 doi:10.20944/preprints201802.0009.v1

Article

On the Performance of the Cache Coding Protocol

Behnaz Maboudi [1], Hadi Sehat [1], Peyman Pahlevani [1]* and Daniel E. Lucani[2]

[1] Institute of Advanced Studies in Basic Sciences (IASBS); {behnazmaboudi, h.sehat, pahlevani}@iasbs.ac.ir
[2] Aarhus University; daniel.lucani@eng.au.dk
* Correspondence: pahlevani@iasbs.ac.ir Tel.: +98-24-3315-3436

Abstract: Network coding approaches typically consider an unrestricted recoding of coded packets in the relay nodes for increased performance. However, this can expose the system to pollution attacks that cannot be detected during transmission, until the receivers attempt to recover the data. To prevent these attacks while allowing for the benefits of coding in mesh networks, the Cache Coding was proposed. This protocol only allows recoding at the relays when the relay has received enough packets to decode an entire generation of packets. At that point, the relay node recodes and signs the recoded packets with its own private key allowing for the system to detect and minimize the effect of pollution attacks and make relays accountable for changes on the data. This paper analyzes the delay performance of Cache Coding to understand the security-performance trade-off of this scheme. We introduce an analytical model for the case of two relays in an erasure channel relying on an Absorbing Markov Chain and a approximate model to estimate the performance in terms of the number of transmissions before successfully decoding at the receiver. We confirm our analysis using simulation results. We show that Cache Coding can overcome security issues of unrestricted recoding with only a moderate decrease in system performance.

Keywords: Cache Coding, Source Coding, Absorbing Markov Chain

1. Introduction

Mobile Ad-hoc Networks (MANET) constitute an infrastructure-less network architecture for mobile devices. In MANETs, network coding has been shown to increase reliability and throughput [19]. In network coding, data is encoded in a sender and transmitted through a network to be decoded at one or multiple receivers. Intermediate nodes in the network can also recombine coded packets (recode) without decoding the original data. Network coding can be considered as a generalized routing, in which relay nodes not only store and forward packets but also combine packets received from multiple input paths before sending to any output path [1][2]. Random Linear Network Coding (RLNC) is a simple, yet asymptotically optimal method, where packets are linearly combined using random coefficients drawn uniformly at random from the elements of a finite field [1] and then transmitted [3].

However, RLNC with unrestricted and unsupervised recoding is vulnerable to pollution attacks. A pollution attack occurs when a malicious or faulty node injects invalid linear combinations of generation into the network [16]. These invalid coded packets can quickly propagate into other packets via recoding in relay nodes and can prevent destination node from decoding properly [11]. To prevent them, homomorphic signatures, which are preserved through linear combination and recoding, can be used [4] [5] [7]. These method allow the system to check integrity of network coded data and to track and find malicious nodes in the network However, due to high processing cost, this method is not practical [11]. While there exists alternatives for preventing pollution attacks with reduced processing costs, these solutions place limitations on topologies, require loose clock synchronization on the order of 100 ms, limit the hop count, require large field sizes or demand new public keys generated per generation. These requirements are not feasible in dynamic MANET networks. Thus, new integrity mechanism to mitigate pollution attacks [10].

As an alternative, the authors of [16] proposed a protocol solution called Cache Coding. Cache Coding allows certified relay nodes to recode data only when they have decoded an entire generation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

At that time, the relay nodes can generate new recoded packets and sign them with their own identifier. This generates trusted packets from both the data source and the trusted relay nodes, thus allowing the network to detect pollution attacks before they propagate.

Although [16] presents the protocol to cope with this vulnerability and identifies some bounds on the number of transmissions, a precise analytical model to characterize the performance of Cache Coding in multi-relay networks is missing. This performance analysis is critical to understand the security-delay trade-off imposed by the protocol. This paper provides two models for analysis of the Cache protocol The first model is a simple, yet accurate approximation to predict the number of required transmissions. The second model is based on an Absorbing Markov Chain inspired by previous work, e.g., [12], and is used to accurately characterize the total number of transmissions in the system as well as the number of linearly independent packets in all nodes after some constant transmissions from the source node. Although we focus on a simple relay policy from the Cache Coding protocol, future work can exploit the Markov Chain structure of the problem to derive optimal policies using Markov Decision Processes (MDP) as in recent RLNC work, e.g., [15], [17], [18].

We validate our analytical models using simulation results showing the difference between the Absorbing Markov Chain model and the simulation results is negligible for all analyzed generation sizes and packet loss rates. For moderate loss rates, at most 10 % deviation is expected between our heuristic model and the simulation results. However, the heuristic model losses accuracy for high loss rates. Finally, we show that the Cache Coding protocol is at most 25 % worse than RLNC with unrestricted recoding in terms of the number of transmissions. However, in average there is only 7 % deviation between Cache Coding and Unrestricted Coding.

This paper is organized as follows. In Section 2, we discuss system model and different coding methods in relay nodes. In Sections 3 and 4 we discuss our two analytical models, while Section 5 focuses on validating those models using simulation results. We present our conclusions in Section 6.

2. System Model

We consider a topology with one source node S, two relay nodes R_1 and R_2, and a destination node D. A total number of n packets must be sent from S to D. This transmission occurs by using the links shown in Figure 1. Each link has a fixed packet loss rate, e, which are considered constant throughout the transmission and independent from each other.

We define a *Degree of Freedom* (*DoF*) of a node as the number of linearly independent packets received by that node. In any topology, the transmission will be carried on until *DoF* of the destination node is equal to generation size, i.e., when the destination has enough information to decode the original data. We also define the *generation size* (n packets) as the number of original data packets that will be linearly combined to generate network coded packets.

S generates packets using RLNC. It uses data packets p_1, p_2,..., p_n in the generation to create a linear combination with coding coefficients a_1, a_2,..., a_n, i.e., $\sum_{i=1}^{n} a_i p_i$. These coefficients are randomly selected from q elements of a Galois Field, i.e., $GF(q)$ [15].

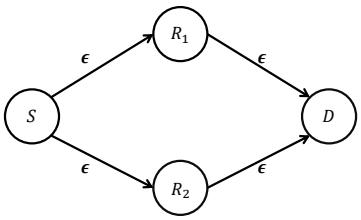

Figure 1. The topology of a two-relay network

Data is sent in three broadcast transmissions in three time slots as in Figure 2 with Tx and Rx indicating the transmission and reception of data, respectively. The first time slot is used for broadcasting from S to both R_1 and R_2. The second and third time slots, are used by R_1 and R_2 to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

broadcast to D, respectively. We assume a shared channel between all nodes for sending data in this topology and the use of a TDMA protocol is used to manage each node's access to this shared channel.

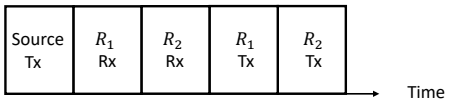

Figure 2. A transmission round

We assume that whenever a relay node receives a packet, it is permitted to transmit a packet. This packet is determined by one of the following three coding methods.

Source Coding ($C-C$)
Upon reception of a transmission from S, R_1 and/or R_2 send a packet to D which is identical to the packet received from S. Thus, the D receives only one independent packet if it receives packets from more than one relay in the same time slot. Repeated coded packets are discarded.

Unrestricted Coding($U-C$)
In this method, R_1 and R_2 receive encoded packets from S. R_1 and R_2 recode packets with previously received coded packets in each relay's buffer and send the newly generated recoded packet to D. Each relay transmits a recoded packet when they receive a coded packet from S.

Cache Coding ($C-C$)
In this method, R_1 and R_2 will act like the Source Coding method until its DoF is equal to the generation size (n). Then, the relay will decode all received packets. After decoding all packets, the mentioned relay will recode the original packet and send recoded data to D. This mechanism continues until the end of transmission, where DoF of D is equal to generation size (n).

3. Heuristic

This section derives an approximation to evaluate the performance of the three coding methods described in Section 2 in a system with the topology of Figure 1. We address each method in different subsections. We focus on the number of transmissions from the source until D decodes a generation of n packets as well as the total number of transmissions in all links in the network.

3.1. Source Coding

The probability that a packet is received by D from either R_1 or R_2 or both R_1 and R_2 is

$$P_r = 2(1-e)^2. \tag{1}$$

As a new packet is innovative in D if and only if it is received from only one of the relays, then we need to subtract probability of receiving a packet from both R_2 and R_2 from Equation (1) The probability that a packet is received by D from both R_1 and R_2 is equal to $(1-e)^4$. Thus, the probability of receiving a linearly dependent packet by D is

$$P_{indep} = 2(1-e)^2 - (1-e)^4. \tag{2}$$

Using Equation (2), the number of required transmissions from S to decode the generation completely in D is

$$k_{sc}(2(1-e)^2 - (1-e)^4) = n, \tag{3}$$

where, n is the size of the generation that D needs to decode and k is number of required transmissions from S to have n linearly independent packets in D. In Source Coding, every

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

transmission round includes a transmission from S, therefore this metric equals the total number of transmission rounds in the system.

The total number of transmissions in all links requires us to also determine the number of transmissions from each relay. As each relay transmits a packet if and only if it receives a packet from S, the number of transmissions from each relay is equal to $k(1-e)$. Hence, the total number of transmissions in all links is

$$T_{sc} = 2k_{sc}(1-e) + k_{sc}. \tag{4}$$

Note that the number of linearly dependent packets received in D, i.e., coded packets received by D from both R_1 and R_2 in the case of this scheme, is given by

$$LD_{sc} = k_{sc}.(1-e)^4. \tag{5}$$

3.2. Unrestricted Coding

The approximations derived for number of transmission rounds and total number of transmissions in this Section are achievable lower bounds to the real value. Our assumption is that the system operates in optimally, i.e., each transmitted coded packet is linearly independent for D as well as for the joint information of both relays. The latter means that if a transmission from the source is successfully received by at least one receiver and it increases the joint DoFs of the relays, then it is considered a useful transmission. Although this optimal scheme is achievable, from a practical perspective it may impose increased signaling requirements in the network. Practical schemes will require the same or more coded packet transmissions.

The total number of transmissions from S to R_1 and R_2 is

$$k_{uc} = \frac{n}{1-e^2}, \tag{6}$$

while the total number of transmissions from R_1 and R_2 to D until D can decode the whole generation is

$$k'_{uc} = \frac{n}{1-e}. \tag{7}$$

Using Equation (6) and (7), results in the total number of transmissions in all links being

$$T_{uc} = \frac{n(2-e)}{(1-e)^2}. \tag{8}$$

3.3. Cache Coding

For the Cache Coding protocol, the total number of transmissions from all links can be separated in three separate contributions, as follows

1. The number of transmissions from S to R_1 and R_2 until the relays decode the generation is equal to $\frac{n}{1-e}$.
2. The number of transmissions from R_1 and R_2 to D until they have decoded the whole generation is equal to $2n$.
3. The number of transmissions from R_1 and R_2 to D after they have decoded the whole generation is equal to
$$\frac{n - \sum_{i=0}^{n}(1-e).n.e^i}{1-e}$$

Summing these three equations, we can calculate total number of transmissions in the whole system as

$$T_{cc} = \frac{n}{e} + 2n + \frac{n - \sum_{i=0}^{n}(1-e)ne^i}{1-e}. \tag{9}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

While the total number of transmission rounds in the system is

$$k_{cc} = \frac{n}{e} + \frac{n - \sum_{i=0}^{n}(1-e)n.e^{i}}{2.(1-e)}, \tag{10}$$

where we have used the three different contributions as well. The key difference is that the number of transmission rounds before R_1 and R_2 have decoded the data is equal to number of transmissions from S to R_1 and R_2, i.e., $\frac{n}{1-e}$, and the fact that the number of transmission rounds after the relays have decoded the data is equal to number of transmissions from the R_1 and R_2 to D divided by number of relays.

In order to determine the number of linearly dependent packets in R_1 and R_2, we use the fact that number of packets received by R_1 or R_2 is equal to $n(1-e)$ and that the number of packets received by only one of them is equal to $ne(1-e)$. Thus, the number of dependent packets is equal to subtraction of these two equations.

$$LD_{cc} = n(1-e)^2. \tag{11}$$

4. Absorbing Markov Chain Model

This section provides a full characterization of the Cache Coding protocol using an analytical model based on an Absorbing Markov Chain. We define the states as a triple (r_1, r_2, r), where r_1 and r_2 are the DoFs in R_1 and R_2, respectively, and r represents the DoFs in D. Note that $r_1 + r_2 \geq r$ based on these states. The absorbing states are the states where $r = n$. This absorbing Markov Chain will also include a cost matrix C alongside the matrix of transition probabilities A. C_{ij} is the total number of transmissions in all links required for the transition between states i and j. We have derived closed-form expressions for each element of these two matrices.

4.1. Transition Probabilities

The transition probabilities between states $\mathcal{S} = (r_1, r_2, r)$ and $\mathcal{S}' = (r_1 + 1, r_2 + j, r + k)$ are dependent on the the Cache Coding protocol's stage in the communication process. Thus, we organize the transition probabilities into five cases as follows.

1. $r_1 \leq n-1$ and $r_2 \leq n-1$

   No single relay node has decoded the data or decodes the data after receiving the new coded packet. Thus, both relays receives the coded packet and forward the received packet to D. Thus,

$$P_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} e.(1-e)^2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=1 \\ e.(1-e)^2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \\ (1+e).(1-e)^3 & \text{if } i=1 \text{ and } j=1 \text{ and } k=1 \\ e^2.(1-e) & \text{if } i=1 \text{ and } j=0 \text{ and } k=0 \\ e^2.(1-e) & \text{if } i=0 \text{ and } j=1 \text{ and } k=0 \\ e^2.(1-e)^2 & \text{if } i=1 \text{ and } j=1 \text{ and } k=0 \\ e^2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ 0 & \text{Otherwise} \end{cases}$$

2. $r_1 = n-1$ and $r_2 \leq n-1$ or $r_1 \leq n-1$ and $r_2 = n-1$

   In this case, if one relay node with DoFs equal to $n-1$ receives a new packet, it decodes the generation completely and starts to send recoded data to D. However, the other relay node still forwards the received coded packet. Thus,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

$$P_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} e.(1-e)^2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=1 \\ e.(1-e)^2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \\ 2e.(1-e)^3 & \text{if } i=1 \text{ and } j=1 \text{ and } k=1 \\ e^2.(1-e) & \text{if } i=1 \text{ and } j=0 \text{ and } k=0 \\ e^2.(1-e) & \text{if } i=0 \text{ and } j=1 \text{ and } k=0 \\ e^2.(1-e)^2 & \text{if } i=1 \text{ and } j=1 \text{ and } k=0 \\ e^2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ (1-e)^4 & \text{if } i=1 \text{ and } j=1 \text{ and } k=2 \\ 0 & \text{Otherwise} \end{cases}$$

3. $r_1 = n$ and $r_2 < n$

   In this case, R_1 recodes and sends data to D, while R_2 forwards the received coded packets. Even if R_2 receives a new packet and decodes data completely(The case $r_2 = n-1$), it recodes data and sends a recoded packet to destination, which is different from the packet sent by R_1. The difference between this case and case 2 is the fact that R_1 sends a recoded packet to D whether or not it receives a packet from S. How ever this event does not occur in case 2. Thus,

$$P_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} (1-e)^3 & \text{if } i=0 \text{ and } j=1 \text{ and } k=2 \\ 2e.(1-e)^2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \\ (1-e).e^2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=0 \\ e^2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ e.(1-e) & \text{if } i=0 \text{ and } j=0 \text{ and } k=1 \\ (1-e^2).(1-e) & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \text{ and} r=n-1 \\ 0 & \text{Otherwise} \end{cases}$$

4. Case 4 : $r_1 < n$ and $r_2 = n$

   In this case, R_2 recodes and sends data to D, while R_1 forwards the received coded packets. Even if R_1 receives a new packet and decodes data completely(The case $r_1 = n-1$), it recodes data and sends a recoded packet to destination, which is different from the packet sent by R_2. The difference between this case and case 2 is the fact that R_2 sends a recoded packet to D whether or not it receives a packet from S. How ever this event does not occur in case 2. Thus,

$$P_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} (1-e)^3 & \text{if } i=1 \text{ and } j=0 \text{ and } k=2 \\ 2e.(1-e)^2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=1 \\ (1-e).e^2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=0 \\ e^2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ e.(1-e) & \text{if } i=0 \text{ and } j=0 \text{ and } k=1 \\ (1-e^2).(1-e) & \text{if } i=1 \text{ and } j=0 \text{ and } k=1 \text{ and} r=n-1 \\ 0 & \text{Otherwise} \end{cases}$$

5. Case 5 : $r_1 = n$ and $r_2 = n$

   In this case, both R_1 and R_2 have decoded the data. Therefore, S will not transmit any longer to R_1 or R_2, while R_1 and R_2 will recode the decoded data and send different coded packets to D.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

$$P_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} e^2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ 2e.(1-e) & \text{if } i=0 \text{ and } j=0 \text{ and } k=1 \\ (1-e^2) & \text{if } i=0 \text{ and } j=0 \text{ and } k=2 \\ 0 & \text{Otherwise} \end{cases}$$

4.2. Cost of transitions

The cost of any transition are 0, 1 or 2 depending on number of broadcast transmissions that occur. Using the cases introduced in Subsection 4.1, we can derive the cost of transitions

1. $r_1 \leq n-1$ and $r_2 \leq n-1$

$$\mathcal{C}_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} 2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=1 \\ 2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \\ 3 & \text{if } i=1 \text{ and } j=1 \text{ and } k=1 \\ 2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=0 \\ 2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=0 \\ 3 & \text{if } i=1 \text{ and } j=1 \text{ and } k=0 \\ 1 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ 0 & \text{Otherwise} \end{cases}$$

2. $r_1 = n-1$ and $r_2 \leq n-1$ or $r_1 \leq n-1$ and $r_2 = n-1$

$$\mathcal{C}_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} 2 & \text{if } i=1 \text{ and } j=0 \text{ and } k=1 \\ 2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \\ 3 & \text{if } i=1; \text{and } j=0 \text{ and } k=0 \\ 2 & \text{if } i=0 \text{ and } j=1 \text{ and } k=0 \\ 3 & \text{if } i=1 \text{ and } j=1 \text{ and } k=0 \\ 1 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ 3 & \text{if } i=1 \text{ and } j=1 \text{ and } k=2 \\ 0 & \text{Otherwise} \end{cases}$$

3. $r_1 = n$ and $r_2 < n$

$$\mathcal{C}_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} 3 & \text{if } i=0 \text{ and } j=1 \text{ and } k=2 \\ 3 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \\ 3 & \text{if } i=0 \text{ and } j=1 \text{ and } k=0 \\ 2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=0 \\ 2 & \text{if } i=0 \text{ and } j=0 \text{ and } k=1 \\ 3 & \text{if } i=0 \text{ and } j=1 \text{ and } k=1 \text{ and} r=n-1 \\ 0 & \text{Otherwise} \end{cases}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

4. $r_1 < n$ and $r_2 = n$

$$\mathcal{C}_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} 3 & \text{if } i = 1 \text{ and } j = 0 \text{ and } k = 2 \\ 3 & \text{if } i = 1 \text{ and } j = 0 \text{ and } k = 1 \\ 3 & \text{if } i = 1 \text{ and } j = 0 \text{ and } k = 0 \\ 2 & \text{if } i = 0 \text{ and } j = 0 \text{ and } k = 0 \\ 2 & \text{if } i = 0 \text{ and } j = 0 \text{ and } k = 1 \\ 3 & \text{if } i = 1 \text{ and } j = 0 \text{ and } k = 1 \text{ and} r = n - 1 \\ 0 & \text{Otherwise} \end{cases}$$

5. $r_1 = n$ and $r_2 = n$

$$\mathcal{C}_{\mathcal{S}\to\mathcal{S}'} = \begin{cases} 2 & \text{if } i = 0 \text{ and } j = 0 \text{ and } k = 0 \\ 2 & \text{if } i = 0 \text{ and } j = 0 \text{ and } k = 1 \\ 2 & \text{if } i = 0 \text{ and } j = 0 \text{ and } k = 2 \\ 0 & \text{Otherwise} \end{cases}$$

4.3. Performance Analysis using the Markov Chain

4.3.1. Calculation of number of transitions

After establishing all transition probabilities in a transition matrix, we can build the fundamental matrix F as follows [20]:

$$F = \begin{bmatrix} I_{r\times r} & Z_{r\times t} \\ R_{t\times r} & Q_{t\times t}, \end{bmatrix}$$

where, t is number of transient (non-absorbing) states and r is number of absorbing states. Q is the matrix of transition probabilities between transient states. R is the matrix of transition probabilities from transient states to absorbing states. I is the identity matrix and Z is an all-zero matrix. After building this matrix, the mean number of transitions can be calculated using Definition 1.

Definition2. *The mean number of transitions before being absorbed when starting in a transient state i, is the i^{th} element of the column vector*

$$M = (I - Q)^{-1}\Gamma,$$

where I is the identity matrix with same dimensions as Q and Γ is all-one vector. [23]

4.3.2. Calculation of number of transmissions from all nodes

The number of transmissions from all node is calculated using both transition matrix A and cost matrix C.

Definition2. *For an absorbing Markov Chain in state i, the mean number of transmissions in its next transition is equal to* [22]

$$E[Tr_i] = \sum_{j=0}^{n} A_{ij}C_{ij}$$

Theorem 1. *For an absorbing Markov chain with state probability π, the mean number of transmissions in next transition is equal to*

$$E[Tr] = \pi(A.C)\Gamma,$$

where $A.C$ represents the element to element multiplication between two matrices A and C and Γ is all-one column vector.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

Proof-sketch of Theorem 1. Using Definition 2, for a system in state I, average number of transmission is equal to $\sum_{i=0}^{n} A_{ij}.C_{ij}$. As π_i shows the probability for the system to be in state i; So average number of transmissions in system equal to $\sum_{i=0}^{n} pi_i \sum_{j=0}^{n} A_{ij}.C_{ij}$.

□

5. Results

In this section, we assess the validity of proposed models and compare the three recoding schemes using various performance measures. We simulate the topology of Figure 1 using the KODO library [21] in C++ to perform encoding/decoding operations. We have carried 1000 independent experiments for each generation size and packet loss rate and report the average of this measurements. This experiments were carried out by using GF(2^8) for encoding/decoding operations.

We have calculated the total number of transmissions in Source Coding. Using equation (4) in our heuristics, we can approximate the total number of transmission in all links for Source Coding. Figure 3 compares simulation results with our heuristics, showing that the approximated model can estimate this metric within 10 % of the real value for a wide range of generation sizes and packet loss rates. This difference comes from the fact that additional transmissions from S to both R_1 and R_2 occur, while these packets are not innovative for D.

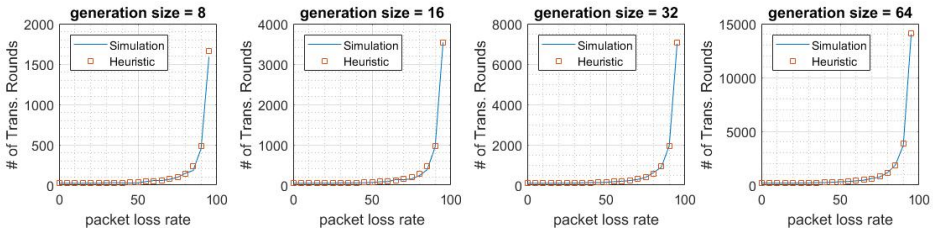

Figure 3. Average number of transmissions in all links in Source Coding

Figure 4 shows the total number of transmissions in all links for Unrestricted Coding. This figure shows that our heuristic model has at most 20 percent deviation with simulation results for all generation sizes and packet loss rates. However in moderate loss rate, there is only 10 percent derivation between simulation and heuristic results. This gap is created because the heuristic is a lower bound for Unrestricted Coding.

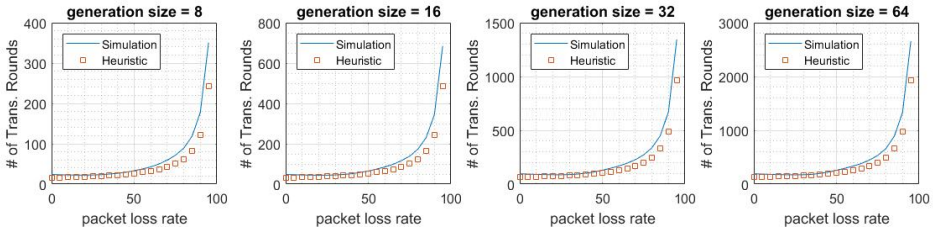

Figure 4. Average number of transmissions in all links in Unrestricted Coding

Figure 5 illustrates the average number of transmission rounds in order to decode the data in D for different generation sizes in Cache Coding. We compare our simulation results with the two proposed models. Figure 6 shows the average number of transmissions for different generation sizes in Cache Coding. In both these metrics, there is at most 10 % derivation between heuristic and simulation. The derivation between Absorbing Markov Chain and simulation is at most 6 %. Abosrbing Markov Chain model is more precise than heuristic, because we did not consider number of relay nodes in our heuristic model. Tables 1 and 2 summarize our key results for Cache Coding.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

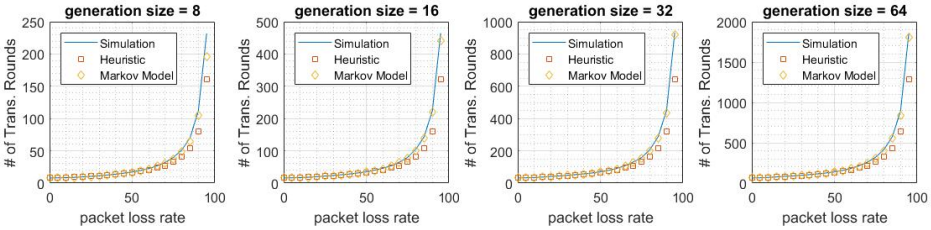

Figure 5. Average number of transmission rounds in Cache Coding

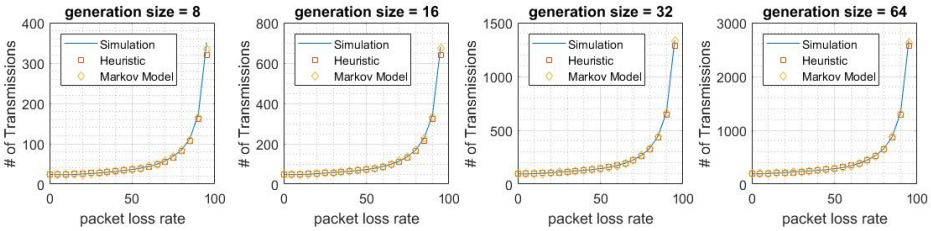

Figure 6. Average number of transmissions in all links in Cache Coding

Table 1. Total Number of transmission rounds in Cache Coding

Generation Size	error = 0.5			error = 0.75		
	simulation	Heuristic	Markov	simulation	Heuristic	Markov
8	10.4695	10.854200	10.387300	40.071900	32.562500	37.576200
16	21.913100	21.708300	21.261800	79.952000	65.125000	76.481200
32	44.574400	43.416700	43.685100	158.715000	130.250000	152.631800
64	89.816200	86.833300	84.539200	322.970000	260.500000	308.481900

Table 2. Total Number of transmissions in all links in Cache Coding

Generation Size	error = 0.5			error = 0.75		
	simulation	Heuristic	Markov	simulation	Heuristic	Markov
8	26.373600	27.333300	26.527300	69.231800	66.000000	68.246300
16	54.644400	54.666700	54.638100	135.878000	132.000000	134.491700
32	112.070000	109.333000	111.984200	268.762000	264.000000	265.871200
64	226.947000	218.667000	223.247100	534.164000	528.000000	528.796300

Figures 7 and 8 show the number of linearly dependent packets received in D for Cache Coding and Source Coding, respectively. As shown, this number decreases by increasing packet loss rate. This comes from the fact that there is a lower probability of receiving the same packets by R_1 and R_2 in higher packet loss rates, which decreases the probability of receiving non-innovative packets in D.

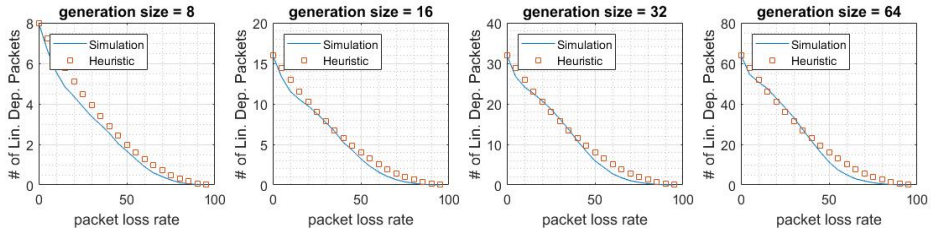

Figure 7. Number of linearly dependent packets in D in Cache Coding

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2018 doi:10.20944/preprints201802.0009.v1

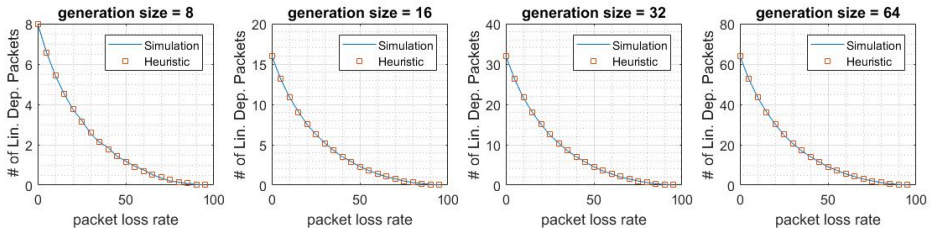

Figure 8. Number of linearly dependent packets in D in Source Coding

We also compare the performance of Cache Coding and Unrestricted Coding. In Figures 9 and 10 we show the average number of transmissions in all links for these coding schemes. The results for simulation confirm that Cache Coding is only 7 % worse than Unrestricted Coding in terms of total number of transmissions in worst case. This gap is larger for results of heuristic. There is at most 12 % derivation in heuristic models of Unrestricted Coding and Cache Coding.

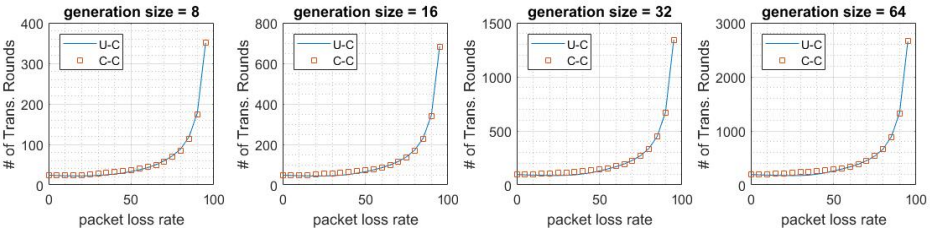

Figure 9. Average number of transmissions in Cache Coding and Unrestricted Coding by simulation results

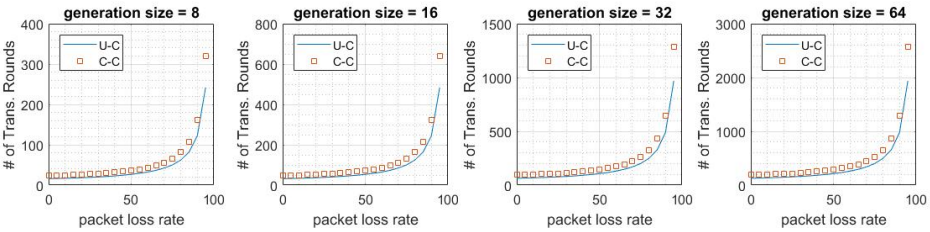

Figure 10. Average number of transmissions in Cache Coding and Unrestricted Coding by Heuristic results

6. Conclusions and Future Work

This paper presents a model and full characterization for the Cache Coding protocol presented in [16] as well as an approximated, yet accurate model for this and other protocols of interest. This analysis also considers an achievable lower bound for the number of transmissions in the network, which serves as the gold standard to measure various protocols, including the Cache coding protocol.

Using our models and simulation results, we have confirmed that the Cache Coding protocol, which is originally designed to overcome the security issues of Unrestricted Coding, can overcome the security issues with negligible decrease in performance in the two-relay system analysed in this paper. Although this fact had been shown by simulations [16], our paper provides the first confirmation using analytical models to describe the system.

Our future work will consider closed-form expressions for other performance metrics, such as number of linearly dependent packets received by the Destination Node. Future work will consider extensions of the analytical model to systems with more than two relays.

Author Contributions: The mathematical analysis of source coding and cache coding and the simulation of all the methods is done by Behnaz Maboudi. The analytical part for Markov Chain method was carried out by Hadi

Sehat and Behnaz Maboudi. Hadi Sehat was responsible for writing and editing the paper using Behnaz's help. The analysis and simulation were taken under supervision of Dr. Pahlevani and Dr. Lucani.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahlswede, R.; Cai, N.; and YENG, R.W., Network information flow.*IEEE T Inform Theory*, **2000** *46.4* 1204-1216, doi:10.1109/18.850663.
2. Koetter, R.; Medard, M., An algebraic approach to network coding. *IEEE ACM T Network*, **2003** *46.4* 782-795,doi:10.1109/TNET.2003.818197.
3. Ho, T.; Medrad, M.; Koetter, D.; Karager, D.R., A Random Linear Network Coding Approach to Malticast. *IEEE T Inform Theory*, **2006** *11.5* 4413-4430,doi:10.1109/TIT.2006.881746.
4. Lee, S.H; Gerla, M.; Krawczyk, H.; Lee, K.W.; Quaglia, E., Performance evaluation of secure network coding using homomorphic signature. In *Proceedings of International Symposium on Network Coding*, Beijing, China, 25-27 July 2011, pp.1-6.
5. Gennaro, R,; Katz, J.; Krawczyk, H.; Rabin, T.,Secure network coding over the integers. In *Proceedings of 13th International Conference on Practice and Theory in Public Key Cryptography*, Paris, France, 26-28 May 2010, pp. 142–160.
6. Boneh, D.; Freeman, D.; Katz, J.; Waters, B.,Signing a linear subspace: Signature schemes for network coding. In *Proceedings of 12th International Conference on Practice and Theory in Public Key Cryptography*, Irvine, CAL, USA, 18-20 March 2009, pp. 68–87.
7. Agrawal S.; Boneh D.,Homomorphic MACs: MAC-Based Integrity for Network Coding. In *Proceedings of 7th Applied Cryptography and Network Security. ACNS 2009*, Paris-Rocquencourt, France, 2-5 June 2009, pp. 292-305.
8. Zhao, F.; Kalker, T.; Medard, M.; Han, K. J.,Signatures for content distribution with network coding. In *Proceedings of IEEE International Symposium on Information Theory, ISIT 2007*, Nice, France, 24-29 June 2007, pp. 556-560.
9. Han, K.; Ho, T.; Koetter, R.; Medard, M.; Zhao, F., On network coding for security. In *Proceedings of Military Communications Conference, MILCOM 2007*, Orlando, FL, USA, 29-31 October 2007, pp. 1-6.
10. Dong, J.; Curtmola, R.;Nita-Rotaru, C., Practical defenses against pollution attacks in intra-flow network coding for wireless mesh networks. In *Proceedings of the second ACM conference on Wireless network security*, Zurich, Switzerland, 16-19 March 2009, pp. 111-122.
11. Esfahani, A.; Mantas, G.;, Rodriguez, J.; Neves, J. C. ,An efficient homomorphic MAC-based scheme against data and tag pollution attacks in network coding-enabled wireless networks. *INT J INF SECUR*, **2017** *16.6* 4413-4430, doi:10.1007/s10207-016-0351-z.
12. Lucani, D. E.; Médard, M.; Stojanovic, M., Broadcasting in time-division duplexing: A random linear network coding approach. In *Proceedings of Network Coding, Theory, and Applications, 2009. NetCod'09*, Lausanne, Switzerland, 15-16 June 2009, pp. 62-67.
13. Tsimbalo, E.; Tassi, A.; Piechocki, R. J., Novel Performance Analysis of Network Coded Communications in Single-Relay Networks. In *Proceedings of Global Communications Conference (GLOBECOM), 2016 IEEE*, Washington, DC, USA, 4-8 December 2016, pp. 1-6.
14. Tsimbalo, E.; Tassi, A.; Piechocki, R. J., Reliability of Multicast under Random Linear Network Coding. In *arXiv preprint*, arXiv preprint arXiv:1709.05477.
15. Khamfroush, H.; Pahlevani, P.; Lucani, D. E., Hundeboll, M., On the coded packet relay network in the presence of neighbors: Benefits of speaking in a crowded room. In *Proceedings of IEEE International Conference on Communications (ICC), 2014*, Sydney, Australia, 10-14 June 2014, pp. 1928-1933.
16. Joy, J.; Yu-Ting, Y.; Perez, V.; Lu, D.; Gerla, M., A new approach to coding in content-based MANETs. In *Proceedings of 2014 International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA , 3-6 February 2014, pp. 173-177.
17. Khamfroush, H.; Lucani, D.E.; Pahlevani, P.; Barros, J., On optimal policies for network-coded cooperation: Theory and implementation. *IEEE J SEL AREA COMM*, **2015** *33.2* 199-212, doi:10.1109/JSAC.2014.2384291.
18. Khamfroush, H.; Lucani, D. E.; Barros, J.; Pahlevani, P., Network-Coded Cooperation Over Time-Varying Channels. *IEEE T COMMUN*, **2014** *62.12* 4413-4425,doi:10.1109/TCOMM.2014.2367016.

19. Zhang, C.; Fang, Y.; Zhu, X., Throughput-Delay Tradeoffs in Large-Scale MANETs with Network Coding *IEEE INFOCOM 2009* **2009** Rio de Janeiro, Brazil, 19-25 April 2009, pp.199-207, doi: 10.1109/INFCOM.2009.5061922.
20. Kemeny, J. G.; Snell J. L.; et al., Finite markov chains. *van Nostrand Princeton*, *356*, NJ, US, 1960.
21. Pedersen, M. V.; Heide, J.; Fitzek, F. H., Kodo: An open and research oriented network coding library. In *International Conference on Research in Networking*, Springer, Berlin, Heidelberg, 9 May 2011, pp. 145-152.
22. Lucani, D. E.; Medard, M.; Stojanovic, M. On coding for delay—Network coding for time-division duplexing. *IEEE Transactions on Information Theory*, **2012**, *58.4*, pp. 2330-2348.
23. Garrido, P.; Lucani, D. E.; Agüero, R., Markov chain model for the decoding probability of sparse network coding.*IEEE Transactions on Communications*, **2017** *65.4*, pp. 1675-1685.