

# Improved bounds on the decoding failure probability of network coding over multi-source multi-relay networks

Amjad Saeed Khan and Ioannis Chatzigeorgiou

**Abstract**—This letter considers a multi-source multi-relay network, in which relay nodes employ a coding scheme based on random linear network coding on source packets and generate coded packets. If a destination node collects enough coded packets, it can recover the packets of all source nodes. The links between source-to-relay nodes and relay-to-destination nodes are modeled as packet erasure channels. Improved bounds on the probability of decoding failure are presented, which are markedly close to simulation results and notably better than previous bounds. Examples demonstrate the tightness and usefulness of the new bounds over the old bounds.

**Index Terms**—Network coding, sparse random matrices, probability of decoding failure, linear dependence.

## I. INTRODUCTION

The exploitation of cooperative diversity and the inclusion of network coding in multi-source multi-relay networks, in order to achieve excellent performance and high diversity gain, has attracted the interest of the research community. For example, it has been demonstrated in [1] that the use of network coding combined with cooperative diversity not only increase network reliability but also improve network throughput. Moreover, it has been shown in [2] and [3] that they can be significantly useful for wireless networks with disruptive channel and connectivity conditions.

This letter considers linear network coding over a multi-source multi-relay network, where  $N$  source nodes are supported by  $M$  relay nodes for the delivery of packets over packet erasure channels. To the best of our knowledge, an exact expression for the probability of decoding failure at a destination is not available but an effort has been made in [4], in which the author derives upper and lower bounds. However, the bounds presented in [4] are tight only for a certain range of parameters, including erasure probabilities, the values of  $N$ ,  $M$  and the size of the finite field. As shown in Section V of this letter, the existing upper bound is poor for a large number of source nodes and for large finite fields. Moreover, the existing lower bound is independent of the field size and is loose for small finite fields and low erasure probabilities.

The motivation for this work is to derive improved bounds on the probability of decoding failure. To this end, the main contributions of this letter can be summarized as follows: (i) we have reformulated the problem statement in order to associate the probability of decoding failure with the probability of attaining singular sparse random matrices, and (ii) we have revisited the expressions for upper and lower bounds in [4] and obtained alternative expressions, which are better than the

previous bounds for any field size and any value of  $N$  and  $M$ . Furthermore, the proposed lower bound incorporates the effect of the field size in contrast to the previous lower bound.

## II. SYSTEM MODEL

We consider a system with  $N$  source nodes and  $M$  relay nodes,  $\{S_1, S_2, \dots, S_N\}$  and  $\{R_1, R_2, \dots, R_M\}$ , respectively, as shown in Fig. 1, where  $M \geq N$ . Each source node  $S_i$  has a packet  $x_i$  to transmit to a destination D via  $M$  relay nodes. No source-to-destination links are assumed. The links connecting source-to-relay and relay-to-destination nodes are modeled as independent packet erasure channels characterized by erasure probability  $\epsilon_{SR}$  and  $\epsilon_{RD}$ , respectively.

The communication process is split into two phases. In the first phase, all the source nodes transmit their information packets simultaneously to the relay nodes over orthogonal broadcast channels. In the second phase, each relay node generates a coded packet by randomly combining the successfully received packets from the  $N$  source nodes. The  $M$  coded packets are then forwarded to the destination D over orthogonal channels. The coded packet  $y_i$ , which is transmitted by the  $i^{\text{th}}$  relay node, can be expressed as  $y_i = \sum_{j=1}^N c_{i,j} x_j$ , where  $c_{i,j}$  is a coding coefficient selected independently at random over a finite field  $F_q$  of size  $q$ . Because of the link condition  $\epsilon_{SR}$  between the source node  $S_j$  and the relay node  $R_i$ , each relay node receives packets from different source nodes. In contrast to [5] where coding coefficients are chosen uniformly at random, our system model imposes that the zero coefficient is assigned to erased packets and the remaining  $q - 1$  non-zero coefficients are selected uniformly at random by each relay for successfully received packets. Consequently, the coding coefficient distribution is given by

$$P[c_{i,j} = t] = \begin{cases} \epsilon_{SR}, & \text{if } t = 0 \\ \frac{1 - \epsilon_{SR}}{q - 1}, & \text{if } t \in F_q \setminus \{0\} \end{cases} \quad (1)$$

where  $0 \leq \epsilon_{SR} \leq 1$ . For a given relay node  $i$ , the sequence  $c_{i,1}, \dots, c_{i,N}$  forms a row vector, which is known as the coding vector of the coded packet  $y_i$ . As is commonly assumed in network coding [6], coding vectors are transmitted along with the corresponding coded packets. When the destination D receives  $N$  linearly independent coded packets, the packets of all source nodes can be recovered. Transmission of source packets over erasure channels and random linear coding at relay nodes is analogous to sparse random linear Network Coding (NC), which uses sparse random matrices [7], [8]. Based on the work of Blömer [7] and Cooper [8], this paper derives improved upper and lower bounds on the probability that the destination will fail to recover the source packets.

A. S. Khan and I. Chatzigeorgiou are with the School of Computing and Communications, Lancaster University, Lancaster, United Kingdom (e-mail: {a.khan9, i.chatzigeorgiou}@lancaster.ac.uk).

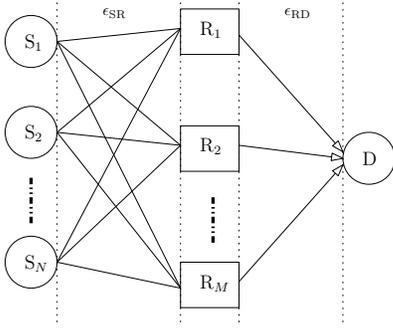


Figure 1. A network consisting of  $N$  source nodes,  $M \geq N$  relay nodes and a destination  $D$ . The packet erasure probability of a source-to-relay link and a relay-to-destination link is represented by  $\epsilon_{SR}$  and  $\epsilon_{RD}$ , respectively.

### III. PRELIMINARY RESULTS AND FORMER BOUNDS ON THE PROBABILITY OF DECODING FAILURE

Consider a matrix  $\mathbf{A} \in F_q^{M \times N}$ , whose elements are the coding coefficients  $c_{i,j}$  such that the  $i^{\text{th}}$  row of  $\mathbf{A}$  represents the coding vector associated with the  $i^{\text{th}}$  coded packet received by the destination  $D$ . The destination can recover the packets of the  $N$  source nodes if and only if  $\text{rank}(\mathbf{A}) = N$ . Thus, the decoding failure probability at the destination  $D$  can be defined as  $P_{\text{fail}} := \Pr\{\text{rank}(\mathbf{A}) < N\}$ . It is related to the linear dependence of the vectors of matrix  $\mathbf{A}$  and is defined as:

**Definition 1.** The vectors of matrix  $\mathbf{A} \in F_q^{M \times N}$  are said to be *linearly dependent* if and only if there exists a column vector  $\mathbf{x} \in F_q^{N \times 1} \setminus \{\mathbf{0}\}$  such that

$$\mathbf{A}\mathbf{x} = \mathbf{0}. \quad (2)$$

When there is no packet loss between the relay-to-destination channels, i.e.,  $\epsilon_{RD} = 0$ , the probability that the elements of the  $i^{\text{th}}$  row of matrix  $\mathbf{A}$  add up to zero, i.e.,  $c_{i,1} + c_{i,2} + \dots + c_{i,N} = 0$ , is given by [7]

$$\gamma_N = q^{-1} + (1 - q^{-1})\left(1 - \frac{1 - \epsilon_{SR}}{1 - q^{-1}}\right)^N. \quad (3)$$

Taking into account that matrix  $\mathbf{A}$  consists of  $M$  rows, the probability  $\Pr(\mathbf{A}\mathbf{x} = \mathbf{0})$  can be obtained as

$$\Pr(\mathbf{A}\mathbf{x} = \mathbf{0}) = \gamma_N^M = \left(q^{-1} + (1 - q^{-1})\left(1 - \frac{1 - \epsilon_{SR}}{1 - q^{-1}}\right)^N\right)^M. \quad (4)$$

The expected number of decoding failures at the destination  $D$  is given by the following theorem, which is a straightforward adaptation of [7, Theorem 3.3], [8, Theorem 3] to the system model under consideration.

**Theorem 1.** For a linear network coding scheme over  $N$  source nodes,  $M \geq N$  relay nodes and a single or multiple destinations, which are interconnected by links characterized by packet erasure probabilities  $0 \leq \epsilon_{SR} \leq 1$  and  $\epsilon_{RD} = 0$ , the expectation of the decoding failures can be obtained as

$$\mu_0(N, M) = E(\mathbf{A}\mathbf{x} = \mathbf{0}) = \frac{1}{q-1} \sum_{w=1}^N \binom{N}{w} (q-1)^w \gamma_w^M \quad (5)$$

where  $\mathbf{A} \in F_q^{M \times N}$  is the coding matrix at a destination.

Following the same line of reasoning, a direct extension of (5) for  $\epsilon_{RD} \geq 0$  has been made in [4, Theorem 1] and was used to upper bound the probability of decoding failure.

**Corollary 1.** The probability of decoding failure at a destination is bounded from above as:

$$P_{\text{fail}} \leq \frac{1}{q-1} \sum_{w=1}^N \binom{N}{w} (q-1)^w [\epsilon_{RD} + (1 - \epsilon_{RD})\gamma_w]^M \quad (6)$$

where  $N$  is the number of source nodes,  $M \geq N$  is the number of relay nodes and  $\epsilon_{SR}$ ,  $\epsilon_{RD}$  represent the packet erasure probabilities between the network nodes.

However, (6) is only tight for limited values of erasures  $\epsilon_{SR}$  and  $\epsilon_{RD}$ , depending on  $N$ ,  $M$  and  $q$ . In particular, the upper bound takes values greater than 1 when either the field size is big or the difference between the number of source and relay nodes is small. This disparity between the probability of decoding failure and the upper bound will be demonstrated in Section V. In an effort to improve the tightness of (6), Seong *et al.* proposed the selection of the minimum value between the upper bound in (6) and 1 [9]. A lower bound on the probability of decoding failure has also been obtained by Seong in [4, Theorem 2]:

**Theorem 2.** Consider a network comprising  $N$  source nodes and  $M \geq N$  relay nodes, assume that links are modeled as packet erasure channels with erasure probabilities  $\epsilon_{SR}$  and  $\epsilon_{RD}$ , and let  $\mathbf{A} \in F_q^{M \times N}$  be the coding matrix at a destination node. The probability of decoding failure  $P_{\text{fail}}$  is lower bounded by

$$P_{\text{fail}} \geq \sum_{k=1}^N \binom{N}{k} \left( (\epsilon_{SR} + \epsilon_{RD} - \epsilon_{SR}\epsilon_{RD})^M \right)^k \times (1 - (\epsilon_{SR} + \epsilon_{RD} - \epsilon_{SR}\epsilon_{RD})^M)^{N-k}. \quad (7)$$

The bounds in (6) and (7) are used in [9] and [10]. For example, (6) is employed in [10] to evaluate the performance gains introduced by linear NC in a practical network architecture for emergency communications. However, the following section will derive new bounds, which are considerably tighter than the previous bounds and can significantly improve the quality and accuracy of results presented in the literature.

### IV. IMPROVED BOUNDS ON THE PROBABILITY OF DECODING FAILURE

#### A. Upper Bound

For  $\epsilon_{RD} = 0$ , an upper bound on the decoding failure probability can be obtained by extending and adapting [7, Theorem 6.3] as follows:

**Lemma 1.** Let  $\mathbf{A} \in F_q^{M \times N}$  be the coding matrix at a destination node of a network consisting of  $N$  source nodes and  $M$  relay nodes. If the internode erasure probabilities are  $0 \leq \epsilon_{SR} \leq 1$  and  $\epsilon_{RD} = 0$ , the probability of decoding failure is upper bounded by

$$\eta_{\text{max}}(N, M) = 1 - \prod_{i=1}^N (1 - \beta_{\text{max}}^{M-i+1}) \quad (8)$$

where  $\beta_{\text{max}} = \max(\epsilon_{SR}, \frac{1 - \epsilon_{SR}}{q-1})$ .

*Proof:* Let us assume that the first  $i - 1$  columns of  $\mathbf{A}$ , denoted by  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{i-1}$ , are linearly independent. This implies that by using elementary column operations, matrix  $\mathbf{A}$  can be transformed into a matrix that contains an  $(i - 1) \times (i - 1)$  identity matrix. Without loss of generality, let us assume that the first  $i - 1$  rows form the identity matrix. The columns of this matrix represent the basis for the vector space spanned by  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{i-1}$ . Therefore, the probability that  $\mathbf{A}_i$  is linearly independent from  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{i-1}$  depends only on the last  $M - i + 1$  elements of  $\mathbf{A}_i$ . This probability is lower bounded by  $1 - \beta_{\max}^{M-i+1}$ , where  $\beta_{\max}$  specifies the maximum probability of obtaining an element from  $F_q$ . Hence, matrix  $\mathbf{A}$  contains an  $N \times N$  non-singular matrix with probability at least  $\prod_{i=1}^N (1 - \beta_{\max}^{M-i+1})$ . As a result, the probability that matrix  $\mathbf{A}$  does not contain an invertible matrix and, consequently, a decoding failure will occur is upper bounded by subtracting this product from one, which completes the proof. ■

Lemma 1 will be used to obtain a tighter upper bound on  $P_{\text{fail}}$ . Before we invoke it, we shall first revisit (6) and rewrite it as:

$$P_{\text{fail}} \leq \sum_{r=0}^M \binom{M}{r} \epsilon_{\text{RD}}^{M-r} (1 - \epsilon_{\text{RD}})^r \mu_0(N, r). \quad (9)$$

This change is possible if  $[\epsilon_{\text{RD}} + (1 - \epsilon_{\text{RD}})\gamma_w]^M$  is expanded into a sum, as per the binomial theorem.

**Theorem 3.** For a network coding scheme over multi-source multi-relay networks, composed of  $N$  source nodes and  $M$  relay nodes with packet erasures  $\epsilon_{\text{SR}}$  and  $\epsilon_{\text{RD}}$ , the probability of decoding failure is upper bounded by

$$P_{\text{fail}} \leq \sum_{r=0}^M \binom{M}{r} \epsilon_{\text{RD}}^{M-r} (1 - \epsilon_{\text{RD}})^r \min\{\eta_{\max}(N, r), \mu_0(N, r)\}. \quad (10)$$

*Proof:* As inferred from (9), the number of packet deliveries by the relays follows the binomial distribution. If we employ Theorem 1 and Lemma 1 on the number of received coded packets  $r$ , a tight upper bound can be obtained by taking the minimum of outcomes and multiply with the probability distribution of  $r$ . Summing the resultant quantity gives (10), which concludes the proof. ■

**Remark.** It is worth noting that the upper bound is not simply the minimum between two cumulative probability distributions (CDFs), that is, the right-hand of (6) and the CDF of (8) for *all* possible numbers of relay nodes. Instead, the right hand of (6) has been rewritten in the form of (9), which enabled us to identify the minimum between  $\mu_0$  and  $\eta_{\max}$  for *each* possible number of relay nodes, and use it in the computation of the CDF shown in (9).

### B. Lower Bound

The bound that was derived in [7, Theorem 6.3] was extended to an upper bound on the probability that an  $M \times N$  matrix  $\mathbf{A}$  does not contain an invertible  $N \times N$  matrix in Lemma 1. The same approach can be followed to obtain a lower bound as follows:

**Lemma 2.** Let  $\mathbf{A} \in F_q^{M \times N}$  be the coding matrix at a destination of a network consisting of  $N$  source nodes and  $M$  relay nodes. If the internode erasure probabilities are  $0 \leq \epsilon_{\text{SR}} \leq 1$  and  $\epsilon_{\text{RD}} = 0$ , the probability of decoding failure is lower bounded by

$$\eta_{\min}(N, M) = 1 - \prod_{i=1}^N (1 - \beta_{\min}^{M-i+1}) \quad (11)$$

where  $\beta_{\min} = \min(\epsilon_{\text{SR}}, \frac{1 - \epsilon_{\text{SR}}}{q - 1})$ .

*Proof:* The proof follows exactly the same line of reasoning as that of Lemma 1. ■

An improved lower bound on  $P_{\text{fail}}$  can be obtained if the right-hand side of (8) is denoted by  $P_0(N, M)$  for  $\epsilon_{\text{RD}} = 0$ , that is

$$P_0(N, M) = \sum_{k=1}^N \binom{N}{k} (\epsilon_{\text{SR}})^k (1 - \epsilon_{\text{SR}})^{N-k} \quad (12)$$

and then combined with (11) in Lemma 2. In particular:

**Theorem 4.** For a linear network coding scheme over  $N$  source nodes and  $M \geq N$  relay nodes, let  $\epsilon_{\text{SR}}$  and  $\epsilon_{\text{RD}}$  be the packet erasure probabilities of the internode links. The probability of decoding failure is lower bounded by

$$P_{\text{fail}} \geq \sum_{r=0}^M \binom{M}{r} \epsilon_{\text{RD}}^{M-r} (1 - \epsilon_{\text{RD}})^r \max\{\eta_{\min}(N, r), P_0(N, r)\}. \quad (13)$$

*Proof:* In contrast to Theorem 3, here we employ Lemma 2 and (12) on the number of received coded packets  $r$ , and we select the maximum of outcomes. The rest of the proof follows the same reasoning as that presented in the proof of Theorem 3. ■

## V. RESULTS

This section compares the analytical expressions of the proposed bounds to simulation results. In addition, the proposed upper bound and lower bound, which shall be referred to as UB-new and LB-new, are contrasted with the old bounds represented by (6) and (7), which shall be referred to as UB-old and LB-old. To obtain simulation results, each scenario was run over  $10^4$  realizations, failures by the destination to recover the packets of all source nodes were counted, and the decoding failure probability was measured.

Fig. 2 shows numerical results of the upper bounds obtained from (6) and (10) and labeled UB-old and UB-new, respectively. We observe that, in contrast to UB-old, UB-new is significantly tighter to the simulated performance. When the number of source nodes and the number of relay nodes increase to  $N = 30$  and  $M = 35$ , respectively, it can be clearly seen that the UB-old curve moves far away from the simulated curve but the proposed UB-new expression still provides a tight bound. This reveals the fact that UB-old produces a worse approximation error for large values of  $N$ .

Fig. 3 evaluates the probability of decoding failure for  $q = \{4, 64\}$ , and contrasts the proposed bounds (UB-new and LB-new) with the old bounds (UB-old and LB-old). The figure demonstrates that for  $\epsilon_{\text{SR}} \in [0.1, 0.7]$ , the network experiences

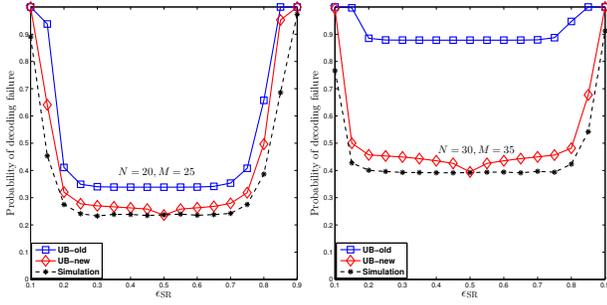


Figure 2. Comparison between simulation results and the theoretical upper bounds obtained from (6) and (10) for different values of  $N$  and  $M$ , when  $q = 2$ ,  $\epsilon_{RD} = 0.1$  and  $\epsilon_{SR} \in [0.1, 0.9]$ .

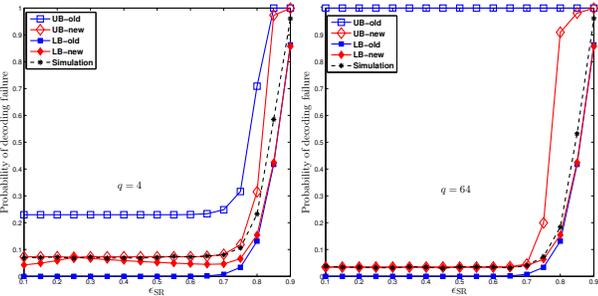


Figure 3. Effect of field size  $q$  on network performance and comparison between the proposed bounds and the old bounds for  $\epsilon_{SR} \in [0.1, 0.9]$ , when  $N = 20$ ,  $M = 25$  and  $\epsilon_{RD} = 0.1$ .

only a small probability of decoding failure. Furthermore, the figure shows that UB-new and LB-new are very close to the simulated performance and outperform UB-old and LB-old, respectively. In particular, when  $q = 64$ , UB-old and LB-old are markedly loose while UB-new and LB-new are very tight to the actual simulation results. The performance of the network deteriorates for values of  $\epsilon_{SR}$  greater than 0.75. Moreover it is interesting to notice that, for large values of  $q$ , the upper bounds deviate from the simulation results and the simulations can be better approximated by the lower bounds.

Figs. 4 and 5 plot the probability of decoding failure versus the number of relays  $M$  with  $N = 10$  and  $q = \{2, 4\}$ . It is evident that the probability of decoding failure decreases with an increasing number of relays and field size. The figures also demonstrate that, when  $M < 2N$ , UB-new and LB-new are close to the simulated outcomes, compared to UB-old and LB-old, respectively. It follows from (7) that LB-old depends only on the erasures  $\epsilon_{SR}$  and  $\epsilon_{RD}$ , and does not depend on the field size  $q$ , thus shows no improvement for  $q = 4$ . However, LB-new approaches the simulation results, when  $q$  increases to 4. For example in Fig. 5, when  $q = 4$  and  $M \leq 14$ , both UB-new and LB-new are very tight, while UB-old and LB-old are noticeably far from the simulated performance.

## VI. CONCLUSIONS

We presented improved upper and lower bounds on the probability of decoding failure in a multi-source multi-relay network, which employs linear network coding. The proposed analysis for counting failures provided significantly tighter bounds, which outperform existing bounds, derived in [4].

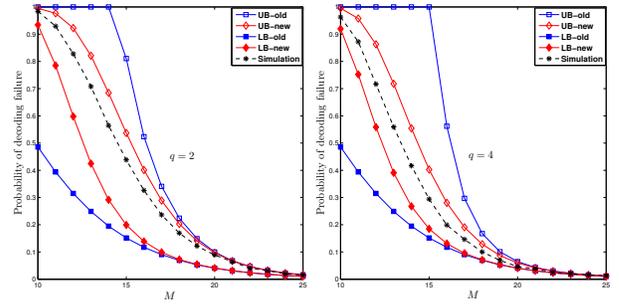


Figure 4. Performance of the network for an increasing number of relays  $M$ . The proposed bounds and the old bounds have been plotted for  $N = 10$ ,  $\epsilon_{SR} = 0.7$ ,  $\epsilon_{RD} = 0.2$  and different values of field size  $q$ .

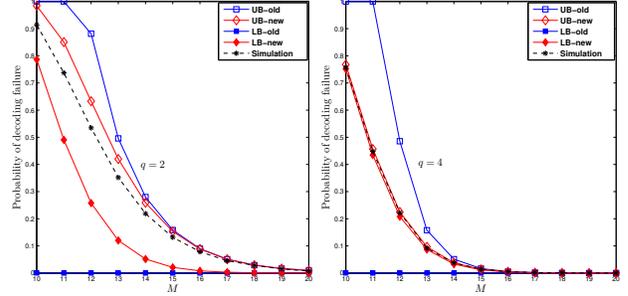


Figure 5. Network performance and comparison between the proposed bounds and the old bounds for  $N = 10$ , an increasing number of relays  $M$ ,  $\epsilon_{SR} = 0.3$ ,  $\epsilon_{RD} = 0.1$  and different field size  $q$ .

Several examples, which considered various numbers of source nodes and relay nodes, different field sizes and a range of erasure probabilities, established the shortcomings of the existing bounds and demonstrated the tightness of the proposed improved bounds. Finally, we assert that the proposed bounds can also be used to better estimate the performance of systems employing sparse random linear network coding schemes, presented in the literature.

## REFERENCES

- [1] Z. Ding, Z. Ma, and K. K. Leung, "Impact of network coding on system delay for multisource-multidestination scenarios," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 831–841, Feb. 2010.
- [2] J.-S. Park, M. Gerla, D. S. Lun, Y. Yi, and M. Médard, "Codecast: A network-coding-based ad hoc multicast protocol," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 76–81, Oct. 2006.
- [3] S. R. Katti, "Network coded wireless architecture," *Ph.D Dissertation, Massachusetts Institute of Technology, USA*, Sep. 2008.
- [4] 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.
- [5] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [6] J. Heide, M. V. Pedersen, F. H. Fitzek, and M. Médard, "On code parameters and coding vector representation for practical RLNC," in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011.
- [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] C. Cooper, "On the distribution of rank of a random matrix over a finite field," *Rand. Struct. Alg.*, vol. 17, no. 3–4, pp. 197–212, Oct. 2000.
- [9] J.-T. Seong and H.-N. Lee, "Predicting the performance of cooperative wireless networking schemes with random network coding," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2951–2964, Aug. 2014.
- [10] T. Do-Duy and M. A. Vázquez-Castro, "Efficient communication over cellular networks with network coding in emergency scenarios," in *Proc. ICT-DM*, Rennes, France, Nov. 2015.