CODED COMMUNICATION OVER TIMING CHANNELS

Rajesh Sundaresan

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ELECTRICAL ENGINEERING

November 1999

Abstract

In this thesis, we consider the problem of reliable transmission of information via the times of arrival to a queueing system. On the single-server queue, we show that the exponential server channel's maximum-likelihood decoder is a robust decoder. We give guarantees on this decoding criterion's performance in situations when the service times are stationary and ergodic, or when certain unmodeled phenomena, such as when an adversary deliberately hinders communication, affect the channel. We also show analogous results on the discrete-time single-server queue.

Using a point-process approach, we give a conceptually simple proof for the capacity of the exponential server queue. Our results indicate an alternative strategy with complete feedback that achieves capacity on the point-process channel. Furthermore, the point-process approach enables us to study timing channels that arise in multiserver queues, queues in tandem, and other simple configurations. Although the capacities of such channels remain to be found, we provide some bounds obtained either analytically or from simulations.

We then consider the problem of finding coding schemes that have good performance with computationally feasible decoding strategies. On the exponential server queue, we show the existence of a tree code that performs well in conjunction with the sequential decoding technique. The expected number of computations before moving one step ahead in the correct direction is upperbounded by a finite number. The rate of information transfer for this code is one half of the capacity.

Acknowledgments

I sincerely thank Professor Sergio Verdú for his guidance and valuable advice throughout the course of my research. He has been a tireless source of fresh ideas. Many questions tackled in this thesis were his suggestions.

I express my thanks to the readers of this thesis, Professor Sanjeev Kulkarni, Professor H. Vincent Poor and Professor Sergio Verdú, for their comments and suggestions that helped improve the presentation.

I thank Karthik Visweswariah and Sathyakama Sandilya for discussions on many aspects of this thesis. Their ideas and insights have gone into many proofs.

I thank the faculty members in the Electrical Engineering, Computer Science, Mathematics and the erstwhile Civil Engineering & Operations Research departments for the numerous courses I attended. They have contributed significantly to my education.

I thank all my friends in Princeton, especially the members of the cricket team and the 3 O' Clock coffee club, for their friendship. They have made my stay in Princeton a very pleasant one.

The research for this thesis was supported by the National Science Foundation under Grant NCR 9523805.

Contents

Chapter 1

Introduction

1.1 Timing Channels

Often, we transmit information not only in the content of messages, but also in the times at which we send these messages. Pauses in spoken language for example convey information that adds to the information contained in the speech itself. We can also convey information through the intervals between telephone calls. If we encode information in the time interval between telephone calls, the receiver can extract this information without answering the calls. This strategy uses only the telephone signaling channel; the voice channel is unused. Unanswered calls are toll-free, and information can therefore be transmitted at no extra cost over the telephone signaling channel.

Consider a single-server queue where packets carrying information suffer some service delay before they exit the system. Unserved packets are stored in a queue. On this system, information can be encoded in the times of arrival of packets. Assume that the queue is empty at time 0 seconds. If the service time for each packet is deterministic, say s seconds, then a packet input to the queue at time x seconds exits at time $x + s$ seconds. The receiver, knowing s, can extract the value x by observing the exit time. If $x \in [0, 1]$, an infinite number of bits can be transmitted in finite time, leading to infinite capacity.

In many situations, however, random mechanisms obfuscate the timing information. In the above example, the service times may be random, and the queue may not be empty at time 0 seconds. The times of departure thus contain a noisy version of the information encoded in the times of arrival. An interesting problem then is to assess rates at which information can be transmitted reliably across such noisy channels via timing.

Such studies are of interest because timing information can be "piggybacked" on existing data networks to boost "bandwidth". Consider the single-server queue where packets or customers are single bits. If the service rate is μ bits per second, and each bit is received noiselessly, it would seem at first glance that the capacity of this communication link is μ bits per second. The capacity is however strictly larger than μ bits per second [1], regardless of the service distribution. This surprising result, reliable transfer of information at a rate faster than the rate at which bits exit the system, holds because information can be encoded in the times of arrival. This idea is well-illustrated in the queue with deterministic service times where the capacity is infinite.

Timing channels are also of interest because they may violate system security requirements. The headquarters of an intelligence agency is connected by e-mail to its outside agents. Strict monitoring procedures are in place to prevent the flow of unauthorized information from the agency to the outside world. It is however possible to encode covert information in the timing of innocuous e-mail messages.

Consider another example where two processors at different levels of security are supported on the same system. Both send jobs to a time-shared facility, for e.g., they may share a common memory system and the jobs may be memory accesses. The two processors, being at different levels of security, should not communicate with each other. They can however violate this constraint by using the following strategy. One processor encodes information by varying the rate at which it sends jobs to the time-shared facility. The response time of the facility depends on its instantaneous load; the other processor therefore gets a noisy version of the information by measuring the response time to its own jobs. Studies of such channels are done while designing computer systems for high security applications [2], and are of considerable interest to the computer security community.

Before we describe the problems tackled in this thesis, we describe some prior results on timing channels.

1.2 Previous Results

Studies of information contained in timing were pioneered in [3]. We describe the results of [3] in the following simple setting. Consider a network that consists of two nodes and a single link between these nodes. K sources are connected to the first node and K receivers are connected to the second node. Source k sends messages only to receiver k ($1 \leq k \leq K$). Each message is a string of data bits, followed by a string of idle time slots. The statistical description of the lengths of the message and the idle period are known. If the expected delay per message for reproduction of the message at the receiver cannot exceed a certain value, then some protocol information about the start of messages and message lengths has to be transmitted in addition to the data bits. This protocol information is the overhead required to meet the expected delay constraint. Upper and lower bounds on the required amount of protocol information were given in [3].

In timing channels, however, the arrival times of message are seen as a means of communicating information. From this point of view, many interesting results on the capacity of the single-server queue were derived in [1]. These results were extended to the discrete-time single-server queue in [4] and [5]. Such channels are closely related to channels with point-process observations (cf. [6], [7], [8]). We now describe these results in some detail.

We focus on the simpler model in which all packets are identical and do not contain information (cf. [1], [4], [5]). Information is therefore contained only in the times of arrival to the queue. Every message is encoded by a different sequence of n arrival times to the queue. The decoder, knowing the codebook and the statistical description of the queue, selects one of the possible messages upon observation of the corresponding n departure times. The rate of the code is equal to the logarithm of the number of messages divided by the average time it takes to receive all n packets. The mechanisms that blur the timing information are the randomness in service times and the queueing delays. The problem of obtaining the capacity of this channel is interesting because of the following challenges. The channel has memory due to the queueing of packets, and is nonstationary because the queue is initially empty. Furthermore, even for the single-server queue, simple queueing-theoretic results [9] are known only when the queue is in steady state and the input is amenable to steady-state analysis; to compute capacity we cannot restrict the encoder to such inputs.

The single-server queue is characterized by its service distribution. In queueing theory, the most tractable service distribution is the exponential distribution. In the usual convention, the exponential server is denoted by $\cdot/M/1$, and a single-server with "generic" distribution (independent and identically distributed service times) is denoted by $\cdot/G/1$. The following results on the capacity of the single-server queue are known.

- The capacity of the $\cdot/G/1$ queue with service rate μ packets per second is greater than or equal to $e^{-1}\mu$ nats per second [1].

- Among all $\cdot/G/1$ queues with given service rate, the $\cdot/M/1$ queue has the lowest capacity, equal to $e^{-1}\mu$ nats per second [1]. The exponential service therefore plays the same role that Gaussian noise plays in additive noise channels; it is therefore the "noisiest" service distribution.

- Consider a simple model of a telephone signaling channel where, upon placing the ith call, the transmitter listens until the first ring is heard, at which time the transmitter hangs up and then, after a period of time that depends on the message (and possibly on previous transit times), places the $(i+1)$st call. The sources of randomness in this model are the call transit times. If these are independent and identically distributed (i.i.d) with a certain distribution, this channel is equivalent to the $\cdot/G/1$ queue with complete feedback. The capacity of the telephone signaling channel is greater than or equal to $e^{-1}\mu$ nats per second, where $1/\mu$ seconds is the average transit time [1]. In this model the ability to send information in the number of rings is not exploited.

- The capacity of the $\cdot/M/1$ queue does not increase even if the encoder has complete feedback information of the output of the queue [1]. It does not decrease if there is an unknown number of packets in the queue initially [1].

- Let each packet contain C_0 nats of information. Packets now are no longer identical. We assume for simplicity that the packet information (C_0 nats per packet) is received noiselessly by the receiver. The capacity of the information bearing queue is [1]
$$C_I = \sup_{\lambda \leq \mu} \left[C(\lambda) + \lambda C_0 \right],$$
where μ is the service rate of the queue, and $C(\lambda)$ is the capacity of the queue at output rate λ. This equation reflects an inherent tradeoff; if we input packets to the queue at a rate λ very close to the service rate, then we destroy the information carried in the arrival times. If C_0 is sufficiently large, it may not

be worth sacrificing input rate in order to convey information via timing. For binary-valued packets, however, the capacity of the queue is equal to

$$\begin{aligned}\sup_{\lambda\le\mu}\left[\lambda\log\frac{\mu}{\lambda}+\lambda\log 2\right] &= 2e^{-1}\mu \text{ nats per second}\\ &= 1.0615\mu \text{ bits per second}\end{aligned}$$

for the exponential server. The capacity is larger for any other service distribution, i.e., we can transmit reliably at a rate strictly larger than μ bits per second.

- Analogous results hold for the discrete-time queue (cf. [4], [5]). Packets arrive and depart only at discrete time instants called slots. The geometric server with service time distribution $\Pr\{S = k \text{ slots}\} = \mu(1-\mu)^{k-1}$ for $k \geq 1$, where $0 < \mu < 1$, plays the role of the exponential server. The capacity of the exponential server queue is $\log\left[1+\mu(1-\mu)^{(1-\mu)/\mu}\right]$ nats per slot.

- The channel with point-process observations (cf. [6], [7], [8]) where the transmitter controls the instantaneous rate of the point process, is intimately related to the single-server queue. Such channels model the direct-detection optical communication channels [8]. Let the rate controlled by the transmitter be $\lambda = (\lambda_t : t \in [0,T])$, where $\lambda_t \in [0,\mu]$. The receiver observes an inhomogeneous Poisson process on $[0,T]$ with rate λ. The capacity of this channel is $e^{-1}\mu$ nats per second. Furthermore, the capacity of this channel does not increase in the presence of complete feedback.

1.3 Motivation

In deriving the aforementioned results, an assumption is that both the encoder and the decoder have full knowledge of the channel, i.e., the distribution of service times.

There are several situations where the decoder must be designed without this knowledge. The decoder may be used either in different settings with differing channel statistics, or in situations where unmodeled random mechanisms affect the channel. It is then of interest to address attainable rates of transmission on a given channel with a decoding rule that is robust to variations in the channel statistics.

Considerable attention has been given to the single-server queue. Little is known, however, for other queueing systems such as the multiserver queues, queues in tandem, or even the single-server queue with finite buffer. For such systems, the approaches of [1], [4] and [5] do not seem useful; using their techniques, even the likelihood function of the output given the input is difficult to obtain. New approaches are therefore required to understand timing channels in networks.

Attention in [1], [4] and [5] is focused on obtaining achievable rates disregarding the practical issues of complexity and computational feasibility. To build real-life communication systems, however, we need coding schemes where the decoder has good performance while being computationally feasible.

In this thesis, we address these issues. The next section gives a preview of our results and indicates how the thesis is organized.

1.4 Organization

In Chapter 2, we consider using a specific decoding criterion to transmit information by timing arrivals to a single-server queue. This criterion is the maximum-likelihood decoding rule for the exponential server queue. For any server with service times that are stationary and ergodic with mean $1/\mu$ seconds, we show that the rate $e^{-1}\mu$ nats per second (the capacity of the exponential server timing channel) is achievable using this decoder. We show that a similar result holds for the timing channel with feedback.

In some situations, it is necessary to hamper the communication capability over the timing channel. On the single-server queue, service times blur timing information. A natural means of hindering communication is to delay the exit of the packets (in addition to the nominal service times). On this jammed timing channel, we show that the rate $e^{-1}\mu$ nats per second is achievable with a random strategy, where the nominal service times are stationary and ergodic with mean $1/\mu_1$ seconds, the arithmetic mean of the delays added by the server does not exceed $1/\mu_2$ seconds, and $\mu = \mu_1\mu_2/(\mu_1 + \mu_2)$. In this random strategy, the encoder picks a codebook at random. The result of this random choice is known to the decoder, but not to the server. For the discrete-time jammed timing channel we show the existence of a nonrandom strategy that transmits reliably at $\log\left[1 + \mu(1-\mu)^{(1-\mu)/\mu}\right]$ nats per slot, where $\mu = \mu_1\mu_2/(\mu_1 + \mu_2)$, if the packets themselves carry information. We also show converses for exponential-server and geometric-server queues.

In Chapter 3, we prove the capacity formula of [1] for the exponential server queue from a point-process perspective. Our results imply an alternative strategy with feedback that achieves capacity on the point-process channel with no background intensity. Our results for the discrete-time queue imply an alternative strategy with feedback that achieves capacity on a discrete memoryless channel called the Z-channel. This point-process route suggests a procedure to analyze the capacity of simple queueing networks. Specifically, we consider multiserver queues, queues in tandem, and a single-server queue whose output is merged with a stream of Poisson arrivals. We provide bounds, some analytical and some obtained from simulations, on the capacities of these systems.

In Chapter 4, we address the question of finding good coding schemes that have computationally feasible decoding procedures. We show the existence of a good tree code with a sequential decoder for the exponential server timing channel. The expected number of node expansions per decoded bit is upperbounded by a constant.

The rate of information transfer for this code is $\mu/(2e)$ nats per second, i.e., one half of the capacity.

In Chapter 5, we collect a few open questions and indicate some future directions.

In the appendix, we discuss one particular open problem. On the single-server queue, it is possible to discard some codewords as incompatible with the observed sequence of departure times. Suppose that the receiver outputs a list of compatible codewords for the received sequence of departure times. We would like this list size to be 1, i.e., the transmitted message is recovered without any error. Under the constraint that the average list-size be close to 1, the largest information rate that can be supported is called the (zero-error) average list size capacity (cf. [10], [11]). There are strong indications that the (zero-error) average list size capacity of the single-server queue (without feedback) is 0.

Chapter 2

Robust Decoding for Timing Channels

2.1 Introduction

Consider the problem of transmitting information through the epochs at which packets arrive at a single-server queue [1]. All packets are identical and information is contained only in the times of arrival of these packets. The service times cause delays that corrupt the input information. Let us recall some results stated in Chapter 1. If the service times are independent and exponentially distributed with mean $1/\mu$ seconds, the capacity of this channel is $e^{-1}\mu$ nats per second when the cost of transmission is the expected time for the last packet to exit the system [1]. Furthermore, if the service times are independent and identically distributed (i.i.d) with mean $1/\mu$ seconds, but are not exponentially distributed, then we can communicate reliably at a rate $e^{-1}\mu$ nats per second [1]. Thus among all servers with i.i.d service times of mean $1/\mu$ seconds, the exponential server has the least capacity. These results in [1] assume that both the encoder and the decoder know the distribution of the service times.

When the service times are independent and exponentially distributed, maximum-likelihood decoding is easy to implement. Given the sequence of times at which packets depart from the queue, the decoder finds the codeword that explains the sequence of departures with the smallest sum of service times. To do this, the decoder needs only additions, subtractions and comparisons. Since the exponential server has the least capacity, and its maximum-likelihood decoder uses simple functions, we consider using this decoding strategy when the service times are not exponentially distributed. In this case, although the above decoder is suboptimal, its simplicity and general applicability are appealing.

In this chapter, we show that we can communicate reliably at a rate $e^{-1}\mu$ nats per second using the above decoding strategy when the distribution of service times, known to the encoder, is stationary and ergodic with mean $1/\mu$ seconds. In other words, the decoder need not know this distribution to achieve $e^{-1}\mu$ nats per second. A similar result is known to hold for the discrete-time additive noise channel; the Gaussian channel's capacity is achievable using the minimum Euclidean distance decoder (cf. [12, Theorem 1] for a version of the result). We describe this result in more detail later in this section.

Consider the following definition of the cost of transmission. Suppose that the decoder has to make decisions based only on departures that occur within a certain time window. If the cost of transmission is the length of the time window of observation, then we show that we can communicate reliably at $e^{-1}\mu$ nats per second. The service times are stationary and ergodic with mean $1/\mu$ seconds. Under this new definition of the cost of transmission, we also show that $e^{-1}\mu$ nats per second is the largest rate achievable on the exponential server channel. We do this by mapping any strategy on the timing channel to an equivalent strategy with complete feedback on the point-process channel [6].

Discrete-time queues were studied in [4] and [5]. The maximum-likelihood decoder for the server with independent and geometrically distributed service times is simple. We argue that using this decoder, the capacity of the geometric server channel is achievable when the distribution of service times is stationary and ergodic with mean $1/\mu$ slots. If the cost of transmission is the length of the observation window, then we show the converse for the geometric server channel by mapping any communication strategy on this timing channel to an equivalent strategy with complete feedback on a binary memoryless channel.

Timing information can be transmitted covertly by transmitting innocuous information in the contents of packets, which may be subject to eavesdropping. Since service times corrupt information encoded in the arrival epochs of packets, we consider the following *jamming strategy* employed by the server to hamper covert communication. Every packet suffers a delay (extra service time) in addition to the nominal service time (which is stationary and ergodic with mean $1/\mu_1$ seconds). If these delays are without limits, then communication in the timing channel can be jammed completely at the expense of information throughput in packet contents. We therefore require that the arithmetic mean of these delays be smaller than $1/\mu_2$ seconds. We call the resulting channel the *jammed timing channel.* This channel is similar to the arbitrarily varying channel (AVC) introduced in [13]. An important distinction between the jammed timing channel and the memoryless AVC ([13], [14], [15], [16] and references therein) is that in the jammed timing channel current input and delay can affect future outputs.

We prove an achievability result in the situation where the jammer does not know the true codebook in use, but knows only a distribution from which the codebook is selected. In particular, the rate $e^{-1}\mu_1\mu_2/(\mu_1+\mu_2)$ nats per second is achievable with random codes on the jammed timing channel. When the nominal service times are independent and exponentially distributed, we show that the rate $e^{-1}\mu_1\mu_2/(\mu_1+\mu_2)$ nats per second is also the largest achievable with random codes, giving us a reduction in capacity by a factor $\mu_2/(\mu_1+\mu_2)$.

We now briefly survey previous works relevant to our study. The use of the exponential server's maximum-likelihood decoder when the service times are not exponentially distributed is an instance of decoder mismatch. In the context of discrete memoryless channels (DMC), suppose that the communication system operates under a channel with transition probability matrix $W(\cdot|\cdot)$. The decoder performs maximum-likelihood decoding assuming that the DMC is characterized by $V(\cdot|\cdot)$, i.e., for a received sequence y^n, it chooses the codeword x^n that maximizes $V(y^n|x^n)$, where n is the number of uses of the channel. Reference [17] showed that using the mismatched decoder, we can communicate reliably at a rate

$$\sup_{P_X} \quad E\left[\log \frac{V(Y|X)}{Q_Y(Y)}\right], \tag{2.1}$$

where Q_Y is the marginal distribution of the output under the mismatch channel V and the input distribution P_X. The expectation in (2.1) is with respect to the joint distribution under the true channel W and the input distribution P_X. This result was extended to discrete channels with memory in [18]. Since these results have not been proved for channels with memory that have continuous inputs and outputs, we first show the achievability of (2.1) for such channels and then apply this result to the timing channel. The proof, though different from the proofs in [17] and [18], is a simple extension of the proof of [19, Lemma 6.9].

Although rates possibly larger than (2.1) are achievable with mismatched decoding ([20], [21], [22] and references therein), achievability of a rate that is analogous to (2.1) is enough to show the results in this chapter.

This chapter extends the parallelism found in [1] between the exponential server timing channel and the discrete-time additive white Gaussian noise channel with an input power constraint. Consider the additive noise channel. For n uses of the channel, each codeword is a point in $\mathcal{R}^n$ having power smaller than nP. It is well-known that for any stationary, ergodic, zero-mean noise process with variance σ^2, the rate $(1/2)\log[1 + P/\sigma^2]$ nats per channel use is achievable using the minimum Euclidean distance criterion for decoding. A version of this result is the direct part of [12, Theorem 1]. A stronger version of the direct part when $\sigma^2 < P$ is given in [16]. The minimum Euclidean distance criterion for decoding is the maximum-likelihood decoding when the noise is independent and has the Gaussian distribution; the capacity in this case is $(1/2)\log[1 + P/\sigma^2]$ nats per channel use. The timing channel counterparts of this result are Propositions 1 and 2 in Section 2.2. As in [1], the analogy is rooted in the fact that the exponential distribution and the Gaussian distribution are similar mutual information saddle-points [23].

A similar result is known for a convex and compact family Θ of DMCs. For an input distribution P and a DMC W, let $I(P, W)$ denote the mutual information. Let P^* and $W^* \in \Theta$ attain the saddle-point of the mutual information functional, i.e.,

$$\max_{P} \min_{W \in \Theta} I(P, W) = \min_{W \in \Theta} \max_{P} I(P, W) = I(P^*, W^*).$$

Suppose now that the channel is characterized by $W \in \Theta$. Then $I(P^*, W^*)$ is achievable over the DMC W using a maximum-likelihood decoder for the DMC with stochastic matrix W^* [24] (see also [25, Section IV-B-4]).

The jammed timing channel is similar in spirit to the Gaussian arbitrarily varying channel (Gaussian AVC) [26], [16], in which a jammer changes the mean of the Gaussian noise subject to a power constraint. Proposition 3 in Section 2.2 is related to results in [26] for random codes in the Gaussian AVC. The capacity of the Gaussian AVC, when the jammer knows the codebook, is known [16]. We do not know if an analogous result holds on the jammed timing channel, when the jammer knows the codebook. In the discrete-time case however we can apply the "elimination" technique of [14] to get a non-random coding strategy if a certain amount of information can be transmitted by the packet contents. Only a negligible fraction of packet contents need be used.

The jammed channel considered in [27] imposes a constraint on the overall delay suffered by the packets. Let the packet be input into the system at time t and let it depart from the system at time $t+d$, where d includes the queueing delay and the service time. Then d is the overall delay suffered by the packet. Let us further assume that more than one arrival and more than one departure can occur at an instant. Let P be the distribution of the arrival process such that the arrival rate is λ. Let Q be the channel such that the maximum of overall delays suffered by the packets does not exceed D. Let $I(P,Q)$ denote the mutual information. Then [27]

$$\max_P \min_Q I(P,Q) = \min_Q \max_P I(P,Q) = \frac{1}{D} H(P^*),$$

where P^* is the geometric distribution on $\mathcal{Z}_+ = \{0,1,2,\cdots\}$ having mean λD. Bounds on the minimax and maximin of the mutual information were given for the constraint where the average overall delay is constrained to be below D. In the model of the jammed timing channel studied in this thesis, the constraint is on the service times of the packets. The model in [27] is therefore significantly different from the jammed timing channel.

The rest of this chapter is organized as follows. Section 2.2 states the basic definitions and results. Section 2.2.1 covers the mismatched decoding problems for the continuous-time single-server queue. Section 2.2.2 studies the jammed timing channel. Section 2.2.3 discusses the signaling channel, or the timing channel with feedback. Section 2.2.4 describes the discrete-time single-server queue. Section 2.2.5 shows the converses for the exponential and the geometric server channels. Section 2.2.6 collects several observations on our results. The proofs are in Section 2.3.

2.2 Definitions and Results

2.2.1 Continuous-Time Single-Server Queue

This subsection deals with mismatched decoding for the continuous-time single-server queueing system without feedback. The definitions of the relevant quantities are as in [1], but written in our notation. Let $\mathcal{R}_+ = [0,\infty), \mathcal{Z}_+ = \{0,1,\cdots\}$ and $\mathcal{N} = \{1,2,\cdots\}$. We assume that the following conditions hold:

- The queue is work-conserving, i.e., if a packet departs after service and another one is in the queue, then the server begins to serve the packet in the queue;
- The queue is initially empty, and the server follows a first-in-first-out service discipline;
- The sequence $(S_k : k \in \mathcal{N})$ of service times is a stationary and ergodic process with mean $1/\mu$ seconds.

For each $n \in \mathcal{N}$, the input to the queueing system is a vector $x^n = (x_1, \cdots, x_n)$ of n nonnegative interarrival times, such that the kth arrival occurs at time $\sum_{i=1}^k x_i,\ k = 1, \cdots, n$. The decoder observes $y^n = (y_0, y_1, \cdots, y_n)$, where $y_0 = 0$, and y_k is the time between the $(k-1)$st and the kth departures, $k = 1, \cdots, n$.

For each $n \in \mathcal{N}$, the input alphabet is $\mathcal{R}_+^n$, and the output alphabet is $\mathcal{R}_+^{n+1}$. The σ-algebras associated with the alphabets are the product Borel σ-algebras. Let $E \subset \mathcal{R}_+^{n+1}$ be a Borel set and $x^n \in \mathcal{R}_+^n$. A *transition probability function* [28, p.315], $P_{Y^n|X^n}$, from the input space to the output space, is a mapping $(x^n, E) \to P_{Y^n|X^n}(E \mid x^n)$ having the following measurability properties: (a) for each $x^n \in \mathcal{R}_+^n$, the mapping $E \to P_{Y^n|X^n}(E \mid x^n)$ is a probability measure on the output space; (b) for each Borel set $E \subset \mathcal{R}_+^{n+1}$, the mapping $x^n \to P_{Y^n|X^n}(E \mid x^n)$ is measurable with respect to the input space. A *channel* is a sequence (parametrized by n) of transition probability functions from the input space to the output space.

Fix $n \in \mathcal{N}$. Let s_k be the service time of the kth packet, $k = 1, \cdots, n$. The observable y^n can be described as follows. Let w_k be the amount of time for which the server is idle between the $(k-1)$st departure and the kth arrival, i.e.,

$$w_k = \max\left\{0, \ \sum_{i=1}^{k} x_i - \sum_{i=0}^{k-1} y_i\right\}, \quad k = 1, \cdots, n. \tag{2.2}$$

Thus if the kth arrival occurs before the $(k-1)$st departure, the idling time w_k is 0. The interdeparture times are then given by

$$y_k = \begin{cases} 0, & k = 0, \\ w_k + s_k, & k = 1, \cdots, n. \end{cases} \tag{2.3}$$

The stationary and ergodic process $(S_k : k \in \mathcal{N})$ and the queue equations (2.2) and (2.3) induce the true channel $\left(P_{Y^n|X^n} : n \in \mathcal{N}\right)$, which is a sequence of transition probability functions from the input space to the output space.

Definition 1: *An* (n, M, T, ε)-code *consists of a codebook of* M *codewords and a decoder. Each codeword is a vector of* n *nonnegative interarrival times* $(x_1, \cdots, x_n)$. *The decoder, after observing the* n *departures, selects the correct codeword with probability greater than* $1 - \varepsilon$, *under equiprobable codewords and* $P_{Y^n|X^n}$. *The* n*th departure occurs on the average (under equiprobable codewords and* $P_{Y^n|X^n}$*) no later than* T. *The rate of the code is* $(\log M)/T$. *Rate* R *is* achievable *if, for every* $\gamma > 0$, *there exists a sequence of* $(n, M_n, T_n, \varepsilon_n)$*-codes that satisfies* $(\log M_n)/T_n > R - \gamma$ *for all sufficiently large* n, *and* $\lim_{n\to\infty} \varepsilon_n = 0$.

We now describe the mismatch channel $\left(Q_{Y^n|X^n} : n \in \mathcal{N}\right)$ according to which the decoder performs maximum-likelihood decoding. Let a synchronizing zeroth packet be sent at $t = 0$ and interpret y_0 as the amount of unfinished work at $t = 0$, including the service time of the zeroth packet (i.e., the time at which the zeroth packet departs from the system). Let the number of packets in the queue at $t = 0$ have the equilibrium distribution that is associated with an $M/M/1$ queue [9, pp. 48-49] having input rate $\lambda' < \mu$ packets per second. The mismatch channel is then the channel induced by the process $(S_k : k \in \mathcal{N})$ that is independent and exponentially distributed with mean $1/\mu$ seconds. It will soon be clear that the decoding strategy does not depend on the parameter λ'.

Let $e_\mu(x)$ denote the exponential density function $\mu e^{-\mu x}, x \in \mathcal{R}_+$, having mean $1/\mu$. The random variable Y_0 has the exponential density $e_{\mu-\lambda'}(y_0)$ under $Q_{Y^n|X^n}$ [1], for every $n \in \mathcal{N}$. In contrast, $Y_0 = 0$ under $P_{Y^n|X^n}$, for every $n \in \mathcal{N}$.

Let π denote the Lebesgue measure (the argument will indicate the appropriate space). Fix $n \in \mathcal{N}$. Using the queue equations (2.2) and (2.3), and the density for exponentially distributed service times, $Q_{Y^n|X^n}$ can be written as

$$dQ_{Y^n|X^n}(y^n|x^n) = d\pi(y^n)\ p(x^n, y^n) \tag{2.4}$$

for every $x^n \in \mathcal{R}_+^n$, where

$$p(x^n, y^n) \triangleq e_{\mu-\lambda'}(y_0) \prod_{k=1}^{n} e_\mu(y_k - w_k), \quad \lambda' < \mu. \tag{2.5}$$

Let the distribution P_{X^n} on the input space be given by

$$dP_{X^n}(x^n) = d\pi(x^n) \prod_{k=1}^{n} e_{\lambda'}(x_k), \quad \lambda' < \mu. \tag{2.6}$$

This is the distribution of the first n arrivals induced by the Poisson arrival process with rate λ'. Let Q_{X^n,Y^n} denote the joint distribution under the input distribution P_{X^n} (cf. (2.6)) and $Q_{Y^n|X^n}$ (cf. (2.4)). The joint distribution Q_{X^n,Y^n} can then be written unambiguously as

$$dQ_{X^n,Y^n}(x^n, y^n) = \ dP_{X^n}(x^n)\ d\pi(y^n)\ p(x^n, y^n), \tag{2.7}$$

due to Fubini's Theorem [29, Theorem 18.3, p.238]. Let Q_{Y^n} denote the marginal distribution of $Y^n = (Y_0, \cdots, Y_n)$ under Q_{X^n,Y^n}. Let $P_{X^n} \times Q_{Y^n}$ denote the joint distribution under which the random variables X^n and Y^n are independent, and have marginal distributions P_{X^n} and Q_{Y^n}, respectively.

As a consequence of (2.7), we have that $Q_{X^n,Y^n} \ll P_{X^n} \times Q_{Y^n}$ [30, Corollary 5.3.1, p.112], and that $Q_{Y^n} \ll \pi_{Y^n}$. A version of the Radon-Nikodym derivative $dQ_{X^n,Y^n}/d(P_{X^n} \times Q_{Y^n})$ is the function f given by

$$f(x^n, y^n) = \begin{cases} p(x^n, y^n)/m(y^n), & \text{if } \ m(y^n) > 0 \\ \qquad 1, & \text{if } \ m(y^n) = 0 \end{cases} \tag{2.8}$$

where

$$m(y^n) \triangleq \int_{\mathcal{R}_+^n} dP_{X^n}(x^n)\ p(x^n, y^n), \quad y^n \in \mathcal{R}_+^{n+1}. \tag{2.9}$$

We can easily verify that

$$dQ_{Y^n}(y^n) = d\pi(y^n)\ m(y^n). \tag{2.10}$$

Clearly, the function f (cf. (2.8)) satisfies

$$\int_{\mathcal{R}_+^n} dP_{X^n}(x^n) f(x^n, y^n) = Ef(X^n, y^n) = 1 \tag{2.11}$$

for every $y^n \in \mathcal{R}_+^{n+1}$. The output of an $M/M/1$ system with input rate $\lambda' < \mu$ is a Poisson process with rate λ' (see, for e.g., [9, Fact 2.8.2, p.60]). Consequently, under Q_{X^n,Y^n}, the random vector $(Y_1, \cdots, Y_n)$ is a vector of independent and exponentially distributed random variables with mean $1/\lambda'$ seconds, i.e.,

$$dQ_{Y_1,\cdots,Y_n}(y_1, \cdots, y_n) = d\pi(y_1, \cdots, y_n) \prod_{k=1}^{n} e_{\lambda'}(y_k). \tag{2.12}$$

We will use (2.10), (2.11) and (2.12) in the proof of our results.

We now describe the mismatched decoder. The decoder makes a decision based on f (cf. (2.8)) as follows. Let the codebook be $\{\mathbf{x}_1, \cdots, \mathbf{x}_M\}$, where $\mathbf{x}_i \in \mathcal{R}_+^n$ for $i = 1, \cdots, M$. The decoder $\phi_f : \mathcal{R}_+^{n+1} \to \{0, 1, \cdots, M\}$ maps the observed interdeparture times y^n to

$$\phi_f(y^n) \triangleq \begin{cases} i, & \text{if } \max_{j \neq i} f(\mathbf{x}_j, y^n) < f(\mathbf{x}_i, y^n) \\ 0, & \text{if no such } i \text{ exists.} \end{cases} \tag{2.13}$$

We interpret the output 0 as an error in decoding.

From Lemma 2 in Section 2.3, $m(y^n) = 0$ if and only if $y_0 + y_1 = 0$, in which case $\phi_f(y^n) = 0$. But $m(Y^n) = 0$ with zero probability under Q_{Y^n}. When $m(y^n) > 0$, which is the case almost always, the decoder ϕ_f tries to pick the unique codeword that maximizes $p(\cdot, y^n)$ (cf. (2.5)). This is the same as picking the unique codeword (among the compatible ones) that minimizes the sum of service times, $\sum_{k=1}^{n}(y_k - w_k)$, or equivalently maximizes the sum of idling times of the server, $\sum_{k=1}^{n} w_k$. When the decoder cannot find such a unique codeword, it declares 0, an error. The only functions required to make this decision are additions, subtractions and comparisons. Although these functions are simple, $\sum_{k=1}^{n} w_k$ must be evaluated for every codeword before a decision is made. Since the number of codewords is exponential in time, the number of operations performed to decode is exponential in time.

Proposition 1: *Let the queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. The rate $e^{-1}\mu$ nats per second is then achievable using the decoding rule in (2.13).*

The result [1, Theorem 7] on the achievability of $e^{-1}\mu$ nats per second for i.i.d service times is a special case of Proposition 1. To transmit reliably at $e^{-1}\mu$ nats per second on such a channel, maximum-likelihood decoding is not required; the decoder ϕ_f is sufficient. This decoder is therefore robust to the distribution of service times. The decoder's robustness, however, does not imply that a single sequence of codes works for all stationary and ergodic distributions of the service times. Furthermore, Proposition 1 does not give rates at which the probability of error goes to zero. The term "robust" should therefore be interpreted with caution. We only argue that, knowing the true channel, a sequence of good codes with decoder ϕ_f and rate close to $e^{-1}\mu$ nats per second can be selected.

Suppose that the codebook is such that for every codeword, the last arrival occurs before $t = T'$. Then the decoder, ϕ_f, need not observe departures beyond $t = T'$. This is because of the following. Suppose that y^n satisfies $\sum_{k=0}^{n} y_k > T'$. Given any candidate codeword, the server is not idle beyond T', i.e., the quantity that is required to make a decision, $\sum_{k=1}^{n} w_k$, can be evaluated upon observation of departures in the time window $[0, T']$. Departures in $[0, T']$ therefore constitute a set of sufficient statistics for determining the input codeword. This is not surprising because of the memoryless property of exponential service times.

Now suppose that the decoder observes only the departures that occur in $[0, T]$, where T is known to the encoder. Clearly, it is useless to have arrivals after time T. This motivates the following definition.

Definition 2: *An* (n, M, T, ε)-window-code *consists of a codebook of* M *codewords and a decoder. Each codeword is a vector* $(x_1, \cdots, x_n)$ *of* n *nonnegative interarrival times. The* n*th arrival of every codeword occurs before time* T*. The decoder, after observing departures in* $[0, T]$*, selects the correct codeword with probability greater than* $1 - \varepsilon$*, under equiprobable codewords and* $P_{Y^n|X^n}$*. The rate of the window-code is* $(\log M)/T$*. Rate* R *is* achievable with window-codes *if, for every* $\gamma > 0$*, there exists a sequence of* $(n, M_n, T_n, \varepsilon_n)$*-window-codes that satisfies* $(\log M_n)/T_n > R - \gamma$ *for all sufficiently large* n*, and* $\lim_{n\to\infty} \varepsilon_n = 0$*.*

Thus the term "window-code" refers to a code whose decoder observes only those departures that occur in the window $[0, T]$; the cost of transmission for the window-code is T seconds. In contrast, the code in Definition 1 has a decoder that observes all the n departures; the cost of transmission in that case is the expected time of the nth departure, i.e., the expected time the decoder waits to gather the output data.

Proposition 2: *Let the queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times* *of mean $1/\mu$ seconds. The rate $e^{-1}\mu$ nats per second is then achievable with window-codes using the decoding rule in (2.13).*

2.2.2 Jammed Timing Channel: Random Codes

We now consider the jammed timing channel where the server acts as an adversary. The queue is initially empty, and the server follows a first-in-first-out service discipline. The process $(S_k : k \in \mathcal{N})$ of nominal service times is stationary and ergodic with mean $1/\mu_1$ seconds. Fix $n \in \mathcal{N}$. The server (jammer) includes a delay of z_k seconds to the service time of the kth packet, $k = 1, \cdots, n$. We call $\mathbf{z} = (z_1, \cdots, z_n) \in \mathcal{R}_+^n$ the *state sequence*, since it determines the state of the channel. The resulting service time for the kth packet is $S_k + z_k$ seconds, $k = 1, \cdots, n$. If no constraints are imposed on the state sequence, communication in the timing channel can be jammed completely at the expense of information throughput in packet contents. We impose the following constraint. For a code with n packets, we allow only those state sequences $\mathbf{z}$ that satisfy

$$l(\mathbf{z}) \triangleq \frac{1}{n} \sum_{k=1}^{n} z_k \leq \frac{1}{\mu_2},$$

i.e., a total delay of at most n/μ_2 seconds is allowed for all the n packets. Each state sequence, $\mathbf{z}$, induces a transition probability function from the input space to the output space, denoted by $W^n(\mathbf{z})$. We need communication strategies that perform well for every state sequence $\mathbf{z}$ that satisfies $l(\mathbf{z}) \leq 1/\mu_2$.

The problem of finding achievable rates for deterministic codes, i.e., when the codebook is known to the jammer, appears to be nontrivial. Instead of fixing a single good (deterministic) codebook, we allow communication strategies with random codes. The encoder chooses the codebook that is used for transmission from a set of possible codebooks. The decoder knows this selection. The jammer, however, is ignorant of the selected codebook. Its partial knowledge is modeled by a distribution on the set of codebooks. Such a code is usually called in the AVC literature, somewhat deceptively, a *random code.*

Given a selected codebook $\mathbf{c}$, the decoder is ϕ_f (cf. (2.13) and (2.8)). For the codebook $\mathbf{c}$, the average probability of error (over equiprobable codewords) is denoted by $P_e(\mathbf{c}, \phi_f, W^n(\mathbf{z}))$ when the state sequence is $\mathbf{z}$.

Let $\mathbf{C}$ be a random variable taking values in the family of all codebooks that have M codewords, and such that the nth arrival in each codeword occurs before T. The parameters (n, M, T) of the random variable $\mathbf{C}$ will be clear from the context. The following definition is an extension of window-codes and achievability with window-codes (Definition 2) for the jammed timing channel.

Definition 3: *An* (n, M, T, ε)-random window-code *consists of a probability distribution for* $\mathbf{C}$*, and a decoder* ϕ *that depends on the codebook realization. Each realization* $\mathbf{c}$ *is a set of* M *codewords. Each codeword is a vector of* n *nonnegative interarrival times. The* n*th arrival of every codeword occurs before time* T*. The decoder, knowing the codebook realization* $\mathbf{c}$*, makes a decision after observing departures in* $[0, T]$*. The average probability of error satisfies* $E[P_e(\mathbf{C}, \phi, W^n(\mathbf{z}))] \leq \varepsilon$ *for every* $\mathbf{z}$ *with* $l(\mathbf{z}) \leq 1/\mu_2$*. Rate* R *is* achievable with random window-codes *if, for every* $\gamma > 0$*, there exists a sequence of* $(n, M_n, T_n, \varepsilon_n)$*-random window-codes that satisfies* $(\log M_n)/T_n > R - \gamma$ *for all sufficiently large* n*, and* $\lim_{n\to\infty} \varepsilon_n = 0$*.*

Proposition 3: *On the jammed timing channel, the rate* $e^{-1}\mu_1\mu_2/(\mu_1 + \mu_2)$ *nats per second is achievable with random window-codes using the decoding rule in (2.13)*

2.2.3 Signaling Channel

In a telephone signaling channel [1], the encoder knows exactly when the packet is removed from service, i.e., complete feedback is available. The encoder can make use of this information to avoid queueing altogether, and the resulting channel is a simple additive noise channel.

On this channel, the rate $e^{-1}\mu$ nats per second is clearly achievable in the presence of complete feedback. Indeed, the encoder can ignore feedback completely, and use a code suggested by Proposition 1. Making use of feedback, however, leads to another decoder that achieves $e^{-1}\mu$ nats per second. It will be clear from the following definition that feedback is used only to avoid queueing.

The sequence of service times is stationary and ergodic with mean $1/\mu$ seconds. An (n, M, T, ε)-*feedback code* consists of codebook of M codewords and a decoder. Each message is an n-vector $(x_1, \cdots, x_n)$ of positive real numbers. The first arrival occurs at $t = x_1$. The kth component, x_k, is the amount of time the encoder will wait after the $(k-1)$st departure, before sending the kth arrival, $k = 2, \cdots, n$. The last packet exits on the average before T. The encoder thus makes use of feedback to avoid queueing and to control completely the idling times of the server. Feedback however is not used to choose the waiting times x_k. The rate of the feedback code is $(\log M)/T$ nats per second. The decoder, after observing the interdeparture times $y^n = (y_1, \cdots, y_n)$, makes the correct decision with probability larger than $1 - \varepsilon$ when averaged over equiprobable codewords.

Rate R is *achievable with feedback* if, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-feedback codes that satisfies $(\log M_n)/T_n > R - \gamma$ for all sufficiently large n, and $\lim_{n\to\infty} \varepsilon_n = 0$.

The decoder chooses the codeword that explains the received sequence of departures with the minimum sum of service times. For a candidate codeword, $\mathbf{x} = (x_1, \cdots, x_n)$, let $d(\mathbf{x}, y^n) \triangleq \sum_{k=1}^n d'(y_k - x_k)$, where

$$d'(r) \triangleq \begin{cases} r, & \text{if } r \geq 0, \\ +\infty, & \text{if } r < 0. \end{cases}$$

Given the codebook $\{\mathbf{x}_1, \cdots, \mathbf{x}_M\}$, the decoder φ_d maps y^n to

$$\varphi_d(y^n) \triangleq \begin{cases} i, & \text{if } d(\mathbf{x}_i, y^n) < \min_{j \neq i} d(\mathbf{x}_j, y^n) \\ 0, & \text{if no such } i \text{ exists,} \end{cases} \tag{2.14}$$

where an output 0 is interpreted as an error in decoding.

Proposition 4: *Let the queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. The rate $e^{-1}\mu$ nats per second is then achievable with feedback using the decoding rule in (2.14).*

We remark that Proposition 4 is not implied by previous results on mismatched decoding. In particular, we cannot apply [20, Theorem 2] because it precludes input distributions with infinite differential entropy. For the input distribution that we choose, differential entropy does not exist. This input distribution is the one that attains the mutual information saddle-point [23, Theorem 1].

2.2.4 Discrete-Time Single-Server Queue

In this subsection, we describe the discrete-time single-server queueing system [4], [5], and state the counterparts of Propositions 1, 2 and 3. The proofs are omitted since they are analogous to the continuous-time case.

In the discrete-time model, arrivals and departures occur only at integer-valued epochs, called slots. At most one packet arrives and at most one packet departs in each slot. Unserved packets are stored in a queue. The queue is work-conserving, and the server follows a first-in-first-out service discipline. The sequence $(S_k : k \in \mathcal{N})$ of nominal service times is an $\mathcal{N}$-valued, stationary and ergodic process with mean $1/\mu_1$ slots, $0 < \mu_1 < 1$. Each packet requires at least one slot of service.

For each $n \in \mathcal{N}$, the input is a vector of n interarrival times, $x^n = (x_1, \cdots, x_n)$. The decoder observes $y^n = (y_0, y_1, \cdots, y_n)$, where $y_0 = 1$, and y_k is the time (in slots) between the $(k-1)$st and the kth departures, $k = 1, \cdots, n$. We set $y_0 = 1$ because $x_1 \geq 1$ and $s_1 \geq 1$, i.e., the first slot does not give any information about the transmitted message. The input alphabet is $\mathcal{N}^n$ and the output alphabet is $\mathcal{N}^{n+1}$. The σ-algebras associated with the alphabets are the collection of all the corresponding subsets.

Fix $n \in \mathcal{N}$. Given the sequence of service times $(s_1, \cdots, s_n) \in \mathcal{N}^n$, the interdeparture times in $y^n = (y_0, y_1, \cdots, y_n)$ are

$$y_k = \begin{cases} 1, & k = 0, \\ s_k + w_k, & k = 1, \cdots, n, \end{cases} \tag{2.15}$$

where $w_k = \max\left\{0, \sum_{j=1}^{k} x_j - \sum_{j=0}^{k-1} y_j\right\}$ is the server's idling time before serving the kth packet. The stationary and ergodic process $(S_k : k \in \mathcal{N})$ and the queue equation (2.15) induce the true channel $\left(P_{Y^n|X^n} : n \in \mathcal{N}\right)$.

The definitions of (n, M, T, ε)-code, achievability, (n, M, T, ε)-window-code and achievability with window-codes are analogous to those in Definitions 1-2.

Fix $n \in \mathcal{N}$. For the jammed timing channel (discrete-time), the state sequence $\mathbf{z} = (z_1, \cdots, z_n) \in \mathcal{Z}_+^n$ satisfies the constraint $l(\mathbf{z}) \stackrel{\Delta}{=} \left(\sum_{k=1}^{n} z_k\right)/n \leq 1/\mu_2$, $\mu_2 > 0$. As in the continuous-time case, each $\mathbf{z}$ induces a transition probability function, $W^n(\mathbf{z})$, from the input space to the output space.

The definitions of (n, M, T, ε)-random window-code and achievability with random window-codes are analogous to those in Definition 3.

We now describe the mismatch channel $\left(Q_{Y^n|X^n} : n \in \mathcal{N}\right)$ based on which the decoder performs maximum-likelihood decoding. We say that a random variable X has the Geo$^+(\lambda)$ distribution, $0 < \lambda < 1$, if $P\{X = x\} = g_\lambda(x) \stackrel{\Delta}{=} \lambda(1-\lambda)^{x-1}$, $x \in \mathcal{N}$. Let a synchronizing zeroth packet be sent at $t = 0$ and interpret y_0 as the amount of unfinished work at $t = 0$, including the service time of the zeroth packet. Let the number of packets in the queue at $t = 0$ have the equilibrium distribution that is associated with the queue having Geo$^+(\lambda')$-distributed arrivals, $0 < \lambda' < \mu_1 < 1$, and $\text{Geo}^+(\mu_1)$-distributed service times. This queueing system is the discrete-time counterpart of the $M/M/1$ system. The mismatch channel is then the channel induced by the process $(S_k : k \in \mathcal{N})$ of independent and $\text{Geo}^+(\mu_1)$-distributed service times. Fix $n \in \mathcal{N}$. Using the queue equation (2.15), we see that the mismatch transition probability function $Q_{Y^n|X^n}$ is the probability mass function (pmf) on $\mathcal{N}^{n+1}$ given by

$$Q_{Y^n|X^n}(y^n|x^n) = g_{\mu-\lambda'}(y_0) \prod_{k=1}^{n} g_\mu(y_k - w_k)$$

for every $x^n \in \mathcal{N}^n$. Let the pmf on the input alphabet $\mathcal{N}^n$ be

$$P_{X^n}(x^n) = \prod_{k=1}^{n} g_{\lambda'}(x_k), \quad 0 < \lambda' < \mu_1 < 1.$$

For $(x^n, y^n) \in \mathcal{N}^n \times \mathcal{N}^{n+1}$, let $f(x^n, y^n) \triangleq Q_{Y^n|X^n}(y^n|x^n)/m(y^n)$, where

$$m(y^n) \triangleq \sum_{x^n \in \mathcal{N}^n} P_{X^n}(x^n) \cdot Q_{Y^n|X^n}(y^n|x^n), \quad y^n \in \mathcal{N}^{n+1}.$$

The function f satisfies $Ef(X^n, y^n) = 1$ for every $y^n \in \mathcal{N}^{n+1}$; the expectation is with respect to P_{X^n}. Given a codebook with M codewords, the decoder is the function $\phi_f : \mathcal{N}^{n+1} \to \{0, 1, \cdots, M\}$ defined in (2.13).

Proposition 5: *Let the discrete-time queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic nominal service times of mean $1/\mu_1$ slots. The following statements then hold:*

(*a*) *The rate* $\log[1 + \mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}]$ *nats per slot is achievable using the (discrete-time) decoding rule in (2.13),*

(*b*) *The rate* $\log[1 + \mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}]$ *nats per slot is achievable with window-codes using the decoding rule in (2.13),*

(*c*) *On the jammed timing channel, the rate* $\log[1 + \mu(1-\mu)^{(1-\mu)/\mu}]$ *nats per slot, where* $\mu = \mu_1\mu_2/(\mu_1 + \mu_2)$, *is achievable with random window-codes using the decoding rule in (2.13).*

On the discrete-time jammed timing channel, if each packet carries a non-zero amount of information, we can apply the *elimination* technique of [14] to get a non-random communication strategy that has rate $\log[1+\mu(1-\mu)^{(1-\mu)/\mu}]$ nats per slot. The first step in this technique is *random code reduction* [19, Lemma 6.8], which we now describe.

Given a codebook $\mathbf{c}$ and the decoder ϕ_f, let $P_{e,i}(\mathbf{c}, \phi_f, W^n(\mathbf{z}))$ be the probability of error when the state sequence is $\mathbf{z}$ and the transmitted message is i, where $1 \leq i \leq M$. In the rest of this subsection, let the definition of (n, M, T, ε)-random window-code be analogous to that in Definitions 3 with the condition on the average probability of error replaced by

$$\max_{1\leq i\leq M_n} E\left[P_{e,i}(\mathbf{C}, \phi_f, W^n(\mathbf{z}))\right] \leq \varepsilon,$$

i.e., a condition on the maximum probability of error.

We can easily modify the proof to show the following extension to Proposition 5(*c*). On the jammed timing channel, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-random window-codes that satisfies

(i) $(\log M_n)/T_n > \log[1+\mu(1-\mu)^{(1-\mu)/\mu}] - \gamma$ for all sufficiently large n,

(ii) $\max_{\mathbf{z}\,:\,l(\mathbf{z})\leq 1/\mu_2} \max_{1\leq i\leq M_n} E\left[P_{e,i}(\mathbf{C}, \phi_f, W^n(\mathbf{z}))\right] \leq \varepsilon_n$ for every $n \in \mathcal{N}$, and

(iii) $\lim_{n\to\infty} \varepsilon_n = 0$.

Now suppose that the error probabilities need not vanish. Fix $n \in \mathcal{N}$. On the discrete-time channel, the cardinality of $\{W^n(\mathbf{z}) : l(\mathbf{z}) \leq 1/\mu_2\}$ is upperbounded by $(1+n/\mu_2)^n$. We can therefore apply random code reduction [19, Lemma 6.8] to get the following. Given $\varepsilon > 0$ and $\gamma > 0$, for all sufficiently large n, we can find a set of n^2 codebooks $\{\mathbf{c}_j : j = 1, \cdots, n^2\}$, where each codebook has parameters (n, M_n, T_n), the set of codebooks satisfies

$$\max_{\mathbf{z}\,:\,l(\mathbf{z})\leq 1/\mu_2} \max_{1\leq i\leq M_n} \frac{1}{n^2}\sum_{j=1}^{n^2} P_{e,i}(\mathbf{c}_j, \phi_f, W^n(\mathbf{z})) < \varepsilon, \tag{2.16}$$

and $(\log M_n)/T_n > \log[1 + \mu(1-\mu)^{(1-\mu)/\mu}] - \gamma$.

If each packet carries C_0 nats of information, $C_0 > 0$, then we can employ the elimination technique of [14] as follows. Fix $n \in \mathcal{N}$. Let the set of equiprobable messages be $\{1, \cdots, n^2\} \times \{1, \cdots, M_n\}$. Given a message (j, i) from this set, choose the codebook $\mathbf{c}_j$. Transmit i on the jammed timing channel using codebook $\mathbf{c}_j$. Convey the codebook index j to the receiver using the first $2 \log n$ nats (of the nC_0 nats) of packet contents. We thus use only a negligible fraction, $(2 \log n)/(nC_0)$, of the packet contents. The *average* probability of error over equiprobable codewords is smaller than ε for this (non-random) communication strategy because of (2.16).

2.2.5 Converses

In this subsection we state converse results for the continuous-time (resp. discrete-time) queueing system with independent and exponentially (resp. geometrically) distributed service times. Converses to Propositions 1, 4 and 5(a) were shown in [1] and [4].

Proposition 6: *For the continuous-time system, let the queue be work-conserving and initially empty. Furthermore, let the nominal service times be independent and exponentially distributed with mean $1/\mu_1$ seconds.*

(a) *The largest achievable rate with window-codes is $e^{-1}\mu_1$ nats per second.*

(b) *On the jammed timing channel, the largest achievable rate with random window-codes is $e^{-1}\mu_1\mu_2/(\mu_1 + \mu_2)$ nats per second.*

Similarly, for the discrete-time system, let the nominal service times be independent and Geo$^+(\mu_1)$-distributed with mean $1/\mu_1$ slots.

(c) *The largest achievable rate with window-codes is $\log[1 + \mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}]$ nats per slot.*

(*d*) *On the jammed timing channel, the largest achievable rate with random window-codes is* $\log[1+\mu(1-\mu)^{(1-\mu)/\mu}]$ *nats per slot, where* $\mu = \mu_1\mu_2/(\mu_1+\mu_2)$.

Proofs of (a) and (b) are in Section 2.3. The key idea in the proof of (a) is to map any window-code on the timing channel to an equivalent strategy with complete feedback on the point-process channel. We now prove (c). We omit the proof of (d) since it is analogous to the proof of (b).

Proof of (c): The service times are independent and have the $\text{Geo}^+(\mu_1)$ distribution. The key idea here is to map any window-code on the timing channel to a strategy with complete feedback on a binary memoryless channel.

Fix $n \in \mathcal{N}$. Suppose that the codebook is designed to transmit M_n messages. Each message maps to a codeword $(x_1, \cdots, x_n) \in \mathcal{N}^n$ of interarrival times that satisfies $\sum_{k=1}^n x_k < T_n$. The decoder observes departures until slot T_n. Let $1\{\cdot\}$ denote the indicator function of an event. Let the output be the binary-valued vector $(v_1, \cdots, v_{T_n})$ given by

$$v_k = 1\{\text{ Departure in slot } k\}, \ k = 1, \cdots, T_n. \tag{2.17}$$

Clearly, $v_1 = 0$ since $x_1 \geq 1$ and each packet requires at least one slot of service.

Fix a codeword $(x_1, \cdots, x_n)$. Let $(A_k : 1 \leq k < T_n)$ be the cumulative arrival process;

$$A_k = \sum_{i=1}^{k} 1\{x_1 + \cdots + x_i \leq k\}, \quad 1 \leq k < T_n,$$

denotes the number of arrivals in the first k slots. Analogously, the cumulative departure process is $(D_k : 1 \leq k < T_n)$, where $D_k = \sum_{i=1}^k v_i$ is the number of departures in the first k slots, $1 \leq k < T_n$. The number of packets that remain in the system at the end of the kth slot is $A_k - D_k$, $1 \leq k < T_n$.

If $A_k - D_k = 0$, the queue is empty at the end of the kth slot, and hence no packet exits in the $(k+1)$st slot. If $A_k - D_k > 0$, a packet is served in the $(k+1)$st slot. Using the memoryless property of geometric service times, this packet departs in the

$(k+1)$st slot with probability μ_1, or stays in the system with probability $1-\mu_1$, independent of the past.

The timing channel therefore behaves like a binary memoryless Z-channel, W, with $W(1|1)=\mu_1$ and $W(1|0)=0$. The inputs to the Z-channel are $u_1=0$, and $u_{k+1}=1\{A_k-D_k>0\}$, $k=1,\cdots,T_n-1$. The output sequence is $(v_1,\cdots,v_{T_n})$ given by (2.17).

Any window-code on the timing channel is therefore equivalent to the above strategy with complete feedback on the memoryless Z-channel. Complete feedback is necessary because the $(k+1)$st input, u_{k+1}, depends on the past departures (outputs) through D_k.

The capacity of the timing channel (for window-codes) is therefore upperbounded by the capacity of the memoryless Z-channel with complete feedback. Since feedback does not increase the capacity of the memoryless Z-channel, the upperbound is found using standard techniques to be

$$\max_{0\leq p\leq 1}\ [h(p\mu_1)-ph(\mu_1)]=\log\left[1+\mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}\right].$$

This completes the proof. ■

We can in fact say more about the converse. For window-codes with rate above $\log[1+\mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}]$ nats per slot, the probability of error goes to 1. This follows from the strong converse for DMCs with feedback [19, P.2.5.16(c)].

2.2.6 Discussion

Propositions 1 and 2 show that the exponential server's maximum-likelihood decoder, ϕ_f (cf. (2.13) and (2.8)), is a robust decoder. Suppose that the service times are stationary and ergodic with mean $1/\mu$ seconds. When the cost of transmission is the expected departure time of the last packet, rate $e^{-1}\mu$ nats per second is *achievable* using the decoder ϕ_f (Proposition 1). A *window-code* is one where the decoder makes a decision based on departures in a certain time window, and all arrivals fall within this time window. The decoder ϕ_f does not have to look beyond the time window to make a decision. Rate $e^{-1}\mu$ nats per second is *achievable with window-codes* using the decoder ϕ_f (Proposition 2). Furthermore, when the service times are independent and exponentially distributed, this rate is the largest achievable with window-codes (Proposition 6(a)). We prove this result by mapping any window-code on the timing channel to an equivalent strategy with complete feedback on the point-process channel. Using feedback on the timing channel to avoid queueing, rate $e^{-1}\mu$ nats per second is *achievable with feedback* using the decoder φ_d (Proposition 4). The mutual information saddle-point inequality plays a crucial role in the proof of our achievability results under mismatch.

On the jammed timing channel, the jammer (server) includes an arbitrary amount of delay to the service time of each packet. This is done to diminish the transmission capabilities of the timing channel. The total delay for all the n packets cannot exceed n/μ_2 seconds. The nominal service times are stationary and ergodic with mean $1/\mu_1$ seconds. Let $\mu = \mu_1\mu_2/(\mu_1+\mu_2)$. Rate $e^{-1}\mu$ nats per second is *achievable with random window-codes* using the decoder ϕ_f (Proposition 3). Furthermore, when the service times are independent and exponentially distributed, rate $e^{-1}\mu$ nats per second is the largest achievable with random window-codes (Proposition 6(b)).

Analogous results hold for the discrete-time single-server queueing system (Propositions 5 and 6(c, d)). Furthermore, if each packet carries a non-zero amount of information, there is a non-random communication strategy to transmit information reliably on the jammed timing channel (cf. discussion following Proposition 5). This strategy uses only a negligible fraction of packet contents. We do not know if a similar result holds for the continuous-time system. Suppose that the jammer is now aware of the codebook in use, and there is no side channel available. We do not know the (deterministic-code) capacity of such a jammed timing channel.

We conclude this section with the following observation. We can map any window-code on the timing channel to an equivalent strategy with complete feedback on the point-process channel (binary memoryless Z-channel in the discrete-time case). Proposition 2 (resp. Proposition 5(b)) therefore gives us an alternative capacity-achieving strategy with complete feedback on the point-process channel (resp. Z-channel). It is well-known that the capacities of the point-process channel and the discrete memoryless channel do not increase with feedback. This fact gives a simple explanation of why the capacity of the timing channel does not increase with feedback.

2.3 Proofs

In this section we prove Propositions 1-4 and 6(a,b). We begin with a simple extension of [19, Lemma 6.9]. We provide the proof because of some minor variations in the statement of the lemma and its applicability to standard measurable spaces. A *standard measurable space* is a measurable space $(A, \mathcal{A})$ that is isomorphic to $(F, \mathcal{F})$, where F is a Borel subset of $[0,1]$ and $\mathcal{F}$ is the Borel σ-algebra on F.

Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be two standard measurable spaces. Let $P_{Y|X}$ be a transition probability function from $(A, \mathcal{A})$ to $(B, \mathcal{B})$. Let P_X be a probability measure on $(A, \mathcal{A})$. P_X and $P_{Y|X}$ induce a joint probability measure $P_{X,Y}$ (cf. [29, P.18.25(b), p.247]), and a marginal probability measure P_Y (cf. [29, P.18.25(d), p.247]), on the appropriate spaces.

Let $\mathbf{c} = \{x_1, \cdots, x_M\}$ be a codebook of M codewords, $x_i \in A$ for $i = 1, \cdots, M$. Let $g : A \to \mathcal{R}_+$ be a measurable function that represents a constraint on the inputs. For a fixed $\Gamma \in \mathcal{R}_+$, we require that $g(x) \leq \Gamma$ for every $x \in \mathbf{c}$. Fix a set $H \in \mathcal{B}$. Let $f : A \times B \to \mathcal{R}_+$ be a measurable function. Let $\phi_{f,H} : B \to \{0, 1, \cdots, M\}$ denote the mapping

$$\phi_{f,H}(y) \triangleq \begin{cases} i, & \text{if } y \in H \text{ and } \max_{j\neq i} f(x_j,y) < f(x_i,y) \\ 0, & \text{if } y \notin H \text{ or if no such } x_i \text{ exists in } \mathbf{c}. \end{cases}$$

In other words, when $y \in H$, the decoder looks for a unique codeword that maximizes $f(\cdot, y)$. Given a codebook $\mathbf{c}$, the encoder and the decoder are thus fixed. An error occurs whenever the decoder declares a symbol that is different from the transmitted symbol. Let $P_{e,i}(\mathbf{c}, \phi_{f,H}, P_{Y|X})$ denote the probability of error when the codebook is $\mathbf{c}$ and the transmitted message is i.

Lemma 1: *Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be standard alphabet spaces. Let $\Gamma \in \mathcal{R}_+$ and $\delta \in (0,1)$. Let P_X be a probability measure on $(A, \mathcal{A})$ that satisfies*

$$P_X\{g(X) \leq \Gamma\} > 1 - \delta. \tag{2.18}$$

Let $f : A \times B \to \mathcal{R}_+$ be a measurable function that satisfies

$$Ef(X, y) = 1 \tag{2.19}$$

for every $y \in B$. Let $H \in \mathcal{B}$. There exists a random variable $\mathbf{C}$ that takes values in the set of codebooks with M codewords of blocklength 1, such that for any realization $\mathbf{c}$ of this random variable, $g(x) \leq \Gamma$ for every $x \in \mathbf{c}$. Furthermore, for every $\beta > 0$, every $i \in \{1, \cdots, M\}$,

$$EP_{e,i}(\mathbf{C}, \phi_{f,H}, P_{Y|X}) \leq \frac{P_{X,Y}\{f(X,Y) \leq \beta\}}{1-\delta} + \frac{M}{\beta(1-\delta)^2} + \frac{P_Y\{Y \notin H\}}{1-\delta}.$$

Proof: Let $A_\Gamma = \{x \in A : g(x) \leq \Gamma\}$ and $\mathcal{A}_\Gamma = \{G \cap A_\Gamma : G \in \mathcal{A}\}$. From (2.18), $P_X\{A_\Gamma\} > 1 - \delta$. Let $P_X^{A_\Gamma}$ be the restriction of P_X to $(A_\Gamma, \mathcal{A}_\Gamma)$. Let P'_X be the probability measure on $(A_\Gamma, \mathcal{A}_\Gamma)$ given by

$$dP'_X(x) = \frac{dP_X^{A_\Gamma}(x)}{P_X\{A_\Gamma\}}, \qquad x \in A_\Gamma. \tag{2.20}$$

Fix $M \in \mathcal{N}$. Each codebook is a subset of A_Γ that contains M elements. Let the codebook random variable $\mathbf{C} = (X_1, \cdots, X_M)$ have the product distribution

$$dP_{\mathbf{C}}(x_1, \cdots, x_M) = dP'_X(x_1)dP'_X(x_2) \cdots dP'_X(x_M), \tag{2.21}$$

on $(A_\Gamma^M, \mathcal{A}_\Gamma^M)$, where $\mathcal{A}_\Gamma^M$ is the product Borel σ-algebra on A_Γ^M. Each codeword therefore is drawn independently from A_Γ according to P'_X. Clearly, for any realization $\mathbf{c}$ of the codebook random variable, $g(x) \le \Gamma$ for every $x \in \mathbf{c}$.

Consider an auxiliary threshold decoder that declares i as the transmitted message only if, for the received $y \in H$, x_i is the only codeword that satisfies $f(x_i, y) > \beta$. Otherwise this auxiliary decoder declares 0, an error. Whenever this auxiliary decoder declares a non-zero symbol, it agrees with the output of the decoder $\phi_{f,H}$. The auxiliary decoder therefore performs worse than $\phi_{f,H}$.

The error probability given that message i is transmitted, averaged over the random selection of the codebook, is the same for all $i = 1, \cdots, M$. We therefore assume without loss of generality that the message $i = 1$ is transmitted. For a fixed codebook $(x_1, \cdots, x_M)$, we are interested in the probability of error when x_1 is transmitted. The error event depends on the entire codebook. Let $E \in \mathcal{B}$. It can be verified that the mapping defined by $((x_1, \cdots, x_M), E) \to P_{Y|X}(E|x_1)$ is a transition probability function from $(A_\Gamma^M, \mathcal{A}_\Gamma^M)$ to $(B, \mathcal{B})$. It represents the probability of an event E given that the codebook is $(x_1, \cdots, x_M)$ and x_1 is transmitted.

For any set $F \subset A_\Gamma^M \times B$ that is measurable relative to the product σ-algebra on $A_\Gamma^M \times B$, let

$$\Pr\{F\} \triangleq \int_{A_\Gamma^M} dP_{\mathbf{C}}(x_1, \cdots, x_M) \int_B dP_{Y|X}(y|x_1)\, 1\{(x_1, \cdots, x_M, y) \in F\}, \tag{2.22}$$

i.e., probability of the event F given that message $i = 1$ is transmitted, averaged over the random selection of codebooks.

If the auxiliary decoder makes an error, then one of the following events should have occurred: (i) $f(x_1, y) \leq \beta$, (ii) $f(x_j, y) > \beta$, for some $j \neq 1$, or (iii) $y \notin H$. It is therefore sufficient to upperbound the probabilities of these events.

Using (2.22), (2.21), (2.20), (2.18), we upperbound the probability of event (i) by

$$\Pr\left\{f(X_1, Y) \leq \beta\right\} \leq P_{X,Y}\left\{f(X,Y) \leq \beta\right\}/(1-\delta).$$

To upperbound the probability of event (ii), observe that

$$\begin{aligned}
&\Pr\left\{f(X_j, Y) > \beta, \text{for some } j \neq 1\right\} \\
\overset{(a)}{\leq}\ & M \cdot \Pr\left\{f(X_2, Y) > \beta\right\} \\
\overset{(b)}{=}\ & M \int_{A_\Gamma \times A_\Gamma} dP'_X(x_1)\, dP'_X(x_2) \int_B dP_{Y|X}(y|x_1)\, 1\left\{f(x_2, y) > \beta\right\} \\
\overset{(c)}{\leq}\ & \frac{M}{(1-\delta)^2} \int_{A \times A} dP_X(x_1)\, dP_X(x_2) \int_B dP_{Y|X}(y|x_1)\, 1\left\{f(x_2, y) > \beta\right\} \\
\overset{(d)}{=}\ & \frac{M}{(1-\delta)^2} \int_A dP_X(x_2) \int_B dP_Y(y)\, 1\left\{f(x_2, y) > \beta\right\} \\
\overset{(e)}{\leq}\ & \frac{M}{(1-\delta)^2} \int_A dP_X(x_2) \int_B dP_Y(y)\, \frac{f(x_2, y)}{\beta} \\
\overset{(f)}{=}\ & \frac{M}{\beta(1-\delta)^2} \int_B dP_Y(y) \\
=\ & \frac{M}{\beta(1-\delta)^2}.
\end{aligned}$$

In the above chain, (a) follows from the union bound for probabilities, (b) from (2.22) and (2.21), (c) from (2.20) and (2.18), (d) from Fubini's theorem and [29, P.18.25(d), p.247], (e) from the fact $1\left\{f(x_2, y) > \beta\right\} \leq f(x_2, y)/\beta$, and (f) from Fubini's theorem and (2.19).

Under maximum-likelihood decoding, $f = dP_{X,Y}/d(P_X \times P_Y)$; step (f) would then be unnecessary, and the last equality would follow immediately after (e). Under mismatched decoding, (2.19) is sufficient to obtain the last equality.

The probability of event (iii) is upperbounded by

$$\Pr\left\{Y \notin H\right\} \leq P_{X,Y}\left\{Y \notin H\right\}/(1-\delta) = P_Y\left\{Y \notin H\right\}/(1-\delta).$$

This completes the proof of Lemma 1. ■

Fix $n \in \mathcal{N}$. We apply Lemma 1 to the timing channel with $A = \mathcal{R}_+^n$ and $B = \mathcal{R}_+^{n+1}$. Let β^n play the role of β. Let $0 < \lambda'' < \lambda' < \mu$. Fix $\delta \in (0,1)$. Let P_{X^n} be as defined in (2.6). $P_{Y^n|X^n}$ is the transition probability function from the input space to the output space that is induced by a stationary and ergodic process of service times and the queue equations (2.2) and (2.3). Let P_{X^n,Y^n} denote the joint distribution under P_{X^n} and $P_{Y^n|X^n}$. Let P_{Y^n} denote the marginal of Y^n under P_{X^n,Y^n}.

Fix $M \in \mathcal{N}$. A codebook is a subset of $\mathcal{R}_+^n$ that contains M elements. Let the function g that denotes the input constraint in Lemma 1 be $g(x^n) \triangleq \left(\sum_{k=1}^n x_k\right)/n$, where $x^n = (x_1, \cdots, x_n)$. We require that $g(x^n) \le 1/\lambda''$ for every x^n in the codebook. By the weak law of large numbers for i.i.d random variables, we have that for every $\delta \in (0,1)$,

$$P_{X^n}\left\{g(X^n) \le 1/\lambda''\right\} \ge 1 - \delta \tag{2.23}$$

for all sufficiently large n.

The function f in (2.8) satisfies $E[f(X^n, y^n)] = 1$ for every $y^n \in \mathcal{R}_+^{n+1}$. We deal with two decoders. The decision set H for the decoder $\phi_{f,H}$ will be either $\left\{y^n \in \mathcal{R}_+^{n+1} : \sum_{k=0}^n y_k \le n/\lambda''\right\}$ or $\mathcal{R}_+^{n+1}$. When $H = \mathcal{R}_+^{n+1}$, the entire output set, we denote the corresponding decoder by ϕ_f after omitting the subscript H.

Corollary 1: *Fix $\delta \in (0,1)$ and $M \in \mathcal{N}$. Fix $n \in \mathcal{N}$ so that $P_{X^n}\left\{g(X^n) \le 1/\lambda''\right\} \ge 1 - \delta$. There exists a random variable $\mathbf{C}$ that takes values in the set of codebooks with M codewords, such that for any realization $\mathbf{c}$ of this random variable, $g(x^n) \le 1/\lambda''$ for every $x^n \in \mathbf{c}$. Furthermore, for every $\beta > 0$,*

$$(a) \quad E\frac{1}{M}\sum_{i=1}^{M} P_{e,i}(\mathbf{C}, \phi_f, P_{Y^n|X^n}) \le \frac{P_{X^n,Y^n}\left\{f(X^n,Y^n) \le \beta^n\right\}}{1-\delta} + \frac{M}{\beta^n(1-\delta)^2},$$

and with $H = \left\{y^n \in \mathcal{R}_+^{n+1} : \sum_{k=0}^n y_k \le n/\lambda''\right\}$,

$$\begin{aligned}(b) \quad E\frac{1}{M}\sum_{i=1}^{M} P_{e,i}(\mathbf{C}, \phi_{f,H}, P_{Y^n|X^n}) \quad &\le \quad \frac{P_{X^n,Y^n}\left\{f(X^n,Y^n) \le \beta^n\right\}}{1-\delta} + \frac{M}{\beta^n(1-\delta)^2} \\ &\quad + \frac{P_{Y^n}\left\{\sum_{k=0}^n Y_k > n/\lambda''\right\}}{1-\delta}.\end{aligned}$$

Proof: The corollary follows from Lemma 1 after averaging the probability of error over the M equiprobable codewords. ■

Since the function f depends on the quantity m (cf. (2.9)), we need the following lemma to evaluate $P_{X^n,Y^n}\{f(X^n,Y^n) \leq \beta^n\}$.

Lemma 2: *Let $0 < \lambda' < \mu$. For $y^n \in \mathcal{R}_+^{n+1}$, the function m satisfies $m(y^n) = 0$ if and only if $y_0 + y_1 = 0$.*

Proof: Fix $y^n \in \mathcal{R}_+^{n+1}$ so that $y_0 + y_1 > 0$ and $\sum_{k=0}^n y_k < \infty$. Let

$$S(y^n) \triangleq \left\{ x^n \in \mathcal{R}_+^n : \sum_{k=1}^n x_k \leq y_0 + y_1 \right\}.$$

Fix $x^n \in S(y^n)$. Then $0 \leq w_1 = \max\{0,\ x_1 - y_0\} \leq y_1$ and $w_k = 0, k = 2, \cdots, n$. These two conditions and (2.5) imply that $p(x^n, y^n) \geq u(y^n)$, where $u(y^n) \triangleq (\mu - \lambda')\mu^n \exp\{-\mu \sum_{k=0}^n y_k\}$. Observe that $u(y^n) > 0$. Moreover, for any $x^n \in S(y^n)$, $\prod_{k=1}^n e_{\lambda'}(x_k) \geq v(y^n) > 0$, where $v(y^n) \triangleq (\lambda')^n \exp\{-\lambda'(y_0 + y_1)\}$. Furthermore, $\int_{S(y^n)} d\pi(x^n) = (y_0 + y_1)^n / n! > 0$. After substitution of these quantities in (2.9), we get

$$\begin{aligned} m(y^n) &\geq \int_{S(y^n)} d\pi(x^n) \prod_{k=1}^n e_{\lambda'}(x_k) \ \ p(x^n, y^n) \\ &\geq v(y^n)\, u(y^n) \int_{S(y^n)} d\pi(x^n) \\ &> 0. \end{aligned}$$

Conversely, if $y_0 + y_1 = 0$, then either $x_1 = 0$ or $p(x^n, y^n) = 0$, and thus

$$m(y^n) = \int_{\mathcal{R}_+^n} dP_{X^n}(x^n)\, p(x^n, y^n)\, 1\{x_1 = 0\} = 0.$$

This completes the proof of the lemma. ■

We now show that under a mild condition on the process of service times the quantity $P_{X^n,Y^n}\{f(X^n,Y^n) \leq \beta^n\}$ goes to zero as $n \to \infty$ if β is chosen judiciously.

Lemma 3: *Let the process $(S_k : k \in \mathcal{N})$ of service times (not necessarily stationary or ergodic) satisfy for every $\alpha > 0$, the condition*

$$\lim_{n\to\infty} P_{S^n}\left\{\frac{1}{n}\sum_{k=1}^{n} S_k > \frac{1}{\mu} + \alpha\right\} = 0. \tag{2.24}$$

Then for every $\gamma > 0$,

$$\lim_{n\to\infty} P_{X^n,Y^n}\left\{\frac{1}{n}\log f(X^n,Y^n) \le \log\frac{\mu}{\lambda'} - \gamma\right\} = 0. \tag{2.25}$$

Proof: Observe that $Y_0 = 0$ under P_{X^n,Y^n} for every $n \in \mathcal{N}$. Let

$$T \triangleq \left\{(x^n, y^n) \in \mathcal{R}_+^n \times \mathcal{R}_+^{n+1} : \quad y_0 = 0 \ \text{ and } \ \sum_{k=1}^{j}(y_k - x_k) \ge 0, \text{ for } j = 1,\cdots,n\right\},$$

i.e., the set of all pairs (x^n, y^n) such that y^n is a possible sequence of interdeparture times (with $y_0 = 0$) when the sequence of interarrival times is x^n. From the queue equations (2.2) and (2.3), we have $P_{X^n,Y^n}\{T\} = 1$. We therefore have from (2.8) that

$$\begin{aligned} & P_{X^n,Y^n}\left\{\frac{1}{n}\log f(X^n,Y^n) \le \log\frac{\mu}{\lambda'} - \gamma\right\} \\ & \quad \le \ P_{X^n,Y^n}\{m(Y^n) = 0; \ \ (X^n,Y^n) \in T\} \\ & \qquad + P_{X^n,Y^n}\left\{\frac{1}{n}\log\frac{p(X^n,Y^n)}{m(Y^n)} \le \log\frac{\mu}{\lambda'} - \gamma; \ \ m(Y^n) > 0; \ \ (X^n,Y^n) \in T\right\}. \end{aligned} \tag{2.26}$$

From Lemma 2, (2.2) and (2.3), we have that

$$\{m(y^n) = 0; \ \ (x^n,y^n) \in T\} = \{y_0 + y_1 = 0; \ \ (x^n,y^n) \in T\} \subset \{x_1 = 0\},$$

and therefore

$$P_{X^n,Y^n}\{m(Y^n) = 0; \ \ (X^n,Y^n) \in T\} \le P_{X^n}\{X_1 = 0\} = 0.$$

We now upperbound the term in (2.26). Let $(x^n,y^n) \in T$ and $m(y^n) > 0$. From (2.10) and (2.12) we can write

$$m(y^n) = \prod_{k=1}^{n} e_{\lambda'}(y_k) \cdot g_{Y_0|Y_1,\cdots,Y_n}(y_0|y_1,\cdots,y_n),$$

where the function g is the conditional density of Y_0 given $(Y_1, \cdots, Y_n)$ under Q_{Y^n}. Using (2.5), we get

$$\frac{1}{n}\log\frac{p(x^n, y^n)}{m(y^n)} = \log\frac{\mu}{\lambda'} + \frac{\lambda'}{n}\sum_{k=1}^{n} y_k - \frac{\mu}{n}\sum_{k=1}^{n}(y_k - w_k) - \frac{1}{n}i_{Y_0;Y_1,\cdots,Y_n}(y_0\ ;\ y_1, \cdots, y_n), \tag{2.27}$$

where $(1/n)i_{Y_0;Y_1,\cdots,Y_n}$ (cf. [1, Lemma 3]) is the normalized information density of the relevant quantities, under Q_{Y^n}. Observe that

$$\sum_{k=1}^{n} y_k \geq \sum_{k=1}^{n} x_k,$$

and that

$$\sum_{k=1}^{n}(y_k - w_k) = \sum_{k=1}^{n} s_k,$$

the sum of service times (cf. (2.3)). Using these facts and (2.27), we get that

$$\begin{aligned}
&\left\{\frac{1}{n}\log\frac{p(x^n, y^n)}{m(y^n)} \leq \log\frac{\mu}{\lambda'} - \gamma;\quad m(y^n) > 0;\quad (x^n, y^n) \in T\right\} \\
\subset\ &\left\{\frac{\lambda'}{n}\sum_{k=1}^{n} x_k - \frac{\mu}{n}\sum_{k=1}^{n} s_k - \frac{1}{n}i_{Y_0;Y_1,\cdots,Y_n}(y_0\ ;\ y_1, \cdots, y_n) \leq -\gamma;\quad (x^n, y^n) \in T\right\} \\
\subset\ &\left\{\frac{1}{n}i_{Y_0;Y_1,\cdots,Y_n}(y_0\ ;\ y_1, \cdots, y_n) \geq \frac{\gamma}{3}\right\} \\
\cup\ &\left\{\frac{1}{n}\sum_{k=1}^{n} x_k < \frac{1}{\lambda'} - \frac{\gamma}{3\lambda'}\right\} \cup \left\{\frac{1}{n}\sum_{k=1}^{n} s_k > \frac{1}{\mu} + \frac{\gamma}{3\mu};\quad (x^n, y^n) \in T\right\}.
\end{aligned}$$

Using the above inclusions, the term in (2.26) is upperbounded by

$$P_{Y^n}\left\{\frac{1}{n}i_{Y_0;Y_1,\cdots,Y_n} \geq \frac{\gamma}{3}\right\} + P_{X^n}\left\{\frac{1}{n}\sum_{k=1}^{n} X_k < \frac{1}{\lambda'} - \frac{\gamma}{3\lambda'}\right\} + P_{S^n}\left\{\frac{1}{n}\sum_{k=1}^{n} S_k > \frac{1}{\mu} + \frac{\gamma}{3\mu}\right\}, \tag{2.28}$$

where the last term follows after a change of variables. The first term in (2.28) goes to 0 by [1, Lemma 3], since $P_{X^n,Y^n}\{Y_0 = 0\} = 1$, for every $n \in \mathcal{N}$. In fact, from the proof of [1, Lemma 3], this term equals 0 for all sufficiently large n. The second term in (2.28) goes to 0 by the weak law of large numbers for i.i.d random variables. The third term in (2.28) goes to 0 by the assumption of the lemma. ■

We are now ready to prove Propositions 2 and 3.

Proof of Proposition 2: Let the assumptions preceding Corollary 1 hold. Fix arbitrary $\gamma > 0$. Let $\log \beta = \log(\mu/\lambda') - \gamma$. We now apply Corollary 1(a). Since the process $(S_k : k \in \mathcal{N})$ is stationary and ergodic, it satisfies (2.24) by the ergodic theorem (see for e.g., [9, Theorem 7.3.3, pp.236-237]), and therefore (2.25) holds by Lemma 3. Choosing M_n so that

$$\log(\mu/\lambda') - 2\gamma < (\log M_n)/n < \log(\mu/\lambda') - 3\gamma/2$$

ensures that $\lim_{n\to\infty} M_n/(\beta^n(1-\delta)^2) = 0$. We can therefore find a sequence $(\varepsilon_n : n \in \mathcal{N})$, such that

$$E\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e,i}(\mathbf{C}, \phi_f, P_{Y^n|X^n}) \leq \varepsilon_n$$

for every $n \in \mathcal{N}$, and $\lim_{n\to\infty} \varepsilon_n = 0$.

Consequently, for each $n \in \mathcal{N}$, there is a codebook realization $\mathbf{c}$ that satisfies

$$\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e,i}(\mathbf{c}, \phi_f, P_{Y^n|X^n}) \leq \varepsilon_n.$$

Furthermore, every codeword $x^n \in \mathbf{c}$ satisfies $\sum_{k=1}^{n} x_k \leq n/\lambda''$. The decoder ϕ_f observes only those departures that occur in the time window $[0, n/\lambda'']$. We therefore have a sequence of $(n, M_n, n/\lambda'', \varepsilon_n)$-window-codes that satisfies

$$\frac{\log M_n}{n/\lambda''} \geq \lambda'' \log\frac{\mu}{\lambda'} - 2\gamma\lambda''$$

for every $n \in \mathcal{N}$, and $\lim_{n\to\infty} \varepsilon_n = 0$. Setting $\lambda' = e^{-1}\mu$ and $\lambda'' = e^{-1}\mu/(1+\gamma)$, we get

$$\lambda''(\log M_n)/n \geq e^{-1}\mu - 3\gamma e^{-1}\mu/(1+\gamma)$$

for every $n \in \mathcal{N}$. This means that rate $e^{-1}\mu$ nats per second is achievable with window-codes (since γ was arbitrary). ■

Proof of Proposition 3: Fix arbitrary $\gamma > 0$. The process $(S_k : k \in \mathcal{N})$ of nominal service times is stationary and ergodic. Consider $n \in \mathcal{N}$ and $\mathbf{z} \in \mathcal{R}_+^n$ satisfying

$l(\mathbf{z}) \leq 1/\mu_2$. Let $W^n(\mathbf{z})$ be the corresponding transition probability function. Set $1/\mu = 1/\mu_1 + 1/\mu_2$. Since $l(\mathbf{z}) \leq 1/\mu_2$, for every $\alpha > 0$, we can write

$$\left\{ \frac{1}{n} \sum_{k=1}^{n} (s_k + z_k) > \frac{1}{\mu} + \alpha \right\} \subset \left\{ \frac{1}{n} \sum_{k=1}^{n} s_k > \frac{1}{\mu_1} + \alpha \right\}. \tag{2.29}$$

With $W^n(\mathbf{z})$ in place of $P_{Y^n|X^n}$ in the proof of Lemma 3, and by using (2.29), we obtain

$$\begin{aligned} & P_{X^n,Y^n} \left\{ \frac{1}{n} \log f(X^n, Y^n) \leq \log \frac{\mu}{\lambda'} - \gamma \right\} \\ \leq \; & P_{Y^n} \left\{ \frac{1}{n} i_{Y_0;Y_1,\cdots,Y_n} \geq \frac{\gamma}{3} \right\} + P_{X^n} \left\{ \frac{1}{n} \sum_{k=1}^{n} X_k < \frac{1}{\lambda'} - \frac{\gamma}{3\lambda'} \right\} \\ + \; & P_{S^n} \left\{ \frac{1}{n} \sum_{k=1}^{n} S_k > \frac{1}{\mu_1} + \frac{\gamma}{3\mu} \right\} \end{aligned} \tag{2.30}$$

Observe that $P_{X^n,Y^n}\{\cdot\}$ and $P_{Y^n}\{\cdot\}$ depend in general on the state sequence $\mathbf{z}$. For sufficiently large n, however, the term $P_{Y^n}\left\{(1/n) i_{Y_0;Y_1,\cdots,Y_n} \geq \gamma/3\right\}$ does not depend on $\mathbf{z}$ because of the following. Indeed, for every $\mathbf{z}$, $P_{Y^n}\{Y_0 > 0\} = 0$. In this case, we can extend the proof of [1, Lemma 1] to get $P_{Y^n}\left\{(1/n) i_{Y_0;Y_1,\cdots,Y_n} \geq \gamma/3\right\} = 0$ for every $n > n_0(\gamma, \lambda', \mu)$. The quantity $n_0(\gamma, \lambda', \mu)$, which can be taken as $-(3/\gamma)\log(1-\lambda'/\mu)$, is independent of $\mathbf{z}$. Furthermore, for a fixed n, the remaining two terms in (2.30) do not depend on $\mathbf{z}$ if $l(\mathbf{z}) \leq 1/\mu_2$; they go to 0 as $n \to \infty$.

Choose $\lambda', \lambda'', \beta$ and the sequence $(M_n : n \geq 1)$ as in the proof of Proposition 2. From Corollary $1(a)$, there is a sequence $(\varepsilon_n : n \in \mathcal{N})$, such that for all sufficiently large n,

$$E \frac{1}{M_n} \sum_{i=1}^{M_n} P_{e,i}(\mathbf{C}, \phi_f, W^n(\mathbf{z})) \leq \varepsilon_n$$

for every $\mathbf{z}$ with $l(\mathbf{z}) \leq 1/\mu_2$, and $\lim_{n\to\infty} \varepsilon_n = 0$. We have therefore obtained a sequence of $(n, M_n, n/\lambda'', \varepsilon_n)$-random window-codes. The sequence also satisfies

$$\lambda'' \frac{\log M_n}{n} \geq e^{-1}\mu - 3e^{-1}\mu \frac{\gamma}{(1+\gamma)}$$

for every $n \in \mathcal{N}$. ■

Proving Proposition 1 requires more work. We need to the find an upperbound on the expected time of departure of the nth packet. We first prove the following lemma.

Lemma 4: *Let the single-server queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. Let the queue be driven by a Poisson process of rate $\lambda' < \mu$. Then $(1/n)\sum_{k=1}^{n} Y_k \to 1/\lambda'$ in probability.*

Proof: Let $(X_n : n \in \mathcal{N})$ be the process of independent and exponentially distributed interarrival times, and $(S_n : n \in \mathcal{N})$ the stationary and ergodic process of service times. Let $P\{\cdot\}$ denote the probability of an event with respect to the joint process. Let R_n be the waiting time of the nth packet. Observe that $\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} X_k + R_n + S_n$. Since $R_n, S_n \geq 0$, we have that for any $\gamma > 0$,

$$P\left\{\frac{1}{n}\sum_{k=1}^{n} Y_k < \frac{1}{\lambda'} - \gamma\right\} \leq P\left\{\frac{1}{n}\sum_{k=1}^{n} X_k < \frac{1}{\lambda'} - \gamma\right\} \to 0$$

as $n \to \infty$. On the other hand, using $\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} X_k + R_n + S_n$ and the union bound for probabilities, we get

$$P\left\{\frac{1}{n}\sum_{k=1}^{n} Y_k > \frac{1}{\lambda'} + \gamma\right\} \leq P\left\{\frac{R_n}{n} > \frac{\gamma}{3}\right\} + P\left\{\frac{S_n}{n} > \frac{\gamma}{3}\right\} + P\left\{\frac{1}{n}\sum_{k=1}^{n} X_k > \frac{1}{\lambda'} + \frac{\gamma}{3}\right\}. \tag{2.31}$$

The last two terms on the right side of (2.31) go to 0 as $n \to \infty$. We now upperbound the first term in (2.31).

It can be shown that $((X_n, S_n) : n \in \mathcal{N})$ is a stationary and ergodic process. R_n converges in distribution to a finite random variable ϕ that satisfies $P\{\phi < \infty\} = 1$ because $\lambda' < \mu$ [9, Theorem 7.4.5, p.241]. Observe that

$$\lim_{n \to \infty} P\{R_n \leq C\} = P\{\phi \leq C\}$$

if C is a point of continuity of the cumulative distribution function (cdf) of ϕ. A cdf can have only a countable number of discontinuities. Fix arbitrary $\varepsilon > 0$. Choose

$C \in (0,\infty)$ large enough so that $P\{\phi > C\} \le \varepsilon$ and C is a point of continuity of the cdf of ϕ. Then, for all sufficiently large n,

$$P\{R_n > n\gamma/3\} \le P\{R_n > C\} \le P\{\phi > C\} + \varepsilon \le 2\varepsilon.$$

This proves the lemma. ■

Proof of Proposition 1: Let the assumptions preceding Corollary 1 hold. Fix $\gamma > 0$. Let $\lambda' = e^{-1}\mu$, $\lambda'' = e^{-1}\mu/(1+\gamma)$, $\lambda = e^{-1}\mu/(1+2\gamma)$ and $\log\beta = \log(\mu/\lambda')-\gamma$. Let $H = \left\{y^n \in \mathcal{R}_+^{n+1} : \sum_{k=0}^n y_k \le n/\lambda''\right\}$. Consider the decoder $\phi_{f,H}$. Clearly, this decoder cannot outperform ϕ_f. The process $(S_k : k \in \mathcal{N})$ is stationary and ergodic with mean $1/\mu$ seconds; it therefore satisfies (2.24) and therefore (2.25) holds by Lemma 3. Choosing M_n so that

$$\log(\mu/\lambda') - 2\gamma < (\log M_n)/n < \log(\mu/\lambda') - 3\gamma/2$$

ensures that $\lim_{n\to\infty} M_n/(\beta^n(1-\delta)^2) = 0$. Since $P_{Y^n}\{Y_0 = 0\} = 1$ for every $n \in \mathcal{N}$ and $1/\lambda'' = 1/\lambda' + \gamma e/\mu$, we can apply Lemma 4 to get $\lim_{n\to\infty} P_{Y^n}\{Y^n \notin H\} = 0$. From Corollary 1(*b*), we can find a sequence $(\varepsilon_n : n \in \mathcal{N})$, such that

$$E\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e,i}(\mathbf{C}, \phi_{f,H}, P_{Y^n|X^n}) \le \frac{\varepsilon_n}{2}$$

for every $n \in \mathcal{N}$, and $\lim_{n\to\infty} \varepsilon_n = 0$.

Fix $n \in \mathcal{N}$. We can find a codebook $\mathbf{c}(n)$ such that

$$\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e,i}(\mathbf{c}(n), \phi_{f,H}, P_{Y^n|X^n}) \le \frac{\varepsilon_n}{2}.$$

Furthermore, every codeword $x^n \in \mathbf{c}(n)$ satisfies $\sum_{k=1}^n x_k \le n/\lambda''$. By removing the $M_n - \lfloor M_n/2 \rfloor$ worst codewords, relabeling the remaining codewords from 1 through $\lfloor M_n/2 \rfloor$, and denoting the resulting codebook as $\mathbf{c}'(n)$, we get

$$P_{e,i}(\mathbf{c}'(n), \phi_{f,H}, P_{Y^n|X^n}) \le \varepsilon_n,$$

for $i = 1, \cdots, \lfloor M_n/2 \rfloor$. In particular, this implies that

$$P_{Y^n|X^n}\left\{\frac{1}{n}\sum_{k=0}^{n} Y_k > \frac{1}{\lambda''} \mid x^n\right\} \le \varepsilon_n \tag{2.32}$$

for every $x^n \in \mathbf{c}'(n)$. The next lemma shows that the expected time of the nth departure given each codeword is smaller than n/λ for all sufficiently large n.

Lemma 5: *Fix* $0 < \lambda < \lambda'' < \lambda' < \mu$. *Let* $(\varepsilon_n : n \in \mathcal{N})$ *satisfy* $\lim_{n\to\infty} \varepsilon_n = 0$. *Let* $(\mathbf{c}'(n) : n \in \mathcal{N})$ *be a sequence of codebooks that satisfies for each* $n \in \mathcal{N}$, *the condition (2.32) and the condition* $\sum_{k=1}^{n} x_k \le n/\lambda''$ *for every* $x^n \in \mathbf{c}'(n)$. *Then for all sufficiently large* n,

$$E\left[\frac{1}{n}\sum_{k=0}^{n} Y_k \mid x^n\right] \le 1/\lambda$$

for every $x^n \in \mathbf{c}'(n)$.

Proof: Fix a $\nu > 0$ so that $1/\lambda'' + \nu < 1/\lambda$. Let

$$F_n \triangleq \{y^n \in \mathcal{R}_+^{n+1} : \sum_{k=0}^{n} y_k > n/\lambda''\}.$$

Let F_n^c denote the complement of the set F_n, and 1_{F_n} the indicator function of F_n. From (2.32), $P_{Y^n|X^n}\{F_n \mid x^n\} \le \varepsilon_n$ for every $x^n \in \mathbf{c}'(n)$.

Let S_k be the service time of the kth packet. For a $C > 0$, let

$$G_k \triangleq \{s_k \in \mathcal{R}_+ : s_k > C\}.$$

For every $x^n \in \mathbf{c}'(n)$, every $k = 1, \cdots, n$, we have

$$\begin{aligned} E[S_k 1_{F_n} \mid x^n] &= E[S_k 1_{F_n} 1_{G_k} \mid x^n] + E[S_k 1_{F_n} 1_{G_k^c} \mid x^n] \\ &\le E[S_k 1_{G_k} \mid x^n] + C P_{Y^n|X^n}\{F_n \mid x^n\}. \end{aligned} \tag{2.33}$$

Observe that $E[S_k 1_{G_k} \mid x^n]$ is independent of x^n and of k because the process $(S_k : k \in \mathcal{N})$ is stationary and independent of the arrivals. Furthermore,

$$E[S] = E[S1_G] + E[S1_{G^c}] = 1/\mu < \infty.$$

Using the monotone convergence theorem, we can choose a C large enough so that $E[S1_G] < \nu/3$. Pick n large enough so that $\max\{1/\lambda'', C\} \cdot \varepsilon_n < \nu/3$. Using (2.33), we therefore have that for all sufficiently large n,

$$E[S_k 1_{F_n} \mid x^n] \leq \nu/3 + C\varepsilon_n \leq 2\nu/3.$$

Since $\sum_{k=1}^n x_k \leq n/\lambda''$ for every $x^n \in \mathbf{c}'(n)$, we get

$$E\left[\left(\frac{1}{n}\sum_{k=1}^n (x_k + S_k)\right) \cdot 1_{F_n} \mid x^n\right] \leq \varepsilon_n/\lambda'' + 2\nu/3 \leq \nu.$$

Since we can assume $\sum_{k=0}^n Y_k \leq \sum_{k=1}^n (x_k + S_k)$ under $P_{Y^n|X^n}\{\cdot \mid x^n\}$, it follows that $E[((1/n)\sum_{k=0}^n Y_k) \cdot 1_{F_n} \mid x^n] \leq \nu$. Therefore, for every $x^n \in \mathbf{c}'(n)$,

$$\begin{aligned} E\left[\frac{1}{n}\sum_{k=0}^n Y_k \mid x^n\right] &= E\left[\left(\frac{1}{n}\sum_{k=0}^n Y_k\right) \cdot 1_{F_n} \mid x^n\right] + E\left[\left(\frac{1}{n}\sum_{k=0}^n Y_k\right) \cdot 1_{F_n^c} \mid x^n\right] \\ &\leq \nu + 1/\lambda'' \\ &< 1/\lambda. \end{aligned}$$

This completes the proof of the lemma. ■

Continuing with the proof of Proposition 1, for all sufficiently large n, the expected time of the nth departure is not greater than n/λ (cf. Lemma 5). We therefore have a sequence of $(n, \lfloor M_n/2 \rfloor, n/\lambda, \varepsilon_n)$-codes that satisfies

$$\lambda\left(\log\lfloor M_n/2\rfloor\right)/n > e^{-1}\mu - 5\gamma e^{-1}\mu/(1+2\gamma)$$

for all sufficiently large n, and $\lim_{n\to\infty} \varepsilon_n = 0$. This proves that rate $e^{-1}\mu$ nats per second is achievable. ■

Proof of Proposition 4: Fix $0 < \lambda < \lambda' < \mu$. Consider an $n \in \mathcal{N}$. We apply Lemma 1 with $A = B = \mathcal{R}_+^n$. Observe that $Y_k = X_k + S_k$, $k = 1, \cdots, n$. Let $P_{Y^n|X^n}$ denote the transition probability function from the input space to the output space. Choose M_n as in the proof of Proposition 2. Let $g(x^n) \triangleq (1/n)\sum_{k=1}^n x_k$. We require that $g(x^n) \leq 1/\lambda - 1/\mu$ for every codeword x^n. Since the mean service time is $1/\mu$

seconds, it follows that the expected time of the nth departure is not greater than n/λ.

Let P_X be the mixture of a point mass and an exponential distribution given by

$$\begin{aligned} P_X\{X=0\} &= \lambda'/\mu, \\ P_X\{X>x\} &= \left(1-\frac{\lambda'}{\mu}\right)e^{-\lambda' x}, \quad x \geq 0. \end{aligned}$$

Note that P_X is the input distribution that attains the mutual information saddle point [1, Theorem 3], [23, Theorem 1]. Let P_{X^n} be the distribution under which $X^n = (X_1, \cdots, X_n)$ is a vector of i.i.d random variables with distribution P_X. Observe that if the service times are independent and exponentially distributed with mean $1/\mu$ seconds, then the outputs are independent and exponentially distributed with mean $1/\lambda'$ seconds. Let

$$f(x^n, y^n) \triangleq \prod_{k=1}^{n} \frac{e_\mu(y_k - x_k)}{e_{\lambda'}(y_k)}. \tag{2.34}$$

This function satisfies $Ef(X^n, y^n) = 1$ for every $y^n \in \mathcal{R}_+^n$. Let $H = \mathcal{R}_+^n$. Observe that the decoder ϕ_f (cf. (2.13)) with f as in (2.34) is the same as the decoder φ_d (cf. Section 2.2.3).

Let P_{X^n,Y^n} be the joint distribution under P_{X^n} and $P_{Y^n|X^n}$. We only need to consider $(x^n, y^n) \in \mathcal{R}_+^n \times \mathcal{R}_+^n$ that satisfy $0 \leq x_k \leq y_k$, $k = 1, \cdots, n$. For such an (x^n, y^n),

$$\frac{1}{n}\log f(x^n, y^n) = \log\frac{\mu}{\lambda'} + \frac{\mu}{n}\sum_{k=1}^{n} x_k - \frac{(\mu - \lambda')}{n}\sum_{k=1}^{n} y_k.$$

From this and the stationarity and ergodicity of $((X_n, Y_n) : n \geq 1)$, we get

$$\lim_{n\to\infty} P_{X^n,Y^n}\left\{\frac{1}{n}\log f(X^n, Y^n) \leq \log\frac{\mu}{\lambda'} - \gamma\right\} = 0.$$

The rest of the proof is similar to that of Proposition 2. ■

Proof of Proposition 6(a): Fix $n \in \mathcal{N}$. There are M_n messages. Each message corresponds to a sequence of interarrival times; the nth arrival occurs before time T_n.

This sequence maps to a (right-continuous with left limits) point process of arrivals $(A_t : t \in [0, T_n])$. A_t is the number of arrivals in $[0, t]$, $t \in [0, T_n]$. Analogously, the observed departures form a (right-continuous with left limits) point process $(D_t : t \in [0, T_n])$, where D_t is the number of departures in $[0, t]$, $t \in [0, T_n]$. Let

$$\mathcal{F}_t^A \triangleq \sigma(A_s : s \in [0, t]), \quad \mathcal{F}_t^D \triangleq \sigma(D_s : s \in [0, t]), \quad \mathcal{F}_t^{A,D} \triangleq \sigma\{\mathcal{F}_t^A, \mathcal{F}_t^D\}.$$

For $t = 0$, let $\nu(0, A, D) \triangleq 0$. For $t > 0$, fix an increasing sequence of rational numbers $(r_n : n \in \mathcal{N})$ such that $r_n \uparrow t$, and let $\nu(t, A, D) \triangleq \lim_{r_n \uparrow t} (A_{r_n} - D_{r_n})$. The quantity $\nu(t, A, D)$ represents the number of packets that remain in the system at time $t-$, i.e., just prior to t. Clearly $\nu(t, A, D)$ is $\mathcal{F}_t^{A,D}$-measurable for every $t \in [0, T_n]$ and $(\nu(t, A, D) : t \in [0, T_n])$ is a left-continuous process. Let

$$\lambda(t, A, D) \triangleq \mu \, 1\{\nu(t, A, D) > 0\}, \quad t \in [0, T_n].$$

Observe that $\lambda(t, A, D)$ is $\mathcal{F}_t^{A,D}$-measurable for every $t \in [0, T_n]$ and that $(\lambda(t, A, D) : t \in [0, T_n])$ is a left-continuous process; it is therefore a *predictable* process [31, Definition 3, p.173].

Fix arbitrary $t \in [0, T_n]$. If $\lambda(t, A, D) = 0$, there is no packet in the system at time $t-$, and therefore no packet can depart at time t; the intensity of the point process of departures is 0 at time t. If $\lambda(t, A, D) = \mu$, there is at least one packet in the system at time $t-$. 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, when $\lambda(t, A, D) = \mu$, the intensity of the point process of departures also takes the value μ at time t. The process of departures is therefore a point process with intensity $(\lambda(t, A, D) : t \in [0, T_n])$.

Any window-code on the timing channel is therefore equivalent to the above strategy with complete feedback on the point-process channel with maximum intensity μ and no background intensity. Complete information about the past departures is necessary to determine the intensity of the departures at time t. The capacity of the timing channel (for window-codes) is therefore upperbounded by the capacity of the point-process channel with complete feedback. From the corollary to [32, Theorem 19.10, pp. 318-320] and the converse proofs in [6] and [7], this upperbound is $e^{-1}\mu$ nats per second. ■

Proof of Proposition 6(b): The process $(S_k : k \in \mathcal{N})$ of nominal service times is a sequence of independent and exponentially distributed random variables with mean $1/\mu_1$ seconds. Let $\mu_2 \in (0, \infty)$. Let $\mu \triangleq \mu_1\mu_2/(\mu_1 + \mu_2)$. Suppose that there were a sequence of $(n, M_n, T_n, \varepsilon_n)$-random window-codes that satisfies for some $\alpha > 0$, $(\log M_n)/T_n > e^{-1}\mu + \alpha$ for all sufficiently large n, and $\lim_{n\to\infty} \varepsilon_n = 0$. Then, for some ϕ and $\mathbf{C}$,

$$\sup_{\mathbf{z}\,:\,l(\mathbf{z})\le 1/\mu_2} E\left[P_e(\mathbf{C}, \phi, W^n(\mathbf{z}))\right] \le \varepsilon_n. \tag{2.35}$$

Choose μ_2' so that $1/\mu_2' < 1/\mu_2$ and $e^{-1}\mu > e^{-1}\mu' - \alpha/2$, where $\mu' \triangleq \mu_1\mu_2'/(\mu_1 + \mu_2')$. Let P_Z be the distribution given by

$$\begin{aligned} P_X\{Z = 0\} &= \mu'/\mu_1, \\ P_X\{Z > z\} &= \left(1 - \frac{\mu'}{\mu_1}\right) e^{-\mu' z}, \; z \ge 0. \end{aligned}$$

Note that Z has mean $1/\mu_2' = 1/\mu' - 1/\mu_1$ seconds. Furthermore, if S is independent of Z and exponentially distributed with mean $1/\mu_1$ seconds, then $S + Z$ is exponentially distributed with mean $1/\mu'$ seconds.

Let $\mathbf{Z} = (Z_1, \cdots, Z_n)$ be a vector of i.i.d random variables with common distribution P_Z, independent of the codebook distribution and the nominal service times. We then have

$$\begin{aligned} E\left[P_e(\mathbf{C}, \phi, W^n(\mathbf{Z}))\right] &\le \sup_{\mathbf{z}\,:\,l(\mathbf{z})\le 1/\mu_2} E\left[P_e(\mathbf{C}, \phi, W^n(\mathbf{z}))\right] + P_{\mathbf{Z}}\{l(\mathbf{Z}) > 1/\mu_2\} \\ &\overset{(a)}{\le} \varepsilon_n + P_{\mathbf{Z}}\{l(\mathbf{Z}) > 1/\mu_2\}, \end{aligned}$$

where (a) follows from (2.35). Let $\delta_n \triangleq \varepsilon_n + P_{\mathbf{Z}}\{l(\mathbf{Z}) > 1/\mu_2\}$. From the weak law of large numbers, we get $\lim_{n\to\infty} \delta_n = 0$ since $1/\mu_2' < 1/\mu_2$. Observe that $E\left[P_e(\mathbf{C}, \phi, W^n(\mathbf{Z}))\right]$ is also the expected probability of error (expectation over the codebook distribution) for the exponential server channel with mean service time $1/\mu'$ seconds. We can therefore find for this channel, a sequence of (n, M_n, T_n, δ_n)-window-codes with $(\log M_n)/T_n \geq e^{-1}\mu' + \alpha/2$ for all sufficiently large n, and $\lim_{n\to\infty} \delta_n = 0$. Since $e^{-1}\mu'$ nats per second is the largest rate achievable with window-codes, we reach a contradiction. ■

Chapter 3

Point Process Channels and Timing Channels

3.1 Introduction

We recall the following results from Chapter 1:

- The capacity of the exponential server queue 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. [6], [7]).
- In both channels, the capacity does not increase in the presence of complete feedback.

In [1], the connection between both channels in the presence of complete feedback was discussed briefly. In Chapter 2, this connection was further explored. We saw that any strategy on the exponential server channel can be mapped to an equivalent strategy that uses feedback on the point-process channel. This observation implies that the capacity of the exponential server channel is upperbounded by the capacity of the point-process channel with complete feedback, i.e., $e^{-1}\mu$ nats per second.

From [1] and Chapter 2, we know that $e^{-1}\mu$ nats per second is indeed achievable on the exponential server channel. The route taken in Chapter 2 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 single-server queue (cf. [1], [4], [5]). Not much is known however about 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], [4], [5] and Chapter 2 do not seem useful; using their techniques, it is difficult to write the likelihood function of the output given the input.

In this chapter, we take a point-process approach to timing channels. We restrict ourselves to exponential servers. We first show the direct part for the single-server queue using the point-process approach. This approach is then used to study timing channels in some simple networks of exponential servers. In particular, we obtain bounds, either analytically or from simulations, on the capacities of multiserver queues, the single-server queue with spurious departures, and a pair of queues connected in tandem.

In Section 3.2, we derive the capacity formula for the exponential server queue. We also find the capacity region when two or more users access this channel. We obtain lowerbounds on the capacity of multiserver queues. In Section 3.3, we study two examples to see how the point-process approach can be used. In the first example, the output of a single-server queue is merged with an independent stream of Poisson departures. The receiver cannot distinguish between the two types of departures. In the second example, we consider the channel where two queues with identical service rates are connected in tandem. In both these examples, we find achievable rates through simulations. Section 3.4 is a discussion of our results.

3.2 Single Queueing Station

The departures from the exponential queueing system form a point process whose rate depends on whether the queue is empty or not. If the queue is empty at a certain time, no departure can occur in the immediate future. If the queue is nonempty at a certain time, the residual time for the next departure is exponentially distributed with mean $1/\mu$ seconds. This observation was made in Chapter 2 to prove the converse for window-codes.

In this section, we re-derive the capacity formula for the exponential server queue using the above observation. We also find the capacity region when two or more users transmit on the channel. We give lower bounds on achievability rates for the $\cdot/M/m$ queue. We first give some mathematical preliminaries before we describe the channel in Section 3.2.2.

3.2.1 Preliminaries

Let $\overline{\mathcal{Z}}_+ = \{0, 1, \cdots, \infty\}$. Let $(\Omega, \mathcal{F})$ 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 $\mathcal{F}_t^X = \sigma\{X_s : s \in [0, t]\}$ for $t \in [0, T]$, and $\mathcal{F}_{t-}^X = \sigma\{X_s : s \in [0, t)\}$ for $t \in (0, T]$.

A stochastic process X is *predictable* with respect to $(\mathcal{F}_t : t \in [0, T])$, if X is measurable with respect to $([0, T] \times \Omega, \mathcal{P})$, where $\mathcal{P}$ is the σ-algebra generated by all left-continuous processes adapted to $(\mathcal{F}_t : t \in [0, T])$. Any adapted left-continuous process is therefore predictable.

The input space is $\left(\mathcal{X}, \mathcal{F}_T^X\right)$, where the alphabet $\mathcal{X}$ is the set of functions x :

$[0,T] \to \overline{\mathcal{Z}}_+$ that are non-decreasing and right-continuous with left limits. $\mathcal{X}$ 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 $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$, where $\mathcal{Y}$ is the set of functions $y : [0,T] \to \overline{\mathcal{Z}}_+$ that are non-decreasing, right-continuous with left limits, 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 ($(\Omega, \mathcal{F}_T)$ could be a larger measurable space than $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$). 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 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) $\int_0^t \lambda_s \, ds < \infty, \quad P - a.s$, 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 $\omega \in \Omega$. It is well-known that the quantities under the two expectations are random variables [33].

The above definition of rate process, though slightly more restrictive than the definition in [33], serves our purposes the best.

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

3.2.2 The $\cdot/M/1$ Queue

Channel Model

Let $P(x, dy)$ be the *transition probability function* [28, p.315] from the space $\left(\mathcal{X}, \mathcal{F}_T^X\right)$ to the space $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$, satisfying the following measurability properties: (a) for each $x \in \mathcal{X}$, the mapping $B \rightarrow P(x, B)$ from $\mathcal{F}_T^Y$ to $[0,1]$ is a probability measure on $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$, and (b) for each $B \in \mathcal{F}_T^Y$, the mapping $x \rightarrow P(x, B)$ 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 exponential server timing channel with information encoded in the arrival times of packets. 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 [33, Ex. 2.6] and [33, Ex. 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 $\left(\mathcal{F}_t^Q : t \in [0, T]\right)$.

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

$$Q_t = x_t - Y_t, \quad t \in [0, T]. \tag{3.1}$$

Fix arbitrary $t \in [0, T]$. If $Q_{t-} = 0$, then 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 $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$ is such that Y, the point process of departures in $[0,T]$, admits the rate process

$$\lambda = (\mu 1\{Q_{t-} > 0\} : t \in [0,T]) \tag{3.2}$$

with respect to $\left(\mathcal{F}_t^Y : t \in [0,T]\right)$, where Q_t is as defined in (3.1).

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

$$\frac{dP(x,\cdot)}{dP_0}(y) = p(x,y), \tag{3.3}$$

where

$$p(x,y) \triangleq \exp\left\{\int_0^T [\log(\lambda_t)\ dy_t + (1-\lambda_t)\ dt]\right\}. \tag{3.4}$$

The results [33, VI T2-T4], ensure that the point-process of departures Y admits the rate process (3.2) with respect to $\left(\mathcal{F}_t^Y : t \in [0,T]\right)$ under the probability measure $P(x,\cdot)$. The function $p(x,y)$ is measurable with respect to $\mathcal{F}_T^X \vee \mathcal{F}_T^Y$ [33], which implies that $P(x,dy)$ is a transition probability function. Note that (3.3) and (3.4) are related to the sample function density for a self-exciting point process with rate λ [34, Theorem 5.2.2], the difference being that (3.3) 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 element 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 $\lim_{n\to\infty} T_n = \infty$, $(\log M_n)/T_n \geq R - \gamma$ for all sufficiently large n, and $\lim_{n\to\infty} \varepsilon_n = 0$. The *capacity* is the largest achievable rate.

The codes considered in Chapter 2 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 is without such restrictions. As we will see the capacity of the single-server queueing system does not increase.

$P(x, dy)$, in addition to modeling the exponential server timing channel, is also the resulting transition probability function under the coding technique (3.2) 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 [6] 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 rate values on the set $\{0, \mu\}$.

Mutual Information

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

$$\pi(dx, dy) = \nu(dx) P_0(dy) p(x, y). \tag{3.5}$$

We can verify from Fubini's theorem [29, 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 $\left(\mathcal{F}_T^X \vee \mathcal{F}_t^Y : t \in [0, T]\right)$.

Let π^X and π^Y be the restrictions of π to $\mathcal{F}_T^X$ and $\mathcal{F}_T^Y$, respectively. From (3.5), we get that $\pi \ll \pi^X \times \pi^Y \ll \nu \times P_0$ [30, Corollary 5.3.1, p.112], and that $\pi^Y \ll P_0$. Furthermore, [33, VI R8] gives

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

where Y has rate $\hat{\lambda}$ with respect to $\left(\mathcal{F}_t^Y : t \in [0, T]\right)$ under the probability measure π^Y. More specifically, we can take $\hat{\lambda}_t = E\left[\lambda_t | \mathcal{F}_{t-}^Y\right]$, for each $t \in [0, T]$ [32, Theorem 18.3]. Thus the normalized information density $(1/T)\ i_T(x, y)$ is

$$\frac{1}{T} \log\left(\frac{d\pi}{d(\pi^X \times \pi^Y)}(x, y)\right) =$$
$$\frac{1}{T} \int_0^T \left[(\log \lambda_t)\, dy_t + (1 - \lambda_t) dt - \left(\log \hat{\lambda}_t\right) dy_t - \left(1 - \hat{\lambda}_t\right) dt\right]. \tag{3.6}$$

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

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

where $\phi(u) = u \log u$ (see [6], [32], [7], [33]). We take $\phi(0) = 0$. The function ϕ is strictly convex on $[0, \infty)$.

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 (3.2) and (3.1). 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.

Converse

The following converse was proved in [7], [6]. 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. (3.2) and (3.1)) is only a specific case.

Proposition 7: *([7], [6]) The capacity of the point-process channel with maximum intensity μ cannot exceed $e^{-1}\mu$ nats per second, even in the presence of complete feedback.*

Proof: This proof is taken from [7], [6]. 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.

$$\begin{aligned} \frac{1}{T} I_T(X;Y) &= \frac{1}{T} E \int_0^T dt \ \left[\phi(\lambda_t) - \phi\left(\hat{\lambda}_t\right)\right] \\ &\overset{(a)}{\leq} \frac{1}{T} E \int_0^T dt \ \phi(\lambda_t) - \phi\left(\frac{1}{T} E \int_0^T dt \ \hat{\lambda}_t\right) \\ &\overset{(b)}{=} \frac{1}{T} E \int_0^T dt \ \phi(\lambda_t) - \phi\left(\frac{1}{T} \int_0^T dt \ E\hat{\lambda}_t\right) \\ &\overset{(c)}{=} \frac{1}{T} E \int_0^T dt \ \phi(\lambda_t) - \phi\left(\frac{1}{T} \int_0^T dt \ E\lambda_t\right) \\ &\overset{(d)}{\leq} \max_{0 \leq a \leq \mu} \quad \max_{F: \int_0^\mu zF(dz) = a} \quad \int_0^\mu \phi(z) F(dz) - \phi(a) \\ &\overset{(e)}{=} \max_{0 \leq a \leq \mu} \quad \max_{F: \int_0^\mu zF(dz) = a} \quad \left(h(a) - \int_0^\mu h(z) F(dz)\right). \end{aligned} \tag{3.7}$$

Inequality (a) follows from Jensen's inequality applied to the strictly convex function ϕ. Equality (b) follows from Fubini's theorem. 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 $(1/T) I_T(X;Y) \leq e^{-1}\mu$.

■

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

Direct Part

Let $\mathcal{N} = \{1, 2, \cdots\}$. The *liminf in probability* of a sequence of random variables $(Z_n : n \in \mathcal{N})$ is the supremum of all reals α such that $\lim_{n\to\infty} P\{Z_n \leq \alpha\} = 0$. If the limit does not exist for any real α, we take the liminf in probability to be $-\infty$. It was shown in [35, Theorem 2] that the liminf in probability of the normalized information density is an achievable rate.

Proposition 8: *Fix finite $T > 0$. For the exponential server timing channel 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 $(T_n : n \in \mathcal{N})$ with $\lim_{n\to\infty} T_n = \infty$, the liminf in probability of the sequence $((1/T_n)\ i_{T_n}(X; Y) : n \in \mathcal{N})$ is $e^{-1}\mu$.*

Proof: From [1] and Chapter 2, we know that the maximum mutual information $e^{-1}\mu$ nats per second is attained using Poisson input following equilibrium at time $t = 0$. We now show this directly from the converse (3.7).

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., $\nu\{X_0 = k\} = (1 - a/\mu)(a/\mu)^k$ for $k \in \mathcal{Z}_+$. 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 (3.2) and (3.1). Let π be as defined in (3.5). 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 for e.g., [33, V T1]), which states that

$$\pi\left\{Q_t = k | \mathcal{F}_t^Y\right\} = \nu\left\{Q_0 = k\right\} \tag{3.8}$$

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 (3.7) 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

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

This implies that inequality (d) in (3.7) is an equality.

For inequality (a) to be an equality, we need $\hat{\lambda}_t = E\left[\lambda_t | \mathcal{F}_{t-}^Y\right] = a$, 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 $\mathcal{F}_{t-}^Y$ [31, 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. 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

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

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 (3.1). If a packet exits at time t, we must have that $\lambda_t = \mu 1\{Q_{t-} > 0\} = \mu$. The first integral is therefore $(Y_{T_n}/T_n) \log(\mu/a)$; it converges to $a \log(\mu/a)$ in probability. The second integral converges to 0 in probability because the time average of the quantity $\mu 1 \left\{Q_{t-} > 0\right\}$ converges in probability to a [36, Theorem 6.1]. By setting $a = e^{-1}\mu$, we get that the normalized information density converges in probability to $e^{-1}\mu$. ■

Remark: From this result and the converse, the capacity of the exponential server timing channel is $e^{-1}\mu$ nats per second. Moreover, the coding scheme utilizing feedback described in (3.2) and (3.1) achieves capacity on the point-process channel.

3.2.3 Multiuser Capacity Region

Suppose now that two users input packets to a single-server queue. 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 [37]. 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.

3.2.4 The $\cdot/M/m$ Queue

The model for the $\cdot/M/m$ queue timing channel is given by (3.1) and (3.3) with the rate $\lambda = (\lambda_t : t \in [0, T])$ being

$$\lambda_t = \mu \min\{m, Q_{t-}\}, \tag{3.9}$$

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 (3.9), 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 upperbounded by $m\mu/e$ nats per second (this follows straightforwardly as in (3.7)).

For the direct part, although inequality (a) in (3.7) can be made an equality by choosing Poisson input, inequality (d) in (3.7) 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 9: *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 8, with input rate a. We therefore have an $M/M/m$ queueing system. We now apply Burke's result for the $M/M/m$ queue [33, V T1] to get $\hat{\lambda}_t = a$ for every $t \in [0, T]$. Consequently, inequality (a) of (3.7) 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 $\lambda_t \in \{0, \mu, 2\mu, \cdots, m\mu\}$. 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 $p_{m,a}(i)$. Upon optimization over a, we obtain that $g(m)$ is the maximum mutual information under *Poisson arrivals*.

To show achievability, however, we need to show that the liminf in probability of the normalized information density is larger than $g(m)$. Fix any sequence of $(T_n : n \in \mathcal{N})$ such that $\lim_{n\to\infty} T_n = \infty$. Substitution of $\hat{\lambda}_t = a$ in (3.6) simplifies the normalized information density to

$$\frac{1}{T_n}i_{T_n}(x;y) = \frac{1}{T_n}\left[\int_0^{T_n}\log(\lambda_t/a)dy_t + \int_0^{T_n}(\lambda_t - a)dt\right]. \qquad (3.10)$$

The second integral converges in probability to 0 as in the proof of Proposition 8. The first integral can be written as

$$\left(\frac{y_{T_n}}{T_n}\right)\left(\frac{1}{y_{T_n}}\sum_{i=1}^{y_{T_n}}\log(\lambda_{\tau_k}/a)\right), \qquad (3.11)$$

where $\tau_1, \cdots, \tau_{y_{T_n}}$ denote the departure times in $[0, T_n]$. It is enough to show that the two terms within parenthesis in (3.11) converge in probability to appropriate constants. The convergence in probability of the first term to the constant a is clear. We now show the convergence in probability of the second term.

Observe that the process $(Q_{\tau_1 -}, Q_{\tau_2 -}, \cdots)$ of the number of packets in the system preceding the kth departure forms a discrete-time Markov chain on the state space $\mathcal{N}$. Indeed, we have $Q_{\tau_{k+1}-} = Q_{\tau_k -} - 1 + A_k$ for $k \in \mathcal{N}$, where A_k is the number of arrivals between departures k and $k+1$. The distribution of A_k does depend on the previous state $Q_{\tau_k -}$ because new arrivals can possibly change the departure rate. The value 0 for the state is impossible because at $\tau_k -$, there is at least one packet in the system. The Markov property follows from the memoryless property of the exponentially distributed interarrival times. Clearly, this Markov chain is irreducible, aperiodic, and starts at equilibrium. Furthermore, the equilibrium distribution is $p(i) = p_{m,a}(i-1)$ because the departing packets leave the queue in equilibrium.

The second term in (3.11) can be written as a weighted average of the m quantities $\log(i\mu/a)$, $i = 1, \cdots, m$. The weights are the fractions of departures for which the system is in one of $1, 2, \cdots, m-1$ or $\geq m$ states. These fractions converge in probability to the appropriate steady state probability values. The normalized information density (3.10) therefore converges in probability to

$$a \left[\sum_{i=1}^{m-1} p_{m,a}(i-1) \log(i\mu/a) + \left(\sum_{i=m}^{\infty} p_{m,a}(i-1) \right) \log(m\mu/a) \right]$$

Simple algebraic manipulation that uses the fact $p_{m,a}(i) = a \; p_{m,a}(i-1)/(i\mu)$ for $i = 1, \cdots, m-1$, results in the stated expression for $g(m)$. ■

Table 3.1 gives the achievable rate $g(m)$ in Proposition 9 as a function of m and compares it to the upperbound 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 second row.

Table 3.1: 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

Remarks: Setting $m = 1$ we get Proposition 8. Proposition 9 is thus a generalization. Note that the lowerbound for $m = 2$ and $m = 3$ give significant improvements over the single-server queue. For higher values of m, our lowerbound saturates at about 0.5804μ. The capacity of the $\cdot/M/\infty$ queue is however infinite. We therefore believe that both the upper and the lower bounds are quite loose. The capacity of the multiserver queue remains an interesting open problem.

3.3 Other Examples

In this section, we analyze two simple configurations of exponential servers and get bounds on the timing channel capacity. Our steps to find these bounds can be applied to other queueing systems too.

3.3.1 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. Figure 3.1). The departures from the two streams are indistinguishable to the receiver. In situations where the capacity of the single server queue is to be reduced, the above scheme that merges the departures from the queue with departures from another system is useful. It is therefore of interest to study the reduction in capacity offered by this scheme. The resulting model is similar to a point-process channel with background intensity α.

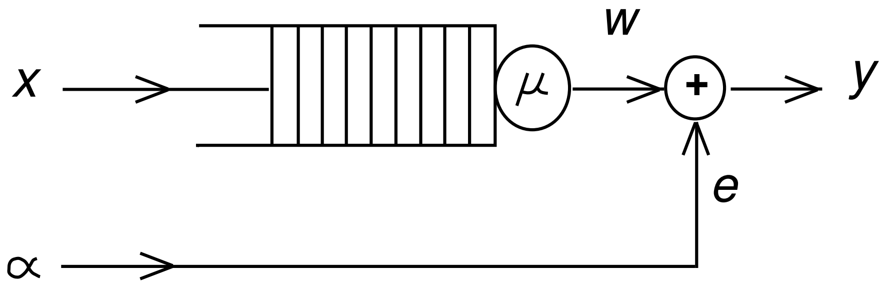

Figure 3.1: Single server queue with spurious departures

We assume single arrivals at the input. Fix $T \in (0, \infty)$. The input space is $\left(\mathcal{X}, \mathcal{F}_T^X\right)$, where $\mathcal{X}$ is the set of counting functions $x : [0, T] \rightarrow \overline{\mathcal{Z}}_+$ (non-decreasing and right-continuous) with unit jumps and $x_0 \in \mathcal{Z}_+$. Let $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$ be the output space where $\mathcal{Y}$ is the set of counting functions $y : [0, T] \rightarrow \overline{\mathcal{Z}}_+$ 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 $\mathcal{G}_t = \mathcal{F}_t^W \vee \mathcal{F}_t^E$ for $t \in [0, T]$.

Step 2: Observe that $\mathcal{F}_t^Y \subset \mathcal{G}_t$ for $t \in [0,T]$. The measure $P(x,\cdot)$ on $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$ that models the channel is represented by (cf. [33, VI R8])

$$\frac{dP(x,\cdot)}{dP_0}(y) = \exp\left\{\int_0^T \left[(\log \hat{\lambda}_t)dy_t + (1-\hat{\lambda}_t)dt\right]\right\}, \tag{3.12}$$

where Y has rate $\hat{\lambda}$ with respect to $\left(\mathcal{F}_t^Y : t \in [0,T]\right)$ under $P(x,\cdot)$. Furthermore, we may assume that $\hat{\lambda}$ satisfies (cf. [32, Theorem 18.3], [31, Theorem 1.6])

$$\begin{aligned} \hat{\lambda}_0 &= \alpha, \\ \hat{\lambda}_t &= \lim_{s\uparrow t} E\left[\alpha + \mu 1\{Q_s > 0\} \mid \mathcal{F}_s^Y\right], \quad t \in (0,T]. \end{aligned} \tag{3.13}$$

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 $E\left[1\{Q_t > 0\} \mid \mathcal{F}_t^Y\right]$ for $t \in [0,T]$. Substitution of this evaluation in (3.13) and (3.12) yields the transition probability function $P(x,dy)$ from $\left(\mathcal{X}, \mathcal{F}_T^X\right)$ to $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$. Given (3.12), the maximum-likelihood criterion for decoding is then straightforward.

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 $\hat{Z}_t(n) \triangleq E\left[1\{Q_t = n\}|\mathcal{F}_t^Y\right]$ for $n \in \mathcal{Z}_+$.

Proposition 10: *Consider the single-server queue with spurious departures. Fix $x \in \mathcal{X}$. The process $\left(\hat{Z}_t(n) : t \in [0,T]\right)$ for $n \in \mathcal{Z}_+$ can be recursively evaluated using the following update rules.*

(*a*) *Initialize $\hat{Z}_0(n) = 1\{x_0 = n\}$ for $n \in \mathcal{Z}_+$.*

(*b*) *If an arrival occurs at time τ, i.e., $dx_\tau = x_\tau - x_{\tau-} = 1$, then*

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

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

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

(*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*

$$\begin{aligned} \hat{Z}_t(0) &= \frac{\hat{Z}_{\tau_k}(0)\ \exp\{\mu(t-\tau_k)\}}{\hat{Z}_{\tau_k}(0)\ \exp\{\mu(t-\tau_k)\} + \left(1 - \hat{Z}_{\tau_k}(0)\right)}, \\ \hat{Z}_t(n) &= \frac{\hat{Z}_{\tau_k}(n)}{\hat{Z}_{\tau_k}(0)\ \exp\{\mu(t-\tau_k)\} + \left(1 - \hat{Z}_{\tau_k}(0)\right)}, \quad n \in \mathcal{N}. \end{aligned}$$

Proposition 10 solves $\left(\hat{Z}_t(n) : t \in [0,T]\right)$ explicitly for $n \in \mathcal{Z}_+$. Such explicit solutions are however hard to obtain in most cases because of the difficulty in solving a system of non-linear 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 3.3.2 are two exceptions where an explicit solution can indeed be found.

Step 4: We use [33, VI R8] to obtain an expression for the mutual information. We proceed as in Section 3.2.2. Let ν be any probability measure on $\left(\mathcal{X}, \mathcal{F}_T^X\right)$. The input measure ν and the transition probability function $P(x, dy)$ in (3.12) define the joint probability measure π on $\left(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_T^X \vee \mathcal{F}_T^Y\right)$, 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 $\left(\mathcal{F}_T^X \vee \mathcal{F}_t^Y : t \in [0,T]\right)$.

With π^X and π^Y, 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, [33, VI R8] gives

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

where Y has rate $\hat{\hat{\lambda}}$ with respect to $\left(\mathcal{F}_t^Y : t \in [0,T]\right)$ under the probability measure π^Y. We may take $\hat{\hat{\lambda}}_t = E\left[\hat{\lambda}_t | \mathcal{F}_{t-}^Y\right]$, for each $t \in [0,T]$ [32, Theorem 18.3]. The normalized information density $(1/T)\ i_T(x,y)$ is therefore given by

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

the normalized mutual information can be written as

$$\begin{aligned}\frac{1}{T}I_T(X;Y) &= \frac{1}{T}E\int_0^T \left[\log\left(\hat{\lambda}_t/\hat{\hat{\lambda}}_t\right)\right] dy_t && (3.14)\\ &= \frac{1}{T}E\int_0^T dt \left[\phi\left(\hat{\lambda}_t\right) - \phi\left(\hat{\hat{\lambda}}_t\right)\right], && (3.15)\end{aligned}$$

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

Step 5: To find achievable information rates, we choose an appropriate input distribution ν and evaluate the normalized mutual information under π. To do this we need to evaluate $\hat{\hat{\lambda}} = \left(\hat{\hat{\lambda}}_t : t \in [0,T]\right)$.

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

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

for $k \in \mathcal{Z}_+$, 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

$$\begin{aligned}\hat{\hat{\lambda}}_t &= E\left[\hat{\lambda}_t | \mathcal{F}^Y_{t-}\right]\\ &= E\left[E\left[\lambda_t | \mathcal{F}^X_T \vee \mathcal{F}^Y_{t-}\right] | \mathcal{F}^Y_{t-}\right]\\ &= E\left[\lambda_t | \mathcal{F}^Y_{t-}\right] && (3.16)\\ &= \alpha + \mu E\left[1\{Q_{t-} > 0\} | \mathcal{F}^Y_{t-}\right]\\ &= \alpha + a, && (3.17)\end{aligned}$$

where (3.16) follows from [29, Theorem 34.4], and (3.17) follows from the fact that Q_s and (E_s, D_s) are independent, which implies that Q_s and Y_s are independent. Substitution of (3.17) in (3.14) yields

$$\frac{1}{T}I_T(X;Y) = E\left[\left(\frac{Y_T}{T}\right)\frac{1}{Y_T}\sum_{i=1}^{Y_T}\log\left(\frac{\hat{\lambda}_t}{\alpha + a}\right)\right]. \qquad (3.18)$$

Although we are unable to evaluate this normalized mutual information analytically, (3.18) is amenable to numerical evaluation through simulations. The results are plotted in the Figure 3.2.

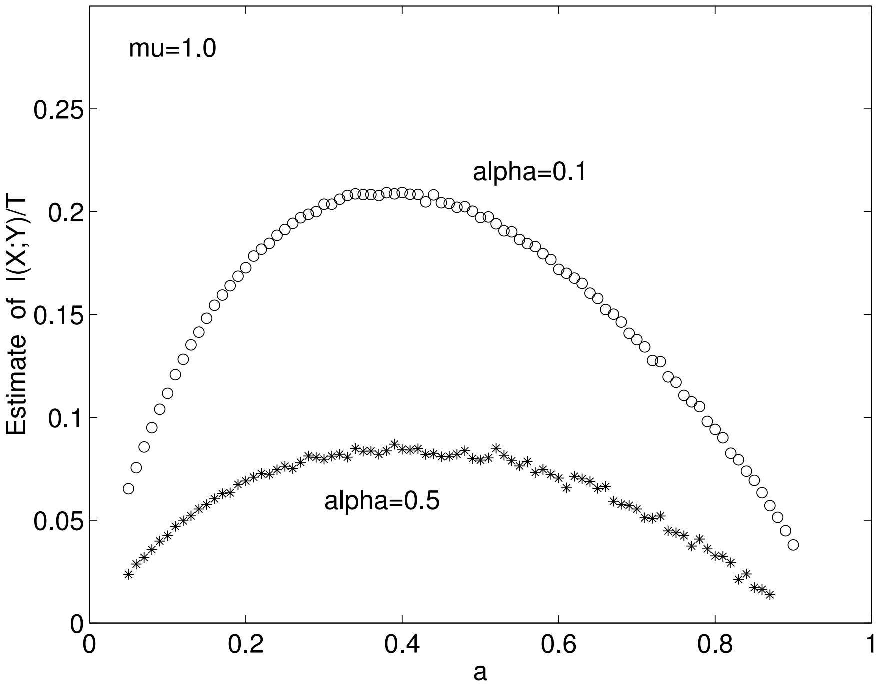

Figure 3.2: In this figure $\mu = 1$ packet per second. The results are for $\alpha = 0.1$ and $\alpha = 0.5$ packets per second. The abscissa is the input rate to the queue. The ordinate is the estimate of $(1/T)I(X;Y)$ in (3.18), the normalized mutual information for the single-server queue with spurious departures. The reported value for each a is an average of 500 values of $(1/T)i(x;y)$, 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 upperbounds C_α are 0.251 nats per second and 0.132 nats per second, respectively.

Step 6: We can find an upperbound on the capacity of the single-server queue with spurious departures. Recall that $P(x, dy)$ (cf. (3.12)) is the transition probability function for the self-exciting point process having rate $\hat{\lambda}$ (cf (3.13)), 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 complete feedback, and where $\hat{\lambda}_t \in [\alpha, \alpha + \mu]$, is (cf. [6], [7])

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

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

3.3.2 Two Queues in Tandem

In Section 3.3.1 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. Figure 3.3).

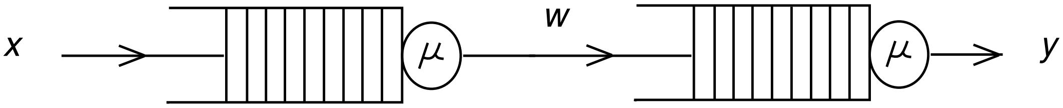

Figure 3.3: Two exponential-server queues in tandem

Step 1: Let $\left(\mathcal{X}, \mathcal{F}_T^X\right)$ and $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$ be as in Section 3.3.1. 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 state of the queue at any time $t \in [0,T]$ is $\left(Q_t^{(1)}, Q_t^{(2)}\right)$, where $Q_t^{(1)} = x_t - W_t$ and $Q_t^{(2)} = W_t - Y_t$ for $t \in [0,T]$.

We choose $\mathcal{G}_t = \mathcal{F}_t^W \vee \mathcal{F}_t^Y$ for $t \in [0,T]$. Then for each fixed $x \in \mathcal{X}$, the process of departures Y is a point process having rate $\lambda = \left(\mu 1\{Q_{t-}^{(2)} > 0\} : t \in [0,T]\right)$ 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 (3.12), where we take $\hat{\lambda}_0 = 0$ and for $t \in (0,T]$,

$$\begin{aligned} \hat{\lambda}_t &= E\left[\lambda_t | \mathcal{F}_{t-}^Y\right] \\ &= \mu \lim_{s \uparrow t} E\left[1\{Q_s^{(2)} > 0\} | \mathcal{F}_s^Y\right]. \end{aligned} \tag{3.19}$$

Step 3: The estimates for the queue states can be evaluated from Proposition 11 below. Let

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

and

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

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

Proposition 11: *Consider two queues in tandem with identical service rates of μ packets per second. Fix $x \in \mathcal{X}$. The process $\left(\hat{Z}_t(n_1,n_2) : t \in [0,T]\right)$ can be recursively evaluated using the following update rules.*

(a) Initialize $\hat{Z}_0(n_1,n_2) = 1\{x_0 = n_1,\ w_0 = n_2\}$ for $(n_1,n_2) \in \mathcal{Z}_+^2$.

(b) If an arrival occurs at time τ, i.e., $dx_\tau = 1$, then

$$\hat{Z}_\tau(n_1,n_2) = \hat{Z}_{\tau-}(n_1-1,n_2)1\{n_1>0\}, \quad (n_1,n_2) \in \mathcal{Z}_+^2.$$

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

$$\hat{Z}_\tau(n_1,n_2) = \frac{\hat{Z}_{\tau-}(n_1,n_2+1)}{1-\hat{Z}_{\tau-}(n_1+n_2+1,0)}, \quad (n_1,n_2) \in \mathcal{Z}_+^2.$$

(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,

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

for $1 \le i \le n-1$, and

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

Step 4: An input measure ν on $\left(\mathcal{X}, \mathcal{F}_T^X\right)$ and the transition probability function $P(x, dy)$ from $\left(\mathcal{X}, \mathcal{F}_T^X\right)$ to $\left(\mathcal{Y}, \mathcal{F}_T^Y\right)$ define the joint measure π on $\left(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_T^X \vee \mathcal{F}_T^Y\right)$ as before. With other analogous definitions, the normalized mutual information $(1/T)I_T(X;Y)$ is given by (3.14) and (3.15).

Step 5: As expected we can evaluate $\hat{\hat{\lambda}}_t$ when ν is such that the arrivals are Poisson with rate $a < \mu$ in $(0, T]$, and the queue size is in equilibrium at time $t = 0$. Under these circumstances, $\hat{\hat{\lambda}}_t = a$ for each $t \in [0, T]$, which leads to

$$\frac{1}{T} I_T(X;Y) = E\left[\left(\frac{Y_T}{T}\right) \frac{1}{Y_T} \sum_{i=1}^{Y_T} \log\left(\frac{\hat{\lambda}_t}{a}\right)\right]. \tag{3.20}$$

Yet again, being unable to evaluate this expectation analytically, we obtain estimates for the normalized mutual information from simulations. The results are shown in Figure 3.4. The capacity of this system is an open problem.

3.3.3 Proofs of Propositions 10 and 11

Proof of Proposition 10: Recall that the process $w = (w_t : t \in [0,T])$ has rate $(\mu(1 - Z_{s-}(0)) : t \in [0,T])$ with respect to $(\mathcal{G}_t : t \in [0,T])$. Moreover, we can write

$$\begin{aligned}
Z_t(n) = Z_0(n) \;&+\; \int_0^t \left[Z_{s-}(n-1)1\{n>0\} - Z_{s-}(n)\right] dx_s \\
&+\; \int_0^t \left[Z_{s-}(n+1) - Z_{s-}(n)1\{n>0\}\right] dw_s \\
= Z_0(n) \;&+\; \int_0^t \left[Z_{s-}(n-1)1\{n>0\} - Z_{s-}(n)\right] dx_s \\
&+\; \int_0^t \left[Z_{s-}(n+1) - Z_{s-}(n)1\{n>0\}\right] \mu(1 - Z_{s-}(0))\, ds \\
&+\; u_t(n),
\end{aligned} \tag{3.21}$$

where

$$u_t(n) \triangleq \int_0^t \left[Z_{s-}(n+1) - Z_{s-}(n)1\{n>0\}\right] \left[dw_s - \mu(1 - Z_{s-}(0))\, ds\right]. \tag{3.22}$$

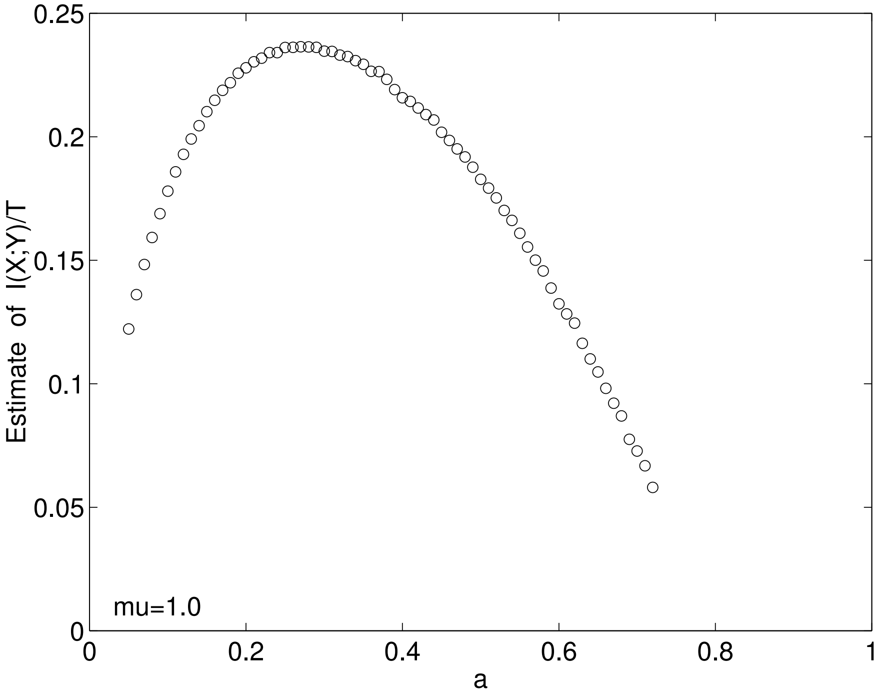

Figure 3.4: In this figure $\mu = 1$ packet per second. The abscissa is the input rate to the first queue. The ordinate is the estimate of $(1/T)I(X;Y)$ in (3.20), the normalized mutual information between the input to the first queue and the output of the second queue. The reported value for each a is an average of 200 values of $(1/T)i(x;y)$, 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.

Observe that

$$[Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}]\,(1 - Z_{s-}(0)) = [Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\}] \tag{3.23}$$

for every $n \in \mathcal{Z}_+$. Furthermore, the process $Z(n) = (Z_t(n) : t \in [0,T])$ is bounded (in fact, either 0 or 1). From [33, II T8(β)], the process $u(n) = (u_t(n) : t \in [0,T])$ is a $(\mathcal{G}_t : t \in [0,T])$ martingale because

$$E\left[\int_0^t (Z_{s-}(n+1) - Z_{s-}(n)1\{n > 0\})\,\mu\,(1 - Z_{s-}(0))\,ds\right] < 2\mu t$$

for every $t \in [0,T]$. Furthermore, $u(n)$ is almost surely of bounded variation on $[0,T]$. It follows from (3.21), (3.23) and the proofs of [33, IV T8] and [32, Theorem 19.5], that

$$\begin{aligned} \hat{Z}_t(n) = \hat{Z}_0(n) \quad &+ \int_0^t \left[\hat{Z}_{s-}(n-1)1\{n>0\} - \hat{Z}_{s-}(n)\right] dx_s \\ &+ \int_0^t \left[\hat{Z}_{s-}(n+1) - \hat{Z}_{s-}(n)1\{n>0\}\right] \mu ds \\ &+ \int_0^t \left[\Psi_{1,s}(n) + \Psi_{2,s}(n) - \hat{Z}_{s-}(n)\right]\left[dy_s - \hat{\lambda}_s ds\right], \end{aligned} \tag{3.24}$$

where $\hat{\lambda}_s = \alpha + \mu\left(1 - \hat{Z}_{s-}(0)\right)$ for $s \in [0,T]$, and $(\Psi_{i,t} : t \in [0,T])$, $i = 1,2$, are predictable with respect to $\left(\mathcal{F}_t^Y : t \in [0,T]\right)$, and satisfy

$$\begin{aligned} E\left[\int_0^t C_s Z_s(n)\lambda_s \, ds\right] &= E\left[\int_0^t C_s \Psi_{1,s}(n)\hat{\lambda}_s \, ds\right], \\ E\left[\int_0^t C_s \, \Delta u_s(n) \, dy_s\right] &= E\left[\int_0^t C_s \Psi_{2,s}(n)\hat{\lambda}_s \, ds\right]. \end{aligned}$$

It is easy to verify that

$$\begin{aligned} \Psi_{1,s}(n) &= \frac{\hat{Z}_{s-}(n)(\alpha + \mu 1\{n>0\})}{\alpha + \mu\left(1 - \hat{Z}_{s-}(0)\right)}, \\ \Psi_{2,s}(n) &= \frac{\mu\left(\hat{Z}_{s-}(n+1) - \hat{Z}_{s-}(n)1\{n>0\}\right)}{\alpha + \mu\left(1 - \hat{Z}_{s-}(0)\right)}. \end{aligned} \tag{3.25}$$

Substitution of (3.25) in (3.24) leads to

$$\begin{aligned} \hat{Z}_t(n) = \hat{Z}_0(n) \quad &+ \int_0^t \left[\hat{Z}_{s-}(n-1)1\{n>0\} - \hat{Z}_{s-}(n)\right] dx_s \\ &+ \int_0^t \hat{Z}_{s-}(n)\left(1\{n=0\} - \hat{Z}_{s-}(0)\right) \mu ds \\ &+ \int_0^t \mu \left[\frac{\hat{Z}_{s-}(n+1) - \hat{Z}_{s-}(n)\left(1 - \hat{Z}_{s-}(0)\right)}{\alpha + \mu\left(1 - \hat{Z}_{s-}(0)\right)}\right] dy_s. \end{aligned} \tag{3.26}$$

The update rules (a), (b) and (c) follow straightforwardly from (3.26). 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

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

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

Proof of Proposition 11: We give only a brief outline of the proof; most of the steps are analogous to the proof of Proposition 10. The state $Z_t(n_1, n_2)$ can be represented by

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

This leads to the following equation for the estimates,

$$\begin{aligned} \hat{Z}_t(n_1,n_2) = \hat{Z}_0(n_1,n_2) + \int_0^t &\quad \left[\hat{Z}_{s-}(n_1-1,n_2)1\{n_1>0\} - \hat{Z}_{s-}(n_1,n_2)\right] dx_s \\ + \int_0^t &\quad \left[-\hat{Z}_{s-}(n_1,n_2) + \frac{\hat{Z}_{s-}(n_1,n_2+1)}{1-\hat{Z}_{s-}(\cdot,0)}\right] dy_s \\ + \int_0^t &\quad \Big[\hat{Z}_{s-}(n_1,n_2)\left(1-\hat{Z}_{s-}(\cdot,0)\right) \\ &\quad -\hat{Z}_{s-}(n_1,n_2)1\{n_2>0\} - \hat{Z}_{s-}(n_1,n_2)1\{n_1>0\} \\ &\quad +\hat{Z}_{s-}(n_1+1,n_2-1)1\{n_2>0\}\Big]\, \mu ds, \end{aligned} \tag{3.27}$$

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

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

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

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

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

$$\frac{1}{\mu}\frac{d\hat{Z}_t(n-i,i)}{dt} = -\hat{Z}_t(n-i,i)\left(1+\hat{Z}_t(n,0)\right) + \hat{Z}_t(n-i+1,i-1).$$

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

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

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

3.4 Discussion

We gave a conceptually simple proof of the capacity of the single-server queue (Proposition 8). Our proof emphasizes the connection between the point-process channel and the single-server queue. We also observed that the capacity region of the two-user single-server queue is a triangle. We then showed a lowerbound on the capacity of multiserver queues (Proposition 9 and Table 3.1) by optimizing the mutual information over Poisson inputs. We however do not know the capacity of such queues.

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 10-11, (3.13), (3.19) and (3.12)). 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 an integral equation for updating the estimates for queue sizes along the lines of (3.26) and (3.27); we are however unable to explicitly solve them as in Propositions 10-11.

In the two examples considered, because we were unable to evaluate the mutual information analytically, we resorted to evaluations based on simulations. For the single-server queue with spurious departures of rate 0.1μ, our simulations indicated that the capacity is at least 0.21μ nats per second (Figure 3.2). We showed that the capacity is upperbounded by C_α which is 0.251μ nats per second. For the pair of queues connected in tandem, our simulations indicated that the capacity is at least 0.23μ nats per second (Figure 3.4). The capacities of these channels are however not known.

Chapter 4

Sequential Decoding for the Exponential Server Timing Channel

4.1 Introduction

In Chapters 2 and 3, we concentrated on obtaining rates that are achievable when no constraint is placed on the complexity of the encoder and the decoder. To build practical communication systems, however, we need coding schemes where the decoder has good performance while being computationally feasible. In this chapter, we see that tree coding schemes can transmit reliably at rates below half the capacity on the single-server queue while maintaining computational feasibility.

Sequential decoding of convolutional codes and tree codes ([38], [39], [40], [41], [42], etc.) is a useful decoding technique wherein the average number of computations performed is linear in block length as compared to an exponential number of computations for the maximum-likelihood decoder. A vast majority of the literature on sequential decoding deals with memoryless channels. A few papers, (for e.g., [43], [44]) extend the sequential decoding technique to a class of channels with memory, namely, finite-state channels. In this chapter we show that the sequential decoding technique can be used on timing channels (for e.g., [1] and [4]). Interestingly, this timing channel is a channel with memory and cannot be described within the class of finite-state channels.

Specifically, we want to transmit information reliably through a single-server queue [1], [4], at rates below *half* the capacity, but with manageable decoding complexity. In [1], [4] and Chapter 2 a decoding technique for block codes was described where the number of computations is exponential in n, the number of packets. By imposing a tree structure on the codes and using the sequential decoding technique, we save on computations at the expense of the rate at which information is reliably transmitted. This chapter is perhaps a first step in the direction of finding good codes for communication over timing channels.

There are many versions of the sequential decoding technique. The basic idea behind the Fano algorithm [40] is to move forward in the decoding tree so long as we seem to be (based on a metric) on the right track. Once the metric falls below a certain threshold, we backtrack and explore other paths, possibly changing the value of the threshold to account for the changed circumstances. The stack algorithm [41], [42], extends the node with the highest metric at each stage, until the end of the tree is reached. There is a relation between the number of computations in both these algorithms.

We are interested in finding bounds on the average number of computations before proceeding one step forward in the correct path. The difficulty with analyzing the performance of the sequential decoding technique for communication systems with memory is the following. When comparing two paths that are the same up to a certain node, the choice of one or the other depends on the branches common to both paths in a way that is typically difficult to handle. For memoryless channels, however, the metric that determines this choice can be selected so that the choice does not depend on the common branches.

We can also get over this difficulty for timing channels. We show that the first m branches can be summed up by one quantity that lends itself to a simple analysis. Our proof is based on the proof in [39] for multiple-access channels, restricted to single-user channels. Burke's output theorem for an $M/M/1$ queue plays an important role in determining a suitable metric. The main contributions of this chapter are the choice of this metric, and a simple analytical artifice (used earlier in [1] in a different context) that shows how the elegant technique in [39] can be modified to prove the existence of a good tree code for this system with memory.

Section 4.2 introduces the problem in the appropriate notation and states the result. Section 4.3 contains the proof. We conclude with a brief discussion in Section 4.4.

4.2 Tree Codes for Single-Server Queue

Before describing the tree code and our result, we briefly describe the channel. The queue is initially empty. The encoder inputs a certain (non-zero) number of packets at time $t = 0$. The last packet input at time $t = 0$ is called the *zeroth* packet. Let y_0 be the time at which the zeroth packet exits the queue after service. The quantity y_0 is therefore the amount of unfinished work at time $t = 0$. Depending on the message to be transmitted, the encoder then sends the first packet at time x_1 seconds, the second packet at time x_2 after the first packet, and so on. Thus the interarrival times of packets are $x_1, x_2, \cdots$. The receiver observes the interdeparture times, $y_1, y_2, \cdots$, following the departure of the zeroth packet. Let $\mathcal{R}_+ = [0, \infty)$. Let $e_\mu(s) = \mu e^{-\mu s}$, $s \in \mathcal{R}_+$. The conditional probability density of the output $y^n = (y_1, \cdots, y_n)$ given x^n and y_0 is

$$f_\mu(y^n | x^n, y_0) = \prod_{i=1}^{n} e_\mu(y_i - w_i), \tag{4.1}$$

where

$$w_i = \max\left\{0, \sum_{j=1}^{i} x_j - \sum_{j=0}^{i-1} y_j\right\} \tag{4.2}$$

is the server's idling time before serving the ith packet.

We now describe the tree code. We follow the notation in [39] with a few modifications. At each instant of time t, the source generates a letter $u_t \in \{0, 1, \cdots, M-1\}$, and the sequence $\mathbf{u} = (u_1, u_2, \cdots)$ is encoded by a tree code $\mathbf{g}$. The tree $\mathbf{g}$ is such that M edges leave each node of the code tree. Each edge is labelled by an N-tuple of nonnegative real numbers. The root node is labelled by the number of packets input at time $t = 0$ including the zeroth packet. We denote by $u^t = (u_1, u_2, \cdots, u_t)$, the path leading from the root node to the tth level. The code corresponding to the source sequence u^t is given by $x^{Nt}(u^t) \in \mathcal{R}_+^{Nt}$, where

$$x^{Nt}\left(u^t\right) = \left(x_1\left(u^1\right), \cdots, x_N\left(u^1\right), x_{N+1}\left(u^2\right), \cdots, x_{Nt}\left(u^t\right)\right)$$

is the sequence of interarrival times of the Nt packets for message sequence u^t. Furthermore, we denote the entire codeword corresponding to the source sequence $\mathbf{u}$ by

$$\mathbf{x}(\mathbf{u}) = \left(x_1\left(u^1\right), \cdots, x_N\left(u^1\right), x_{N+1}\left(u^2\right), \cdots\right).$$

The source sequence from m to l is defined to be

$$u_m^l = (u_m, u_{m+1}, \cdots, u_l).$$

Similarly we define

$$x_{Nm+1}^{Nl}\left(u^l\right) = \left(x_{Nm+1}\left(u^{m+1}\right), \cdots, x_{Nl}\left(u^l\right)\right).$$

The set of all paths in $\mathbf{g}$ that diverge from $\mathbf{u}$ at the mth level is called the mth incorrect subtree for the path $\mathbf{u}$, i.e.,

$$\mathcal{U}_m(\mathbf{u}) = \{\hat{\mathbf{u}} = (u_1, \cdots, u_{m-1}, \hat{u}_m, \hat{u}_{m+1}, \cdots) : \hat{u}_m \neq u_m\}.$$

Let $\mathbf{g}$ be a tree code. We characterize the source as follows. The source sequence $\mathbf{U} = (U_1, U_2, \cdots)$ is an independent and identically distributed (i.i.d) sequence of random variables where each source letter U_t is uniformly distributed on the set $\{0, 1, \cdots, M-1\}$. The tree code $\mathbf{g}$ then transmits information at a rate R nats per unit time, where

$$R = \lim_{t\to\infty} \frac{\log M^t}{E[\sum_{i=0}^{Nt} Y_i]}, \tag{4.3}$$

if the limit exists. The quantity $E\left[\sum_{i=0}^{Nt} Y_i\right]$ is the average time to receive the Nt packets, when the tree code is $\mathbf{g}$. This rate can also be written as

$$R = \left(\frac{\log M}{N}\right) / \left(\lim_{t\to\infty} E\left[\frac{1}{Nt}\sum_{i=0}^{Nt} Y_i\right]\right).$$

The quantity $r = (\log M)/N$ depends only on the structure of the tree, and is a measure of the number nats of information transmitted per packet.

We now define the metric. This metric depends on the quantity r. Fix $0 < \lambda < \mu/2 = \mu'$. We take

$$\Gamma\left(u^t | y_0, y^{Nt}\right) = M\left(x^{Nt}\left(u^t\right) | y_0, y^{Nt}\right), \tag{4.4}$$

where

$$M(x^n | y_0, y^n) \triangleq \log\left(\frac{f_{\mu'}(y^n | x^n, y_0)}{\prod_{i=1}^n e_\lambda(y_i)}\right) - nr(1+\varepsilon). \tag{4.5}$$

The bias term $nr(1+\varepsilon)$ in (4.5) is to make a fair comparison between paths of different lengths. $M(\cdot|\cdot,\cdot)$ in (4.5) is similar to the metric in [39, Equations (4.1), (4.4)]. Note the dependence on the quantity μ' rather than μ. This is because the quantity $\sqrt{f_\mu(\cdot|\cdot,\cdot)}$ which determines the metric [39, Equation (4.4)] normalizes to $f_{\mu'}(\cdot|\cdot,\cdot)$.

The function M in (4.5) is further related to [39, Equation (4.4)] due to the following special case of Burke's output theorem [9, Fact 2.8.2, p.60]. Let $\lambda < \mu'$. Let the number of packets Q_0 at time $t = 0$, excluding the zeroth packet, be distributed according to $\Pr\{Q_0 = k\} = (1 - \lambda/\mu')(1 - \lambda/\mu')^k$, $k \in \mathcal{Z}_+$. In addition to these packets, the zeroth packet is sent. Thus the zeroth packet sees the queue in steady state upon arrival. Let the arrivals thereafter form a Poisson process of rate λ. The zeroth packet departs the queue at time Y_0, whose probability density is $e_{\mu'-\lambda}$. Furthermore, at the moment of its departure, the queue is in equilibrium. The output starting from time Y_0 is then a Poisson process of rate λ [9, Fact 2.8.2, p.60]. In other words,

$$E\left[f_{\mu'}(y^n|X^n, Y_0)\right] = \prod_{i=1}^{n} e_\lambda(y_i), \tag{4.6}$$

where the expectation is with respect to X^n and Y_0. X^n is a random vector of i.i.d exponential random variables with mean $1/\lambda$ seconds, Y_0 is independent of X^n, and is exponentially distributed with mean $1/(\mu' - \lambda)$. The right hand side of (4.6) is the normalizing denominator within the log function in (4.5).

The decoder follows the stack algorithm. From a stack containing some paths in $\mathbf{g}$, the decoder selects a path with the largest metric, extends it to the next level in M possible ways and stores the M new paths in the stack. A sorting is done as soon as the new paths are added. The stack algorithm terminates for a tree code with finite depth as soon as the last level of the tree reaches the stack top. As mentioned in [38] we shall consider only infinite trees because the average complexity of sequential decoding is most cleanly formalized and conservatively estimated in the framework of infinite trees. For finite trees, we also need to evaluate the probability of error, which occurs when the last level of the tree to reach the stack top is not the correct message sequence. The proof in Section 4.3 applies to finite tree codes with simple modifications.

Using (4.1), the first term in the right hand side of (4.5) can be expanded as

$$\log\left(\frac{f_{\mu'}(y^n|x^n, y_0)}{\prod_{i=1}^{n} e_\lambda(y_i)}\right) = \begin{cases} n\log\frac{\mu'}{\lambda} \quad - \quad (\mu'-\lambda)\sum_{i=1}^{n} y_i \quad + \quad \mu'\sum_{i=1}^{n} w_i, \\ \qquad\qquad \text{if } y_i \geq w_i, \text{ for } i = 1, \cdots, n, \\ -\infty, \qquad \text{otherwise,} \end{cases} \tag{4.7}$$

where w_i is the idling time defined in (4.2).

We now make the following important observation. Suppose we compare two paths of lengths j and l, respectively, that are identical for the first $m-1$ nodes and diverge at the mth node. The past up to the first $m-1$ nodes can be summarized by one quantity,

$$\tilde{y}_{m-1} = \sum_{i=0}^{N(m-1)} y_i \quad - \sum_{i=1}^{N(m-1)} x_i.$$

This quantity $\tilde{y}_{m-1}$ is the amount of unfinished work at the instant when the $N(m-1)$st packet arrives. To decide which of the two paths is placed higher on the stack, we can simply treat the $(m-1)$st node as the root node with $\tilde{y}_{m-1}$ playing the role of y_0. The terms in (4.7) common to both paths are the same up to the $(m-1)$st node. Furthermore, the w_i's for branches from node m and beyond are unchanged with $\tilde{y}_{m-1}$ in place of y_0. This is because, for $k > N(m-1)$, we can rewrite (4.2) as

$$w_k = \max\left\{0, \sum_{i=N(m-1)+1}^{k} x_i \quad - \sum_{i=N(m-1)+1}^{k-1} y_i \; - \; \tilde{y}_{m-1}\right\}.$$

Thus the path metric depends on the common nodes only through the unfinished work at the instant of the arrival of the last common packet. This observation is summarized by

$$\Gamma\left(u^j | y_0, y^{Nj}\right) = \Gamma\left(u^{m-1} | y_0, y^{N(m-1)}\right) + \Gamma\left(u_m^j | \tilde{y}_{m-1}, y_{N(m-1)+1}^{Nj}\right), \tag{4.8}$$

if $j \geq m$. Of course, Γ for the root node is taken to be 0. Comparing $\Gamma\left(u^j | y_0, y^{Nj}\right)$ and $\Gamma\left(u^l | y_0, y^{Nl}\right)$, where $l, j \geq m$, and when the two source sequences have identical initial $m-1$ branches, is therefore equivalent to comparing

$$\Gamma\left(u_m^j | \tilde{y}_{m-1}, y_{N(m-1)+1}^{Nj}\right) \text{ and } \Gamma\left(u_m^l | \tilde{y}_{m-1}, y_{N(m-1)+1}^{Nl}\right).$$

Let $C_m(\mathbf{g}, \mathbf{u}, \mathbf{y})$ denote the number of nodes in $\mathcal{U}_m(\mathbf{u})$ that reach the top of the stack for a given tree code $\mathbf{g}$ and a received sequence $\mathbf{y}$. This is precisely the number of computations made in the mth incorrect subtree. Let

$$C_m(\mathbf{g}) = E\left[C_m(\mathbf{g}, \mathbf{U}, \mathbf{Y})\right]$$

be the average number of computations (averaged over the source sequence and output of the channel). The random variables over which the expectation is taken are indicated in uppercase letters. For each $L \geq 1$, let

$$D_L(\mathbf{g}) \triangleq \frac{C_1(\mathbf{g}) + \cdots + C_L(\mathbf{g})}{L}.$$

$D_L(\mathbf{g})$ is therefore a measure of the average number of computations required to move one step ahead on the correct path [38].

Proposition 12: *For every $\delta > 0$, there exists a tree code $\mathbf{g}$ and a constant $A < \infty$ such that the rate of information transfer is R nats per second where $R(1+\delta) > \mu/(2e)$, and $D_L(\mathbf{g}) \leq A$ for every $L \geq 1$.*

4.3 Proof of Proposition 12

4.3.1 Main Steps

Our proof technique to show the existence of a good tree code with sequential decoding is the well-known random coding technique. A tree is characterized by the number of packets at time $t = 0$, and the labels for all the branches. A suitable distribution on these quantities induces a distribution on the set of infinite trees (using extension theorems in probability theory). We state some bounds over this ensemble of trees and thence argue the existence of a good tree. We then prove the stated bounds in the following subsection.

Choose $\varepsilon > 0$ so that $(1+\delta) > (1+\varepsilon)^3$. Fix $\lambda = e^{-1}\mu' = \mu/(2e)$. Fix M and N so that $r = (\log M)/N$ satisfies $r(1+\varepsilon) < \log(\mu'/\lambda) = 1 < r(1+\varepsilon)^2$. Each realization $\mathbf{g}$

is a tree of infinite depth having M branches per node, the root node is labelled by a positive integer $Q_0 + 1$, and every branch of the tree is labelled by an N-tuple in $\mathcal{R}_+^N$. $Q_0 + 1$ is the number of arrivals (including the zeroth packet) at time $t = 0$.

The distribution $\mathbf{G}$ on the set of infinite trees is described as follows. Q_0 is selected independent of the other branch labelings according to the distribution $\Pr\{Q_0 = k\} = (1 - \lambda/\mu)(\lambda/\mu)^k$ for $k \in \mathcal{Z}_+$. Furthermore, each N-tuple is i.i.d, and such that each component of the N-tuple is independent and has density e_λ. This induces a distribution $\mathbf{G}$ on the set of infinite trees.

Let

$$T_L(\mathbf{g}, u^L, y^{NL}) = \frac{1}{NL} \sum_{i=1}^{NL} y_i$$

denote the average time for a packet to exit, given the input message is u^L, and the output stream is y^{NL}. Let $T_L(\mathbf{g}) = ET_L(\mathbf{g}, U^L, Y^{NL})$. Consider the random variable $T_L(\mathbf{G})$. The queue is in equilibrium at time $t = 0$, and the arrivals thereafter are Poisson with rate λ. By Burke's output theorem, the departures are also Poisson with rate λ. Hence, for every $L \geq 1$,

$$ET_L(\mathbf{G}) = 1/\lambda, \tag{4.9}$$

where the expectation in (4.9) is with respect to the distribution $\mathbf{G}$.

From the argument in the introduction, while finding the expected number of computations in the mth incorrect subtree, the past up to $m - 1$ nodes can be summarized by one quantity, $\tilde{y}_{m-1}$. Equilibrium at $t = 0$ and Poisson arrivals thereafter ensures that the $N(m - 1)$st packet (the last common packet to the paths under consideration) sees the queue in equilibrium upon arrival. $\tilde{Y}_{m-1}$ therefore has the same distribution as Y_0. Consequently, the random variables $C_m(\mathbf{G}),\ \ m \geq 1$, are identically distributed. Recall that $D_L(\mathbf{G}) = (C_1(\mathbf{G}) + \cdots + C_L(\mathbf{G}))/L$. In Section 3.2 we show the following result.

Lemma 6: *If* $r(1 + \varepsilon) < \log(\mu'/\lambda)$*, there is a finite* K *such that* $EC_1(\mathbf{G}) \leq K$.

Stationarity and the ergodic theorem [45, p. 374] imply that, as $L \to \infty$, both $T_L(\mathbf{G})$ and $D_L(\mathbf{G})$ converge almost surely to random variables $T'(\mathbf{G})$ and $D(\mathbf{G})$, respectively, such that $ET'(\mathbf{G}) = 1/\lambda$, and $ED(\mathbf{G}) = EC_1(\mathbf{G}) \leq K$. Furthermore, because $Y_0(\mathbf{G})$ has a finite expectation, dominated convergence theorem implies that

$$E\left[\lim_{L\to\infty} \frac{1}{NL} Y_0(\mathbf{G})\right] = 0.$$

Hence, with $T(\mathbf{G}) = T'(\mathbf{G}) + \lim_{L\to\infty}(Y_0(\mathbf{G})/(NL))$, we get $ET(\mathbf{G}) = 1/\lambda$.

From Chebyshev's inequality and the union bound on probabilities, we obtain

$$P\left\{\left\{T(\mathbf{G}) > \frac{1+\varepsilon}{\lambda}\right\} \cup \left\{D(\mathbf{G}) > \frac{2K(1+\varepsilon)}{\varepsilon}\right\}\right\} \leq \frac{1+\varepsilon/2}{1+\varepsilon},$$

which implies that

$$\begin{aligned} P\left\{\left\{T(\mathbf{G}) \leq \frac{1+\varepsilon}{\lambda}\right\} \cap \left\{D(\mathbf{G}) \leq \frac{2K(1+\varepsilon)}{\varepsilon}\right\}\right\} &\geq \frac{\varepsilon/2}{1+\varepsilon} \\ &> 0. \end{aligned}$$

Hence, there exists a tree code $\mathbf{g}$ such that $T(\mathbf{g}) \leq (1+\varepsilon)/\lambda$ and $D(\mathbf{g}) \leq 2K(1+\varepsilon)/\varepsilon$.

Following the argument in [38], we then get $\limsup D_L(\mathbf{g}) \leq 2K(1+\varepsilon)/\varepsilon$, and therefore $\sup\{D_L(\mathbf{g}) : L \geq 1\} < A$ for some finite A. Moreover, because $r(1+\varepsilon)^2 > 1$, we get

$$R(1+\varepsilon)^3 = r(1+\varepsilon)^3/T(\mathbf{g}) \geq \lambda r(1+\varepsilon)^2 > \mu/(2e).$$

This concludes the proof of Proposition 12. ■

4.3.2 Expected Number of Computations over the Tree Ensemble

In this subsection, we prove Lemma 6. Fix the first incorrect subtree $\mathcal{U}_1(\mathbf{u})$. The number of computations in this subtree is upperbounded by (cf. [39, Equation (3.1)])

$$C_1(\mathbf{g}, \mathbf{u}, \mathbf{y}) \leq \sum_{l\geq 0}\sum_{j\geq 1}\sum_{\hat{u}^j \in \mathcal{U}_1(\mathbf{u})} \exp\left\{\Gamma\left(\hat{u}^j | y_0, y^{Nj}\right) - \Gamma\left(u^l | y_0, y^{Nl}\right)\right\}.$$

Our aim is to find the expected value of this upperbound over the code ensemble and the output. Clearly, this average value does not depend on the source sequence due to symmetry.

We now look at a $\hat{u}^j$ in the first incorrect subtree. The distribution of $\mathbf{G}$ is such that the choice of $x_1(\hat{u}^1), \cdots, x_{Nj}(\hat{u}^j)$, is independent of the choice of $\mathbf{x}(\mathbf{u})$. Consequently, taking the expectation with respect to the choice of $x_1(\hat{u}^1), \cdots, x_{Nj}(\hat{u}^j)$, and denoting that expectation by $\hat{E}[\cdot]$ as in [39], we get

$$\begin{aligned}\hat{E}C_1(\mathbf{G},\mathbf{u},\mathbf{y}) &\leq \sum_{l\geq 0}\sum_{j\geq 1}\exp\left\{-\Gamma\left(u^l|y_0, y^{Nl}\right)\right\} \\ &\quad\cdot \sum_{\hat{u}^j\in\mathcal{U}_1(\mathbf{u})} \hat{E}\left[\exp\left\{\Gamma\left(\hat{u}^j|y_0, y^{Nj}\right)\right\}\right]. \end{aligned} \tag{4.10}$$

The last summation in (4.10) can be upperbounded as follows. This would have been straightforward if it were not for the memory represented by y_0.

Lemma 7:

$$\sum_{\hat{u}^j\in\mathcal{U}_1(\mathbf{u})} \hat{E}\left[\exp\left\{\Gamma\left(\hat{u}^j|y_0, y^{Nj}\right)\right\}\right] \leq e^{-Njr\varepsilon}\cdot\frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')}.$$

Proof: There are $\exp\{jNr\}$ nodes at depth j in the set $\mathcal{U}_1(\mathbf{u})$. The left hand side is therefore equal to

$$e^{jNr}\cdot e^{-jNr(1+\varepsilon)}\cdot\hat{E}\left[\frac{f_{\mu'}\left(y^{Nj}|X^{Nj}, y_0\right)}{\prod_{i=1}^{Nj} e_\lambda(y_i)}\right], \tag{4.11}$$

where the expectation $\hat{E}[\cdot]$ is with respect to X^{Nj}, which represents the branch labelings for a generic path in the first incorrect subtree.

We now introduce an auxiliary random variable Z which denotes the number of packets in the system when the zeroth packet departs after service. The conditional distribution of Z given $Y_0 = y_0$ is,

$$P'_{Z|Y_0}(z|y_0) = \frac{(\lambda y_0)^z\ e^{-\lambda y_0}}{z!},$$

for $z \in \mathcal{Z}_+$. The marginal of Z when the service times are independent and have density $e_{\mu'}$ is given by

$$P'_Z(z) = \left(1 - \frac{\lambda}{\mu'}\right)\left(\frac{\lambda}{\mu'}\right)^z,$$

for $z \in \mathcal{Z}_+$. The prime indicates that the service times have density $e_{\mu'}$.

Observe that

$$\begin{aligned} P'_{Z|Y_0}(z|y_0) &= P'_Z(z)\,\frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')}\frac{(\mu' y_0)^z e^{-\mu' y_0}}{z!} \\ &\leq P'_Z(z)\,\frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')}, \qquad (4.12) \end{aligned}$$

where (4.12) follows from $(\mu' y_0)^z e^{-\mu' y_0}/z! \leq 1$ for every $z \in \mathcal{Z}_+$ and $y_0 \in \mathcal{R}_+$.

Let

$$P'_{Y^{Nj}|X^{Nj},Y_0},\ P'_{Y^{Nj}|Y_0},\ P'_{Y^{Nj}|Y_0,Z},\ P'_{Y^{Nj}},$$

denote the conditional densities of Y^{Nj} given the indicated random variables. We then have the following sequence of inequalities,

$$\begin{aligned} \hat{E}\left[f_{\mu'}\left(y^{Nj}|X^{Nj},y_0\right)\right] &= \hat{E}\left[P'_{Y^{Nj}|X^{Nj},Y_0}\left(y^{Nj}|X^{Nj},y_0\right)\right] \\ &= P'_{Y^{Nj}|Y_0}\left(y^{Nj}|y_0\right) \\ &= \sum_{z\in\mathcal{Z}_+} P'_{Z|Y_0}(z|y_0)\ \ P'_{Y^{Nj}|Y_0,Z}\left(y^{Nj}|y_0,z\right) \\ &\overset{(a)}{=} \sum_{z\in\mathcal{Z}_+} P'_{Z|Y_0}(z|y_0)\ \ P'_{Y^{Nj}|Z}\left(y^{Nj}|z\right) \\ &\overset{(b)}{\leq} \sum_{z\in\mathcal{Z}_+} P'_Z(z)\ \ P'_{Y^{Nj}|Z}\left(y^{Nj}|z\right)\ \frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')} \\ &\overset{(c)}{=} P'_{Y^{Nj}}\left(y^{Nj}\right)\ \frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')}, \\ &= \frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')}\prod_{i=1}^{Nj} e_\lambda(y_i), \qquad (4.13) \end{aligned}$$

where (a) follows because Y_0 and Y^{Nj} are conditionally independent given Z, a consequence of the memoryless property of the interarrival times; (b) follows from (4.12);

in equality (c), the dependence on y_0 has been successfully separated; equality (4.13) follows from (4.6).

Substitution of (4.13) in (4.11) yields Lemma 7. ■

We continue with the proof of Lemma 6. Observe that the random variables in the right hand side of (4.10) are Y_0 and $(\mathbf{X}(\mathbf{u}),\mathbf{Y})$. Substitution of (4.5) and the result of Lemma 7 in (4.10), followed by the expectation operation with respect to Y_0 and $(\mathbf{X}(\mathbf{u}),\mathbf{Y})$, yields

$$\begin{aligned} EC_1(\mathbf{G},\mathbf{u},\mathbf{Y}) \;&\le\; \sum_{l\geq 0}\sum_{j\geq 1} e^{-jNr\varepsilon} e^{-lNr(1+\varepsilon)} \cdot \int_{\mathcal{R}_+} dy_0 \; P_{Y_0}(y_0)\, \frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')} \\ &\quad \cdot \left[\int_{\mathcal{R}_+^{Nl}} dy^{Nl}\; E\left[f_\mu\left(y^{Nl}|X^{Nl},y_0\right)\left(\frac{\prod_{i=1}^{Nl} e_\lambda(y_i)}{f_{\mu'}\left(y^{Nl}|X^{Nl},y_0\right)}\right)\right]\right], \end{aligned} \tag{4.14}$$

where the expectation in the innermost integral in (4.14) is with respect to X^{Nl}. Observe that

$$\begin{aligned} E\left[\frac{f_\mu\left(y^{Nl}|X^{Nl},y_0\right)}{f_{\mu'}\left(y^{Nl}|X^{Nl},y_0\right)}\right] \;&=\; E\left[f_{\mu'}\left(y^{Nl}|X^{Nl},y_0\right)\right]\left(\frac{4}{\mu}\right)^{Nl} \\ &\le\; \left(\prod_{i=1}^{Nl} e_\lambda(y_i)\right)\frac{e^{(\mu'-\lambda)y_0}}{(1-\lambda/\mu')}\left(\frac{4}{\mu}\right)^{Nl}, \end{aligned} \tag{4.15}$$

where (4.15) follows from (4.13). Furthermore, because $2\mu' = \mu$ and $P_{Y_0}(y_0) = e_{\mu-\lambda}(y_0)$, we obtain

$$\int_{\mathcal{R}_+} dy_0\; P_{Y_0}(y_0)\,\frac{e^{2(\mu'-\lambda)y_0}}{(1-\lambda/\mu')^2} = \frac{(\mu/\lambda-1)}{(1-\lambda/\mu')^2}, \tag{4.16}$$

and

$$\begin{aligned} \int_{\mathcal{R}_+^{Nl}} dy^{Nl}\left(\prod_{i=1}^{Nl} e_\lambda(y_i)\right)^2 \;&=\; \left(\int_{\mathcal{R}_+} dy\; \lambda^2 e^{-2\lambda y}\right)^{Nl} \\ &=\; (\lambda/2)^{Nl}. \end{aligned} \tag{4.17}$$

Substitution of (4.15), (4.16) and (4.17) in (4.14) yields

$$\begin{aligned} EC_1(\mathbf{G}, \mathbf{u}, \mathbf{Y}) \ \le \ & \sum_{l\ge 0}\sum_{j\ge 1} e^{-jNr\varepsilon} \cdot \frac{(\mu/\lambda - 1)}{(1-\lambda/\mu')^2} \\ & \cdot \exp\left\{ lN \left[r(1+\varepsilon) - \log\left(\frac{\mu}{2\lambda}\right)\right]\right\}. \end{aligned}$$

The summation over j is finite. The summation over l is finite because $r(1+\varepsilon) < \log(\mu/(2\lambda))$. Consequently, $EC_1(\mathbf{G}) \le K$, for some finite K. This concludes the proof of Lemma 6. ■

4.4 Discussion

We have shown that for every $\delta > 0$, there is a tree code such that the rate of information transfer, R, using the sequential decoding technique, satisfies $R(1+\delta) > \mu/(2e)$ nats per second, and the average number of computations to move one step forward in the correct direction is upperbounded by a finite number. The quantity $\mu/(2e)$ nats per second is one half of the capacity, and is a lower bound on the cutoff rate for sequential decoding. Some open questions remain. For example, we do not know the cutoff rate for this exponential server timing channel.

Although we have not dealt with discrete-time timing channels [4] in this chapter, analogous results follow straightforwardly. However, we do not know a closed form expression for the rate achievable using sequential decoding with an analogous metric. For the geometric service time distribution $P(S = k) = \mu(1-\mu)^{k-1}, \quad k \ge 1$, the corresponding achievable rate in nats per slot is

$$\max_{\lambda \in [0,\ 1-\sqrt{1-\mu}\ )} \quad \lambda \left[\log\left(\frac{1-\sqrt{1-\mu}}{1+\sqrt{1-\mu}} \right) + \log\left(\frac{2-\lambda}{\lambda} \right) \right].$$

Let λ^* be the maximizing λ. To remove the dependence on Y_0 as in the continuous-time case (cf. (4.13)), λ^* should satisfy

$$(1 - (\mu - \lambda^*)) \cdot \left(2 - \lambda^* + \sqrt{1-\mu}\right)^2 < 1.$$

Although we have not proved that this holds for all $\mu \in (0,1)$, numerical evidence indicates that this is so.

In practice, we need trees with finite depth having extra terminating branches. These tail branches ensure that the last few source symbols can also be decoded correctly with high probability. While this causes a loss in rate, the loss is negligible if the number of additional branches is small in comparison to the block length of the code. In this case, we can easily show that the number of computations in each incorrect subtree is upperbounded by a constant that is independent of the code length. Furthermore, the probability of error, when one of the other terminating leaves reaches the top of the stack, can be made small by choosing a sufficiently long tail [41]. We omit proofs for the rationale of these simple modifications.

If all terminating leaves have the same $\sum_{i=1}^{Nt} x_i$, where t is the maximum depth of the tree, then the state represented by $\tilde{y}_t$ is the same for all terminating leaves, given a sequence of received interdeparture times. All states have therefore merged into a single one. Transmission can then begin afresh, with a decision up to depth t not affecting future decisions.

We finally remark that λ, the net throughput in packets per second, should be smaller than $\mu/2$ for the sequential decoding scheme to work with finite per-branch computational complexity. Therefore, in already existing systems, information can be piggy-backed through timing in the above tree-code form only if the system is lightly loaded. Moreover, unlike convolutional codes, we need to store the labels for the entire tree at the decoder. Despite these drawbacks, this chapter is a positive step in the direction of finding good codes with computationally feasible decoding techniques for communication over timing channels.

Chapter 5

Open Questions

In this chapter, we collect the open problems mentioned in the previous chapters. We include a few more on which this thesis did not focus attention. In the appendix, we look at one other interesting problem in detail.

We described a robust decoder in Chapter 2 for the single-server queue. Given any stationary and ergodic sequence of service times with mean $1/\mu$ seconds, we showed that $e^{-1}\mu$ nats per second is an achievable rate using this decoding criterion. Given a particular stationary and ergodic sequence of service times, is there a systematic way to give tighter lower and upper bounds on the capacity under this specified decoding rule?

Suppose that the codebook random variable $\mathbf{C}$ with parameters (n, M_n, T_n) is chosen as in Chapter 2 (cf. (2.21), (2.20) and (2.6)). We call $\mathbf{C}$ a *Poisson codebook* because of the following. A realization $\mathbf{c}$ of this random variable consists of M_n codewords; each codeword is a realization of the Poisson process such that the nth arrival occurs before time T_n. Let $P_{Y^n|X^n}$ be the channel induced by a stationary and ergodic process of service times with mean $1/\mu$ seconds. Let $E\left[P_e\left(\mathbf{C}, \phi_f, P_{Y^n|X^n}\right)\right]$ be the average probability of error for window-codes employing the robust decoder, averaged over the ensemble of Poisson codebooks. Is it true that for any stationary and ergodic sequence of service times,

$$E\left[P_e\left(\mathbf{C}, \phi_f, P_{Y^n|X^n}\right)\right] \to 1$$

as $T_n \to \infty$ for rates above $e^{-1}\mu$ nats per second? A similar result holds for *Gaussian codebooks* on the additive noise channel [12, Theorem 1]. The parallelism between the single-server queue and the additive noise channel seems to indicate this might be true.

We showed that for the discrete-time queue random-code reduction is possible and that we can apply the elimination technique of [14]. This implied that on the jammed timing channel with geometrically distributed nominal service times, the largest rate achievable with random window-codes is in fact achievable with a nonrandom strategy, provided the packets themselves carry a certain amount of information noiselessly. Is there a similar strategy on the continuous-time queue?

What is the capacity of the jammed timing channel with exponentially (resp. geometrically) distributed nominal service times, when no side channel is available, and when the jammer has complete knowledge of the codebook?

We mentioned in Chapter 2 that the discrete-time queue satisfies the strong converse (cf. paragraph following proof of Proposition 6(c)). This means that for every $\delta > 0$ and every sequence of $(n, M_n, T_n, \varepsilon_n)$-window-codes with rate

$$\frac{\log M_n}{T_n} > \log\left[1 + \mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}\right] + \delta,$$

we have that $\lim_{n\to\infty} \varepsilon_n = 1$. Does the continuous-time queueing system satisfy the strong converse? The answer to this question will be affirmative if we can show that the ideal Poisson channel with noiseless feedback [46] satisfies the strong converse.

In Chapter 3, we saw how the point-process approach suggested a method to give bounds on the capacity of some simple queueing systems. We do not however know the capacity of such systems. In particular, we would like to know answers to the following questions. How does the capacity of the multiserver queue increase to infinity with the number of servers? How does the capacity of queues connected in tandem decrease to 0 as the number of queues increases?

Are there systematic ways to develop codes with computationally feasible decoding criteria on the single-server queue? We saw one approach in Chapter 4; we showed that tree codes with sequential decoding can achieve rates below half the capacity on the single-server queue. It would be interesting to find techniques that attain rates larger than half the capacity.

Consider the single-server queueing system with noiseless feedback. Suppose now that coding schemes having arbitrarily small probability of error with random transmission times are allowed. Are there any transmission strategies that achieve capacity? What is the capacity if we insist on zero probability of error?

Another interesting question for further work is the following. On the continuous-time queueing system, consider the (n, M_n, T_n, P_n)-window-code. Let $P^*(n, M_n, T_n)$ be the infimum of those P_n for which a window-code with parameters (n, M_n, T_n, P_n) exists. For rates below the capacity, $0 < R \leq e^{-1}\mu$, let the *reliability function* $E(R)$ be the optimal error-exponent for window-codes, i.e.,

$$E(R) \triangleq \limsup_{T_n \to \infty} -\frac{1}{T_n} \log P^* \left(n, \lfloor e^{RT_n} \rfloor, T_n \right),$$

maximized over the best choice of $(T_n : n \in \mathcal{N})$. What is the $E(R)$ of the single-server queue for $0 < R \leq e^{-1}\mu$? Bounds on the reliability function for the ideal Poisson channel with noiseless feedback, a channel closely related to the single-server queue, were given in [46]. It would be interesting to compare the reliability function of the single-server queue with the reliability function of the ideal Poisson channel with noiseless feedback.

Appendix A

(Zero-Error) Average List Size Capacity

In this appendix we discuss the following open problem. We restrict ourselves to the $\cdot/M/1$ queue. The channel is $(Q_{Y^n|X^n} : n \in \mathcal{N})$, where $Q_{Y^n|X^n}$ is given by (2.4) and (2.5). Fix $n \in \mathcal{N}$. It is possible that certain codewords are incompatible with the observed sequence of departures y^n. This naturally raises the following question.

Let $n \in \mathcal{N}$. Let $\mathbf{c}$ be the codebook containing M_n codewords. Recall that $y^n = (y_0, y_1, \cdots, y_n)$ in Chapter 2. We now assume that $y_0 = 0$, and henceforth we take $y^n = (y_1, \cdots, y_n)$. The output alphabet is therefore $\mathcal{R}_+^n$. Upon observing y^n, let the decoder output the set

$$L(y^n, \mathbf{c}) \triangleq \{x^n \in \mathbf{c} : p(x^n, y^n) > 0\},$$

i.e., the list of codewords that could explain the received sequence of departures y^n. Observe that $p(x^n, y^n) > 0$ is equivalent to

$$\begin{aligned} y_1 &\geq x_1, \\ y_1 + y_2 &\geq x_1 + x_2, \\ &\vdots \\ y_1 + y_2 + \cdots + y_n &\geq x_1 + x_2 + \cdots + x_n, \end{aligned}$$

i.e., a codeword is compatible with y^n if only if the ith arrival in the codeword occurs before the ith departure, $i = 1, \cdots, n$. If $|L(Y^n, \mathbf{c})| = 1$, the decoder recovers the transmitted message without error.

Given an $\varepsilon > 0$, suppose that we require $E\left[|L(Y^n, \mathbf{c})|\right] \leq 1 + \varepsilon$, where the expectation is over uniform inputs and the transition probability function $Q_{Y^n|X^n}$. The largest asymptotic rate $\lim_{n\to\infty}(\log M_n)/n$ that can be supported is called the *(zero-error) average list size capacity* (cf. [10], [11]). What is the (zero-error) average list size capacity of the single-server queue?

We believe that the answer is 0 due to the following. Fix $n \in \mathcal{N}$ and $M_n \in \mathcal{N}$. We pick the codebook (of M_n codewords) according to a distribution $P_{\mathbf{C}}$, and show that, if the (zero-error) average list size is smaller than $1 + \varepsilon$ when averaged over the choice of codebooks, then

$$M_n \leq 1 + \varepsilon(2n + 1), \tag{A.1}$$

for a fairly large class of distributions $P_{\mathbf{C}}$.

Let P_{X^n} be an arbitrary distribution on $\mathcal{R}_+^n$ and its product Borel σ-algebra. We pick the codebook of M_n codewords according to $P_{\mathbf{C}}$, where

$$dP_{\mathbf{C}}(\mathbf{x}_1, \cdots, \mathbf{x}_{M_n}) = dP_{X^n}(\mathbf{x}_1)dP_{X^n}(\mathbf{x}_2) \cdots dP_{X^n}(\mathbf{x}_M).$$

Let $E\left[|L(Y^n, \mathbf{C})|\right]$ denote the (zero-error) average list size, averaged over the choice of codebooks. We can easily verify that

$$p(\mathbf{x}_2, \mathbf{x}_1) > 0 \text{ and } p(\mathbf{x}_1, y^n) > 0 \Rightarrow p(\mathbf{x}_2, y^n) > 0. \tag{A.2}$$

We then have the following sequence of inequalities,

$$\begin{aligned} 1 + \varepsilon &> E\left[|L(Y^n, \mathbf{C})|\right] \\ &= 1 + E\left[\sum_{i=2}^{M_n} 1\left\{p(\mathbf{X}_i, Y^n) > 0\right\}\right] \end{aligned}$$

$$\begin{aligned}
&= 1+(M_n-1)\int_{\mathcal{R}_+^n\times\mathcal{R}_+^n} dP_{X^n}(\mathbf{x}_1)\;dP_{X^n}(\mathbf{x}_2) \\
&\qquad\qquad\qquad \cdot\int_{\mathcal{R}_+^n} dQ_{Y^n|X^n}(y^n)\;1\left\{p(\mathbf{x}_2,y^n)>0\right\} \\
&\overset{(a)}{\geq} 1+(M_n-1)\int_{\mathcal{R}_+^n\times\mathcal{R}_+^n} dP_{X^n}(\mathbf{x}_1)\;dP_{X^n}(\mathbf{x}_2)\;1\left\{p(\mathbf{x}_2,\mathbf{x}_1)>0\right\},
\end{aligned}$$

where (a) follows from (A.2). Let $z_i = x_{i,1}-x_{i,2}$, and $s_i = \sum_{j=1}^{i} z_j$. Then, $p(\mathbf{x}_2,\mathbf{x}_1)>0$ is equivalent to saying that the sequence $(s_i : i=1,\cdots,n)$ stays above 0, i.e., $s_i \geq 0$ for $i=1,\cdots,n$. We then have

$$M_n \leq 1+\frac{\varepsilon}{P_{S^n}\left\{S_i\geq 0 : i=1,\cdots,n\right\}}. \tag{A.3}$$

Let P_{X^n} be such that $dP_{X^n}(x^n)=\prod_{i=1}^{n} dP_X(x_i)$, where the cdf of X is continuous. This implies that the cdf of Z is continuous and symmetric, that $(S_1,\cdots,S_n)$ is a random walk, and that [47, pp. 396-397]

$$P_{S^n}\left\{S_i\geq 0 : i=1,\cdots,n\right\} = \binom{2n}{n}\frac{1}{2^{2n}}. \tag{A.4}$$

Substitution of (A.4) in (A.3) and the observation

$$\binom{2n}{n}\frac{1}{2^{2n}} \geq \frac{1}{2n+1}$$

yields (A.1).

Under the input measure P_{X^n} and the transition probability function $Q_{Y^n|X^n}$, let the output measure be P_{Y^n}. Furthermore, let $P_{X^n\times Y^n}$ be the joint probability measure where X^n and Y^n are independent with marginals P_{X^n} and P_{Y^n}. Observe that for some codebook $\mathbf{c}$ with M_n codewords, we require

$$\begin{aligned}
1+\varepsilon &\geq \sum_{\mathbf{x}\in\mathbf{c}}\frac{1}{M_n}\int_{\mathcal{R}_+^n} dQ_{Y^n|X^n}(y^n|\mathbf{x})\left(\sum_{\mathbf{a}\in\mathbf{c}} 1\left\{p(\mathbf{a},y^n)>0\right\}\right), \\
&= M_n\int_{\mathcal{R}_+^n} dP_{Y^n}(y^n)\left(\sum_{\mathbf{a}\in\mathbf{c}}\frac{1}{M_n}1\left\{p(\mathbf{a},y^n)>0\right\}\right) \\
&= M_n\cdot P_{X^n\times Y^n}\left\{p(X^n,Y^n)>0\right\},
\end{aligned}$$

where P_{X^n} is the uniform distribution on the set $\mathbf{c}$ and P_{Y^n} is the corresponding output distribution. We then have that

$$\begin{aligned} M_n &\leq \frac{1+\varepsilon}{P_{X^n \times Y^n}\{p(X^n, Y^n) > 0\}} \\ &\overset{(a)}{\leq} \frac{1+\varepsilon}{\min_{P_{X^n}} \quad P_{X^n \times Y^n}\{p(X^n, Y^n) > 0\}}; \end{aligned}$$

in (a), P_{X^n} is any arbitrary distribution on the input space, not necessarily restricted to uniform distribution on a finite set with M_n points.

Is there a sequence $(P_{X^n} : n \in \mathcal{N})$ for which $P_{X^n \times Y^n}\{p(X^n, Y^n) > 0\}$ goes to 0 exponentially fast? It seems unlikely in the light of (A.1).

Bibliography

[1] V. Anantharam and S. Verdú, “Bits through queues,” *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 4–18, Jan. 1996.

[2] W.M. Hu, “Reducing timing channels with fuzzy time,” in *Proc. 1991 Computer Society Symposium on Research in Security and Privacy*, Oakland, CA, May 1991.

[3] R.G. Gallager, “Basic limits on protocol information in data communication networks,” *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 385–395, Jul. 1976.

[4] A. Bedekar and M. Azizoğlu, “The information-theoretic capacity of discrete-time queues,” *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 446–461, Mar. 1998.

[5] J.A. Thomas, “On the Shannon capacity of discrete-time queues,” in *Proc. 1997 IEEE Int. Symp. on Information Theory*, Ulm, Germany, Jul. 1997, p. 333.

[6] Y.M. Kabanov, “The capacity of a channel of the Poisson type,” *Theory Prob. Appl.*, vol. 23, pp. 143–147, 1978.

[7] M.H.A. Davis, “Capacity and cutoff rate for Poisson-type channels,” *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 710–715, Nov. 1980.

[8] A.D. Wyner, “Capacity and error exponent for the direct detection photon channel,” *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1449–1471, Nov. 1988.

[9] J. Walrand, *An Introduction to Queueing Networks*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[10] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 55–62, Jan. 1996.

[11] İ.E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1977–1982, Nov. 1997.

[12] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1520–1529, Sept. 1996.

[13] D. Blackwell et al., "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.

[14] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitsth. Verw. Gebiete*, vol. 44, pp. 159–175, 1978.

[15] I. Csiszár and P. Narayan, "The capacity of arbitrarily varying channels revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 181–193, Mar. 1988.

[16] I. Csiszár and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 18–26, Jan. 1991.

[17] T.R.M. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," in *Trans. 8th Prague Conf. on Information Theory*, 1978, pp. 211–226.

[18] G. Kaplan and S. Shamai (Shitz), "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU*, vol. 47, no. 4, pp. 228–239, 1993.

[19] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Budapest, second edition, 1986.

[20] N. Merhav et al., "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 1953–1967, Nov. 1994.

[21] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 35–43, Jan. 1995.

[22] A. Lapidoth, "On mismatched decoding," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1439–1452, Sept. 1996.

[23] S. Verdú, "The exponential distribution in information theory," Invited paper, *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 100–111, Jan.-Mar. 1996 (In Russian). English version in *Problems of Information Transmission*, vol. 32, no. 1, pp. 86-95, Jan.-Mar. 1996.

[24] I. Stiglitz, "Coding for a class of unknown channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 189–195, Apr. 1966.

[25] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 2148–2177, Oct. 1998.

[26] B. Hughes and P. Narayan, "Gaussian arbitrarily varying channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 267–284, Mar. 1987.

[27] J. Giles and B.E. Hajek, "The jamming game for timing channels," in *Proc. 1999 IEEE Information Theory and Networking Workshop*, Metsovo, Greece, Jun. 1999.

[28] K.L. Chung, *A Course in Probability Theory*, Academic Press, New York, second edition, 1974.

[29] P. Billingsley, *Probability and Measure*, John Wiley & Sons, New York, second edition, 1986.

[30] R.M. Gray, *Entropy and Information Theory*, Springer-Verlag, New York, 1990.

[31] R.S. Liptser and A.N. Shiryayev, *Statistics of Random Processes*, vol. 1, Springer, New York, 1977.

[32] R.S. Liptser and A.N. Shiryayev, *Statistics of Random Processes*, vol. 2, Springer, New York, 1978.

[33] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York, 1981.

[34] D.L. Snyder, *Random Point Processes*, John Wiley & Sons, New York, 1975.

[35] S. Verdú and T.S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 1147–1157, Jul. 1994.

[36] R.G. Gallager, *Discrete Stochastic Processes*, Kluwer Academic Publishers, Boston, 1996.

[37] S. Verdú, "Information theory of queueing systems," Featured Invited Talk*, 1995 IEEE Information Theory Workshop on Information Theory, Multiple Access and Queueing,* St. Louis, Missouri, April 19-21, 1995.

[38] E. Arikan, "Sequential decoding for multiple access channels," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 246–259, Mar. 1988.

[39] V.B. Balakirsky, "An upper bound on the distribution of computation of a sequential decoder for multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 399–408, Mar. 1996.

[40] R.G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

[41] F. Jelinek, "A fast sequential decoding algorithm using a stack," *IBM J. Res. Develop.*, vol. 13, pp. 675–685, 1969.

[42] K. Zigangirov, "Some sequential decoding procedures," *Problemy Peredachi Informatsii*, vol. 2, no. 4, pp. 13–25, 1966.

[43] A. Lapidoth and J. Ziv, "Universal sequential decoding," in *Proc. 1998 IEEE Information Theory Workshop*, Killarney, Ireland, Jun. 1998.

[44] J. Ziv, "Universal decoding for finite-state channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 453–460, Jul. 1985.

[45] G.R. Grimmett and D.R. Stirzaker, *Probability and Random Processes*, Oxford University Press, Inc., New York, second edition, 1992.

[46] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 491–503, Mar. 1993.

[47] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley Eastern Limited, New Delhi, Wiley Eastern University edition, 1991.