Exact Decoding Probability Under Random Linear Network Coding

Oscar Trullols-Cruces, *Student Member, IEEE,* Jose M. Barcelo-Ordinas, *Member, IEEE,* and Marco Fiore, *Member, IEEE*

***Abstract*—In this letter, we compute the exact probability that a receiver obtains N linearly independent packets among $K \geq N$ received packets, when the sender/s use/s random linear network coding over a Galois Field of size q. Such condition maps to the receiver's capability to decode the original information, and its mathematical characterization helps to design the coding so to guarantee the correctness of the transmission. Our formulation represents an improvement over the current upper bound for the decoding probability, and provides theoretical grounding to simulative results in the literature.**

***Index Terms*—Random linear network coding, wireless networks, error control.**

I. Introduction

NETWORK coding allows efficient transmission from a set of sources to a set of destinations, allowing nodes to manipulate the information before forwarding it [1]. Random linear network coding is a class of network coding, that operates on data through linear combinations of random codes [2]. Random linear network coding has been shown to improve the latency, capacity and energy efficiency of the communication in loss-prone and intermittently-connected wireless networks, either ad-hoc [3], delay-tolerant [4], or satellite and underwater [5].

Within these contexts, when one or more sources want to transmit N packets to one or more mobile nodes, the channel unreliability and the fluctuation of connectivity force the adoption of reliable communication techniques. However, traditional retransmission-based mechanisms easily lead to excessive overhead, even in presence of coordination between the sources. If random linear network coding is employed, the reception of a (limited) amount S of excess packets can eliminate the need for a feedback channel. Indeed, to be able to decode the original data, a destination node simply has to acquire N linearly independent packets over the K=N+S it received from the source(s) and intermediate relay(s).

The effectiveness of random linear network coding thus depends on the the probability that at least N packets reach their destination, and that they represent linearly independent encodings of the original data. While the former condition relates to the error probability on the channel or to the network topology, on which one does not typically have control, the latter concerns the coding design, that is instead configurable. A common approach is to assume the size q of the Galois Field over which the coding is performed to be very large, so that any received packet is independent from those previously obtained, with high probability. As an example, using large encoding coefficients of 20 bits, that correspond to values of q in the order of 2^{20}, allows to exclusively dimension S on the packet loss probability. However, using large field sizes has a drawback in terms of computational complexity: it would thus be desirable to determine the exact impact of the field size on the decoding probability, so to properly dimension q. The only such result to date is the upper bound in [6], stating that the average number of coded packets K to be received before the original data can be decoded is, for a field size q, equal to

$$K = min\{N\frac{q}{q-1}, N+1+\frac{1-q^{-N+1}}{q-1}\}. \quad (1)$$

According to (1), when q=2, S=2 excess packets are sufficient, on average, for N+S received packets to be linearly independent, no matter the value of N.

In this letter, we improve the upper bound in (1), by deriving the exact formulation of the probability that N out of N+S received packets are linearly independent, under random linear network coding, as a function of the field size q. Our formulation evidences that a value of q equal to four allows a correct decoding with just one excess packet on average.

II. Exact Decoding Probability

Let $GF(q)$ be a Galois Field of size q and let us assume that a set of uncoordinated sources transmit K=N+S packets. Each k-th packet is constructed using random linear network coding; i.e., each new packet is associated with a random encoding vector g_k over $GF(q)$ of dimension N, and it is the result of the linear combination of the N original packets, [2]. Let us call G_q the matrix containing the encoding vectors g_k. The N original packets can be decoded if G_q has rank N, i.e., the receiver node has obtained N linear independent packets over the K sent packets. We denote as P_{ns} the probability that matrix G_q has rank N.

In order to derive the exact expression of P_{ns} we employ an urn model. Consider an urn with all the vectors that can be generated by codes in a Galois Field of size q, $GF(q)$, over N packets. There are q^N possible vectors. For the sake of clarity, we will first analyze the simple case in which K=N, and then derive the case in which $K > N$.

Case K=N: consider the extraction of linearly independent vectors from the urn. In the first extraction any vector that is not the zero vector[1] will be a suitable vector. The probability of extracting a vector which is not the zero vector is equal to

Manuscript received August 12, 2010. The associate editor coordinating the review of this letter and approving it for publication was I. Maric.

O. Trullols-Cruces and J. M. Barcelo-Ordinas are with the Department of Computer Architecture, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain (e-mail: {trullols, joseb}@ac.upc.edu).

M. Fiore is with the Université de Lyon, INRIA, INSA-Lyon, CITI Lab, F-69621, France (e-mail: marco.fiore@insa-lyon.fr).

Digital Object Identifier 10.1109/LCOMM.2010.110310.101480

[1]Discarding the zero vector is later treated.

1089-7798/11$25.00 © 2011 IEEE

$(1-\frac{q^0}{q^N})=(1-\frac{1}{q^N})$. After each extraction, we reinsert the extracted vector in the urn since in the model we are randomly extracting vectors over a Galois Field $GF(q)$. In the second extraction, there are q vectors that are dependent among them, so the probability of having two linearly independent vectors is $(1-\frac{1}{q^N})(1-\frac{q}{q^N})$. In the third extraction there are q^2 vectors that are linearly dependent with the previously extracted ones, and so on. As we perform exactly N extractions, and we need N independent vectors, we must not fail in any of the extractions. Thus, the probability of having N linear independent vectors over K=N extractions is given by:

$$P_{ns}(K,N)=\prod_{j=0}^{N-1}(1-\frac{q^j}{q^N})=\prod_{j=1}^{N}(1-\frac{1}{q^j}). \quad (2)$$

CASE $K>N$: Let us first assume K=N+1 extractions from the urn. Again, during the first extraction, any vector that is not the zero vector is acceptable. However, in the second extraction, there is room for exactly one failure. If such failure occurs, with probability $\frac{q}{q^N}$, we will be left with exactly N-1 extractions to obtain N-1 independent vectors. If the newly extracted vector is instead independent with respect to the first one, with probability $(1-\frac{q}{q^N})$, we will still have N-1 extractions to get N-2 independent vectors. In the third extraction, we can fail with probability $\frac{q^2}{q^N}$ and must not fail with probability $(1-\frac{q^2}{q^N})$ in the fourth and so on. Note that if a failure occurs in the k-th extraction, there must not be any failure in the future extractions, and thus these events are exclusive. Iterating, we obtain the probability that N linear independent vectors are extracted, given K=N+1 extractions, as:

$$P_{ns}(K,N)=\prod_{j=1}^{N}(1-\frac{1}{q^j})\sum_{i=0}^{N}\frac{q^i}{q^N}. \quad (3)$$

Following the same reasoning for a generic $K>N$, we obtain the following formula:

$$P_{ns}(K,N)= \frac{1}{q^{N(K-N)}}\prod_{j=1}^{N}(1-\frac{1}{q^j}) \sum_{r_1=0}^{N}q^{r_1}\sum_{r_2=r_1}^{N}q^{r_2}\cdots\sum_{r_{K-N}=r_{(K-N-1)}}^{N}q^{r_{K-N}}. \quad (4)$$

Note that there are K-N summatories in the formula. However, this formula can be reduced considering the q-binomial coefficients, also called Gauss Coefficients. The q-binomial of two non-negative integers m and n is defined as:

$$\begin{bmatrix}m\\n\end{bmatrix}_q=\frac{(q^m-1)(q^{m-1}-1)...(q^{m-n+1}-1)}{(q^n-1)(q^{n-1}-1)...(q-1)}. \quad (5)$$

Note that, if n = 0, the q-binomial has value 1 by definition, while, if q = 1, the Gauss Coefficient becomes the well known binomial coefficients. Although Gauss Coefficients appear as rational functions, they are in fact polynomial, since the denominator is always a factor of the numerator. It is not surprising that Gauss Coefficients appear in eq. (4), since, among others, a Gauss Coefficient counts the number $V\{m,n;q\}$ of different n-dimensional vector subspaces of an m-dimensional vector space over $GF(q)$.

Using the Gauss Coefficient properties $\sum_{i=0}^{m-1}q^i=\begin{bmatrix}m\\1\end{bmatrix}_q$ and $\begin{bmatrix}m\\n\end{bmatrix}_q=q^n\begin{bmatrix}m-1\\n\end{bmatrix}_q+\begin{bmatrix}m-1\\n-1\end{bmatrix}_q$, we may show for K-N=2 that $\begin{bmatrix}K\\K-N\end{bmatrix}_q=\begin{bmatrix}N+2\\2\end{bmatrix}_q=\sum_{r=0}^{N}q^r\sum_{s=r}^{N}q^s$. Now, it can be easily shown using these recursions that the embedded summatories are equal to $\begin{bmatrix}K\\K-N\end{bmatrix}_q$, and grouping terms:

$$P_{ns}(K,N)=\begin{cases}0 & \text{if } K<N\\ \prod_{j=1}^{N}(1-\frac{1}{q^j})\frac{\begin{bmatrix}K\\K-N\end{bmatrix}_q}{q^{N(K-N)}} & \text{if } K\geq N,\end{cases} \quad (6)$$

and, by applying eq. (5), we can reduce eq. (6) to:

$$P_{ns}(K,N)=\begin{cases}0 & \text{if } K<N\\ \prod_{j=0}^{N-1}(1-\frac{1}{q^{K-j}}) & \text{if } K\geq N.\end{cases} \quad (7)$$

In eq. (7), $P_{ns}(K,N)$ represents the cumulative distribution function of the probability of receiving N linearly independent packets, given the transmission of $K\geq N$ packets under random linear network coding. The probability density function can then be computed as $p_{ns}(K,N)=P_{ns}(K,N)-P_{ns}(K-1,N)$, and the average number of packets to be sent in order to decode the N original ones is:

$$E[K]=\sum_{k=N}^{\infty}k\cdot p_{ns}(k,N). \quad (8)$$

As a further point, in a real implementation, the zero vector would be explicitly excluded from the extraction urn. When accounting for this aspect in the model, eq. (2) becomes:

$$P_{ns}(K,N)=\prod_{j=0}^{N-1}(1-\frac{q^j-1}{q^N-1})=(\frac{q^N}{q^N-1})^N\prod_{j=1}^{N}(1-\frac{1}{q^j}), \quad (9)$$

whereas eq. (4) results in:

$$P_{ns}(K,N)=\frac{q^{N^2}}{(q^N-1)^K}\prod_{j=1}^{N}(1-\frac{1}{q^j}) \sum_{r_1=0}^{N}(q^{r_1}-1)\sum_{r_2=r_1}^{N}(q^{r_2}-1)\cdots\sum_{r_{K-N}=r_{(K-N-1)}}^{N}(q^{r_{K-N}}-1). \quad (10)$$

Expressing eq. (10) in terms of q-binomial coefficients, with $P_{ns}^0=\frac{q^{N^2}}{(q^N-1)^K}\prod_{j=1}^{N}(1-\frac{1}{q^j})$, results in:

$$P_{ns}(K,N)= \begin{cases}0 & \text{if } K<N\\ P_{ns}^0\cdot\left(\begin{bmatrix}K\\K-N\end{bmatrix}_q+\sum_{n=1}^{K-N}(-1)^n\binom{K}{n}\begin{bmatrix}K-n\\K-N-n\end{bmatrix}_q\right) & \text{if } K\geq N.\end{cases} \quad (11)$$

As a final remark, we stress that the formulation above only accounts for the decoding probability due to the actual random coding, in terms of excess packets and field size. Packet losses and network topology also impact on the number of packets required for a correct transmission, but they are independent of how random vectors are chosen. For example, for a packet

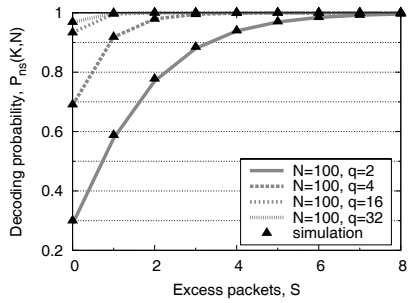

Fig. 1. P_{ns} versus S, for N=100.

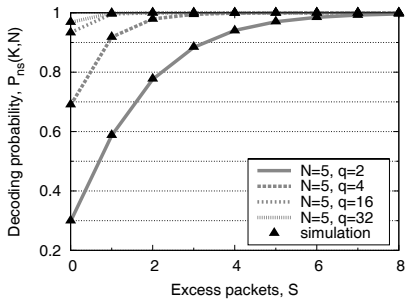

Fig. 2. P_{ns} versus S, for N=5.

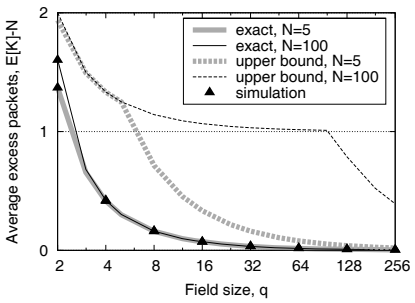

Fig. 3. Mean excess packets to decode versus q.

delivery probability p at the receiver, the probability of decoding correctly N packets over K coded packets will be:

$$p_{dec} = \sum_{j=N}^{K} \binom{K}{j} p^j (1-p)^{K-j} \cdot p_{ns}(j,N) \qquad (12)$$

III. Numerical Results

As discussed above, our analysis concerns the decoding probability as a function of q and S, and is thus limited to the linear independence of the random vectors employed for the encoding of received packets. Figure 1 shows $P_{ns}(K,N)$ as a function of the number of excess packets, S, for N=100 and $q \in \{2,4,16,32\}$. Our analysis, represented by continuous and dashed lines, perfectly matches simulation results, portrayed as points, that are obtained via an actual extraction of random vectors, and are averaged over 10^5 runs: differences between analytical and simulative values are in the order of 0.1%.

We can also observe that larger values of the field size q allow to reach higher decoding probabilities with a same number of excess packets, or, conversely, less excess packets are required to reach a high decoding probability. However, increasing the field size only pays out for low values of q, since considering very large field sizes induces greater computational complexity, but no real advantage in terms of decoding probability.

On the other hand, the number of packets N has a negligible impact on the results, as depicted in Figure 2. There, the number of packets is N=5, but the results are virtually identical to those obtained for the case of N=100. Again, the analysis perfectly matches simulation.

Figure 3 shows the average number of excess packets $E[K]-N$ required to decode the N original packets. The plot portrays the outcome of our analytical formulation in eq. (11), the upper bound in [6] described in eq. (1), as well as the results obtained from simulation. Once more, our analysis provides a perfect matching with the simulation results.

Moreover, the exact formulation shows that the average number of excess packets required for the decoding is noticeably lower than that indicated by the upper bound. Indeed, the exact solution demonstrates that, if a field size of q equal to 3 or 4 is considered, just one extra packet is sufficient for having N linearly independent packets. Such result also holds when considering large blocks of coded packets (i.e., high values of N), a situation that the upper bound cannot reproduce and that was only discussed via simulation in [6].

As a final remark, we note that numerical results on the decoding probability in presence of the zero vector, as in eq. (7), returned values similar to those shown for eq. (11), unless very small values of q and N are selected.

IV. Conclusions

We have computed the exact probability that a receiver obtains N linear independent packets over $K \geq N$ received packets under random linear network coding over a Galois Field of size q. The derivation makes use of an urn model, and employs Gauss Coefficients to achieve a simple formulation.

Acknowledgment

This work was partially supported by the EuroNF NoE and by Spanish grants TIN2010-21378-C02-01 and 2009-SGR-1167

References

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

[2] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413-4430, 2006.

[3] T. Ho, B. Leong, M. Medard, R. Koetter, Y. Chang, and M. Effros, "On the utility of network coding in dynamic environments," *IWWAN'04*, Oulu, Finland, 2004.

[4] J. Widmer and J.-Y. Le Boudec, "Network coding for efficient communication in extreme networks," *ACM WDTN'05*, Philadelphia, USA, 2005.

[5] D. Lucani, M. Medard, and M. Stojanovic, "Random linear network coding for time division duplexing: when to stop talking and start listening," *IEEE INFOCOM'09*, Rio de Janeiro, Brazil, 2009.

[6] D. Lucani, M. Medard, and M. Stojanovic, "Random linear network coding for time division duplexing: field size considerations," *IEEE GLOBECOM'09*, Hawaii, USA, 2009.