An Information Theory for Erasure Channels

Ramin Khalili, Kavé Salamatian
IEEE Member
LIP6-CNRS, Université Pierre et Marie Curie, France
Khalili@lss.supelec.fr, Salamatian@rp.lip6.fr

Abstract

We propose an information theory analysis of the wireless erasure channels. We derive the classical cut-set bound for the single-sender single-relay erasure channel and we show that this bound is not achievable using a deterministic coding scheme. We propose a new coding converse which is tight enough to be reached by an MDS based coding scheme. The analysis is based on the nature of the erasure channel that enables us to find the optimal relaying scheme over the channel. The developed achievability scheme is a linear coding scheme which provides a practical and simple way of doing Slepian-Wolf coding [1] in the context of erasure channels.

I. INTRODUCTION

Wireless networks consist of senders, receivers, and intermediate nodes that roughly cooperate to achieve a communication. An important problem in the context of wireless networks consists of finding the best possible cooperation scheme between nodes on the network that maximizes the transferred information rate. Information theory aims to find the set of transfer rate that is ultimately achievable for any given scenario as well as the coding/cooperation schemes achieving or approaching these bounds .

In this paper we assume a class of wireless channels, called erasure channels, where links connecting nodes are erasure channels. In this setting symbols sent over the channel are received without error or erased and replaced by erasure symbol $\{\mathtt{e}\}$. These channels are very relevant for modeling communication networks as felt by applications sitting at upper layers. A receiver might receive symbols sent by any sender with a probability that depends on the topological settings (the distance between senders and receivers), emission power, resulting level of interference (*e.g.* collision), coding characteristics and, *etc*. An underlying assumption of the erasure channel is that concurrent senders send over separated channels and no interference out of collisions can occurs between transmissions. Using separate physical channels (multiple interface card at nodes) or implementing a scheduling mechanism (centralized or decentralized) analogous to CSMA/CA achieves this separation. In this last setting residual collisions result in erasures.

Classically most of multi-user information theory research has dealt with error channels have left out erasure channels. However, this class of channels is very important from different perspectives. Actually used layered communication network assumes an erasure channel as the communication medium of layers sitting over the physical layer.Moreover, the relative analytic simplicity of erasure channels makes them attractive as an initial playground to extract essential characteristics of a communication problem and apply the insight gained to tackle with the more complex error channels.

In this paper, we present a rapid review of known results in the context of erasure channels. We will also prove a strong coding converse for the single-sender single-relay erasure channel showing that the cut-set bound is not attainable in general by any deterministic coding scheme (this result has been firstly presented in [2]). This bound shows the distinction between the degraded situation where the cut-set bound is attainable and some specific case of the non-degraded situation where one cannot reach the cut-set bound. The main blocking situation, in which the cut-set bound is not attainable, is when the capacity of the sender to the relay channel is larger than the capacity of the relay to the destination channel and the relay cannot decode the information sent by the source. In this situation, the relay have to devote a non negligible and non vanishing amount of information to signal to the destination the coding scheme used at the relay. This restricts the achievable rates to a region smaller than the region predicted by the cut-set bound. This obstacle has been circumvented in [3] and [4] by adding side information at the receiver. The paper [3] assumes a perfect knowledge of erasure patterns over all erasure channels at receivers and [4] uses an in band signaling mechanism to transfer the coding scheme used at the relay. The results presented here differentiate from the cited works by no assumption about any type of side information.

We obtain the converse bound thanks to the nature of erasure channels that relates directly the amount of information transferred over an erasure channel to the proportion of packets received at the receiver. This results in the optimality of the Maximal Distance Separable codes for erasure channels that is the main ingredient of the converse bound. We further propose a coding scheme that achieves this bound. Up to our knowledge, the proposed converse bound and the resulting capacity regions are the first valid for general non-degraded channels.

In following section we will develop on the capacity of a point to point erasure channel. We study the broadcast erasure channel with degraded message set in section III. In the section IV we the single-sender single-relay erasure channel which is the simplest scenario of node collaboration is analysed.

In the forthcoming we will use bold symbols represent vectors; capitals denotes random variables and lower cases are realizations; the superscripts for node index and subscripts for time, *e.g.* $\mathbf{X}^k$ is a vector containing random variables $X_1, \ldots, X_k$.

II. Point to Point erasure channel

A point to point erasure channel is defined with an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y} = \mathcal{X} \cup \{\mathtt{e}\}$ and a conditional input/output channel distribution $\mathcal{E}_k$. There are only two possible outcomes at the output of an erasure channel: the sent symbol or an erasure $\mathtt{e}$. One can therefore characterize the erasure channel by the probability of observing an erasure at the k^{th} transmission, *i.e.* $\mathcal{E}_k(\mathtt{e}|\mathbf{X}^k)$ the probability that an erasure occurs at the output of the channel given that the sequence of symbols $\mathbf{X}^k = \{x_1, \ldots, x_k\}$. When $\mathcal{E}_k$ do not depend on $\mathbf{X}^k$, one can characterize the erasure channel by the erasure process $\{\mathbf{Z}\}$. This process is defined as $z_k = 1$, if an erasure occurs at the output of the channel at time k; 0 otherwise. Therefore, one can characterize an erasure channel by the sequence of probability measures $\mathcal{E}_k = \mathbb{P}\mathrm{rob}\big\{\mathbf{Z}^k\big\}$.

A. Capacity Results

We denote by $\mathbf{X}^n(\mathbf{Z}^n)$ a subsequence of $\mathbf{X_n}$ containing all indices k such that $Z_k = 0$. The following theorem is at the basis of the analysis of erasure channels.

Theorem 1—Shearer Theorem [5]: Let $\mathbf{X}^n$ be a collection of n random variables and $\mathbf{Z}^n$ be a collection of n boolean random variable, such that for each k, $1 \leq k \leq n$, $\mathbb{E}\big\{Z_k\big\} = p$. Then

$$\mathbb{E}\big\{H\left(\mathbf{X}^n(\mathbf{Z}^n)\right)\big\} = (1-p)H(\mathbf{X}^n),$$

where average is taken over the statistics of the random variables Z_k, $1 \leq k \leq n$.

This theorem can be easily extended to the conditional case.

Corollary 1—Conditional Shearer Theorem: Suppose that U is a random variable and $\mathbb{P}\mathrm{rob}\big\{\mathbf{X}^n|U\big\}$ a conditional probability. We have :

$$\mathbb{E}\{H\left(\mathbf{X}^n(\mathbf{Z}^n) \mid U\right)\} = (1-p)H(\mathbf{X}^n|U),$$

where the average is taken over the statistics of the conditional random variable $\mathbf{Z}^n|U$. □

The capacity of the memoryless channel have been proved initially by Elias in [**?**] where he prove that a random code can achieve the capacity. Here we present an extension of the capacity to the stationary ergodic erasure channel as a direct consequence of the Shearer theorem and its conditional version :

Theorem 2—Capacity of point to point erasure channel: The capacity of a stationary and ergodic point to point erasure channel is given by:

$$C = (1 - \mathbb{E}\big\{\mathbf{Z}\big\})H(X) \ \ \text{bits/trans},$$

where $\{\mathbf{Z}\}$ is the erasure process of the channel.

Proof: The received sequence at the output of a point to point erasure channel, $\mathbf{y}n$, contains the same information that $\mathbf{X}^n(\mathbf{Z}^n)$. Moreover, $I(\mathbf{X}^n;\mathbf{y}^n) = \mathbb{E}\{H(\mathbf{X}^n(\mathbf{Z}^n)) - H(\mathbf{X}^n(\mathbf{Z}^n)|\mathbf{X}^n)\}$. By using conditional Shearer theorem with $U = \mathbf{X}^n$ and defining $p = \mathbb{E}\big\{Z\big\}$ we have :

$$I(\mathbf{X}^n;\mathbf{y}^n) = (1-p)H(\mathbf{X}^n).$$

The mutual information is maximized when the sent symbols $\mathbf{X}^n$ are chosen independently, *i.e.* $H(\mathbf{X}^n) = nH(X)$. ■

This capacity result could be even extended to channels where the loss proportion accepts a strong law of large numbers.

Corollary 2: The capacity of the erasure channel with a loss process $\{Z\}$ is :

$$C = 1 - \limsup_{n \to \infty} \frac{1}{n} \sum_{i \leq n} Z_i.$$

□

We will show in next section a converse if $\frac{1}{n} \sum_{i \leq n} Z_i$ converges in probability to a fixed value. Moreover if the channel accept a large deviation principle for the empirical proportion of losses, we have a strong converse.

B. *Converse bounds*

One important characteristic of the erasure channel is the availability of tight converse bounds that are valid under a large set of hypothesis. Let's $(\mathcal{C}, \Phi, \Psi)$ be an erasure correction scheme with encoding and decoding function Φ and Ψ. The encoding function Φ maps the message set $\mathcal{M} = \{1, \ldots, |\mathcal{X}|^{n\mathbf{R}}\}$ to a codeword set $\mathcal{C} \subseteq \mathcal{X}^n$. To simplify the presentation we will assume without loss of generality than that Φ implements a one to one mapping, meaning that decoding at receiver the correct sent sequence $\mathbf{X}^m = \Phi(m)$ for a sent message m is enough to retrieving uniquely the sent message. Conversely if ambiguity remains at receiver about the sent sequence, the same level of ambiguity remains about the sent message. Under this hypothesis the decoding function is defined as $\Psi : \mathcal{Y}^n \to \mathcal{C}$.

By the particular structure of erasure channel the decoder of erasure correction codes should follow a particular and generic structure. Let's define the set $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$ as the set of all codewords of code $\mathcal{C}$ that are in agreement with the symbols in the received sequence $\mathbf{y}^n$, *i.e.* the set contains all codewords that have the same symbols at the same position as the symbols received in $\mathbf{y}^n$. We will name this set the agrement set of $\mathbf{y}^n$ for code $\mathcal{C}$. Let's define the decoding function Ψ^* that assigns to any received sequence $\mathbf{y}^n$ a codeword in the agreement set of $\mathbf{y}^n$, $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$. If the agreement set contains a single codeword the sent message is decoded correctly without ambiguity. If the agreement set contains more than a single codeword ambiguity subsists about the sent message. As the loss process is independent of the message sent and as all messages are equally likely any particular decoded codeword assignment in $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$ will results in a decoding error probability $(1 - \frac{1}{|\mathcal{A}(\mathcal{C}, \mathbf{y}^n)|})$ when $\mathbf{y}^n$ is received. A list decoding version of this decoder can be obtained by returning the agreement set $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$ as the decoded list.

It is noteworthy that the above decoding function is sufficient to study as its optimal. This could be explained through the following lemma that shows that that any code not following the above decoding function can be supersed by a code using the decoding function Ψ^* that will have better erasure correction performances.

Lemma 1: The decoding function Ψ^* is the best decoding function for a code $\mathcal{C}$ in term of erasure correction performances.

Proof: Let's assume a coding scheme $(\mathcal{C}, \Phi, \Psi)$ where the decoded codeword $\hat{\mathbf{X}}^n$ corresponds to a received sequence $\mathbf{y}^n$, *i.e.* $\Psi(\mathbf{y}^n) = \hat{\mathbf{x}}^n$ is not in agreement with the received sequence $\mathbf{y}^n$. By the structure of erasure channels there will be a decoding error in this case as the probability that a received sequence $\mathbf{y}^n$ is not compatible with the sent message is null. Now let's assume we use the decoding function ψ^*. In this case the decoded sequence $\Psi^*(\mathbf{y}^n)$ will be in agreement with the received sequence $\mathbf{y}^n$. We can easily compare the decoding error probability of the two coding schemes.

The decoding error probability of the decoding schemes, *i.e.* the probability of the decoding error event $\mathfrak{E}$ when using the decoder Ω ($\mathbb{P}rob_\Omega\{\mathfrak{E}\}$) is calculated as:

$$\mathbb{P}rob_\Omega\{\mathfrak{E}\} = \sum_{\mathbf{z}^n \in \mathcal{Y}^n} \mathbb{P}rob_\Omega\{\mathfrak{E}|\mathbf{Y}^n = \mathbf{z}^n\}\mathbb{P}rob\{\mathbf{Y}^n = \mathbf{z}^n\}.$$

The error probability for the two decoder Ψ and Ψ^* differ just by the term where $\mathbf{z}^n = \mathbf{y}^n$. For this term we have:

$$\begin{aligned}
\mathbb{P}rob_\Psi\{\mathfrak{E}|\mathbf{Y}^n = \mathbf{y}^n\} &= 1 \\
\mathbb{P}rob_{\Psi^*}\{\mathfrak{E}|\mathbf{Y}^n = \mathbf{y}^n\} &= (1 - \frac{1}{|\mathcal{A}(\mathcal{C}, \mathbf{y}^n)|})
\end{aligned}$$

As $\mathbb{P}rob_\Psi\{\mathfrak{E}|\mathbf{Y}^n = \mathbf{y}^n\} > \mathbb{P}rob_{\Psi^*}\{\mathfrak{E}|\mathbf{Y}^n = \mathbf{y}^n\}$, we have $\mathbb{P}rob_\Psi\{\mathfrak{E}\} > \mathbb{P}rob_{\Psi^*}\{\mathfrak{E}\}$. ■

The performance of erasure correction codes depends on Hamming distance properties of the codeword set $\mathcal{C}$; if $\mathcal{C}$ has a minimal Hamming distance $d_{\min}$, the code should be able to decode up to $d_{\min} - 1$ erasures. The decoding function Ψ^* achieves this decoding performance.

Lemma 2: An erasure correcting code $(\mathcal{C}, \Phi, \Psi^*)$ can decode up to $d_{\min}(\mathcal{C}) - 1$ erasures.

Proof: Let's assume that we have received a sequence $\mathbf{y}^n$. By its definition, the agreement set $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$ contains at least the sent codeword and all codewords that are in Hamming distance less than $e(\mathbf{y}^n)$ of the send sequence. However if $e(\mathbf{y}^n) < d_{\min}(\mathcal{C})$, per definition of the minimal distance there is no codeword with Hamming distance less than $d_{\min}(\mathcal{C})$ of a given sequence. The agreement set $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$ contains therefore only one codeword, the sent one. ■

This lemma have an important result. By Singleton bound, it is known that the largest possible minimal Hamming distance in a set containing K n-dimensional codewords is $n - \frac{\log(K)}{\log|\mathcal{X}|} + 1$. A Maximal Distance Separable (MDS) is a code attaining the Singleton Bound. A MDS code of rate R will have $K = |\mathcal{X}|^{nR}$ codeword and a miminal Hamming distance $n(1-R)+1$. By using the decoder Ψ^* one can decode up to $n(1-R)$ erasures in a block of n transmissions.

Let's assume a channel that accepts a law of large number for erasure proportion, *i.e.*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i\leq n} Z_i \to p$$

where $\{Z_i\}$ is the loss process of the channel. The typical set $\mathcal{A}_\epsilon^{(n)}$ containing typical erasure patterns at the output of an erasure channel contains all erasure patterns with an erasure proportion that is is within a distance $\delta^n()$ of the asymptotic erasure probability p, *i.e.* $\forall \mathbf{y}^n \in \mathcal{Y}^n \in \mathcal{A}_\epsilon^{(n)},\ |\frac{e(\mathbf{y}^n)}{n} - p| < \delta_\epsilon(n)$ with $\lim_{n\to\infty} \delta_\epsilon(n) = 0$. This means that if a MDS code $\mathcal{C}_n$ of length n with rate $R = (1-p) - \frac{\delta_\epsilon(n)}{n}$ exists it can surely correct all typical error patterns of length n at the output of such a channel. A sequence of MDS codes $\mathcal{C}_n$ of rate $R = (1-p) - \frac{\delta_\epsilon(n)}{n}$ achieve therefore the capacity of such a channel.

Reed Solomon codes are a well-known family of MDS code and for every finite block length n it is possible to choose an alphabet size q and a finite field $GF(q)$ to build a Reed Solomon code (n, nR, q). However other MDS code applicable for erasure codes exists [**?**]. Almost-MDS [6] codes with linear decoding complexity are also applicable in this context. The decoding of all these codes can be reduced to solving a set of nR linear equations among n equation with nR variables.

The above analysis lead to the following strong converse theorem

Theorem 3—Strong Converse: Over an erasure channel following a strong law of large numbers for the proportion of erasures, the decoding error probability of any coding scheme with a rate $R > C$ where C is the capacity goes to 1. If the channel follows a large deviation principle, the decoding error probability goes to 1 exponentially with a speed that depends on $R - C$.

Proof: To prove the converse it is enough to prove that no code with rate $R > C$ can achieve a vanishing decoding error probability. We have proven in lemma 1 that we can restrict ourself to the agreement set decoding s any other decoding function will lead to higher decoding error probability. Let's assume a code $\mathcal{C}$ of length n and rate R with a minimal distance $d_{\min}(\mathcal{C}_n)$. The code contains $|\mathcal{X}|^{nR}$ codewords. we will denote as $\mathfrak{E}$ the decoding error event. We are also partionning the set of erasure patterns $\{0,1\}^n$ into n subsets $\mathcal{Z}_k,\ k = 0, \ldots, n$ where we define $\mathcal{Z}_k = \{\mathbf{z}^n \in \{0,1\}^n | \sum_{i=1}^n z_i = k\}$, *i.e.* $\mathcal{Z}_k$ is the set of all erasure pattern containing k erasures. Obviously $\cup_{i=1}^n \mathcal{Z}_i = \{0,1\}^n$. Now we have :

$$\mathbb{P}\text{rob}\{\mathfrak{E}\} = \sum_{i=1}^{n} \sum_{\mathbf{z}^n \in \mathcal{Z}_i} \mathbb{P}\text{rob}\{\mathfrak{E}|\mathbf{z}^n\}\mathbb{P}\text{rob}\{\mathbf{z}^n\}$$

Now let's assume that an erasure pattern $\mathbf{z}^n \in \mathcal{Z}_k$ has happened, *i.e.* $\mathbf{y}^n = \mathbf{x}^n(\mathbf{z}^n)$ have been received. Based on agreement set decoding, unless the agreement set $\mathcal{A}(\mathcal{C}, \mathbf{y}^n)$ contains only one codeword we can have a decoding error. However with k erasure we can define in absolute terms, a maximum of $|\mathcal{X}|^{(n-k)}$ disjoint agreement set relative to different choice over the symbols in the $n-k$ received symbols when an erasure pattern $\mathbf{z}^n$ happens. Now if the number of codewords in the code ($|\mathcal{X}|^{nR}$) is larger than this value, by pigeon hole priciple, we will have at least one of the agreement sets that will have more than 1 members leading to a possible decoding error for at least $|\mathcal{X}|^{nR} - |\mathcal{X}|^{(n-k)} + 1$ codewords. This will surely happen if $k > n(1-R)$. Nevertheless when $k \leq n(1-R)$ it is still possible to choose codewords in such a way than the agreement sets for all received sequences with k erasures are disjoints and have a single member that is the sent codeword.

We can give a lower bound for decoding error when $k > n(1-R)$ as $\mathbb{P}\text{rob}\{\mathfrak{E}|\mathbf{z}^n \in \mathcal{Z}_k\} \geq \frac{1}{2}\frac{|\mathcal{X}|^{nR} - |\mathcal{X}|^{(n-k)}}{|\mathcal{X}|^{nR}}$, where the first fraction term comes from the fact that we need at least 2 members in the agreement set to have an error and the second terms comes from the fact that $|\mathcal{X}|^{nR} - |\mathcal{X}|^{(n-k)}$ codewords among the $|\mathcal{X}|^{nR}|$ leads to a decoding error. We can therefore lower bound the decoding error probability of a code of rate R as :

$$\mathbb{P}\text{rob}\{\mathfrak{E}\} \geq \frac{1}{2} \sum_{i=n(1-R)+1}^{n} (1 - |\mathcal{X}|^{-(i-n(1-R))}) \sum_{\mathbf{z}^n \in \mathcal{Z}_i} \mathbb{P}\text{rob}\{\mathbf{z}^n\}$$

where the term $\sum_{\mathbf{z}^n \in \mathcal{Z}_i} \mathbb{P}\text{rob}\{\mathbf{z}^n\}$ is simply the probability of observing i erasure in a block of n transmission, *i.e.* $\mathbb{P}\text{rob}\{e(\mathbf{z}^n) = i\}$. By regrouping terms and by adding negative terms we can simplify the lower bound as :

$$\mathbb{P}\text{rob}\{\mathfrak{E}\} \geq \frac{1}{2}\left(\mathbb{P}\text{rob}\{e(\mathbf{z}^n) > n(1-R)\} - \sum_{i=0}^{n}(|\mathcal{X}|^{-(i-n(1-R))})\mathbb{P}\text{rob}\{e(\mathbf{z}^n) = i\}\right)$$

We can analyse the asymptotic behaviour of the second term by using Jensen inequality :

$$\sum_{i=0}^{n} |\mathcal{X}|^{-i}\mathbb{P}\text{rob}\{e(\mathbf{z}^n) = i\} \geq |\mathcal{X}|^{-\bar{e}(\mathbf{z}^n)}$$

This leads to the following upper bound for the second term :

$$-\sum_{i=0}^{n}(|\mathcal{X}|^{-(i-n(1-R))})\mathbb{P}\text{rob}\{e(\mathbf{z}^n) = i\} \leq -|\mathcal{X}|^{-n(R-(1-\frac{\bar{e}(\mathbf{z}^n)}{n}))}$$

This means that the second term converges to 0 when $n \to \infty$ with a speed controlled by $(R - \frac{\bar{e}(\mathbf{z}^n}{n})$, meaning that asymptocally the lower bound on decoding error probability goes to :

$$\lim_{n\to\infty} \mathbb{P}\mathrm{rob}\{\mathfrak{E}\} \geq \frac{1}{2}\left(\mathbb{P}\mathrm{rob}\{\frac{1}{n}e(\mathbf{z}^n) > (1-R)\}\right)$$

Therefore the lower bound on decoding error probability do not depend on the coding scheme used but rather on the behaviour of the empirical proportion of loss observed over the channel. If we assume that the channel follows a strong law of large number, *i.e.* $\lim_{n\to\infty} \frac{1}{n}e(\mathbf{z}^n) = p$, we will have :

$$\forall\epsilon > 0, \delta > 0, \ \exists N \text{ such that } \mathbb{P}\mathrm{rob}\{\frac{1}{N}e(\mathbf{z}^N) > p + \epsilon\} \leq \delta$$
$$\forall\epsilon > 0, \delta > 0, \ \exists N \text{ such that } \mathbb{P}\mathrm{rob}\{\frac{1}{N}e(\mathbf{z}^N) > p - \epsilon\} \geq 1 - \delta$$

Now, if $R = 1 - p + \epsilon, \ \epsilon > 0$ using the above expression for any δ you can find a value N such that $\mathbb{P}\mathrm{rob}\{\frac{1}{N}e(\mathbf{z}^N) > (1-R)\} \geq 1 - \delta$ and if $R = 1 - p - \epsilon, \ \epsilon > 0$ for any δ you can find a value N such that $\mathbb{P}\mathrm{rob}\{\frac{1}{N}e(\mathbf{z}^N) > (1-R)\} \leq \delta$. This shows that the capacity value C acts as a sharp qualitative boundary; for $R < C$ the lower bound on the decoding error probability goes to 0, for $R > C$ this probability will be bounded away and the error decoding probability goes to 1. This proves the converse.

Moreover this argument can be strengthened if the channel follows a large deviation principle with rate function $r(.r)$, *i.e.*

$$\mathbb{P}\mathrm{rob}\{\frac{1}{N}e(\mathbf{z}^N) > p + \epsilon\} \leq e^{Nr(\epsilon)}$$
$$\mathbb{P}\mathrm{rob}\{\frac{1}{N}e(\mathbf{z}^N) > p - \epsilon\} \geq 1 - e^{Nr(\epsilon)}$$

Under this condition the channel accept a strong converse and for $R < C$ the lower bound on the decoding error probability goes to 0 exponentially fast with a speed that is a function of C_R, and for $R > C$ the decoding error probability goes to 1 with a speed that is a function of $R - C$, resulting in the strong converse. ■

This strong converse bound is quite general and can be applied to a very large spectrum of practical channel. In particular stationary ergodic channels that contains Markov Modulated channels [**?**] as well as erasure channel with Long Range dependences fall in this class. It is noteworthy that the converse is valid even for non-ergodic channel on the condition of existence of an asymptotic loss proportion. This last setting is interesting for analysis of channels with failures, *i.e.* channel that fails after an amount of time and do not get back online.

The above proof shows also that even if the capacity of erasure channels only depends on the asymptotic loss probability, but the empirical performances of erasure coding schemes depends largely on the memory structure through the existence of not of a large deviation principle and through the specific form of $r(.)$.

Properties of the point to point erasure channel give a strong basis for the analysis of multi-user erasure channels. We have shown for point to point channel that the form of the capacity region is the same for memoryless and more general channels with memory. Moreover we showed that the same type of codes, namely MDS codes, achieve the capacity of all erasure channels. We will therefore in the forthcoming describe the memoryless broadcast and the relay channels and say how the memoryless can be extended to the general case of channel that accept a law of large numbers.

III. Broadcast erasure Channels

In this section, for the sake of completeness, we wil describe the broadcast erasure channel with *degraded message set*. This channel have been fully characterized in [7]. For this channel we have a single source that want to transmit information to a set of m receivers. We want to transmit over this channel a "*Degraded Message Set*" (DMS) [8]. For such a message set the capacity region is the set of rate tuples $(\mathbf{R}_1, \cdots, \mathbf{R}_m)$ such that the source can reliably transmit with rate $\mathbf{R}_i$ to the receiver i the i^{th} level of the DMS.

The capacity region for the erasure broadcast has been derived in [7]

Theorem 4—Capacity of Degraded Message Set over a Broadcast Erasure channel: A tuple of rates $(\mathbf{R}_0, \cdots, \mathbf{R}_{m-1})$ is achievable over a general broadcast erasure channel with degraded message set if and only if :

$$\sum_{i=1}^{m-1} \frac{\mathbf{R}_i}{1-p_i} < 1$$

where p_i is the asymptotic packet loss rate between sender and receiver i. ■

It is first noteworthy that the above capacity region do not assume any degradedness assumption. Moreover, the above theorem proves that for broadcast erasure channels we cannot go beyond the timesharing, *i.e.* capacity region is not going beyond the timesharing through superposition as it ithe case for error channels. Any coding schemes implementing a time sharing and achieving a rate $\mathbf{R}_i$ in the above defined region achieves the capacity.

In particular the capacity region can be achieved through a Priority Encoding Transmission (PET) mechanism introduced by Albanese in [9]. In this coding scheme the information flow is partitionned into m different priority levels. Each priority level is encoded using a MDS code with the same length n and defined over the same alphabet $\mathcal{X}$. The rate of each MDS code is set to an rate equal to $\mathbf{R}_i$,

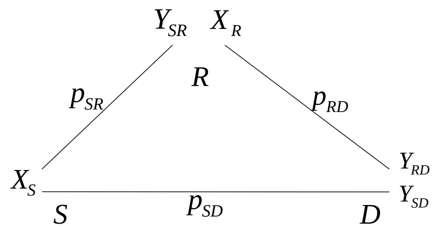

Fig. 1. Single relay erasure channel

with $\mathbf{R}_1 < \mathbf{R}_2 < \cdots < \mathbf{R}_m$. The encoded data are interleaved and encapsulated in symbols to be transmitted over the channel. Each transmitted symbol contains a concatenation of all the m different levels of encoding and is therefore defined over the alphabet $\mathcal{X}^m$.Now, if a receiver receives at least $n\mathbf{R_i}$ packets sent over the channel it is able to decode up to the i^{th} priority level resulting in the reception of the Degraded Message Set. The above coding scheme is attracting as the broadcast encoding scheme is constructed as m parallel and independent point to point MDS coding schemes. It is interesting to investigate if any other more complex coding scheme achieving the capacity can have better performance in term of error exponents. It is proven in [7] that this encoding scheme is indeed optimal, *i.e.* no other coding scheme can achieve better error exponents that this scheme over broadcast erasure channels. This last theorem provides also strong converse bound for the erasure broadcast channel.

In next section we will deal with the relay erasure channel. This channel is indeed interesting as it describes the simplest type of cooperation in a network.

IV. Single-sender Single-relay Erasure Channel

The simplest scenario of node cooperation is the single relay channel shown in figure 1. In this setting we have three nodes: one node (S) acting as a sender, one relay node (R) receiving information from sender and collaborating with it to transmit information to the destination node (D). The destination decodes jointly the symbols received from the sender and the relay to figure out the message symbols sent by the sender. This channel have been initially investigated in [**?**].

In this paper we assume that nodes communicate over non-interfering erasure channels, *i.e.* the relay to destination channel is separated from the source to destination channel. This is relevant and applicable to computer networks where the separation between these two channels is achieved by using two different physical channel (interface card) at the destination node or through logical separation by a time-sharing mechanism (centralized or distributed as CSMA/CA access control). This channel can be described with five random variables X_S, X_R, Y_{SD},Y_{SR} and Y_{RD} and a sequence of conditional channel transfer function $\mathcal{E}\left(\mathbf{y}_{SD}^n, \mathbf{y}_{SR}^n, \mathbf{y}_{RD}^n | \mathbf{x}_S^n, \mathbf{x}_R^n\right)$. We further define $Y_D = (Y_{SD}, Y_{RD})$ as the received variable from the sender and the relay at the receiver. The sent symbols are defined in a set $\mathcal{X}$ and as we have erasure channels the received symbols are in $\mathcal{Y} = \mathcal{X} \cup \{\mathtt{e}\}$. In general setting the transfer function of the relay channel $\mathcal{E}\left(\mathbf{y}_{SD}^n, \mathbf{y}_{SR}^n, \mathbf{y}_{RD}^n | \mathbf{x}_S^n, \mathbf{x}_R^n\right)$ gives the probability that when the sequence of n symbols $\mathbf{x}_S^n$ is sent by sender and $\mathbf{x}_R^n$ by the relay, $\mathbf{y}_{SD}^n$ and $\mathbf{y}_{RD}^n$ are received at the receiver and the relay receives $\mathbf{y}_{SR}^n$.

The erasure nature of the channel leads to some simplification. Under the assumption that the losses do not depend on the sent messages, the transfer function is characterized by the erasure process $\mathbf{z}^n$ probability. Moreover, if we assume that the erasure between channels are non-correlated, *i.e.* erasures are not spatially correlated, the relay erasure channel can simply be characterized by the transfer function of three point to point channels as :

$$\mathcal{E}\left(\mathbf{y}_{SD}^n, \mathbf{y}_{SR}^n, \mathbf{y}_{RD}^n | \mathbf{x}_S^n, \mathbf{x}_R^n\right) = \mathcal{E}_{SR}\left(\mathbf{z}_{SR}^n | \mathbf{x}_S^n\right) \mathcal{E}_{SD}\left(\mathbf{z}_{SD}^n | \mathbf{x}_S^n\right) \mathcal{E}_{RD}\left(\mathbf{z}_{RD}^n | \mathbf{x}_R^n\right)$$

In other terms the relay channel could be simply characterized by three erasures probabilities p_{SR}^k, p_{SD}^k and p_{RD}^k. A situation of interest is the physically degraded channel [**?**], where the signal received at destination y_{SD} from the source S is a random alteration of the signal received at relay y_{SR}. The physically degraded channel condition in the context of erasure channels translates to the condition that **any** symbols that have been received from source at the destination should be received also by the relay. A weaker condition is the virtually degraded condition, where it is only assumed that the erasure rate over the source to relay channel is smaller than the erasure over the source to destination channel. This condition is essentially the same as the more capable channel condition defined in [10].

As explained before we will present here the memoryless case, however all results can be extended to more general channel as stationary and ergodic erasure channels and even to channel that follow a law of large numbers by replacing the memoryless erasure probability with the asymptotic erasure mean calculated over the channel statistics.

A. Capacity bounds

The Min Cut-Max Flow bound , also called cut-set bound [**?**], is the most general bound defined for multi-user channels. This bound can be instantiated for erasure relay channels. It results in a simple closed form that is given in the following theorem :

Theorem 5— Min Cut-Max Flow bound for erasure relay channel: The capacity region of a single-sender single-relay erasure channel is bounded by :

$$\mathbf{R} \leq \min\{(1-p_{SR}p_{SD});(1-p_{SR})+(1-p_{RD})\}$$

where the first and second terms are respectively the sender side and the receiver side cut-set bounds.

Proof: The cut-set bound for the relay channel described in figure 1 is

$$\mathbf{R} \leq \sup_{p(X_S,X_R)} \min\{I(X_S;Y_{SR},Y_{SD}|X_R), I(X_S,X_R;Y_{SD},Y_{RD})\} \tag{1}$$

The first term in the brackets represents the sender side cut set and the second term is corresponding to the receiver side cut-set.

Let's begin with the first term. To deal with this term we have first to refer to another problem: the entropy characterization problem as defined in [11]. The entropy characterization problem consists in determining the closure of the set of all vector of form

$$\bigcup_{k=1}^{\infty}\left\{\frac{1}{k}\left(H(\mathbf{X}_S^k|f(\mathbf{X}_S^k)), H(\mathbf{Y}_{SR}^k|f(\mathbf{X}_S^k)), H(\mathbf{Y}_{SD}^k|f(\mathbf{X}_S^k))\right)\right\}$$

with f running over all functions with domain defined on $\mathcal{X}^k$.

The solution to this problem is described in [11] where it is shown that the above set is equal to the following set:

$$\bigcup_{k=1}^{\infty}\left\{\frac{1}{k}\left(H(\mathbf{X}_S^k|U), H(\mathbf{Y}_{SR}^k|U), H(\mathbf{Y}_{SD}^k|U)\right)\right\}$$

where U is running over all finite random variables defined on a finite support $\mathcal{U}$ and where $U \to \mathbf{X}_S^k \to (\mathbf{Y}_{SR}^k, \mathbf{Y}_{SD}^k)$ make a Markov chain. The cardinality of the set $\mathcal{U}$ is bounded by $|\mathcal{U}| \leq |\mathcal{X}|+2$. Now we can return back to the relay channel.

The symbol sent by the relay node at time step k, X_R^k, is a function of the received sequence $\mathbf{y}_{SR}^{k-1}$ up to time $k-1$, *i.e.* $X_R^k = \psi_R^k(\mathbf{y}_{SR}^{k-1})$. Moreover, the symbol received at the relay at time k, y_{SR}^k, is a stochastic function of x_S^k for every $k \leq n$. Therefore the symbol vector sent by the relay in the first k transmission can be expressed as a stochastic function of the previous symbols transmitted by sender, *i.e.* $\mathbf{x}_R^k = f(\mathbf{x}_S^{k-1})$. Thanks to the entropy characterization problem, there exists a random variable U such that $U \to X_S \to (Y_{SR}, Y_{SD})$ is a Markov chain and

$$\begin{aligned} I(X_S;Y_{SR},Y_{SD}|X_R) &= I(X_S;Y_{SR},Y_{SD}|U) \\ &= I(X_S;Y_{SR},Y_{SD}) - I(U;Y_{SR},Y_{SD}) \end{aligned}$$

The first term above depends only on the symbols sent by source S, where the second term depends on the random variable U that replaces X_R, and represents therefore the relay action. The above term can be maximized when the relay acts such that $I(U; Y_{SR}, Y_{SD}) = 0$. However as Y_{SR} and Y_{SD} depends only on X_S, we need to chose U independent from X_S to satisfy this condition.

Now let's look at the receiver side cut-set, *i.e.* $I(X_S, X_R; Y_{SD}, Y_{RD})$. As explained before, we consider an erasure channel where symbols transmitted by source and relay are not interfering. Therefore from the view point of the receiver, the relay channel consists of two point-to-point channels $(X_S; Y_{SD})$ and $(X_R; Y_{RD})$ with independent erasures. The mutual Information is maximized when X_S and X_R are chosen independently. This means that the receiver side term is bounded by $I(X_S; Y_{SD}) + I(X_R; Y_{RD})$.

The previous independence condition will therefore jointly maximize the two terms of the cut-set bound. Thus the largest upperbound for capacity is :

$$\mathbf{R} \leq \sup_{p(x_S, x_R)} \min\{I(X_S; Y_{SR}, Y_{SD}), I(X_S; Y_{SD}) + I(X_R; Y_{RD})\}$$

Now it remains to derive this term for the erasure channels. Using the Shearer theorem we have :

$$I(X_S; Y_{SD}) + I(X_R; Y_{RD} = (1 - p_{SD}) + (1 - p_{RD})$$

and

$$I(X_S; Y_{SR}, Y_{SD}) = H(X_S) - H(X_S | Y_{SR}, Y_{SD}) = 1 - p_{SD} p_{SR}$$

that gives the researched upper bound. ■

The achievability of the cut-set bound have been a major research question during the past two decades. Up to now the cut-set bound is known to be achievable only for the physically degraded channel and the question of its achievability in more general setting (non degraded channels) is still a challenging and open question.

An inspection of the above proof gives direction to find coding schemes that can achieve the cut-set bound. one needs to ensure that the variables sent by the sender X_S and the relay X_R are independent but still complementary to enable a maximal rate at the receiver. However a high level of collaboration between nodes in the channel is needed to satisfy this condition. We know that in the physically and virtually degraded situation this high level of cooperation is possible. In **??** a coding scheme using linear erasure correcting code is presented that achieves the cutset bound under physically degraded conditions. In this coding scheme the relay decodes the messages sent by source and cooperates in the communication by transmitting re encoded symbols with a code that is chosen so that the encoding matrices have linearly independent rows. The full cooperation in this case consists of the ability of the relay to fully decode the message and to re encode the message using an independent code. As we will show this coding scheme can achieve the cutset bound even for non degraded scenarios (the virtually degraded) case as a the relay node will still be able to fully cooperate. There are however some non degraded situation where full cooperation is not possible as the relay might not be able to decode the messages sent by the source. In this case only a restricted level of cooperation is possible. We will see that in some situation this cooperation is not possible at all and the relay is not anymore useful. The conditions that lead to this situation will be described through a strong converse for relay erasure channels.

In following section we will describe erasure correcting codes suitable for the relay channel. This description is the main ingredient of the strong converse bound presented in section IV-C.

B. Erasure coding for the relay channel

The description given for erasure correction codes in section II-B can be extended to relay channels. Let's assume that the source in a relay channel is using a code $\mathcal{C}_S$ of length n and rate $\mathtt{R}$ with an encoding function $\phi_S : \{1, \ldots, |\mathcal{X}|^{n\mathtt{R}}\} \rightarrow \mathcal{C}_S$ and sends x_S over the channel. To simplify the presentation we will as before assume that the sent message m can be retrieved uniquely by decoding at receiver the correct sent sequence $\phi_S(m)$.

At time k a sequence $\mathbf{y}_{SR}^k \in \mathcal{Y}^k$ is received at the relay node. The relay maps the received sequence $\mathbf{y}_{SR}^k$ into the symbol $x_R^k \in \mathcal{X}$ transmitted at time k through an encoding function $\phi_R^k(\mathbf{y}_{SR}^k)$, *i.e.* $\mathbf{x}_R^k = \phi_R^k(\mathbf{y}_{SR}^k)$. As we are analysing the transmission of a block of n symbols from source to destination and as erasures are assumed to be independent between channels, we might transform any relay coding scheme to a a Markov block code, where the relay wait until the reception of $\mathbf{y}_{SR}^n$ and generate thereafter a codeword $\mathbf{x}_R^n = \phi_R(\mathbf{y}_{SR}^n)$ from a code $\mathcal{C}_R$ of length n and rate $\mathtt{R}'$, *i.e.* $|\mathcal{C}_R| = |\mathcal{X}|^{n\mathtt{R}'}$. If a rate $\mathtt{R}'$ is achieved by the initial code the resulting markov block code will have the same erasure decoding performance and a rate equal to $\mathtt{R}'\frac{B-1}{B}$ after the transmission $B-1$ blocks of n symbols. With larger block number B the resulting code will have the same rate as the initial code. Therefore for the purpose of erasure correction performance we might assume that a Markov block code is used and get ride of the k superscript in the encoding function $\phi_R^k(.)$. However in general terms the encoding function ϕ_R could be chosen arbitrarily. Nevertheless, the number of possible received sequence at relay , $\mathbf{y}_{SR}^n$ will be larger than $|\mathcal{C}_S|$. Therefore unless the rate $\mathtt{R}'$ is large enough, it is not possible to assign to each received sequence a codeword of $|\mathcal{C}_R|$, one have therefore to assign several received sequence $\mathbf{y}_{SR}^n$ to at least one codeword of $|\mathcal{C}_R|$.

At the destination node two sequences of symbols are received: $\mathbf{y}_{SD}^n$ from the source and $\mathbf{y}_{RD}^n$ from the relay. The destination should implement a joint decoding function $\psi_D : \mathcal{Y}^n \times \mathcal{Y}^n \to \{1, \cdots, |\mathcal{X}|^{n\mathrm{R}}\}$, that will get the two received sequence $\mathbf{y}_{SD}^n$ and $\mathbf{y}_{RD}^n$ and map them to the message sent by source.

The above description is generic and not particular to a specific erasure correcting code for the relay channel, *i.e.* any specific code can be reduced to the above structure. A general erasure correcting code for the relay channel will be characterized by $(\mathcal{C}_S, \mathcal{C}_R, \phi_R, \psi_D)$. However, similarly to point to point erasure channel, we will prove through a lemma that among the very large set of possible relay decoding function there is a specific choice ψ_D^* that superseed all other functions in term of probability of decoding error. We will describe in the forthcoming the function decoding function ψ_D^*.

As described in lemma1, the ambiguity about the sequence sent by the source is represented by the agreement set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n)$. Let's define for each codeword $\mathbf{x}_{RD}^n \in \mathcal{C}_R$, the ambiguity set $\mathcal{B}(\mathcal{C}_S, \mathbf{x}_{RD}^n) \subset \mathcal{C}_S$ defined as $\mathcal{B}(\mathcal{C}_S, \mathbf{x}_{RD}^n) = \bigcup_{\mathbf{y}^n \in \phi_R^{-1}(\mathbf{x}_{RD}^n)} \mathcal{A}(\mathcal{C}_S, \mathbf{y}^n)$ that will contains all sequences sent by the source that might have resulted in the transmission of $\mathbf{x}_{RD}^n$ by the relay. The set $\mathcal{B}(\mathcal{C}_S, \mathbf{x}_{RD}^n)$ represents the ambiguity about the sent message as inferred by the codeword sent by the relay. It is noteworthy that even if the sets $\mathcal{P}(\phi_R^{-1}(\mathbf{x}_{RD}^n))$ are disjoint, the sets $\mathcal{B}(\mathbf{x}_{RD}^n)$ might not be disjoint in general.

Obviously, the sequence $\mathbf{y}_{SD}^n$ should be in agreement with the sent sequence $\mathbf{x}_S$, *i.e.* $\mathbf{x}_S \in \mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n)$. The sequence coming from the relay $\mathbf{y}_{RD}^n$ is more complex to exploit as it is affected by the relay encoding function ϕ_R. By observing $\mathbf{y}_{RD}^n$, we know that the sequence sent by the relay will be in the agreement set $\mathcal{A}(\mathcal{C}_R, \mathbf{y}_{RD}^n)$. Now we define the ambiguity set relative to the sequence $\mathbf{y}_{RD}^n$, $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ as as a union of ambiguity sets:

$$\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n) = \bigcup_{\mathbf{x}_R^n \in \mathcal{A}(\mathcal{C}_R, \mathbf{y}_{RD}^n)} \mathcal{B}(\mathbf{x}_R^n)$$

By its construction the set $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ contains the agreement set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n)$ and more generally all sequences $\mathbf{x}_S^n \in \mathcal{C}_S$ that could have lead to the observation of $\mathbf{y}_{RD}^n$.

We can define the decoding function ψ_D^* as

$$\psi_D^*(\mathbf{y}_{RD}^n, \mathbf{y}_{SD}^n) = \Omega\left(\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)\right)$$

where the function $\Omega(.)$ represents a random and uniform choice among the member of the set given as variable. By the erasure nature of the channel $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{B}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ is never empty. Decoding errors occurs when this intersection contains more than a single codeword.

We show in figure 2 an illustration explaining the decoding in single relay erasure channels.

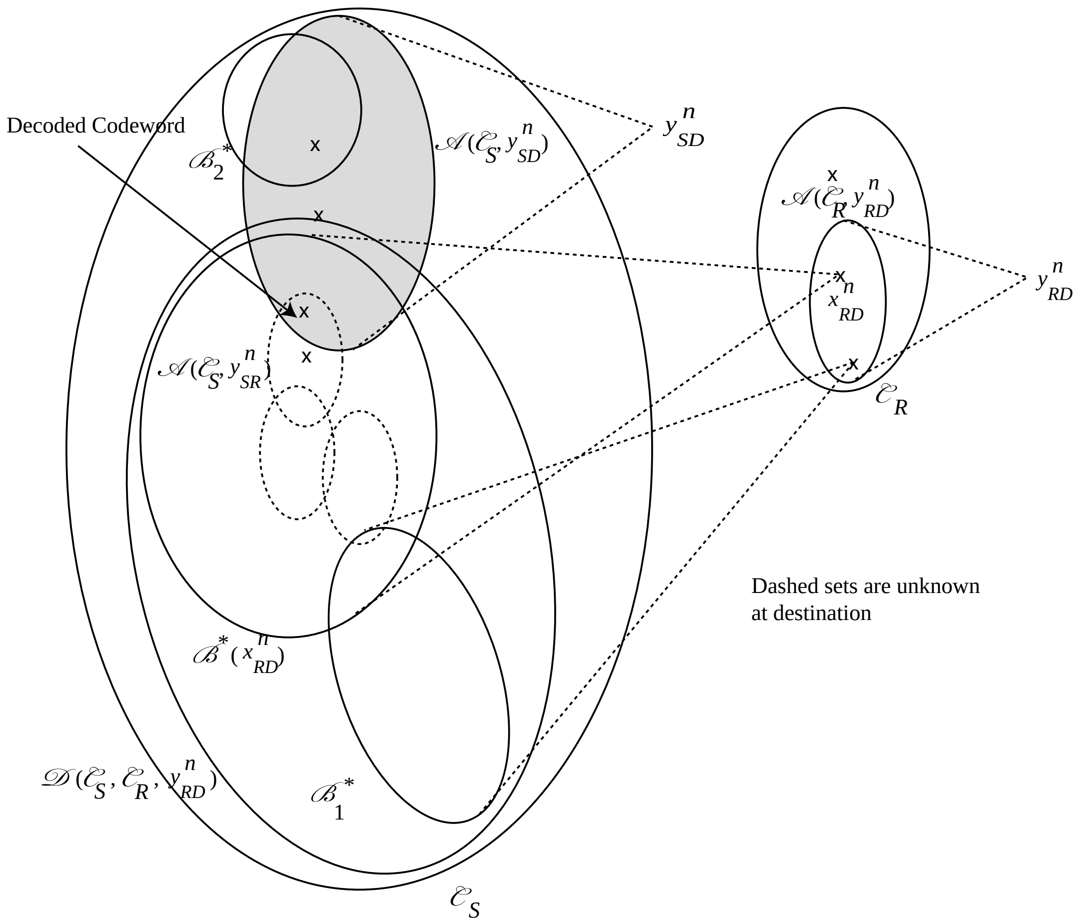

Fig. 2. Structure of joint erasure decoding of $\mathbf{y}_{RD}^n$ and $\mathbf{y}_{SD}^n$ code

Lemma 3: The decoding functions ψ_D^* is the best decoding function for a relay code $(\mathcal{C}_S, \mathcal{C}_R, \phi_R)$ in term of erasure correction performances.

Proof: By definition the set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ contains all sequences that are in agreement with the two sequences $\mathbf{y}_{SD}^n$ and $\mathbf{y}_{RD}^n$ and only these sequences. The error probability when $(\mathbf{y}_{SR}^n, \mathbf{y}_{SD}^n, \mathbf{y}_{RD}^n)$ is received over different channels and when using the decoding function ψ_D^*, will be $\left(1 - \frac{1}{|\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)|}\right) < 1$. Now let's assume that another decoding function ψ_D is used that such that $\psi(\mathbf{y}_{RD}^n, \mathbf{y}_{SD}^n) = \mathbf{x} \notin \mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$. There will be a decoding error in this case as the probability that the sent sequence do not belongs to $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ is null and the error probability will be 1. Resulting in a larger error probability for this case. ■

Based on the above lemma and for the purpose of lower bounding the error probability, one can always assume that the decoding functions ψ_D^* are used. A deterministic coding scheme for relay channel is therefore defined by fixing $(\mathcal{C}_S, \mathcal{C}_R, \phi_R)$. Similarly to the point to point case, the performance of a particular erasure correcting scheme depends on minimal distance properties of $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n)$ and $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R\mathbf{y}_{RD}^n)$.

An interesting interpretation of coding schemes for erasure relay schemes is coming from its equivalence with a point to point coding scheme with randomly varying codesets. One might see the sets $\mathcal{B}(\mathbf{x}_R^n)$ as codesets indexed by $\mathbf{x}_R^n$. Let's assume a point to point coding scheme over the Source to Destination point to point erasure channel that choose randomly for each code block the codeset to be used among the $|\mathcal{X}|^{n\mathrm{R}'}$ codesets $\mathcal{B}(\mathbf{x}_R^n)$ the one that will be used. Intially the Destination is not aware of the codeset to be used and can just assume that $\mathcal{C}_S = \bigcup_{\mathbf{x}_R^n \in \mathcal{C}_R} \mathcal{B}(\mathbf{x}_R^n)$ has been in use. However, by receiving $\mathbf{y}_{RD}^n$, the destination will reduce its ambiguity about the codeset to $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ and use this codeset to decode the particular message send by using the received sequence $\mathbf{y}_{SD}^n$. This interpretation is also generic and not particular to a specific relay coding scheme, *i.e.* any coding scheme for erasure relay channel can be translated to a point to point coding scheme with randomly varying codesets with the choice of appropriate coding sets. This equivalence will be very useful to derive converse bounds for the relay erasure channel.

C. Strong Converse for the erasure relay channels

We have now all the ingredients for proving a strong converse for erasure relay channels. Similarly to point to point case we will process by proving a lower bound for decoding error probability and deriving condition where this lower bound is bounded away from 0. The converse bound will be proven using the interpretation of the equivalence of a relay coding scheme to a point to point coding scheme with varying codeset at source. However, we need first to relate properly the error probability of these two approaches. This is done by the following lemma.

Lemma 4: A relay coding scheme for erasure channel can attain a vanishing error probability if and only if the probability that a codeset $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R\mathbf{y}_{RD}^n)$ with a non-zero probability of error over the source to destination erasure channel is designated by the the symbol sequence $\mathbf{y}_{RD}^n$, is vanishing.

Proof: The lemma proceeds directly from the fact that any relay coding scheme for erasure can be reduced to a point to point erasure correcting scheme with varying codesets. Through this equivalence the error probability of a relay coding scheme can be calculated as :

$$\mathbb{P}\mathrm{rob}\{\mathcal{E}\} = \sum_{(\mathbf{y}_{RD}^n, \mathbf{y}_{SD}^n) \in \mathcal{Y}} \mathbb{P}\mathrm{rob}\{\mathcal{E}|\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n), \mathbf{y}_{SD}^n\} \mathbb{P}\mathrm{rob}\{\mathbf{y}_{RD}^n\} \mathbb{P}\mathrm{rob}\{\mathbf{y}_{SD}^n\}$$

where $\mathbb{P}\mathrm{rob}\{\mathcal{E}|\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n), , \mathbf{y}_{SD}^n\}$ stands for the error probabiliy using the codeset $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ when the sequence $\mathbf{y}_{SD}^n$ is received from source at destination.

Obviously the above decoding error probability will not go to zero unless the probability that a codeset $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R\mathbf{y}_{RD}^n)$ with a non-zero decoding probability, *i.e.* a code with $\mathbb{P}\text{rob}\{\mathcal{E}|\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n), \mathbf{y}_{SD}^n\}$, is chosen by the source is vanishing, *i.e.* $\mathbb{P}\text{rob}\{\mathbf{y}_{RD}^n\} \to 0$. Inversely if the probability that a codeset $\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R\mathbf{y}_{RD}^n)$ with a non-zero decoding probability is chosen by the source is vanishing the decoding error probability will also to zero by a dominated convergence theorem. ■

This lemma is the main ingredient of the strong converse we are proposing for the relay erasure channel.

Theorem 6—Strong converse for the erasure relay channel: No relay coding scheme can achieve a rate larger than $(1 - p_{SR})$ with a vanishing decoding error probability. This converse is a strong converse.

Proof: We have proven in lemma 3 that we can restrict ourself to the the decoding function ψ_D^* as any other decoding function will lead to higher decoding error probability. Let's assume a relay code $(\mathcal{C}_S, \mathcal{C}_R, \phi_{RD})$ of length n and rate $(\mathtt{R}, \mathtt{R}')$ with a minimal distance couple $(d_{\min}(\mathcal{C}_S), d_{\min}(\mathcal{C}_R))$. The code $\mathcal{C}_S$ contains $|\mathcal{X}|^{n\mathtt{R}}$ codewords and the code $\mathcal{C}_R$ contains $|\mathcal{X}|^{n\mathtt{R}'}$ codewords. We will denote as $\mathfrak{E}$ the decoding error event, *i.e.* the event that the set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n) \bigcap \mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ contains more than a single codeword. Using the equivalence described before between the relay coding scheme and a point to point code with varying codeset one might rewrite the error probability as :

$$\mathbb{P}\text{rob}\{\mathcal{E}\} = \sum_{(\mathbf{y}_{RD}^n, \mathbf{y}_{SD}^n) \in \mathcal{Y}} \mathbb{P}\text{rob}\{\mathcal{E}|\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n), \mathbf{y}_{SD}^n\}\mathbb{P}\text{rob}\{\mathbf{y}_{RD}^n\}\mathbb{P}\text{rob}\{\mathbf{y}_{SD}^n\}$$

Based on lemma 4, it is enough to show that for any relay coding scheme with $\mathtt{R} > (1 - p_{RS})$ there is some probable situation where the codeset indicated by the received sequence $\mathbf{y}_{RD}^n$ have a none vanishing probability of error.

Let's assume that a relay coding scheme is used with $\mathtt{R} > (1-p_{RS})$. Under this situation and following the theorem 2 no coding scheme can ensure that all messages sent by source will be received without ambiguity at relay. This means that with high probability the set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n$ will contain more than a single codeword. Let's assume that

We are also partionning the set of erasure patterns $\{0,1\}^n$ into n subsets $\mathcal{Z}_k,\ k = 0, \ldots, n$ where we define $\mathcal{Z}_k = \{\mathbf{z}^n \in \{0,1\}^n | \sum_{i=1}^n z_i = k\}$, *i.e.* $\mathcal{Z}_k$ is the set of all erasure pattern containing k erasures. Now we have :

$$\mathbb{P}\text{rob}\{\mathfrak{E}\} = \sum_{i,j,k=1}^{n} \mathbb{P}\text{rob}\{\mathfrak{E}|e(\mathbf{z}_{RD}^n) = i, e(\mathbf{z}_{SR}^n) = j, e(\mathbf{z}_{SD}^n) = k\}\mathbb{P}\text{rob}\{e(\mathbf{z}_{RD}^n) = i, e(\mathbf{z}_{SR}^n) = j, e(\mathbf{z}_{SD}^n) = k\} \tag{2}$$

However as we have assume that losses are uncorrelated between channels we have $\mathbb{P}\text{rob}\{e(\mathbf{z}_{RD}^n) = i, e(\mathbf{z}_{SR}^n) = j, e(\mathbf{z}_{SD}^n) = k\} = \mathbb{P}\text{rob}\{e(\mathbf{z}_{RD}^n) = i\}\mathbb{P}\text{rob}\{e(\mathbf{z}_{SR}^n) = j\}\mathbb{P}\text{rob}\{e(\mathbf{z}_{SD}^n) = k\}$.

Now let's work on the term $\mathbb{P}\text{rob}\{\mathfrak{E}|e(\mathbf{z}_{RD}^n)=i, e(\mathbf{z}_{SR}^n)=j, e(\mathbf{z}_{SD}^n)=k\}$. Let's define a function $c:\{0,1\}^n\times\{0,1\}^n \to \{1,\ldots,n\}$ that returns the number of positions in two erasure patterns where an erasure occured simultaneously. We will have $\max\{0,((j+k)-n\} \leq c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n) \leq \min\{j,k\}$. We can therefore rewrite $\mathbb{P}\text{rob}\{\mathfrak{E}|\mathbf{z}_{RD}^n,\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n\}$ as :

$$\mathbb{P}\text{rob}\{\mathfrak{E}|e(\mathbf{z}_{RD}^n)=i, e(\mathbf{z}_{SR}^n)=j, e(\mathbf{z}_{SD}^n)=k\} = \sum_{l=0}^{n} \mathbb{P}\text{rob}\{\mathfrak{E}|i,j,k,l\}\mathbb{P}\text{rob}\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l|i,j,k\}$$

Now we have to derive

$$\mathbb{P}\text{rob}\{\mathfrak{E}|i,j,k,l\} = \mathbb{P}\text{rob}\{\mathfrak{E}|e(\mathbf{z}_{RD}^n)=i, e(\mathbf{z}_{SR}^n)=j, e(\mathbf{z}_{SD}^n)=k, c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l\}$$

As we are assuming that j erasures have occured on source to relay link there is a maximum of $\min\{|\mathcal{X}|^{(n-j)}, |\mathcal{X}|^{nR}\}$ possible values for $\mathbf{y}_{SR}^n$ compatible with $\mathbf{z}_{SD}^n$. Based on theorem 3 it might be possible to construct a coding scheme wih vanishing probability of decoding error to transfer from relay to destination the received sequence $\mathbf{y}_{SR}^n$ at relay if $\min\{|\mathcal{X}|^{(n-j)}, |\mathcal{X}|^{nR}\} < |\mathcal{X}|^{n(1-p_{RD})}$. Let's assume that such a coding scheme exists. As this assumption is simplifying the activity of the decoder at destination (knowing the received sequence at relay is always better than not knowing it), it will give us a lower bound on the decoding error probability. Now if we are able to decode at destination the received sequence at relay $\mathbf{y}_{SR}^n$, this means that the set $\mathcal{B}(\mathcal{C}_S,\mathcal{C}_R,\mathbf{y}_{RD}^n)$ will contain a single set $\mathcal{A}(\mathcal{C}_S,\mathbf{y}_{SR}^n)$. Now we have assume that there is l erasures in common between $\mathbf{z}_{SR}^n$ and $\mathbf{z}_{SD}^n$, meaning that the the member of the intersection set $\mathcal{A}(\mathcal{C}_S,\mathbf{y}_{SD}^n)\bigcap\mathcal{B}(\mathcal{C}_S,\mathcal{C}_R,\mathbf{y}_{RD}^n)$ will have in common $n-l$ symbols. However with l common erasures we can define in absolute terms, a maximum of $|\mathcal{X}|^{(n-l)}$ disjoint agreement set relative to different choice over the symbols in the $n-l$ received symbols. Now if the number of codewords in the code ($|\mathcal{X}|^{nR}$) is larger than this value, by pigeon hole priciple, we will have at least one of the agreement sets that will have more than 1 members leading to a possible decoding error for at least $|\mathcal{X}|^{nR}-|\mathcal{X}|^{(n-l)}+1$ codewords. This will surely happen if $l > n(1-R)$. Under this condition the error probability will be at least $\frac{1}{2}\frac{|\mathcal{X}|^{nR}-|\mathcal{X}|^{(n-l)}}{|\mathcal{X}|^{nR}}$. From this point the proof follows the same way as theorem 3.

Based on the above argument we can define a lower bound for the decoding error probability as :

$$\mathbb{P}\text{rob}\{\mathfrak{E}|i,j,k\} \geq \sum_{l=n(1-R)+1}^{n} \frac{1}{2}\frac{|\mathcal{X}|^{nR}-|\mathcal{X}|^{(n-l)}}{|\mathcal{X}|^{nR}}\mathbb{P}\text{rob}\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l|i,j,k\}$$

By replacing this term in equation 3, we have:

$$\begin{aligned}
\mathbb{P}\text{rob}\{\mathfrak{E}\} &\geq \frac{1}{2}\sum_{l=n(1-R)+1}^{n}(1-|\mathcal{X}|^{-(l-n(1-R))})\mathbb{P}\text{rob}\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l\} \\
&\geq \frac{1}{2}\left(\mathbb{P}\text{rob}\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)\geq n(1-R)+1\} - \sum_{l=n(1-R)+1}^{n}|\mathcal{X}|^{-(l-n(1-R))})\mathbb{P}\text{rob}\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l\}\right)
\end{aligned}$$

We can analyse the asymptotic behaviour of the negative term by using Jensen inequality :

$$\sum_{l=0}^{n}|\mathcal{X}|^{-l}\mathbb{P}\text{rob}\big\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l\big\}\geq|\mathcal{X}|^{-\bar{c}(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)}$$

This leads to the following upper bound for the second term :

$$-\sum_{l=0}^{n}(|\mathcal{X}|^{-(l-n(1-R))})\mathbb{P}\text{rob}\big\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)=l\big\}\leq-|\mathcal{X}|^{-n(R-(1-\frac{\bar{c}(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)}{n}))}$$

This means that the second term converges to 0 when $n\rightarrow\infty$ with a speed controlled by $\left(1-\frac{\bar{c}(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)}{n}\right)$, meaning that asymptocally the lower bound on decoding error probability goes to :

$$\lim_{n\rightarrow\infty}\mathbb{P}\text{rob}\big\{\mathfrak{E}\big\}\geq\frac{1}{2}\left(\mathbb{P}\text{rob}\big\{c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)\geq n(1-R)+1\big\}\right)$$

Now if the channel follows a large deviation principle with rate function $r(.)$ for proportion of joint error on source-relay and source-destination we will have

$$\mathbb{P}\text{rob}\big\{\frac{1}{N}c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)>\mathbb{E}\big\{Z_{SR}Z_{SD}\big\}+\epsilon\big\}\leq e^{Nr(\epsilon)}$$
$$\mathbb{P}\text{rob}\big\{\frac{1}{N}c(\mathbf{z}_{SR}^n,\mathbf{z}_{SD}^n)>\mathbb{E}\big\{Z_{SR}Z_{SD}\big\}-\epsilon\big\}\geq 1-e^{Nr(\epsilon)}$$

Under this condition the channel accept a strong converse and for $R<1-\mathbb{E}\big\{Z_{SR}Z_{SD}\big\}$ the lower bound on the decoding error probability goes to 0 exponentially fast with a speed that is a function of $C-R$, and for $R>\mathbb{E}\big\{Z_{SR}Z_{SD}\big\}$ the decoding error probability goes to 1 with a speed that is a function of $R-C$, resulting in the strong converse for this situation where $(1-p_{SR})<(1-p_{RD})$. ■

We proved in the above lemma the converse for erasure relay channels where $(1-p_{SR})>(1-p_{RD})$. To fully characterize the erasure relay channel we have to deal with cases where $(1-p_{SR})>(1-p_{RD})$. This is treated in following lemma.

Lemma 5: If $(1-p_{SR})>(1-p_{RD})$, no coding scheme for erasure relay channel with rate $R>\max\{(1-p_{SR}),(1-p_{SD})\}$ can transmit with a vanishing decoding error probability. Moroever if there exists large deviation principles for the empirical mean of processes $\{Z_{SR}\},\{Z_{SD}\}$ and $\{Z_{RD}\}$ this converse is a strong converse.

Proof: We will first prove the converse for the case where $(1-p_{SR})>(1-p_{SD})$ and $(1-p_{SR})>(1-p_{RD})$. We will prove than under these hypothesis no code with rate $R>(1-p_{SR})$ can achieve a vanishing decoding error probability. We have proven in lemma 3 that we can restrict ourself to the the decoding function ψ_R^* and ψ_D^* as any other decoding function will lead to higher decoding error probability.

Similarly, to the previous lemma, case II and III are equivalent to point to point channels and we have proven in theorem 3 that one cannot send with a rate larger than $(1-p_{SD})$ for case II and a rate larger than $\min\{(1-p_{RD}),(1-p_{SR})\}$ for case III. Because of the hypothesis for case III we cannot send with a rate larger than $(1-p_{RD})$. This means that if the used coding scheme falls into case II or III it will verify the lemma. It remains to prove the lemma for case I.

Let's assume a relay code $(\mathcal{C}_S, \mathcal{C}_R, \phi_{RD})$ of length n and rate (R, R') with a minimal distance couple $(d_{\min}(\mathcal{C}_S), d_{\min}(\mathcal{C}_R))$. The code $\mathcal{C}_S$ contains $|\mathcal{X}|^{nR}$ codewords and the code $\mathcal{C}_R$ contains $|\mathcal{X}|^{nR'}$ codewords. Applying the generalization of the pigeon hole principle to the ambiguity set of each codeword $\mathbf{x}_R^n$, $\mathcal{B}^*(\mathbf{x}_R^n)$, we can state that at least the ambiguity set one of the codeword of $\mathcal{C}_R$ must hold no fewer than $\lceil |\mathcal{X}|^{n(R-R')}\rceil$ codewords. We are interested in the largest ambiguity set as errors are more likely to happen with messages leading to the transmission by relay of the codeword relative to this set. It is noteworthy that over all choice of erasure relay codes the minimum size of the largest ambiguity set is equal to $\lceil |\mathcal{X}|^{n(R-R')}\rceil$, and this happens when all ambiguity sets have the same number of elements. Best performances are obtained when the size of the ambiguity set is minimized, *i.e.* the maximum size of the ambiguity set is equal to $\lceil |\mathcal{X}|^{n(R-R')}\rceil$. In the forthcoming we will assume that a codeword $\mathbf{x}_S^n$ have been sent by the source and that leads to the transmission by relay of the codeword relative to an ambiguity set of size $\lceil |\mathcal{X}|^{n(R-R')}\rceil$ is sent. To simplify the notation we will assume that the size of the ambiguity set is equal to $|\mathcal{X}|^{n(R-R')}$.

We will denote as $\mathfrak{E}$ the decoding error event , *i.e.* the event that the set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SD}^n)\bigcap\mathcal{D}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ contains more than a single codeword . Assuming that the source and relay codes are fixed, one can decompose the decoding error probability as :

$$\mathbb{P}\text{rob}\big\{\mathfrak{E}\big\} = \sum_{\mathbf{x}_S^n, \mathbf{y}_{SR}^n, \mathbf{y}_{SD}^n, \mathbf{y}_{RD}^n}$$

We are also partionning the set of erasure patterns $\{0,1\}^n$ into n subsets $\mathcal{Z}_k,\ k=0,\dots,n$ where we define $\mathcal{Z}_k = \{\mathbf{z}^n \in \{0,1\}^n | \sum_{i=1}^n z_i = k\}$, *i.e.* $\mathcal{Z}_k$ is the set of all erasure pattern containing k erasures. Now we have :

$$\mathbb{P}\text{rob}\big\{\mathfrak{E}\big\} = \sum_{i,j,k=1}^{n} \mathbb{P}\text{rob}\big\{\mathfrak{E}|e(\mathbf{z}_{RD}^n)=i, e(\mathbf{z}_{SR}^n)=j, e(\mathbf{z}_{SD}^n)=k\big\}\mathbb{P}\text{rob}\big\{e(\mathbf{z}_{RD}^n)=i, e(\mathbf{z}_{SR}^n)=j, e(\mathbf{z}_{SD}^n)=k\big\} \tag{3}$$

However as we have assume that losses are uncorrelated between channels we have $\mathbb{P}\text{rob}\big\{e(\mathbf{z}_{RD}^n) = i, e(\mathbf{z}_{SR}^n)=j, e(\mathbf{z}_{SD}^n)=k\big\} = \mathbb{P}\text{rob}\big\{e(\mathbf{z}_{RD}^n)=i\big\}\mathbb{P}\text{rob}\big\{e(\mathbf{z}_{SR}^n)=j\big\}\mathbb{P}\text{rob}\big\{e(\mathbf{z}_{SD}^n)=k\big\}$.

Based on the hypothesis that $(1-p_{SR}) > (1-p_{RD})$ and because of theorem 3, we know that we cannot transmitting with a vanishing error probability the ambiguity existing at relay to destination (as the number of agreement sets at relay will at least $|\mathcal{X}|^{n(1-R)} > |\mathcal{X}|^{n(1-p_{SR})} > |\mathcal{X}|^{n(1-p_{RD})}$). This means that at least one of the sets $\mathcal{B}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ will contain more than the single agreement set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n)$. Let's assume that $\mathcal{B}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n)$ contains also the agreement set $\mathcal{A}(\mathcal{C}_S, \mathbf{u}^n)$ relative to another sequence $\mathbf{u}^n$ with the same erasure pattern as $\mathbf{y}_{SR}^n$, *i.e.* $\mathcal{B}(\mathcal{C}_S, \mathcal{C}_R, \mathbf{y}_{RD}^n) = \mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n) \bigcup \mathcal{A}(\mathcal{C}_S, \mathbf{u}^n)$.

and their will exist at least a received sequence $\mathbf{y}_{SR}^n$ such that its agreeement set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n)$

Let's assume that a relay coding scheme is transmitting with a rate $R > (1-p_{SR})$. Because of theorem 3, we know that we cannot transmitting with a vanishing error probability and their will exist at least a received sequence $\mathbf{y}_{SR}^n$ such that its agreeement set $\mathcal{A}(\mathcal{C}_S, \mathbf{y}_{SR}^n)$ ■

Performance of a specific coding scheme will depend on the properties of codewords chosen in each one of cosets $\mathcal{C}_S$ and $\mathcal{C}_R$. Since a codeword set with a minimal distance $d_{\min}$ can correct up to $d_{\min}-1$ erasures, it is desirable to increase the minimal distance in the codeword sets as possible. However, by Singleton bound the minimum distance of a codeword set containing $2^{n\mathbf{R}}$ symbols is bounded by $d_{\min} < n(1-\mathbf{R})+1$. MDS codes attains this bound and are therefore very suitable for erasure channels as they are sphere packing codes for the erasure channels. As a consequence, if a MDS code exists no coding scheme has a better performance than this code. As we wish to provide reverse bound we will assume in the forthcoming that the codeword sets $\mathcal{C}_S$ and $\mathcal{C}_R$ are MDS.

The coset $\psi_{SD}(\mathbf{y}_{SD}^n)$ has a simple structure. It contains all codeword that are compatible with a received sequence $\mathbf{Y}_{SD}^n$. Under typical erasure pattern, with a probability higher than $1-\epsilon$, $n(1-p_{SD}-\delta(n))$ symbols will be received in $\mathbf{Y}_{SD}^n$ at the destination[1]. The coset $\psi_{SD}(\mathbf{y}_{SD}^n)$ contains all codewords in $\mathcal{C}_S$ that are in agreement with this received sequence of symbols. Since the initial codewords set $\mathcal{C}_S$ is chosen to be an MDS set the number of codewords in $\psi_{SD}(\mathbf{y}_{SD}^n)$ is equivalent to $2^{n\mathbf{R}-n(1-p_{SD}-\delta(n))}$. Moreover, by Singleton bound, the largest minimal distance in $\psi_{SD}(\mathbf{y}_{SD}^n)$ is bounded by $d_{\min}(\psi_{SD}(\mathbf{y}_{SD}^n)) \leq n(1-\mathbf{R})+1$.

Analysis of the coset $\psi_{RD}(\mathbf{y}_{RD}^n)$ is more challenging. As illustrated in the previous section, the coset $\lambda_D(\mathbf{y}_{RD}^n)$ contains all sequences in $\mathcal{C}_R$ that are in agreement with the symbols received in $\mathbf{Y}_{RD}^n$. Since $\mathcal{C}_R$ is chosen to be an MDS code $|\lambda_D(\mathbf{y}_{RD}^n)| = \min\{1, 2^{n\mathbf{R}'-n(1-p_{RD}-\delta(n))}\}$. Moreover if $(1-p_{SR}) < (1-p_{RD})$ each codeword in $\mathcal{C}_R$ can be related to only one coset $\psi_R(\mathbf{y}_{SR}^n)$; while if $(1-p_{SR}) > (1-p_{RD})$ each codeword in $\mathcal{C}_R$ should be related to $\min\{1, 2^{n\mathbf{R}-n\mathbf{R}'}\}$ cosets $\psi_R(\mathbf{y}_{SR}^n)$. Therefore

[1]In the forthcoming we will use $\delta(n)$ as the generic term with the property that $\lim_{n\to\infty}\delta(n)=0$

$\psi_{RD}(\mathbf{y}_{RD}^n)$ contains a single coset $\psi_R(\mathbf{y}_{SR}^n)$ if $(1-p_{SR}) < (1-p_{RD})$ and $2^{n\mathbf{R}-n(1-p_{RD}-\delta(n))}$ cosets $\psi_R(\mathbf{y}_{SR}^n)$ if $(1-p_{SR}) > (1-p_{RD})$ (see equation (**??**)). It is also worthy noting that for transmission rate $\mathbf{R} < (1-p_{SR})$ the coset $\psi_R(\mathbf{y}_{SR}^n)$ consists of a single point and so the message sent by the source can be decoded at the relay. However if the transmission rate $\mathbf{R} > (1-p_{SR})$ the coset $\psi_R(\mathbf{y}_{SR}^n)$ will contains $2^{n\mathbf{R}-n(1-p_{SR}-\delta(n))}$ codewords that are all compatible with $\mathbf{Y}_{SR}^n$.

The following theorem gives a reverse coding bound for the relay channel.

Theorem 7—Converse theorem for relay erasure channel: No deterministic erasure coding scheme can exceed the following bound over an erasure relay channel :

$$\begin{cases} \mathbf{R} & < 1-p_{SD}p_{SR}, \qquad\qquad if \;\; (1-p_{SR}) \leq (1-p_{RD}) \\ \mathbf{R} & < \max\{T,(1-p_{SD})\}, \quad if \;\; (1-p_{SR}) > (1-p_{RD}) \end{cases}$$

with $T = \min\{(1-p_{SR}),(1-p_{SD})+(1-p_{RD})\}$

□

Proof: The proof proceeds by using the list decoding interpretation for erasure coding schemes and using properties of the decoding set $\psi_{RD}(\mathbf{y}_{RD}^n)$ under different situations. The keystone argument is based on the universality of the list decoding based joint decoding and the fact that that no code can have a better performance in term of probability of decoding error than an MDS code

- $(1-p_{SR}) < (1-p_{RD})$:
  Under this hypothesis $\psi_{RD}(\mathbf{y}_{RD}^n)$ contains a single coset but with $\min\{2^{n\mathbf{R}-n(1-p_{SR}-\delta(n))},1\}$ codewords inside (for $\mathbf{R} < (1-p_{SR})$ the coset contains only one point). The intersection $\psi_{SD}(\mathbf{y}_{SD}^n)\bigcap\psi_{RD}(\mathbf{y}_{RD}^n)$ is a set composed of all codewords compatible with $\mathbf{Y}_{SD}^n$ and $\mathbf{Y}_{SR}^n$. These codewords have, under typical erasure pattern and without any spatial correlation in erasures, $n(1-p_{SD}p_{SR}-\delta(n))$ symbols in common. Therefore, to ensure that there is a single codeword in the intersection we should have $d_{\min}(\psi_{SD}(\mathbf{y}_{SD}^n)) > n-n(1-p_{SD}p_{SR}-\delta(n))$, *i.e.* $\mathbf{R} < (1-p_{SD}p_{SR})-\delta(n)$.
- $(1-p_{SR}) > (1-p_{RD})$ and $\mathbf{R} < (1-p_{SR})$:
  Under this hypothesis $\psi_{RD}(\mathbf{y}_{RD}^n)$ contains $2^{n\mathbf{R}-n(1-p_{RD}-\delta(n))}$ cosets with a single point inside. The codewords are not constrained to be compatible with each other. The maximal minimum distance in the set $\psi_{RD}(\mathbf{y}_{RD}^n)$ is therefore $d_{\min}(\psi_{RD}(\mathbf{y}_{RD}^n)) = n-(n\mathbf{R}-n(1-p_{RD}))+1$. To ensure that under such extreme condition there is a single codeword in the intersection we should have $d_{\min}(\psi_{RD}(\mathbf{y}_{RD}^n)) > n-n(1-p_{SD}-\delta(n))$, *i.e.* $\mathbf{R} < (1-p_{SD})+(1-p_{RD})-\delta(n)$.
- $(1-p_{SR}) > (1-p_{RD})$ and $\mathbf{R} > (1-p_{SR})$:
  Under this hypothesis $\psi_{RD}(\mathbf{y}_{RD}^n)$ contains several cosets with $2^{n\mathbf{R}-n(1-p_{SR}-\delta(n))}$ codewords in

each. These cosets might have intersection or being separated. For this case we should avoid being in the situation depicted in figure **??**(b) where ambiguity will remains between the correct codeword in $\psi_R(\mathbf{y}_{SR}^n)$ and another codeword not compatible with $\mathbf{Y}_{SR}^n$ but still in $\psi_{SD}(\mathbf{y}_{SD}^n) \bigcap \psi_{RD}(\mathbf{y}_{RD}^n)$. This incorrect codeword is in agreement with the $n(1-p_{SD}-\delta(n))$ symbols received in $\mathbf{y}_{SD}^n$ but is not in agreement with non of the $n(p_{SD}(1-p_{SR})-\delta(n))$ symbols received at the relay but not at the destination (under typical erasure patterns).

If the source S is sending at a rate $\mathbf{R} > (1-p_{SD})$, the maximal minimum distance of $\mathcal{C}_S$ will be smaller than $np_{SD}+1$. The set $\phi_{RD}^{-1}(\lambda^{-1}(\mathbf{y}_{RD}^n)) - \psi_R(\mathbf{y}_{SR}^n)$ has also a minimal distance smaller than $np_{SD}+\delta(n)$. This comes from the fact that $\phi_{RD}^{-1}(\lambda^{-1}(\mathbf{y}_{RD}^n))$ contains at least another coset $\psi_R(\mathbf{z}^n)$ and it was shown that the minimal distance in any coset is smaller than $n(1-\mathbf{R})$. Meaning that it is impossible to design a set $\phi_{RD}^{-1}(\mathbf{x}^n) - \psi_R(\mathbf{y}_{SR}^n)$ where one can guarantee that for all possible reception patterns containing $n(1-p_{SD}-\delta(n))$ received symbols there is no codeword in $\psi_{SD}(\mathbf{y}_R^n)$ since the minimal distance in $\phi_{RD}^{-1}(\mathbf{x}^n)$ is smaller than np_{SD}. There exists therefore a typical erasure pattern containing $n(1-p_{SD}-\delta(n))$ received symbols such that there is at least one codewords in $\phi_{RD}^{-1}(\lambda^{-1}(\mathbf{y}_{RD}^n)) - \psi_R(\mathbf{y}_{SR}^n)$ that is compatible with $\mathbf{y}_{SD}^n$. Pointing out the fact that whenever transmission rate goes higher than $(1-p_{SD})$ and the set $\phi_{RD}^{-1}(\lambda^{-1}(\mathbf{y}_{RD}^n))$ contains more than a single coset with more than one codeword inside we will be in situation depicted in figure **??**(b) with a probability larger than 0. In summary, it is impossible to transfer reliably over an erasure relay channel using a deterministic coding scheme with a rate higher than $(1-p_{SD})$ if $(1-p_{SR}) > (1-p_{RD})$ and $\mathbf{R} > (1-p_{SR})$.

Therefore the proof is completed. ■

□

Remarks

- The physically degraded situation is the case that all symbols received at the destination are also received at the relay. Under this condition $\psi_{SD}(\mathbf{y}_{SD}^n) \subseteq \psi_R(\mathbf{y}_{SR}^n)$ and the joint decoding process at the receiver is successful if $\psi_R(\mathbf{y}_{SR}^n)$ contains a single codeword, *i.e.* when $\mathbf{R} < (1-p_{SR})$. The converse bound for physically degraded channel conditions is therefore equivalent to $\mathbf{R} < \min\{(1-p_{SR}), (1-p_{SD})+(1-p_{RD})\}$. In [12] a milder situation, in which supersedes the physical degraded condition and is enough to make the cut-set bound tight, is defined as "*virtually degraded*" condition. Under this condition $\psi_R(\mathbf{y}_{SR}^n)$ contains only a single codeword and therefore the relay is able to decode the message sent by the source. The converse bound for virtually degraded channel has the same form that the bound under physically degraded situation.

- The proof of the converse shows three types of collaboration for the relay node. The first type of collaboration "*active collaboration*" is possible if the channel is in the degraded (virtually of physically) situation in which the relay can decode the message sent by the source. Active collaboration of relay consists of arranging points in $\phi_{RD}^{-1}(\mathbf{x}^n)$ such that it makes an MDS set. This type of collaboration is possible when $\mathbf{R} < (1 - p_{SR})$. In this case a full cooperation, in the sense described in [13], is possible between the sender and the relay. The second type of collaboration is the "*no collaboration*" that occurs when $(1-p_{SR}) > (1-p_{RD})$ and $\mathbf{R} > (1-p_{SR})$. Under this setting the relay cannot ensure any fruitful collaboration for sending information to the final destination and it is useless to forward any information. This is a milder condition supersedes the reversely degraded channel situation described in [13]. The last type of collaboration is the "*passive collaboration*" in which the relay cannot decode the message sent by the source, but nevertheless can forward all its received symbols to the destination without rearranging the points in $\phi_{RD}^{-1}(\mathbf{x}^n)$. This type of collaboration occurs when $(1 - p_{SR}) < (1 - p_{RD})$ and $\mathbf{R} > (1 - p_{SR})$. Therefore since a perfect cooperation is not accessible; however, the alternative information sent by the relay is not zero and thus facilitate the communication.
- This theorem shows that the well-known and classical cut-set bound is not achievable when $p_{SD}(1-p_{SR}) < (1 - p_{RD}) < (1 - p_{SR})^2$. A precise examination of the proof shows that if the decoding set $\phi_{RD}^{-1}(\mathbf{x}^n)$ contains more than one coset information coming from the relay might be non useful for decoding purposes. Two solutions have been proposed to overcome this shortage. One solution proposed in [3] uses side information in the form of the erasure pattern over the source-relay channel. This side information could be used when the decoding set $\psi_{RD}(\mathbf{y}_{RD}^n)$ is designed to contain different cosets relative to different erasure patterns over the source to relay channel, *i.e.* for every erasure pattern there is only one coset in $\psi_{RD}(\mathbf{y}_{RD}^n)$ compatible with this pattern. We therefore fall back to the situation depicted in Fig. **??**(a) and the cut-set bound can be one more time attained. Nevertheless, the amount of extra information needed to transfer the side information should be assessed. A simple evaluation shows that one needs $n(1 - h(p_{SR}))$ bits of extra information to transfer the side information which is the erasure pattern over the sender to relay channel. Accounting the information rate needed for side information the scheme proposed in [3] attains a proportion

[2] As shown in the proof if $\mathbf{R} < (1-p_{SR})$ and if $(1-p_{SR}) > (1-p_{RD})$ the joint decoding at the receiver can be successfully done if $\mathbf{R} < (1 - p_{RD}) + (1 - p_{SD})$. Therefore if $(1 - p_{SR}) > (1 - p_{RD}) + (1 - p_{SD})$ then the cut-set bound is attainable since $\mathbf{R}_{cut-set} = (1 - p_{RD}) + (1 - p_{SD})$

$\frac{\mathbf{R}\log(|\mathcal{X}|)}{\mathbf{R}\log(|\mathcal{X}|)+(1-h(p_{SR}))}$ of the cut-set bound.

- Another solution to expand the converse bound is proposed in [4]. Classically in information theory, one chooses a random code at the beginning of the communication and informs the receiver about the used coding scheme. The used coding scheme might be seen as side information that is given at the beginning of the communication. However, in [4], the use of a randomly changing coding scheme is proposed, where the relay node chooses randomly at each transmission how to mix the received symbols and the decoding set $\phi_{RD}^{-1}(\mathbf{x}^n)$ is not fixed (as in classical settings) but changes randomly during the transmission. By averaging over erasure channel as well as random code statistics, the probability that $\psi_{RD}(\mathbf{y}_{RD}^n)$ becomes an MDS set goes to 1. A naive evaluation of the amount of extra information needed to transfer the side information leads to $\mathcal{O}(n\mathbf{R})$ bits per transferred packet (this information placed in the packet headers). By making the packet lengths enough large the over head fraction per each packet can be made arbitrarily small. Moreover, the error exponent presented in [4] shows that this over head is not significant for practical error bounds and when the transmission links are not so lossy. However the overhead will grow if the transmission links get lossier, since to achieve the same error probability a larger block size (larger $n\mathbf{R}$) is needed.

In the forthcoming we propose a linear deterministic coding schemes achieving the capacity region bounded by theorem 7. This code can solve the joint decoding problem described previously by solving a linear system of equation with $n\mathbf{R}$ variables with a complexity $\mathcal{O}(n\log(n))$. Almost-MDS [6] codes with linear decoding complexity are also applicable in this context. This code does not need any type of side information and just knowing the average packet loss over the transmission links is sufficient for constructing the code. We describe this scheme for the degraded situation, where the relay is a full collaborated node, as well as the non-degraded situation.

D. Achievability Scheme

In this section we first show that by using a decoding and forwarding scheme the capacity of the active collaborating (physically and virtually degraded) situation and the non collaborating situation, in which relaying is useless, are attainable. We will then propose an estimating and forwarding coding scheme which attains the capacity of the channel in the passive collaborating situation.

1) Decoding and Forwarding: Let suppose we have designed an $(n, n\mathbf{R})$ MDS code with an encoding matrix $\mathcal{C} = \left[I_{n\mathbf{R}\times n\mathbf{R}}|A_{n\mathbf{R}\times(n-n\mathbf{R})}|B_{n\mathbf{R}\times l}\right]$. At the sender we encode the message symbols through the $(n, n\mathbf{R})$ MDS code with encoding matrix $\left[I_{n\mathbf{R}\times n\mathbf{R}}|A_{n\mathbf{R}\times(n-n\mathbf{R})}\right]$ leading to $(n - n\mathbf{R})$ redundant packets. These redundant packets will help the receiver and the relay to retrieve the erased packets over the channel. Under the condition that $\mathbf{R} < (1 - p_{SR})$ a MDS code will, asymptotically with large n, ensure perfect communication between the sender and relay as the MDS codes achieve the capacity of the erasure channels. Such a code will therefore validate the conditions of the degraded situation.

Following the coding structure proposed in [13] the relay should transfer only side information (indices) reducing the ambiguity at the receiver. Every packets of an MDS code can be used as an index that will reduce the ambiguity about the initial message. We therefore generate at the relay some redundant packets coming from multiplying the message symbols by the encoding matrix $B_{n\mathbf{R}\times l}$. Using this coding scheme relay only send useful index information to the receiver and the fact that $\mathcal{C}$ is the generator of an MDS code guarantee that every packet received from the relay will reduce the ambiguity about the message symbols sent by the sender.

At the receiver side we will, asymptotically with large n, receive around $n(1 - p_{SD})$ packets from the sender and $l(1 - p_{RD})$ packets from the relay. The MDS code with encoding matrix $\mathcal{C}$ can be decoded at the receiver if $n(1 - p_{SD}) + l(1 - p_{RD}) > n\mathbf{R}$. Let's denote by $\alpha = \frac{l}{n}$ as the relay load sharing of relay node ($0 \leq \alpha \leq 1$). The rate $\mathbf{R} \leq (1 - p_{SD}) + \alpha(1 - p_{RD})$ is therefore achievable with an error probability approaching to zero where n and l are arbitrary large. This comes from the fact that $\mathcal{C}$ is an MDS code and as shown in section II the error decoding probability of an MDS code goes to zero if the number of erasures is lower than the minimal distance of the code. It is worthy to say that if l is chosen equivalent to 0, the relay is bypassed and we fall back into the point to point situation with the achievability rate $\mathbf{R} \leq (1 - p_{SD})$.

Therefore the capacity of the degraded situation and the case that the relaying is useless, where $(1 - p_{SR}) > (1 - p_{RD})$ and $(1 - p_{SD}) > (1 - p_{SR})$, is achievable by the proposed scheme.

2) Estimating and Forwarding: Here we present coding schemes applicable without the degraded assumption. The main difference between this situation and the degraded case is that in the non-degraded situation the receiver might have some information that have not been received by the relay, where under the degraded assumption all information available at receiver are also available at the relay.

Let's suppose that the sender use a $(n, n\mathbf{R})$ MDS code $\mathcal{C} = \left[I_{n\mathbf{R}\times n\mathbf{R}} | A_{n\mathbf{R}\times(n-n\mathbf{R})}\right]$. Under the general erasure relay channel scenario defined previously if an asymptotically large number n of packets is sent by the sender a number $n(1-p_{SD})$ (resp. $n(1-p_{SR})$) packets are received at receiver (resp. relay). Out of these packets $n(1 - p_{SD})(1 - p_{SR})$ are received at both the relay and sender and $np_{SR}(1 - p_{SD})$ (resp. $np(1 - p_{SR})$) only at the receiver (resp. the relay). The relay has to forward a sufficient number of packets only received by it to eliminate ambiguity at receiver. However, the relay and receiver are not aware of which packets they respectively received. The solution consists of forwarding all received packets at the relay to the receiver, *i.e.* the relay assumes that the $n(1-p_{SR})$ received packets are the information packets.

The coding scheme complete $K = n(1-p_{SR})$ packets received at relay with $n-K$ redundant packets obtained using an MDS code of rate $(1-p_{SR})$. Since the capacity of the relay to the receiver channel is $1-p_{RD}$ and when $(1-p_{SR}) < (1-p_{RD})$, these K packets will be decoded at receiver with an error decoding probability which goes to zero if n is sufficiently large. Out of these packets, Kp_{SD} are packets that have not been received by the receiver through the direct link. Accomplishing that $n(1-p_{SD}) + Kp_{SD}$ packets out of n packets sent by sender are received at the receiver. If $n(1-p_{SD}) + Kp_{SD} > n\mathbf{R}$ the MDS code with encoding matrix $\mathcal{C}$ can be decoded at the receiver. Therefore, asymptotically with large n, a rate equal to $\mathbf{R} = (1-p_{SD}p_{SR})$ is achieved by this coding scheme with an error probability approaching to zero. This shows that the region bounded by theorem 7 is in fact the capacity of the erasure relay channel.

V. CONCLUSION

We derived the classical cut-set bound for the single-sender single-relay erasure channel and we showed that this bound is not achievable in general. The main reason is that attaining the bound supposes that a full level of collaboration exists between the sender and the relay in the channel. It might be happen only if the channel is in the degraded situation, while under some scenarios the full collaboration is not possible. We defined three types of the node collaboration in the channel : active, passive and without collaboration. Based on this precise analysis we are able to find the optimum relaying scheme in the channel which leads us to a strong coding converse bound for this simple scenario.

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseles coding of correlated information sources," *IEEE. Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.

[2] K. Salamatian and R. Khalili, "An information theory for erasure channels (invited paper)," *Allerton*, 2005.

[3] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *submitted to Trans. Inform. Theory*.

[4] D. S. Lun, M. Mdard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," *submitted to Trans. Inform. Theory*.

[5] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer, "Some intersection theorem for ordered set and graphs," *Journ. of Comb. Theory*, vol. 43, no. Series A, no. Series A, pp. 23–37, 1986.

[6] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," pp. 150–159, 1997.

[7] S.Boucheron and M.R.Salamatian, "About priority encoding transmission," *IEEE. Trans. Inform. Theory*, vol. on processing.

[8] J. Korner and K. Marton, "General broadcast channels with degraded messages sets," *IEEE. Trans. Inform. Theory*, vol. IT-23, pp. 60–64, 1977.

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

[10] I. Csiszar and J. Korner, "Broadcast channel with confidential messages," *IEEE. Trans. Inform. Theory*, vol. IT-24, pp. 339–348, May 1978.

[11] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic press, 1981.

[12] R. Khalili and K. Salamatian, "On the achievability of cut-set bound for a class of erasure relay channels: The non-degraded case," *ISITA*, 2004.

[13] T.M.Cover and A. E. Gamal, "Capacity theorems for relay channels," *IEEE. Trans. Inform. Theory*, vol. IT-25, pp. 572–584, Sep 1979.