Rethinking the Intercept Probability of Random Linear Network Coding

Khan, A. S., Tassi, A. & Chatzigeorgiou, I.

Original citation & hyperlink:
Khan, AS, Tassi, A & Chatzigeorgiou, I 2015, 'Rethinking the Intercept Probability of Random Linear Network Coding' IEEE Communications Letters, vol. 19, no. 10, pp. 1762-1765. https://dx.doi.org/10.1109/LCOMM.2015.2470662

DOI 10.1109/LCOMM.2015.2470662
ISSN 1089-7798
ESSN 1558-2558

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Rethinking the Intercept Probability of Random Linear Network Coding

Amjad Saeed Khan, Andrea Tassi and Ioannis Chatzigeorgiou

***Abstract*—This letter considers a network comprising a transmitter, which employs random linear network coding to encode a message, a legitimate receiver, which can recover the message if it gathers a sufficient number of linearly independent coded packets, and an eavesdropper. Closed-form expressions for the probability of the eavesdropper intercepting enough coded packets to recover the message are derived. Transmission with and without feedback is studied. Furthermore, an optimization model that minimizes the intercept probability under delay and reliability constraints is presented. Results validate the proposed analysis and quantify the secrecy gain offered by a feedback link from the legitimate receiver.**

***Index Terms*—Network coding, fountain coding, physical layer security, secrecy outage probability, intercept probability.**

I. Introduction

In the context of networks and protocols, network coding [1] has been widely recognized as an intriguing technique to improve network performance. It can considerably reduce transmission delay, processing complexity and energy consumption, and has the potential to significantly increase throughput and robustness [2]. Therefore, it has been studied for use in many applications, including large scale content distribution in peer-to-peer networks [3] and data transmission in sensor networks or delay tolerant networks [4]. Due to the broadcast nature of wireless channels, networks are vulnerable to security attacks, such as wiretapping and eavesdropping. The problem of achieving secure communication in systems employing network coding has recently attracted the attention of the research community in wireles networks. Ning and Yeung [5] first formulated the concept of secure network coding, which avoids information leakage to a wiretapper. They imposed a security requirement, that is, the mutual information between the source symbols and the symbols received by the wiretapper must be zero for secure communication. Based on a well-designed precoding matrix, Wang *et al.* [6] proposed a secure broadcasting scheme with network coding to obtain perfect secrecy. Probabilistic weak security for linear network coding was presented in [7], which devised network coding rules that can improve security depending on the adopted field size, the number of transmitted symbols and the ability of the attacker to eavesdrop on one or more independent channels.

This work was carried out under the auspices of COST Action IC1104 and the support of EPSRC under Grant EP/L006251/1.

A. S. Khan and I. Chatzigeorgiou are with the School of Computing and Communications, Lancaster University, LA1 4WA, UK (e-mail: {a.khan9, i.chatzigeorgiou}@lancaster.ac.uk).

A. Tassi is with the Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK (email: a.tassi@bristol.ac.uk).

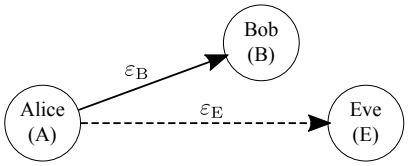

Figure 1. Block diagram of the system model, where ε_{B} and ε_{E} denote the erasure probabilities of the channels linking Alice to Bob and Alice to Eve, respectively.

Recently, the intercept probability of fountain coding, which is equivalent to random linear network coding for wireless broadcast applications, was formulated in [8]. Our work has been inspired by the methodology in [8] but differs in two major points. Firstly, we have revisited the derivation of the intercept probability. More specifically, the decoding probability of a receiver has been taken into account in our calculations. Furthermore, key probability expressions have been revised to accurately reflect (i) the effect of the size of the finite field over which network coding is performed, (ii) the impact of a feedback link between the legitimate receiver and the transmitter, and (iii) the fact that the number of transmitted coded packets cannot be infinite in practice. The second difference is that [8] proposed an optimization model with respect to the number of source packets composing a message. However, the number of source packets and, by extension, their length are often dictated by the provided service. Our objective is to minimize the intercept probability by optimizing the number of transmitted coded packets, under delay and reliability constraints. As part of the optimization process, we prove that awareness of the existence of an eavesdropper is not required by the transmitter and the legitimate receiver.

II. System Model

We consider a network configuration whereby a source (Alice) wishes to transmit a message to a legitimate destination (Bob) in the presence of a passive eavesdropper (Eve), as shown in Fig. 1. Before initiating the communication process, Alice segments the message into K source packets and employs Random Linear Network Coding (RLNC) to generate and broadcast $N \geq K$ coded packets. The links connecting Alice to Bob and Alice to Eve are modeled as packet erasure channels characterized by erasure probabilities ε_{B} and ε_{E}, respectively. As per the RLNC requirements, Bob and Eve can recover the message only if they collect at least K linearly independent coded packets.

Based on this setup and the general condition that $\varepsilon_{\mathrm{B}} < \varepsilon_{\mathrm{E}}$ for physical layer security, we consider two network coded

transmission modes, which we refer to as *Feedback-aided Transmission* (FT) and *Unaided Transmission* (UT). In the FT mode, Alice broadcasts up to N coded packets but ceases transmission as soon as Bob sends a notification over a perfect feedback channel acknowledging receipt of K linearly independent coded packets. In the case of UT, a feedback channel between Bob and Alice is not available, therefore Alice broadcasts exactly N coded packets anticipating Bob to successfully recover her message. In both modes, the communication process is considered to be secure if Eve fails to reconstruct Alice's message. In the rest of this letter, we will investigate the resilience of FT and UT to the interception of K linearly independent coded packets by Eve.

III. Performance Analysis

The physical layer security offered by the two transmission modes will be quantified by the probability that Eve will manage to recover the message. To derive this probability, which is known as the *secrecy outage probability* or the *intercept probability*, we will first consider the general case of point-to-point communication between Alice and a receiver D over an erasure channel with erasure probability ε_D. Note that D can be either Bob or Eve, i.e., $D \in \{\mathrm{B}, \mathrm{E}\}$. If Alice transmits $N \geq K$ coded packets and the receiver retrieves n_{R} coded packets, where $K \leq n_{\mathrm{R}} \leq N$, the probability that the receiver will successfully recover the K source packets is given by [9]

$$P(n_{\mathrm{R}}, K) = \prod_{i=0}^{K-1} \left[1 - q^{-(n_{\mathrm{R}}-i)}\right], \tag{1}$$

where q is the size of the finite field over which network coding operations are performed. Let X be a random variable that represents the number of transmitted coded packets for which the receiver can recover the K source packets. The Cumulative Distribution Function (CDF) of X describes the probability that the receiver will recover the K source packets after n_{T} coded packets have been transmitted, where $K \leq n_{\mathrm{T}} \leq N$. This CDF can be obtained by averaging (1) over all valid values of n_{R}, that is,

$$\begin{aligned} F_D(n_{\mathrm{T}}) &= \Pr\{X \leq n_{\mathrm{T}}\} \\ &= \sum_{n_{\mathrm{R}}=K}^{n_{\mathrm{T}}} \binom{n_{\mathrm{T}}}{n_{\mathrm{R}}} (1-\varepsilon_D)^{n_{\mathrm{R}}} \varepsilon_D^{n_{\mathrm{T}}-n_{\mathrm{R}}} \; P(n_{\mathrm{R}}, K). \end{aligned} \tag{2}$$

The probability that the receiver will recover the K source packets when the n_{T}-th coded packet has been transmitted, but not earlier, is given by the Probability Mass Function (PMF) of X, which can be derived as follows:

$$\begin{aligned} f_D(n_{\mathrm{T}}) &= \Pr\{X = n_{\mathrm{T}}\} \\ &= \begin{cases} F_D(n_{\mathrm{T}}) - F_D(n_{\mathrm{T}}-1), & \text{if } K < n_{\mathrm{T}} \leq N \\ F_D(K), & \text{if } n_{\mathrm{T}} = K. \end{cases} \end{aligned} \tag{3}$$

Let us now return our focus to the considered network configuration operating in the FT mode. Recall that Bob sends an acknowledgement to Alice when he receives K linearly independent coded packets and can thus recover the source message. The intercept probability can be expressed as the sum of two constituent probabilities:

$$P_{\text{int}}^{\text{FT}}(N) = P_{\mathrm{BE}}(N) + P_{\mathrm{E}}(N). \tag{4}$$

The first term of the sum in (4), $P_{\mathrm{BE}}(N)$, denotes the probability that both Bob and Eve will recover the message. This can happen if Bob decodes the message only after the n_{T}-th coded packet has been transmitted, while Eve has already recovered the message or recovers it concurrently with Bob. Invoking the definitions in (2) and (3), and considering all possible values of n_{T}, we can express $P_{\mathrm{BE}}(N)$ as

$$P_{\mathrm{BE}}(N) = \sum_{n_{\mathrm{T}}=K}^{N} f_{\mathrm{B}}(n_{\mathrm{T}})\, F_{\mathrm{E}}(n_{\mathrm{T}}). \tag{5}$$

The second term of the sum in (4), $P_{\mathrm{E}}(N)$, represents the probability that Eve will be successful in recovering the message but Bob will fail to decode it after Alice has transmitted the complete sequence of N coded packets. Using the CDF of the number of coded packets delivered by Alice to Eve and Bob, respectively, we can write $P_{\mathrm{E}}(N)$ as follows:

$$P_{\mathrm{E}}(N) = F_{\mathrm{E}}(N)\, \left[1 - F_{\mathrm{B}}(N)\right]. \tag{6}$$

We should stress that (5) and (6) are exact only if the sequence of coded packets delivered over the Alice-to-Bob link is independent of the sequence delivered over the Alice-to-Eve link. This is a common hypothesis in the literature of broadcast networks, e.g., [8] and [10], and is valid for a non-vanishing product between the number of coded packets transmitted over a channel and the erasure probability of that channel [11]. The accuracy of (4) will also be demonstrated in Section V.

In the case of UT, a feedback channel is not available between Bob and Alice, therefore Alice transmits the complete sequence of N coded packets uninterruptedly. Therefore, the intercept probability is simply equal to the probability that Eve will recover the message after Alice has transmitted N coded packets. Using the definition of the CDF in (2), we obtain

$$P_{\text{int}}^{\text{UT}}(N) = F_{\mathrm{E}}(N). \tag{7}$$

Manipulation of the expression for $P_{\text{int}}^{\text{FT}}(N)$, as shown in Appendix A, and subtraction of $P_{\text{int}}^{\text{UT}}(N)$ from it, yields

$$P_{\text{int}}^{\text{FT}}(N) - P_{\text{int}}^{\text{UT}}(N) = -\sum_{n_{\mathrm{T}}=K+1}^{N} f_{\mathrm{E}}(n_{\mathrm{T}})\, F_{\mathrm{B}}(n_{\mathrm{T}}-1). \tag{8}$$

Expression (8) measures the loss in the intercept capability of Eve or, equivalently, the gain in secrecy by Bob, if Bob can acknowledge the recovery of the source message to Alice using a feedback channel.

Remark. In this letter, we assume that Alice has knowledge of the *average* channel conditions, characterized by the erasure probability, between her and Bob. If Alice could sense the *instantaneous* channel quality and transmitted coded packets only when the channel quality warranted their error-free delivery to Bob, as in [8], [12], the equivalent erasure probability of the link between Alice and Bob would be $\varepsilon_{\mathrm{B}} = 0$. In that case, Alice could generate exactly K linearly independent coded packets in a deterministic manner, as opposed to random, and

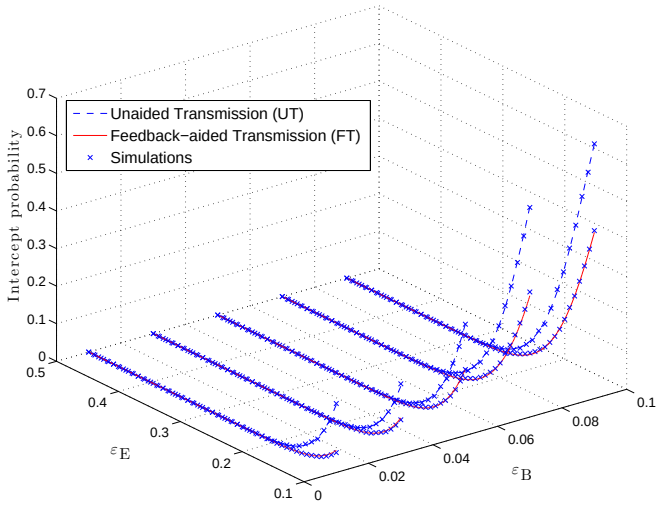

Figure 2. Comparison between analytical and simulation results for FT and UT, when $\varepsilon_E \in [0.1, 0.5]$, $\varepsilon_B = \{0.01, 0.03, 0.05, 0.07, 0.09\}$, $K = 50$, $\hat{N} = 150$, $q = 2$ and $\hat{P} = 90$.

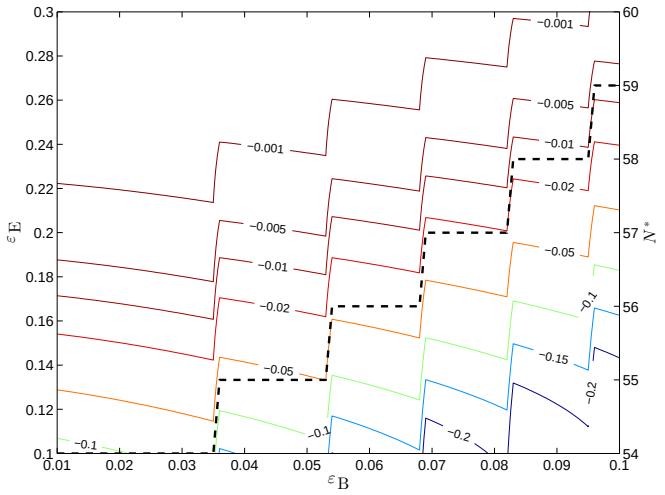

Figure 3. Contour map (solid lines) depicting the loss in intercept probability caused by the change from UT to FT, as a function of ε_E and ε_B. The value of N^* (dashed line) as a function of ε_B has been superimposed on the plot.

forward them to Bob. As a result, the intercept probability would reduce to $(1-\varepsilon_E)^K$ regardless the transmission mode. This remark concurs with the conclusion of [8] that an arbitrarily small intercept probability can be achieved by increasing the value of K, but at the cost of increased delay.

IV. Optimization Model

This section aims to determine the optimum value of N, i.e., the number of coded packet transmissions, that minimizes the intercept probability, provided that a hard deadline is met. This hard deadline, denoted by $\hat{N}$, represents the number of coded packet transmissions that Alice is not allowed to exceed. In addition, the proposed optimization strategy permits Bob to recover the message with a target probability $\hat{P}$. In the rest of this letter, both FT and UT will be optimized by the Resource Allocation Model (RAM), which is defined as follows:

$$\text{(RAM)} \quad \min_{N} \; P_{\text{int}}(N) \tag{9}$$

$$\text{subject to} \quad F_B(N) \geq \hat{P} \tag{10}$$

$$N \leq \hat{N} \tag{11}$$

where the objective function (9) represents the intercept probability when N coded packets have been scheduled for transmission. Constraint (10) ensures that the probability of Bob recovering the message is at least $\hat{P}$, while constraint (11) imposes that the number of planned coded packet transmissions is less than or equal to $\hat{N}$.

The proof of the following proposition will contribute to the solution of the RAM problem.

Proposition 1. *The intercept probability $P_{int}(N)$ is a non-decreasing function of N, i.e.,*

$$P_{int}(N_1) \leq P_{int}(N_2) \quad \textit{for all} \quad N_1 \leq N_2. \tag{12}$$

Proof: One of the properties of CDFs is that they are non-decreasing functions and, as per (7), the intercept probability of UT is equal to a CDF. In the case of FT, the subtraction of $P_{\text{int}}(N_1)$ from $P_{\text{int}}(N_2)$ for $N_2 \geq N_1$ gives a sum of non-negative terms, as shown in Appendix B. Therefore, $P_{\text{int}}(N_2) - P_{\text{int}}(N_1) \geq 0$, which concludes the proof. ■

We can now proceed to Proposition 2 and provide a description of the solution to the RAM problem.

Proposition 2. *If the RAM problem admits a solution, the optimum solution is*

$$N^* = \arg\min \left\{ N \in [K, \hat{N}] \;\middle|\; F_B(N) \geq \hat{P} \right\}. \tag{13}$$

Proof: Let N^* denote the smallest value of N in the interval $[K, \hat{N}]$ for which constraint (10) holds. If an integer value smaller than N^* is selected, for example $N^* - 1$, the intercept probability will reduce, as per Proposition 1, but constraint (10) will not be met. We thus conclude that N^* is the optimum solution to the RAM problem. ■

Root-finding algorithms, such as the bisection method, can be used on the right-hand side of (13) to determine if N^* exists and identify its value. Based on this analysis, we showed that minimization of the intercept probability under delay and reliability constraints can be achieved by minimizing the number of transmitted coded packets. Thus, Alice should know the erasure probability of the channel between her and Bob but knowledge of the presence of an eavesdropper *is not necessary*.

V. Numerical and Analytical Results

This section compares the derived analytical expressions with simulation results, establishes their validity and obtains solutions to the RAM problem for various channel conditions.

Fig. 2 depicts the relationship between the intercept probability and the quality of Bob's and Eve's channels, represented by ε_B and ε_E, respectively. For each point, the value of the N coded packet transmissions was optimized by RAM for $K = 50$ source packets, $\hat{N} = 150$ maximum allowable coded packet transmissions, a field size of $q = 2$ and a target probability of Bob recovering the source message equal to $\hat{P} = 90\%$. In simulations, Alice broadcasts the optimal

number of coded packets determined by RAM. Instances where Eve successfully recovers K linearly independent coded packets are counted and averaged over 10^4 realizations to obtain the intercept probability. We observe the close agreement between analytical and simulation results, which confirms the tightness of (4) and (7). Fig. 2 also shows that when the channel quality between Alice and Eve is significantly worse than the channel quality between Alice and Bob, the intercept probability is close to zero for both FT and UT. As expected, the intercept probability increases when the two channels experience identical or relatively similar conditions but FT offers a clear advantage over UT. For example, for $\varepsilon_{\mathrm{B}} = 0.09$ and $\varepsilon_{\mathrm{E}} = 0.1$, the intercept probability will reduce from 68% to 45% if the mode of operation switches from UT to FT. The reduction in the intercept probability due to the adoption of FT becomes pronounced when ε_{E} drops below 0.25.

Fig. 3 quantifies the loss in intercept probability or, equivalently, the gain in secrecy that occurs by changing the operational mode from UT to FT, as noted in (8). The optimum value of N, denoted by N^*, has also been plotted in Fig. 3 (secondary y-axis on the right-hand side of the plot). Observe that as ε_{B} increases from 0.01 to 0.1, Alice increases the coded packet transmissions from 54 to 59 in an effort to maintain the probability of Bob recovering the source message at $\hat{P} = 90\%$. Notice the abrupt change in the intercept probability each time RAM generates a new optimum value for N, based on ε_{B}.

A way to reduce the intercept probability, especially in settings where the values of ε_{B} and ε_{E} are similar, has been hinted in the Remark. If Alice can measure the instantaneous quality of the channel between her and Bob and transmits coded packets only when the measured quality is above an acceptable threshold, the effective value of ε_{B} will be reduced and the intercept probability will drop at the expense of delay.

VI. Conclusion

We derived accurate expressions for the intercept probability of a network, where a transmitter uses random linear network coding to broadcast information. Both unaided transmission and feedback-aided transmission were investigated and the secrecy gain achieved by the latter approach was computed. We presented a resource allocation model to minimize the intercept probability, while satisfying delay and reliability constraints, and showed that the legitimate receiver is not required to have knowledge of the presence of an eavesdropper. Theoretical and simulation results identified the channel erasure probabilities for which feedback-aided transmission offers a lower intercept probability than unaided transmission when the proposed resource allocation model is applied.

Appendix

A. Reformulation of the intercept probability of FT

Based on the definition of the PMF in (3), the expression for $P_{\mathrm{BE}}(N)$ in (5) can be expanded as follows:

$$\begin{aligned} P_{\mathrm{BE}}(N) = {} & F_{\mathrm{B}}(K)F_{\mathrm{E}}(K) \\ & - F_{\mathrm{B}}(K)F_{\mathrm{E}}(K+1) + F_{\mathrm{B}}(K+1)F_{\mathrm{E}}(K+1) \\ & - \dots \\ & - F_{\mathrm{B}}(N-1)F_{\mathrm{E}}(N) + F_{\mathrm{B}}(N)F_{\mathrm{E}}(N). \end{aligned}$$

If we create pairs from each two consecutive terms, with the exception of the last term, and invoke again the definition of the PMF, we obtain

$$P_{\mathrm{BE}}(N) = \left[-\sum_{n_{\mathrm{T}}=K+1}^{N} f_{\mathrm{E}}(n_{\mathrm{T}})\, F_{\mathrm{B}}(n_{\mathrm{T}}-1) \right] + F_{\mathrm{B}}(N)F_{\mathrm{E}}(N).$$

In (6), we established that $P_{\mathrm{E}}(N) = F_{\mathrm{E}}(N) - F_{\mathrm{B}}(N)F_{\mathrm{E}}(N)$. Using (4), the intercept probability of FT can be expressed as:

$$P_{\mathrm{int}}^{\mathrm{FT}}(N) = F_{\mathrm{E}}(N) - \sum_{n_{\mathrm{T}}=K+1}^{N} f_{\mathrm{E}}(n_{\mathrm{T}})\, F_{\mathrm{B}}(n_{\mathrm{T}}-1). \tag{14}$$

B. Proof of Proposition 1 for the case of FT

In order to prove Proposition 1 for the FT mode, it suffices to set $\Delta = P_{\mathrm{int}}(N_2) - P_{\mathrm{int}}(N_1)$ and show that $\Delta \geq 0$ for all $N_2 \geq N_1$. Using (14), we find that

$$\Delta = F_{\mathrm{E}}(N_2) - F_{\mathrm{E}}(N_1) - \sum_{n_{\mathrm{T}}=N_1+1}^{N_2} f_{\mathrm{E}}(n_{\mathrm{T}})\, F_{\mathrm{B}}(n_{\mathrm{T}}-1). \tag{15}$$

Terms $-F_{\mathrm{E}}(i)$ and $F_{\mathrm{E}}(i)$ for $i = N_1+1, \dots, N_2-1$, which cancel each other out, are added to $F_{\mathrm{E}}(N_2) - F_{\mathrm{E}}(N_1)$ and give

$$\begin{aligned} F_{\mathrm{E}}(N_2) - F_{\mathrm{E}}(N_1) = {} & \left(F_{\mathrm{E}}(N_2) - F_{\mathrm{E}}(N_2-1)\right) + \dots \\ & \dots + \left(F_{\mathrm{E}}(N_1+1) - F_{\mathrm{E}}(N_1)\right) \\ = {} & \sum_{n_{\mathrm{T}}=N_1+1}^{N_2} f_{\mathrm{E}}(n_{\mathrm{T}}). \end{aligned} \tag{16}$$

If we substitute (16) into (15), we obtain

$$\Delta = \sum_{n_{\mathrm{T}}=N_1+1}^{N_2} f_{\mathrm{E}}(n_{\mathrm{T}}) \left[1 - F_{\mathrm{B}}(n_{\mathrm{T}}-1)\right]$$

which is a sum of non-negative terms and is, thus, $\Delta \geq 0$.

References

[1] R. Ahlswede, N. Cai, S. R. Li, and R. W. Weung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] 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.

[3] C. Gkantsidis and P. R. Rodriguez, "Network coding for large scale content distribution," in *Proc. IEEE INFOCOM*, Miami, Mar. 2005.

[4] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proc. ACM SIGCOMM*, Kyoto, Aug. 2007.

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

[6] X. Wang, W. Guo, Y. Yang, and B. Wang, "A secure broadcasting scheme with network coding," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1435–1438, Jul. 2013.

[7] 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.

[8] 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.

[9] 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.

[10] E. Kurniawan, S. Sun, K. Yen, and K. F. E. Chong, "Network coded transmission of fountain codes over cooperative relay networks," in *Proc. IEEE WCNC*, Sydney, Apr. 2010.

[11] 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.

[12] Z. Guan, T. Melodia, and G. Scutari, "To transmit or not to transmit? Distributed queueing games in infrastructureless wireless networks," *IEEE/ACM Trans. Netw.*, to appear in 2015.