About Priority Encoding Transmission

Stéphane Boucheron, *Member, IEEE,* and
Mohammad Reza Salamatian, *Student Member, IEEE*

***Abstract*—Recently, Albanese *et al.* introduced priority encoding transmission (PET) for sending hierarchically organized messages over lossy packet-based computer networks [1]. In a PET system, each symbol in the message is assigned a priority which determines the minimal number of codeword symbols that is required to recover that symbol. This note revisits the PET approach using tools from network information theory. We first outline that priority encoding transmission is intimately related with the broadcast erasure channel with degraded message set. Using the information spectrum approach, we provide an informational characterization of the capacity region of general broadcast channels with degraded message set. We show that the PET inequality has an information-theoretical counterpart: The inequality defining the capacity region of the broadcast erasure channel with degraded message sets. Hence the PET approach which consists in time-sharing and interleaving classical erasure-resilient codes achieves the capacity region of this channel. Moreover, we show that the PET approach may achieve the sphere packing exponents. Finally, we observe that on some simple nonstationary broadcast channels, time-sharing may be outperformed. The impact of memory on the optimality of the PET approach remains elusive.**

***Index Terms*—Broadcast channels, coding exponents, erasure-resilient codes, information spectrum, priority encoding transmission.**

I. Introduction

The quality of packet voice and image on the Internet has been mediocre due, in part, to congestion-induced packet losses. From the end-user viewpoint, the Internet can actually be modeled as an erasure channel acting over the large input alphabet formed by IP packets. On an erasure channel, each input symbol is either faithfully transmitted or erased, independently from its value. The output alphabet contains the input alphabet plus a special symbol denoting erasure. Although retransmission upon request, automatic repeat request (ARQ) has traditionally been the way to turn computer networks into reliable channels, the delay requirement of multimedia applications eliminates the possibility of retransmission and renews the interest for forward error correcting (FEC) [2]. Potential users of FEC also have to take into account that standard multimedia compression techniques [4] introduce a hierarchical structure in the information source.

The priority encoding transmission (PET) approach has been motivated by the search for robust multicasting of digital video sequences conforming to the MPEG standard [1]. In a first approximation, the different sorts of frames constituting a GOP (group of pictures in the MPEG methodology [4]) are assumed to form independent sources (this would be true if compression were perfect). When using PET, the source information is protected in such a way that even receivers undergoing high loss rates can reconstruct essential parts of the source flow (for example, I-frames), while receivers undergoing lower loss rates can reconstruct most of the flow.

This note outlines the connection between the pragmatically motivated PET approach and the *broadcast channel with degraded message set* described in multiuser information theory [5]. In such a broadcast channel, one transmitter tries to send $\boldsymbol{k}$ independent messages $m_0, \cdots m_{\boldsymbol{k}-1}$ (corresponding to different priority levels) to $\boldsymbol{k}$ receivers (enjoying different reception conditions). The messages are multiplexed by a channel encoder into a sequence of input symbols x, $x = f(m_0 \cdots m_{\boldsymbol{k}-1})$. The $\boldsymbol{i}$th receiver $(0 \leq \boldsymbol{i} < \boldsymbol{k})$ gets a corrupted version $y_{\boldsymbol{i}}$ of the input x, and tries to reconstruct messages $m_0 \cdots m_{\boldsymbol{i}}$. At least from an intuitive viewpoint, the task faced by priority encoding transmission and coping with broadcast channels are similar. To elaborate further, let us recall more precisely the PET code description.

A PET code over alphabet $\mathcal{X}$, with message length m, code length n, and nondecreasing priority function β is a pair of mappings f (encoder) from $\mathcal{X}^m$ to $\mathcal{X}^n$ and ϕ (decoder) from $(\mathcal{X} \cup \{\mathsf{e}\})^n$ to $\mathcal{X}^m \cup \{\text{reject}\}$ such that if w' is obtained from $f(w)$ by erasing at most $n - \beta(i)$ symbols, then $\phi(w')$ coincides with w at least on the first i symbols. The values in the range of a priority function are called the priority levels. In the sequel, the $\boldsymbol{i}$th level of the priority function is denoted by $\beta^{\boldsymbol{i}}$ for $0 \leq \boldsymbol{i} < \boldsymbol{k}$.

The rate (resp., normalized rate) of a code represents the fraction of information bits (resp., symbols) per symbol. Formally, the rate R (resp., normalized rate R) of a code of length n, with M codewords over alphabet $\mathcal{X}$ is $\frac{1}{n} \log \mathsf{M}$ (resp., $\frac{1}{n} \log_{|\mathcal{X}|} \mathsf{M}$), all logarithms being given in base 2. Following [1], a tuple of normalized rates $(\tilde{\mathsf{R}}_0, \cdots \tilde{\mathsf{R}}_{k-1})$ is achieved by a PET system if and only if there are exactly $n\tilde{\mathsf{R}}_i$ symbols from the message that are protected at level i. Though PET was not presented in a Shannon-theoretical perspective, the following relevant inequality is proved in [1, Theorems 3.3, 5.4]. For any PET system

$$\sum_{0=i}^{k-1} \frac{n\tilde{\mathsf{R}}_i}{\beta^i} \leq 1. \tag{1}$$

Inequality (1) puts combinatorial limits on finite sets of fixed-length words. When restricted, for example, to the case of one priority level, it coincides with the Singleton bound: $\tilde{\mathsf{R}}_0 \leq \beta^0/n$. Note that the latter bound cannot be achieved by codes of arbitrary length (cf., for example, [1, Theorem 6.1]). On a given alphabet, arbitrarily long codes satisfying the PET inequality do not exist and thus cannot be used to achieve arbitrarily reliable transmission on lossy broadcast channels.

Moreover, the PET approach advocates a very simple method to design broadcast codes. Packets are divided into slots, each slot size defining an alphabet. For each priority level, apply a good point-to-point erasure-resilient code over the small alphabet corresponding to the slot size and then interleave the resulting codewords in the packets. This is a version of the engineering approach called time-sharing in [6]. This approach is known to be nonoptimal on many broadcast channels. Assessing the PET approach from an information-theoretical viewpoint amounts to first checking whether the capacity region of broadcast erasure channels with degraded message sets can be exhausted using priority encoding transmission. In the affirmative, the second natural question is: Does error probability decline as fast as possible when PET is used?

We first use the *information spectrum approach* [7] to provide an informational characterization of the capacity region of general broadcast channels with degraded message set (DMS) in Section II. This characterization illustrates the relevance to general broadcast channels with memory of superposition codes as introduced in [6]. Then we show in Section III that the PET approach achieves the capacity region of memoryless broadcast erasure channels with DMS and, moreover, that it may achieve the sphere-packing exponent of this channel. The last sec-

Manuscript received October 21, 1997; revised September 22, 1999. This research was supported in part by CNET under Grant 96 1-B 212, and by Esprit Working Group RAND II.

The authors are with the LRI, CNRS UMR 8623, Bât 490, Université Paris-Sud, 91405 Orsay, France (e-mail: {bouchero; salamat}@lri.fr.).

Communicated by F. R. Kschischang, Associate Editor for Coding Theory.
Publisher Item Identifier S 0018-9448(00)01689-8.

tion shows that on some simple nonstationary broadcast erasure channels, time-sharing may be outperformed.

II. Broadcast Channel with Degraded Message Set

A. General Definitions

In the sequel $\boldsymbol{X}$, $\boldsymbol{Y}$ denote families of input processes and their corresponding output through a channel $\boldsymbol{W}$. For each n, X^n denotes a random variable over $\mathcal{X}^n$ and Y^n is distributed over $\mathcal{Y}^n$ according to $W^n(.|X^n)$. $X^n(i)$, $Y^n(i)$ denotes the ith element of X^n, Y^n.

The entropy of a random variable X, $H(X)$ is defined by

$$H(X) = -\sum_i \Pr\{X = i\} \log \Pr\{X = i\}.$$

The mutual information between X and Y is defined by

$$I(X;Y) \stackrel{\Delta}{=} -\sum_{x,\,y} \Pr\{X = x\, Y = y\} \times \log \frac{\Pr\{X = x\, Y = y\}}{\Pr\{X = x\} \Pr\{X = y\}}.$$

The *information spectrum approach* has been proposed recently to handle systems with general dependencies [7]. It relies on the notion of *liminf in probability*: $p - \liminf_{n\to\infty} X_n$ is defined as the supremum of α such that $\limsup_{n\to\infty} \Pr\{X_n < \alpha\} = 0$. Given two random processes $(\boldsymbol{X}, \boldsymbol{Y})$, the mutual *information spectrum-inf* is defined by

$$\underline{I}(\boldsymbol{X};\boldsymbol{Y}) \stackrel{\Delta}{=} p - \liminf_{n\to\infty} \frac{1}{n} \times \log \frac{\Pr\{X^n = x^n\, Y^n = y^n\}}{\Pr\{X^n = x^n\} \times \Pr\{Y^n = y^n\}}.$$

B. Broadcast Channels

To avoid confusion, sources (corresponding to requested priority levels) and component channels (corresponding to different levels of transmission reliability) will be indexed using boldface indices $\boldsymbol{i}, \cdots, \boldsymbol{k}$.

Definition 1: A k-ary broadcast channel $\boldsymbol{W}$ consists of a sequence of joint probability transitions $W^n(Y_{\boldsymbol{0}}^n \cdots Y_{\boldsymbol{k-1}}^n|X^n)$ from $\mathcal{X}^n$ toward $\mathcal{Y}^{n\times k}$. The marginal probability transitions $W_{\boldsymbol{i}}(Y_{\boldsymbol{i}}^n|X^n)$ are called the component channels.

A *degraded message set* can be transmitted at rate $(\mathsf{R}_{\boldsymbol{0}}, \cdots \mathsf{R}_{\boldsymbol{k-1}})$ over $\boldsymbol{W}$ with error probability ϵ if and only if there exists a family of (broadcast) codes

$$f^n(m_0, \cdots m_{\boldsymbol{k-1}}) \mapsto x$$

and

$$\phi_{\boldsymbol{i}}^n(y_{\boldsymbol{i}}) = (\hat{m}_{\boldsymbol{0}}, \cdots \hat{m}_{\boldsymbol{i}})$$

such that for almost all block length n, $\liminf(\log |M_{\boldsymbol{i}}|/n) \geq \mathsf{R}_{\boldsymbol{i}}$ and for all $\boldsymbol{i} < \boldsymbol{k}$, the error probability experienced by the ith receiver satisfies

$$\limsup\, \mathsf{e}_{\boldsymbol{i}}(f^n, \phi_i^n) \stackrel{\Delta}{=} \max_{m_0, \cdots m_{k-1}} W[\phi_{\boldsymbol{i}}^n(y_{\boldsymbol{i}}) \neq (m_{\boldsymbol{0}} \cdots m_{\boldsymbol{i}})|f^n(m_{\boldsymbol{0}}, \cdots m_{\boldsymbol{k-1}})] \leq \epsilon.$$

The rate-tuple $(\mathsf{R}_{\boldsymbol{i}})$ is than said to be ϵ-achievable over $\boldsymbol{W}$. A broadcast code with block length n, rates $(\mathsf{R}_{\boldsymbol{i}})$, and error rate smaller than ϵ for all receivers on channel $\boldsymbol{W}$ is called an $(n, \mathsf{R}_{\boldsymbol{0}} \cdots \mathsf{R}_{\boldsymbol{k-1}}, \epsilon)$-code over $\boldsymbol{W}$. A tuple of rates is achievable if it is ϵ-achievable for all $\epsilon > 0$.

Remarks:

1) From the definition, it is immediate that the set of achievable (resp., ϵ-achievable) rates is closed.
2) For broadcast channels, achievable rates do not depend on whether we consider average (over codewords) or worst case error probability (cf. [5], where no assumption on channel memory is made). In the sequel, we will adopt the most convenient viewpoint depending on the situation.

The memoryless broadcast channel with degraded message set has received a single letter characterization [5, Theorems III.4.1 and 3], [8]. The information spectrum approach allows to give an informational (though not computational) characterization of operationally defined achievable rates over general broadcast channels.

Let us now define the class of families of input processes $\mathcal{S}$ as $\boldsymbol{U}_{\boldsymbol{0}}, \cdots \boldsymbol{U}_{\boldsymbol{k-2}}, \boldsymbol{U}_{\boldsymbol{k-1}}$ with $\boldsymbol{U}_{\boldsymbol{k-1}} \stackrel{\Delta}{=} \boldsymbol{X}$ and corresponding output families $\boldsymbol{Y}_{\boldsymbol{i}=\boldsymbol{0}}^{\boldsymbol{k-1}}$ as follows. For every block length n, $U_{\boldsymbol{i}}^n$ is independent from $U_{\boldsymbol{i+2}}^n \cdots X^n$, Y^n conditionally on $U_{\boldsymbol{i+1}}^n$, and $Y_{\boldsymbol{i}}^n$ is distributed according to $W_{\boldsymbol{i}}^n(.|X^n)$. Let us insist on the fact that the variables $U_{\boldsymbol{i}}^n$ $(\boldsymbol{i} < \boldsymbol{k-1})$ do not necessarily live in an n–dimensional product space. Such families of input processes are general since no consistency constraint between processes with different indices n and m is imposed. The structure of input families is inspired from the superposition codes construction [6].

$\mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}})$ is the set of tuple of rates $(\mathsf{R}_{\boldsymbol{0}} \cdots \mathsf{R}_{\boldsymbol{k-1}})$ satisfying

$$0 \leq \mathsf{R}_{\boldsymbol{i}} \leq \underline{I}(\boldsymbol{U}_{\boldsymbol{i}};\boldsymbol{Y}_{\boldsymbol{i}}|\boldsymbol{U}_{\boldsymbol{i-1},\cdots\boldsymbol{0}}), \qquad \text{for } \boldsymbol{i} < \boldsymbol{k}$$

$$0 \leq \sum_{\boldsymbol{j}\leq\boldsymbol{i}} \mathsf{R}_{\boldsymbol{j}} \leq \underline{I}(\boldsymbol{U}_{\boldsymbol{i}};\boldsymbol{Y}_{i}), \qquad \text{for } \boldsymbol{i} < \boldsymbol{k}.$$

Proposition 1: The set of achievable rates $(\mathsf{R}_{\boldsymbol{0}}, \mathsf{R}_{\boldsymbol{1}}, \cdots \mathsf{R}_{\boldsymbol{k-1}})$ over a broadcast channel $\boldsymbol{W}$ with degraded message set is the closure of

$$\bigcup_{\boldsymbol{U}_{\boldsymbol{0}},\cdots\boldsymbol{U}_{\boldsymbol{k-1}}\in\mathcal{S}} \mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}}).$$

Proposition 1 can be completed by the determination of the region of ϵ-achievable rate tuples and by the characterization of those broadcast channels that have the strong converse property.

$\mathsf{R}_{\boldsymbol{W}}(\epsilon, \boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}})$ is the set of tuple of rates $(\mathsf{R}_{\boldsymbol{0}} \cdots \mathsf{R}_{\boldsymbol{k-1}})$ satisfying

$$\limsup_{n\to\infty} \Pr\left\{ \vee_{\boldsymbol{i}}\ \frac{1}{n} \log \frac{\Pr\{Y_{\boldsymbol{i}}^n, U_{\boldsymbol{i}}^n \mid U_{\boldsymbol{0}}^n, \cdots, U_{\boldsymbol{i-1}}^n\}}{\Pr\{Y_{\boldsymbol{i}}^n|U_{\boldsymbol{0}}, \cdots, U_{\boldsymbol{i-1}}^n\}} \leq \mathsf{R}_{\boldsymbol{i}}. \right.$$

$$\left. \vee \qquad \vee_{\boldsymbol{i}} \frac{1}{n} \log \frac{\Pr\{Y_{\boldsymbol{i}}^n, U_{\boldsymbol{i}}^n\}}{\Pr\{Y_{\boldsymbol{i}}^n\}} \leq \sum_{\boldsymbol{j}\leq\boldsymbol{i}} \mathsf{R}_{\boldsymbol{j}} \right\} \leq \epsilon. \quad (2)$$

This set is closed.

Proposition 2: The set of ϵ-achievable rates $(\mathsf{R}_{\boldsymbol{0}}, \mathsf{R}_{\boldsymbol{1}}, \cdots, \mathsf{R}_{\boldsymbol{k-1}})$ over a broadcast channel $\boldsymbol{W}$ with degraded message set is the closure of

$$\bigcup_{\boldsymbol{U}_{\boldsymbol{0}},\cdots\boldsymbol{U}_{\boldsymbol{k-1}}\in\mathcal{S}} \mathsf{R}_{\boldsymbol{W}}(\epsilon, \boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}}).$$

As $\cap_{\epsilon>0}\mathsf{R}_{\boldsymbol{W}}(\epsilon, \boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}})$ equals $\mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}})$, this could serve as a definition of $\mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U}_{\boldsymbol{i}<\boldsymbol{k}})$. Proposition 2 has an intuitive interpretation: for a tuple of rates to be ϵ-achievable it is essential that the probability that the amount of transmitted information is less than the corresponding rate, is smaller than ϵ. The statement of the two technical Lemmas 1 and 2 that are used to prove Propositions 1 and 2 reveals that this intuition is quantitative.

A broadcast channel has the *strong converse property* if for any tuple of rates $(\mathsf{R}_{\boldsymbol{0}}, \mathsf{R}_{\boldsymbol{1}}, \cdots, \mathsf{R}_{\boldsymbol{k-1}})$, either $(\mathsf{R}_{\boldsymbol{0}}, \mathsf{R}_{\boldsymbol{1}}, \cdots, \mathsf{R}_{\boldsymbol{k-1}})$ is ϵ-achievable for all ϵ or any sequence of codes with rates $(\mathsf{R}_{\boldsymbol{0}}, \mathsf{R}_{\boldsymbol{1}}, \cdots, \mathsf{R}_{\boldsymbol{k-1}})$ has error probability converging toward 1.

Let $\mathsf{R}^*_{\boldsymbol{W}}(\boldsymbol{U_{i<k}})$ be the set of rate-tuples defined by

$$\liminf_{n\to\infty} \Pr\left\{ \vee_{\boldsymbol{i}} \frac{1}{n}\log\frac{\Pr\{Y_{\boldsymbol{i}}^n, U_{\boldsymbol{i}}^n \mid U_{\mathbf{0}}^n, \cdots U_{\boldsymbol{i-1}}^n\}}{\Pr\{Y_{\boldsymbol{i}}^n | U_{\mathbf{0}}, \cdots U_{\boldsymbol{i-1}}^n\}} \le \mathsf{R}_{\boldsymbol{i}} \right.$$
$$\left. \vee \vee_{\boldsymbol{i}} \frac{1}{n}\log\frac{\Pr\{Y_{\boldsymbol{i}}^n, U_{\boldsymbol{i}}^n\}}{\Pr\{Y_{\boldsymbol{i}}^n\}} \le \sum_{\boldsymbol{j}\le\boldsymbol{i}} \mathsf{R}_{\boldsymbol{j}} \right\} < 1. \quad (3)$$

The broadcast channels that have the strong converse properties may be characterized by the following condition.

Proposition 3: A broadcast channel with degraded message set has the strong converse property if and only if

$$\mathsf{closure}[\cup_{\boldsymbol{U_i}} \mathsf{R}^*_{\boldsymbol{W}}(\boldsymbol{U_{i<k}})] = \mathsf{closure}[\cup_{\boldsymbol{U_i}} \mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U_{i<k}})].$$

The proof of Propositions 1, 2, and 3 is an exercise in *information spectrum calculus*. It relies on an *information spectrum* version of Feinstein's lemma (for direct parts) and of its dual (for converse parts). To alleviate notations, we assume $\boldsymbol{k} = 2$.

Lemma 1: Let $\boldsymbol{U}$ and $\boldsymbol{X}$ be any input sequence, then for any positive integer $(\mathsf{R}_{\mathbf{0}}, \mathsf{R}_{\mathbf{1}})$, for any positive γ, there exists a broadcast $(n, \mathsf{R}_{\mathbf{0}}, \mathsf{R}_{\mathbf{1}}, \mathsf{e}_n)$-code satisfying

$$\mathsf{e}_n \le \Pr\left\{ \frac{1}{n}\log\frac{\sum_{x^n} \Pr_{X^n|U^n}(x^n|u^n) W_{\mathbf{0}}^n(y_{\mathbf{0}}^n|x^n)}{\Pr_{Y_{\mathbf{0}}^n}\{y_{\mathbf{0}}^n\}} \le \mathsf{R}_{\mathbf{0}} + \gamma \right.$$
$$\text{or } \frac{1}{n}\log\frac{W_{\mathbf{1}}^n(y_{\mathbf{1}}^n|x^n)}{\Pr_{Y_{\mathbf{1}}^n|U^n}\{y_{\mathbf{1}}^n|u^n\}} \le \mathsf{R}_{\mathbf{1}} + \gamma$$
$$\left. \text{or } \frac{1}{n}\log\frac{W_{\mathbf{1}}^n(y_{\mathbf{1}}^n|x^n)}{\Pr_{Y_{\mathbf{1}}^n}\{y_{\mathbf{1}}^n\}} \le \mathsf{R}_{\mathbf{0}} + \mathsf{R}_{\mathbf{1}} + \gamma \right\} + 3\,e^{-n\gamma}.$$

Lemma 2: For every n, any $(n, \mathsf{R}_{\mathbf{0}}, \mathsf{R}_{\mathbf{1}}, \mathsf{e}_n)$ broadcast code satisfies

$$\mathsf{e}_n \ge \Pr\left\{ \frac{1}{n}\log\frac{\sum_{x^n} \Pr_{X^n|U^n}(x^n|u^n) W_{\mathbf{0}}^n(y_{\mathbf{0}}^n|x^n)}{\Pr_{Y_{\mathbf{0}}^n}\{y_{\mathbf{0}}^n\}} \le \mathsf{R}_{\mathbf{0}} - \gamma \right.$$
$$\text{or } \frac{1}{n}\log\frac{W_{\mathbf{1}}^n(y_{\mathbf{1}}^n|x^n)}{\Pr_{Y_{\mathbf{1}}^n|U^n}\{y_{\mathbf{1}}^n|u^n\}} \le \mathsf{R}_{\mathbf{1}} - \gamma$$
$$\left. \text{or } \frac{1}{n}\log\frac{W_{\mathbf{1}}^n(y_{\mathbf{1}}^n|x^n)}{\Pr_{Y_{\mathbf{1}}^n|\boldsymbol{U}^n}\{y_{\mathbf{1}}^n\}} \le \mathsf{R}_{\mathbf{0}} + \mathsf{R}_{\mathbf{1}} - \gamma \right\} - 3\,e^{-n\gamma}$$

where $\boldsymbol{U}^n$ is uniformly distributed on a set of $\mathsf{M}_{\mathbf{0}} = 2^{n\mathrm{R}_{\mathbf{0}}}$ disjoint sets (clouds) of codewords and $\Pr_{X^n|U^n}$ places probability $2^{-n\mathrm{R}_{\mathbf{1}}}$ over each codeword in a cloud.

A sketch of the proof of those two lemmas is given in the Appendix.

Remarks on the Proof of Propositions 1 and 2: The proof of the direct part is an application of Lemma 1 and of the definition of the mutual information spectrum-inf, as in [9, pp. 2782–2784]. The proof of the converse part is an application of Lemma 2 and of the definition of the mutual information spectrum-inf, as in [9, p. 2783–2784]. □

Remarks on the Proof of Proposition 3: As

$$\mathsf{closure} \cup_{\boldsymbol{U_{i<k}}} \mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U_{i<k}}) \subseteq \mathsf{closure} \cup_{\boldsymbol{U_{i<k}}} \mathsf{R}^*_{\boldsymbol{W}}(\boldsymbol{U_{i<k}})$$

always holds, the proof amounts to showing that the strong converse property is equivalent to

$$\mathsf{closure} \cup_{\boldsymbol{U_{i<k}}} \mathsf{R}^*_{\boldsymbol{W}}(\boldsymbol{U_{i<k}}) \subseteq \mathsf{closure} \cup_{\boldsymbol{U_{i<k}}} \mathsf{R}_{W}(\boldsymbol{U_{i<k}}).$$

Proving that the strong converse property implies the inclusion is based on Lemma 1 and simple closure properties. Proving that the proper inclusion implies the strong converse property is based on Lemma 2 and again on simple closure properties. Indeed, if the last inclusion holds, the limiting probability that the amount of transmitted information is below a certain value is either 1 or 0. □

Thus even though the superposition codes idea was first motivated by memoryless broadcast channels, the information spectrum approach shows that it remains fully relevant in face of memory. Indeed, for the broadcast channel, the characterization of the capacity region described in Lemma 1 is closer to the memoryless characterization than in the case of the multiple-access channel (cf. [9]). And it is natural to ask whether something is lost when using the PET method. In the next section, we show that almost nothing is lost in the memoryless case, while in the last section, a simple example shows it may be false in face of memory.

III. Lossy Broadcast Channels

A. General Characterization

When specialized to lossy channels, the information spectrum approach provides a simple characterization of the capacity of single-user lossy channels. Let us first introduce another notation. The loss process $(\boldsymbol{Z})$ is defined as $Z^n(i) = 1$ if $Y^n(i)$ is a loss, 0 otherwise. The loss process completely defines the erasure channel. $\boldsymbol{Z}$ is assumed to be independent from channel inputs. Let $X^n(z^n)$ denote the subsequence of random variables $X^n(i)$, such that $z^n(i) = 0$.

Proposition 4: The capacity of the erasure channel defined by the loss process $\boldsymbol{Z}$ is

$$\tilde{\mathsf{C}} = 1 - \text{p-}\limsup \frac{1}{n}\sum_{i\le n} Z^n(i).$$

Remark: An erasure channel has the strong converse property if and only if $\frac{1}{n}\sum_{i\le n} Z^n(i)$ converges in probability toward a fixed value. For stationary ergodic channels, the capacity only depends on the stationary loss probability, but dependence may dramatically change the reliability function of the channel. Let

$$\underline{H}(\boldsymbol{X}) \triangleq \text{p-}\liminf_n -\frac{1}{n}\log\Pr\{X^n\}.$$

Proposition 1 can be specialized to lossy broadcast channels. $\mathsf{R}_{\boldsymbol{W}}(\boldsymbol{U_{i<k}})$ is now the set of tuple of rates $(\mathsf{R}_{\mathbf{0}} \cdots \mathsf{R}_{\boldsymbol{k-1}})$ satisfying

$$\begin{aligned}
0 \le \mathsf{R}_{\boldsymbol{i}} &\le \underline{I}(\boldsymbol{U_i}; \boldsymbol{X}(\boldsymbol{Z_i}) \,|\boldsymbol{U}_{\boldsymbol{i-1},\cdots\mathbf{0}}) \\
0 \le \mathsf{R}_{\boldsymbol{k-1}} &\le \underline{H}(\boldsymbol{X}(\boldsymbol{Z}_{\boldsymbol{k-1}}) \,|\boldsymbol{U}_{\boldsymbol{k-2},\cdots\mathbf{0}}) \\
0 \le \sum_{\boldsymbol{j}\le\boldsymbol{i}} \mathsf{R}_{\boldsymbol{j}} &\le \underline{I}(\boldsymbol{U_i}; \boldsymbol{X}(\boldsymbol{Z_i})), \qquad \text{for } \boldsymbol{i} < \boldsymbol{k-1} \\
0 \le \sum_{\boldsymbol{i}<\boldsymbol{k}} \mathsf{R}_{\boldsymbol{i}} &\le \underline{H}(\boldsymbol{X}(\boldsymbol{Z}_{\boldsymbol{k-1}})). \qquad (4)
\end{aligned}$$

Remark: If a broadcast erasure channel has exchangeable component loss processes, it is said to be exchangeable. Such broadcast channels provide simple examples of stationary broadcast channels failing to have to strong converse property.

B. Memoryless Broadcast Erasure Channels

1) Capacity Region: The structure of memoryless broadcast erasure channels is almost captured by Han's inequalities [10] or the following theorem due to Shearer [11]

Theorem 1: Let X^n be a collection of n random variables and Z^n be a collection of n Boolean random variables, such that for each i, $1 \le i \le n$, $\mathbb{E}\,Z_i = 1 - \tilde{\mathsf{C}}$.

$$\mathbb{E}\,H(X^n(Z^n)) \ge \tilde{\mathsf{C}}\,H(X^n). \quad (5)$$

This theorem was a key ingredient in the derivation of inequality (1), it enables here a direct derivation of the capacity region of the memoryless broadcast erasure channel without resorting to the single-letter characterization. It could as well have been used to derive the strong converse for memoryless lossy broadcast channels without resorting to the single-letter characterization, since it allows to use directly [5, Ch. III.3, Lemmas 4.2 and 4.3] in the proof of the strong converse for lossy broadcast channels [5, Theorem 4.3, Ch. III.4]. The following proposition was independently pointed out in [12] and in [13], its proof parallels the proof of [1, Inequality 1].

Proposition 5: A tuple of rates $(\mathsf{R}_{bf0}, \mathsf{R}_1, \cdots \mathsf{R}_{k-1})$ is achievable over a memoryless broadcast erasure channel with degraded message set if and only if

$$\sum_{i=1}^{k-1} \frac{\mathsf{R}_i}{\mathsf{C}_i} < 1. \tag{6}$$

Thus achievable rates may always be achieved by time-sharing.

Proof of Proposition 5: Since time-sharing is always feasible over stationary ergodic broadcast channels, only the converse part needs to be proved. We prove it when $k = 2$. Let $(\mathsf{R}_0, \mathsf{R}_1)$ be an achievable rate pair, then by [7, Proposition 1 and Theorem 8], there exists a family of input processes (U, X) such that for any $\gamma > 0$, there exists an n such that

$$\begin{aligned} \mathsf{R}_0 &\le \tfrac{1}{n}(H[X^n(Z_0^n)] - H[X^n(Z_0^n)|U^n]) + \gamma \\ \mathsf{R}_1 &\le \tfrac{1}{n} H[X^n(Z_1^n)|U^n] + \gamma. \end{aligned} \tag{7}$$

Now

$$H[X^n(Z_0^n)] \le \tilde{\mathsf{C}}_0 \sum_{j \le n} H[X^n(j)].$$

On the other hand, since losses on each channel are assumed to be independent, we may consider that the broadcast channel is stochastically or even physically degraded: any symbol lost on W_1 is lost over W_0, other symbols are lost over W_0 with probability $1 - (\tilde{\mathsf{C}}_0/\tilde{\mathsf{C}}_1)$. Then conditionally on Z_1^n, the sequence of random variables $X^n(Z_0^n)$ is a subsequence of $X^n(Z_1^n)$, a conditional use of Shearer's theorem implies

$$\mathbb{E}\,(H[X^n(Z_0^n)]|U^n,\, Z_1^n = z_1^n) \ge \frac{\tilde{\mathsf{C}}_0}{\tilde{\mathsf{C}}_1} H[X^n(z_1^n)|U^n].$$

Deconditioning with respect to z_1^n, we get

$$H([X^n(Z_0^n)]|U^n) \ge \frac{\tilde{\mathsf{C}}_0}{\tilde{\mathsf{C}}_1} H[X^n(Z_1^n)|U^n].$$

Substituting in (7) and adding the two inequalities, we get for arbitrary $\gamma > 0$

$$\frac{\mathsf{R}_0}{\tilde{\mathsf{C}}_0} + \frac{\mathsf{R}_1}{\tilde{\mathsf{C}}_1} \le \log\,|\mathcal{X}| + 2\gamma. \qquad \square$$

Remark: The essential ingredient in the proof is the conditional application of Shearer's theorem (or Han inequalities). It is still relevant to the analysis of exchangeable channels: The capacity region of exchangeable broadcast erasure channels is also described by (6), although those channels do not have the strong converse property.

2) Error Exponents: For good families of codes, error probabilities of memoryless broadcast channels are known to decline exponentially fast with block length n, provided the rate vector is in the capacity region of the channel [14], [15]. E is called an attainable error exponent if there exists a sequence of codes such that for any $\delta > 0$, for almost all block length n: $\max_i \mathsf{e}_i(n, \mathsf{R}_0, \cdots \mathsf{R}_{k-1}, W) < e^{-n(E-\delta)}$. The best error exponent is upper-bounded by the sphere-packing bound $E_{\rm sp}(., W)$ and lower-bounded by the random coding bound $E_{\rm rc}(., W)$, informational characterizations of those quantities are known [14], [15]. We are not aware of the existence of a general algorithm capable of computing the value of the exponents for broadcast channels but, fortunately, the broadcast erasure channel is simple enough so that those two quantities can be determined. A refined assessment of the PET approach consists in comparing the random coding exponent for PET codes with the sphere-packing exponent for memoryless broadcast erasure channels. Let us first recall the form of the sphere-packing and random-coding exponent for the single-user erasure channel. Gallager's approach [16] provides a closed form for single-user erasure channel exponents. Let

$$h(x,\, y) \triangleq x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}$$

denote the relative entropy between two Bernoulli random variables with parameters x and y, then

$$E_{sp}(\mathsf{R},\, W) = h(\tilde{\mathsf{R}},\, \tilde{\mathsf{C}}).$$

Let now the critical rate be

$$\tilde{\mathsf{R}}_{\rm cr} \triangleq \tilde{\mathsf{C}}/(\tilde{\mathsf{C}} + |\mathcal{X}|(1 - \tilde{\mathsf{C}}))$$

then the random coding exponent satisfies

$$\begin{aligned} E_{\rm rc}(\tilde{\mathsf{R}},\, W) &= E_{\rm sp}(\tilde{\mathsf{R}},\, W), \qquad \text{if } \tilde{\mathsf{R}}_{\rm cr} \le \tilde{\mathsf{R}} < \tilde{\mathsf{C}} && (8) \\ &= E_{\rm sp}(\tilde{\mathsf{R}}_{\rm cr},\, W) + \tilde{\mathsf{R}}_{\rm cr} - \mathsf{R}, \qquad \text{if } \tilde{\mathsf{R}} \le \tilde{\mathsf{R}}_{\rm cr}. && (9) \end{aligned}$$

Remarks: For erasure channels, the random coding exponent only depends on $\tilde{\mathsf{R}}$, $\tilde{\mathsf{C}}$, and $|\mathcal{X}|$. It will be denoted $E_{\rm rc}(\tilde{\mathsf{R}},\, \tilde{\mathsf{C}},\, \mathcal{X})$. For large alphabets, the sphere-packing exponent and the random-coding exponents coincide over a wide range of rates. Inspection of the derivation of the coding exponents of erasure channels (cf. [5]) shows that if we consider k erasure channels with capacities $\mathsf{C}_0 < \cdots \mathsf{C}_{k-1}$, it is always possible to find a family of codes that realizes simultaneously the k random coding exponents for the k channels.

The multiplexing strategy described here encompasses the partition refinement idea described in [1]. It is parameterized by the block length n, the bit length l of a packet ($|\mathcal{X}| = 2^l$), the number of priority levels k ($k \le l$), the rates R_i at which the different priority levels should be encoded for $0 \le i < k$, and the capacities of the component channels (W_i): (C_i). We assume that

$$\sum_{i<k} (\mathsf{R}_i/\mathsf{C}_i) < 1.$$

In a PET code, for all $i < k$, l_i bits from each packet will be dedicated to encoding of message m_i at normalized rate $\tilde{\mathsf{R}}_i'$. We will have to satisfy the constraints $\sum_i l_i = l$ and $l_i \tilde{\mathsf{R}}_i' = l\tilde{\mathsf{R}}_i$.

The total bit length of a codeword will be nl, and nl_i bits will be dedicated to the encoding of the ith priority level. The PET code designer still has to trade l_i's and $\tilde{\mathsf{R}}_i'$'s. And one may wonder whether this is the best way to minimize the error probabilities experienced by the different receivers, i.e., whether spreading information from the ith priority level over the whole packet would not improve the error exponents. The following propositions constitute a partial answer to those questions. In the sequel, Λ denotes the set of vectors from $\mathbb{R}^k$ such that $\sum_{i=0}^{k} \lambda_i = 1$ and $\lambda_i \ge 0$.

Proposition 6: For a broadcast erasure channel W, with capacities $(\mathsf{C}_i)_{i<k}$, and degraded message set, the sphere packing exponent is at most

$$\inf_{\lambda \in \Lambda} \max_i h\left(\frac{\tilde{\mathsf{R}}_i}{\lambda_i},\, \tilde{\mathsf{C}}_i\right).$$

Proposition 7: For a broadcast erasure channel W, with degraded message set, the random coding exponent achieved by juxtaposing and interleaving single-user erasure codes is

$$\sup_{\lambda\in\Lambda}\min_{\boldsymbol{i}} E_{\rm rc}\left(\frac{\tilde{\mathsf{R}}_{\boldsymbol{i}}}{\lambda_{\boldsymbol{i}}},\,\tilde{\mathsf{C}}_{\boldsymbol{i}},\,|\mathcal{X}|^{\lambda_{\boldsymbol{i}}}\right).$$

Remarks: The supremum in both propositions is attained for a tuple λ where all terms in the minimization (or maximization) are equal.

Notice that if all $\tilde{\mathsf{R}}_{\boldsymbol{i}}/\lambda_{\boldsymbol{i}}$ are larger than the critical rate of the erasure channel with normalized capacity $\tilde{\mathsf{C}}_{\boldsymbol{i}}$ and alphabet size $|\mathcal{X}|^{\lambda_{\boldsymbol{i}}}$, the upper bound stated in Proposition 7 is achieved. The combination of the two propositions seriously backs the following separation principle: First design good single-erasure codes with the required rates and alphabets, second, multiplex them in the simplest way using juxtaposition and interleaving.

Proof of Proposition 6: The arguments use the change of channel trick and a good guess of the twisted broadcast channel (cf. [5, Ch. II.5, p. 167] or [14]). We will consider the case of two sources that are to be transmitted at rates $\mathsf{R}_{\mathbf{0}}$ and $\mathsf{R}_{\mathbf{1}}$ over a broadcast erasure channel $(W_{\mathbf{0}},\,W_{\mathbf{1}})$. $(\mathsf{R}_0,\,\mathsf{R}_1)$ is assumed to be achievable. Now consider a twisted memoryless broadcast erasure channel $(W_0^{n\prime},\,W_1^{n\prime})$ with pair of capacities

$$(\tilde{\mathsf{C}}_0^{\prime},\,\tilde{\mathsf{C}}_1^{\prime}) = (1-\delta)(\tilde{\mathsf{R}}_0/\lambda,\,\tilde{\mathsf{R}}_1/(1-\lambda))$$

with $\lambda\in(0,\,1)$ and $\delta>0$. The pair $(\tilde{\mathsf{R}}_0,\,\tilde{\mathsf{R}}_1)$ is not achievable over this twisted channel. Hence as the strong converse property holds for memoryless broadcast channels [5], for any sequence of block broadcast codes, for n large enough $\max_{i\in\{0,\,1\}}\mathsf{e}_i(f^n,\,\phi_i^n) > 1-\delta/2$. This means that for some pair of messages $(m_0,\,m_1)$, the events $S_0\stackrel{\Delta}{=}\{y;\phi_0^n(y)\neq m_0\}$ and $S_1\stackrel{\Delta}{=}\{y;\phi_1^n(y)\neq\langle m_0,\,m_1\rangle\}$ satisfy

$$\max_{\boldsymbol{i}} W_{\boldsymbol{i}}^{n\prime}(S_{\boldsymbol{i}}|f^n(m_0,\,m_1)) > 1-\delta/2.$$

Notice that on a fixed input, the distributions of outputs through channels $W_{\boldsymbol{i}}$ and $W_{\boldsymbol{i}}^{\prime}$ have relative entropies $n\times h(\tilde{\mathsf{C}}_{\boldsymbol{i}}^{\prime},\,\tilde{\mathsf{C}}_{\boldsymbol{i}})$. Thus we have [5, Ch. II.5]

$$h[W_{\boldsymbol{i}}^{n\prime}(S_{\boldsymbol{i}}|f^n(m_0,\,m_1)),\,W_{\boldsymbol{i}}^{n}(S_{\boldsymbol{i}}|f^n(m_0,\,m_1))] \le n.h(\tilde{\mathsf{C}}_{\boldsymbol{i}}^{\prime},\,\tilde{\mathsf{C}}_{\boldsymbol{i}})$$

which implies:

$$\max_{\boldsymbol{i}} W_{\boldsymbol{i}}^{n}(S_{\boldsymbol{i}}|f^n(m_0,\,m_1)) \ge \min_{\boldsymbol{i}}\exp\left(-\frac{n\,h(\tilde{\mathsf{C}}_{\boldsymbol{i}}^{\prime},\tilde{\mathsf{C}}_{\boldsymbol{i}})-h(\delta/2)}{1-\delta/2}\right).$$

As δ may be arbitrarily small, by continuity of h, this in turn implies that the sphere-packing exponent for the memoryless broadcast erasure channel is smaller than $\max_{\boldsymbol{i}}(h(\tilde{\mathsf{R}}_{\boldsymbol{i}}/\lambda_{\boldsymbol{i}},\,\tilde{\mathsf{C}}_{\boldsymbol{i}}))$ where $\lambda_{\mathbf{0}}=\lambda$ and $\lambda_{\mathbf{1}}=1-\lambda$. □

Proof of Proposition 7: Let $\lambda_{\boldsymbol{j}}$ denote $l_{\boldsymbol{j}}/l$ for $\boldsymbol{j}\in\{0,\,1\}$. Let multiplexed single-user codes realize simultaneously the random-coding exponents for the relevant erasure channels, for all i

$$\begin{aligned}\mathsf{e}_{\boldsymbol{i}} &\le \sum_{\boldsymbol{j}\le\boldsymbol{i}} e^{-n\,E_{\rm rc}(\tilde{\mathsf{R}}_{\boldsymbol{j}}^{\prime},\,\tilde{\mathsf{C}}_{\boldsymbol{i}},\,|\mathcal{X}|^{\lambda_{\boldsymbol{j}}})}\\ &\le (\boldsymbol{i}+1)\max_{\boldsymbol{j}\le\boldsymbol{i}} e^{-n\,E_{\rm rc}(\tilde{\mathsf{R}}_{\boldsymbol{j}}^{\prime},\,\tilde{\mathsf{C}}_{\boldsymbol{i}},\,|\mathcal{X}|^{\lambda_{\boldsymbol{j}}})}. \qquad (10)\end{aligned}$$

Thus by monotonicity of the random-coding exponent with respect to capacity

$$\begin{aligned}\max_{\boldsymbol{i}<\boldsymbol{k}}\mathsf{e}_{\boldsymbol{i}} &\le \boldsymbol{k}\max_{\boldsymbol{j}\le\boldsymbol{i}<\boldsymbol{k}} e^{-n\,E_{\rm rc}(\tilde{\mathsf{R}}_{\boldsymbol{j}}^{\prime},\,\tilde{\mathsf{C}}_{\boldsymbol{i}},\,|\mathcal{X}|^{\lambda_{\boldsymbol{j}}})}\\ &\le \boldsymbol{k}\max_{\boldsymbol{i}<\boldsymbol{k}} e^{-n\,E_{\rm rc}(\tilde{\mathsf{R}}_{\boldsymbol{i}}^{\prime},\,\tilde{\mathsf{C}}_{\boldsymbol{i}},\,|\mathcal{X}|^{\lambda_{\boldsymbol{i}}})}. \qquad (11)\end{aligned}$$

We get

$$E_{\rm rc}(\mathsf{R}_0\cdots\mathsf{R}_{\boldsymbol{k}-1},\,W) = \sup_{\lambda\in\Lambda}\min_{\boldsymbol{i}} E_{\rm rc}(\tilde{\mathsf{R}}_{\boldsymbol{i}}/\lambda_{\boldsymbol{i}},\,\tilde{\mathsf{C}}_{\boldsymbol{i}},\,|\mathcal{X}|^{\lambda_{\boldsymbol{i}}}).$$

IV. General Broadcast Erasure Channels

The engineering solution, i.e., time sharing and interleaving, provides a good method to design broadcast codes over memoryless broadcast erasure channels. It would be nice if it were also the case over erasure channels with memory. Unfortunately, it is possible to design asymptotically memoryless broadcast erasure channels for which time sharing does not exhaust the capacity region. Let $(W_0,\,W_1)$ be defined in the following way. For every sequence of six symbols $x_{6i}\cdots x_{6i+5}$ for $i=0\cdots\infty$ either the first three symbols $x_{6i}\cdots x_{6i+2}$, or last three symbols $x_{6i+3}\cdots x_{6i+5}$ are erased by the powerful component channel W_1, and four symbols, either $x_{6i}\cdots x_{6i+3},\ x_{6i},\,x_{6i+1},\,x_{6i+4},\,x_{6i+5}$, or $x_{6i+2}\cdots x_{6i+5}$ are erased. Blocks $x_{6i}\cdots x_{6i+5}$ and $x_{6j}\cdots x_{6j+5}$ are handled independently by both component channels. This broadcast channel is not stationary, although it is block-stationary and loss probability is shift-invariant over both component channels. It does not have long memory: Channel memory vanishes after six steps. Component channels have capacities $\tilde{\mathsf{C}}_0=1/3$ and $\tilde{\mathsf{C}}_1=1/2$, nevertheless $(\tilde{\mathsf{R}}_0,\,\tilde{\mathsf{R}}_1)=(1/6,\,1/3)$ is an achievable rate pair: The channel alphabet is assumed to be provided with a group structure, one symbol of common information m_0 and two symbols of private information $(m_1,\,m_2)$ are encoded in six symbols $x_0\cdots x_5$, in the following way: $x_0=x_5=m_1$, $x_2=m_2$, $x_1=x_4=m_0\oplus m_1$, $x_3=m_0\oplus m_2$. The common information can always be recovered from the $\boldsymbol{W_0}$ output, the private information can always be recovered from $\boldsymbol{W_1}$ output. And $\dfrac{\tilde{\mathsf{R}}_0}{\tilde{\mathsf{C}}_0}+\dfrac{\tilde{\mathsf{R}}_1}{\tilde{\mathsf{C}}_1}=7/6>1$. Notice that because losses are not exchangeable, for the input sequence defined by this encoding, the conditional entropy rates per variable $H[\boldsymbol{X}(\boldsymbol{Z_0})|\boldsymbol{U}]$ is $1/2$ while $H[\boldsymbol{X}(\boldsymbol{Z_1})|\boldsymbol{U}]$ is $2/3$.

Hence the class of broadcast erasure channels over which time sharing does not exhaust the capacity region, has still to be determined.

Appendix

Sketch of Proof of Lemma 1: Let $(\boldsymbol{U},\,\boldsymbol{X})$ be a family of input processes. $M_0=2^{\lfloor n\mathbf{R_0}\rfloor}$ elements $u_1,\,\cdots,\,u_{M_0}$ are generated according to U^n, then for each u_i, $M_1=2^{\lfloor n\mathbf{R_1}\rfloor}$ codewords $x_{i,\,j}$ are drawn according to $\Pr_{X^n|U^n}$. The message $(i,\,j)$ is encoded by $x_{i,\,j}$. To describe decoding let us define as in [9] the following information-spectrum typical sets.

$$T_1^n=\left\{(\mathrm{u},\,\mathrm{y_0}):\frac{1}{n}\log\frac{\sum_{x^n}\Pr_{X^n|U^n}(x^n|\mathrm{u})W_{\mathbf{0}}^n(\mathrm{y_0}|x^n)}{\Pr_{Y_{\mathbf{0}}^n}\{\mathrm{y_0}\}} > \mathsf{R}_{\mathbf{0}}+\gamma\right\}$$

$$T_2^n=\left\{(\mathrm{x},\,\mathrm{y_1}):\frac{1}{n}\log\frac{W_{\mathbf{1}}^n(\mathrm{y_1}|\mathrm{x})}{\Pr_{Y_{\mathbf{1}}^n}\{\mathrm{y_1}\}} > \mathsf{R}_{\mathbf{0}}+\mathsf{R}_{\mathbf{1}}+\gamma\right\}$$

$$T_3^n=\left\{(\mathrm{u},\,\mathrm{x},\,\mathrm{y_1}):\frac{1}{n}\log\frac{W_{\mathbf{1}}^n(\mathrm{y_1}|\mathrm{x})}{\Pr_{Y_{\mathbf{1}}^n|U^n}\{\mathrm{y_1}|\mathrm{u}\}} > \mathsf{R}_{\mathbf{1}}+\gamma\right\}.$$

Let F_i and $E_{i,\,j}$ denote the following sets:

$$F_i=\{(u_i,\,\mathrm{y_0})\in T_1^n\}$$
$$E_{i,\,j}=\{(u_i,\,\mathrm{x}_{i,\,j},\,\mathrm{y_1})\in T_2^n\cap T_3^n\}.$$

On receiving $\mathrm{y}_{\mathbf{0}}$ over $W_{\mathbf{0}}^n$ the decoder $\phi_{\mathbf{0}}$ reproduces i if and only if there exists a unique i such that $(u_i,\ \mathrm{y}_{\mathbf{0}}) \in F_i$. On receiving an output $\mathrm{y}_{\mathbf{1}}$ over $W_{\mathbf{1}}^n$ the decoder $\phi_{\mathbf{1}}$ reproduces $(i,\ j)$ if and only if there exists a unique $(i,\ j)$ such that $(\mathrm{u}_i,\ \mathrm{x}_{i,\,j},\ \mathrm{y}_{\mathbf{1}}) \in E_{i,\,j}$.

Averaging the decoding error probability for $W_{\mathbf{0}}^n$ over the random codes, using the exchangeability of messages, and the union bound, we get for the average error probability on channel $W_{\mathbf{0}}$

$$\mathsf{e}_{\mathbf{0}} \le \Pr\{u_{\mathbf{1}},\ x_{\mathbf{1},\,\mathbf{1}},\ y_{\mathbf{0}}\colon (u_{\mathbf{1}},\ y_{\mathbf{0}}) \notin T_1^n\} + M_0 \times \sum_{u_{\mathbf{1}},\ x_{\mathbf{1},\,\mathbf{1}},\ y_{\mathbf{0}}} \Pr\{u_{\mathbf{1}},\ y_{\mathbf{0}}\} \sum_{u'\colon (u',\ y_{\mathbf{0}}) \in T_1^n} \Pr\{u'\}. \quad (12)$$

But by definition of T_1^n for any $(u',\ y_{\mathbf{0}}) \in T_1^n$

$$\sum_{x'} \Pr\{x'|u'\}\, W_{\mathbf{0}}^n(y_{\mathbf{0}}|x') > M_0 \times e^{\gamma n} \times \Pr\{y_{\mathbf{0}}\}$$

hence for any $y_{\mathbf{0}}$

$$\sum_{u'\colon (u',\ y_{\mathbf{0}}) \in T_1^n} \Pr\{u'\} \le \frac{e^{-\gamma n}}{M_0}. \quad (13)$$

Plugging (13) into (12) gives

$$\mathsf{e}_{\mathbf{0}} \le \Pr\{T_1^{n,\,c}\} + e^{-\gamma n}. \quad (14)$$

A similar argument is developed for the decoding error probability over $W_{\mathbf{1}}^n$.

$$\begin{aligned}\mathsf{e}_{\mathbf{1}} \le{}& \Pr\{T_2^{n,\,c} \text{ or } T_3^{n,\,c}\} \\ &+ M_1 M_0 \sum_{u_{\mathbf{1}},\ x_{\mathbf{1},\,\mathbf{1}},\ y_{\mathbf{1}}} \Pr\{u_{\mathbf{1}},\ x_{\mathbf{1},\,\mathbf{1}},\ y_{\mathbf{1}}\} \\ &\times \sum_{u',\ x',\ (x',\ y_{\mathbf{1}}) \in T_2^n} \Pr\{u', x'\} \\ &+ M_1 \sum_{u_{\mathbf{1}},\ x_{\mathbf{1},\,\mathbf{1}},\ y_{\mathbf{1}}} \Pr\{u_{\mathbf{1}},\ x_{\mathbf{1},\,\mathbf{1}},\ y_{\mathbf{1}}\} \\ &\times \sum_{x',\ (u_{\mathbf{1}},\ x',\ y_{\mathbf{1}}) \in T_3^n} \Pr\{x'|u_{\mathbf{1}}\}. \end{aligned} \quad (15)$$

But by definition of T_2^n and T_3^n for any $y_{\mathbf{1}}$ for any $u',\ x'$, if

$$(x',\ y_{\mathbf{1}}) \in T_2^n : W_{\mathbf{1}}^n(y_{\mathbf{1}}|x') > e^{\gamma n}\, M_0\, M_1\, \Pr\{y_{\mathbf{1}}\}$$

and if

$$(u_{\mathbf{1}},\ x',\ y_{\mathbf{1}}) \in T_3^n, W_{\mathbf{1}}^n(y_{\mathbf{1}}|x') > e^{\gamma n}\, M_1\, \Pr\{y_{\mathbf{1}}|u_{\mathbf{1}}\}$$

hence

$$\sum_{u',\ x'(x',\ y_{\mathbf{1}}) \in T_2^n} \Pr\{x',\ u'\} \le \frac{e^{-\gamma n}}{M_0\, M_1} \quad (16)$$

and

$$\sum_{x'(u_{\mathbf{1}},\ x',\ y_{\mathbf{0}}) \in T_3^n} \Pr\{x'|u_{\mathbf{1}}\} \le \frac{e^{-\gamma n}}{M_1}. \quad (17)$$

Plugging (16) and (17) into (15)

$$\mathsf{e}_{\mathbf{1}} \le \Pr\{T_2^{n,\,c} \text{ or } T_3^{n,\,c}\} + 2.e^{-\gamma n}. \quad (18)$$

This terminates the proof of Lemma 1. □

Proof of Lemma 2: Let the broadcast code and input process be defined as in the statement of Lemma 2. Then let

$$\begin{aligned} L_n^1 &\triangleq \{(u,\ y_{\mathbf{0}})\colon \Pr\{y_{\mathbf{0}},\ u\} \le e^{-\gamma n}\ \Pr\{y_{\mathbf{0}}\}\} \\ L_n^2 &\triangleq \{(x,\ y_{\mathbf{1}})\colon \Pr\{x,\ y_{\mathbf{1}}\} \le e^{-\gamma n}\ \Pr\{y_{\mathbf{1}}\}\} \\ L_n^3 &\triangleq \{(u,\ x,\ y_{\mathbf{1}})\colon \Pr\{x,\ y_{\mathbf{1}}|u\} \le e^{-\gamma n}\ \Pr\{y_{\mathbf{1}}|u\}\}. \end{aligned}$$

Notice that $L_n^1 \cup L_n^2 \cup L_n^3$ is the event described in the lemma's statement. By the union bound

$$\begin{aligned}\Pr\{L_n^1 \cup L_n^2 \cup L_n^3\} \le{}& \mathsf{e}_n \\ &+ \sum_{y_{\mathbf{0}}} \sum_{i=1}^{M_0} \Pr\{(u_i,\ y_{\mathbf{0}}) \in L_n^1 \text{ and } \phi_0^n(y_{\mathbf{0}}) = i\} \\ &+ \sum_{y_{\mathbf{1}}} \sum_{i=1,\,j=1}^{M_0,\,M_1} \Pr\{(u_i,\ x_j,\ y_{\mathbf{1}}) \in L_n^2 \text{ and } \phi_1^n(y_{\mathbf{1}}) = (i,\ j)\} \\ &+ \sum_{y_{\mathbf{1}}} \sum_{i=1,\,j=1}^{M_0,\,M_1} \Pr\{(u_i,\ x_j,\ y_{\mathbf{1}}) \in L_n^3 \text{ and } \phi_1^n(y_{\mathbf{1}}) = (i,\ j)\}. \end{aligned}$$

Now, using first the definition of L_n^1, then the fact that each $y_{\mathbf{0}}$ belongs to at most one decoding set

$$\begin{aligned}&\sum_{y_{\mathbf{0}}} \sum_{i=1}^{M_0} \Pr\{(u_i,\ y_{\mathbf{0}}) \in L_n^1 \text{ and } \phi_0^n(y_{\mathbf{0}}) = i\} \\ &\quad= \sum_{i=1}^{M_0} \sum_{y_{\mathbf{0}}\colon\, \phi_{\mathbf{0}}^n(y_{\mathbf{0}})=i} \Pr\{(u_i,\ y_{\mathbf{0}}) \in L_n^1\} \\ &\quad\le \sum_{i=1}^{M_0} \sum_{y_{\mathbf{0}}\colon\, \phi_{\mathbf{0}}^n(y_{\mathbf{0}})=i} \Pr\{y_{\mathbf{0}}\}\, e^{-n\gamma} \\ &\quad\le \sum_{y_{\mathbf{0}}} \Pr\{y_{\mathbf{0}}\}\, e^{-n\gamma}. \end{aligned}$$

Similar arguments are applied to the second and third summand, completing the proof of the lemma. □

Acknowledgment

The authors wish to thank an anonymous reviewer for suggesting a mistake in a previous version of the correspondence, and another reviewer for pointing out [13].

References

[1] A. Albanese, J. Bloemer, J. Edmonds, and M. Luby, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737–1744, Nov. 1996.

[2] E. Biersack, "Performance evaluation of forward error correction in an ATM environment," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 631–640, 1993.

[3] V. Paxson, "Measurements and analysis of end-to-end internet traffic," Ph.D. dissertation, Univ. Calif. Berkeley, Feb. 1997.

[4] J. D. Gibson, T. Berger, and D. Lindbergh, *Digital Compression for Multimedia: Principles and Standards*. San Francisco, CA: Morgan Kaufmann, 1998.

[5] I. Csiszár and J. Körner, "Information theory: coding theorems for discrete memoryless channels," in *Probability and Mathematical Statistics*. New York: Academic, 1981.

[6] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, 1972.

[7] S. Verdù and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, 1994.

[8] J. Körner and K. Marton, "General broadcast channel with degraded message set," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60–64, 1977.

[9] T. S. Han, "An information spectrum approach to capacity theorems for the general multiple-access channel," *IEEE. Trans. Inform. Theory*, vol. 44, pp. 2773–2795, Nov. 1998.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer, "Some intersection theorems for ordered sets and graphs," *J. Comb. Theory, Ser. A*, vol. 43, pp. 23–37, 1986.

[12] S. Boucheron and K. Salamatian, "Codage à protections inégales et diffusion," in *Actes du 16ème GRETSI*, 1997, pp. 547–550.

[13] R. Urbanke and A. D. Wyner, "Packetizing for the erasure broadcast channel with an Internet application," in *Int. Conf. Combinatorics, Information Theory and Statistics*, 1997, p. 93.
[14] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 670–679, Nov. 1980.
[15] G. Poltyrev, "Random coding bounds for some broadcast channels," *Probl. Pered. Inform.*, vol. 19, no. 1, pp. 9–20, 1983.
[16] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1967.

Sequential Decoding for the Exponential Server Timing Channel

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

***Abstract*—We show the existence of a good tree code with a sequential decoder for the exponential server timing channel. The expected number of computations before moving one step ahead is upper-bounded by a finite number. The rate of information transfer for this code is $\mu/(2e)$ nats per second, i.e., one half of the capacity. The cutoff rate for the exponential server queue is therefore at least $\mu/(2e)$ nats per second.**

***Index Terms*—Computation, decoding metric, sequential decoder, single-server queue, timing channel, tree codes.**

I. Introduction

Sequential decoding of convolutional codes and tree codes ([1]–[5], 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 example, [6], [7]) extend the sequential decoding technique to a class of channels with memory, namely, finite-state channels. In this work we show that the sequential decoding technique can be used on timing channels (for example, [8] and [9]). 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 [8], [9], at rates below *half* the capacity, but with manageable decoding complexity. In [8]–[10], 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 work 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 [3] 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 [4], [5], 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 [2] 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 work are the choice of this metric, and a simple analytical artifice (used earlier in [8] in a different context) that shows how the elegant technique in [2] can be modified to prove the existence of a good tree code for this system with memory.

Section II introduces the problem in the appropriate notation and states the result. Section III contains the proof. We conclude with a brief discussion in Section IV.

II. 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 (nonzero) 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

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

where

$$w_i = \max\left\{0, \sum_{j=1}^{i} x_j - \sum_{j=0}^{i-1} y_j\right\} \tag{2}$$

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

We now describe the tree code. We follow the notation in [2] 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 $\boldsymbol{u} = (u_1, u_2, \cdots)$ is encoded by a tree code $\boldsymbol{g}$. The tree $\boldsymbol{g}$ is such that M edges leave each node of the code tree. Each edge is labeled by an N-tuple of nonnegative real numbers. The root node is labeled by the number of

Manuscript received December 21, 1998; revised Septepmber 23, 1999. This work was supported in part by the National Science Foundation under Grant NCR-9523805 002.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

Communicated by T. E. Fuja, Associate Editor At Large.
Publisher Item Identifier S 0018-9448(00)01690-4.