On the Maximum Throughput of Two-Hop Wireless Network Coding

Deze Zeng∗†, Song Guo∗, Hai Jin † and Shui Yu‡

∗School of Computer Science and Engineering, The University of Aizu, Japan
Email: {d8112106, sguo}@u-aizu.ac.jp
†School of Computer Science and Technology, Huazhong University of Science and Technology, China
Email: hjin@hust.edu.cn
‡School of Engineering and Information Technology, Deakin University, Australia
Email syu@deakin.edu.au

Abstract—Network coding has shown the promise of significant throughput improvement. In this paper, we study the throughput of two-hop wireless network coding and explore how the maximum throughput can be achieved under a random medium access scheme. Unlike previous studies, we consider a more practical network where the structure of overhearing status between the intended receivers and the transmitters is arbitrary. We make a formal analysis on the network throughput using network coding upon the concept of network coding cliques (NCCs). The analysis shows that the maximum normalized throughput, subject to fairness requirement, is \( \frac{n}{n+m} \), where \( n \) is the number of transmitters and \( m \) is the number of NCCs in a 2-hop wireless network. We have also found that this maximum throughput can be achieved under a random medium access scheme when the medium access priority of the relay node is equal to the number of NCCs in the network. Our theoretical findings have been validated by simulation as well.

I. INTRODUCTION

Network Coding (NC) origins from the seminal work [1] and has been considered as a promising technique to improve the throughput via encoding the packets at the intermediate relay. Especially in wireless networks, by exploiting the intrinsic broadcast nature of wireless communication, network coding can significantly increase the network throughput compared to the traditional routing transmission schemes [2], [3].

The first practical wireless network coding system COPE is introduced by Katti et al. in [2]. The principle of COPE can be illustrated by Fig. 1(a) in which all transmitters cannot communicate directly with their intended receivers, but via a relay node. Some receivers can overhear the packets sent by their neighbors as shown by dashed lines in Fig. 1(a). Instead of forwarding the packets to the intended receiver one by one, the relay node encodes the packets by an exclusive-or (XOR) operation according to the overhearing status and broadcasts the XORed one. Upon receiving the encoded packet, receiver \( Rx_1 \) and \( Rx_2 \) retrieve their desired packets by XORing with the overheard packets while receiver \( Rx_3 \) and \( Rx_4 \) by XORing with the packets which they just sent before. The former is known as inter-session network coding and the latter as intra-session network coding.

The outstanding performance of COPE has inspired many research activities. Some papers study on the NC-aware optimization to improve the network performance, e.g., [4], [5]. Others focus on the performance analysis that has provided insight into the throughput gain by network coding compared to the traditional store-and-forward scheme, e.g., [2] [6] and [7]. In particular, Katti et al. [2] first derive that the throughput gain is at most 2 and the result is verified by a practical system. Later on, Le et al. show that the upper bound of the throughput is \( \frac{n}{n+1} \) in a network with \( n \) transmitters and a shared bandwidth \( B \) [6].

Is the upper bound \( \frac{n}{n+1} \) tight enough? We re-investigate this problem by taking Fig. 1(a) as an example. Suppose that the shared bandwidth is normalized to one and each packet transmission takes one time unit. One can easily find out that it takes at least 6 time units to complete one transmission at each transmitter. In other words, the throughput is \( \frac{4}{6} = 0.67 \), other than 0.8 calculated by \( \frac{4}{4+1} \). The reason is that the result in [6] is based on a strong assumption that each receiver can overhear all others. This observation motivates us to explore more accurate throughput upper bound. In summary, we make the following contributions in this paper:

- We derive a throughput upper bound in a two-hop wireless network using network coding and show that the throughput is determined by both the number of flows in the networks and the overhearing status of the network. The letter factor is determined by the network connec-

![Fig. 1. A network coding scenario.](image-url)
tivity. According to the overhearing status, we partition the transmitters into Network Coding Cliques (NCCs, as shown in Fig. 1(b)) and prove that the normalized throughput is bounded by \( \frac{n}{n+m} \), where \( n \) is the number of flows and \( m \) is the number of NCCs.

- Under random medium access scheme, we prove that the above throughput upper bound is achievable when the medium access priority of the relay is the same as the number of NCCs in the network and the buffer capacity at relay node is infinity. Furthermore, we find that under random medium channel access, the achievable throughput is affected only by the number of NCCs, but not the clique size. Our analysis has been also verified by simulations.

The rest of this paper is organized as follows. Section II discusses previous studies on network coding. Section III gives a network model and introduces the concept of NCC. Section IV presents our theoretical analysis. The theoretical findings are verified by experiment in Section V. Finally, Section VI concludes our work.

II. RELATED WORK

After COPE has been proposed and implemented by Katti et al. [2], the major research following COPE is toward two directions, the performance analysis of network coding and the NC-aware optimization.

For the performance analysis, Katti et al. [2] first derive that the network coding gain is at most 2, which is also validated by the practical COPE system they proposed. Later, Liu and Xue [8] characterize the achievable rate regions and find the theoretical optimal sum rates for a basic 3-node topology. Liu, Goeckel and Towsley [9] show the benefit of network coding for 2D and 1D networks. Keshavarz-Haddad and Riedi [10] prove that the gain of network coding in terms of transport capacity is bounded by a constant factor \( \pi \) in an arbitrary wireless network under traditional channel models. Le et al. [6] derive an upper bound of the throughput as \( \frac{n}{n+1} \) based on the encoding number. Furthermore, they argue that the throughput upper bound could be achieved under random medium access scheme when all the devices, including the relay, have the same medium access priority. More recently, Iraji et al. [11] propose a new analytical model for throughput evaluation of a wireless tandem network coding based on a multi-class open queueing network. Umehara et al. [12] provide explicit expressions of the throughput for a single-relay two-hop wireless CSMA system and show that the transmission probability of the relay node is a design parameter that is crucial for maximizing the achievable throughput with network coding.

For the network coding aware optimization, a general framework to develop optimal and adaptive schemes of joint network coding and MAC scheduling is proposed in [4], in which network coding will be applied only when it can bring performance benefits. Yomo et al. [13] introduce an opportunistic scheduling which selects an appropriate set of nodes whose packets are network-coded as well as the data rate for the broadcast transmission according to the instantaneous link conditions. Kim and Veciana consider joint rate adaptation and network coding [14], [15]. They suggest that the relay should have higher access priority than other nodes to achieve higher throughput, but they do not present the exact access priority of the relay. Later, Seferoglu et al. [16] take into consideration for both the network coded flows and the induced conflicts between nodes in the throughput optimization problem of joint rate control and scheduling for intersession wireless network coding. Khreishah et al. [17] consider the joint optimal coding, scheduling, and rate-control scheme for wireless multi-hop networks using pairwise intersession coding.

III. SYSTEM MODEL AND CODING STRUCTURE

A. Network Model

We consider a two-hop wireless network model with a single relay node \( R \) and \( n \) transmitter-receiver pairs. For each flow \( i \) (\( 1 \leq i \leq n \)), the transmitter is denoted by \( Tx_i \) (\( Tx_i \in TX \)) and the corresponding receiver is \( Rx_i \) (\( Rx_i \in RX \)). \( Tx_i \) cannot directly communicate with \( Rx_i \) but with the help of relay \( R \). A pair of transmitter and receiver specifies a session. A session may have one flow in one direction or two flows in both directions, so \( TX \cap RX \neq \emptyset \) is possible.

Let \( R_i \) be the transmission range of transmitter \( Tx_i \). Loss-free channel is assumed and the shared bandwidth is normalized to one. As a result, the throughput discussed in this paper always means the normalized one. To avoid starvation, all transmitters share the common medium with the same priority.

B. Network Coding Clique

For a fixed topology and known transmission ranges, the set of receivers that can overhear a transmitter is determined, resulting in that the packets can be encoded for inter-session network coding. On the other hand, packets from opposite flows within the same session could be also encoded as intra-session network coding. We construct an undirected graph \( G \) by connecting any two transmitters with an edge if and only if 1) each can be overheard by the other’s receiver or 2) the two nodes are in the same session with bi-directional flows.

In formal, we define a coding graph \( G = (TX, E) \), which is composed of a set of transmitters \( TX \) and a set of undirected edges \( E \). The link \( (Tx_i, Tx_j) \in TX \times TX \) is in \( E \) if 1) \( Rx_i \) can overhear \( Tx_j \)’s transmission and \( Rx_j \) can overhear \( Tx_i \)’s transmission or 2) \( Tx_j (= Rx_i) \) is the receiver of \( Tx_i \) and \( Tx_i (=Rx_j) \) is the receiver of \( Tx_j \). So \( E \) is formally defined as:

\[
E = \{ (Tx_i, Tx_j) \in TX \times TX \mid |Tx_i Rx_j| \leq R_i \text{ and } |Tx_j Rx_i| \leq R_j \}
\]

Notice that the intra-session network coding is a special case where the Euclidian distance \( |Tx_i Rx_j| \) and \( |Tx_j Rx_i| \) are equal to 0. For example, in Fig. 1(a) the distance between \( Tx_3 \) and \( Rx_4 \) is 0 because they refer to the same device.

By constructing coding graph \( G \), all the transmitters can be partitioned into one or several cliques. Each clique includes
which should not exceed the input traffic \( \sum \mu \). The corresponding contribution to the overall throughput is \( V \).

In our network, we know that at most one transmission is active at a time, i.e.,

\[
TX = \bigcup_{i=1}^{m} V_i, \quad V_i \cap V_j = \emptyset, \quad 1 \leq i \neq j \leq m.
\]

IV. Coding Performance Analysis

Based on our network model and the coding structure, we first analyze the upper bound of network throughput under fairness requirement. Then we explore how this bound can be achieved under random access scheduling.

A. Throughput Bound

When fairness is not considered, i.e., the throughput of each flow could be arbitrary, the maximum network throughput is obtained when transmissions happen only in an NCC with the maximum size \( |V_{\text{max}}| \). This result is elaborated by the following theorem.

Theorem 1: The maximum network throughput is \( \frac{|V_{\text{max}}|}{|V_{\text{max}}| + 1} \), which is achieved when only transmitters in an NCC \( V \) with \( |V| = |V_{\text{max}}| \) transmit.

Proof: Let \( \lambda_i (1 \leq i \leq n) \) and \( \mu_j (1 \leq j \leq m) \) denote the time share for transmission at \( TX_i \) and relay \( R \) for \( V_j \), respectively. For all the traffic that originates in each \( V_j \), the corresponding contribution to the overall throughput is \( |V_j| \cdot \mu_j \), which should not exceed the input traffic \( \sum_{T_x \in V_j} \lambda_i \), i.e.,

\[
\sum_{T_x \in V_j} \lambda_i \geq |V_j| \cdot \mu_j, \quad (1)
\]

which can be further derived to \( \mu_j \leq \frac{\lambda_i + \sum_{T_x \in V_j} \lambda_i}{|V_j| + 1} \).

By substituting the above result into the throughput expression, we have,

\[
T = \sum_{j=1}^{m} |V_j| \cdot \mu_j \\
\leq \sum_{j=1}^{m} \frac{|V_j|}{|V_j| + 1} \cdot (\mu_j + \sum_{T_x \in V_j} \lambda_i) \\
\leq \max_{1 \leq j \leq m} \left( \frac{|V_j|}{|V_j| + 1} \right) \cdot \sum_{j=1}^{m} (\mu_j + \sum_{T_x \in V_j} \lambda_i) \\
= \frac{|V_{\text{max}}|}{|V_{\text{max}}| + 1} \cdot \left( \sum_{j=1}^{m} \mu_j + \sum_{i=1}^{n} \lambda_i \right).
\]

The above derivation leads to \( T \leq \frac{|V_{\text{max}}|}{|V_{\text{max}}| + 1} \) and the equation holds when all traffic only originates from an NCC with maximum size.

This transmission scheme that achieves the maximum network throughput is sometimes unacceptable because the transmitters in smaller NCCs suffer starvation. To guarantee the fairness, the throughput of each flow should be the same, i.e., \( \lambda_i = \lambda \) for \( 1 \leq i \leq n \) and \( m \mu_j = m \mu \) for \( 1 \leq j \leq m \).

Theorem 2: The maximum network throughput with guaranteed fairness is \( T^* = \frac{n \mu}{n + m} \), where \( n \) is the number of transmitters and \( m \) is the number of NCCs.

Proof: Because each flow has the same throughput, by using (1), we can have the derivations as:

\[
n \mu = \sum_{j=1}^{m} |V_j| \cdot \mu_j \leq \sum_{j=1}^{m} \sum_{T_x \in V_j} \lambda_i = \sum_{i=1}^{n} \lambda_i = n \lambda.
\]

Furthermore, (2) becomes

\[
n \lambda + m \mu = \sum_{i=1}^{n} \lambda_i + \sum_{j=1}^{m} \mu_j \leq 1.
\]

Combining the above two equations, we have \( \mu \leq \frac{1}{n + m} \). As a result, \( T = \mu \sum_{j=1}^{m} |V_j| = n \mu \leq \frac{n}{n + m} \). Finally, we have \( T^* = \frac{n}{n + m} \).

B. Throughput Bound under Random Access and Fairness Scheduling

We can view the buffer at relay logically as a number of separate ones, each of which is maintained for an individual flow. For one specific flow, say \( TX_i \rightarrow R \rightarrow RX_i \), an embedded Markov chain can be used to describe the dynamic of the buffer as shown in Fig. 2.

Let \( \Gamma_i (i = 1, 2, \ldots, n) \) and \( \Gamma_c \) denote the transmission probability of the transmitter \( TX_i \) and the forwarding probability of the relay for flow \( i \), respectively. We note that \( \Gamma_i \) also represents the forwarding probability for the clique \( V_i \). Due to the fairness, packets are encoded and forwarded with the same probability for each clique, resulting in \( \Gamma_c = \Gamma_i / m \), where \( \Gamma_c \) is the transmission probability of the encoding/relay node \( R \) and \( m \) is the number of cliques.

All devices randomly access the medium for transmission whenever they have pending data. We use \( \gamma_i \) and \( \gamma_c \) to denote the probability that the transmitter \( TX_i \) and the relay node \( R \) compete for the medium in a time unit, respectively. The relay
node competes whenever at least one logical buffer (i.e., the buffer for a flow) is not empty. Therefore, we have:

\[
\gamma_c = 1 - \prod_{i=1}^{n}(1 - \zeta_i), \quad (3)
\]

where \(\zeta_i\) denotes the probability that the logical buffer for flow \(i\) is not empty.

Let \(\psi_i^j\) denote the steady-state probability that the logical buffer of flow \(i\) has \(j\) packets. We then have:

\[
\zeta_i = 1 - \psi_i^0. \quad (4)
\]

Note that \(\psi_i^j\) can be obtained by solving the embedded Markov chain given in [Fig. 2] and the details will be discussed in the next section.

Now we examine the network throughput using random access mechanism, in which the medium access priority of the relay node will be given special interest.

We consider a mechanism called \(K\)-priority random medium access, in which all transmitters have the same priority 1 for guaranteed fairness and the relay node has a priority \(K\) \((K \geq 1)\). Notice that “\(K = 1\)” means equal access priority for all devices, including the relay node. A successful transmission (the relay node will be given special interest).

To examine the maximum throughput under random access and fair scheduling when the encoding node priority is the same probability, we have:

\[
\begin{align*}
\Gamma_i &= \gamma_i / K \gamma_c + \sum_{j \neq i} \gamma_j + 1, \\
\Gamma_c &= \gamma_c / K + \sum_i \gamma_i. 
\end{align*}
\]

Recalling that the relay node fairly serves each NCC with the same probability, we have:

\[
\Gamma_c^i = \frac{\Gamma_c}{m} = \gamma_c / m(K + \sum_i \gamma_i). \quad (6)
\]

For a two-hop network with a single relay node, the throughput of the network \(T\) is determined by both the average encoding number \(E\) and the transmission probability of the relay node \(\Gamma_c\), i.e.,

\[
T = ET. \quad (7)
\]

Based on the XOR coding principle, the average encoding number \(E\) is:

\[
E = \frac{1}{m} \sum_{j=1}^{n} |V_j| \sum_{i=1}^{m} \zeta_i. \quad (8)
\]

### C. Maximum Throughput of the \(K\)-priority Mechanism

To examine the maximum throughput under random access and fairness scheduling, we only need to study the case that all the transmitters have saturated traffic. In other words, each transmitter always has pending packets to transmit, i.e., \(\gamma_i = 1\), \(\forall 1 \leq i \leq n\). Applying this to (5) and (6), we have:

\[
\begin{align*}
\Gamma_i^i &= \frac{1}{K \gamma_c + n}, \\
\Gamma_c &= \frac{\Gamma_c}{K + n}, \quad K \geq 1. 
\end{align*}
\]

Assuming that the buffer size is infinite at relay node, we can obtain \(\psi_i^j\) by solving the embedded Markov chain given in Fig. 2 as:

\[
\psi_i^j = \begin{cases} 
1 - \zeta_i & \text{if } j = 0, \\
\zeta_j(1 - \zeta_i) & \text{if } j \neq 0,
\end{cases} \quad (11)
\]

where

\[
\zeta = \zeta_i = \frac{\Gamma_i}{\Gamma_c} = \frac{m(K + n)}{K \gamma_c(K \gamma_c + n)}, \forall 1 \leq i \leq n, K \geq 1. \quad (12)
\]

By substituting (12) into (8), the average encoding number \(E\) becomes:

\[
E = \frac{c}{m} \sum_{j=1}^{m} |V_j| = \frac{n}{m} \cdot \zeta = \frac{n(K + n)}{K \gamma_c(K \gamma_c + n)}. \quad (13)
\]

Combining (7) (9) and (12), the network throughput under random access and fair scheduling can be also expressed as:

\[
T = \frac{\frac{n}{m} + K \gamma_c}{K \gamma_c}. \quad (14)
\]

Finally, the maximum achievable network throughput is given by the following theorem.

**Theorem 3:** The maximum throughput of the \(K\)-priority mechanism is \(\frac{m}{n+K \gamma_c}\), which is achieved when \(K\) is equal to the number of cliques.

**Proof:** Because \(\zeta \leq 1\) is required to achieve the solution given in [11], we have:

\[
1 \geq \zeta = \frac{m(K + n)}{K \gamma_c(K \gamma_c + n)} \geq \frac{m}{K \gamma_c}. \quad (15)
\]

Then, we conclude:

\[
K \gamma_c \geq m. \quad (16)
\]

From (14), the throughput \(T\) can be viewed as a decreasing function of \(K \gamma_c\). The maximum \(T\) is achieved when \(K \gamma_c = m\), i.e., \(T^* = \frac{n}{n+K \gamma_c}\). Furthermore, when \(K \gamma_c = m\), we can derive \(\gamma_c = \zeta = 1\) from (15). It further leads to \(K = m\).

From Theorem 3, we can conclude that the maximum throughput given in Theorem 3 also be achieved in random access and fair scheduling when the encoding node priority is equal to the number of network coding cliques.

### V. Performance Evaluation

In this section, we evaluate the accuracy of our theoretical conclusions in different scenarios by comparing with the simulation results. We have implemented a discrete-event simulator which simulates a two-hop wireless network as discussed in Section III.

In our simulation settings, there is a single relay node and 12 transmitters. Since we are interested in the maximum throughput, all transmitters are under the saturated traffic load. The priority competing for channel is set to one, the same for each transmitter, while the priority of the relay node is varied.
from 1 to 12. The shared bandwidth is normalized to one and any device can transmit one packet in a given time unit whenever it gets the channel access opportunity.

A. Simulation Verification

The 12 transmitters are equally partitioned into different number of equal-sized NCCs in different scenarios, e.g. one NCC, 3 NCCs and 4 NCCs. We vary the priority of the relay node in each scenario to obtain the network throughput and the average encoding number, which will be used to verify the analytical results given in Section IV.

In Fig. 3 and Fig. 4, we depict the network throughput and the average encoding number, respectively, under various medium access priorities ($K$) of the relay node. Both analytical and simulation results are presented and they match tightly. The accuracy of our analysis is thus verified.

As mentioned before, the model in [6] can be viewed a special case in our framework. In the first scenario, we set the transmission range of any transmitter to cover all other devices in the network except for the intended receiver to examine the case studied in [6]. As a result, there is only one ($m = 1$) 12-clique. From Fig. 3(a), we notice that $K = 1$ achieves the highest throughput. This is consistent with the conclusion of [6] and also our Theorem 3.

We also observe that the throughput is a decreasing function of the medium access priority of the relay. This is because a higher priority makes the relay node lose coding opportunities which can be seen from Fig. 4(a). The average encoding number is maximized, approaching to 12, while this number decreases with the increasing of $K$.

When there are three ($m = 3$) 4-cliques, we envision that the maximum throughput shall be obtained when the medium access priority at relay node is 3, according to Theorem 3. This is validated by the simulation results as shown in Fig. 3(b). The throughput first shows as an increasing function of $K$ until $K$ reaches 3, and then becomes an decreasing function when $K \geq 3$.

Due to a relatively low medium access priority $K \leq 3$, the relay node always has opportunities to encode all packets from transmitters belonging to the same clique. As a results, most packets are encoded and transmitted. As shown in Fig. 4(b), the average encoding number always approaches to the maximum value 4 when $K \leq 3$. Under such condition, the more encoding opportunities to the relay node, the higher throughput can be achieved. On the other hand, when $K$ continues to increase, i.e. the relay node obtains even higher chances to relay packets, the throughput decays as shown in Fig. 3(b). This is because the benefit of network coding is not fully exploited due to a relatively lower average encoding number as shown in Fig. 4(b).

The same phenomena when $m = 4$ can be also viewed from Fig. 3(c) and Fig. 4(c) where we obtain the highest throughput when $K = m = 4$. 

![Throughput under various medium access priorities (K) of the relay node in a two-hop network with 12 nodes (n=12) and a different number of equal-sized NCCs (m=1, 3 and 4).](image1)

![Average encoding number under various medium access priorities (K) of the relay node in a two-hop network with 12 nodes (n=12) and a different number of equal-sized NCCs (m=1, 3 and 4).](image2)
K of the relay is set to be the number of NCCs in the network. We also derive the closed-form expression of this maximum throughput. Both analytical results and simulation results are presented. They match well and verify the correctness of our theoretical findings. Our work could be a good starting point for the network coding aware MAC protocol design.

REFERENCES


Fig. 5. Throughput under various medium access priorities (K) of the relay node in a two-hop network with 12 nodes (n = 12), which are randomly partitioned into different number of NCCs, i.e. 4 NCCs (m = 4) and 3 NCCs (m = 3), respectively.

B. Does the Clique Size Matter to the Throughput?

As discussed in Section IV, the size of each clique does not affect the throughput of the network. To verify this conclusion, we randomly partition the 12 nodes into 4 cliques with different sizes in the simulation. We present the throughputs of two cases in Fig. 5(a). The first case consists of one 1-clique, two 2-cliques and one 7-clique, denoted as (7, 2, 2, 1). Another case is (6, 3, 2, 1). Notice that for the 1-clique, no coding opportunity exists according to the definition of NCC. In Fig. 5(a) we observe that the throughputs in both cases are almost the same as the theoretical results under various priorities K. Similarly, the results in Fig. 5(b) where the 12 nodes are partitioned as (7, 3, 2) and (6, 3, 3), also demonstrate that the throughput performance are irrelevant to the clique size.

VI. CONCLUSION

We have analyzed the throughput upper bound for a two-hop wireless network with a single relay in this paper. Based on the overhearing status of the network, we partition the transmitters into network coding cliques. Packets from the transmitters in the same clique can be encoded together. We first present the theoretical maximum throughput with and without fairness. Then we show that the maximum throughput with guaranteed fairness could be achieved by a K-priority medium access scheme, in which the medium access priority

Throughput
Medium Access Priority of the Relay Node K
Analysis
Simulation (7, 2, 2, 1)
Simulation (6, 3, 2, 1)

(b) 3 Network Coding Cliques (m = 3)

Medium Access Priority of the Relay Node K
Analysis
Simulation (7, 3, 2)
Simulation (6, 3, 3)

(a) 4 Network Coding Cliques (m = 4)