

# On the Decoding Failure Probability of Random Network Coded Cooperation

Ioannis Chatzigeorgiou

School of Computing  
and Communications

Lancaster University

Lancaster LA1 4WA, UK

Email: i.chatzigeorgiou@lancaster.ac.uk

Günes Kurt, Semiha Tedik Başaran

Department of Electronics  
and Communication Engineering

Istanbul Technical University

34469 Istanbul, Turkey

Email: {gkurt,tedik}@itu.edu.tr

Amjad Saeed Khan

School of Mechanical, Electrical  
and Manufacturing Engineering

Loughborough University

Loughborough LE11 3TU, UK

Email: a.khan@lboro.ac.uk

**Abstract**—This paper considers a network of source nodes that transmit data packets to a destination node via relay nodes over erasure channels by using random linear network coding. The probability that the destination node will fail to recover the packets of all source nodes has been bounded and approximated in the literature for the case of relay nodes that randomly assign only nonzero values to the coefficients of linear combinations of data packets. The paper argues for the necessity of giving relay nodes the choice to also assign the zero value to coefficients when arithmetic operations are over finite fields of small size, e.g. GF(2). Alternative probability mass functions for the coefficients are considered, and expressions for the decoding failure probability are re-derived.

**Index Terms**—Cooperative communication, decoding probability, random linear network coding, random matrices.

## I. INTRODUCTION

Network coding, proposed in the seminal work of Ahlswede *et al.* [1], can increase data flow rates in a network. When network coding is employed, source nodes transmit data packets, whereas relay nodes generate and forward linear combinations of successfully received packets toward destination nodes. Wireless links between nodes can introduce packet erasures in the transmission process and affect network topology. Ho *et al.* [2] proposed the use of a *random linear network code* (RLNC) for networks with varying or unknown topologies, according to which the coefficients in a linear combination of packets are selected uniformly and independently from a Galois field of size  $q$ , denoted by  $\text{GF}(q)$ .

Chiasserini *et al.* [3] studied a network that consists of  $N$  source nodes,  $M$  relay nodes and one destination node. Each source node transmits one data packet. Source-to-destination links are not available. An approximate expression for the *decoding probability* was obtained, that is, the probability that the destination node will decode the data packets of all source nodes. Başaran *et al.* [4] took into account the source-to-destination links, improved the counting accuracy of the combinatorial solution presented in [3] and derived an exact expression for the decoding probability. The expressions in both [3] and [4] invoke the  $\text{mat}_q(\cdot)$  function, the calculation of which is based on a numerical method by Klein *et al.* [5]. As a result, the exact solution proposed in [3] and [4] is both semi-analytical and computationally intensive, and cannot be

easily integrated in optimization problems. Furthermore, both [3] and [4] assume that each relay node fails to receive at most one data packet, thus limiting the scope of the solution.

Seong [6] obtained closed-form bounds on the *decoding failure probability* (DFP), that is, the probability that the destination node will fail to decode the data packets of all source nodes. The key difference between [6] and [3], [4] is that the whole network in [6] has been modeled as an RLNC in which the values of code coefficients are selected with unequal probability from  $\text{GF}(q)$ . When code coefficients are more likely to be assigned the zero value than any of the other  $q - 1$  values of  $\text{GF}(q)$ , the code is known as *sparse* RLNC, otherwise it is referred to as *dense* RLNC. Based on the work of Blömer *et al.* [7] on sparse random matrices, Khan *et al.* [8] derived tighter bounds than those in [6]. However, both [6] and [8] study networks without direct source-to-destination links. Both [6] and [8] assume that relay nodes randomly assign only *nonzero* values to code coefficients in linear combinations of successfully received data packets. This assumption can introduce flaws when the erasure probability of the source-to-relay links is very low and RLNC is over  $\text{GF}(2)$ , resulting in a high DFP.

In this paper, the work of [6] and [8] is revisited by jointly considering the relationship between the erasure probability of the source-to-relay links and the probability mass function (PMF) of the code coefficients. The bounds on the DFP are re-derived. The work of Brown *et al.* [9], which proposes an approximation for the decoding failure probability, is taken into account and is extended from broadcast to relay-assisted communication. Our analysis initially focuses on transmission via the relay nodes only, referred to as *Random Network Coded Relaying* (RNCR), and is then extended to include the support of source-to-destination links, known as *Random Network Coded Cooperation* (RNCC).

## II. SYSTEM MODEL

As shown in Fig. 1, we consider a network that consists of source nodes  $S_1, \dots, S_N$ , relay nodes  $R_1, \dots, R_M$  and one destination node  $D$ . The transmission process is completed in two phases. In the *broadcast phase*, each source node transmits a data packet to the destination node and the relay nodes.

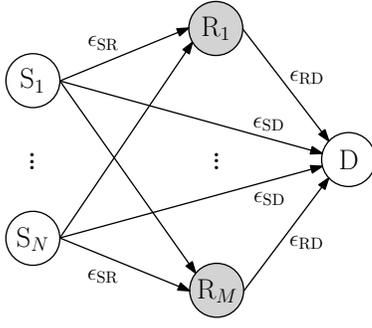


Figure 1. Model of a network consisting of  $N$  source nodes,  $M$  relay nodes and one destination node. The packet erasure probabilities  $\epsilon_{SR}$ ,  $\epsilon_{RD}$  and  $\epsilon_{SD}$  characterize the quality of the source-to-relay, relay-to-destination and source-to-destination links, respectively.

In the *relaying phase*, each relay node randomly combines successfully received data packets, generates a coded packet and forwards it to the destination. Packets are transmitted over orthogonal channels. The average channel conditions of the source-to-relay, relay-to-destination and source-to-destination links are characterized by the packet erasure probabilities  $\epsilon_{SR}$ ,  $\epsilon_{RD}$  and  $\epsilon_{SD}$ , respectively.

The packets that are transmitted during the two-phase process can be written in matrix form as:

$$\mathbf{x} = \mathbf{G}\mathbf{u} \Leftrightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_{N+M} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N \times N} \\ \mathbf{C}_{M \times N} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}. \quad (1)$$

The column vector  $\mathbf{u}$  contains the data packets  $u_1, \dots, u_N$  of the source nodes, where  $u_n$  is the data packet of source node  $S_n$ , for  $1 \leq n \leq N$ . The column vector  $\mathbf{x}$  consists of the transmitted packets  $x_1, \dots, x_{N+M}$ , where  $x_1, \dots, x_N$  are the data packets transmitted by the  $N$  source nodes to the destination node, and  $x_{N+1}, \dots, x_{N+M}$  are the coded packets transmitted by the  $M$  relay nodes to the destination node.  $\mathbf{G}$  is obtained by the vertical concatenation of the  $N \times N$  identity matrix  $\mathbf{I}_{N \times N}$  and the  $N \times M$  coding matrix  $\mathbf{C}_{M \times N}$ . Note that  $\mathbf{I}_{N \times N}$  represents the broadcast phase and  $\mathbf{C}_{M \times N}$  describes the encoding process of the data packets during the relaying phase. Each element of  $\mathbf{C}_{M \times N}$ , denoted by  $c_{m,n}$ , is randomly drawn from  $\text{GF}(q)$ , that is,  $c_{m,n} \in \text{GF}(q)$ . The PMF of  $c_{m,n}$ , represented by  $P(c_{m,n})$ , is influenced by the erasure probability of the source-to-relay links. For example, the event of a data packet being transmitted from source node  $S_n$  to relay node  $R_m$  and subsequently erased by the channel is equivalent to the event of relay node  $R_m$  setting  $c_{m,n} = 0$  in the encoding process. If erasures have an impact on the probability of selecting  $c_{m,n} = 0$ , then non-erasures affect the probability of choosing  $1 \leq c_{m,n} \leq q$ . Hence, the PMF  $P(c_{m,n})$  depends on  $\epsilon_{SR}$ .

At the end of the two-phase process, node D constructs matrix  $\mathbf{A}$  from the rows of  $\mathbf{G}$  that are associated with successfully received packets. The number of rows that sub-matrix  $\mathbf{I}_{N \times N}$  contributes to  $\mathbf{A}$  is affected by  $\epsilon_{SD}$ . Similarly, the number of rows of sub-matrix  $\mathbf{C}_{M \times N}$  that will be appended to  $\mathbf{A}$

depends on  $\epsilon_{RD}$ . Node D will recover  $u_1, \dots, u_N$  if and only if  $\mathbf{A}$  contains  $N$  linearly independent rows, thus, has rank  $N$ .

### III. RANDOM NETWORK CODED RELAYING

Let  $E_{m,n}$  denote the erasure of a data packet transmitted from source node  $S_n$  to relay node  $R_m$ , and let  $\bar{E}_{m,n}$  signify the successful delivery of the same packet, where the probability of each event is  $P(E_{m,n}) = \epsilon_{SR}$  and  $P(\bar{E}_{m,n}) = 1 - \epsilon_{SR}$ . The probability that element  $c_{m,n}$  of the coding matrix  $\mathbf{C}_{M \times N}$  will be assigned a particular value from  $\text{GF}(q)$  is given by

$$P(c_{m,n}) = P(c_{m,n} | E_{m,n})\epsilon_{SR} + P(c_{m,n} | \bar{E}_{m,n})(1 - \epsilon_{SR}). \quad (2)$$

In the event of an erasure, data packet  $u_n$  of source node  $S_n$  will not be stored on the memory of relay node  $R_m$  and will not participate in the encoding process. This is equivalent to setting  $c_{m,n} = 0$  in (1) and writing  $P(c_{m,n} = 0 | E_{m,n}) = 1$ .

Given that  $\epsilon_{SD} = 1$  and  $\mathbf{G} = \mathbf{C}_{M \times N}$  in RNCR, the DFP can be expressed as [6], [8], [9]

$$F_{\text{RNCR}}(N, M) = \sum_{m=0}^M \binom{M}{m} \epsilon_{RD}^{M-m} (1 - \epsilon_{RD})^m F(N | m) \quad (3)$$

where  $F(N | m)$  is the DFP conditioned on the number of coded packets  $m$  that have been received by the destination. If node D receives  $m$  of the  $M$  coded packets, the dimensions of  $\mathbf{A}$  will be  $m \times N$ . The conditional probability  $F(N | m)$ , which has been bounded in [6], [8] and approximated in [9], is dependent on  $P(c_{m,n})$  given in (2).

The analysis in [6] and [8] assumes that if data packet  $u_n$  is received by relay node  $R_m$ , it will *definitely* contribute to the generation of coded packet  $x_m$ . This implies that the relay node will select a value uniformly at random from the  $q - 1$  nonzero elements of  $\text{GF}(q)$  and assign it to  $c_{m,n}$ . We can thus write  $P(c_{m,n} = t | \bar{E}_{m,n}) = 1/(q - 1)$  for  $t \in \text{GF}(q) \setminus \{0\}$ . Substitution into (2) gives

$$P(c_{m,n} = t) = \begin{cases} \epsilon_{SR}, & \text{if } t = 0 \\ (1 - \epsilon_{SR})/(q - 1), & \text{if } t \in \text{GF}(q) \setminus \{0\}. \end{cases} \quad (4)$$

Based on (4), the following upper bound on the conditional DFP was obtained in [8, Theorem 3]:

**Theorem 1** (Khan *et al.* [8]). *If the PMF used in the encoding process of RNCR is defined as in (4), the conditional DFP is upper bounded by*

$$F(N | m) \leq \min \{F_1(N | m), F_2(N | m)\} \quad (5)$$

where

$$F_1(N | m) = 1 - \prod_{n=1}^N \left[ 1 - \left( \max \left\{ \epsilon_{SR}, \frac{1 - \epsilon_{SR}}{q - 1} \right\} \right)^{m - n + 1} \right] \quad (6)$$

$$F_2(N | m) = \sum_{n=1}^N \binom{N}{n} (q - 1)^{n-1} \rho_n^m, \quad (7)$$

with  $\rho_n$  in (7) given by

$$\rho_n = q^{-1} + (1 - q^{-1}) \left( 1 - \frac{1 - \epsilon_{SR}}{1 - q^{-1}} \right)^n. \quad (8)$$

Expression (5) selects the tightest of two different loose upper bounds, namely  $F_1(N|m)$  and  $F_2(N|m)$ . Upper bound  $F_2(N|m)$  in (7) is a function of  $\rho_n$ , which is defined in (8) and provides the probability that the sum over  $\text{GF}(q)$  of the first  $n$  elements of a row of matrix  $\mathbf{A}$  is zero. The same PMF as in (4) was used in [9] but the Stein-Chen method was followed to approximate the DFP. In an effort to facilitate the comparison of the findings in [8] and [9], we present [9, Theorem 3.1] in the context of RNCR, as:

**Theorem 2** (Brown *et al.* [9]). *If the PMF used in the encoding process of RNCR with  $N$  source nodes and  $M \geq N$  relay nodes is defined as in (4), the conditional DFP can be approximated by*

$$F(N|m) \approx 1 - (1 - \epsilon_{\text{SR}}^m)^N \exp\left(-\sum_{n=2}^N \binom{N}{n} \frac{\pi_n}{(1 - \epsilon_{\text{SR}}^m)^n}\right) \quad (9)$$

where  $\pi_n$  is an approximation of the probability that a subset of  $n$  columns of matrix  $\mathbf{A}$  add up to the zero vector in  $\text{GF}(q)$  but no combination of columns from the same subset sum to the zero vector. The following recursion holds for  $\pi_n$ :

$$\pi_n = \rho_n^m - \sum_{s=1}^{n-1} \binom{n-1}{s} \rho_s^m \pi_{n-s} \quad (10)$$

where  $\rho_n$  has been defined in (8) and  $\pi_1 = \rho_1^m$ .

As observed in [8] and also reported in [10], if  $\epsilon_{\text{RD}} < 1$ , the DFP is minimized when the values assigned to the elements of the coding matrix  $\mathbf{C}_{M \times N}$  are selected uniformly at random from  $\text{GF}(q)$ , i.e., with probability  $1/q$ . According to (4), the PMF follows the discrete uniform distribution when  $\epsilon_{\text{SR}} = 1/q$ . For  $q = 2$ , the system design has the undesirable attribute of generating markedly more decoding failures when  $0 \leq \epsilon_{\text{SR}} < 1/2$  than when  $\epsilon_{\text{SR}} = 1/2$ . In order to alleviate this problem and reduce the DFP, we need to revise (4).

#### A. Relay nodes without knowledge of $\epsilon_{\text{SR}}$

If data packet  $u_n$  is received by relay node  $R_m$ , element  $c_{m,n}$  of the coding matrix  $\mathbf{C}_{M \times N}$  can be assigned values from  $\text{GF}(q)$  with equal probability, including the zero value, i.e.,  $P(c_{m,n} | \bar{\mathbf{E}}_{m,n}) = 1/q$ . This will ensure that, if  $0 \leq \epsilon_{\text{SR}} < 1/q$ , data packets that have been received by a relay node will not necessarily be included in the RLNC process. Thus, dense RLNC will be avoided and the chances of generating linearly dependent coded packets at low values of  $\epsilon_{\text{SR}}$  will be reduced. Based on (2), the PMF of  $c_{m,n}$  becomes

$$P(c_{m,n} = t) = \begin{cases} \epsilon_{\text{SR}} + (1 - \epsilon_{\text{SR}})/q, & \text{if } t = 0 \\ (1 - \epsilon_{\text{SR}})/q, & \text{if } t \in \text{GF}(q) \setminus \{0\}. \end{cases} \quad (11)$$

The following proposition is an adjustment of Theorem 1:

**Proposition 1.** *If the PMF used in the encoding process of RNCR with  $N$  source nodes and  $M \geq N$  relay nodes is defined as in (11), the conditional DFP is upper bounded by (5), where*

$$F_1(N|m) = 1 - \prod_{n=1}^N \left[ 1 - \left( \epsilon_{\text{SR}} + \frac{1 - \epsilon_{\text{SR}}}{q} \right)^{m-n+1} \right], \quad (12)$$

$F_2(N|m)$  is defined as in (7) but  $\rho_n$  is given by

$$\rho_n = q^{-1} + (1 - q^{-1}) \epsilon_{\text{SR}}^n. \quad (13)$$

*Proof.* According to the PMF in (11), the zero value will always be selected with a higher probability than any nonzero value from  $\text{GF}(q)$ , for any value of  $\epsilon_{\text{SR}}$ . Therefore, the maximum of the two branches of the PMF is  $\epsilon_{\text{SR}} + (1 - \epsilon_{\text{SR}})/q$ . Substitution into (6) gives (12). The expression of  $\rho_n$  for the revised PMF in (11) can be derived if the same reasoning as in [7, Theorem 3.3] is applied. In particular, the sum of the first  $n$  elements of a row will be zero if (i) the sum of the first  $n-1$  elements is zero and the  $n$ -th element has the zero value, or (ii) the sum of the first  $n-1$  elements is greater than zero but the  $n$ -th element has the appropriate nonzero value that gives a sum of zero. In terms of probabilities, we can write:

$$\begin{aligned} \rho_n &= \rho_{n-1} \left( \epsilon_{\text{SR}} + \frac{1 - \epsilon_{\text{SR}}}{q} \right) + (1 - \rho_{n-1}) \left( \frac{1 - \epsilon_{\text{SR}}}{q} \right) \\ &= \rho_{n-1} \epsilon_{\text{SR}} + (1 - \epsilon_{\text{SR}}) q^{-1} \end{aligned} \quad (14)$$

where  $\rho_0 = 1$ . If we set  $\phi_n = \rho_n - q^{-1}$ , it follows from (14) that  $\phi_0 = 1 - q^{-1}$  and

$$\phi_n = \phi_{n-1} \epsilon_{\text{SR}} = (1 - q^{-1}) \epsilon_{\text{SR}}^n. \quad (15)$$

Replacing  $\phi_n$  with the righthand side of (15) in  $\phi_n = \rho_n - q^{-1}$  and solving for  $\rho_n$  leads to (13).  $\square$

**Corollary 1.** *When RNCR uses the PMF in (11), the conditional DFP can be approximated by (9) if  $\rho_n$  in (10) is defined as in (13).*

We established that the PMF in (11) results in sparse RNCR because selection of the zero element is favored over the other elements of  $\text{GF}(q)$  in the encoding process. Thus, for  $0 \leq \epsilon_{\text{SR}} < 1/q$ , the PMF in (11) is closer to the (optimal) uniform distribution than the PMF in (4) and a lower DFP can be achieved. For  $1/q \leq \epsilon_{\text{SR}} \leq 1$ , both (11) and (4) lead to sparse implementations of RNCR; however, the PMF in (11) deviates from the uniform distribution more than the PMF in (4) and yields a higher DFP. The benefits of both (11) and (4) can be reaped for any value of  $\epsilon_{\text{SR}}$ , if relay nodes know  $\epsilon_{\text{SR}}$ .

#### B. Relay nodes with knowledge of $\epsilon_{\text{SR}}$

The packet erasure probability  $\epsilon_{\text{SR}}$  captures the average channel conditions of the source-to-relay links. If the relay nodes have knowledge of  $\epsilon_{\text{SR}}$ , they can ensure that the values assigned to the elements of  $\mathbf{C}_{M \times N}$  are selected uniformly at random from  $\text{GF}(q)$ , when  $0 \leq \epsilon_{\text{SR}} < 1/q$ . Otherwise, when  $1/q \leq \epsilon_{\text{SR}} < 1$ , the relay nodes can remove the zero value from the pool of candidate values for the elements of  $\mathbf{C}_{M \times N}$ .

For  $0 \leq \epsilon_{\text{SR}} < 1/q$ , the probability that relay  $R_m$  will select a particular value for  $c_{m,n}$  upon successful reception of data packet  $u_n$ , so that  $P(c_{m,n}) = 1/q$ , can be obtained from (2):

$$P(c_{m,n} = t | \bar{\mathbf{E}}_{m,n}) = \begin{cases} \frac{1 - q \epsilon_{\text{SR}}}{q(1 - \epsilon_{\text{SR}})}, & \text{if } t = 0 \\ q(1 - \epsilon_{\text{SR}})^{-1}, & \text{if } t \in \text{GF}(q) \setminus \{0\}. \end{cases} \quad (16)$$

If relay nodes use (16) to assign values to those elements of  $\mathbf{C}_{M \times N}$  that correspond to received data packets, the PMF in (2) will follow the discrete uniform distribution and the conditional DFP will be equal to [11]

$$F(N|m) = 1 - \prod_{n=1}^N (1 - q^{-m+n-1}). \quad (17)$$

For  $1/q < \epsilon_{SR} \leq 1$ , packet erasures can still set the elements of  $\mathbf{C}_{M \times N}$  equal to zero but relay nodes should only assign nonzero values to avoid turning  $\mathbf{C}_{M \times N}$  into an overly sparse matrix. Thus, the PMF in (4) and the upper bound on the conditional DFP given in (5) can be used for  $1/q < \epsilon_{SR} \leq 1$ .

#### IV. RANDOM NETWORK CODED COOPERATION

In the case of RNCC, matrix  $\mathbf{G}$  is the vertical concatenation of  $\mathbf{I}_{N \times N}$  and  $\mathbf{C}_{M \times N}$ , as in (1). Let us assume that the destination node D has collected  $n$  data packets directly from the  $N$  source nodes and  $m$  coded packets from the  $M$  relay nodes. Elementary row and column operations can split the  $(n+m) \times N$  matrix  $\mathbf{A}$  at the destination node into four submatrices, with the top-left being the  $n \times n$  identity matrix, the bottom-right being an  $m \times (N-n)$  random matrix and the remaining entries being zero. As a result, the probability that the destination node will fail to decode the remaining  $N-n$  data packets, given that  $m$  coded packets have been received, is equal to  $F(N-n|m)$ . The DFP can be obtained by averaging  $F(N-n|m)$  over all  $n$  and  $m$ , that is

$$F_{\text{RNCC}}(N, M) = \sum_{n=0}^N \binom{N}{n} \epsilon_{\text{SD}}^{N-n} (1 - \epsilon_{\text{SD}})^n \cdot \sum_{m=0}^M \binom{M}{m} \epsilon_{\text{RD}}^{M-m} (1 - \epsilon_{\text{RD}})^m F(N-n|m) \quad (18)$$

where  $F(N-n|m) = 0$  for  $n = N$ . Note that  $F_{\text{RNCC}}(N, M)$  in (18) reduces to  $F_{\text{RNCR}}(N, M)$  in (3) when  $\epsilon_{\text{SD}} = 1$ . Expression (18) is an extension of the decoding probability of systematic network coding [12, eq. (8)] to the system model of RNCC. Depending on the PMF that is used by relay nodes to assign values to the elements of  $\mathbf{C}_{M \times N}$ , expressions (5), (9) and (17) can be used to upper bound, approximate or compute  $F(N-n|m)$ , respectively, as explained in Section III.

#### V. RESULTS

The proposed modifications in the PMF used in the encoding process of RNCR and RNCC, their impact on the DFP and the tightness of the revised theoretical expressions are explored in this section. To refer to the different implementations of RNCR and RNCC, we have adopted a naming convention that describes the lack or excess of zeros in the coding matrix  $\mathbf{C}_{M \times N}$ , when  $\epsilon_{SR} \in [0 \dots 1/q]$  or  $\epsilon_{SR} \in (1/q \dots 1]$ , respectively. Three cases can be encountered:

- C1: *dense-sparse*, if the PMF in (4) is used to assign values to the elements of  $\mathbf{C}_{M \times N}$ .
- C2: *sparse-very sparse*, if the PMF in (11) is used to assign values to the elements of  $\mathbf{C}_{M \times N}$ .

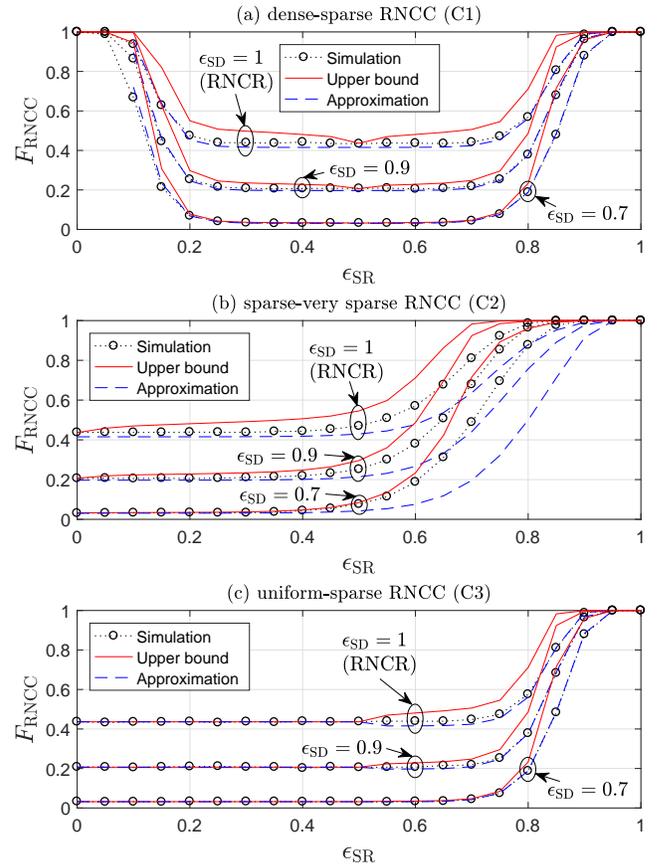


Figure 2. Decoding failure probability of RNCC as a function of  $\epsilon_{SR}$  for  $q = 2$ ,  $N = 20$ ,  $M = 25$ ,  $\epsilon_{RD} = 0.15$  and  $\epsilon_{SD} \in \{0.7, 0.9, 1\}$ .

- C3: *uniform-sparse*, if (16) is used to ensure that the PMF is uniform when  $\epsilon_{SR} \in [0 \dots 1/q]$ , and the PMF in (4) is employed when  $\epsilon_{SR} \in (1/q \dots 1]$ .

Fig. 2 shows the impact of each PMF on the DFP of both RNCR ( $\epsilon_{SD} = 1$  by definition) and RNCC (when  $\epsilon_{SD} \in \{0.7, 0.9\}$ ) for  $N = 20$  source nodes,  $M = 25$  relay nodes,  $q = 2$ ,  $\epsilon_{RD} = 0.15$  and  $\epsilon_{SR} \in [0, 1]$ . Fig. 2(a) illustrates the poor performance of C1, analyzed in [6], [8], for low values of  $\epsilon_{SR}$ . As can be observed in Fig. 2(a) for  $\epsilon_{SR} < 0.2$ , the destination node will fail with a high probability to collect a sufficient number of linearly independent coded packets and recover all of the data packets. The high DFP for low values of  $\epsilon_{SR}$  necessitates the adoption of a different PMF for the encoding process. In C2, relay nodes can assign 0 or 1 with equal probability to code coefficients of received data packets. As can be seen in Fig. 2(b) for low values of  $\epsilon_{SR}$ , the DFP drops to 0.44 for  $\epsilon_{SD} = 1$ , to 0.2 for  $\epsilon_{SD} = 0.9$  and to 0.03 for  $\epsilon_{SD} = 0.7$ . For an increasing value of  $\epsilon_{SR}$ , code coefficients are more likely to be assigned the value 0 as per (11), hence fewer data packets are involved in linear combinations. In C3, the PMF of code coefficients in (16) prompts relay nodes to favor 1 over 0 with an increasing probability, as  $\epsilon_{SR}$  grows from 0 to 0.5. The overall PMF is uniform and the DFP is minimized for  $\epsilon_{SR} \in [0, 0.5]$ , as confirmed by Fig. 2(c). If  $\epsilon_{SR} \in (0.5, 1]$ , both C1 and C3 use the same PMF.

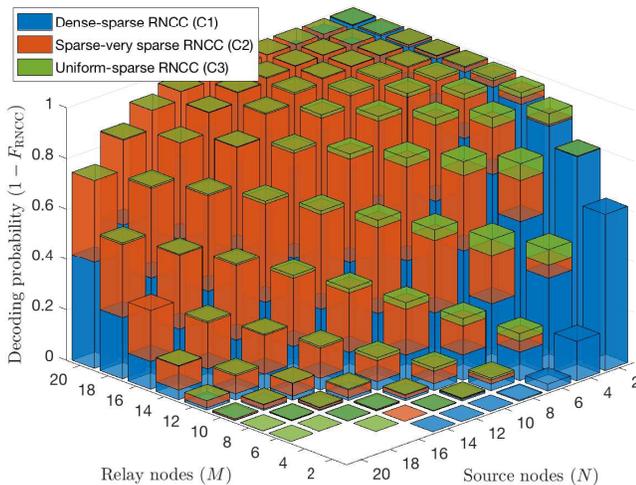


Figure 3. Decoding probability achieved by the three RNCC implementations for  $q = 2$ ,  $\epsilon_{SR} = \epsilon_{RD} = 0.15$ ,  $\epsilon_{SD} = 0.7$  and  $N, M \in \{2, 4, \dots, 20\}$ .

The decoding probability, given by  $1 - F_{RNCC}$ , at the destination node for each of the three cases under consideration is shown in Fig. 3. In particular, the gain in decoding probability that can be achieved by the proposed schemes, C2 and C3, over C1 is depicted as a function of the number of source nodes and relay nodes in the system, where  $N, M \in \{2, 4, \dots, 20\}$ ,  $q = 2$ ,  $\epsilon_{SR} = \epsilon_{RD} = 0.15$  and  $\epsilon_{SD} = 0.7$ . As expected, the decoding probability is low when the source-to-destination links are weak and the relay nodes are fewer than the source nodes ( $M < N$ ); this is because the packets received by the destination node are, occasionally, fewer than the source packets or the linearly independent packets among the received packets are, often, fewer than  $N$ . As only three useful (out of four possible) linear combinations of two source packets can be generated by relay nodes when  $N = 2$  source nodes, C2 and C3 achieve only a marginal increase in the decoding probability, as  $M$  grows. However, we observe that both C2 and C3 significantly improve the decoding probability when  $2 < N \leq M$  but the superiority of C3 is notable only for low values of  $M$ . This observation leads to the practical recommendation that C2 can be used in place of C3 as  $N$  increases,  $N < M$  and  $\epsilon_{SR} < 0.5$ , because the complexity introduced in C3 in order to equip the relay nodes with knowledge of  $\epsilon_{SR}$  provides negligible performance gains.

We note that, as the field size used by RLNC grows, the number of possible linear combinations that can be generated by relay nodes increases. The use or omission of zero coefficients during the encoding process has a diminishing impact on the DFP of RLNC over fields larger than  $GF(2)$ , and C1, C2 and C3 perform similarly, especially at low values of  $\epsilon_{SR}$ .

## VI. CONCLUSIONS

Expressions to bound and approximate the decoding failure probability of random network coded relaying are readily available. This paper has extended these expressions to random network coded cooperation in order to capture the contribution

that links between the source nodes and the destination node can make to the reduction of the decoding failure probability. More crucially, the paper established that inclusion or exclusion of the zero element from  $GF(2)$  in the selection process of code coefficients at the relay nodes should take into consideration the packet erasure probability of the source-to-relay links, i.e.  $\epsilon_{SR}$ . For  $\epsilon_{SR} \leq 0.5$ , relay nodes should be able to choose the zero element with a probability equal to 0.5 or a variable probability that decreases from 0.5 to 0 for an increasing value of  $\epsilon_{SR}$ . For  $\epsilon_{SR} > 0.5$ , the zero element should be excluded from the encoding process at the relay nodes to avoid the generation of overly sparse linear combinations of data packets.

## VII. ACKNOWLEDGEMENTS

This work was partially supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant No. EP/R006385/1. Collaboration of the authors was facilitated by the Erasmus+ programme of the European Union.

## REFERENCES

- [1] R. Ahlswede, N. Cai, S. R. Li, and R. W. Weung, "Network information flow," *IEEE Trans. on Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [2] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Info. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct 2006.
- [3] C.-F. Chiasserini, E. Viterbo, and C. Casetti, "Decoding probability in random linear network coding with packet losses," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2128–2131, Nov. 2013.
- [4] S. T. Basaran, S. Gokceli, G. K. Kurt, E. Ozdemir, and E. Yaraneri, "Error performance analysis of random network coded cooperation systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5325–5337, Aug. 2017.
- [5] A. J. Klein, J. B. Lewis, and A. H. Morales, "Counting matrices over finite fields with support on skew Young diagrams and complements of Rothe diagrams," *J. Algebr. Comb.*, vol. 39, no. 2, pp. 429–456, Mar. 2014.
- [6] J.-T. Seong, "Bounds on decoding failure probability in linear network coding schemes with erasure channels," *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 648–651, Apr. 2014.
- [7] J. Blömer, R. Karp, and E. Welzl, "The rank of sparse random matrices over finite fields," *Rand. Struct. Alg.*, vol. 10, no. 4, pp. 407–420, Jul. 1997.
- [8] A. S. Khan and I. Chatzigeorgiou, "Improved bounds on the decoding failure probability of network coding over multi-source multi-relay networks," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2035–2038, Oct. 2016.
- [9] S. Brown, O. Johnson, and A. Tassi, "Reliability of broadcast communications under sparse random linear network coding," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–1, Jan. 2018.
- [10] A. Tassi, I. Chatzigeorgiou, and D. E. Lucani, "Analysis and optimization of sparse random linear network coding for reliable multicast services," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 285–299, Apr. 2016.
- [11] O. Trullols-Cruces, J. Barcelo-Ordinas, and M. Fiore, "Exact decoding probability under random linear network coding," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 67–69, Jan. 2011.
- [12] B. Shrader and N. M. Jones, "Systematic wireless network coding," in *Proc. IEEE Military Comm. Conf. (MILCOM)*, Boston, USA, Oct. 2009.