Reliability of Relay Networks under Random Linear Network Coding

Evgeny Tsimbalo and Magnus Sandell

Abstract—We consider a single-source, multiple-relay, single-destination lossy network employing Random Linear Network coding at all transmitting nodes. We address the problem of calculating the probability of successful decoding at the destination node. In contrast with some previous studies, we assume the classical RLNC scheme, in which the relaying nodes simply re-encode packets, without resorting to decoding. In addition, we consider an arbitrary field size and take into account correlation between the relay nodes. We derive exact expressions for networks with up to two relays, and propose a novel upper bound for the general case of multiple relays. Using Monte Carlo simulations, we show that the exact results match the simulated ones for the case of two relays. For networks with more than two relays, we show that the proposed bound is very accurate, exhibiting the mean square error as low as $10^{-6}$.

Index Terms—Relay Networks, Reliability, Fountain Coding, RLNC.

I. INTRODUCTION

Relay-based topologies are widely used in modern wireless networks to extend transmission range and to utilise spatial diversity in order to improve communication reliability. For example, in body sensor networks, intermediate nodes are commonly deployed to relay data to monitoring devices [1]. Nevertheless, such networks are often characterised by poor reliability at the physical layer, resulting in frequent retransmissions requested by the upper layers. In this context, the idea of Application Level Forward Error Correction (AL-FEC) [2], where coding is performed over packets rather than bits, has attracted significant interest over the years.

Traditional AL-FEC schemes are based on the so-called fountain codes [3], [4]. However, fountain codes usually operate efficiently with large block sizes and are applied on a single-hop basis, i.e., requiring decoding followed by re-encoding at each hop [5]. By contrast, the idea of linearly combining packets using random coefficients, known in the literature as Random Linear Network Coding (RLNC) [6], [7], dictates simple re-encoding at each hop, thus inherently supporting multi-hop topologies. It was proved that RLNC is capacity-achieving in the context of multicast wireless networks [3], [9].

Several attempts to analyse the performance RLNC applied to relay networks have been in the literature. Multiple-source, multiple-relay two-hop networks were studied in [10]–[12]. In [10], each source node was designed to transmit a single uncoded packet, while each relay node encoded packets it received into a single linear combination and forwarded it to the destination node. An exact expression for the probability of successful decoding was derived, assuming that at most one of the source-relay links failed and that communication between the relays and the destination was error free. More general scenarios were used in [11], for which lower and upper bounds were proposed. More accurate bounds were subsequently derived in [12], where the theory of sparse random matrices [13] was employed. It should be noted that in all three studies [10]–[12], only a single packet was assumed to be transmitted by any source or relay node. In other words, despite having multiple sources and multiple relays, the networks considered in [10]–[12] can be represented by an equivalent single-source single-relay network, in which the source and relay transmit multiple packets experiencing independent packet loss.

A common assumption in [10]–[12] was that the source nodes transmit uncoded packets. Encoding at the source node was considered in [14] for a single-source single-relay network with a direct link between the source and destination nodes. It was demonstrated that encoding at both source and relay nodes has performance gains compared to encoding at the relay only. The analysis, however, was performed assuming an infinitely large field size, which in practice is limited [15], [16]. A finite field size was considered in [17], where the relay node attempts the decoding of the source message, followed by re-encoding. Using the terminology of relay networks, such scheme can be referred to as Decode-and-Forward (DF). However, the analysis in [17] was based on the assumption that the relay and destination nodes receive uncorrelated subsets of packets from the source. As a result, the derived performance bounds were accurate only under certain network scenarios. The correlation was taken into account in [18], where an exact expression for the probability of successful decoding for a single-relay network operating under the DF scheme was derived. Furthermore, the DF scheme was complemented with a passive relay mode, in which if the relay is not able to decode, it simply re-transmits the packets it received from the source to the destination.

In this work, we address the limitations of the previous studies of relay networks employing RLNC and provide the following contributions:

- We generalise the single-relay scenario considered previously in [17], [18] to an arbitrary number of relays. In contrast with [10]–[12], we enable encoding at the source node and allow each relay node to transmit multiple coded packets. We derive an exact expression for the probability of successful decoding in the case of two relays, and a tight upper bound for any number of relays.
• Compared with the DF scheme employed in [17], [18], we consider a different relaying strategy, whereby the relay nodes do not attempt to decode, but simply re-encode the packets they receive from the source. Once again employing the terminology of relay networks, we call such scheme Amplify-and-Forward (AF). The AF strategy follows the original idea of RLNC, in which every transmitting node in the network performs random linear combinations of the packets in its buffer. Comparing with DF, the AF scheme is expected to provide better reliability, since the relay nodes are likely to transmit more packets.

• We perform thorough benchmarking of the proposed analytical results via extensive Monte Carlo simulation for various network and code parameters. In particular, we demonstrate that the bound proposed for an arbitrary number of relays closely follows the simulated results, and its accuracy increases with the number of relays and the source message size.

The remainder of the paper is organised as follows. Section II describes the system model and provides some background results. Section III presents theoretical analysis, starting with the trivial case of a single relay, extending the result to two relays and concluding with the general case of multiple relays. In Section IV, the proposed results are compared with simulated ones. The conclusions are drawn in Section V.

II. SYSTEM MODEL AND BACKGROUND

A single-source network with \( L \) relays is depicted in Fig. 1. The goal is to successfully deliver a message of \( K \) equally-sized source packets from the source node \( S \) to the destination node \( D \) via the relay nodes \( R_j, j = 1, \ldots, L \). Let \( \epsilon_{SR_j} \) denote the PER between \( S \) and \( R_j \), and let \( \epsilon_{RD_j} \) denote the PER between \( R_j \) and \( D \).

In the AF scheme considered in this work, the transmission is performed in two stages. During the first stage, the source node broadcasts \( N_S \geq K \) coded packets to the relay nodes. Each coded packet is a linear combination of the source packets with the coefficients drawn uniformly at random from a finite field \( \mathbb{F}_q \) of size \( q \). This way of encoding, also known as non-systematic RLNC, will be used throughout the paper. Let \( m_j \leq N_S, j = 1, \ldots, L \), denote the number of coded packets received by \( R_j \). It is assumed that each receiving node knows the coding coefficients associated with every packet it receives, which can be achieved by transmitting the coefficients in the packet header [7]. Each relay can therefore construct an \( m_j \times K \) matrix of coefficients \( C_{R_j}, j = 1, \ldots, L \).

During the second stage of the AF scheme, each relay node re-encodes the packets it received from \( S \) into \( N_R \) packets and transmits them to \( D \). For re-encoding, each relay node generates a new \( N_R \times m_j \) matrix of random coefficients \( G_j, j = 1, \ldots, L \). This is equivalent to applying a product of the two matrices, \( G_jC_{R_j}, j = 1, \ldots, L \), to the vector of \( K \) source packets. The new coding coefficients, corresponding to the rows of \( G_jC_{R_j}, j = 1, \ldots, L \), are transmitted in packet headers. Let \( m'_j \leq N_R \) denote the number of packets the destination node receives from \( R_j \). Node \( D \) can then reconstruct \( m'_j \) rows of \( G_jC_{R_j}, j = 1, \ldots, L \), and stack them together into a \( \sum_{j=1}^{L} m'_j \times K \) matrix \( C_D \). The destination node can recover the source message if matrix \( C_D \) has rank \( K \). Our goal will be to characterise the performance of such network in terms of probability of successful decoding, which will be also referred to as decoding probability.

A. Theoretical Background

We now present some relevant results in the area of RLNC that we will use throughout the paper.

Consider first a single-hop point-to-point link with PER \( \epsilon \). The decoding probability of a message of \( K \) packets after \( N \) coded transmissions is expressed as follows [17]:

\[
P(N, \epsilon) = \sum_{m=K}^{N} B(m, N, \epsilon) \mathcal{P}(m, K).
\]

Here, \( B(m, N, \epsilon) \) denotes the probability mass function (PMF) of the binomial distribution, i.e., the probability of \( m \) successes out of \( N \) independent trials with success probability \( \epsilon \) calculated as follows:

\[
B(m, N, \epsilon) = \binom{N}{m} (1-\epsilon)^{m} \epsilon^{N-m}.
\]

Furthermore, \( \mathcal{P}(m, K) \) in [1] denotes the probability that an \( m \times K \) matrix of elements generated uniformly at random from \( \mathbb{F}_q, m \leq K \), has rank \( K \), which is given by [19]:

\[
\mathcal{P}(m, K) = \prod_{i=0}^{K-1} (1-q^{i-m}).
\]

It should be noted that (3) is a specific case of the probability of the binomial distribution, i.e.,

\[
\mathcal{P}(m, K) = \binom{K}{m} \mathbb{B}(m, K).
\]

where

\[
\binom{K}{m} = \prod_{i=0}^{K-1} (1-q^{i-m})
\]

is the Gaussian binomial coefficient [21].

Consider now a single-relay network with PERs \( \epsilon_{SR} \) and \( \epsilon_{RD} \). Such network was analysed under the DF relaying scheme in [13] in the presence of a direct link between \( S \) and \( D \). In the case there is no communication between \( S \) and \( D \), the probability of successful decoding can be expressed using the results of [18] as follows:

\[
P_R^{(1)} = \sum_{m=K}^{N_S} B(m, N_S, \epsilon_{SR}) \sum_{m'=K}^{N_R} B(m', N_R, \epsilon_{RD}) \mathcal{P}(m, K) \mathcal{P}(m', K).
\]

By re-arranging the terms and employing (1), (6) can be rewritten as follows:

\[
P_R^{(1)} = P(N_S, \epsilon_{SR})P(N_R, \epsilon_{RD}).
\]
In other words, the decoding probability for a single-relay DF scheme is the product of the decoding probabilities for each hop.

The notation introduced in this section is summarised in Table I.

### III. Theoretical Analysis

We now turn our attention back to the AF relaying scheme described previously. We aim at calculating the decoding probability $P^{(\ell)}_R$ for the $L$-relay network depicted in Fig. 1. We start with two specific cases of $L = 1$ and $L = 2$, for which we derive exact expressions, and then propose a bound for an arbitrary number of relays.

#### A. Single-relay Case

Consider a single-relay network operating under the AF scheme described in Section II. To simplify the notation, we will omit the relay indices for all relevant parameters of Table I.

Consider the $m' \times K$ matrix of coding coefficients received by D, $C_D$, where $m'$ is the number of packets received by D from R. The decoding probability is then equal to the probability of this matrix having rank $K$. As was described in Section II, $C_D$ is constructed from $m'$ rows of the product $GC_R$, where $G$ is the $N_R \times m$ re-encoding matrix and $C_R$ is the $m \times K$ matrix of coding coefficients received by R from S. Let $\tilde{G}$ denote a matrix composed of $m'$ rows of $G$. Therefore,

$$C_D = \tilde{G}C_R.$$

In other words, matrix $C_D$ is composed of linear combinations of the rows of matrix $C_R$ determined by matrix $\tilde{G}$.

Assume that a row of $\tilde{G}$ is a linear combination of a set of some other rows of $G$, $J \subset \{1, \ldots, m'\}$. In this case, the corresponding row of $\tilde{G}C_R$ will be the same linear combination of the set $J$ of rows of $GC_R$. The number of independently generated rows in $\tilde{G}C_R$ will be determined by the number of linearly independent rows of $G$, i.e., its rank. The probability that matrix $C_D$ is full rank, given $m$ and $m'$, can therefore be expressed by marginalising it over the distribution of the rank of $G$ as follows:

$$\Pr[\text{rank}(C_D) = K] = \sum_{r=K}^{\min(m', m)} \Pr_r(m', m) \Pr(r, K).$$

Here, $\Pr_r(m', m)$ is the probability that $\tilde{G}$ has rank $r$ defined in (9).

We now establish the following result.

**Lemma 3.1:** For the right-hand side of (9), the following identity holds:

$$\sum_{r=K}^{\min(m', m)} \Pr_r(m', m) \Pr(r, K) = \Pr(m, K) \Pr(m', K).$$

**Proof:** See Appendix A.

From (9), it can be observed that the probability that matrix $C_D$ is full rank depends on $m$ and $m'$, the numbers of packets received by R and D, respectively. Considering these numbers as random variables described by the binomial distribution (2), the decoding probability can be calculated by marginalising (9) accordingly and applying the identity of Lemma 3.1.

$$P_R^{(\ell)} = \sum_{m=K}^{N_R} B(m, N_R, \epsilon_{SR}) \sum_{m'=K}^{N_R} B(m', N_R, \epsilon_{RD}) \Pr[\text{rank}(C_D) = K]$$

where $P(\cdot)$ is given by (1). It can be observed that the decoding probability of the single-relay AF scheme is exactly the same as that of the DF scheme (7). Indeed, for both schemes, the relay node has to be able to decode the message. As a result, the decoding probability is the product of those for each individual link (hop). Naturally, the result can be generalised to networks with any number of hops, also known as multiple-link tandem networks (8).

We will now apply the methodology presented in this section for networks with two and more relays.
B. Two-relay Case

As was mentioned in Section II during the first stage the source node multicasts to the relay nodes. The traditional approach in the literature to the analysis of RLNC over multicast communication is to assume either a large field size \(q\) or that each receiver has an independent subset of coded packets. In both cases, the multicast network can be approximated as a set of independent unicast, point-to-point connections. Applied to the network in question, it means that the matrices of coding coefficients at the relay nodes, \(C_{R1} \in \mathbb{F}_q^{m_1 \times K}\) and \(C_{R2} \in \mathbb{F}_q^{m_2 \times K}\), can be assumed uncorrelated. Therefore, the probability that the matrix of coding coefficients collected by \(D\) is full rank, for given \(m_j\) and \(m_j'\), \(j = 1, 2\), can be approximated by the straightforward extension of (9):

\[
\Pr[\text{rank}(C_D) = K] \approx \sum_{r_1, r_2} \prod_{j=1}^{2} \mathbb{P}(r_j \in r_1 + r_2, K), \tag{12}
\]

where the summation over \(r_1, r_2\) is performed such that \(r_1 + r_2 \geq K\) and neither value exceeds the corresponding matrix dimensions. Marginalising (12) over the distribution of numbers of packets received by the relays and the destination, the decoding probability for the AF scheme with \(L = 2\) relays can be expressed as follows:

\[
P_R^{(2)} \approx \sum_{m_1, m_2=0}^{N_S} \sum_{j=1}^{2} B(m_j, N_S, \epsilon_{SR_j}) \cdot \sum_{m_1', m_2'=0}^{N_E} \sum_{j=1}^{2} B(m_j', N_R, \epsilon_{RD_j}) \cdot \prod_{j=1}^{2} \mathbb{P}(r_j \in r_1 + r_2, K), \tag{13}
\]

In (22) it was shown, however, that the receivers in a multicast network are likely to receive common packets, hence their matrices of coding coefficients are correlated. For instance, if the source node transmits \(N_S = 10\) packets to the two relay nodes with the PER of 0.1, each relay is expected to receive 9 packets, which makes the number of commonly received packets either 8 or 9. As a result, the relay nodes will reencode more or less the same packets, thus acting more like a single relay, but generating twice as many linear combinations. By contrast, if approximation (13) is used, the destination node can decode with a much higher probability than in a single-relay case. Therefore, it is expected that (13) is a fairly loose upper bound, which will be later proved in Section IV.

To find an exact expression for \(P_R^{(2)}\), the probability of successful decoding needs to be further marginalised over the distribution of the number of packets received simultaneously by \(R_1\) and \(R_2\). To this end, packets received by the relays can be split into three categories: common to both relays, unique to \(R_1\) and unique to \(R_2\). Let \(\mu\) be a random variable denoting the number of common packets. The joint PMF of \(\mu\), \(m_1\) and \(m_2\) is described by the multinomial distribution (22) as follows:

\[
\mathcal{M}(m_1, m_2, \mu, N_S, \epsilon_{SR1}, \epsilon_{SR2}) = \binom{N_S}{\mu} \binom{N_S}{m_1 - \mu} \binom{N_S}{m_2 - \mu} \prod_{j=1}^{2} (1 - \epsilon_{SR_j})^{m_j'} N_S - m_j'. \tag{14}
\]

Having known \(\mu\), the rows of coding matrices \(C_{R1}\) and \(C_{R2}\) can be stacked together, such that, without loss of generality, the first \(\mu\) rows correspond to the common packets, the next \(m_1 - \mu\) rows correspond to packets received by \(R_1\) only, and the last \(m_2 - \mu\) rows correspond to packets received by \(R_2\) only. Let \(C_R\) denote the resulting matrix. The re-encoding operation can now be expressed as \(GC_R\), where \(G\) is an equivalent re-encoding matrix that has the following structure:

\[
G = \begin{pmatrix}
G^{(1)}_1 & G_{11} & 0 \\
G^{(2)}_2 & 0 & G_{22}
\end{pmatrix}, \tag{15}
\]

Here, the \(N_R \times \mu\) matrices \(G^{(j)}_j\) are composed of the columns of \(G_j\) that correspond to \(\mu\) common packets, while \(N_R \times (m_j - \mu)\) matrices \(G_{jj}\) correspond to \(m_j - \mu\) unique packets received by the \(j\)-th relay, \(j = 1, 2\).

Matrices with the structure as in (15), commonly referred to as block angular matrices, have been studied in the literature, where the probability of such matrices being full-rank was first derived in [23] and later applied in [17] and [18]. Here we generalise the result of Theorem 2 in [23] as follows:

**Lemma 3.2:** Let \(X\) be a block angular matrix with the following structure:

\[
X = \begin{pmatrix}
A_1 & B_1 & 0 \\
A_2 & 0 & B_2
\end{pmatrix}, \tag{16}
\]

where \(A_j \in \mathbb{F}_q^{a_j \times a}\) and \(B_j \in \mathbb{F}_q^{b_j \times b_j}, j = 1, 2\). The probability that matrix \(X\) has rank \(r \leq \min(a_1 + a_2, a + b_1 + b_2)\) is given by

\[
\mathbb{P}_r(a_1, a_2, b_1, b_2) = \sum_{x,y} \mathbb{P}_x(a_1, b_1) \mathbb{P}_y(a_2, b_2) \cdot \mathbb{P}_r-x-y(a_1 - r_1 + a_2 - r_2, a), \tag{17}
\]

where \(\mathbb{P}_x(.)\) is given by (1) and summation is performed over all pairs of \(x, r\) such that \(r - a \leq x + y \leq r, x \leq a_1, b_1\), \(y \leq a_2, b_2\).

**Proof:** See Appendix B.

**Remark 3.1:** The result of Theorem 2 in [23] is a specific case of Lemma 3.2 for which \(r = \min(a_1 + a_2, a + b_1 + b_2)\).

**Remark 3.2:** Lemma 3.2 can be generalised to a block angular matrix with an arbitrary number of blocks \(A\) and \(B\).

Having established (17), we can now formulate the exact expression for the decoding probability of a two-relay network.
**Theorem 3.1**: The probability of successful decoding for the two-relay network described in Section II is given by

\[
P^{(2)}_R = \sum_{m_1, m_2, \mu} \mathcal{M}(m_1, m_2, \mu, N_S, \epsilon_{SR1}, \epsilon_{SR2}) \cdot \sum_{m'_1, m'_2, j=1}^2 \mathcal{B}(m'_j, N_R, \epsilon_{RD_j}) \cdot \sum_r \hat{P}_r(m'_1, m'_2, \mu, m_1 - \mu, m_2 - \mu) \mathbb{P}(r, K),
\]

where the summations are performed over the following values:

- \(m_1 + m_2 \geq K; m_1, m_2 \leq N_S\);
- \(\mu = \max(0, m_1 + m_2 - N_S), \ldots, \min(m_1, m_2)\);
- \(m'_1 + m'_2 \geq K; m'_1, m'_2 \leq N_R\);
- \(r = K, \ldots, \min(m'_1 + m'_2, m_1 + m_2 - \mu)\).

**Proof**: Consider the equivalent re-encoding matrix \(\mathbf{G}\) defined by (15) which is applied to the combined matrix of coding coefficients \(\mathbf{C}_R\). By analogy to the single-relay case, let \(\mathbf{G}\) denote an \((m'_1 + m'_2) \times m\) matrix obtained from \(\mathbf{G}\) by removing rows corresponding to lost packets between the relays and the destination. The matrix of coding coefficients of the destination node, \(\mathbf{C}_D\), can be therefore calculated according to (8). The probability that this matrix is full rank, for given \(\mu, m_j\) and \(m'_j\), \(j = 1, 2\), can be computed similarly to (9) as follows:

\[
\Pr[\text{rank}(\mathbf{G}_D) = K] = \sum_r \Pr[\text{rank}(\hat{\mathbf{G}}) = r] \mathbb{P}(r, K),
\]

where the minimum value of \(r\) is equal to \(K\) and the maximum value is limited to the smallest dimension of \(\hat{\mathbf{G}}\), which is equal to \(\min(m'_1 + m'_2, m_1 + m_2 - \mu)\). The first probability under the summation in (19) can be calculated by using Lemma 3.2 where \(a_j = m'_j, b_j = m_j - \mu\) and \(a = \mu, j = 1, 2\):

\[
\Pr[\text{rank}(\hat{\mathbf{G}}) = r] = \hat{P}_r(m'_1, m'_2, \mu, m_1 - \mu, m_2 - \mu).
\]

Substituting (20) into (19) and marginalising the latter over the multinomial distribution (14) of \(m_1, m_2, \mu\) and two independent binomial distributions of \(m'_1, m'_2\), expression (19) can be readily obtained. The starting values of \(m_1\) and \(m_2\) should be chosen so that their sum is at least \(K\), to collect a sufficient number of linearly independent rows of coding coefficients from the source nodes. The starting values of \(m'_1\) and \(m'_2\) are chosen in the same way. Finally, the number of shared packets between the relay nodes \(\mu\) cannot exceed either \(m_1\) or \(m_2\). Its starting value should be 0 if \(m_1 + m_2 \leq N_R\) and \(m_1 + m_2 - N_R\) otherwise.

**C. Multiple-relay Case**

Consider now the general case of \(L\) relays. During the first stage of the transmission, the source node multicasts to \(L\) relay nodes. Each transmitted packet can be received by a single relay, a selection of at least two relays or by none of the relays. In total, the number of possible outcomes is equal to \(\sum_{i=0}^L \binom{L}{i} = 2^L\). By analogy to the case of two relays, the coding matrices of all relays \(\mathbf{C}_R^j, j = 1, \ldots, L\), can be stacked together starting with the rows corresponds received by all \(L\) relays, followed by rows received by all combinations of \(L - 1\) relays, and so on. In total, there are \(2^L - 1\) groups of rows received by all combinations of the relays. The equivalent re-encoding matrix \(\mathbf{G}\) can be constructed accordingly.

**Example 3.1**: Let \(L = 3\). Matrix \(\mathbf{G}\) has the following structure:

\[
\mathbf{G} = \begin{pmatrix}
\mathbf{G}^{(\mu)}_1 & \mathbf{G}^{(\mu)}_2 & \mathbf{G}^{(\mu)}_3 \\
\mathbf{G}^{(\mu)}_1 & \mathbf{G}^{(\mu)}_2 & \mathbf{G}^{(\mu)}_3 \\
\mathbf{G}^{(\mu)}_1 & \mathbf{G}^{(\mu)}_2 & \mathbf{G}^{(\mu)}_3 \\
\end{pmatrix}
\]

Here, matrices \(\mathbf{G}^{(\mu)}_j, j = 1, 2, 3\), are composed of the columns of \(\mathbf{G}_j\) that correspond to \(\mu\) common packets received by all relays, by analogy to the two-relay network, while matrices \(\mathbf{G}_{ij}, i, j = 1, 2, 3\), are composed of the columns of \(\mathbf{G}_j\) that correspond to packets shared between the \(i\)-th and \(j\)-th relays.

The zero matrices correspond to packets not received by at least one, but fewer than three relays. In total, matrix \(\mathbf{G}\) has \(2^L - 1\) vertical blocks corresponding to all possible combinations of the relays.

By analogy to the two-relay case, we denote by \(\hat{\mathbf{G}}\) a \((\sum_{j=1}^L m'_j) \times m\) matrix obtained from \(\mathbf{G}\) by removing the rows corresponding to lost packets between the relay nodes \(D\), where \(m\) denotes the total number of columns in \(\mathbf{G}\). The probability that the coding matrix at \(D\) will be full rank can be then computed by (19), which requires the knowledge of the probability distribution of the rank of matrix \(\mathbf{G}\). As can be observed from Example 3.1 even for the case of three relays, matrix \(\mathbf{G}\) (and hence \(\hat{\mathbf{G}}\)) has a complex block-angular structure, to which the result of Lemma 3.2 cannot be extended in a straightforward way. Even if the exact expression was found, the result would likely to be too complex for practical use.

Instead of finding the exact expression for the probability that \(\mathbf{G}\) has a certain rank, we establish the following bound:

**Lemma 3.3**: The probability that the \(\sum_{j=1}^L m'_j\) matrix \(\hat{\mathbf{G}}\), obtained from the equivalent re-encoding matrix \(\mathbf{G}\) by removing the rows due to packet loss, has a certain rank \(r\), is upper bounded as follows:

\[
\Pr[\text{rank}(\hat{\mathbf{G}}) = r] \leq \hat{P}_r \left( \sum_{j=1}^L m'_j, m \right).
\]

**Proof**: It can be observed that bound (21) is based on replacing deterministic zero elements in \(\mathbf{G}\) arising from packet loss with random elements generated from \(\mathbb{F}_p\). Intuitively, allowing those elements to be non-zero increases the probability of the matrix to have the same rank, hence the bound is an upper one. The equality in (21) will correspond to the case when all relays share the same set of received packets; in other words, when there are no deterministic zero elements in \(\mathbf{G}\).

Bound (21) is expected to be tighter for lower PER values between the source and relay nodes. Indeed, in such case a packet transmitted by \(S\) is more likely to be received by all
relays, thus making the number of zero elements in $\tilde{G}$ smaller. Employing \cite{21}, we now formulate the bound for the decoding probability for a network with $L$ relays:

**Theorem 3.2:** The decoding probability for an $L$-relay network operated under the AF scheme is upper bounded as follows:

$$P_R^{(L)} \leq P(N_S, \epsilon_{SR}) \cdot \sum_{m_1', \ldots, m_L'} B(m_j', N_R, \epsilon_{RD}) \mathbb{P}\left(\sum_{j=1}^L m_j', K\right).$$  \hspace{1cm} (22)

where $\epsilon_{SR} = \prod_{j=1}^L \epsilon_{SR,j}$ and the summation is performed such that $\sum_{j=1}^L m_j' \geq K$ and $m_j' \leq N_R$.

**Proof:** Consider the probability that the matrix of coding coefficients of the destination node $C_D$ has rank $K$, which is given by \cite{19}. Applying Lemma 5.3 to \cite{19} yields the following bound:

$$\Pr[\text{rank}(C_D) = K] \leq \sum_r \Pr(r) \mathbb{P}\left(\sum_{j=1}^L m_j', m\right) \mathbb{P}(r, K).$$  \hspace{1cm} (23)

Furthermore, employing Lemma 3.3 \cite{23} can be rewritten as follows:

$$\Pr[\text{rank}(C_D) = K] \leq \mathbb{P}(m, K) \mathbb{P}\left(\sum_{j=1}^L m_j', K\right).$$  \hspace{1cm} (24)

From \cite{24}, it can be observed that the probability that matrix $C_D$ has rank $K$ depends on $m$, the total number of unique packets received by all $L$ relays, and the sum of the numbers of packets $m_j'$ received by $D$ from the $j$-th relay, $j = 1, \ldots, L$. The first random variable has the binomial distribution, in which the probability of success is the probability that a packet transmitted by the source is received by at least one relay. In other words, the PMF of $m$ is equal to $B(m, N_S, \epsilon_{SR})$, where $\epsilon_{SR} = \prod_{j=1}^L \epsilon_{SR,j}$. On the other hand, the PMF of $m_j'$, $j = 1, \ldots, L$, is equal to $B(m_j', N_R, \epsilon_{RD})$. By marginalising \cite{24} accordingly, rearranging terms and employing \cite{1}, \cite{22} can be readily obtained.

Comparing the bound of Theorem 3.2 to the decoding probability of a single-relay network \cite{11}, some similarities can be observed. Indeed, the first factor in \cite{22} is the decoding probability for a point-to-point link, in which the set of $L$ relays is replaced by a single equivalent node characterised by the PER equal to $\epsilon_{SR} = \prod_{j=1}^L \epsilon_{SR,j}$. Similarly, the remainder of the right-hand side of \cite{22} can be thought of as the probability of another point-to-point link, in which the source performs $L$ rounds of $N_R$ coded transmissions with the PER $\epsilon_{RD}$ in each round, $j = 1, \ldots, L$. In other words, bound \cite{22} approximates the $L$-relay network as an equivalent single-relay network. For a single-relay case, bound \cite{22} becomes exact and equal to \cite{11}.

### IV. Numerical Results

In this section, we analyse the performance of the $L$-relay network described in Section II via Monte Carlo simulation and compare the results with those predicted by the derived analytical framework. The performance in all cases is evaluated in terms of the probability of successful decoding at the destination node as a function of the number of coded transmissions. For Monte Carlo simulation, we employed the Kodo C++ network coding library \cite{24}. Each simulated result was obtained by averaging over $10^7$ iterations. For simplicity, all links were assumed to have the same PER $\epsilon = \epsilon_{SR,j} = \epsilon_{RD,j}$, $j = 1, \ldots, L$. In addition, each node was assumed to transmit the same number of coded packets $N = N_S = N_R$. The network and code parameters used throughout this section are summarised in Table IV.

We start with the two-relay case operated under the AF scheme and compare the results predicted by the exact expression \cite{18} with simulated ones. In addition, we also simulate the performance of the two-relay DF scheme in order to compare the two schemes. Finally, we plot the performance predicted by bound \cite{13}.

The results for $L = 2$ relays and the binary field are presented in Fig. 3 for two values of PER $\epsilon$ and various sizes of the source message $K$. As predicted in Section III-B, it can be seen that bound \cite{13} is too loose, especially for the lower PER. At the same time, the performance predicted by the exact expression \cite{18} perfectly matches the simulated one in all cases. Comparing the two relaying schemes, it is clear that the AF mechanism analysed in this paper is superior to the DF scheme, especially for $\epsilon = 0.4$ and larger values of $K$. Indeed, for such scenarios, the relays operating under the DF scheme are not likely to decode and therefore will simply retransmit a fraction of packets transmitted by the source. By contrast, the relays operating under the AF scheme will always transmit $N$ coded packets, thus increasing the probability of collecting $K$ linearly independent vectors of coding coefficients at $D$.

![Table II](image)

| Parameter                        | Values                       |
|----------------------------------|------------------------------|
| Number of relays $L$             | $\{2, 3, 4\}$              |
| PER $\epsilon$                  | $\{0.1, 0.4\}$             |
| Number of source packets $K$     | $\{5, 10, 15, 20\}$        |
| Number of transmissions $N$      | $\{K, K + 1, \ldots, K + 10\}$ |
| Finite field size $q$            | $\{2, 256\}$              |

1 Due to the lack of the analytical results for the DF scheme in the literature, only simulated results will be plotted.
Fig. 2. Probability of successful decoding as a function of the number of transmissions $N$ for $L = 2$ relays, binary field and the PER equal to (a) $0.1$ and (b) $0.4$.

Fig. 3. Probability of successful decoding as a function of the number of transmissions $N$ for $L = 2$ relays, non-binary field ($q = 256$) and the PER equal to $0.4$.

We now consider the cases of more than two relays. We employ bound (22) and compare the results with the simulated ones. Only the AF scheme is considered, since the DF scheme proved to be inferior even for two relays. Fig. 4 illustrates the results for $L = 3$ and $4$. It can be observed that bound (22) accurately predicts the performance for all considered values of $\epsilon$, $K$ and $q$.

Based on the results so far, bound (22) proposed in Section III-C appears to be tight for all considered network scenarios. To further illustrate the accuracy of the bound, Fig. 5 plots the Mean Squared Error (MSE) between (22) and simulated results as a function of source message size $K$, for different numbers of relays and different values of PER. The MSE was obtained by averaging the squared absolute difference between (22) and simulated results over all possible numbers of transmissions, i.e., from $N = K$ to $K + 10$. First, it can be seen that the bound is tighter for the smaller PER, which is due to the fact that the relays are likely to share more packets, thus making approximation (21) closer. Second, the bound becomes tighter as the number of relays $L$ increases. To understand this phenomenon, consider the equivalent re-encoding matrix $G$ described in Section III-C. The expected number of its rows and columns is equal to $LN(1 - \epsilon)$ and $N(1 - \epsilon L)$, respectively. Clearly, the difference between the two increases with $L$, and for a sufficiently large $L$ matrix $G$ is full rank with a high probability, despite the presence of fixed zeros due to packet loss. In other words, both sides of inequality (21) will get closer to 1 as $L$ increases, thus making bound (22) tighter. For the same reason, the accuracy of the bound increases with $K$. If fact, for a sufficiently large source message size, bound (22) is tight enough even for the case...
of two relays, which means that it can be used as a simpler alternative to the exact expression \([23]\).

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have addressed the theoretical performance analysis of a single-source relay-assisted network operating under RLNC. We employed the classical network coding scheme, in which the relay nodes simply re-encode packets they receive, without resorting to decoding. For networks with up to two relays, we derived an exact expression for the probability of successful decoding. For the general case of an arbitrary number of relays, we proposed a novel bound.

The theoretical results were verified via Monte Carlo simulation for various network and code parameters. In the case of two relays, it was shown that the results predicted by the exact expression perfectly match the simulated ones. In addition, the considered relaying scheme was shown to significantly outperform the DF scheme proposed in the literature previously. In the case of more than two relays, the proposed bound was shown to be very tight, with the accuracy increasing as the number of relays or the source message size grows.

The novel theoretical framework proposed in this paper can be used to predict the performance of relay networks and to select optimum network and code parameters without resorting to complex simulations. In the future, it is planned to extend the framework to networks with multiple sources and destinations, as well as to consider other performance metrics, such as delay and energy efficiency.

APPENDIX A

PROOF OF LEMMA 3.2

To show that \([10]\) holds, we employ \([4]\) and \([5]\) and assume, without loss of generality, that \(m \leq m'\):

\[
\sum_{r=K}^{m} \mathbb{P}_r(m', m) \mathbb{P}(r, K) = \mathbb{P}(m, K) \mathbb{P}(m', K) \\
\cdot \sum_{r=K}^{m} q^{K(m'+m)} q^{-K} \prod_{i=K}^{m-1} \left( \frac{q^m - q^i}{q^r - q^i} \right) \left( q^{m'} - q^i \right).
\]

The last sum can be shown to be equal to 1 by changing the summation index to \(r' = r - K\) and realising that the summation operand is equal to \(\mathbb{P}_r(m' - K, m - K)\).

APPENDIX B

PROOF OF LEMMA 3.2

We follow the same line of thought as in the proof of Theorem 2 in \([23]\) and Corollary 1 in \([18]\). Let us assume that matrices \(B_1\) and \(B_2\) have ranks \(x \leq a_1, b_1\) and \(y \leq a_2, b_2\), respectively. We can therefore identify a set of \(x\) linearly independent rows in the first \(a_1\) rows of \(X\) and a set \(y\) linearly independent rows in the last \(a_2\) rows of \(X\). Given \(x\) and \(y\), the probability in question does not depend on the \(x\) rows of \(A_1\) and on the \(y\) rows of \(A_2\), but depends only on whether the combination of the remaining \(a_1 - x\) rows of \(A_1\) and the remaining \(a_2 - y\) rows of \(A_2\) contain \(r - x - y\) linearly independent rows. The result \([21]\) then directly follows. The starting values of \(x\) and \(y\) are determined by the third term in \([21]\), based on which the rank \(r - x - y\) cannot be smaller than one of the dimensions, \(a\). The maximum values of \(x\) and \(y\) are determined by the dimensions of \(B_1\) and \(B_2\) and also by the fact that the total rank cannot be larger than \(r\).

REFERENCES

[1] X. Fafoutis, E. Tsimbalo, E. Mellios, G. Hilton, R. Piechocki, and I. Craddock, “A Residential Maintenance-Free Long-Term Activity Monitoring System for Healthcare Applications,” EURASIP Journal on Wireless Communications and Networking, vol. 2016, no. 1, pp. 1–20, 2016.
[2] D. Gomez-Barquero, D. Gozalvez, and N. Cardona, “Application Layer FEC for Mobile TV Delivery in IP Datacast Over DVB-H Systems,” IEEE Transactions on Broadcasting, vol. 55, pp. 396–406, June 2009.
[3] J. W. Byers, M. Luby, and M. Mitzenmacher, “A Digital Fountain Approach to Asynchronous Reliable Multicast,” IEEE Journal on Selected Areas in Communications, vol. 20, pp. 1528–1540, Oct. 2002.
[4] P. Wang, G. Mao, Z. Lin, M. Ding, W. Liang, X. Ge, and Z. Lin, “Performance Analysis of Raptor Codes Under Maximum Likelihood Decoding,” IEEE Transactions on Communications, vol. 64, pp. 906–917, Mar. 2016.
[5] E. Magli, M. Wang, P. Frossard, and A. Markopoulou, “Network Coding Meets Multimedia: A Review,” IEEE Transactions on Multimedia, vol. 15, pp. 1195–1212, Aug. 2013.
[6] T. Ho, R. Koetter, M. Méard, D. Karger, and M. Effros, “The Benefits of Coding over Routing in a Randomized Setting,” in Proc. of IEEE ISIT 2003, p. 7803, June 2003.
[7] T. Ho, M. Méard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, “A Random Linear Network Coding Approach to Multicast,” IEEE Transactions on Information Theory, vol. 52, pp. 4413–4430, Oct. 2006.
[8] D. S. Lun, M. Méard, R. Koetter, and M. Effros, “On Coding for Reliable Communication over Packet Networks,” in Proc. of Proc. 42nd Annu. Allerton Conf. Communication, Control, and Computing, Jan. 2004.
[9] D. S. Lun, M. Méard, R. Koetter, and M. Effros, “Further results on coding for reliable communication over packet networks,” in Proc. of ISIT 2005, pp. 1848–1852, Sept 2005.
[10] C. F. Chiasserini, E. Viterbo, and C. Casetti, “Decoding probability in random linear network coding with packet losses,” IEEE Communications Letters, vol. 17, pp. 1–4, November 2013.
[11] J. T. Seong and H. N. Lee, “Predicting the performance of cooperative wireless networking schemes with random network coding,” IEEE Transactions on Communications, vol. 62, pp. 2951–2964, Aug 2014.
[12] A. S. Khan and I. Chatzigeorgiou, “Improved bounds on the decoding failure probability of network coding over multi-source multi-relay networks,” IEEE Communications Letters, vol. 20, pp. 2035–2038, Oct 2016.
[13] C. Cooper, “On the distribution of rank of a random matrix over a finite field,” Random Struct. Algorithms, vol. 17, no. 3–4, pp. 197–212, 2000.
[14] X. Shi, M. Medard, and D. E. Lucani, “Whether and where to code in the wireless packet erasure relay channel,” *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 1579–1389, Aug 2013.

[15] M. V. Pedersen, D. E. Lucani, F. H. P. Fitzek, C. W. Sorensen, and A. S. Badr, “Network coding designs suited for the real world: What works, what doesn’t, what’s promising,” in *Proc. of IEEE ITW 2013*, pp. 1–5, Sept. 2013.

[16] D. Ferreira, R. A. Costa, and J. Barros. “Real-Time Network Coding for Live Streaming in Hyper-Dense WiFi Spaces,” *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 773–781, Apr. 2014.

[17] A. S. Khan and I. Chatzigeorgiou, “Performance Analysis of Random Linear Network Coding in Two-Source Single-Relay Networks,” in *Proc. IEEE ICC 2015*, pp. 991–996, June 2015.

[18] E. Tsimbalo, A. Tassi, and R. J. Piechocki, “Novel Performance Analysis of Network Coded Communications in Single-Relay Networks,” in *Proc. of IEEE GLOBECOM 2016*, pp. 1–6, Dec. 2016.

[19] O. Trullols-Cruces, J. M. Barcelo-Ordinas, and M. Fiore, “Exact Decoding Probability Under Random Linear Network Coding,” *IEEE Communications Letters*, vol. 15, pp. 67–69, Jan. 2011.

[20] A. S. Khan and I. Chatzigeorgiou, “Opportunistic Relaying and Random Linear Network Coding for Secure and Reliable Communication,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 223 – 234, 2018.

[21] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*. Cambridge University Press, 2nd ed., 2001.

[22] E. Tsimbalo, A. Tassi, and R. J. Piechocki, “Reliability of Multicast under Random Linear Network Coding,” *IEEE Transactions on Communications*, vol. 1, pp. 1 – 1, Feb. 2018.

[23] P. J. S. G. Ferreira, B. Jesus, J. Vieira, and A. J. Pinho, “The Rank of Random Binary Matrices and Distributed Storage Applications,” *IEEE Communications Letters*, vol. 17, pp. 151–154, Jan. 2013.

[24] M. V. Pedersen, J. Heide, and F. H. P. Fitzek, *Kodo: An Open and Research Oriented Network Coding Library*, vol. 6827 of *Lecture Notes in Computer Science*, pp. 145–152. Springer, 2011.