AND DISCUSSION

We gave a conceptually simple proof of the capacity of the ESTC (Proposition 2). Our proof emphasizes the connection between the point-process channel and the ESTC. We also observed that the capacity region of the two-user ESTC is a triangle. We then showed a lower bound on the capacity of multiserver queues (Proposition 3 and Table I) by optimizing the mutual information over Poisson inputs. The bounds are quite loose when there are several servers, and it would be interesting to find tighter lower and upper bounds.

Estimates for the queue size, given partial information (either the departures alone, or the departures and arrivals) play a key role in determining the sample function densities and therefore the mutual information between the input and the output. This observation leads to a methodology to study timing channels that arise in some simple networks. We looked at two examples, the single-server queue with spurious departures, and a pair of queues connected in tandem. In these two special cases, we could explicitly write an expression for the sample function density (cf. Propositions 4 and 5, (15), (23), and (14)). In other examples, such as the single-server queue with finite buffer size, or the single-server queue where the spurious packets are *input* to the queue, we can write a system of integral equations for updating the estimates for queue sizes along the lines of (40) and (41); however, an explicit solution analogous to Propositions 4 and 5 is not known.

In the two examples considered, under a conjecture that the normalized information density converges in L^1 to a constant real number, we identified achievable rates using simulations with Poisson inputs. For the single-server queue with spurious departures of rate 0.1μ , our simulations indicate that the capacity is at least 0.21μ nats per second (Fig. 2). The capacity is upper-bounded by C_α which is 0.251μ nats per second. For the pair of queues connected in tandem, our simulations indicate that the capacity is at least 0.23μ nats per second (Fig. 4). The capacity is upper-bounded by the ESTC's capacity 0.3679μ nats per second.

APPENDIX I PROOF OF LEMMA 1

Proof

We first show that $W_n \rightarrow c$, almost surely. Recall that W_n is given by (13). When a departure occurs, λ_t is one of $\{\mu, 2\mu, \dots, m\mu\}$. In particular, just prior to the departing epoch, if r packets are in the queue, $\lambda_t = \mu \min\{m, r\}$, $r \geq 1$. We can therefore rewrite W_n as

$$W_n = \sum_{r=1}^{m-1} \left(\log \frac{r\mu}{a} \right) \frac{1}{T_n} \sum_{i=1}^{y_{T_n}} N_{T_n}(r) + \left(\log \frac{m\mu}{a} \right) \frac{1}{T_n} \sum_{i=1}^{y_{T_n}} N_{T_n}(m) \quad (42)$$

where $N_{T_n}(r)$ for $0 < r < m$ is the number of departures in $[0, T_n]$ that saw r packets in queue prior to their departing epochs, and $N_{T_n}(m)$ is the number of departures that saw m or more packets in queue prior to their departing epochs. Clearly

$$Y_{T_n} = \sum_{r=1}^{m-1} N_{T_n}(r) + N_{T_n}(m). \quad (43)$$

By the renewal theorem, $Y_{T_n}/T_n \rightarrow a$, almost surely, because the departures are Poisson. Moreover, because the $M/M/m$ queue states form a continuous-time Markov chain, the time between two successive departures that see r packets in the queue prior to their departures ($0 < r < m$), is independent and identically distributed with mean duration $\geq a$. Another application of the renewal theorem indicates that $N_{T_n}(r)/T_n \rightarrow a_r$, almost surely, where $a_r < \infty$. (Note that this argument holds even if $a_r = 0$.) The almost sure convergences of Y_{T_n}/T_n and $N_{T_n}(r)/T_n$ for $0 < r < m$ to finite constants, and (43), imply that $N_{T_n}(m)/T_n$ also converges almost surely to a real number, thereby establishing that W_n in (13) converges to a $c \in \mathbb{R}$ almost surely.

We next show that the sequence $(W_n : n \in \mathbb{Z}_+)$ is uniformly integrable ([7, p. 322]), i.e.,

$$\sup_n E[|W_n| 1\{|W_n| > \alpha\}] \rightarrow 0, \quad \text{as } \alpha \rightarrow \infty. \quad (44)$$

From (42) and (43), we can bound W_n as follows:

$$|W_n| \leq M \cdot \frac{Y_{T_n}}{T_n}$$

where

$$M = \max\{|\log(r\mu/a)| : 1 \leq r \leq m\} < \infty.$$

Furthermore, for every $\alpha > 0$, we have

$$|W_n| 1\{|W_n| > \alpha\} \leq M \cdot \frac{Y_{T_n}}{T_n} \cdot 1\left\{\frac{Y_{T_n}}{T_n} > \frac{\alpha}{M}\right\}.$$

Observe that Y_{T_n} is a Poisson random variable with mean $1/(aT_n)$. Consequently

$$\begin{aligned} E[|W_n| 1\{|W_n| > \alpha\}] &\leq M \cdot E\left[\frac{Y_{T_n}}{T_n} \cdot 1\left\{\frac{Y_{T_n}}{T_n} > \frac{\alpha}{M}\right\}\right] \\ &= \frac{M}{T_n} \sum_{k > \alpha T_n/M} k \frac{(aT_n)^k e^{-aT_n}}{k!} \end{aligned}$$

$$\begin{aligned} &= aM \sum_{k > \alpha T_n/M} \frac{(aT_n)^{k-1} e^{-aT_n}}{(k-1)!} \\ &= aM \Pr\{Y_{T_n} > \alpha T_n/M - 1\} \end{aligned}$$

$$\leq aM \frac{E[Y_{T_n}]}{\alpha T_n/M - 1} \quad (45)$$

$$\begin{aligned} &= aM \frac{aT_n}{\alpha T_n/M - 1} \\ &\leq \frac{2a^2 M^2}{\alpha} \end{aligned} \quad (46)$$

where (45) follows from Markov's inequality, and (46) is true for all $n \in \mathbb{N}$, if $\alpha > 2M/T_0$. Since (46) goes to 0 as $\alpha \rightarrow \infty$, the sequence $(W_n : n \in \mathbb{N})$ is uniformly integrable.

A uniformly integrable sequence of random variables that converges in probability also converges in L^1 , and this ends the proof. \square

APPENDIX II •/M/∞ QUEUE

Proposition 6: The capacity of a $\cdot/M/\infty$ queue is ∞ .

Proof: Let $(T_n : n \in \mathbb{N})$ be an arbitrary sequence with $T_n \in [1, \infty)$ and $\lim_{n \rightarrow \infty} T_n = \infty$. Let $(\varepsilon_n : n \in \mathbb{N})$ be an arbitrary sequence with $\varepsilon_n \in (0, 1)$ and $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Fix n . In the following, we will not explicitly show the dependence on n . So, we will work with the indexed T_n and ε_n , but will denote them by T and ε , respectively.

Fix an arbitrary target rate $R \in \mathbb{R}_+$. Let $M = 2^{\lceil RT \rceil}$ be the number of codewords. Fix $p \in \mathbb{N}$ such that $e^{-p\mu/M} < \varepsilon$. The queue is assumed to be empty at $t = 0$. To transmit message m , the transmitter floods the queue with p packets precisely at the instant m/M seconds, where $m = 0, \dots, M-1$.

The receiver identifies the instant of the first departure from the queue. Let this epoch be Y . The decoding rule is described as follows:

$$\hat{m} = \begin{cases} \lfloor MY \rfloor, & Y \in [0, 1) \\ M-1, & Y \geq 1 \end{cases}$$

i.e., the receiver does not observe beyond $[0, 1]$.

It is easy to see that if an error occurs, then the minimum of the service times for the p packets is at least $1/M$. This minimum random variable is exponentially distributed with mean $1/(p\mu)$; the probability that the minimum is at least $1/M$ is $e^{-p\mu/M} < \varepsilon$. Thus, the probability of error is upper-bounded by ε , and we have a (T, M, ε) -code (indeed, a $(1, M, \varepsilon)$ -code). Since R was arbitrary, the proposition follows. \square

ACKNOWLEDGMENT

The authors would like to thank an anonymous reviewer for pointing out a gap in the proof of Proposition 3 in the original submission. The first author would like to thank Prof. Anurag Kumar for a discussion that led to the corrected proof.

REFERENCES

- [1] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.
- [2] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory Probab. Appl.*, vol. 23, pp. 143–147, 1978.

- [3] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 710–715, Nov. 1980.
- [4] R. Sundaresan and S. Verdú, "Robust decoding for timing channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 405–419, Mar. 2000.
- [5] A. Bedekar and M. Azizoğlu, "The information-theoretic capacity of discrete-time queues," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 446–461, Mar. 1998.
- [6] J. A. Thomas, "On the Shannon capacity of discrete-time queues," in *Proc. 1997 IEEE Int. Symp. Information Theory*, Ulm, Germany, Jul. 1997, p. 333.
- [7] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 2nd ed. New York: Oxford Univ. Press, 1992.
- [8] R. Sundaresan, "Coded communication over timing channels," Ph.D. dissertation, Princeton Univ., Princeton, NJ, 1999.
- [9] E. Arikan, "On the reliability exponent of the exponential timing channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1681–1689, Jun. 2002.
- [10] A. B. Wagner and V. Anantharam, "Zero-rate reliability of the exponential-server timing channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 447–465, Feb. 2005.
- [11] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*. New York: Springer-Verlag, 1981.
- [12] K. L. Chung, *A Course in Probability Theory*, 2nd ed. New York: Academic, 1974.
- [13] D. L. Snyder, *Random Point Processes*. New York: Wiley, 1975.
- [14] P. Billingsley, *Probability and Measure*, 2nd ed. New York: Wiley, 1986.
- [15] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
- [16] R. S. Liptser and A. N. Shirayev, *Statistics of Random Processes*. New York: Springer-Verlag, 1978, vol. 2.
- [17] D. L. Snyder, "Filtering and detection for doubly stochastic Poisson processes," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 91–102, Jan. 1972.
- [18] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 350–361, May 1969.
- [19] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," in *Proc. 2004 IEEE Workshop on Information Theory*, San Antonio, TX, Oct. 2004, pp. 265–270.
- [20] S. Verdú and T. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
- [21] R. S. Liptser and A. N. Shirayev, *Statistics of Random Processes*. New York: Springer-Verlag, 1977, vol. 1.
- [22] R. G. Gallager, *Discrete Stochastic Processes*. Boston, MA: Kluwer Academic, 1996.
- [23] S. Verdú, "Information theory of queueing systems," presented at the 1995 IEEE Information Theory Workshop on Information Theory, Multiple Access and Queueing, St. Louis, MO, Apr. 19–21, 1995.