On the Minimum Number of Transmissions Required for Universal Recovery in Broadcast Networks

Thomas A. Courtade, *Student Member, IEEE,* and Richard D. Wesel, *Senior Member, IEEE*

***Abstract*—Consider an arbitrarily connected broadcast network of N nodes that all wish to recover k desired packets. Each node begins with a subset of the desired packets and broadcasts messages to its neighbors. In a previous paper we established necessary and sufficient conditions on the number of transmissions from each node required for universal recovery (in which each node recovers all k packets). However, these conditions are numerous and cumbersome. The present paper gives a series of relatively simple conditions for universal recovery that apply when the number of packets is large and the distribution of packets among the nodes is well behaved.**

Our first results, which apply to any fixed network topology, use only simple cuts in the network to characterize a set of transmission strategies such that for any $\epsilon > 0$ these strategies require at most $k\epsilon$ transmissions above the minimum required for universal recovery. For certain topologies including d-regular d-connected networks, we explicitly construct transmission strategies that achieve universal recovery while using at most N transmissions above the minimum even when the total number of required transmissions is very large. These explicit constructions essentially resolve the problem completely for many canonical networks (e.g. cliques, rings, grids on tori, etc.).

I. Introduction

CONSIDER an arbitrary network of N nodes that all wish to recover k desired packets. Each node begins with a subset of the desired packets and broadcasts messages to its neighbors over discrete, memoryless, and interference-free channels. Furthermore, every node knows which packets are already known by each node and knows the topology of the network. How many transmissions are required to permit universal recovery (i.e. to disseminate the k packets to every node in the network)? This is the central question addressed. In fact, answering this question allows one to efficiently design a communication scheme which permits universal recovery using exactly the minimum required number of transmissions.

For the special case when the network is a clique, upper and lower bounds for this problem have been recently described in [14]. Our paper provides feasibilty results and asymptotic bounds on the minimum number of transmissions required for the general case, and it also provides bounds (within a constant number of transmissions) on the minimum required number of transmissions for a general class of networks which includes many of the canonical networks frequently considered. In doing so, it significantly extends the recent work by the present authors [1].

The authors are with the Electrical Engineering Department, University of California, Los Angeles, CA, 90095 USA (email: tacourta@ee.ucla.edu; wesel@ee.ucla.edu).

The problem addressed in this paper is related to the index coding problem originally introduced by Birk and Kol in [4]. Specifically, generalizing the index coding problem to permit each node to be a transmitter (instead of having a single server) and further generalizing so that the network need not be a single hop network leads to a class of problems that includes our problem as a special case in which each node desires to receive all packets.

It has been noted that, in general, nonlinear coding is required for the index coding problem (See [2], [12].). As discussed above, our problem is distinct from the index coding problem, and it turns out that linear encoding does achieve the minimum number of transmissions required for universal recovery.

This paper is organized as follows. Section II defines the problem and introduces basic definitions and notation. Section III describes our main results: (1) necessary and sufficient conditions describing all transmission strategies allowing universal recovery, (2) the characterization (in terms of simple cuts in the network) of a set of transmission strategies requiring at most $k\epsilon$ transmissions above the minimum number required for universal recovery , and (3) for certain topologies (including many canonical networks), we explicitly construct transmission strategies that achieve universal recovery while using at most N transmissions above the minimum number required. Section IV sketches the proofs for these results. Section V delivers the conclusions.

II. System Model

This paper considers an arbitrary network $\mathcal{T}$ of N nodes. The network must be connected, but it need not be fully connected (a clique). A graph $\mathcal{G}_{\mathcal{T}} = (V, E)$ describes the specific connections in the network, where V is the set of vertices $\{v_i : i \in \{1, \ldots, N\}\}$ (each corresponding to a node) and E is the set of edges connecting nodes. We assume that the edges in E are undirected, but our work easily extends to directed graphs.

Each node wishes to recover the same k desired packets, and each node begins with a (possibly empty) subset of the desired packets. $\mathcal{P}_i \subseteq \{p_1, \ldots, p_k\}$ is the set of packets originally available at node i, and $\{\mathcal{P}_i\}_{i=1}^N$ satisfies $\bigcup_{i=1}^N \mathcal{P}_i =$

Preprint submitted to Forty-Eighth Annual Allerton Conference.
Received July 7, 2010.

$\{p_1, \ldots, p_k\}$. Each $p_j \in \mathbb{F}$, where $\mathbb{F}$ is some finite field (e.g. $\mathbb{F} = \mathrm{GF}(2^m)$).

In this paper, it will often be necessary to discuss sequences of networks indexed by the total number of packets k. In these sequences of networks, we assume the network topology is fixed, but that the distribution of packets is dependent on k. In order to emphasize this dependence on k, we will often refer to a sequence of packet distributions indexed by k as $\{\mathcal{P}_i(k)\}_{i=1}^N$. Naturally, these sets must satisfy $\left|\bigcup_{i=1}^N \mathcal{P}_i(k)\right| = k$. We say that a sequence of packet distributions is *well-behaved* if the limit

$$\mathbb{P}_{\mathcal{S}} \triangleq \lim_{k\to\infty} \frac{1}{k} \left| \bigcup_{i\in\mathcal{S}} \mathcal{P}_i(k) \right| \quad (1)$$

exists for all subsets $\mathcal{S} \subseteq \{1, \ldots, N\}$. In other words, $\mathbb{P}_{\mathcal{S}}$ is the limit of the empirical probability that any node in $\mathcal{S}$ receives a particular packet. We define $\mathbb{P}^c_{\mathcal{S}} = 1 - \mathbb{P}_{\mathcal{S}}$. The authors would like to comment that condition (1) can be replaced with convergence in probability; this is discussed in subsection III-D.

Let the set $\mathcal{B}(i)$ be the neighbors of node i. For convenience, $i \in \mathcal{B}(i)$. There exists an edge $e \in E$ connecting two vertices $v_i, v_j \in V$ iff $i \in \mathcal{B}(j)$. Node i broadcasts messages to its neighbors $\mathcal{B}(i)$ over discrete, memoryless, and interference-free channels. If $\mathcal{S}$ is a set of nodes, then we define $\mathcal{B}(\mathcal{S}) = \cup_{i\in\mathcal{S}}\mathcal{B}(i)$. The neighbors of a set $\mathcal{S}$ not included in $\mathcal{S}$ itself are denoted $\tilde{\mathcal{B}}(\mathcal{S}) = \mathcal{B}(\mathcal{S}) \cap \mathcal{S}^c$.

Throughout this paper, we adopt the conventional notation $[k] \triangleq \{1, \ldots, k\}$. An indexed set (or vector) $\{x_1, \ldots, x_k\}$ is referred to by the shorthand notation $\{x_i\}_{i=1}^k$. When the range of the index is apparent from context, we will sometimes abbreviate it further as $\{x_i\}$.

This paper seeks to determine the minimum amount of communication required to achieve universal recovery of the k packets. An important consideration is whether packets are considered indivisible (so that the smallest unit of transmission is a packet) or packets may be split into chunks so that a fraction of a packet may be transmitted. The implications of divisible packets were considered previously in [1]. In this paper, we maintain a combinatorial approach throughout and assume that transmissions must consist of an integer number of packets unless othwerwise stated. However, it should be noted that divisible packets can be modeled by choosing an appropriate sequence of well-behaved packet distributions.

A. Indivisible Packets

When packets are deemed indivisible, a single *transmission* by user i consists of sending a packet (some $z \in \mathbb{F}$) to all nodes $j \in \mathcal{B}(i)$. For the remainder of this section, let M be the minimum required number of such transmissions that allow universal recovery. Throughout this paper, we assume error-free broadcast channels and orthogonal multiple-access channels (i.e. there is no interference from simultaneous transmissions). Thus, the task of determining M is combinatorial.

Example 1 (Line Network): Suppose $\mathcal{T}$ is a network of nodes connected along a line as follows: $V = \{v_1, v_2, v_3\}$, $E = \{(v_1, v_2), (v_2, v_3)\}$, $\mathcal{P}_1 = \{p_1\}$, $\mathcal{P}_2 = \emptyset$, and $\mathcal{P}_3 = \{p_2\}$. Note that each node must transmit at least once in order for all nodes to recover $\{p_1, p_2\}$, hence $M \geq 3$. Suppose node 1 transmits p_1 and node 3 transmits p_2. Then (upon receipt of p_1 and p_2 from nodes 1 and 3) node 2 transmits $p_1 \oplus p_2$ where $\oplus$ indicates addition in the finite field $\mathbb{F}$). This strategy requires 3 transmissions and allows each user to recover $\{p_1, p_2\}$. Hence $M = 3$.

Example 1 demonstrates a transmission strategy that uses two *rounds* of communication. The broadcasts by node i in a particular round of communication can depend only on the information available to node i prior to that round (i.e. $\mathcal{P}_i$ and previously received transmissions from neighboring nodes). In other words, the broadcasts are causal.

Example 2 (Clique): Suppose $\mathcal{T}$ is a network of nodes that are fully connected so as to form a clique as follows: $\mathcal{P}_i = \{p_1, p_2, p_3\} \backslash p_i$, and $\mathcal{G}_{\mathcal{T}}$ is a clique of size 3. Clearly one transmission is not sufficient, thus $M \geq 2$. It can be seen that two transmissions suffice: let node 1 transmit p_2 which lets node 2 have $\mathcal{P}_2 \cup p_2 = \{p_1, p_2, p_3\}$. Now, node 2 transmits $p_1 \oplus p_3$, allowing nodes 1 and 3 to each recover all three packets. Thus $M = 2$. Since each transmission was only a function of the packets originally available at the corresponding node, this transmission strategy can be accomplished in a single round of communication.

III. Main Results

In this section, we state our main results. In subsection III-A, we define a set of conditions which are both necessary and sufficient for universal recovery. In subsection III-B, we derive asympotic bounds on the number of transmissions required for universal recovery by considering a slice of the region defined in subsection III-A. Subsection III-C specializes the results of subsection III-A for a wide class of networks. In particular, we demonstrate how to explicitly construct a transmission strategy which uses at most a constant number of transmissions more than the minimum number of transmissions required for universal recovery. Subsection III-D summarizes the results in the context of a probabilistic framework (i.e., when random packet distributions are considered). Proofs of all results are delayed until Section IV.

A. Transmission Strategies Permitting Universal Recovery

Let b_i^j be the number of transmissions from node i during round j. In this way, the total number of packet transmissions summing over all rounds is $\sum_{i=1}^N \sum_{j=1}^r b_i^j$. Also, let $\{b_i^j\}$ denote the set of b_i^j values for $i \in [N]$ and $j \in [r]$.

Define the region $\mathcal{R}_r \subset \mathbb{Z}_+^{n\times r}$ as follows:

$$\begin{aligned}&\{b_i^j\} \in \mathcal{R}_r \text{ if and only if:}\\ &\quad \forall \emptyset \subsetneq \mathcal{S}_0 \subseteq \cdots \subseteq \mathcal{S}_r \subsetneq [N] \text{ satisfying } \mathcal{S}_j \subseteq \mathcal{B}(\mathcal{S}_{j-1})\\ &\quad \text{for each } j \in [r], \text{ the following inequalities hold :}\\ &\qquad \sum_{j=1}^r \sum_{i\in\mathcal{S}_j^c \cap \mathcal{B}(\mathcal{S}_{j-1})} b_i^{(r+1-j)} \geq \left| \bigcap_{i\in\mathcal{S}_r} \mathcal{P}_i^c \right|. \end{aligned} \quad (2)$$

Theorem 1 (From [1]): For a fixed number of communication rounds r, a transmission strategy in which node i makes exactly b_i^j transmissions during the j^{th} round of communication permits universal recovery if and only if $\{b_i^j\} \in \mathcal{R}_r$.

Thus, the family of inequalities given by (2) are necessary and sufficient for universal recovery. These inequalities utilize sequences of sets of the form $\emptyset \subsetneq \mathcal{S}_0 \subseteq \cdots \subseteq \mathcal{S}_r \subsetneq [N]$ satisfying $\mathcal{S}_j \subseteq \mathcal{B}(\mathcal{S}_{j-1})$ for each $j \in [r]$. For convenience, we refer to this type of sequence of sets as an *appropiate sequence of sets*. The inequalities defined by these appropriate sequences of sets can be thought of as "generalized cutset bounds" which govern the information flow in the multihop broadcast networks under consideration.

In this paper, we are interested in the minimum number of transmissions required to permit universal recovery in a network. Since the number of transmissions required is finite, there exists some finite r_0 for which a vector $\{b_i^j\} \in \mathcal{R}_{r_0}$ defines a transmission strategy requiring the minimum number of transmissions.

For any feasible vector $\{b_i^j\}$ (i.e. one that satisfies $\{b_i^j\} \in \mathcal{R}_r$ for some r), a corresponding transmission strategy (in which user i makes exactly b_i^j transmissions during round j) can be computed in polynomial time using the algorithm described in [15]. Therefore, the difficulty in describing a transmission strategy which achieves (or approaches) the minimum number of transmissions lies solely in solving a minimization problem over $\mathcal{R}_r$.

Remark 1: Theorem 1 (and the subsequent results) can be readily extended to include the case where only a subset of the nodes wishes to reconstruct the original k packets and the other nodes serve only as helpers or relays. However, in order to keep the arguments and notation simple, we restrict our attention to the case where universal recovery is the objective.

B. Asymptotic Results for General Networks

Theorem 1 characterizes all transmission strategies by means of a very complicated sequence of cuts. The goal of this paper is to simplify this result to yield a more meaningful region which still characterizes feasible transmission strategies of interest. Somewhat surprisingly, a set of constraints which only considers simple cuts in the network characterizes a sufficient condition permitting universal recovery in the asymptotic regime under a per-round constraint on the number of transmissions by each node. In other words, for any $\epsilon > 0$, we can find transmission strategies requiring at most $k\epsilon$ more transmissions than the minimum number allowing universal recovery. This asymptotic result is analagous to a recent result of Mohajer et. al. [16] for the networks under consideration.

Theorem 2: For a fixed network topology, $\tau \triangleq k/r$ fixed, and a well-behaved sequence of packet distributions $\{\mathcal{P}_i(k)\}_{i=1}^N$, if $\{b_i\}$ satisfies:

$$\sum_{i \in \tilde{\mathcal{B}}(\mathcal{S})} b_i > \tau \mathbb{P}_{\mathcal{S}}^c, \quad \forall\, \emptyset \subsetneq \mathcal{S} \subsetneq [N] \tag{3}$$

then any vector $\{b_i^j\}$ satisfying $b_i^j \geq b_i$ for all $i \in [N], j \in [r]$ permits universal recovery for all sufficiently large k.

The interpretation of Theorem 2 is as follows: if node i is allowed to make at most b_i transmissions per communication round, then universal recovery is possible if the b_i's satisfy the conditions in the theorem. In other words, the b_i's can be thought of as the capacities of the broadcast links. In this context, Theorem 1 also yields the following converse.

Theorem 3: If node i is allowed to make at most b_i transmissions per communication round, $\tau = k/r$ is fixed, and there is some set $\emptyset \subsetneq \mathcal{S} \subsetneq [N]$ for which

$$\sum_{i \in \tilde{\mathcal{B}}(\mathcal{S})} b_i < \tau \mathbb{P}_{\mathcal{S}}^c, \tag{4}$$

then universal recovery is never possible for all k greater than some finite k_0 (assuming a well-behaved sequence of packet distributions).

A key ingredient in the proof of Theorem 2 is that we allow the number of communication rounds r grow linearly with k as $k \to \infty$. This yields a potential gap of $k\epsilon$ transmissions between the number of transmissions required by any transmission strategy satisfying the constraints of Theorem 2 and a truly optimum transmission strategy computed by optimizing over the region $\mathcal{R}_{r_0}$ directly. It is perhaps more interesting to approximate the minimum number of transmissions required within a constant factor (not depending on k). In the next section, we accomplish this for a wide class of networks by constructing transmission strategies that require at most N more transmissions than the required minimum.

C. Transmission Strategies Within a Constant Number of Transmissions of the Optimum

As shown in [1], it is sometimes possible to analytically compute the minimum number of transmissions required. In [1], this was only demonstrated for clique networks and a special type of packet distribution.

For convenience, we will say that a vector $\{b_i^j\}$ is "within ℓ transmissions of optimal" if the minimum number of transmissions required for universal recovery is M and $\sum_{i,j} b_i^j \leq M + \ell$. In this section, we show for a wide class of networks that we can construct vectors $\{b_i^j\}$ that are within N transmissions of optimal. Thus, we can bound the minimum number of transmissions by the network parameter N, instead of by a linear factor of k as in the previous section.

Since we plan to determine (or approximate) the minimum required number of transmissions analytically, it is necessary to introduce some notational machinery based on the network topology and the limiting packet distribution $\{\mathbb{P}_{\mathcal{S}}\}_{\mathcal{S} \subseteq [N]}$. In what follows, we shall always assume that we are working with a well-behaved sequence of packet distributions with limiting distribution $\{\mathbb{P}_{\mathcal{S}}\}_{\mathcal{S} \subseteq [N]}$.

Define A to be the adjacency matrix of the graph $\mathcal{G}_{\mathcal{T}}$ (i.e. $a_{i,j} = 1$ if $(i,j) \in E$ and $a_{i,j} = 0$ otherwise). For convenience, if A is nonsingular, we say that the network is nonsingular.

Preprint submitted to Forty-Eighth Annual Allerton Conference.
Received July 7, 2010.

For a particular k and subset $\mathcal{S}$, define

$$\epsilon_{\mathcal{S},k} = \frac{1}{k}\left|\bigcap_{i\in\mathcal{S}} \mathcal{P}_i^c(k)\right| - \mathbb{P}_{\mathcal{S}}^c. \tag{5}$$

Then, one can write:

$$\left|\bigcap_{i\in\mathcal{S}} \mathcal{P}_i^c(k)\right| = k\left(\mathbb{P}_{\mathcal{S}}^c + \epsilon_{\mathcal{S},k}\right). \tag{6}$$

Let $\vec{\mathbb{P}}^c = [\mathbb{P}_1^c, \ldots, \mathbb{P}_N^c]^T$ and $\vec{\epsilon}_k = [\epsilon_{1,k}, \ldots, \epsilon_{N,k}]^T$. Assuming A is nonsingular, define $\delta_i = \left[A^{-1}\vec{\mathbb{P}}^c\right]_i$ and $\epsilon'_{i,k} = \left[A^{-1}\vec{\epsilon}_k\right]_i$, where $[\vec{x}]_i$ denotes the i^{th} coordinate of the vector $\vec{x}$.

The following theorem gives sufficient conditions that, when met, allow us to explicitly construct a transmission strategy $\{b_i^j\}$ from the quantities k, r, δ_i, and $\epsilon'_{i,k}$.

Theorem 4: For all k sufficiently large, if r is a constant not depending on k, $\delta_i + \epsilon'_{i,k} \geq 0$ and

$$\frac{1}{r}\sum_{j=1}^{r} \sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} \delta_i > \mathbb{P}_{\mathcal{S}_r}^c \tag{7}$$

for all appropriate sequences of sets with $|\mathcal{S}_r| \geq 2$, then choosing b_i^j so that $b_i^j \geq \lfloor \frac{k}{r}(\delta_i+\epsilon'_{i,k})\rfloor$ and $\sum_{j=1}^{r} b_i^j = \lceil k(\delta_i+\epsilon'_{i,k})\rceil$ yields a vector $\{b_i^j\}$ that is within N transmissions of optimal.

Remark 2: If packets are divisible, then we can always choose $\{b_i^j\}$ so that $\sum_{j=1}^{r} b_i^j = k(\delta_i+\epsilon'_{i,k})$. Thus, the optimum can be acheived by packet splitting. See [1] for a detailed discussion regarding divisible packets.

The main application of Theorem 4 is to networks which have some structure allowing the δ_i's to be easily computed. In this paper, we are particularly interested in the case of d-regular networks (i.e. every node has d neighbors) and packet distributions where $\mathbb{P}_i = \rho$ for all $i \in [N]$ (i.e. each node has approximately $\rho \times k$ packets). The choice of d-regularity includes many canonical networks (e.g. cliques, rings, grids on tori, Cayley graphs, etc.), and the choice of the uniform-type packet distribution is inspired by real-world type applications where the networks are more or less homogeneous.

In the following theorem, we also require our networks to be d-connected. A d-connected network is a network in which at least d nodes must be removed in order to disconnect the network. Clearly the connectivity of a d-regular network is at most d since one can remove the nodes in $\hat{\mathcal{B}}(i)$ to disconnect i from the rest of the network. It turns out (see [17]) that almost every large random d-regular network is d-connected, therefore our choice to enforce the d-connectivity criterion serves to eliminate certain pathological realizations of networks which would rarely (if ever) appear in practice (e.g. extreme bottlenecks between large pieces of the network).

Theorem 5: All nonsingular d-regular, d-connected networks with $\rho > 0$ fixed,

$$\mathbb{P}_i = \rho < \mathbb{P}_{\mathcal{S}}, \quad \forall i \in [N],\ \forall \mathcal{S} : |\mathcal{S}| > 1, \tag{8}$$

and

$$(N - |\mathcal{S}|)\cdot \mathbb{P}_i^c > d \cdot \mathbb{P}_{\mathcal{S}}^c, \quad \forall \mathcal{S} : |\mathcal{S}| > N - d \tag{9}$$

satisfy the conditions of Theorem 4. Accordingly, one can analytically compute a vector $\{b_i^j\}$ that is within N transmissions of optimum. Moreover, for this $\{b_i^j\}$, we have that:

$$\frac{1}{d}\sum_{i\in[N]} |\mathcal{P}_i^c(k)| \leq \sum_{\substack{i\in[N]\\ j\in[r]}} b_i^j \leq N + \frac{1}{d}\sum_{i\in[N]} |\mathcal{P}_i^c(k)|. \tag{10}$$

Theorem 5 has some important implications which we now state as corollaries.

Corollary 1: If $\mathbb{P}_i < \frac{1}{d+1}$, then condition (9) in Theorem 5 can be omitted.

Corollary 2: If $\mathbb{P}_{\mathcal{S}} = 1$ whenever $|\mathcal{S}| \geq N - d + 1$, then condition (9) in Theorem 5 can be omitted.

Corollary 3: If the packet distribution satisfies

$$\mathbb{P}_{\mathcal{S}} = \frac{1-(1-q)^{|\mathcal{S}|}}{1-(1-q)^N}, \quad \forall \mathcal{S} \subseteq [N], \tag{11}$$

for any $0 < q < 1$, then conditions (8-9) in Theorem 5 can be omitted.

In other words, subject to the qualifications of Theorem 5, we can explicitly construct a transmission strategy $\{b_i^j\}$ that is within N transmissions of optimal. Thus, we have an analytical expression for transmission strategies that approximate the performance of the optimum strategy computable by solving an Integer Linear Program over the region $\mathcal{R}_{r_0}$.

D. Random Distributions of Packets

Instead of considering well-behaved sequences of packet distributions as previously defined, we can instead consider random distributions of packets which satisfy

$$\frac{1}{k}\left|\bigcup_{i\in\mathcal{S}} \mathcal{P}_i(k)\right| \to \mathbb{P}_{\mathcal{S}} \quad \text{in probability} \tag{12}$$

for all subsets $\mathcal{S} \subseteq [N]$. Then, by standard arguments, it can be shown that all the previous results hold with arbitrarily high probability as $k \to \infty$.

One particularly important random distribution that we consider is when each node has a packet with probability q independent from other nodes. Necessarily, we must condition on the fact that each packet is available to at least one node, yielding:

$$\mathbb{P}_{\mathcal{S}} = \frac{1-(1-q)^{|\mathcal{S}|}}{1-(1-q)^N}, \quad \forall \mathcal{S} \subseteq [N]. \tag{13}$$

Abusing terminology slightly, we call this the *independent distribution*.

Due to the importance of this interpretation, we summarize some of the previous results in this probabilistic context. Note that regardless of the random distribution considered, we always require that $\mathbb{P}_{[N]} = 1$ (i.e., all N nodes collectively have all the packets). Let $\epsilon'_{i,k}$ be as defined previously.

Preprint submitted to Forty-Eighth Annual Allerton Conference.
Received July 7, 2010.

Theorem 6: Suppose that each of k total packets is distributed in a nonsingular, d-regular, d-connected network according to an i.i.d. process, and that the marginal probability of a node having a packet is ρ for all nodes. If any one of the following conditions are true:

1) $\rho < \frac{1}{d+1}$, or
2) the probability of any subset of nodes of size greater than $N-d$ missing a packet is zero, or
3) packets are distributed according to the independent distribution

then the transmission strategy $\{b_i^j\}$ obtained by choosing b_i^j so that $b_i^j \geq \lfloor \frac{k}{r}(\frac{1-\rho}{d} + \epsilon'_{i,k})\rfloor$ and $\sum_{j=1}^r b_i^j = \lceil k(\frac{1-\rho}{d} + \epsilon'_{i,k})\rceil$ yields a vector $\{b_i^j\}$ that is within N transmissions of optimal and is feasible with arbitrarily high probability (as $k \to \infty$). Moreover, for this $\{b_i^j\}$, we have that:

$$\frac{1}{d}\sum_{i\in[N]} |\mathcal{P}_i^c(k)| \leq \sum_{\substack{i\in[N]\\ j\in[r]}} b_i^j \leq N + \frac{1}{d}\sum_{i\in[N]} |\mathcal{P}_i^c(k)|. \quad (14)$$

IV. PROOFS

This section contains all the proofs of the results given in the previous section. The reader is free to skip this section and continue on to the Concluding Remarks if he/she wishes.

Proof of Theorem 1 : See [1]. ∎

Proof of Theorem 2: Fix $\{b_i\}$ and assume that

$$\sum_{i\in\mathcal{S}^c\cap\mathcal{B}(\mathcal{S})} b_i \geq \tau\left(\mathbb{P}_{\mathcal{S}}^c + \epsilon\right) \quad (15)$$

for some sufficiently small $\epsilon > 0$ and all subsets $\mathcal{S} \subseteq [N]$. Take k sufficiently large so that $\mathbb{P}_{\mathcal{S}}^c \geq \frac{1}{k}\left|\cap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right| - \frac{\epsilon}{2}$ for all $\mathcal{S}$ and also so that $k\frac{\epsilon}{2} \geq \sum_{i=1}^N b_i$. Then, if $b_i^j \geq b_i$, the following string of inequalities together with Theorem 1 proves that $\{b_i^j\}$ permits universal recovery.

$$\begin{aligned}
\sum_{j=1}^r \sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} b_i^j &\geq \sum_{j=1}^r \sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} b_i && (16)\\
&= \sum_{j=1}^r \sum_{i\in\mathcal{S}_{j-1}^c\cap\mathcal{B}(\mathcal{S}_{j-1})} b_i - \sum_{i\in\mathcal{S}_0^c\cap\mathcal{S}_r} b_i && (17)\\
&\geq \tau\sum_{j=1}^r\left(\mathbb{P}_{\mathcal{S}_{j-1}}^c + \epsilon\right) - \sum_{i=1}^N b_i && (18)\\
&\geq r\tau\left(\mathbb{P}_{\mathcal{S}_r}^c + \epsilon\right) - \sum_{i=1}^N b_i && (19)\\
&\geq k\left(\frac{1}{k}\left|\bigcap_{i\in\mathcal{S}_r}\mathcal{P}_i^c(k)\right| + \frac{\epsilon}{2}\right) - \sum_{i=1}^N b_i && (20)\\
&\geq \left|\bigcap_{i\in\mathcal{S}_r}\mathcal{P}_i^c(k)\right|. && (21)
\end{aligned}$$

∎

Proof of Theorem 3: By assumption, we can find some $\epsilon > 0$ so that the following inequality holds:

$$\sum_{i\in\tilde{\mathcal{B}}(\mathcal{S})} b_i \leq \tau\mathbb{P}_{\mathcal{S}}^c - \epsilon. \quad (22)$$

Take $\mathcal{S}_i = \mathcal{S}$ for $i \in [r]\cup\{0\}$, and write:

$$\mathbb{P}_{\mathcal{S}}^c = \frac{1}{k}\left|\bigcap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right| + \epsilon_{\mathcal{S},k}. \quad (23)$$

Then we have the following string of inequalities:

$$\begin{aligned}
\sum_{j=1}^r \sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} b_i^j &= \sum_{j=1}^r \sum_{i\in\mathcal{S}^c\cap\mathcal{B}(\mathcal{S})} b_i^j && (24)\\
&\leq \sum_{j=1}^r \sum_{i\in\tilde{\mathcal{B}}(\mathcal{S})} b_i && (25)\\
&\leq r\left(\tau\mathbb{P}_{\mathcal{S}}^c - \epsilon\right) && (26)\\
&= k\left(\frac{1}{k}\left|\bigcap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right| + \epsilon_{\mathcal{S},k} - \frac{\epsilon}{\tau}\right) && (27)\\
&= \left|\bigcap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right| - k\left(\frac{\epsilon}{\tau} - \epsilon_{\mathcal{S},k}\right) && (28)\\
&< \left|\bigcap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right|. && (29)
\end{aligned}$$

Where the final inequality follows since $\lim_{k\to\infty}|\epsilon_{\mathcal{S},k}| = 0$ by definition. Thus $\{b_i^j\} \notin \mathcal{R}_r$, implying that universal recovery is not possible by Theorem 1. ∎

Proof of Theorem 4: Suppose $\sum_{j=1}^r \hat{b}_i^j = k(\delta_i + \epsilon'_{i,k})$, then $\sum_{i,j}\hat{b}_i^j$ is a lower bound on the minimum number of transmissions required (because each node receives exactly the number of packets it is missing). Therefore, if $\sum_{j=1}^r b_i^j = \lceil k(\delta_i + \epsilon'_{i,k})\rceil$, and $\{b_i^j\}$ is feasible, then it describes a transmission strategy within N transmissions of optimal since $\sum_{i,j} b_i^j \leq \sum_{i,j}\hat{b}_i^j + N$. Thus, it just remains to be shown that $\{b_i^j\}$ is feasible when the conditions of the theorem are met.

Note that we can always choose b_i^j so that $b_i^j \geq \lfloor\frac{k}{r}(\delta_i + \epsilon'_{i,k})\rfloor$ and $\sum_{j=1}^r b_i^j = \lceil k(\delta_i + \epsilon'_{i,k})\rceil$. Set $\epsilon_k = \max_i|\epsilon'_{i,k}|$. By this definition of $\{b_i^j\}$, we have that $\{b_i^j\} \in \mathcal{R}_r$ when $|\mathcal{S}_r| = 1$, therefore, in what follows we only consider $|\mathcal{S}_r| \geq 2$.

By hypothesis, for k sufficiently large we have that the following holds:

$$\frac{1}{r}\sum_{j=1}^r \sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} \delta_i \geq \mathbb{P}_{\mathcal{S}_r}^c + N(\epsilon_k + \frac{r}{k}) + \epsilon_k. \quad (30)$$

Equivalently, we have:

$$\begin{aligned}
\frac{k}{r}\sum_{j=1}^r \sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} \delta_i - N(k\epsilon_k + r) &\geq k\left(\mathbb{P}_{\mathcal{S}_r}^c + \epsilon_k\right) && (31)\\
&\geq \left|\bigcap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right|. && (32)
\end{aligned}$$

Putting everything together yields the following string of inequalities:

$$\sum_{j=1}^{r}\sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} b_i^j \geq \sum_{j=1}^{r}\sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})} \left\lfloor \frac{k}{r}(\delta_i+\epsilon'_{i,k})\right\rfloor \quad (33)$$

$$\geq \frac{k}{r}\sum_{j=1}^{r}\sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})}\delta_i - N(k\epsilon_k+r) \quad (34)$$

$$\geq \left|\bigcap_{i\in\mathcal{S}}\mathcal{P}_i^c(k)\right|. \quad (35)$$

Therefore, $\{b_i^j\}\in\mathcal{R}_r$, and permits universal recovery by Theorem 1. ∎

Proof of Theorem 5: Note that $\delta_i = \frac{1}{d}\mathbb{P}_i^c$. By the hypothesis that the network is d-connected, $|\tilde{\mathcal{B}}(\mathcal{S})| \geq d$ whenever $|\mathcal{S}| \leq N-d$ and $|\tilde{\mathcal{B}}(\mathcal{S}_1)| \geq N-|\mathcal{S}_2|$ whenever $|\mathcal{S}_2| \geq N-d$ and $\mathcal{S}_1\subseteq\mathcal{S}_2$. We consider the cases where $2\leq|\mathcal{S}_r|\leq N-d$ and $N-d<|\mathcal{S}_r|\leq N-1$ separately.

First, note the following inequality:

$$\sum_{j=1}^{r}\sum_{i\in\mathcal{S}_j^c\cap\mathcal{B}(\mathcal{S}_{j-1})}\delta_i = \sum_{j=1}^{r}\sum_{i\in\tilde{\mathcal{B}}(\mathcal{S}_{j-1})}\delta_i - \sum_{i\in\mathcal{S}_r\cap\mathcal{S}_0^c}\delta_i \quad (36)$$

$$> \sum_{j=1}^{r}\sum_{i\in\tilde{\mathcal{B}}(\mathcal{S}_{j-1})}\delta_i - N\delta_i \quad (37)$$

$$= \frac{1}{d}\mathbb{P}_i^c\left(\sum_{j=1}^{r}\left|\tilde{\mathcal{B}}(\mathcal{S}_{j-1})\right| - N\right). \quad (38)$$

Thus, in order to verify the conditions of Theorem 4, it is sufficient to check that

$$\frac{1}{d}\mathbb{P}_i^c\left(\sum_{j=1}^{r}\left|\tilde{\mathcal{B}}(\mathcal{S}_{j-1})\right| - N\right) \geq r\cdot\mathbb{P}_{\mathcal{S}_r}^c. \quad (39)$$

Or equivalently, that the following holds:

$$\frac{1}{rd}\sum_{j=1}^{r}\left|\tilde{\mathcal{B}}(\mathcal{S}_{j-1})\right| \geq \frac{\mathbb{P}_{\mathcal{S}_r}^c}{\mathbb{P}_i^c}+\frac{N}{rd}. \quad (40)$$

Assume that $1<|\mathcal{S}_r|\leq N-d$, then we have the following:

$$\frac{1}{rd}\sum_{j=1}^{r}\left|\tilde{\mathcal{B}}(\mathcal{S}_{j-1})\right| \geq \frac{1}{rd}\sum_{j=1}^{r} d \quad (41)$$

$$= 1 \quad (42)$$

$$\geq \frac{\mathbb{P}_{\mathcal{S}_r}^c}{\mathbb{P}_i^c}+\frac{N}{rd}. \quad (43)$$

Where the last inequality holds if r is taken sufficiently large (how large depends only on N, d, and $\max_{\mathcal{S}:|\mathcal{S}|>1}\mathbb{P}_{\mathcal{S}}^c$).

Next, assume that $|\mathcal{S}_r|>N-d$, then by hypothesis, we have:

$$\frac{1}{rd}\sum_{j=1}^{r}\left|\tilde{\mathcal{B}}(\mathcal{S}_{j-1})\right| \geq \frac{N-|\mathcal{S}_r|}{d} \quad (44)$$

$$\geq \frac{\mathbb{P}_{\mathcal{S}_r}^c}{\mathbb{P}_i^c}+\frac{N}{rd}. \quad (45)$$

Where the last inequality again holds for sufficiently large r (depending only on N, d, and $\max_{\mathcal{S}:|\mathcal{S}|>N-d}\mathbb{P}_{\mathcal{S}}^c$).

The bounds on $\sum_{i,j} b_i^j$ are a result of direct computation. ∎

Proof of Corollary 1: Note that $\mathbb{P}_i < \frac{1}{d+1}$ implies that $\mathbb{P}_i^c > \frac{d}{d+1}$. Then, note that we have the following string of inequalities:

$$d\cdot\mathbb{P}_{\mathcal{S}}^c \leq d\cdot\mathbb{P}_{\mathcal{S}^c} \quad (46)$$

$$\leq d\cdot\sum_{i\in\mathcal{S}^c}\mathbb{P}_i \quad (47)$$

$$= d\cdot(N-|\mathcal{S}|)\cdot\mathbb{P}_i \quad (48)$$

$$< \frac{d}{d+1}\cdot(N-|\mathcal{S}|) \quad (49)$$

$$< \mathbb{P}_i^c\cdot(N-|\mathcal{S}|). \quad (50)$$

Thus, $d\cdot\mathbb{P}_{\mathcal{S}}^c < (N-|\mathcal{S}|)\cdot\mathbb{P}_i^c$ and the conditions of Theorem 5 are satisfied. ∎

Proof of Corollary 2: This trivially implies that $\mathbb{P}_{\mathcal{S}}^c = 0$. Thus, $d\cdot\mathbb{P}_{\mathcal{S}}^c = 0 < (N-|\mathcal{S}|)\cdot\mathbb{P}_i^c$ and the conditions of Theorem 5 are satisfied. ∎

Lemma 1: If $0<q<1$ is fixed, then there exists some $\delta>0$ such that the following inequality holds for all $\ell\in\{2,\ldots,N-1\}$:

$$\frac{N-\ell}{N-1} \geq \frac{(1-q)^\ell-(1-q)^N}{1-q-(1-q)^N}+\delta. \quad (51)$$

Proof: We write $\ell = \theta\cdot 1+(1-\theta)\cdot N$, where $\theta=\frac{N-\ell}{N-1}$. Since $(1-q)^x$ is strictly convex in x, by Jensen's inequality we have:

$$\frac{(1-q)^\ell-(1-q)^N}{1-q-(1-q)^N} < \frac{N-\ell}{N-1}. \quad (52)$$

Taking

$$\delta = \min_{\ell\in\{2,\ldots,N-1\}}\left[\frac{N-\ell}{N-1}-\frac{(1-q)^\ell-(1-q)^N}{1-q-(1-q)^N}\right] \quad (53)$$

completes the proof. ∎

Proof of Corollary 3: Under the assumption that

$$\mathbb{P}_{\mathcal{S}} = \frac{1-(1-q)^{|\mathcal{S}|}}{1-(1-q)^N}, \quad \forall\mathcal{S}\subseteq[N], \quad (54)$$

we have the following by Lemma 1:

$$\frac{N-|\mathcal{S}|}{d} \geq \frac{N-|\mathcal{S}|}{N-1} \quad (55)$$

$$> \frac{(1-q)^{|\mathcal{S}|}-(1-q)^N}{(1-q)-(1-q)^N} \quad (56)$$

$$= \frac{\mathbb{P}_{\mathcal{S}}^c}{\mathbb{P}_i^c}. \quad (57)$$

Thus, $d \cdot \mathbb{P}_{\mathcal{S}}^c < (N - |\mathcal{S}|) \cdot \mathbb{P}_i^c$ and the conditions of Theorem 5 are satisfied. ■

Proof of Theorem 6: The theorem is an immediate consequence of Corollaries 1-3 and the weak law of large numbers. ■

V. Concluding Remarks

This paper studies an arbitrarily connected broadcast network of N nodes that all wish to recover k desired packets originally dispersed among the nodes. In this paper, we present a series of relatively simple conditions for universal recovery that apply when the number of packets is large and the distribution of packets among the nodes is well behaved.

For any fixed network topology, it suffices to consider only simple cuts in the network to characterize a set of transmission strategies that requires at most $k\epsilon$ transmissions above the minimum required for universal recovery (ϵ can be arbitrarily small). For certain topologies including d-regular d-connected networks, we explicitly construct transmission strategies that achieve universal recovery while using at most N transmissions above the minimum number required, thus essentially resolving the problem completely for many canonical networks (e.g. cliques, rings, grids on tori, etc.).

Acknowledgment

This research was supported by Rockwell Collins through contract #4502769987. The authors would like to thank Kent Benson of Rockwell Collins for proposing the investigation that led to this paper. We also enjoyed numerous useful conversations with Pavan Datta and Chi-Wen Su.

References

[1] T. Courtade, B. Xie, and R. Wesel, "Optimal Exchange of Packets for Universal Recovery in Broadcast Networks," To appear at MILCOM 2010, San Jose, CA, October 31 - November 3, 2010. Preprint available online at http://www.ee.ucla.edu/∼csl/files/publications.html.

[2] N. Alon, A. Hassidim, E. Lubetzky, U. Stav and A. Weinstein, Broadcasting with side information, Proc. of the 49th IEEE FOCS (2008), 823-832.

[3] Z. Bar-Yossef, Y. Birk, T.S. Jayram and T. Kol, Index coding with side information, Proc. of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pp. 197-206.

[4] Y. Birk and T. Kol, Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients, IEEE Transactions on Information Theory 52 (2006), 2825-2830.

[5] H. S. Witsenhausen, The zero-error side information problem and chromatic numbers, IEEE Transactions on Information Theory, 22(5) (1976), 592-593.

[6] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network information flow", IEEE Transactions on Information Theory, July 2000.

[7] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding", IEEE Transactions on Information Theory, Feb. 2003.

[8] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, "A network coding approach to energy efficient broadcasting: From theory to practice", in IEEE INFOCOM, Barcelona, Spain, Apr. 2006.

[9] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, "Efficient Broadcasting Using Network Coding," IEEE/ACM Transactions on Networking, Vol. 16, No. 2, April 2008, 450-463.

[10] Y. Wu, J. Padhye, R. Chandra, V. Padmanabhan, and P. A. Chou. The local mixing problem. In Proc. Information Theory and Applications Workshop, San Diego, Feb. 2006.

[11] S. El Rouayheb, M.A.R. Chaudhry, and A. Sprintson. On the minimum number of transmissions in single-hop wireless coding networks. In IEEE Information Theory Workshop (Lake Tahoe), 2007.

[12] E. Lubetzky and U. Stav. Non-linear index coding outperforming the linear optimum. In Proc. of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 161167, 2007.

[13] S. El Rouayheb, A. Sprintson, and C. N. Georghiades. On the relation between the index coding and the network coding problems. Proc. of IEEE International Symposium on Information Theory (ISIT08), 2008.

[14] S El Rouayheb, A Sprintson, P Sadeghi, "On Coding for Cooperative Data Exchange" - Arxiv preprint arXiv:1002.1465, 2010.

[15] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, "Polynomial Time Algorithms for Multicast Network Code Construction, IEEE Transactions on Information Theory, vol. 51, no. 6, pp. 19731982, 2005.

[16] S. Mohajer, C. Tian, and S. Diggavi, "On Source Transmission over Deterministic Relay Networks," IEEE Information Theory Workshop, Cairo, Egypt, January 6-8, 2010.

[17] B. Bollobás, Random Graphs, 2nd edition, Cambridge University Press (2001).

[18] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.

[19] Fekete, M. "Uber die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit. ganzzahligen Koeffizienten." Mathematische Zeitschrift 17 (1923), pp. 228-249.