

# On the Zero-Rate Error Exponent of the Exponential-Server Timing Channel \*

Aaron B. Wagner and Venkat Anantharam

Dept. of Electrical Engineering and Computer Sciences

University of California

Berkeley, CA 94720 USA

{awagner, ananth}@eecs.berkeley.edu

## Abstract

We prove that the zero-rate error exponent of the exponential-server timing channel is at least  $\mu/2$ , where  $1/\mu$  is the mean service time. This proves that at low data rates, the timing channel is strictly more reliable than the related Poisson channel with peak intensity  $\mu$  and zero dark current. This answers a question recently posed by Arikan, who proved that the timing channel's reliability function is at least as large as the Poisson channel's at all rates, and that the two reliability functions coincide at high rates. Our proofs provide insight into the construction of good codes for the timing channel by revealing a useful distance metric between codewords.

## 1 Introduction

The exponential-server timing channel (ESTC) is one in which the sender chooses the times at which identical jobs arrive at a  $\cdot/M/1$  queue, while the receiver observes their departure times. We shall assume a first-in-first-out service discipline, but clearly other disciplines resulting in the same conditional law of the departure times given the arrival times, such as last-in-first-out, may be substituted. We shall also assume that the queue is initially empty, although this assumption is also not crucial.

Our interest in this channel is due to its canonical nature and the important role it plays in the study of covert timing channels, in which the sender sends messages containing irrelevant or misleading information, and actually communicates with the receiver using the timing of the messages, unknown to a casual observer [1, 2].

The ESTC is unusual among channels in that the difference between the transmission time and reception time of a codeword is not negligible, and both the transmission time and reception time depend on the codeword that is sent. This makes the cost of transmission unclear *a priori*; see Anantharam and Verdú [3] for a discussion of this point. We resolve this ambiguity by adopting the *window code* definition of Sundaresan and Verdú [1].

A  $(n, M, T)$  (window) code is a collection of  $M$  codewords, each a vector of  $n$  nonnegative interarrival times  $(a_1, \dots, a_n)$  such that the  $k$ th arrival occurs at time  $\sum_{i=1}^k a_i$ , and the  $n$ th

---

\*This research was supported by DARPA Grants F30602-00-2-0538 and N66001-00-C-8062, by Grant N00014-1-0637 from the Office of Naval Research, and by Grant SBR-9873086 and a Graduate Research Fellowship from the National Science Foundation.

arrival occurs before time  $T$ ; and a decoder that observes the output over  $[0, T]$  then selects a codeword or declares an error. The data rate of a  $(n, M, T)$  code is  $(1/T) \log M$  (throughout we use logarithms with base  $e$ ). The arrival rate is  $n/T$ .

Let  $a^n = (a_1, \dots, a_n)$  be a codeword. Then the conditional density of the channel output, the interdeparture times  $d^n = (d_1, \dots, d_n)$ , given  $a^n$  is

$$P(d^n | a^n) = \prod_{i=1}^n e_{\mu}(d_i - w_i),$$

where  $e_{\mu}(\cdot)$  is the exponential density with mean  $1/\mu$  and  $w_i$  is the time that the server is idle between servicing jobs  $i-1$  and  $i$ ,

$$w_i = \max \left( \sum_{j=1}^i a_j - \sum_{j=1}^{i-1} d_j, 0 \right),$$

and  $w_1 = a_1$  by convention.

The Shannon capacity of this channel, when using window codes, is  $\mu/e$  nats/time [1]. The channel was introduced by Anantharam and Verdú [3], who found an identical capacity while using a slightly different block code definition.

We shall study the more refined characterization of the channel known as the reliability function [4], continuing the work of Arikan [5]. Let  $P_e(R, T)$  be the infimum of the average probability of error over all  $(n, M, T)$  codes with  $n \geq 1$  and  $M \geq e^{RT}$ . Then the *reliability function* (or *error exponent*) at rate  $R > 0$  of the channel is

$$E(R) = \limsup_{T \rightarrow \infty} -\frac{1}{T} \log P_e(R, T),$$

and

$$E(0) = \sup_{R > 0} E(R).$$

By modifying the random-coding and sphere-packing bounds [4] to handle the unique features of this channel, Arikan proved the following.

**Proposition 1 ([5])** *Let  $E_{sp}(R)$  for  $0 < R < \mu/e$  be defined parametrically by*

$$E_{sp}(R) = \frac{\mu}{(1 + \rho)^{(1+\rho)/\rho}} [\rho - \log(1 + \rho)],$$

$$R = \frac{\mu}{(1 + \rho)^{(1+\rho)/\rho}} \log(1 + \rho)^{1/\rho},$$

as  $\rho$  ranges over  $(0, \infty)$ , and let  $E_{sp}(0) = \mu$ . For  $R_c \triangleq (\mu/4) \log 2 < R < \mu/e$ , let  $E_r(R) = E_{sp}(R)$ , while for  $0 \leq R \leq R_c$ , let  $E_r(R) = \mu/4 - R$ . Then the reliability function of the exponential-server timing channel with mean service time  $1/\mu$  satisfies

$$E_r(R) \leq E(R) \leq E_{sp}(R)$$

for all  $0 \leq R < \mu/e$ .

Since  $E_r(R)$  and  $E_{sp}(R)$  coincide for  $R_c \leq R \leq C$ , the reliability function is determined at these rates. The reliability function is currently not known for rates below  $R_c$ , and in particular  $E(0)$  is not known; one only has the bounds  $\mu/4 \leq E(0) \leq \mu$ . We wish to determine  $E(0)$  for two reasons.

The first is to compare the reliability function to the Poisson channel's. The input to the Poisson channel is a waveform  $\lambda(t)$  while the output is a Poisson process with rate  $\lambda(t) + \lambda_0$ . It models an optical channel with direct detection; the parameter  $\lambda_0$  models a "dark current." Generally,  $\lambda(t)$  is constrained in both peak and mean. We shall restrict our attention to the case of a peak constraint on the input:  $0 \leq \lambda(t) \leq \mu$ , no constraint on the average, and zero dark current:  $\lambda_0 = 0$ .

The capacity of the Poisson channel under these constraints is also  $\mu/e$  [6]. Furthermore, the reliability function, which was determined by Wyner [7, 8], coincides with Arikan's random-coding lower bound,  $E_r$ , for the reliability function of the ESTC. Thus the two reliability functions are equal at rates above  $R_c$ . A natural question, posed by Arikan, is whether the two reliability functions coincide at all rates. The answer likely hinges on their zero-rate asymptotes since this is where the random-coding and sphere-packing exponents tend not to be tight. In particular, one can show using Gallager's expurgated exponent [4] that the random-coding bound lies strictly below the reliability function at all sufficiently low rates for any discrete-memoryless channel with nonzero capacity. Indeed, we prove, using a modified expurgated exponent bound, that  $E(0) \geq \mu/2$  for the ESTC, so that the two reliability functions differ at low rates: the ESTC is strictly more reliable. These results are contained in Section 2. Some remarks about the tightness of our bound are made in Section 3.

Our second source of interest in  $E(0)$  is that low-rate error exponents typically reveal considerable insight into the structure of good low-rate codes. The calculation of  $E(0)$  for the BSC (resp. Gaussian) channel, for instance, reveals that the error probability of a low-rate code is governed by the minimum Hamming (resp. Euclidean) distance between pairs of codewords. Although we do not determine  $E(0)$ , we exhibit an analogous notion of distance for the ESTC and show that the error probability of a low rate code<sup>1</sup> is determined by the minimum of this distance between codeword pairs.

## 2 An Expurgated-Exponent Bound

The exponentially-distributed service times make the maximum-likelihood (ML) decoder for the channel simple to describe. Given the received vector of interdeparture times, the ML decoder first excludes those codewords for which a departure occurs before the corresponding arrival—these codewords could not have been sent, and we call them *infeasible*. For each feasible codeword, the decoder computes the service times that would cause the observed departure times. The ML decoder then selects the codeword with the smallest *sum* of service times.

Unfortunately, the ML decoder is not as simple to analyze as it is to describe. We proceed by approximating ML detection with a suboptimal rule. To any interevent times  $(u_1, \dots, u_n)$ , we associate the function

$$u(t) = \sup \left\{ 0 \leq i \leq n : \sum_{j=1}^i a_j \leq t \right\}.$$

Observe that a codeword  $x^n$  is feasible for an output  $y^n$  if and only if  $y(t) \leq x(t)$  for all  $t$ . For any pair of interevent times  $u^n$  and  $v^n$ , we say that  $u^n$  *leads*  $v^n$  by

$$\mathcal{L}(u^n, v^n) = \int_0^\infty 1(u(t) > v(t)) dt.$$

---

<sup>1</sup>More precisely, a code with a small data rate *and* arrival rate.

If  $x^n$  are the interarrival times for a single-server queue, and  $y^n$  are the interdeparture times, then  $\mathcal{L}(x^n, y^n)$  is the time that the server is busy. Thus the ML rule can be stated as

$$\arg \min_{x^n \text{ feasible}} \mathcal{L}(x^n, y^n).$$

Consider the ML detection of two codewords,  $u^n$  and  $v^n$ . Define the codeword  $w^n$  by  $w(t) = \min(u(t), v(t))$ . If the output  $y^n$  satisfies  $y(t) > w(t)$  for some  $t$ , then only one of  $u^n$  and  $v^n$  is feasible for  $y^n$ . If  $y(t) \leq w(t)$ , then both codewords are feasible, and a comparison of  $\mathcal{L}(u^n, y^n)$  to  $\mathcal{L}(v^n, y^n)$  is required. Let  $W = \{y^n : y(t) \leq w(t)\}$ . In general, there are outputs in  $W$  that  $u^n$  leads less than  $v^n$  does, and different outputs in  $W$  that  $v^n$  leads less than  $u^n$  does. In this case,  $W$  is not wholly contained in the decision region of either  $u^n$  or  $v^n$ . If  $u^n$  and  $v^n$  are well-separated, however, one expects the probability of  $W$  to be small under both codewords, and dominated by the chance that  $y^n$  is close to  $w^n$ . This suggests that there would be little loss in placing the region  $W$  entirely within the decision region of the codeword that leads  $w^n$  by the smallest amount. It is easily verified that  $\mathcal{L}(u^n, w^n) = \mathcal{L}(u^n, v^n)$  and similarly  $\mathcal{L}(v^n, w^n) = \mathcal{L}(v^n, u^n)$ , which leads us to examine the following suboptimal decoder.

**Definition 1** A codeword  $u^n$  in a  $(n, M, T)$  code dominates another codeword  $v^n$  in the same code given the output  $y^n$  if at least one of two conditions holds:

1.  $u^n$  is feasible for  $y^n$  and  $v^n$  is not, or
2.  $u^n$  and  $v^n$  are both feasible for  $y^n$  and  $\mathcal{L}(v^n, u^n) \geq \mathcal{L}(u^n, v^n)$ .

The pseudo-ML decoder operates as follows: given the channel output  $y^n$ , if there exists a unique codeword  $u^n$  that dominates all other codewords given  $y^n$ , then  $u^n$  is chosen. Otherwise, an error is declared.

Typically the output  $y^n$  is clear from the context and we abbreviate “ $u^n$  dominates  $v^n$  given  $y^n$ ” as simply  $u^n \geq_d v^n$ .

The next two lemmas relate the lead time between codewords to their error probabilities. The first lemma follows immediately from the Chernoff bound.

**Lemma 1** If  $\{S_i\}_{i=1}^{\infty}$  are i.i.d. exponential with mean  $1/\mu$ , then for all  $a > 1/\mu$  and all  $n \geq 1$ ,

$$\Pr \left( \frac{1}{n} \sum_{i=1}^n S_i \geq a \right) \leq \exp(-n(a\mu - 1 - \log(a\mu))).$$

Let  $\bar{\mathcal{L}}(u^n, v^n) = \max(\mathcal{L}(u^n, v^n), \mathcal{L}(v^n, u^n))$ . It is readily verified that  $\bar{\mathcal{L}}$  is a metric on the space of codewords with  $n$  points, for each  $n$ .

**Lemma 2** Let  $u^n$  and  $v^n$  be two codewords. If

$$\bar{\mathcal{L}}(u^n, v^n) > \frac{n}{\mu},$$

then

$$P(v^n \geq_d u^n | u^n) \leq \exp(-\mu \bar{\mathcal{L}}(u^n, v^n) + n + n \log(\bar{\mathcal{L}}(u^n, v^n) \mu / n)),$$

where  $1/\mu$  is the mean service time.

*Proof.* If  $\mathcal{L}(v^n, u^n) > \mathcal{L}(u^n, v^n)$ , then it is impossible for  $v^n$  to dominate  $u^n$  given an output caused by  $u^n$ , so  $P(v^n \geq_d u^n | u^n) = 0$ . If  $\mathcal{L}(u^n, v^n) \geq \mathcal{L}(v^n, u^n)$ , then  $\mathcal{L}(u^n, v^n) = \bar{\mathcal{L}}(u^n, v^n)$ . Now  $v^n \geq_d u^n$  only if  $v^n$  is feasible, and this requires that the server be busy during the entire time that  $u^n$  leads  $v^n$ . Thus if  $S_1, \dots, S_n$  are the  $n$  service times,

$$P(v^n \geq_d u^n | u^n) \leq \Pr \left( \sum_{i=1}^n S_i \geq \bar{\mathcal{L}}(u^n, v^n) \right).$$

The result then follows from Lemma 1.  $\square$

Our approach will be to select codewords randomly as i.i.d. Poisson processes. We will then bound the minimum  $\bar{\mathcal{L}}$ -time between pairs of codewords in these codes and use Lemma 2 to estimate their error probability. This will require the following lemma about the local time of random walks.

**Lemma 3** *Let  $\{X_t\}_{t=0}^\infty$  be a continuous-time random walk on  $\mathbb{Z}$  with  $X_0 = 0$  and transition rates*

$$q(n, n+1) = q(n, n-1) = \lambda > 0 \text{ for all } n \in \mathbb{Z}.$$

*Let  $\tau_t = \int_0^t 1(X_s = 0) ds$ . Then for all  $\delta > 0$  there exists  $C, \gamma > 0$  such that for all  $t \geq 0$ ,*

$$\Pr(\tau_t/t > \delta) \leq C \exp(-\gamma t).$$

*Proof.* Let  $U_0 = 0$  and for  $i \geq 1$ , let

$$\begin{aligned} V_i &= \inf\{t > U_{i-1} : X_t \neq 0\}, \\ U_i &= \inf\{t > V_i : X_t = 0\}, \end{aligned}$$

and

$$\begin{aligned} Y_i &= V_i - U_{i-1}, \\ Z_i &= U_i - V_i. \end{aligned}$$

Let  $N_t = \inf\{n : U_n \leq t\}$ . Then

$$\Pr(\tau_t/t > \delta) = \Pr(N_t > \delta\lambda t/2) + \Pr(\tau_t/t > \delta, N_t \leq \delta\lambda t/2).$$

Now  $E[Z_1] = \infty$ , so for any  $M \in \mathbb{N}$ , there exists  $M'$  such that if  $Z'_i = \min(Z_i, M')$  then  $M < E[Z'_1] < \infty$  and there exists  $\epsilon > 0$  such that

$$\Pr \left( \sum_{i=1}^n Z_i \leq Mn \right) \leq \Pr \left( \sum_{i=1}^n Z'_i \leq Mn \right) \leq \exp(-\epsilon n).$$

by Hoeffding's inequality [9]. Write  $T = \lceil \delta\lambda t/2 \rceil$ . Then there exists  $\epsilon > 0$  such that

$$\Pr(N_t > \delta\lambda t/2) \leq \Pr \left( \sum_{i=1}^T Z_i \leq t \right) \leq \exp(-\epsilon t).$$

And by Lemma 1, there exists  $\alpha > 0$  such that for all sufficiently large  $t$ ,

$$\begin{aligned} \Pr(\tau_t/t > \delta, N_t \leq \delta\lambda t/2) &\leq \Pr \left( \sum_{i=1}^{T+1} Y_i > t\delta \right) \\ &\leq \exp(-\alpha t). \square \end{aligned}$$

Define a  $n$ -point Poisson process with rate  $\lambda$  to be a random vector  $U^n$  such that  $\{U_i\}_{i=1}^n$  are i.i.d. exponential with mean  $1/\lambda$ . Define a *constrained*  $n$ -point Poisson process with rate  $\lambda$  to be a random vector  $V^n$  with distribution

$$\Pr(V^n \in A) = \Pr\left(U^n \in A \mid n/\lambda - \sqrt{n} < \sum_{i=1}^n U_i < n/\lambda + \sqrt{n}\right),$$

where  $U^n$  is a (unconstrained)  $n$ -point Poisson process with rate  $\lambda$ .

**Lemma 4** *For all  $\lambda, \delta > 0$ , there exists  $C, \gamma > 0$  such that for all  $n$  if  $U^n$  and  $V^n$  are independent constrained  $n$ -point Poisson processes with rate  $\lambda$ , then*

$$\Pr(\bar{\mathcal{L}}(U^n, V^n) < (1/2 - \delta)T) \leq C \exp(-\gamma n),$$

where  $T = n/\lambda + \sqrt{n}$ .

*Proof.* Let  $\tilde{U}^n$  and  $\tilde{V}^n$  be independent (unconstrained)  $n$ -point Poisson processes with rate  $\lambda$ , and define the event

$$A_n = \left\{ \sum_{i=1}^n \tilde{U}_i \geq n/\lambda - \sqrt{n}, \sum_{i=1}^n \tilde{V}_i \geq n/\lambda - \sqrt{n} \right\}.$$

Then

$$\Pr(\bar{\mathcal{L}}(U^n, V^n) < (1/2 - \delta)T) \leq \frac{\Pr(\bar{\mathcal{L}}(\tilde{U}^n, \tilde{V}^n) < (1/2 - \delta)T, A_n)}{\Pr(n/\lambda - \sqrt{n} < \sum_{i=1}^n \tilde{U}_i \leq n/\lambda + \sqrt{n})^2}.$$

Let  $X_t$  and  $\tau_t$  be as in the previous lemma, constructed so that  $\tilde{U}^n$  and  $\tilde{V}^n$  are the interevent times of the first  $n$  up-movements and the first  $n$  down-movements of  $X_t$ , respectively. Then

$$\Pr(\bar{\mathcal{L}}(\tilde{U}^n, \tilde{V}^n) < (1/2 - \delta)T, A_n) \leq \Pr(\tau_{n/\lambda - \sqrt{n}} > n/\lambda - \sqrt{n} - (1 - 2\delta)(n/\lambda + \sqrt{n})).$$

By the previous lemma, there exists  $C, \gamma > 0$  so that

$$\Pr(\tau_{n/\lambda - \sqrt{n}} > n/\lambda - \sqrt{n} - (1 - 2\delta)(n/\lambda + \sqrt{n})) \leq C \exp(-\gamma n).$$

This combined with the observation that

$$\inf_n \Pr\left(n/\lambda - \sqrt{n} < \sum_{i=1}^n \tilde{U}_i < n/\lambda + \sqrt{n}\right) > 0$$

completes the proof. □

**Proposition 2** *The zero-rate error exponent of the exponential-server timing channel with mean service time  $1/\mu$  satisfies  $E(0) \geq \mu/2$ .*

*Proof.* Let  $0 < \delta < 1/2$ . For any  $(n, M, T)$  code with codewords  $u_1^n, \dots, u_M^n$  that uses the pseudo-ML decoder, the probability of error of the  $i$ th codeword satisfies

$$P_{e,i} \leq \sum_{j \neq i} P(u_j^n \geq_d u_i^n | u_i^n).$$

Suppose in addition that  $T = n/\lambda + \sqrt{n}$  for some  $\lambda < (1/2 - \delta)\mu$ . Let

$$I_{ij} = \begin{cases} 1 & \text{if } \bar{\mathcal{L}}(u_i^n, u_j^n) \geq (1/2 - \delta)T \\ 0 & \text{otherwise.} \end{cases}$$

Then using Lemma 2 and the inequality  $(\sum_i x_i)^s \leq \sum_i x_i^s$  for all nonnegative  $x_i$  and all  $s \in (0, 1]$ , we have

$$P_{e,i}^s \leq \sum_{j \neq i} \left[ (1 - I_{ij}) + I_{ij} \exp \left( -n \left( (1/2 - \delta) \frac{\mu}{\lambda} - 1 - \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right) \right) \right) \right]^s,$$

Now fix  $\lambda < (1/2 - \delta)\mu$ , and consider a random ensemble of  $(n, M, n/\lambda + \sqrt{n})$  codes in which each codeword is chosen independently as a constrained  $n$ -point Poisson process with rate  $\lambda$ , the distribution of which we denote by  $Q_\lambda^n(\cdot)$ . Each code in the ensemble uses the pseudo-ML decoding rule. For each  $1 \leq i \leq M$ ,  $P_{e,i}^s$  averaged over the ensemble satisfies

$$\begin{aligned} \overline{P_{e,i}^s} &\leq \int \sum_{j \neq i} \left[ (1 - I_{ij}) + I_{ij} \exp \left( -n \left( (1/2 - \delta) \frac{\mu}{\lambda} - 1 - \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right) \right) \right) \right]^s \\ &\quad dQ_\lambda^n(u_1^n) \cdots dQ_\lambda^n(u_M^n). \end{aligned}$$

Exchanging the summation and the integral, then observing that all terms in the sum are equal, we apply Lemma 4 to obtain

$$\overline{P_{e,i}^s} \leq (M - 1) \left[ C \exp(-\gamma n) + \exp \left( -ns \left( (1/2 - \delta) \frac{\mu}{\lambda} - 1 - \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right) \right) \right) \right].$$

To translate this into a statement about the probability of error for a particular code, one can double the number of codewords in the codes and use Markov's inequality to upper bound the probability of error of the best half. The details are given in Gallager [4, p. 151], and the result is that for each  $s \in (0, 1]$  there exists a  $(n, M, n/\lambda + \sqrt{n})$  code for each  $n$  such that the probability of error of the  $i$ th codeword satisfies

$$\begin{aligned} P_{e,i} &\leq \\ &2^{1/s} (2M - 2)^{1/s} \left[ C \exp(-\gamma n) + \exp \left( -ns \left( (1/2 - \delta) \frac{\mu}{\lambda} - 1 - \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right) \right) \right) \right]^{1/s} \end{aligned}$$

for all  $1 \leq i \leq M$ . Equivalently, for all rates  $R > 0$ , there exists a  $(n, M, n/\lambda + \sqrt{n})$  code for each  $n$  with rate at least  $R$  such that

$$\begin{aligned} P_{e,i} &\leq \\ &4^{1/s} \exp(RT_n/s) \left[ C \exp(-\gamma n) + \exp \left( -ns \left( (1/2 - \delta) \frac{\mu}{\lambda} - 1 - \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right) \right) \right) \right]^{1/s}, \end{aligned}$$

where  $T_n = n/\lambda + \sqrt{n}$ . Thus

$$\begin{aligned} E(R) &\geq -R/s \\ &- \limsup_{n \rightarrow \infty} \frac{1}{sT_n} \log \left[ C \exp(-\gamma n) + \exp \left( -ns \left( (1/2 - \delta) \frac{\mu}{\lambda} - 1 - \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right) \right) \right) \right]. \end{aligned}$$

Thus for all sufficiently small  $s$ ,

$$E(R) \geq -R/s + (1/2 - \delta)\mu - \lambda - \lambda \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right).$$

So

$$E(0) \geq (1/2 - \delta)\mu - \lambda - \lambda \log \left( (1/2 - \delta) \frac{\mu}{\lambda} \right).$$

Since  $\lambda > 0$  and  $\delta > 0$  were arbitrary, the result follows.  $\square$

### 3 Remarks on Tightness

For a code with a large number of codewords chosen as Poisson processes, the minimum  $\bar{\mathcal{L}}$ -distance between pairs of distinct codewords is near  $T/2$  with high probability, where  $T$  is the blocklength. The next result implies that  $E(0) > \mu/2$  only if there is an  $\alpha > 1/2$  and a sequence of codes whose blocklengths tend to infinity while the data rates do not tend to zero, each with the property that if  $T$  is the blocklength then

$$\bar{\mathcal{L}}(u_i^n, u_j^n) \geq \alpha T,$$

for all distinct codewords  $u_i^n$  and  $u_j^n$ .

**Proposition 3** *If  $(u_1^n, \dots, u_M^n)$  is a  $(n, M, T)$  code (with any decoder) and*

$$\alpha = \inf_{i \neq j} \bar{\mathcal{L}}(u_i^n, u_j^n)/T,$$

*and  $P_e$  is the maximum probability of error over the  $M$  codewords then*

$$P_e \geq (1/2) \exp(-\mu\alpha T),$$

*where  $1/\mu$  is the mean service time.*

*Proof.* Suppose  $u_i^n$  and  $u_j^n$  have  $\bar{\mathcal{L}}(u_i^n, u_j^n)/T = \alpha$ . Let  $D_i \subset \mathbb{R}_+^n$  be the decision region for  $u_i^n$  and let  $D_j \subset \mathbb{R}_+^n$  be the decision region for  $u_j^n$ . Define the codeword  $u_0^n$  by  $u_0(t) = \min(u_i(t), u_j(t))$ , and let

$$Y_i = \{y^n \in \mathbb{R}_+^n : u_i^n \text{ is feasible for the output } y^n\}$$

and define  $Y_j$  and  $Y_0$  similarly. Then the probability of error for codeword  $i$  satisfies

$$\begin{aligned} P_i &= \int_{D_i^c \cap Y_i} \mu^n \exp(-\mu \mathcal{L}(u_i^n, y^n)) dy^n, \\ &\geq \int_{D_i^c \cap Y_0} \mu^n \exp(-\mu (\mathcal{L}(u_i^n, u_0^n) + \mathcal{L}(u_0^n, y^n))) dy^n, \\ &= \exp(-\mu \mathcal{L}(u_i^n, u_0^n)) \int_{D_i^c \cap Y_0} \mu^n \exp(-\mu \mathcal{L}(u_0^n, y^n)) dy^n, \\ &= \exp(-\mu\alpha T) \int_{D_i^c \cap Y_0} \mu^n \exp(-\mu \mathcal{L}(u_0^n, y^n)) dy^n, \end{aligned}$$

where we have used the fact that  $\bar{\mathcal{L}}$  is a metric and that  $\bar{\mathcal{L}}(a^n, b^n) = \mathcal{L}(a^n, b^n)$  if  $a(t) \geq b(t)$  for all  $t$ . Similarly,

$$P_j \geq \exp(-\mu\alpha T) \int_{D_j^c \cap Y_0} \mu^n \exp(-\mu \mathcal{L}(u_0^n, y^n)) dy^n.$$

But

$$\int_{Y_0} \mu^n \exp(-\mu \mathcal{L}(u_0^n, y^n)) dy^n = 1.$$

So since  $D_i$  and  $D_j$  are disjoint,

$$\int_{D_i^c \cap Y_0} \mu^n \exp(-\mu \mathcal{L}(u_0^n, y^n)) dy^n + \int_{D_j^c \cap Y_0} \mu^n \exp(-\mu \mathcal{L}(u_0^n, y^n)) dy^n \geq 1.$$

Therefore at least one of these integrals must be at least  $1/2$ .  $\square$

Thus to show  $E(0) = \mu/2$ , it suffices to show that all positive-rate codes have a minimum  $\bar{\mathcal{L}}$ -distance between pairs of  $T/2$ , or more precisely that for all  $\delta > 0$ , there exists a function  $f_\delta(n)$  that grows subexponentially with  $n$  such that for all  $(n, M, T)$  codes with  $M \geq f_\delta(n)$  there exists a pair of codewords  $u^n, v^n$  such that  $\bar{\mathcal{L}}(u^n, v^n)/T \leq 1/2 + \delta$ . We are unable to prove or disprove this assertion, but we note that a code described by Arikan [5] shows that  $f_\delta(n)$  must grow at least linearly in order for the assertion to hold: consider the  $(M-1, M, T_M)$  code where  $T_M = M/\mu$  such that the  $m$ th codeword places  $m-1$  jobs at time 0 and  $M-m$  jobs at time  $M/\mu$ , for  $1 \leq m \leq M$ . This code has a  $\bar{\mathcal{L}}$  distance of  $T_M$  between all pairs of codewords and the number of codewords grows linearly with the number of points.

A reader interested in investigating this problem should note that the normalized  $\bar{\mathcal{L}}$ -distance between two randomly chosen Poisson codewords has a wide distribution: its spread is on the same scale as its mean. This is in contrast to codewords for the BSC chosen as fair coin flips, and codewords for the Gaussian channel chosen as independent Gaussian random variables, for which the distance distribution is concentrated around its mean. For the ESTC then, either the Poisson distribution is not the optimum one with which to choose codewords at low rates, or the best low-rate codes have a wide distance spectrum, so that the usual Plotkin-type bound on the minimum distance fails to establish the tightness of our lower bound on  $E(0)$ .

## References

- [1] R. Sundaresan and S. Verdú, “Robust decoding for timing channels,” *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 405–419, Mar. 2000.
- [2] J. Giles and B. Hajek, “An information-theoretic and game-theoretic study of timing channels,” *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2455–2477, 2002.
- [3] V. Anantharam and S. Verdú, “Bits through queues,” *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.
- [4] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
- [5] E. Arikan, “On the reliability exponent of the exponential timing channel,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1681–1689, Jun. 2002.
- [6] Y. M. Kabanov, “The capacity of a channel of the Poisson type,” *Theory Probab. Appl.*, vol. 23, pp. 143–147, 1978.
- [7] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel—Part I,” *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.
- [8] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel—Part II,” *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1462–1471, Nov. 1988.
- [9] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford University Press, Oxford, 2nd edition, 1992.