Capacity Region of the Permutation Channel

John MacLaren Walsh and Steven Weber

Abstract— **We discuss the capacity region of a degraded broadcast channel (DBC) formed from a channel that randomly permutes input packets by selecting a permutation according to a probability distribution. Starting from the known capacity region expression for the DBC, we give an explicit outer and inner bound to the capacity region which are shown to be equal for the cases of 2 and 3 packets. The work extends previous results which considered the case where the permutation was selected uniformly from the set of all permutations. The results are useful in determining fundamental rate delay tradeoffs when transmitting temporally ordered content over multipath routed networks.**

I. INTRODUCTION

Consider a point to point channel that randomly permutes a set of packets applied at the input, which we call a *permutation channel*. In particular, the input is an ordered set of M different K-bit packets, $(x_1, \ldots, x_M)$, and the output is the packets in permuted order, $(x_{\pi(1)}, \ldots, x_{\pi(M)})$ (the bits in the packets are unaffected by the channel). The likelihood of each possible permutation is specified by a probability distribution on all possible $M!$ permutations. The M packet arrival instants at the destination correspond to M receivers in a *degraded broadcast channel* (DBC). In particular, we consider the channel to have M outputs, with the mth ouput consisting of the all of the packets that have arrived until just after the arrival of the mth received packet at the sink. Although not necessary, it is convenient, and in our opinion natural, to assume the packets are *labeled* from 1 to M. Note that the total overhead of these labels, $\lceil \log_2 M \rceil$ bits per packet, is amortized through the use of packets of long length, K. The channel is clearly degraded as the information available after m packet arrivals is a strict subset of the information available after $m + 1$ arrivals. In this paper, we seek a closed form description for the capacity region of this DBC.

A. Motivation: Delay Mitigating Codes for Multipath Routed Networks

A major motivation for studying the permutation channel is the problem of using coding to mitigate delay in multipath routed networks. In order to make the correspondence between a simple model for this problem and the permutation channel, consider a network in which a source wishes to communicate with a sink, and has a diversity of M independent network paths through which it can communicate with the sink. Because there are other packets in the queues along these paths from other independent flows, M packets transmitted at the same time from the source along the M independent paths will arrive at the destination in a random order at different times.

Suppose also that the information that the source wishes to transmit to the sink is temporally ordered. This could occur, for instance, if the source is streaming a multimedia clip to the sink composed of frames of multimedia data corresponding to different evenly spaced time instants. Alternatively, the source could be sending a sequence of control instructions to the sink through the network which must be executed in order, and within a short amount of time after transmission (e.g., to ensure stability of the associated control loop). In both instances, the unknown order of arrival of the M packets introduces ambiguity as to when data will be available for use at the sink. The purpose of delay mitigating codes is to encode the contents of the packets sent along the M paths in such a manner as to allow for successive decoding of information at the receiver at the schedule dictated by the content (i.e., playing out the multimedia or control frames when they need to be played). Such delay mitigating codes must exhibit a rate-delay tradeoff. A low delay guarantee requires extra coding redundancy, which in turn limits the number of distinct source data bits, which lowers the channel rate.

Our initial work [1], [2], [3] has characterized this rate delay tradeoff when the permutation distribution is *uniform*. In this instance, it is possible to build codes that achieve points on the boundary of the capacity region of the associated DBC with only M transmitted packets (i.e., only *one* use of the associated broadcast channel).[1] These codes are constructed by time-sharing maximum distance separable (MDS) codes to get overall codes

Both authors are with the Department of Electrical and Computer Engineering of Drexel University. The contact author is J. Walsh (jwalsh@ece.drexel.edu). Manuscript date: September 26, 2008.

[1]This comes as somewhat of a surprise given the asymptotically large number of channel uses that are required when using random codes to prove the associated rate regions.

known as priority encoded transmission (PET) codes. This allows delay to be defined directly in terms of the arrival instants of the M packets and the amount of source information bourne in them regarding each of the original N source frames is indicated by the rate vector on the boundary of the rate region of the corresponding DBC. Our work specifies how to design a PET code to achieve a particular delay guarantee.

By contrast, the analysis in the present paper allows us to consider the case in which some packet arrival orderings are more likely than others. With this shift to the general case comes the possible requirement that $B > 1$ blocks of M packets be simultaneously encoded in order to achieve points near the boundary of the associated DBC capacity region. Thus, in determining the achievable rate-delay tradeoffs for this general case it will be necessary to consider simultaneously the block length B required to achieve different points in the capacity region along with their associated rate vectors. The rate region presented and determined here is a necessary prerequisite for a complete analysis of trading rate for delay on such channels, hence the motivation for the present problem.

B. Summary of Related Work

Aside from our work mentioned in the previous section, a large amount of other related work deals with priority encoded transmission, [4] which are a class of efficient unequal erasure protection codes. The efficiency of these PET codes relative to a DBC equivalent to a uniform permutation channel was discussed in [5]. Our uniform permutation channel codes optimize the levels of unequal protection with delay mitigation in mind. A similar discussion appeared in [6] and an associated patent [7], but for a slightly different DBC formed from concatenated Q-ary probabilistic erasure channels. Because we are considering a non-uniform permutation channel, this work in this paper can be viewed as an extension of this prior work, although with a focus shift from successive refinements encoded source data in the PET work to temporally ordered source data in our work.

Far more recent approaches to the problem of content transmission over networks have involved fountain codes [8]. While these codes are useful for multicasting multimedia content over networks with destinations suffering from varying unpredictable packet erasure levels, from the perspective of the decoder they are block codes which should have large length to make the construction efficient. This requires a large amount of packet information to be received before the first (in time) source bit can be decoded. Whereas [9] measures the average number of bits that may be decoded upon reception of K packets using a fountain code, the concern in the context of mitigating delay is to also jointly specify at the encoder which bits (not just the fraction) can be decoded at a sequence of reception instants.

II. The permutation channel for M packets

The channel input is the M labeled, coded K-bit packets $u_M \equiv (x_1, \ldots, x_M)$, and the output is a permutation of the packets, $y_M \equiv (x_{\pi(1)}, \ldots, x_{\pi(M)})$. Define $[M] \equiv \{1, \ldots, M\}$. For a collection of generic packet indices, $A \subseteq [M]$, we use the shorthand $\mathbf{x}_A \equiv \{x_i,\ i \in A\}$ to denote the contents of the packets with those indices. The permutation π is selected from the set Π of $M!$ possible permutations of the elements $[M]$, according to a specified probability distribution $\mathbf{p} = (p(\pi),\ \pi \in \Pi)$. Consider the DBC with mth receiver observing $y_m \equiv (x_{\pi(1)}, \ldots, x_{\pi(m)})$. The capacity region of the discrete memoryless DBC with M receivers is known, and we restate it here below.

Theorem 1: [10], [11], [12]. The capacity region of the discrete memoryless DBC, both with and without feedback, is the closure of the convex hull of the region $\mathcal{R}$ of rates satisfying

$$R_m \leq I(y_m; u_m | u_{m-1}),\ m \in [M], \tag{1}$$

for a sequence of dummy discrete random variables $u_1, \ldots, u_M$ with bounded support, where the rvs $u_1 \to \cdots \to u_M \to y_M \to \cdots \to y_1$ form a Markov chain, and $u_0 \equiv 0$.

In other words, $\mathbf{R} = (R_1, \ldots, R_M)$ is an achievable rate vector if there exists a collection of rvs $u_1, \ldots, u_M$ such that the M rate inequalities are satisfied. The following proposition is obtained by applying the definition of the permutation channel to Theorem 1.

Proposition 1: In the particular case of the permutation channel, the generic DBC capacity region expression (1) can be simplified to:

$$\begin{aligned} R_m &\leq \sum_{G \in \mathcal{G}_m} p_m(G) I(\mathbf{x}_G; u_m | u_{m-1}),\ m \in [M-1] \\ R_M &\leq H(u_M | u_{M-1}) \end{aligned} \tag{2}$$

where $\mathcal{G}_m = \{G \subseteq [M] \ :\ |G| = m\}$ is the set of all subsets of $[M]$ of cardinality m. The probabilities $p_m(G)$, defined as

$$p_m(G) \equiv \sum_{\pi \in \mathcal{P}_m(G)} p(\pi),\ G \in \mathcal{G}_m, \tag{3}$$

are the probabilities that the first m packets to arrive are those with indices in G. The set $\mathcal{P}_m(G)$, defined as

$$\mathcal{P}_m(G) = \{\pi \in \Pi \ :\ \{\pi(1), \ldots, \pi(m)\} = G\}, \tag{4}$$

contains those permutations $\pi \in \Pi$ with their first m elements equal to the indices in G.

Our earlier work [3] also gives the capacity region for two specific permutation channels: the deterministic channel, where $p(\pi) = 1$ for some particular $\pi \in \Pi$, and the uniform channel, where $p(\pi) = 1/M!$ for each $\pi \in \Pi$. We collect these results in the following theorem.

Theorem 2: For the deterministic permutation channel, where only a single permutation $\pi \in \Pi$ is possible, the capacity region is specified by the M inequalities

$$\sum_{i=1}^{m} R_i \leq mK, \quad m \in [M]. \tag{5}$$

For the uniform permutation channel, where all permutations are equally likely, the capacity region is specified by the inequality:

$$\sum_{m=1}^{M} \frac{R_m}{m} \leq K. \tag{6}$$

It is easy to see that the deterministic permutation channel is an outer bound and the uniform permutation channel is an inner bound on the capacity region for all other permutation distributions $\mathbf{p}$.

III. A CONJECTURE ON THE CAPACITY REGION OF THE PERMUTATION CHANNEL WITH M PACKETS

Here, we give an inner bound, $\mathcal{R}_M^{\text{in}}$, and an outer bound, $\mathcal{R}_M^{\text{out}}$, on the capacity region, $\mathcal{R}_M$, of the permutation channel, and conjecture that the two bounds are in fact equal to the actual capacity region.

A. Achievability (inner bound)

In order to characterize a region of achievable rates, consider all partitions of the M packets among M different users, $\mathcal{A}_M$, defined as:

$$\mathcal{A}_M \equiv \left\{ (A_1, \ldots, A_M) \;\middle|\; \begin{array}{c} A_m \subseteq [M], \\ \bigcup_{m=1}^{M} A_m = [M], \\ A_m \cap A_{m'} = \emptyset, \ m \neq m' \end{array} \right\}. \tag{7}$$

A member of this set of assignments $\mathbf{A} = (A_1, \ldots, A_M) \in \mathcal{A}_M$ has elements $A_m \subseteq [M]$ assigning packet indices to each user $m \in [M]$. An assignment of packets to user m means all of the K bits in each of those packets are used to encode the source bits for user m.

The input to the permutation channel as viewed by a typical user is the pair $X \equiv (A, \mathbf{x}_A)$, i.e., the packet contents ($\mathbf{x}_A$) and their associated labels (A). The output of the channel is the pair $Y \equiv (\mathcal{B}, \mathbf{x}_{\mathcal{B}})$ for a random subset $\mathcal{B} \subseteq A$ indicating those packet indices in A that are among the first packets to arrive at the receiver (i.e., among the first m packets for user m). The capacity of this channel is given in the following proposition.

Proposition 2: The capacity of a point to point channel with input $X \equiv (A, \mathbf{x}_A)$ and output $Y \equiv (\mathcal{B}, \mathbf{x}_{\mathcal{B}})$ for a random $\mathcal{B} \subseteq A$ is K bits per packet times the expected number of packets to arrive under the distribution on $\mathcal{B}$:

$$C(A) = K \sum_{B \subseteq A} p_A(B)|B| = K\mathbb{E}_A[|\mathcal{B}|], \tag{8}$$

where $p_A(B)$ and $\mathbb{E}_A[|\mathcal{B}|]$ indicate the support and distribution of $\mathcal{B}$ depend upon the index set A.

Proof: Let $\mathcal{Q}$ be the collection of distributions $\mathbf{q}$ on the input X, meaning the distribution on the contents of packets $\mathbf{x}_A$ (since A is a fixed, not random, set). Thus:

$$C(A) \equiv \max_{\mathbf{q} \in \mathcal{Q}} H(Y) - H(Y|X). \tag{9}$$

Conditioning on the random subset of received packets $\mathcal{B}$:

$$\begin{aligned} H(Y) &= H(\mathcal{B}, \mathbf{x}_{\mathcal{B}}) = H(\mathcal{B}) + H(\mathbf{x}_{\mathcal{B}}|\mathcal{B}) \\ H(Y|X) &= H(\mathcal{B}, \mathbf{x}_{\mathcal{B}}|A, \mathbf{x}_A) \\ &= H(\mathcal{B}|A, \mathbf{x}_A) + H(\mathbf{x}_{\mathcal{B}}|\mathcal{B}, A, \mathbf{x}_A) = H(\mathcal{B}). \end{aligned} \tag{10}$$

Here, $H(\mathcal{B}|A, \mathbf{x}_A) = H(\mathcal{B})$ since the channel's random selection of the subset of the received packets, indexed by $\mathcal{B}$, is independent of the contents of the packets (recall A is a fixed, not random, set). Moreover, $H(\mathbf{x}_{\mathcal{B}}|\mathcal{B}, A, \mathbf{x}_A) = 0$ as there is no uncertainty as to the contents of the received packets, $\mathbf{x}_{\mathcal{B}}$, when conditioned on the set of received packet indices, $\mathcal{B}$, and the contents of all transmitted packets, $\mathbf{x}_A$. Combining these observations and conditioning on all possible values of $\mathcal{B}$:

$$\begin{aligned} C(A) &\equiv \max_{\mathbf{q} \in \mathcal{Q}} H(\mathbf{x}_{\mathcal{B}}|\mathcal{B}) \\ &= \max_{\mathbf{q} \in \mathcal{Q}} \sum_{B \subseteq A} p_A(B) H(\mathbf{x}_B|\mathcal{B} = B) \\ &\leq \sum_{B \subseteq A} p_A(B) K |B|. \end{aligned} \tag{11}$$

Equality is achieved by the uniform distribution on the contents of packets in A, i.e., on $\mathbf{x}_A$. ■

By Proposition 2, the following theorem gives a polytope region of achievable rates. Informally, the theorem states that a packet assignment $\mathbf{A}$ can achieve an associated rate point $\mathbf{R}(\mathbf{A})$, where each user m receives rate $R_m(A_m)$ equal to K bits per packet times the number of packets from A_m expected to arrive by time m.

Theorem 3: For a packet assignment $\mathbf{A} = (A_1, \ldots, A_M)$, consider M point-to-point channels, one for each user $m \in [M]$, where the input on channel m is $X_m \equiv (A_m, \mathbf{x}_{A_m})$ and the output is $Y_m \equiv (\mathcal{B}_m, \mathbf{x}_{\mathcal{B}_m})$, where the support of $\mathcal{B}_m$ is the set of indices from A_m

of length at least $(|A_m|+m-M)^+$ and at most m (since user m sees exactly m packets). The achievable rate point associated with assignment A is

$$\begin{aligned} \mathbf{R}(\mathbf{A}) &= (R_1(A_1),\ldots,R_M(A_M)), \qquad (12)\\ R_m(A_m) &= K\sum_{B\in\mathcal{S}(A_m,m)} p_{A_m}(B)|B|,\ m\in[M], \end{aligned}$$

where the support of $\mathcal{B}_m$ is defined as

$$\mathcal{S}(A_m,m)=\{B\subseteq A_m : (|A_m|+m-M)^+ \le |B| \le m\}, \qquad (13)$$

and the distribution of $\mathcal{B}_m$ is given by

$$p_{A_m}(B)=\sum_{\pi\in\mathcal{P}(B,A_m,m)} p(\pi),\ B\in\mathcal{S}(A_m,m), \qquad (14)$$

and where the set of permutations $\mathcal{P}(B,A_m,m)\subseteq\Pi$ is defined as:

$$\left\{\pi\in\Pi \,\middle|\, \begin{array}{c} B\subseteq\{\pi(1),\ldots,\pi(m)\}, \\ A_m\setminus B\subseteq\{\pi(m+1),\ldots,\pi(M)\} \end{array}\right\}. \qquad (15)$$

A region of achievable rates is defined by the polytope formed by the convex hull of the rate points for each possible packet assignment:

$$\mathcal{R}_M^{\text{in}} = \text{conv}\left(\{\mathbf{R}(\mathbf{A}),\ \mathbf{A}\in\mathcal{A}_M\}\right). \qquad (16)$$

Non-vertex points in this set are achievable via time-sharing across channel uses. The region $\mathcal{R}_M^{\text{in}}$ is an inner bound on the capacity of the permutation channel.

The minimum number of received packets for user m, i.e., $(|A_m|+m-M)^+$, follows from the fact that the number of received packets for *other* users among the first m received packets is upper bounded by the total number of transmitted packets for other users: $m-|\mathcal{B}_m|\le M-|A_m|$. The set of permutations $\mathcal{P}(B,A_m,m)$ consists of those permutations where the first m elements contain all elements of B, and the last $M-m$ elements contain all elements from A_m not in B.

B. Converse (outer bound)

The M DBC equations given in Proposition 1 may be viewed as a linear map from the set of feasible vectors of mutual information to the set of feasible rate points, where the coefficients of the linear map are determined by the permutation distribution $\mathbf{p}$. In particular, the DBC equation for user m depends upon $n_m\equiv\binom{M}{m}$ distinct mutual informations $I(x_G;u_m|u_{m-1})$, one for each $G\in\mathcal{G}_m$. It follows that the rate point $\mathbf{R}=(R_1,\ldots,R_M)$ depends upon a total of $\sum_{m=1}^{M}\binom{M}{m}=2^M-1$ distinct mutual information variables (the sole variable for R_M, the entropy $H(u_M|u_{M-1})$, is a self-information). This linear map may be written as a $M\times 2^M-1$ matrix:

$$\mathbf{T}=\text{BlockDiag}\left(\mathbf{V}_1,\ldots,\mathbf{V}_{M-1},1\right) \qquad (17)$$

where the BlockDiag function creates a matrix whose non-zero block diagonal sub-matrices are

$$\begin{aligned} \mathbf{V}_1 &= (p_1(G_{1,1})\cdots p_1(G_{1,n_1})) \qquad (18)\\ &\ \ \vdots \\ \mathbf{V}_{M-1} &= (p_{M-1}(G_{M-1,1})\cdots p_{M-1}(G_{M-1,n_{M-1}})) \end{aligned}$$

In particular, a feasible mutual information vector for a particular joint distribution on the message r.v.s

$$\mathbf{I}=(I(\mathbf{x}_G;u_m|u_{m-1}),\ G\in\mathcal{G}_m,\ m\in[M]) \qquad (19)$$

maps to a feasible DBC rate point $\mathbf{R}$ via

$$\mathbf{TI}=\mathbf{R}\in\mathcal{R}_M, \qquad (20)$$

where $\mathcal{R}_M$ is the capacity region of the DBC. It is simpler to express the feasible region for mutual information vectors in terms of the corresponding entropy vector

$$\mathbf{H}=(H(\mathbf{x}_G|u_{m-1}),H(\mathbf{x}_G|u_m),\ G\in\mathcal{G}_m,\ m\in[M]). \qquad (21)$$

To that end, the $2^M-1\times 2(2^M-1)-1$ matrix $\mathbf{V}$ maps entropy vectors to mutual information vectors:

$$\mathbf{V}=\begin{bmatrix} 1 & -1 & & & & \\ & & \ddots & & & \\ & & & 1 & -1 & \\ & & & & & 1 \end{bmatrix}, \qquad (22)$$

where only the non-zero elements are listed. That is, if $\mathbf{H}$ is a feasible entropy vector (i.e., it corresponds to some joint distribution on the message random variables) then $\mathbf{I}=\mathbf{VH}$ is a feasible mutual information vector. The space of feasible entropy vectors with entries of the form in (21) lies within the polytope formed by the following set of linear constraints. We first ensure that the (unconditioned) entropy of each packet is at most K bits:

$$H(x_m)\le K,\ m\in[M] \qquad (23)$$

Second, we ensure non-negativity of mutual information, or, equivalently, that conditioning reduces entropy. Here, for brevity conditioning on u_0 is equivalent to no conditioning.

$$H(\mathbf{x}_G|u_m)\le H(\mathbf{x}_G|u_{m-1}),\ G\in\mathcal{G}_m,\ m\in[M] \qquad (24)$$

The third constraint is a conditioned form of Han's inequality ([13], (16.38), p491):

$$\begin{aligned} H(\mathbf{x}_G|u_{m-1}) &\le \frac{1}{m-1}\sum_{F\subseteq G:|F|=m-1} H(\mathbf{x}_F|u_{m-1}),\\ & G\in\mathcal{G}_m,\ m\in[M] \qquad (25) \end{aligned}$$

Fourth, we ensure that the entropy of any collection of m rvs exceeds the entropy of any subset of $m-1$ of those rvs:

$$\max_{F \subseteq G:|F|=m-1} \{H(\mathbf{x}_F|u_{m-1})\} \leq H(\mathbf{x}_G|u_{m-1}), \quad G \in \mathcal{G}_m,\ m \in [M] \tag{26}$$

Fifth, and finally, we require the entropy of a collection of m rvs is at most the sum of the entropies of a partition of those rvs into subsets of size m' and m'' with $m' + m'' = m$, and the conditioned rvs in the rhs change from u_m to $u_{m'}$ and $u_{m''}$ by the second group of inequalities

$$\begin{aligned} H(\mathbf{x}_G|u_m) \leq\ & H(\mathbf{x}_{G'}|u_{m'}) + H(\mathbf{x}_{G''}|u_{m''}), \\ & G \in \mathcal{G}_m,\ G' \in \mathcal{G}_{m'},\ G'' \in \mathcal{G}_{m''} \\ & G = G' \cup G'' \\ & m' + m'' = m,\ m \in [M] \end{aligned} \tag{27}$$

The nonnegative vectors obeying (23–27) form a polytope, denoted by $\mathcal{H}$. It follows that there exists a minimal set of entropy vector vertices $\{\mathbf{H}\}$ whose convex hull equals $\mathcal{H}$:

$$\mathcal{H} = \mathrm{conv}(\{\mathbf{H}\}). \tag{28}$$

Convert each entropy vector vertex to a mutual information vector so that the set of feasible mutual information vectors lies within the convex hull of the vertices $\mathbf{I} = \mathbf{VH}$ for each $\mathbf{H} \in \{\mathbf{H}\}$:

$$\mathcal{I} = \mathrm{conv}(\{\mathbf{I}\}),\ \mathbf{I} = \mathbf{VH},\ \mathbf{H} \in \{\mathbf{H}\}. \tag{29}$$

That $\mathcal{I}$ contains the set of mutual information vectors follows from the fact that the polytope $\mathcal{H}$ is a superset to the set of feasible entropy vectors, and the set of feasible mutual information vectors is the image of the set of feasible entropy vectors under $\mathbf{V}$. Finally, convert each mutual information vector vertex in $\{\mathbf{I}\}$ to a rate vector vertex $\mathbf{R} = \mathbf{TI}$ for each $\mathbf{I} \in \{\mathbf{I}\}$, and define $\mathcal{R}_M^{\mathrm{out}}$ as the convex hull of these rate points. We have shown the following outer bound.

Theorem 4: An outer bound on the capacity region of the permutation channel is the polytope given by the convex hull of the rate vector vertices $\{\mathbf{R}\}$:

$$\mathcal{R}_M^{\mathrm{out}} = \mathrm{conv}(\{\mathbf{R}\}),\ \mathbf{R} = \mathbf{TVH},\ \mathbf{H} \in \{\mathbf{H}\}. \tag{30}$$

Finally, we conjecture that the inner bound from Theorem 3 (16) equals the outer bound from Theorem 4 (30).

Conjecture 1: The inner bound bound and outer bound are tight, and thus the capacity of the permutation channel is given by both (16) and (30):

$$\mathcal{R}_M = \mathcal{R}_M^{\mathrm{in}} = \mathcal{R}_M^{\mathrm{out}}. \tag{31}$$

Moreover, we conjecture the (minimal) set of vertices generating the two polytopes are equal:

$$\{\mathbf{R}(\mathbf{A}),\ \mathbf{A} \in \mathcal{A}_M\} = \{\mathbf{R}(\mathbf{H}),\ \mathbf{H} \in \{\mathbf{H}\}\}. \tag{32}$$

IV. Capacity region for $M = 2$ packets

We now show Conjecture 1 is true for $M = 2$ packets (this section) and $M = 3$ packets (Section V). For $M = 2$ there are only two permutations: $\pi_1 = (1, 2)$ (packets arrive in order) and $\pi_2 = (2, 1)$ (packets arrive out of order), let the probability of the former be p. Further, we define $r_i = R_i/K$ for $i = 1, 2$ as the normalized rate per packet bit. By Theorem 3, the achievable rate points for the different packet assignments are

$$\begin{aligned} \mathbf{R}(\emptyset, \{1,2\}) &= (0, 2), \\ \mathbf{R}(\{1,2\}, \emptyset) &= (1, 0), \\ \mathbf{R}(\{1\}, \{2\}) &= (p, 1), \\ \mathbf{R}(\{2\}, \{1\}) &= (1-p, 1) \end{aligned} \tag{33}$$

It is clear that a minimal set of vertices generating the inner bound is the four points:

$$\mathcal{R}_2^{\mathrm{in}} = \mathrm{conv}\left(\{(0,0), (0,2), (1,0), (\max\{p, 1-p\}, 1)\}\right). \tag{34}$$

For the converse, the program `lrs` [14] is used to enumerate[2] the vertices of $\mathcal{H}$ for $M = 2$. Passing these points through $\mathbf{V}$ and then $\mathbf{T}$ produces the five points

$$(0,0), (0,2), (1,0), (p,1), (1-p,1). \tag{35}$$

We have shown the capacity region is given by (34), confirming Conjecture 1 for $M = 2$. The capacity region is illustrated in Figure 1.

V. Capacity region for $M = 3$ packets

Table I lists the achievable rate vectors for $M = 3$ discussed in Theorem 3, where a_i is the *number* of packets assigned to each user, and $r_i = R_i/K$ is the corresponding rate per packet bit. Moreover, p_i^1 is the probability that packet i is the first to arrive, and p_{ij}^2 is the probability that the *unordered* pair of packets $\{i, j\}$ is the first pair of packets to arrive. For the converse `lrs` yields 152 entropy vectors which generate $\mathcal{H}$ for $M = 3$. Mapping these through $\mathbf{V}$, and removing the redundant points not needed for the convex hull (using `redund` from [14]) gives 70 distinct mutual information vectors, which when mapped through $\mathbf{T}$, after manually removing points not on the boundary yields the points in the Table I. This confirms the conjecture for $M = 3$. Some example capacity region plots are provided in Figure 2.

[2]There are easier ways to prove the converse for $M = 2$, but we use a method that highlights the general technique.

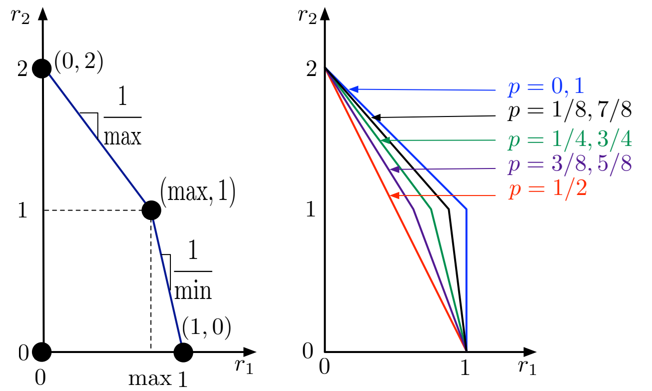

Fig. 1. The capacity region for the permutation channel with $M = 2$ packets. Here, $\max \equiv \max\{p, 1-p\}$ and $\min \equiv \min\{p, 1-p\}$.

(a_1, a_2, a_3)	(r_1, r_2, r_3)
$(3, 0, 0)$	$(1, 0, 0)$
$(0, 3, 0)$	$(0, 2, 0)$
$(0, 0, 3)$	$(0, 0, 3)$
$(0, 1, 2)$	$(0, \max\{p_{12}^2 + p_{13}^2, p_{12}^2 + p_{23}^2, p_{13}^2 + p_{23}^2\}, 2)$
$(0, 2, 1)$	$(0, \max\{1 + p_{12}^2, 1 + p_{13}^2, 1 + p_{23}^2\}, 1)$
$(1, 0, 2)$	$(\max\{p_1^1, p_2^1, p_3^1\}, 0, 2)$
$(1, 2, 0)$	$(p_1^1, 1 + p_{23}^2, 0)$ $(p_2^1, 1 + p_{13}^2, 0)$ $(p_3^1, 1 + p_{12}^2, 0)$
$(2, 0, 1)$	$(\max\{p_1^1 + p_2^1, p_1^1 + p_3^1, p_2^1 + p_3^1\}, 0, 1)$
$(2, 1, 0)$	$(p_1^1 + p_2^1, p_{13}^2 + p_{23}^2, 0)$ $(p_1^1 + p_3^1, p_{12}^2 + p_{23}^2, 0)$ $(p_2^1 + p_3^1, p_{12}^2 + p_{13}^2, 0)$
$(1, 1, 1)$	$(p_1^1, \max\{p_{12}^2 + p_{23}^2, p_{13}^2 + p_{23}^2\}, 1)$ $(p_2^1, \max\{p_{12}^2 + p_{23}^2, p_{13}^2 + p_{23}^2\}, 1)$ $(p_3^1, \max\{p_{12}^2 + p_{13}^2, p_{12}^2 + p_{23}^2\}, 1)$

TABLE I

THE ACHIEVABLE RATE VERTICES FOR $M = 3$.

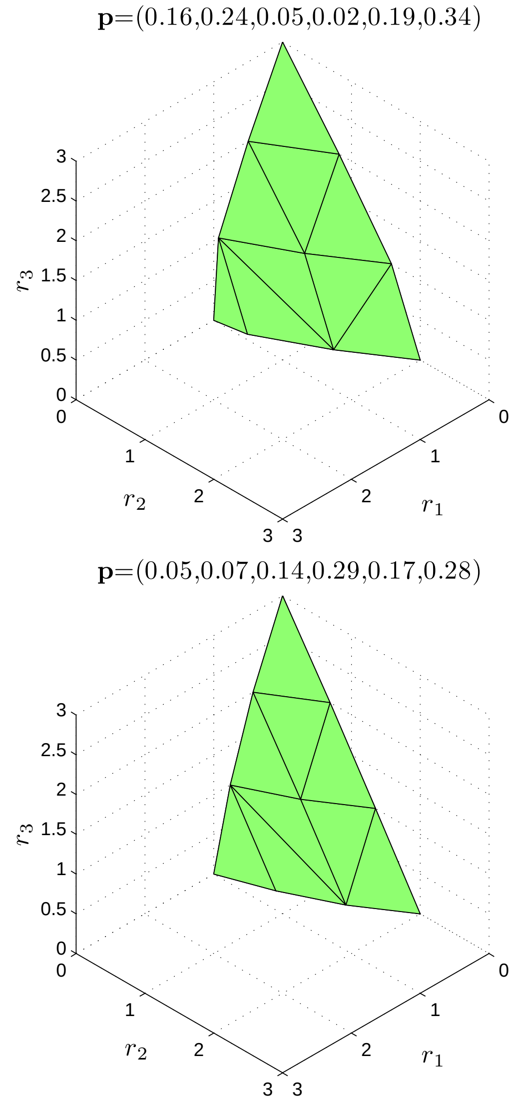

Fig. 2. The permutation channel capacity region for $M = 3$ and two different permutation distributions $\mathbf{p} = (p(123), p(132), p(213), p(231), p(312), p(321))$ on the six possible permutations.

REFERENCES

[1] J. Walsh and S. Weber, "Coding to reduce delay on a permutation channel," in *45th Allerton Conference on Communication, Control, and Computing*, September 2007.

[2] J. Walsh, S. Weber, and C. wa Maina, "Optimal rate delay tradeoffs for multipath routed and network coded networks," in *International Symposium on Information Theory (ISIT)*, July 2008.

[3] ——, "Optimal rate delay tradeoffs and delay mitigating codes for multipath routed and network coded networks," *submitted to the IEEE Transactions on Information Theory*, April 2008.

[4] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoded transmission," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[5] S. Boucheron and M. R. Slamatian, "About priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 699–705, Mar. 2000.

[6] R. L. Urbanke and A. D. Wyner, "Packetizing for the erasure broadcast channel with an internet application," 1997, preprint. [Online]. Available: citeseer.ist.psu.edu/urbanke97packetizing.html

[7] ——, "Methods and apparatus for packetizing data for transmission through an erasure broadcast channel," U.S. Patent 6 175 944, 1997.

[8] A. Shokrollahi, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[9] S. Sanghavi, "Intermediate performance of rateless codes," in *Information Theory Workshop (ITW)*, September 2007, pp. 478–482.

[10] A. E. Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Transactions on Information Theory*, vol. IT-24, no. 3, pp. 379–381, May 1978.

[11] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. IT-18, no. 1, pp. 2–14, January 1972.

[12] ——, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. IT-44, no. 6, pp. 2524–2530, October 1998.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley and Sons, 1991.

[14] D. Avis, "lrslib ver 4.2." [Online]. Available: http://cgm.cs.mcgill.ca/∼avis/C/lrs.html