

# The Impact of Partial Packet Recovery on the Inherent Secrecy of Random Linear Coding

Ioannis Chatzigeorgiou

School of Computing and Communications

Lancaster University, United Kingdom

Email: i.chatzigeorgiou@lancaster.ac.uk

**Abstract**—This paper considers a source, which employs random linear coding (RLC) to encode a message, a legitimate destination, which can recover the message if it gathers a sufficient number of coded packets, and an eavesdropper. The probability of the eavesdropper accumulating enough coded packets to recover the message, known as the intercept probability, has been studied in the literature. In our work, the eavesdropper does not abandon its efforts to obtain the source message if RLC decoding has been unsuccessful; instead, it employs partial packet recovery (PPR) offline in an effort to repair erroneously received coded packets before it attempts RLC decoding again. Results show that PPR-assisted RLC decoding marginally increases the intercept probability, compared to RLC decoding, when the channel conditions are good. However, as the channel conditions deteriorate, PPR-assisted RLC decoding significantly improves the chances of the eavesdropper recovering the source message, even if the eavesdropper experiences similar or worse channel conditions than the destination.

**Index Terms**—Information theoretic security, network coding, partial packet recovery, syndrome decoding, spark.

## I. INTRODUCTION

Random linear coding (RLC), which encompasses fountain coding and network coding, generates coded packets that are random linear combinations of input data packets. RLC has the potential to increase throughput and improve robustness [1] but has also been considered for information-theoretic secure communications because eavesdroppers cannot necessarily retrieve source data from a subset of received coded packets. Ning and Yeung [2] first formulated the concept of secure network coding, while Adeli and Liu [3] studied probabilistic weak security for linear network coding. Niu *et al.* [4] computed the intercept probability of RLC for wireless broadcast applications, that is, the probability that an eavesdropper will accumulate the required number of coded packets for the reconstruction of the source data. The analysis in [4] was later revisited and refined by Khan *et al.* [5]. Since then, the intercept probability has been analyzed in different settings; for example, layered RLC was studied in [6], sparse RLC was investigated in [7] and applications of secure RLC-encoded data transfer were explored in [8], [9].

RLC decoding at a legitimate destination or an eavesdropper discards received coded packets that have been corrupted by errors and attempts to reconstruct the source packets from correctly received coded packets. In this paper, we assume that an eavesdropper does not discard corrupted coded packets but stores them and utilizes them offline, i.e., at a later instance, if

RLC decoding has been unsuccessful. Partial packet recovery (PPR), proposed in [10], is employed to repair stored packets in polynomial time and improve the chances of the eavesdropper recovering the source packets. The motivation for this paper is to investigate RLC decoding combined with PPR from the perspective of an eavesdropper, study the impact of PPR-assisted RLC decoding on the intercept probability, contrast it with conventional RLC decoding and observe tradeoffs between secrecy and reliability.

The remainder of this paper has been organized as follows: Section II presents the system model, including definitions and notation. Section III proposes full-rank nonsystematic RLC encoding at the source, while Section IV studies PPR-assisted RLC decoding at an eavesdropper and briefly describes two methods for repairing corrupted packets. The chances that a legitimate destination or an eavesdropper will recover the source message are discussed in Section V. Results, which give insights into secrecy-reliability tradeoffs, are presented in Section VI and key findings are summarized in Section VII.

## II. SYSTEM MODEL

We consider a network setup whereby a source  $S$  transmits a message to a legitimate destination  $D$  in the presence of an eavesdropper  $E$ , as depicted in Fig. 1. Before initiating transmission, the source segments the message into  $K$  packets. Each source packet has been modeled as a sequence of  $L$  symbols from a finite field of  $q$  elements, denoted by  $\mathbb{F}_q$ , where  $q$  is a prime power. The  $K$  source packets of length  $L$  can be expressed as a matrix  $\mathbf{U} \in \mathbb{F}_q^{K \times L}$ , where  $\mathbb{F}_q^{K \times L}$  denotes the set of all  $K \times L$  matrices over  $\mathbb{F}_q$ . RLC [11] is used to encode the  $K$  source packets into  $N \geq K$  coded packets, each of length  $L$  symbols. The  $N$  coded packets can also be expressed in matrix form as  $\mathbf{X} \in \mathbb{F}_q^{N \times L}$ . The relationship between matrices  $\mathbf{X}$  and  $\mathbf{U}$  is:

$$\mathbf{X} = \mathbf{G}\mathbf{U}, \quad (1)$$

where  $\mathbf{G} \in \mathbb{F}_q^{N \times K}$  is known as the *generator matrix*. Typically, the entries of  $\mathbf{G}$  are chosen uniformly at random from  $\mathbb{F}_q$ . As explained in [10], a seed can be used to initialize the pseudo-random number generator that outputs the entries of  $\mathbf{G}$ . Given that the value of the seed can be conveyed in the headers of the coded packets, we assume that both the destination and the eavesdropper have knowledge of  $\mathbf{G}$ . Note that the eavesdropper could be a legitimate destination in a different occasion –

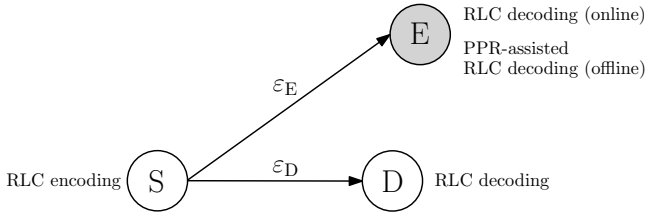


Fig. 1. Diagram of the system model, where  $\varepsilon_D$  and  $\varepsilon_E$  denote the packet error probabilities of the channels linking the source S to the destination D and the eavesdropper E. Both the destination and the eavesdropper employ RLC decoding to recover the source message. The eavesdropper resorts to PPR-assisted RLC decoding, if RLC decoding alone has been unsuccessful.

hence, it is equipped with the same pseudo-random number generator – but is not authorized to receive the private source message in this occasion. The link connecting the source to the destination has been modeled as a memoryless channel characterized by packet error probability  $\varepsilon_D$  or, equivalently, by bit error probability  $1 - (1 - \varepsilon_D)^{1/L}$ . Similarly, the link connecting the source to the eavesdropper is characterized by packet error probability  $\varepsilon_E$ . We assume that the eavesdropper experiences, on average, similar or worse channel conditions than the destination, that is,  $\varepsilon_E \geq \varepsilon_D$ .

Let  $N_D$  be the number of coded packets that have been successfully received by the destination, and  $\mathcal{D}$  be the ordered set of the row indices of  $\mathbf{X}$  that correspond to correctly received coded packets, i.e.,  $|\mathcal{D}| = N_D \leq N$ . The destination constructs matrix  $\mathbf{X}_{\mathcal{D}}$  from the  $N_D$  correctly received coded packets and matrix  $\mathbf{G}_{\mathcal{D}}$  from  $N_D$  rows of  $\mathbf{G}$  with indices in  $\mathcal{D}$ . The source message, represented by matrix  $\mathbf{U}$ , can be obtained from  $\mathbf{X}_{\mathcal{D}} = \mathbf{G}_{\mathcal{D}} \mathbf{U}$  if the rank of  $\mathbf{G}_{\mathcal{D}}$  is  $K$ . In that case, the relationship  $\mathbf{X}_{\mathcal{D}} = \mathbf{G}_{\mathcal{D}} \mathbf{U}$  can be seen as a system of  $N_D \geq K$  linear equations, which can be reduced to a system of  $K$  linearly independent equations with  $K$  unknowns, i.e., source packets. This  $K \times K$  system can be solved, e.g., using Gaussian elimination, and return a unique solution for  $\mathbf{U}$ . Similarly, the overhearing eavesdropper collects  $N_E$  coded packets, builds the ordered set  $\mathcal{E}$  of row indices, constructs matrices  $\mathbf{X}_{\mathcal{E}}$  and  $\mathbf{G}_{\mathcal{E}}$ , and attempts to obtain  $\mathbf{U}$  from  $\mathbf{X}_{\mathcal{E}} = \mathbf{G}_{\mathcal{E}} \mathbf{U}$ .

*Khan et al.* considered this system in [5] and derived closed-form expressions for the probability that the eavesdropper will gather enough coded packets to recover the source message. The authors also quantified the gain in secrecy when the source broadcasts up to  $N_{\max}$  coded packets but ceases transmission as soon as the destination sends a notification over a perfect feedback channel confirming receipt of the source message. In this paper, the eavesdropper attempts to decode the  $N_E$  correctly received coded packets concurrently with the legitimate destination, as in [5], but does not discard the  $N - N_E$  erroneously received coded packets. Instead, the eavesdropper employs PPR [10] offline in an effort to repair received coded packets that have been corrupted by errors and improve the chances of the RLC decoder recovering the source message. Section III discusses RLC encoding at the source, while Section IV adapts the PPR process presented in [10]

to the system model under consideration and describes PPR-assisted RLC decoding at the eavesdropper.

### III. RLC ENCODING AT THE SOURCE

RLC encoding is typically *non-systematic*, that is, the  $K$  source packets are encoded into  $N$  packets, which are random linear combinations of the source packets, as described in Section II. The probability that the legitimate destination will successfully recover the  $K$  source packets from  $N_D \leq N$  received coded packets is given by [12]:

$$P_{\text{ns}}(N_D, K) = \prod_{i=0}^{K-1} \left[ 1 - q^{-(N_D-i)} \right]. \quad (2)$$

RLC encoding can also be *systematic*, in which case the first  $K$  of the  $N$  transmitted packets are identical to the  $K$  source packets while the remaining  $N - K$  packets are random linear combinations of the source packets. In systematic RLC, the generator matrix  $\mathbf{G}$  can be expressed in standard form:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_K \\ \mathbf{P} \end{bmatrix}, \quad (3)$$

where  $\mathbf{I}_K$  is the  $K \times K$  identity matrix and  $\mathbf{P}$  is a  $(N - K) \times K$  randomly-generated matrix with entries from  $\mathbb{F}_q$ . The probability of the destination recovering the  $K$  source packets, upon receipt of  $N_D \leq N$  error-free packets, can be calculated from:

$$P_s(N, N_D, K) = \sum_{h=h_{\text{lim}}}^K \frac{\binom{K}{h} \binom{N-K}{N_D-h}}{\binom{N}{N_D}} P_{\text{ns}}(N_D - h, K - h), \quad (4)$$

where  $h_{\text{lim}} = \max(0, N_D + K - N)$  [13].

In systematic RLC, the eavesdropper could receive some of the source packets and could thus gain direct access to parts of the source message even before it attempts to decode the collected packets. Non-systematic RLC offers the advantage of inherent secrecy but exhibits a lower probability of message recovery at the destination than systematic RLC, as proven in [13]. For example, let us assume that the destination successfully received the first  $K$  transmitted packets. If systematic RLC has been used, the destination will recover the source message. If non-systematic RLC has been employed, recovery of the source message is not guaranteed, as the first  $K$  transmitted packets may not be linearly independent. In that case, the destination will need more than  $K$  packet transmissions at the risk of allowing the eavesdropper to collect more packets and possibly recover the source message.

In order to combine the inherent secrecy of non-systematic RLC with the reliability of systematic RLC, we could use non-systematic RLC constructions defined by a generator matrix of the form:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_K \\ \mathbf{P} \end{bmatrix}. \quad (5)$$

To obtain  $\mathbf{G}$ , the seed of the pseudo-random number generator should be chosen to ensure that the top  $K$  randomly-generated rows of  $\mathbf{G}$  are linearly independent, that is,  $\text{rank}(\mathbf{G}_K) = K$ .

Column-wise Gaussian elimination can transform  $\mathbf{G}$  into standard form and convert the proposed *full-rank non-systematic RLC* into an equivalent systematic RLC, i.e.,

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_K \\ \mathbf{P} \end{bmatrix} \xrightarrow[\text{column operations}]{\text{elementary}} \mathbf{G}' = \begin{bmatrix} \mathbf{I}_K \\ \mathbf{P}' \end{bmatrix}. \quad (6)$$

Full-rank non-systematic RLC retains the inherent secrecy features of non-systematic RLC, while its equivalence to systematic RLC suggests that (4) provides the probability of the destination recovering the  $K$  source packets when  $N_D \leq N$  coded packets have been received without errors. Expression (6) will be used in the following section, which looks into how the eavesdropper attempts to retrieve the source message.

#### IV. EAVESDROPPING ENHANCED BY PPR

We explained in Section II that  $\mathcal{E}$  is the ordered set of  $N_E$  row indices of  $\mathbf{X}$  that correspond to error-free coded packets received by the eavesdropper. The indices of the remaining  $N - N_E$  rows of  $\mathbf{X}$ , which identify erroneously received coded packets stored at the eavesdropper, form  $\bar{\mathcal{E}} = \{1, \dots, N\} \setminus \mathcal{E}$ . Based on  $\mathcal{E}$ , the eavesdropper constructs  $\mathbf{X}_{\mathcal{E}}$  and  $\mathbf{G}_{\mathcal{E}}$ , and uses RLC decoding to solve  $\mathbf{X}_{\mathcal{E}} = \mathbf{G}_{\mathcal{E}} \mathbf{U}$  for  $\mathbf{U}$  and recover the source message. If  $\text{rank}(\mathbf{G}_{\mathcal{E}}) = K$ , the source message can be recovered. If  $\text{rank}(\mathbf{G}_{\mathcal{E}}) < K$ , the eavesdropper turns to set  $\bar{\mathcal{E}}$ , and utilizes sparse recovery principles to repair erroneously received coded packets and complement RLC decoding.

More specifically, let  $\mathbf{Y}$  be an erroneous copy of  $\mathbf{X}$  that has been received by the eavesdropper, defined as:

$$\mathbf{Y} = \mathbf{X} + \mathbf{E}. \quad (7)$$

The *error matrix*  $\mathbf{E}$  contains non-zero elements in positions where errors have occurred and zero elements in the remaining positions. Note that the eavesdropper has knowledge of  $\mathbf{Y}$  and  $\mathbf{G}$  only. The  $N \times (N - K)$  *parity-check matrix*  $\mathbf{H}$  can be derived from the  $N \times K$  generator matrix  $\mathbf{G}$ , as follows:

$$\mathbf{H} = \left[ -\mathbf{P}' \mid \mathbf{I}_{N-K} \right]^T, \quad (8)$$

so that:

$$\mathbf{H}^T \mathbf{G} = \mathbf{0}. \quad (9)$$

Matrix  $\mathbf{P}'$  can be obtained from (6). Matrix negation in (8) is performed in  $\mathbb{F}_q$ , e.g.,  $-\mathbf{P}' = \mathbf{P}'$  in  $\mathbb{F}_2$ . Multiplication of  $\mathbf{Y}$  by  $\mathbf{H}^T$  produces the  $(N - K) \times L$  *syndrome matrix*  $\mathbf{S}$ , i.e.,  $\mathbf{S} = \mathbf{H}^T \mathbf{Y}$ . Using (9), we find that the relationship between the syndrome matrix  $\mathbf{S}$  and the error matrix  $\mathbf{E}$  is:

$$\mathbf{S} = \mathbf{H}^T \mathbf{Y} = \mathbf{H}^T (\mathbf{G} \mathbf{U} + \mathbf{E}) = \mathbf{H}^T \mathbf{E}. \quad (10)$$

Given that our focus is on the  $N - N_E$  erroneously received coded packets, we use the set  $\bar{\mathcal{E}}$  to isolate the  $N - N_E$  of the  $N$  rows of  $\mathbf{Y}$ ,  $\mathbf{X}$ ,  $\mathbf{E}$  and  $\mathbf{H}$ , and construct  $\mathbf{Y}_{\bar{\mathcal{E}}}$ ,  $\mathbf{X}_{\bar{\mathcal{E}}}$ ,  $\mathbf{E}_{\bar{\mathcal{E}}}$  and  $\mathbf{H}_{\bar{\mathcal{E}}}$ . As a result, expressions (7) and (10) change to:

$$\mathbf{Y}_{\bar{\mathcal{E}}} = \mathbf{X}_{\bar{\mathcal{E}}} + \mathbf{E}_{\bar{\mathcal{E}}} \quad (11)$$

and

$$\mathbf{S} = (\mathbf{H}_{\bar{\mathcal{E}}})^T \mathbf{E}_{\bar{\mathcal{E}}}, \quad (12)$$

respectively. If the  $j$ -th column of  $\mathbf{S}$  and  $\mathbf{E}_{\bar{\mathcal{E}}}$  is denoted by  $[\mathbf{S}]_{*,j}$  and  $[\mathbf{E}_{\bar{\mathcal{E}}}]_{*,j}$ , expression (12) can be re-written as  $L$  independent systems of  $N - K$  linear equations with  $N - N_E$  unknowns per equation:

$$[\mathbf{S}]_{*,j} = (\mathbf{H}_{\bar{\mathcal{E}}})^T [\mathbf{E}_{\bar{\mathcal{E}}}]_{*,j} \quad \text{for } j = 1, \dots, L. \quad (13)$$

Mohammadi *et al.* [10] observed that erroneously received coded packets usually contain large error-free segments, thus  $\mathbf{E}_{\bar{\mathcal{E}}}$  is a *sparse* matrix, that is,  $\mathbf{E}_{\bar{\mathcal{E}}}$  has more zero entries than non-zero entries. Based on this observation, the solution to (13) can be formulated as an ' $\ell_0$  minimization' problem:

$$[\hat{\mathbf{E}}_{\bar{\mathcal{E}}}]_{*,j} = \arg \min_{\mathbf{w}^T} \|\mathbf{w}\|_0 \quad (14a)$$

$$\text{subject to } (\mathbf{H}_{\bar{\mathcal{E}}})^T \mathbf{w}^T = [\mathbf{S}]_{*,j} \quad (14b)$$

where  $\mathbf{w} \in \mathbb{F}_q^{N-N_E}$  is a row vector that should satisfy constraint (14b) and have the minimum possible number of non-zero entries. To count the non-zero entries in  $\mathbf{w}$ , the  $\ell_0$  norm is used, which is defined as  $\|\mathbf{w}\|_0 = |w_1|^0 + \dots + |w_{N-N_E}|^0$  assuming that  $0^0 = 0$  [14].

Two PPR methods for the evaluation of  $\hat{\mathbf{E}}_{\bar{\mathcal{E}}}$  were proposed in [10]. The first approach, which is inspired by the compressed sensing (CS) literature, replaces the non-convex  $\ell_0$  norm in (14a) with the convex  $\ell_1$  norm, solves the optimization problem over the set of real numbers, and rounds off the derived values to the nearest elements of  $\mathbb{F}_q$ . The second approach, referred to as *syndrome decoding*, considers (14) and initiates an exhaustive search for a candidate solution; the sparsity of the row vector  $\mathbf{w}$  is gradually reduced, i.e., the number of non-zero entries in  $\mathbf{w}$  increases, and the search concludes when the sparsest vector  $\mathbf{w}$  that satisfies constraint (14b) has been identified.

When candidate solutions for the  $L$  columns of  $\hat{\mathbf{E}}_{\bar{\mathcal{E}}}$  have been obtained and  $\hat{\mathbf{E}}_{\bar{\mathcal{E}}}$  has been evaluated, an estimate of  $\mathbf{X}_{\bar{\mathcal{E}}}$ , denoted by  $\hat{\mathbf{X}}_{\bar{\mathcal{E}}}$ , can be derived using (11):

$$\hat{\mathbf{X}}_{\bar{\mathcal{E}}} = \mathbf{Y}_{\bar{\mathcal{E}}} - \hat{\mathbf{E}}_{\bar{\mathcal{E}}}. \quad (15)$$

Matrix  $\hat{\mathbf{X}}_{\bar{\mathcal{E}}}$  is the outcome of the eavesdropper's attempt to repair the received matrix  $\mathbf{Y}_{\bar{\mathcal{E}}}$ . Let  $\nu$  denote the number of rows in  $\hat{\mathbf{X}}_{\bar{\mathcal{E}}}$  that correspond to successfully repaired coded packets, e.g., packets that passed Cyclic Redundancy Check (CRC) verification. The indices of the  $\nu$  repaired coded packets are removed from set  $\bar{\mathcal{E}}$  and added to set  $\mathcal{E}$ , and the corresponding rows of  $\hat{\mathbf{X}}_{\bar{\mathcal{E}}}$  are moved to  $\mathbf{X}_{\mathcal{E}}$ . The cardinalities of sets  $\bar{\mathcal{E}}$  and  $\mathcal{E}$  change to  $N - N_E - \nu$  and  $N_E + \nu$ , respectively, while the dimensions of  $\hat{\mathbf{X}}_{\bar{\mathcal{E}}}$  and  $\mathbf{X}_{\mathcal{E}}$  change to  $(N - N_E - \nu) \times L$  and  $(N_E + \nu) \times L$ , respectively. The indices of the  $\nu$  repaired coded packets are also used to identify the rows of the generator matrix  $\mathbf{G}$  that should be inserted in  $\mathbf{G}_{\mathcal{E}}$ . If PPR increases the rank of the enlarged  $(N_E + \nu) \times K$  matrix  $\mathbf{G}_{\mathcal{E}}$  to  $K$ , the RLC decoder at the eavesdropper will recover the source message.

The computational complexity that PPR adds to RLC decoding for operations in  $\mathbb{F}_2$  has been discussed in [10]. Essentially, PPR allows the eavesdropper to improve the chances of repairing and decoding the collected coded packets at the cost

of increased complexity. This fact makes PPR suitable for offline use. The probability that the eavesdropper will recover the source message, referred to as the *intercept probability* [5], when RLC decoding is used, and the challenges in computing the intercept probability when PPR-assisted RLC decoding is employed, are discussed in the following section.

## V. DECODING AND INTERCEPT PROBABILITIES

We initially investigated the inherent secrecy of RLC in [5] and obtained a closed-form expression for the intercept probability when both the destination and the eavesdropper employ RLC decoding. This expression will serve as a benchmark for the intercept probability when the eavesdropper uses PPR-assisted RLC decoding. For this reason, and for the sake of completeness, we summarize below the analysis in [5].

### A. RLC Decoding

Recall that the source broadcasts  $N \geq K$  coded packets, where  $N$  is constrained by an upper limit  $N_{\max}$ . The destination and the eavesdropper receive  $N_D \leq N$  and  $N_E \leq N$  error-free coded packets, respectively. Let  $R$  denote a receiver, either the destination D or the eavesdropper E, that is,  $R \in \{D, E\}$ . Also, let  $X$  be a random variable, which represents the number of transmitted coded packets that are required for  $R$  to recover the  $K$  source packets. The Cumulative Distribution Function (CDF) of  $X$ , which describes the probability that the receiver  $R$  will recover the  $K$  source packets after  $N$  coded packets have been transmitted, for  $K \leq N \leq N_{\max}$ , takes the form:

$$\begin{aligned} F_R(N) &= \Pr\{X \leq N\} \\ &= \sum_{N_R=K}^N \binom{N}{N_R} (1 - \varepsilon_R)^{N_R} \varepsilon_R^{N-N_R} P_s(N, N_R, K). \end{aligned} \quad (16)$$

Note that  $P_s(N, N_R, K)$  has been used in place of  $P_{ns}(N_R, K)$  in (16) because typical non-systematic RLC, which was considered in [5], has been replaced by full-rank non-systematic RLC, which was defined in Section III. Based on the definition of  $F_R(N)$ , the probability that the destination D will recover the source message by the time  $N_{\max}$  coded packets have been transmitted or sooner, is given by  $F_R(N_{\max})$  for  $R = D$ , i.e.,

$$P_{\text{dec}}(N_{\max}) = F_D(N_{\max}), \quad (17)$$

and is commonly referred to as the *decoding probability*.

The Probability Mass Function (PMF) of  $X$ , which is the probability that the receiver  $R$  will recover the  $K$  source packets only after the  $N$ -th coded packet has been transmitted, but not sooner, can be obtained from the CDF as follows:

$$f_R(N) = \Pr\{X = N\} = \begin{cases} F_R(N) - F_R(N-1), & \text{if } K < N \leq N_{\max} \\ F_R(K), & \text{if } N = K. \end{cases} \quad (18)$$

The intercept probability can be expressed as the sum of the following two terms:

$$\begin{aligned} P_{\text{int}}(N_{\max}) &= \sum_{N=K}^{N_{\max}} f_D(N) F_E(N) \\ &\quad + F_E(N_{\max}) [1 - F_D(N_{\max})], \end{aligned} \quad (19)$$

where  $F_E(\cdot)$  and  $F_D(\cdot)$  can be obtained from (16) for  $R = E$  and  $R = D$ , respectively, and  $f_D(\cdot)$  can be obtained from (18) for  $R = D$ . The first term in (19) denotes the probability that the destination will recover the source message after the  $N$ -th coded packet has been transmitted, for  $K \leq N \leq N_{\max}$ , and the eavesdropper will recover it sooner or concurrently with the destination. The second term in (19) represents the probability that only the eavesdropper will be successful in recovering the source message when the maximum number of coded packet transmissions,  $N_{\max}$ , is reached.

Expression (19) assumes that the sequence of coded packets delivered over the source-to-destination link is independent of the sequence delivered over the source-to-eavesdropper link. In practice, the interdependence of the two sequences becomes weaker as the value of the product  $N\varepsilon_D$  or  $N\varepsilon_E$  increases [15] and, hence, the accuracy of expression (19) improves.

### B. PPR-assisted RLC Decoding at the Eavesdropper

Whereas the rank of  $\mathbf{G}_E$  is important in RLC decoding at the eavesdropper, the *spark* of the transpose of  $\mathbf{H}_{\bar{E}}$  plays a key role in PPR-assisted RLC decoding. The term ‘spark’ of a matrix was first coined by Donoho and Elad in [16] and is defined as follows for the case of  $(\mathbf{H}_{\bar{E}})^T$ :

**Definition 1** ([17], p.23). *The spark of  $(\mathbf{H}_{\bar{E}})^T$  is the smallest number of columns from  $(\mathbf{H}_{\bar{E}})^T$  that are linearly dependent. Using mathematical notation, we can write:*

$$\begin{aligned} \text{spark} \left( (\mathbf{H}_{\bar{E}})^T \right) &:= \min_{\mathbf{z}} \{ \|\mathbf{z}\|_0 \} \\ &\text{subject to } (\mathbf{H}_{\bar{E}})^T \mathbf{z}^T = \mathbf{0} \text{ for } \mathbf{z} \neq \mathbf{0}. \end{aligned}$$

If we consider the  $L$  systems of equations defined in (13) and focus on the solution to the  $j$ -th system described in (14), the spark of  $(\mathbf{H}_{\bar{E}})^T$  gives the following criterion for the uniqueness of sparse solutions:

**Theorem 1** ([17], p.24). *If a system of equations, as in (14b), has a solution  $\mathbf{w}^T$  obeying:*

$$\text{spark} \left( (\mathbf{H}_{\bar{E}})^T \right) > 2\|\mathbf{w}\|_0, \quad (20)$$

*then this solution is necessarily the sparsest possible.*

Computation of the intercept probability for an eavesdropper employing online RLC decoding and offline PPR-assisted RLC decoding requires the inclusion of additional terms in (19). These terms should account for the probability that the RLC decoder at the eavesdropper will fail to recover the source message at first, but will succeed if PPR repairs a sufficient number of erroneously received coded packets. More specifically, a correctly estimated column of  $\mathbf{E}_{\bar{E}}$ :

- either satisfies (14b) and obeys (20), i.e., the estimated column is the sparsest possible solution,
- or satisfies (14b) without obeying (20).

PPR-assisted RLC decoding will recover the  $K$  source packets:

- if PPR correctly estimates all of the  $L$  columns of  $\mathbf{E}_{\bar{E}}$ , because of a combination of (a) and (b), thus repairing

all of the erroneously received coded packets and reconstructing the full-rank generator matrix  $\mathbf{G}$ , or

- if PPR correctly estimates some of the  $L$  columns of  $\mathbf{E}_{\bar{\mathcal{E}}}$ , because of a combination of (a) and (b), and repairs some of the erroneous coded packets, leading to an increase in the rank of  $\mathbf{G}_{\mathcal{E}}$  to  $K$ , as described in Section IV.

Derivation of the probability of fully or partially reconstructing matrix  $\mathbf{E}_{\bar{\mathcal{E}}}$  and subsequently obtaining an increased-size full-rank matrix  $\mathbf{G}_{\mathcal{E}}$ , where  $\mathbf{G}_{\mathcal{E}}$  and  $\mathbf{E}_{\bar{\mathcal{E}}}$  are random matrices over  $\mathbb{F}_q$ , is a non-trivial problem and a closed-form expression is not available in the literature, to the best of our knowledge. For this reason, we use the analysis in Section V-A to develop and validate the simulation results of the system in Fig. 1, when the eavesdropper employs RLC decoding only. Then, we equip the eavesdropper with PPR capabilities and carry out simulations to measure the impact of PPR on the intercept probability, as we explain in the following section.

## VI. RESULTS AND DISCUSSION

In order to investigate the inherent secrecy of RLC, we consider a source employing full-rank non-systematic RLC over  $\mathbb{F}_2$  to encode a message that has been segmented into  $K = 20$  packets, each consisting of  $L = 128$  bits. The intercept probability  $P_{\text{int}}$  has been measured for different values of packet error probabilities  $\varepsilon_{\text{D}}$  and  $\varepsilon_{\text{E}}$ , when the destination employs RLC decoding and the eavesdropper also relies on RLC decoding or combines PPR with RLC decoding. Of the two PPR methods summarized in Section IV, we implemented syndrome decoding (SD). The CS-based method can recover the source message with a slightly higher probability than SD for a broad range of channel conditions but at a significantly higher computational cost [10]. Given that the information value of the source message may decline with time, we opted for the less time-consuming SD method to complement RLC decoding. We refer to this joint scheme as ‘RLC with SD’.

We first study the case where the source broadcasts coded packets until the destination recovers the source message, that is, the number of transmitted coded packets  $N$  is not constrained by an upper limit  $N_{\text{max}}$ . This is equivalent to setting  $N_{\text{max}} \rightarrow \infty$  in order to achieve  $P_{\text{dec}} = 1$ . The intercept probability has been plotted against  $\varepsilon_{\text{D}}$  in Fig. 2, where  $\varepsilon_{\text{D}}$  varies from 0.01 to 0.1 and  $\varepsilon_{\text{E}}$  takes values in  $\{0.1, 0.15, 0.2\}$ . For a fixed value of  $\varepsilon_{\text{E}}$ , we observe that as the value of  $\varepsilon_{\text{D}}$  increases and the source is compelled to broadcast more coded packets to ensure that the destination will recover the source message, the chances of the eavesdropper also reconstructing the source message increase. Note that expression (19), which has been adapted for full-rank non-systematic RLC encoding proposed in Section III, generates theoretical predictions for the intercept probability that closely match simulation results when the eavesdropper employs RLC decoding. Fig. 2 shows that if SD is used offline to complement RLC decoding, a gain in the intercept probability will be attained, which increases with  $\varepsilon_{\text{D}}$  and, in particular, with  $\varepsilon_{\text{E}}$ .

Let us now consider the case where the source sets a limit  $N_{\text{max}}$  on the number of transmitted coded packets, so that the

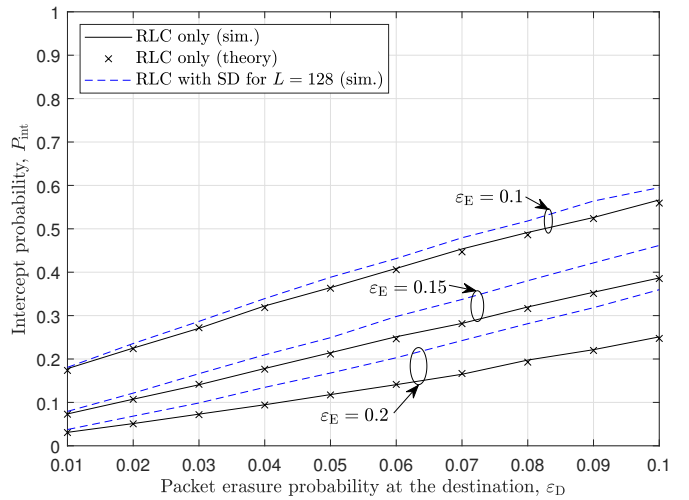


Fig. 2. Intercept probabilities achieved by RLC and by RLC with SD as a function of  $\varepsilon_{\text{D}}$  for  $K = 20$  and  $\varepsilon_{\text{E}} \in \{0.1, 0.15, 0.2\}$ . The source ceases transmission when the destination recovers the source message ( $P_{\text{dec}} = 1$ ).

decoding probability at the destination is at least 0.99, that is,  $P_{\text{dec}}(N_{\text{max}}) \geq 0.99$  but  $P_{\text{dec}}(N_{\text{max}} - 1) < 0.99$ . Expression (17) was used to determine  $N_{\text{max}}$  for different values of  $\varepsilon_{\text{D}}$ , and their relationship has been plotted in Fig. 3 (bottom). The intercept probability for increasing values of  $\varepsilon_{\text{D}}$  and  $N_{\text{max}}$  is shown in Fig. 3 (top), where  $\varepsilon_{\text{E}} = \varepsilon_{\text{D}} + \delta$  for  $\delta \in \{0, 0.05, 0.1\}$ . We observe that, as channel conditions deteriorate, the inherent secrecy of RLC is not severely compromised when the eavesdropper relies on RLC decoding to obtain the source message; for  $\varepsilon_{\text{E}} > \varepsilon_{\text{D}}$ , the intercept probability increases marginally but then plateaus and reduces slowly, as  $\varepsilon_{\text{D}}$  shifts from 0 to 0.5. However, when the eavesdropper combines RLC with SD, the intercept probability increases noticeably and approaches 1 for large values of  $\varepsilon_{\text{D}}$ . This result establishes that a poor source-to-destination channel, which could have been intentionally degraded by an eavesdropper who generates artificial noise, greatly improves the decoding capability of the eavesdropper when RLC with SD is used, even if the eavesdropper experiences similar or worse channel conditions than the destination.

Given that a scenario where  $\varepsilon_{\text{E}} = \varepsilon_{\text{D}}$  for  $\varepsilon_{\text{D}} \rightarrow 0$  is not likely to occur, limiting the value of the intercept probability for all other cases is of importance for systems that rely on information theoretic security. Instead of raising the value of  $N_{\text{max}}$  as  $\varepsilon_{\text{D}}$  increases in order to maintain a fixed decoding probability at the destination, as in Fig. 3, we fix the value of  $N_{\text{max}}$  in Fig. 4 and trade reliability for secrecy. We observe that a decrease in the value of  $N_{\text{max}}$  suppresses information leakage and reduces the intercept probability, when the eavesdropper combines SD with RLC decoding, but inevitably lowers the decoding probability. The value of  $N_{\text{max}}$  also determines the range of  $\varepsilon_{\text{D}}$  values over which the system could operate. For example, if  $N_{\text{max}} = 25$ , the destination will achieve  $P_{\text{dec}} = 0.8$  for  $\varepsilon_{\text{D}} = 0.1$  while the worst-case intercept probability ( $\varepsilon_{\text{E}} = \varepsilon_{\text{D}}$ ) will be limited to  $P_{\text{int}} = 0.57$ . However, the system may be required to support the same decoding probability ( $P_{\text{dec}} = 0.8$ )

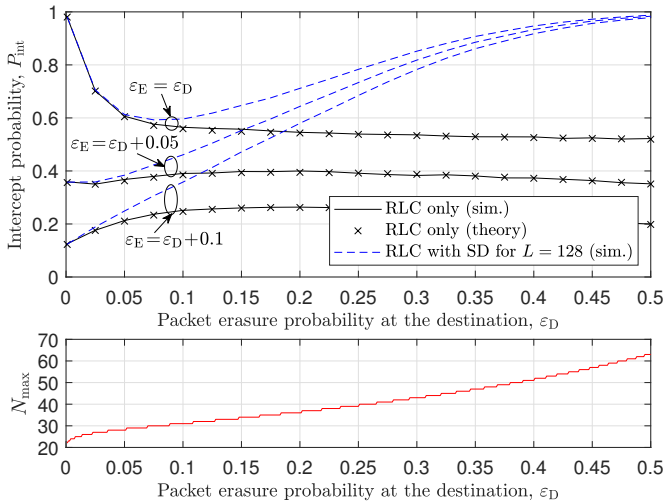


Fig. 3. Top: Intercept probabilities achieved by RLC and by RLC with SD as a function of  $\varepsilon_D$  for  $\varepsilon_E = \varepsilon_D + \delta$ , where  $\delta \in \{0, 0.05, 0.1\}$ ,  $K = 20$  and a varying  $N_{\max}$ . Bottom: For each value of  $\varepsilon_D$ , the value of  $N_{\max}$  ensures that the destination will recover the source message with probability  $P_{\text{dec}} \geq 0.99$ .

for  $\varepsilon_D = 0.2$ ; Fig. 4 indicates that  $N_{\max}$  will have to increase from 25 to 29 as long as the subsequent increase in the worst-case intercept probability ( $P_{\text{int}} = 0.7$ ) is tolerable, e.g., because additional security measures have been implemented or the source message does not carry sensitive information.

## VII. CONCLUSIONS

This paper investigated the impact of partial packet recovery on the inherent secrecy of random linear coding. Results demonstrated that syndrome decoding combined with random linear decoding can boost the probability of an eavesdropper recovering a message transmitted from a source to a legitimate destination. Of particular interest was the observation that this probability increases as the channel conditions between the source and the destination, but also between the source and the eavesdropper, deteriorate. The eavesdropper has thus an incentive to introduce artificial noise and actively impair data packets transmitted by the source. If information theoretic security is essential, system requirements should dictate a limit on the number of packet transmissions and a threshold on the channel conditions, beyond which transmission should cease in order to keep low the probability of the eavesdropper reconstructing the source message.

## ACKNOWLEDGEMENT AND RESEARCH REPRODUCIBILITY

This work was motivated by the outcomes of the third-year project completed by Adam Matthews and supervised by the author. The MATLAB<sup>®</sup> code for the simulations and the generation of the figures in Section VI can be downloaded from <https://github.com/loannisChatzigeorgiou/VTC2022-Spring>.

## REFERENCES

[1] R. Bassoli, H. Marques, J. Rodriguez, K. W. Shum, and R. Tafazolli, "Network coding theory: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1950–1978, 2013.

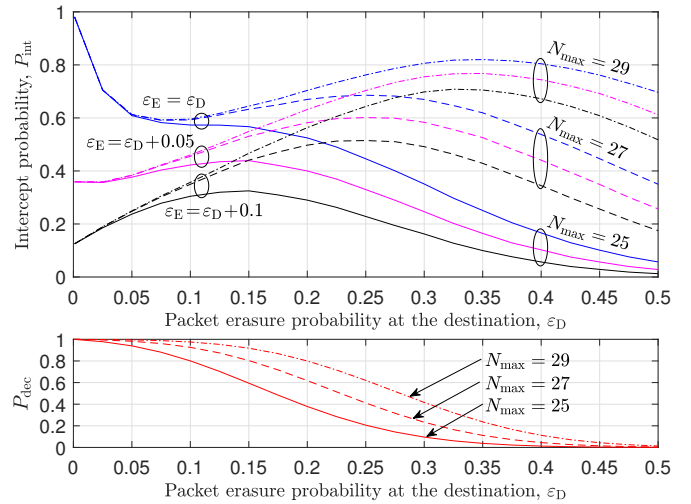


Fig. 4. Top: Intercept probabilities achieved by RLC with SD as a function of  $\varepsilon_D$  for  $\varepsilon_E = \varepsilon_D + \delta$ , where  $\delta \in \{0, 0.05, 0.1\}$ ,  $K = 20$  and fixed values of  $N_{\max}$ . Bottom: Probability that the destination will recover the source message for  $N_{\max} \in \{25, 27, 29\}$ .

[2] C. Ning and R. W. Yeung, "Secure network coding," in *Proc. IEEE ISIT*, Lausanne, Jun. 2002.

[3] M. Adeli and H. Liu, "On the inherent security of linear network coding," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1668–1671, Aug. 2013.

[4] H. Niu, M. Iwai, K. Sezaki, L. Sun, and Q. Du, "Exploiting fountain codes for secure wireless delivery," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 777–780, May 2014.

[5] A. S. Khan, A. Tassi, and I. Chatzigeorgiou, "Rethinking the intercept probability of random linear network coding," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1762–1765, Oct. 2015.

[6] M. S. Karim, M. Esmailzadeh, and P. Sadeghi, "On reducing intercept probability for unsubscribed video layers using network coding," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1385–1388, Jun. 2017.

[7] A. Tassi, R. J. Piechocki, and A. Nix, "On intercept probability minimization under sparse random linear network coding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6137–6141, Jun. 2019.

[8] A. Tassi, I. Mavromatis, R. J. Piechocki, and A. Nix, "Secure data offloading strategy for connected and autonomous vehicles," in *Proc. IEEE VTC2019-Spring*, Kuala Lumpur, Malaysia, Apr. 2019.

[9] L. Sun, D. Huang, and A. L. Swindlehurst, "Fountain-coding aided secure transmission with delay and content awareness," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7992–7997, Jul. 2020.

[10] M. S. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Reading damaged scripts: Partial packet recovery based on compressive sensing for efficient random linear coded transmission," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3296–3310, Aug. 2016.

[11] 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.

[12] 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.

[13] A. L. Jones, I. Chatzigeorgiou, and A. Tassi, "Binary systematic network coding for progressive packet decoding," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, UK, Jun. 2015.

[14] D. L. Donoho, "Sparse components of images and optimal atomic decompositions," *Constr. Approx.*, vol. 17, no. 1, pp. 353–382, Jan. 2001.

[15] A. S. Khan and I. Chatzigeorgiou, "Performance analysis of random linear network coding in two-source single-relay networks," in *Proc. IEEE ICC workshops*, London, Jun. 2015.

[16] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[17] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*. New York, USA: Springer, 2010.