The Rate-Distortion Function of a Poisson Process with a Queueing Distortion Measure

Todd P. Coleman   Negar Kiyavash   Vijay G. Subramanian

Abstract

This paper characterizes the rate distortion function of a Poisson process with a queuing distortion measure that is in complete analogy with the proofs associated with the rate distortion functions of a Bernoulli source with Hamming distortion measure and a Gaussian source with squared-error distortion measure. Analogous to those problems, the distortion measure that we consider is related to the logarithm of the conditional distribution relating the input to the output of a well-known channel coding problem, specifically the Anantharam and Verdú "Bits through Queues" [1] coding problem. We show this problem is equivalent to a standard rate-distortion problem due to: i) the independent increments property of the Poisson process ii) the numerical entropy rate of any finite-rate point process tending to 0 , iii) the existence of a reproduction with finite expected distortion, iv) the additive structure of the distortion measure. Our Shannon lower bound involves a number of mutual information inequalities, one of which exploits the maximum-entropy property of the exponential distribution. We also show that the rate-distortion functions pertaining to expected distortion and deviation probability are equivalent. We conclude with a comparison to other rate-distortion formulations of the Poisson process in the literature.

I. INTRODUCTION

In their landmark "Bits through Queues" paper Anantharam & Verdú [1] illustrated the fundamental limits of coding through the *timing* of packets in queuing systems. For a first-come, first-serve (FCFS)

queuing system with exponentially distributed service times - also termed the exponential server timing channel (ESTC), the capacity was elegantly established in closed form. The remarkable aspect of [1] is that despite a queueing system being nonlinear, and non-stationary with infinite memory, there still exists a closed-form solution. This fact, along with the fact that an ESTC can be seen as an "entropy-increasing operator" [2, Sec. 2],[3, Thm 1] and also the saddle-point property of mutual information for the ESTC [1, Thm 3], makes the ESTC appear to be analogous to the binary symmetric channel (BSC) and additive white Gaussian noise (AWGN) channels.

The latter two channels have canonical rate-distortion problems associated with them, where the distortion measure is related to the logarithm of the conditional distribution relating the input to the output of the channels. A thorough discussion of this and the applicability of this to optimality of uncoded transmissions can be found in [4]. In these two cases (AWGN and BSC), for the associated rate-distortion problems, simple mutual information inequalities exploiting maximum-entropy distributions and entropy manipulations lead to a closed-form lower bound - which can be shown to be tight by developing elegant "test-channel" achievability arguments [5, Sec. 10.3].

This paper considers the continuous-time development of the rate-distortion function of a Poisson process, with a queuing distortion measure. The distortion measure is dual - in the same sense as the BSC/Hamming and AWGN/squared-error dualities - to the ESTC. Specifically, the distortion measure that we consider is proportional to the logarithm of the conditional distribution relating the input to the output of an ESTC. This illustrates that one can reason about this rate-distortion problem completely from a traditional memoryless, expectation-based additive expected distortion measure perspective, by exploiting four key properties:

1) the independent increments property of the Poisson process,
2) the numerical entropy rate of any finite-rate point process tending to 0 (Lemma 2.1),
3) the existence of a reproduction with finite expected distortion (Lemma 2.2), and
4) the additivity of the discrete-time distortion measure (25).

This allows us to equate the rate-distortion function - with an expected distortion criterion, as well as a probability-of-deviation distortion criterion - to the standard Shannon lower bound. We then exploit a number of simple mutual information inequalities - one of which exploits the maximum-entropy property of the exponential distribution - to develop a Shannon lower bound. The 'test-channel' to characterize the joint distribution achieving the lower bound with equality exploits Burkes' theorem from queuing theory [6], [7]. We conclude with a comparison to other rate-distortion formulations of the Poisson process, with

a variety of distortion measures, in the research literature.

A. Notation on Point Processes

As this discussion revolves around considering timings, we consider a point process $\mathcal{P}$ with occurrence times $\{P_1, P_2, \ldots\}$. The counting function $N_\mathcal{P}(t)$ of a point process depicts the number of arrivals that have occurred up to and including time t:

$$N_\mathcal{P}(t) = \sup\{n \in \mathbb{N} : P_n \leq t\}.$$

Define $\mathcal{S}_T$ to be the set of all counting functions on $[0, T]$:

$$\mathcal{S}_T \triangleq \{N : [0, T] \to \mathbb{Z}_+ \text{ s.t. } N \text{ is nondecreasing and right-continuous }\}. \tag{1}$$

The entropy on $[0, T]$ of a point process $\mathcal{P}$ with arrival times $\{P_1, P_2, \ldots\}$ is defined [8] as the sum of its *numerical entropy* and its *positional entropy*:

$$h_T(\mathcal{P}) := H(N_\mathcal{P}(T)) + h\left(P_1, \ldots, P_{N_\mathcal{P}(T)} | N_\mathcal{P}(T)\right),$$

where $H(\cdot)$ is discrete entropy, $h(\cdot)$ is differential entropy, and $\{P_1, \ldots, P_{N_\mathcal{P}(T)}\}$ are the locations (in time) of the arrivals. The entropy of a Poisson process of rate λ over $(0, T]$ is given in closed form by [8]

$$h_T(\mathcal{P}) = T\lambda(1 - \log \lambda).$$

We can define the joint entropy on $[0, T]$ of two point processes $\mathcal{P}^a$ and $\mathcal{P}^b$ by

$$h_T(\mathcal{P}^a, \mathcal{P}^b) := H(M_a, M_b) + E_{M_a, M_b}[h(P_1^a, \ldots, P_{m_a}^a, P_1^b, \ldots, P_{m_b}^b | M_a = m_a, M_b = m_b)],$$

where M_a and M_b are the random variables denoting the number of arrivals of processes $\mathcal{P}^a$ and $\mathcal{P}^b$ respectively, and $\{P_1^a, \ldots, P_{m_a}^a\}$ and $\{P_1^b, \ldots, P_{m_b}^b\}$ are the locations (in time) of the arrivals of processes $\mathcal{P}^a$ and $\mathcal{P}^b$ given $M_a = m_a$ and $M_b = m_b$. From this we can define the conditional point process entropy simply as

$$h_T(\mathcal{P}^a | \mathcal{P}^b) := h_T(\mathcal{P}^a, \mathcal{P}^b) - h_T(\mathcal{P}^b).$$

B. Related Work

Given the entropy-maximizing property of the exponential distribution and the pre-eminence of the Poisson process in point-process theory [9], rate distortion functions under various distortion measures have been sought in the literature [10], [11], [12], [13], [14]. A large class of distortion measures, including

those that consider a magnitude error criterion and the reproduction of the number of events in a given interval and the times between successive events, lead to non-causal reproductions [10], [11]. Gallager [12] considered transmission of point processes through a causal system such that the output point process always follows the input but the order of messages need not be preserved. Considering the output of such a system to be a reproduction of the input point process and using average delay per message as the distortion measure, Gallager [12] derived a minimum information rate for the Poisson process for a given fidelity of reproduction that included both the timing information and the message ordering. Verdú [13] considered a different approach where the message inter-arrival times were reproduced to a fidelity of D seconds under the constraint that the last arrival is declared no sooner than it occurs. Bedekar [14] insisted upon in-order delivery of the messages through the system and defining the distortion measure to be the average service-time D of a hypothetical FCFS queue that would result in the output. Related work pertaining to minimizing mutual information across special classes of causal systems is the jamming work of Giles and Hajek [15].

C. Queuing Timing Channels

Now consider a FCFS queuing system with an arrival point process $\mathcal{A} = \{a_1, a_2, \ldots\}$ at its input, service times $\mathcal{S} = \{s_1, s_2, \ldots\}$, and a departure process $\mathcal{D} = \{d_1, d_2, \ldots\}$ at its output, as depicted in Figure 1(a). As an example, consider Figure 1(b). Note that the service time s_1 for the first packet is given by $d_1 - a_1 = 6$. Note that the first packet does not depart from the queue until after the second packet arrives. Thus, the service time s_2 for the second packet is given by $s_2 = d_2 - d_1$, because the server starts working on the second packet once the first packet departs. The second packet departs before the third arrival a_3. Thus, the third service time is simply $s_3 = d_3 - a_3$. So in general, for any FCFS queuing system [7], [16]:

$$s_i = d_i - \max(a_i, d_{i-1}) \tag{2}$$

A special case of FCFS queuing timing channel is the "exponential server timing channel (ESTC)" where service times are i.i.d. and exponentially distributed of rate μ. On the interval $[0, T]$, the likelihood is known to have the structure [6], [17]

$$P\left(\underline{d}|\underline{a}\right) \quad \propto \quad \begin{cases} \exp\left\{-\mu \int_0^T 1_{\{N_\mathcal{D}(t) > N_\mathcal{A}(t)\}} dt\right\} & \text{if } N_\mathcal{D}(t) \geq N_\mathcal{A}(t) \; \forall \, t \; \in \; [0, T] \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In their landmark "Bits through Queues" paper Anantharam & Verdú [1] considered the information-theoretic problem of coding through the *timing* of packets in packet systems. For a memoryless arrival
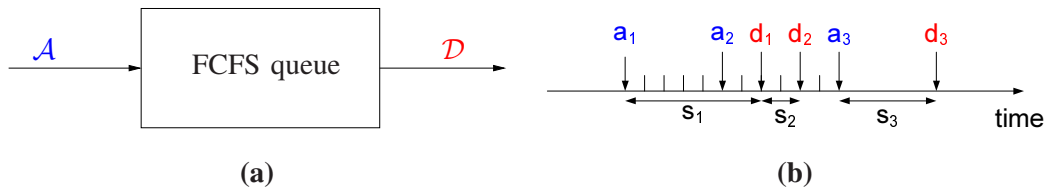
Fig. 1. FCFS queue: **(a)** The arrivals (blue) and departures (red) from a FCFS $\cdot/M/1$ Queue. **(b)** Note that the service time s_1 for the first packet is given by $d_1 - a_1 = 6$. Note that the first packet does not depart from the queue until after the second packet arrives. Thus, the service time s_2 for the second packet is given by $s_2 = d_2 - d_1$, because the server starts working on the second packet once the first packet departs. The second packet departs before the third arrival a_3. Thus, the third service time is simply $s_3 = d_3 - a_3$.

process of rate λ, the First-Come First-Serve (FCFS) $\cdot/M/1$ continuous-time queue with service rate $\mu > \lambda > 0$ has a capacity $C(\lambda)$ given by

$$C(\lambda) = \lambda \log \frac{\mu}{\lambda}, \quad \lambda < \mu \text{ nats/s.} \tag{4}$$

Then the capacity of the FCFS $\cdot/M/1$ continuous-time queue with service rate μ is given by the maximum of $C(\lambda)$ over all possible arrival rates $\lambda < \mu$, namely,

$$C = e^{-1}\mu \text{ nats/s,} \tag{5}$$

where the maximum corresponding to (5) is achieved in (4) at $\lambda = e^{-1}\mu$.

D. Canonical Rate-Distortion Formulations

The remarkable aspect of the "Bits through Queues" result [1] is that despite a queueing system having non-linearities, memory, and non-stationarities, there still exists a closed-form solution for the FCFS $\cdot/M/1$ queue. This fact, along with the fact that the exponential service timing channel (ESTC) can be seen as an "entropy-increasing operator" [2, Sec. 2],[3, Thm 1] and also the saddle-point property of mutual information for the ESTC [1, Thm 3], makes the ESTC appear to be analogous to the binary symmetric channel (BSC) and additive white Gaussian noise (AWGN) channels. The BSC and AWGN dualities are precise: the distortion measure is related to the channel probability model by

$$\rho(x, \hat{x}) \propto -\log p(x|\hat{x}) \tag{6}$$

The BSC and AWGN dual distortion problems have proofs involving simple mutual information inequalities exploiting maximum-entropy distributions, leading to a closed-form lower bound, which is shown to be tight by developing an elegant "test-channel" [5, Sec. 10.3]. This dual relationship has been recently
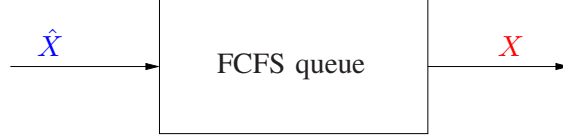
Fig. 2. The diagram of interpreting $\mathcal{X}$ as the output of a FCFS queueing system with $\hat{\mathcal{X}}$ as the input.

shown to have additional implications for uncoded transmission [4]. Wagner and Anantharam [17] embarked upon understanding the similarities between the BSC, AWGN, and ETSC channels by developing a distance metric for the $\cdot/M/1$ queuing system that is analogous to Hamming/Euclidean distances for BSC/AWGN channels; in that, it is related to the logarithm of the channel's conditional density.

II. THE CANONICAL RATE-DISTORTION FUNCTION IN ANALOGY WITH "BITS THROUGH QUEUES"

A. Distortion Measure

Consider two point processes $\mathcal{X}$ and $\hat{\mathcal{X}}$ over $[0, T]$. Denote the occurrence times (counting functions) of $\mathcal{X}$ by $\{X_i\}$ ($N_{\mathcal{X}}(t)$) and the occurrence times (counting functions) of $\hat{\mathcal{X}}$ by $\{\hat{X}_i\}$ ($N_{\hat{\mathcal{X}}}(t)$). For any $\mathcal{X}$ and $\hat{\mathcal{X}}$ such that

$$N_{\hat{\mathcal{X}}}(t) \geq N_{\mathcal{X}} \qquad \forall t \in [0, T], \tag{7}$$

define

$$\mathcal{S} = \mathcal{X} \diamond \hat{\mathcal{X}} \tag{8}$$

as the sequence of induced service times $\{S_i\}$ of a FCFS queuing system with $\hat{\mathcal{X}}$ as the input and $\mathcal{X}$ as the output, as shown in Figure 2. Specifically, $S_i = X_i - \max(X_{i-1}, \hat{X}_i)$; see Figure 1 for the validity of this relationship.

With this definition, define the distortion between any two realizations $\mathbf{x}$ and $\hat{\mathbf{x}}$ of point processes $\mathcal{X}$ and $\hat{\mathcal{X}}$ as:

$$\rho_T(\mathbf{x}, \hat{\mathbf{x}}) \quad = \quad \begin{cases} \frac{1}{T} \int_0^T 1_{\{N_{\hat{\mathbf{x}}}(t) > N_{\mathbf{x}}(t)\}} dt & \text{if } N_{\hat{\mathbf{x}}}(t) \geq N_{\mathbf{x}}(t) \; \forall t \in [0, T] \\ \infty & \text{otherwise} \end{cases} \tag{9}$$

We note that $\int_0^T 1_{\{N_{\hat{\mathbf{x}}}(t) > N_{\mathbf{x}}(t)\}} dt$ is intimately related to the sum of the service times. If $N_{\hat{\mathcal{X}}}(T) = N_{\mathcal{X}}(T)$, then we note from Figure 3 that $\sum_{i=1}^{N_{\mathcal{X}}(T)} s_i = \int_0^T 1_{\{N_{\hat{\mathbf{x}}}(t) > N_{\mathbf{x}}(t)\}} dt$. Otherwise, if $N_{\hat{\mathcal{X}}}(T) >$
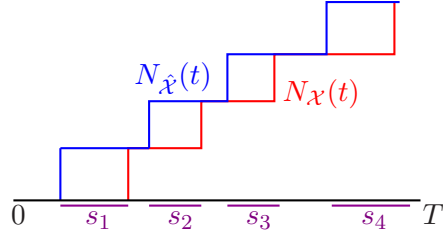
Fig. 3. An example sample path realization of the counting functions for $\mathcal{X}$ and $\hat{\mathcal{X}}$ as well as the induced service times

$N_{\mathcal{X}}(T)$, then $\sum_{i=1}^{N_{\mathcal{X}}(T)} s_i$ contribute to the integral, along with the remaining time from the final departure time, $d_{N_{\mathcal{X}}(T)}$, until T. Thus it follows that

$$\int_0^T 1_{\{N_{\hat{\mathcal{X}}}(t) > N_{\mathcal{X}}(t)\}} dt = \sum_{i=1}^{N_{\mathcal{X}}(T)} s_i + 1_{\{N_{\hat{\mathcal{X}}}(T) > N_{\mathcal{X}}(T)\}}(T - d_{N_{\underline{d}}(T)}) \tag{10}$$

Figure 3 illustrates an example sample path realization of the counting functions for $\mathcal{X}$ and $\hat{\mathcal{X}}$ as well as the induced service times. Note that the measure (9) satisfies $\rho(x, \hat{x}) \propto -\log p(x|\hat{x})$, where the channel model is given by (3).

We now illustrate that our distortion measure is of the "single-letter" type [10]. By defining

$$\rho(n_x, n_{\hat{x}}) \triangleq \begin{cases} 0, & n_x = n_{\hat{x}} \\ 1, & n_x < n_{\hat{x}} , \\ \infty, & n_x > n_{\hat{x}} \end{cases} \tag{11}$$

it follows that

$$\rho_T(\mathcal{X}, \hat{\mathcal{X}}) = \frac{1}{T} \int_0^T \rho\left(N_{\mathcal{X}}(t), N_{\hat{\mathcal{X}}}(t)\right) dt. \tag{12}$$

Also note from (11) that for any c:

$$\rho(n_x, n_{\hat{x}}) = \rho(n_x + c, n_{\hat{x}} + c). \tag{13}$$

B. Rate-Distortion Definitions

For the rate-distortion problem, consider a class of codebooks $\{\mathcal{C}_T = (e_T, d_T) : T \geq 0\}$ consisting of an encoder map $e_T : \mathcal{S}_T \to \{1, \dots, M_T\}$ and a decoder map $d_T : \{1, \dots, M_T\} \to \mathcal{S}_T$. We denote $|\mathcal{C}_T| \triangleq M_T$. We say the class of codebooks $\{\mathcal{C}_T = (e_T, d_T) : T \geq 0\}$ is of rate R if

$$\lim_{T \to \infty} \frac{\log |\mathcal{C}_T|}{T} = R.$$

In this paper we will consider two types of rate-distortion formulations:

- *Expected Value Criterion*: For a fixed $D \geq 0$, we define $R_E(D)$ to be the infimum of all R for which there exists a class of codebooks $\{\mathcal{C}_T = (e_T, d_T) : T \geq 0\}$ of rate R such that

$$\limsup_{T \to \infty} E\left[\rho_T(\mathcal{X}, d_T(e_T(\mathcal{X})))\right] \leq D. \tag{14}$$

- *Deviation Probability Criterion*: For a fixed $D \geq 0$, we define $R_P(D)$ to be the the infimum of all R for which there exists a class of codebooks $\{\mathcal{C}_T = (e_T, d_T) : T \geq 0\}$ of rate R such that

$$\limsup_{T \to \infty} P\left(\rho_T(\mathcal{X}, d_T(e_T(\mathcal{X}))) > D\right) = 0. \tag{15}$$

In general, in most scenarios, $R_P(D) \geq R_E(D)$ [18, Thm 1]. It will be shown later that for the problem of interest in this setting, $R_E(D) = R_P(D) = -\lambda \log D$.

  a) Shannon Information definitions: Define the Shannon Information $\mathcal{I}(D)$ as

$$\mathcal{I}(D) = \limsup_{T \to \infty} \inf_{P_{\mathcal{X}, \hat{\mathcal{X}}} : E[\rho_T(\mathcal{X}, \hat{\mathcal{X}})] \leq D} \frac{1}{T} I_T(\mathcal{X}; \hat{\mathcal{X}}). \tag{16}$$

Now given that

$$
\begin{aligned}
I_T(\mathcal{X}; \hat{\mathcal{X}}) &= h_T(\mathcal{X}) - h_T(\mathcal{X}|\hat{\mathcal{X}}) \\
&= h_T(\mathcal{X}) - H\left(N_{\mathcal{X}}(T)|\hat{\mathcal{X}}\right) - h_T\left(X_1, \ldots, X_{N_{\mathcal{X}}(T)}|N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right),
\end{aligned} \tag{17}
$$

we define $\tilde{I}_T(\mathcal{X}; \hat{\mathcal{X}})$ as

$$\tilde{I}_T(\mathcal{X}; \hat{\mathcal{X}}) = h_T(\mathcal{X}) - h_T\left(X_1, \ldots, X_{N_{\mathcal{X}}(T)}|N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right). \tag{18}$$

We now show that the term $H\left(N_{\mathcal{X}}(T)|\hat{\mathcal{X}}\right)$ that is present in $I_T(\mathcal{X}; \hat{\mathcal{X}})$ and missing in $\tilde{I}_T(\mathcal{X}; \hat{\mathcal{X}})$ is asymptotically negligible [10]:

  Lemma 2.1: For any joint distribution on point processes $\mathcal{X}$ and $\hat{\mathcal{X}}$ such that $\lim_{T \to \infty} \frac{1}{T} E[N_{\mathcal{X}}(T)] = \lambda < \infty$,

$$\lim_{T \to \infty} \frac{1}{T} H\left(N_{\mathcal{X}}(T)|\hat{\mathcal{X}}\right) = \lim_{T \to \infty} \frac{1}{T} H\left(N_{\mathcal{X}}(T)\right) = 0.$$

  Proof: Note that a geometric random variable of parameter p has mean given by $m = \frac{1}{p} - 1$, entropy given by $-\frac{1-p}{p} \log(1-p) - \log p$, and thus its entropy expressed in terms of m is given by

$$H_G(m) = m \log\left(1 + \frac{1}{m}\right) + \log(m + 1).$$

Now, for any $\epsilon > 0$ and for all $T > T_\epsilon$, $\frac{E[N_{\mathcal{X}}(T)]}{T} < \lambda + \epsilon$. So since the geometric random variable maximizes the entropy of all discrete non-negative random variables of fixed mean, we have that for all

$T > T_\epsilon$:

$$0 \leq \frac{1}{T} H\left(N_\mathcal{X}(T)|\hat{\mathcal{X}}\right) \leq \frac{1}{T} H\left(N_\mathcal{X}(T)\right) \leq \frac{1}{T} H_G\left((\lambda + \epsilon)T\right)$$

$$= (\lambda + \epsilon)\log\left(1 + \frac{1}{(\lambda + \epsilon)T}\right) + \frac{1}{T}\log\left((\lambda + \epsilon)T + 1\right).$$

$$\Rightarrow \lim_{T \to \infty} \frac{1}{T} H\left(N_\mathcal{X}(T)|\hat{\mathcal{X}}\right) = \lim_{T \to \infty} \frac{1}{T} H\left(N_\mathcal{X}(T)\right) = 0. \tag{19}$$

∎

The above Lemma will become useful in developing two-stage coding schemes, where one of the stages has asymptotically negligible rate, and the other stage corresponds to standard source coding of an i.i.d. process with an additive distortion measure. Also, note that with Lemma 2.1, we have that

$$\tilde{\mathcal{I}}(D) \triangleq \limsup_{T \to \infty} \inf_{P_{\mathcal{X},\hat{x}}:E[\rho_T(\mathcal{X},\hat{x})] \leq D} \frac{1}{T} \tilde{I}_T(\mathcal{X};\hat{\mathcal{X}}) \tag{20}$$

$$= \mathcal{I}(D). \tag{21}$$

C. Source Coding Theorems for the Poisson Process with Queuing Distortion Measure

We now note that even thought the Poisson process is neither stationary nor ergodic, it has independent increments. Secondly, the cost in rate to communicate $\{N_\mathcal{X}(T) - N_\mathcal{X}(0), N_\mathcal{X}(2T) - N_\mathcal{X}(T), \ldots, N_\mathcal{X}(nT) - N_\mathcal{X}((n-1)T)\}$ in bits per second, by virtue of Lemma 2.1, is arbitrarily small for increasing T. This motivates the following coding scheme [10]:

- First communicate the sequence of n duration-T jumps $\{N_\mathcal{X}(T) - N_\mathcal{X}(0), N_\mathcal{X}(2T) - N_\mathcal{X}(T), \ldots, N_\mathcal{X}(nT) - N_\mathcal{X}((n-1)T)\}$ using a standard lossless entropy code. Since each duration-T jump is a Poisson random variable of rate λT, from Lemma 2.1, this will require an arbitrarily small bits per second for sufficiently large T.

- Given knowledge of the n duration-T jumps:

  - By defining the duration-T increments of the counting process,

$$\underline{X}_k \triangleq (N_\mathcal{X}(t) - N_\mathcal{X}((k-1)T) : t \in ((k-1)T, kT]) \in \mathcal{S}_T, \quad k = 1, 2, \ldots, \tag{22a}$$

$$\underline{\hat{X}}_k \triangleq \left(N_{\hat{\mathcal{X}}}(t) - N_\mathcal{X}((k-1)T) : t \in ((k-1)T, kT]\right) \in \mathcal{S}_T, \quad k = 1, 2, \ldots, \tag{22b}$$

it follows that, since Poisson processes have independent increments, the sequence $\{\underline{X}_k\}$ is i.i.d.

– The discrete-time distortion measure $\underline{\rho}_n(\underline{X}^n, \underline{\hat{X}}^n) \triangleq \rho_{nT}(\mathcal{X}, \hat{\mathcal{X}})$ is additive:

$$
\begin{aligned}
\underline{\rho}_n(\underline{X}^n, \underline{\hat{X}}^n) =& \rho_{nT}(\mathcal{X}, \hat{\mathcal{X}}) \\
=& \frac{1}{nT} \int_0^{nT} \rho\left(N_{\mathcal{X}}(t), N_{\hat{\mathcal{X}}}(t)\right) dt & (23) \\
=& \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \int_{(i-1)T}^{iT} \rho\left(N_{\mathcal{X}}(t), N_{\hat{\mathcal{X}}}(t)\right) dt \\
=& \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \int_{(i-1)T}^{iT} \rho\left(N_{\mathcal{X}}(t) - N_{\mathcal{X}}((i-1)T), N_{\hat{\mathcal{X}}}(t) - N_{\mathcal{X}}((i-1)T)\right) dt & (24) \\
=& \frac{1}{n} \sum_{i=1}^n \underline{\rho}_1(\underline{X}_i, \hat{\underline{X}}_i). & (25)
\end{aligned}
$$

where (23) follows from (12), (24) follows from (13), and (25) follows from (22) and (12). So we have developed a standard rate-distortion problem with additive distortion measure over abstract alphabet $\mathcal{S}_T$.

Given the above developed approach, we can reason about this rate-distortion problem in terms of classical discrete-time rate-distortion theory with additive distortion measure and abstract symbol alphabet $\mathcal{S}_T$. In order for those classical theorems to apply [19, Sec 7.2], the following Lemma establishing a required technical condition is presented:

Lemma 2.2: There exists an $\hat{\underline{x}}^* \in \mathcal{S}_T$ such that for $\underline{X}_1 \in \mathcal{S}_T$:

$$
E[\underline{\rho}_1(\underline{X}_1, \hat{\underline{x}}^*)|\underline{X}_1(T) = m] < \infty \tag{26}
$$

Proof: Simply select $\hat{\underline{x}}^*$ to be the counting function for which $\hat{\underline{x}}^*(0) = \hat{\underline{x}}^*(T) = m$. Since for any $\tau \in [0, T)$, $\underline{X}_1(\tau) = N_{\mathcal{X}}(\tau) \leq \underline{X}_T(\tau) = N_{\mathcal{X}}(T) = m$, it follows that $\hat{\underline{x}}^*$ will always lead $\underline{X}_1$ on $(0, T]$. Thus, from the distortion measure, (12), the conditional expectation in (26) will be upper bounded by 1. ∎

Now enabled with the additivity of the distortion measure (25), the required technical condition given by Lemma 2.2, the independent increments property of the Poisson process, and the negligible numerical entropy of the duration-T jumps given by Lemma 2.1, we are in position to state the following theorem, due to to Rubin:

Theorem 2.3 ([10], Thm 1): Consider the Poisson source $\mathcal{X} = \{N_{\mathcal{X}}(t), t \geq 0\}$ having rate-distortion function $R_E(D)$ with respect to the single-letter fidelity criterion (12). Then $R_E(D) = \mathcal{I}(D)$.

The essence of the proof of [10, Thm 1] is the fact that a Poisson process has independent increments, and that Lemma 2.1 holds. So for additive distortion measures for which the technical condition pertaining

to Lemma 2.2 holds, this is simply a standard rate-distortion problem over an abstract alphabet - where it is known that the information rate $R_E(D) = \mathcal{I}(D)$ [19, Sec 7.2].

Next, we show that the rate-distortion functions $R_E(D)$ and $R_P(D)$, pertaining to expected distortion and deviation probability of distortion, are equivalent. The essence of this statement is that even though the distortion measure given by (25) can possibly be infinite, if it is finite, then it is upper bounded by 1. Also note that there is a simple way to generate encoders such that ∞ is never attained: use the reproduction specified in the proof of Lemma 2.2. Specifically, given the n duration-T jumps $\{N_{\mathcal{X}}(iT) - N_{\mathcal{X}}((i-1)T), \;\; i = 1, 2, \ldots\}$ at the encoder and decoder, simply use a codebook with one possible reproduction being $(\underline{x}_1^*, \ldots, \underline{x}_n^*)$, where $\underline{x}_i^*(0) = \underline{x}_i^*(T) = N_{\mathcal{X}}(iT) - N_{\mathcal{X}}((i-1)T)$. Thus we arrive at:

Theorem 2.4: Consider the Poisson source $\mathcal{X} = \{N_{\mathcal{X}}(t), t \geq 0\}$ having rate-distortion function $R_E(D)$ with respect to the single-letter fidelity criterion (12). Then $R_E(D) = R_P(D) = \mathcal{I}(D)$.

The proof basically follows in in two parts. The strong converse statement, that $R_P(D) \geq R_E(D)$ follows from Kieffer [18]. The opposite direction, that $R_P(D) \leq R_E(D)$, directly follows Steinberg and Verdú 's proof of [20, Thm 11] and basically exploits the strong converse and the continuity of $\mathcal{I}(D)$. The details are in Appendix I.

At this point, we have shown that $R_E(D) = R_P(D) = \mathcal{I}(D)$, namely that the rate-distortion function operationally, for this distortion measure, is given by the Shannon bound on mutual information. We are now able to characterize the structure of $\mathcal{I}(D)$:

Theorem 2.5: Consider the Poisson source $\mathcal{X} = \{N_{\mathcal{X}}(t), t \geq 0\}$ having rate-distortion function $R_E(D)$ with respect to the single-letter fidelity criterion (12). Then

$$R_E(D) = R_P(D) = \mathcal{I}(D) = \begin{cases} -\lambda \log D \text{ bits/second} & D \in (0,1); \\ 0 & D \geq 1. \end{cases}$$

Proof: Fix a distortion value $D \in (0,1)$. Define

$$\mu \triangleq \frac{\lambda}{D}. \tag{27}$$

Note that for any joint distribution $P_{\mathcal{X}, \hat{\mathcal{X}}}$ such that

$$E[\rho_T(\mathcal{X}, \hat{\mathcal{X}})] \leq D \tag{28}$$

holds, we have:

$$\tilde{I}_T(\mathcal{X}; \hat{\mathcal{X}}) = h_T(\mathcal{X}) - h_T\left(X_1, \ldots, X_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right)$$

$$= \lambda T(1 - \log \lambda) - h_T\left(X_1, \ldots, X_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right) \tag{29}$$

$$= \lambda T(1 - \log \lambda) - h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right) \tag{30}$$

$$\geq \lambda T(1 - \log \lambda) - h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T)\right) \tag{31}$$

$$\geq \lambda T(1 - \log \lambda) - \lambda T(1 - \log \mu) \tag{32}$$

$$= -\lambda T \log D. \tag{33}$$

where (29) follows because $\mathcal{X}$ is a Poisson(λ) process; (30) follows from Appendix II, (31) follows because conditioning reduces entropy; (32) follows from the maximum-entropy argument in Appendix III, and (33) follows from (27). Thus $\tilde{\mathcal{I}}(D) = \mathcal{I}(D) \geq -\lambda \log D$.

Now, to show that $-\lambda \log D$ is tight, consider the output of an ESTC with rate $\mu > \lambda$ in steady-state with input process $\hat{\mathcal{X}}$, and denote the output process as $\mathcal{X}$. Note that by Burke's theorem [6], [7] the departure process $\mathcal{X}$ is also a Poisson process with rate λ. Note then that:

$$\tilde{I}_T(\mathcal{X}; \hat{\mathcal{X}}) = \lambda T(1 - \log \lambda) - h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right)$$

$$= \lambda T(1 - \log \lambda) - h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T)\right) \tag{34}$$

$$= \lambda T(1 - \log \lambda) - \lambda T(1 - \log \mu) \tag{35}$$

$$= -\lambda T \log D,$$

where (34) holds since service and arrival times are independent in M/M/1 queues; and (35) holds because M/M/1 queues have i.i.d. exponential service times - so the bound in Appendix III is tight. ∎

III. DISCUSSION

The rate distortion function in bits per second given here is equivalent to that in [14]. The distortion measure (9) and the measure in [14] are related and the achievability of the minimum rate using Poisson code-words passed through an $M/M/1$ queue in steady-state is also similar. However, in Bedekar [14]

a) considers the distortion formulation in terms of n arrivals and n departures, rather than the $[0, nT]$ interval as we do here.

b) relies completely on the probability of deviation distortion criterion,

c) requires the reproduction codewords $\hat{\mathcal{X}}$ to always *follow* $\mathcal{X}$, i.e. $\mathcal{X}$ is the arrival process to the queue, and $\hat{\mathcal{X}}$ is the output.

Distinction a) forces his approach to rely upon the information spectrum and sup-information rates [20] to characterize the rate-distortion function. We were able to reason about this problem completely from a traditional memoryless, expectation-based rate-distortion perspective, by exploiting four key properties: i) the independent increments of the Poisson process, ii) the numerical entropy rate tending to 0 (Lemma 2.1), iii) the technical condition pertaining to existence of a reproduction with finite expected distortion (Lemma 2.2), and iv) the additivity of the discrete-time distortion measure (25).

Distinction b) can be relaxed: we have shown here that the deviation probability criterion and the expected distortion criterion lead to the same rate-distortion function.

Distinction c) is the opposite of our requirement, but it is in essence equivalent. Specifically, if the roles of $\mathcal{X}$ and $\hat{\mathcal{X}}$ in the distortion measure are reversed, i.e. if we are to interpret $\mathcal{X}$ as the input to a queue and $\hat{\mathcal{X}}$ as the output, we arrive at the same rate-distortion function:

1) Since the rate of the departure process of a stable queuing system is equal to the rate of the arrival process of any stable queuing system [16], and since the arrival process will now be Poisson of rate λ, the departure process $\hat{\mathcal{X}}$ will also have rate λ, i.e. $\lim_{T\to\infty} \frac{N_{\hat{\mathcal{X}}}(T)}{T} = \lambda$. Thus Lemma 2.1 still holds.

2) Lemma 2.2 still holds, by simply selecting $\underline{x}^*$ to now be the all-zero counting function: $\underline{x}^*(0) = \underline{x}^*(T) = 0$.

3) Theorems 2.3 and 2.4 directly still hold.

4) Theorem 2.5 still holds by simply replacing the functional roles of $\mathcal{X}$ as now the input to the queue, $\hat{\mathcal{X}}$ as the output of the queue; defining the service time accordingly: $s_i = \hat{x}_i - \max(x_i, \hat{x}_{i-1})$; and modifying (29)-(33) as follows:

$$\tilde{I}_T(\mathcal{X}; \hat{\mathcal{X}}) = \lambda T(1 - \log \lambda) - h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{\mathcal{X}}\right)$$

$$\geq \lambda T(1 - \log \lambda) - h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\hat{\mathcal{X}}}(T)\right) \tag{36}$$

$$\geq \lambda T(1 - \log \lambda) - E[N_{\hat{\mathcal{X}}}(T)](1 - \log \mu) \tag{37}$$

where (37) follows from the maximum-entropy argument in Appendix III. As mentioned above, for any stable queuing system, the departure process will have rate equal to the arrival process (which

is Poisson), so $\lim_{T\to\infty} \frac{N_{\tilde{\mathcal{X}}}(T)}{T} = \lambda$. Thus we have:

$$\limsup_{T\to\infty} \frac{1}{T}\tilde{I}_T(\mathcal{X};\hat{\mathcal{X}}) \geq \lambda(1 - \log\lambda) - \lim_{T\to\infty}\frac{N_{\hat{\mathcal{X}}}(T)}{T}(1 - \log\mu)$$

$$= \lambda(1 - \log\lambda) - \lambda(1 - \log\mu)$$

$$= -\lambda\log D \tag{38}$$

where (38) follows from (27). Thus $\tilde{\mathcal{I}}(D) = \mathcal{I}(D) \geq -\lambda\log D$. The tightness of $-\lambda\log D$ follows directly from Burke's Theorem.

The rate-distortion function, in bits per second, is also equivalent to that of [13]. However, the motivation for [13] appears to be the extremal properties of the exponential distribution rather than on queuing. Note that all the code-words in [13] are such that the arrival points (except for the last arrival) of the reproductions can either lead or follow the arrival points of $\mathcal{X}$. There are also other subtle differences in the constraints and distortion measures in our approach *vis-a-vis* [13], namely,

1) <u>Distortion measure</u>: [13] uses the following distortion measure for the reproductions given the source realization

$$\hat{X}_i - \hat{X}_{i-1}1_{\{i>1\}} \leq X_i - X_{i-1}1_{\{i>1\}} + D, \ \forall\, i \in \{1, \ldots, m\},$$

where D is a length of time while we consider $\frac{1}{T}E[\sum_{i=1}^{N_{\mathcal{X}}(T)} S_i] \leq D$ where D is a dimension-less number;

2) <u>Timing constraint</u>: [13] imposes the constraints that the reproductions produce the same number of arrivals as the source realization with the last arrival of the reproduction being after the last arrival of the source realization, i.e.,

$$N_{\mathcal{X}}(T) = N_{\hat{\mathcal{X}}}(T) \text{ and } x_{N_{\mathcal{X}}(T)} \leq \hat{x}_{N_{\tilde{\mathcal{X}}}(T)},$$

whereas we insist that the reproductions completely lead/lag the original, i.e.,

$$\forall\, t \in [0, T]:\ N_{\mathcal{X}}(t) \leq N_{\hat{\mathcal{X}}}(t) \quad \text{(lead)}$$

$$\forall\, t \in [0, T]:\ N_{\hat{\mathcal{X}}}(t) \leq N_{\mathcal{X}}(t) \quad \text{(lag)}$$

3) <u>Nature of Reproductions</u>: A final but important distinction is that our reproductions are always Poisson processes whereas those in [13] are not even renewal processes, although they are, in a sense, close to being renewal processes.

IV. CONCLUSIONS

In conclusion we would like to emphasize that our main motivation for this work was to develop a mutual information lower-bounding technique using maximum-entropy arguments, and show achievability with an appropriate test-channel - to be in complete analogy to the methodology developed for the AWGN and BSC dualized rate-distortion problems. Also note that the exact same line of analysis follows for the discrete-time queueing case [21], [3].

ACKNOWLEDGEMENTS

REFERENCES

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

[2] B. Prabhakar and N. Bambos, "The entropy and delay of traffic processes in ATM networks," in *Proceedings of the Conference on Information Science and Systems (CISS)*, Baltimore, Maryland, 1995, pp. 448–453.

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

[4] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[5] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

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

[7] R. Gallager, *Discrete Stochastic Processes*. Boston, MA: Kluwer, 1996.

[8] J. McFadden, "The entropy of a point process," *SIAM Journal of Applied Mathematics*, vol. 13, pp. 988–994, 1965.

[9] D. Daley and D. Vere-Jones, *An Introduction to the theory of point processes*. New York: Springer-Verlag, 1988.

[10] I. Rubin, "Information rates and data-compression schemes for Poisson processes," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 200–210, 1974.

[11] H. Sato and T. Kawabata, "Information rates for Poisson point processes," *Transactions of IEICE*, vol. E70, no. 9, pp. 817–822, Sept. 1987.

[12] R. G. .Gallager, "Basic limits on protocol information in data communication networks," *IEEE Transactions on Information Theory*, vol. 22, no. 4, pp. 385–398, 1976.

[13] S. Verdú, "The exponential distribution in information theory," *Problems of Information Transmission*, vol. 32, no. 1, pp. 86–95, Jan-Mar 1996.

[14] A. S. Bedekar, "On the information about message arrival times required for in-order decoding," in *Proceedings of the Internation Symposium on Information Theory (ISIT)*, Washington, D.C., 2201, p. 227.

[15] 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.

[16] L. Kleinrock, *Queueing Systems. Vol 1: Theory.* New York, NY: Wiley, 1975.

[17] A. B. Wagner and V. Anantharam, "Zero-rate reliability of the exponential-server timing channel," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 447–465, February 2005.

[18] J. Kieffer, "Strong converses in source coding relative to a fidelity criterion," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 257–262, 1991.

[19] T. Berger, *Rate distortion theory.* Prentice-Hall Englewood Cliffs, NJ, 1971.

[20] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *Information Theory, IEEE Transactions on*, vol. 42, no. 1, pp. 63–86, 1996.

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

[22] A. Berger, V. Della Pietra, and S. Della Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[23] J. Jeon and R. Manmatha, "Using Maximum Entropy for Automatic Image Annotation," *Image and Video Retrieval: Third International Conference, CIVR 2004, Dublin, Ireland, July 21-23, 2004: Proceedings*, 2004.

APPENDIX I

PROOF OF THEOREM 2.4

We discuss the proof in two parts. The strong converse statement, that $R_P(D) \geq R_E(D)$ follows from Kieffer [18]. The opposite direction, that $R_P(D) \leq R_E(D)$, directly follows Steinberg and Verdú 's proof of [20, Thm 11] and basically exploits the strong converse and the continuity of $\mathcal{I}(D)$.

- $R_P(D) \geq R_E(D)$:

  Proof: Consider a sequence of codes $\{\mathcal{C}_n : n \geq 0\}$ consisting of encoders $e_n : \mathcal{S}_T^n \to \{1, \ldots, 2^{nRT}\}$ and decoders $d_n : \{1, \ldots, 2^{nRT}\} \to \mathcal{S}_T^n$ where clearly $\{\mathcal{C}_n : n \geq 0\}$ is of rate R. Then note that we can create an augmented code $\{\mathcal{C}_n^* : n \geq 0\}$ also of rate R, consisting of two parts. First, from Lemma 2.1, it encodes the duration-T jumps, $\{N_{\mathcal{X}}(kT) - N_{\mathcal{X}}((k-1)T)\}$, at arbitrarily small rate for large T. Secondly, the sequence of encoders $e_n^* : \mathcal{S}_T^n \to \{1, \ldots, 2^{nRT} + 1\}$ and decoders $d_n^* : \{1, \ldots, 2^{nRT} + 1\} \to \mathcal{S}_T^n$ act as the original code $\mathcal{C}_n$ except for any $\underline{X}^n \in \mathcal{S}_T^n$ for which $\rho_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) = \infty$. For those $\underline{x}^n$, $e(\underline{x}^n) = 2^{nRT} + 1$ and $d(2^{nRT} + 1)$ is given by $(\hat{\underline{x}}_1^*, \ldots, \hat{\underline{x}}_n^*)$ where $\hat{\underline{x}}_i^*$, given in the proof of Lemma 2.2. This allows for $\rho_n\left(\underline{X}^n, d_n^*(e_n^*(\underline{X}^n))\right)$ to always be upper-bounded by 1. Specifically:

  - $e_n^*(\underline{x}^n, \hat{\underline{x}}^n) = e_n(\underline{x}^n, \hat{\underline{x}}^n)$ if $\rho_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) < \infty$. Otherwise, $e_n^*(\underline{x}^n, \hat{\underline{x}}^n) = 2^{nRT} + 1$
  - $d_n^*(1) = d_n(1), \ldots d_n^*\left(2^{nRT}\right) = d_n\left(2^{nRT}\right)$
  - $d_n^*\left(2^{nRT} + 1\right) = (\hat{\underline{x}}_1^*, \ldots, \hat{\underline{x}}_n^*)$, where $\hat{\underline{x}}_k^*(0) = \hat{\underline{x}}_k^*(T) = N_{\mathcal{X}}(kT) - N_{\mathcal{X}}((k-1)T)$.

  Then note that for any $\underline{X}^n \in \mathcal{S}_T^n$, $\rho_n\left(\underline{X}^n, d_n^*(e_n^*(\underline{X}^n))\right) < \infty$ and so from (9) it is upper bounded by 1. Thus from the strong converse results of Kieffer [18, Thm 1], [20, Thm 12] we have that: *for*

the sequence of codes $\tilde{e}_n$ and $\tilde{d}_n$ with uniformly bounded distortions, if $R < R_E(D)$, then

$$\limsup_{n\to\infty} P\left(\underline{\rho}_n\left(\underline{X}^n, d_n^*(e_n^*(\underline{X}^n))\right) > D\right) = 1.$$

Since by construction, $\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \geq \underline{\rho}_n\left(\underline{X}^n, d_n^*(e_n^*(\underline{X}^n))\right)$, and secondly, $\{\mathcal{C}_n : n \geq 0\}$ and $\{\tilde{\mathcal{C}}_n : n \geq 0\}$ are both of rate R, it follows that whenever $R < R_E(D)$:

$$\limsup_{n\to\infty} P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) > D\right) \geq \limsup_{n\to\infty} P\left(\underline{\rho}_n\left(\underline{X}^n, d_n^*(e_n^*(\underline{X}^n))\right) > D\right) = 1.$$

Thus from the definition of $R_P(D)$, we have that $R_P(D) \geq R_E(D)$. ∎

- $R_P(D) \leq R_E(D)$:

Proof: This is essentially the same as the proof of [20, Thm 11]. This will be a proof by contradiction. Suppose that $R_E(D) < R_P(D)$. Consider a sequence of codes $\mathcal{C}_n$ with encoders and decoders e_n, d_n of rate $R_E(D)$ for which $E[\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right)] \leq D$. Since $R_E(D) = \mathcal{I}(D)$ and $\mathcal{I}(D)$ is continuous in D, there exists a sufficiently small $\delta > 0$ such that $R < R_P(D + \delta)$. Thus from the definition of $R_P(D + \delta)$, it must be that

$$\limsup_{n\to\infty} P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) > D + \delta\right) > 0.$$

So there must exist some $\alpha > 0$ and a subsequence $\mathcal{J}$ for which $\alpha_n > \alpha : n \in \mathcal{J}$ for which

$$P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) > D + \delta\right) = \alpha_n, \quad n \in \mathcal{J}.$$

Then note that we have from the law of iterated expectations that

$$
\begin{aligned}
D \;\geq\;& E[\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right)] \\
=\;& E[\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \mid \underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \leq D]P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \leq D\right) \\
+\;& E[\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \mid \underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \in (D, D+\delta]]P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \in (D, D+\delta]\right) \\
+\;& E[\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \mid \underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) > D + \delta]P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) > D + \delta\right) \\
\geq\;& 0 + D\left(1 - \alpha_n - P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \leq D\right)\right) + (D+\delta)\alpha_n
\end{aligned}
$$

and thus we have that

$$P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) \leq D\right) \geq \frac{\delta \alpha_n}{D} > \frac{\delta \alpha}{D}, \quad n \in \mathcal{J}.$$

But this implies then that for this sequence of codes at rate $R = R_E(D)$:

$$\lim_{n\to\infty} P\left(\underline{\rho}_n\left(\underline{X}^n, d_n(e_n(\underline{X}^n))\right) > D\right) \neq 1,$$

which contradicts strong converse statement from above. ∎

APPENDIX II

PROOF OF (30)

Proof: First note from (2), and since $N_{\hat{\mathcal{X}}}(T) \geq N_{\mathcal{X}}(T)$, that $\{S_i\}$, $i \in \{1, \ldots, N_{\mathcal{X}}(T)\}$ are uniquely defined given $\{X_1, \ldots, X_{N_{\mathcal{X}}(T)}\}$ and $\hat{X}$. Also, from the queueing interpretation, it is clear that given $\hat{\mathcal{X}}$, $\mathcal{S}$ uniquely defines X. However, the point process entropy contains both a discrete numerical entropy component as well as a positional entropy, where the latter relies upon a differential entropy term - so just illustrating this bijection does not suffice. We now focus on the differential entropy term pertaining to the positional entropy and show there is no *amplitude scaling*. To do this, we consider conditioning upon $\hat{\mathcal{X}} = \hat{x}$ and investigate the Jacobian of the mapping between $\mathcal{X}$ and $\mathcal{S}$. Now it suffices to show that

$$h(X_1, \ldots, X_m | N_{\mathcal{X}}(T) = m, \hat{X} = \hat{x}) = h(S_1, \ldots, S_m | N_{\mathcal{X}}(T) = m, \hat{X} = \hat{x}).$$

But note that the service process $\{S_i\}$, given $\{\hat{X}_i\}$, is related to the departure process $\{X_i\}$ by the simple (recursive) relationship [7], [16]

$$S_i = X_i - \max(X_{i-1}, \hat{x}_i).$$

Also see Figure 1. Considering specific realizations note that the mapping from $(x_1, \ldots, x_m)$ to $(s_1, \ldots, s_m)$ is differentiable except on a set $\mathcal{A} \in \Re_+^m$ with Lebesgue measure zero (pertaining to the points of discontinuity of the functions $\max(x_i, s_i)$). We can consider the Jacobian at any point $(x_1, \ldots, x_m)$ not in $\mathcal{A}$, and note that

$$|J(x_1, \ldots, x_m)| = \begin{vmatrix} \frac{\partial s_1}{\partial x_1} & \frac{\partial s_1}{\partial x_2} & \cdots & \frac{\partial s_1}{\partial x_m} \\ \frac{\partial s_2}{\partial x_1} & \frac{\partial s_2}{\partial x_2} & \cdots & \frac{\partial s_2}{\partial x_m} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_m}{\partial x_1} & \frac{\partial s_m}{\partial x_2} & \cdots & \frac{\partial s_m}{\partial x_m} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \cdots & 0 & 0 \\ \frac{-1 \pm 1}{2} & 1 & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{-1 \pm 1}{2} & 1 \end{vmatrix} = 1.$$

The $\frac{-1 \pm 1}{2}$ simply denotes that if $x_{i-1} > \hat{x}_i$, then $\max(x_{i-1}, \hat{x}_i) = x_{i-1}$ and thus the corresponding Jacobian entry in the $(i, i-1)$ point of the matrix is equal to -1; otherwise, if $x_{i-1} < \hat{x}_i$ then $\max(x_{i-1}, \hat{x}_i) = \hat{x}_i$, and thus the corresponding Jacobian entry in the $(i, i-1)$ point of the matrix is equal to 0. No matter what the specific values in the $(i, i-1)$ locations of the Jacobian are, note that the determinant is always 1.

Thus it follows that

$$h(X_1, \ldots, X_m | N_{\mathcal{X}}(T) = m, \hat{X} = \hat{x}) = h(S_1, \ldots, S_m | N_{\mathcal{X}}(T) = m, \hat{X} = \hat{x}).$$

Consequently,

$$h\left(X_1, \ldots, X_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{X}\right) = h\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T), \hat{X}\right)$$

■

APPENDIX III

PROOF OF (32)

Proof: Consider the conditional entropy [22], [23] form of the maximum-entropy problem [5, Ch 12]. In general, for a fixed $Q(x)$, we would like to maximize

$$H(P) = -\sum_{x,y} Q(x)P(y|x)\log P(y|x)$$

subject to

$$\sum_{x,y} Q(x)P(y|x)f(x,y) \leq C.$$

By introducing a penalized Lagrangian, we would like to maximize

$$\mathcal{L}(\gamma, P) = H(P) + \gamma\left(\sum_{x,y} Q(x)P(y|x)f(x,y) - C\right).$$

The solution may be obtained by maximizing P and selecting γ so that the constraint is met. The solution of such problems can be formed using Lagrange multipliers. The resultant equations become

$$P^*(y|x) = \frac{1}{Z(x)}\exp\left[\gamma f(x,y)\right]$$

where $Z(x)$ is a normalization obtained to guaranteed that $\sum_y P(y|x) = 1$.

So in going from (32) to (33), we are interesting in upper-bounding the conditional entropy

$$h_T\left(S_1, \ldots, S_{N_{\mathcal{X}}(T)} | N_{\mathcal{X}}(T)\right),$$

subject to the constraint (28). We will consider maximizing according to a slightly weaker constraint:

$$E\left[\frac{1}{T}\sum_{i=1}^{N_{\mathcal{X}}(T)} S_i\right] \leq D \tag{39a}$$

$$\Leftrightarrow E\left[\sum_{i=1}^{N_{\mathcal{X}}(T)} S_i\right] \leq \frac{\lambda T}{\mu} \tag{39b}$$

We note from (9) and (10) that if any joint distribution satisfies (28), then it satisfies (39).

So now consider understanding the optimal structure of that the conditional density of the service time process

$$f_{S^n|N_{\mathcal{X}}(T)}\left(s^n|n\right)$$

in order to maximize

$$h_T\left(S_1,\ldots,S_{N_{\mathcal{X}}(T)}|N_{\mathcal{X}}(T)\right) = -\sum_{n=0}^{\infty}\int_{s^n} P_{N_{\mathcal{X}}(T)}(n)\, f_{S^n|N_{\mathcal{X}}(T)}(s^n|n)\log f_{S^n|N_{\mathcal{X}}(T)}(s^n|n)\, ds^n$$

subject to the constraints (39) that

$$\sum_{n=0}^{\infty}\int_{s^n} P_{N_{\mathcal{X}}(T)}(n)\, f_{S^n|N_{\mathcal{X}}(T)}(s^n|n)\left[\sum_{i=1}^{n} s_i\right] ds^n \leq \frac{\lambda T}{\mu}.$$

Thus, using the maximum-conditional-entropy formulation from above, we have that the conditional-entropy maximizing service time conditional pdf will have the form

$$f^*_{S^n|N_{\mathcal{X}}(T)}(s^n|n) = \frac{1}{Z(n)}\exp\left[\gamma\sum_{i=1}^{n} s_i\right].$$

It follows using straight-forward calculations that $\gamma = -\mu$ and $Z(n) = \frac{1}{\mu^n}$:

$$f^*_{S^n|N_{\mathcal{X}}(T)}(s^n|n) = \mu^n\exp\left[-\mu\sum_{i=1}^{n} s_i\right].$$

With this, it follows then that

$$
\begin{aligned}
h_T\left(S_1,\ldots,S_{N_{\mathcal{X}}(T)}|N_{\mathcal{X}}(T)\right) &\leq \sum_{n=0}^{\infty}\int_{s^n} P_{N_{\mathcal{X}}(T)}(n)\, f^*_{S^n|N_{\mathcal{X}}(T)}(s^n|n)\log f^*_{S^n|N_{\mathcal{X}}(T)}(s^n|n)\, ds^n \\
&= \sum_{n=0}^{\infty}\int_{s^n} P_{N_{\mathcal{X}}(T)}(n)\, f^*_{S^n|N_{\mathcal{X}}(T)}(s^n|n)\left(n\log\mu - \mu\sum_{i=1}^{n} s_i\right) ds^n \\
&= E[N_{\mathcal{X}}(T)]\log\mu - \mu E\left[\sum_{i=1}^{N_{\mathcal{X}}(T)} S_i\right] \\
&= \lambda T(1-\log\mu). \qquad\qquad (40)
\end{aligned}
$$

where (40) follows from (39), (27), and since $\mathcal{X}$ is a Poisson process of rate λ. ■