This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

Optimal Configuration of Network Coding in Ad Hoc Networks

Yi Qin[1], Feng Yang[1], Xiaohua Tian[1], Xinbing Wang[1], *Member, IEEE,* Hanwen Luo[1], Haiquan Wang[2], Mohsen Guizani[3], *Fellow, IEEE*

***Abstract*—In this paper, we analyze the impact of network coding configuration on the performance of ad hoc networks with the consideration of two significant factors: the throughput loss and the decoding loss, which are jointly treated as the overhead of network coding. In particular, physical layer network coding and random linear network coding are adopted in static and mobile ad hoc networks (MANETs), respectively. Furthermore, we characterize the goodput and delay/goodput tradeoff in static networks, which are also analyzed in MANETs for different mobility models (i.e., the random i.i.d. mobility model and the random walk model) and transmission schemes (i.e., the 2-hop relay scheme and the flooding scheme). Moreover, the optimal configuration of network coding, which consists of the data size, generation size and network coding Galois field, is derived to optimize the delay/goodput tradeoff and goodput. The theoretical results demonstrate that network coding does not bring about order gain on delay/goodput tradeoff for each network model and scheme except for the flooding scheme in random i.i.d. mobility model. However, the goodput improvement is exhibited for all the proposed schemes in mobile networks. To our best knowledge, this is the first work to investigate the scaling laws of network coding performance and configuration with the consideration of coding overhead in ad hoc networks.**

I. Introduction

Network coding was initially designed as a kind of source coding [1]. Further studies [2], [3] showed that the capacity of wired networks can be improved by network coding which can fully utilize the network resources. Due to this advantage, how to employ network coding in wireless ad hoc networks had been intensively studied in recent years [4] [5] with the purpose of improving the throughput and delay performance. The main difference between wired networks and wireless networks is that there is non-ignorable interference between nodes in wireless networks [6] [7]. Therefore, it is important to design the network coding in wireless ad hoc networks with interference so as to achieve the improvement on system performance such as goodput and delay/goodput tradeoff.

It was proved in [2] that the maximum flow could be achieved by employing network coding in wired networks. In the last few years, significant efforts had been devoted to designing schemes adopting network coding, aiming at full utilization of network resources in applications such as wireless ad hoc networks [4] [5] [8]-[14], peer-to-peer networks [15], etc. An important work by Liu *et al.* in[4] introduced the observation that only a constant factor of throughput improvement can be brought about to k-dimensional random static networks. Further works by Zhang et al.[5] and [8] analyzed the delay, throughput (including the overhead of network coding) and their tradeoff in fast and slow mobility models for MANETs by employing *Random Linear Network Coding* (RLNC). It was indicated by their results that order improvement of throughput scaling laws can be achieved by adopting RLNC in MANETs. Recent works [9], [10] also focused on the network performance with network coding. [9] presented a theoretical study of the throughput in vehicular ad hoc networks using packet-level network coding and symbol-level network coding. Moreover, [10] demonstrated that the throughput gain of multicast networks by employing network coding was of order $O(\sqrt{\log n})$[1], when the number of destinations m for a source satisfied $m = O\left(\frac{n}{\log^3 n}\right)$ and $m = \Omega\left(\frac{n}{\log n}\right)$. The above works showed that network coding was widely studied in wireless ad hoc networks.

However, a significant factor, say, throughput loss, was not taken into account in these works, while it cannot be ignored in the case of large Galois field as well as large hop counts. For example, if k packets are linearly combined according to network coding, the corresponding k network coding coefficients (in Galois field $\mathbb{F}_q$) are stored in the packet with B bits data. Therefore, the size of each combined packet is $(B + k \log q)$ bits. It takes $(B + k \log q)$ bits to convey B bits data, and this increment cannot be ignored when q and k are large, which is treated as the throughput loss. Another example can be found in [17].

In addition to the throughput loss, the decoding loss is another important factor in network coding, which is caused by decoding failure. It was addressed in some typical works such as [18]. Tracey Ho. *et al.* [18] implied the probability that the random linear network coding was valid for a multicast connection problem on an arbitrary network with independent sources was at least $(1 - d/q)^\eta$, where η was the number of links with associated random coefficients, d was the number of

[1] Yi Qin, Feng Yang, Xiaohua Tian, Xinbing Wang and Hanwen Luo are with the Department of Electric Engineering, Shanghai Jiao Tong University, China. E-mail: {qinyi_33, yangfeng, xtian, xwang8, hwluo}@sjtu.edu.cn Correspondence author: Dr. Xinbing Wang

[2] Haiquan Wang is with the Department of Communication Engineering, Hangzhou Dianzi University, China. E-mail: tx_wang@hdu.edu.cn

[3] Mohsen Guizani is with the Department of Computer Science, Western Michigan University, USA. E-mail: mguizani@ieee.org

[1]We use standard asymptotic notations in our paper. Consider two nonnegative function $f(\cdot)$ and $g(\cdot)$: (1) $f(n) = o(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) = 0$. (2) $f(n) = O(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) < \infty$. (3) $f(n) = \omega(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) = \infty$. (4) $f(n) = \Omega(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) > 0$. (5) $f(n) = \Theta(g(n))$ means $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

receivers and q was the size of Galois field $\mathbb{F}_q$. It was obvious that a large q was required to guarantee that the system with RLNC was valid.

When considering the above two factors, the traditional definition of throughput in ad hoc networks is no longer appropriate since it does not consider the bits of network coding coefficients and the linearly correlated packets which do not carry any valuable data. Instead, the goodput and delay/goodput tradeoff are investigated in this paper, which only take into account the successfully decoded data. Although there were some works focusing on the throughput loss and decoding loss in some other networks [17] [18], the impact of them on scaling laws in ad hoc networks is still a challenging question [19]. Moreover, if we treat the data size of each packet, the generation size (the number of packets which are combined by network coding as a group) and the network coding coefficient Galois field as the configuration of network coding, it is necessary to find the scaling laws of the optimal configuration for given network model and transmission scheme.

To answer these questions, we analyze the goodput and delay/goodput tradeoff performance for both static and mobile models in unicast case. In static model, we consider a static network with physical layer network coding. For mobile model, the 2-hop relay scheme and the flooding scheme are proposed for both random i.i.d. mobility model and random walk model with random linear network coding. The throughput loss and decoding loss of network coding, which are treated as the overhead of network coding, are also considered. All the models and schemes above were widely adopted in the researches of wireless ad hoc networks [4]-[14], [20]- [23]. In particular, some of them (e.g. random i.i.d. mobility model) were good for preliminary research, others (e.g. random walk model) were of good reality. The main *differences* and *contributions* of this paper are summarized as follows:

- *Main differences*: All the following results are based on the scaling laws of network coding overhead which were not considered in most previous works in wireless ad hoc networks. Instead of throughput and delay/throughput tradeoff, this paper mainly focuses on the impact of network coding on goodput and delay/goodput tradeoff performance. Moreover, the corresponding optimal network coding configuration is also derived, which includes the generation size (k), network coding Galois field (u bits) and data size of one packet (B bits).
- *Main contribution I*: For static networks, our theoretical results indicate that the goodput and delay/goodput tradeoff cannot be improved in order sense by employing network coding when considering the throughput loss and decoding loss.
- *Main contribution II*: For 2-hop relay scheme in random i.i.d. mobility model, no goodput gain can be obtained comparing with no replicas case. However, compared to the replicas case, the goodput improvement is $\Theta(\sqrt{n})$ when $k = \Theta(n)$, $u = \Theta(\log n)$ and $B = \Theta(n \log n)$. However, the delay/goodput tradeoff cannot be strengthened in order sense when network coding is employed. For the flooding scheme, goodput gain $\Theta(\log n)$ can be achieved for the configuration $k = \Theta(\log n)$, $u = \Theta(\log n)$ and $B = \Theta(\log^2 n)$. Besides, the delay/goodput tradeoff improvement is $\Theta\left(\sqrt{\log n}\right)$ when $B = \Theta(\log n)$, $k = \Theta(\sqrt{\log n})$ and $u = \Theta(\sqrt{\log n})$.
- *Main contribution III*: For 2-hop relay scheme in random walk model, comparing with no replicas case, there is still no order gain on goodput. Nevertheless, significant goodput improvement $\tilde{\Theta}(n)$ is obtained when compared to the replicas case under the same network coding configuration as in 2-hop relay scheme of random i.i.d. mobility model. Furthermore, for the flooding scheme, goodput gain $\Theta(\sqrt{n})$ can be achieved for the configuration $k = \Theta(\sqrt{n})$, $u = \Theta(\log n)$ and $B = \Theta(\sqrt{n} \log n)$. Finally, it is proved that the network coding cannot improve the order of delay/goodput tradeoff in random walk model for both of the two schemes.

The rest of the paper is organized as follows. In Section II, we describe the characteristics of network coding and network performance metrics. Afterwards, the static and mobile networks with network coding are analyzed in Section III and IV, respectively. In Section V, we discuss the results and obtain the configuration of network coding to optimize the goodput and delay/goodput tradeoff. Finally, this paper is concluded in Section VI.

II. Some Characteristics of Network Coding and Network Performance Metrics

In this section, we present the basic idea of network coding as well as the scaling laws of throughput loss and decoding loss. Furthermore, some useful concepts and parameters are listed. Finally, we give the definitions of some network performance metrics.

A. Basic Idea of Network Coding

We employ the physical layer network coding (PNC) in static networks and random linear network coding in mobile networks.

Physical layer network coding scheme: PNC has been studied in [11] and [24], which is designed based on the *Channel State Information* (CSI) and network topology. The PNC is appropriate for the static networks since the CSI and network topology are pre-known in static case. Furthermore, we will give an example to show the feature of PNC as follow.

There are G nodes in one cell, and node $i (i = 1, 2, \cdots, G)$ holds packet x_i. All of the G packets are independent, and they belong to the same unicast session. The packets are transmitted to a node $i^{'}$ in the next cell simultaneously. $g_{ii'}$ is a complex number which represents the CSI between i and $i^{'}$ in frequency domain. The received signal can be expressed as

$$y_{i'} = \Sigma_{i=1}^{G} g_{ii'} s_i + n_{i'}, \tag{1}$$

where s_i is the modulated x_i and $n_{i'}$ is the noise. To employ network coding, node $i^{'}$ needs to generate a new packet which is the linear combination of x_i as follow

$$x_{i'} = \left(\Sigma_{i=1}^{G} \alpha_i x_i\right) mod(2^B), \tag{2}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

where α_i is the network coding coefficient (fixed) selected from the Galois field $\mathbb{F}_q$, and B is the size of data in one packet. The operation $mod(2^B)$ is introduced to guarantee that the size of the data combination part does not increase, and we will not show it in the following part of this paper for brief. Since the $g_{ii'}$ is independent from each other, node $i^{'}$ can detect $x_{i'}$ based on $y_{i'}$ according to the maximum-likelihood criterion, and the rate of successful decoding is $\Theta(1)$ when $G = \Theta(1)$. The detailed PNC technique can be found in [11] and [24].

Moreover, in order to guarantee that the data can be uniquely decoded, it must be satisfied that $B \geq u$, where u is the size of network coefficient in bits (i.e., $u = \log q$). Furthermore, according to the theoretical analysis in Section V, it is obvious that $B \geq u$ can be satisfied under the optimal configuration.

Random linear network coding: Different from the static networks, the CSI and network topology are dynamic and hard to be obtained in each time-slot. Therefore, RLNC is adopted in mobile networks instead of PNC. For example, if node j has received G packets $X_i (i = 1, 2, \cdots, G)$ which belong to the same source (may be received from different nodes), it will generate a new packet

$$Y = \left[Overhead, \left(\Sigma_{i=1}^{G+1} \alpha_i X_i\right) mod(2^B)\right], \tag{3}$$

where X_{G+1} is the packet that has already been held by node j in its memory. The overhead will be introduced in the next subsection, and the network coding coefficient α_i in (3) is no longer fixed but randomly selected from the Galois field $\mathbb{F}_q$. Moreover, in this paper, G is assumed to be 1 for mobile networks since the node j is hard to obtain multiple CSI.

B. The Overhead of Network Coding

In this paper, the throughput loss and decoding loss are jointly treated as the overhead of the network coding. We introduce both of them as follows.

Throughput loss of network coding: Considering the linear network coding of $\mathbb{F}_q$, a group of k packets $\{X_i\}$ are transmitted from the source to the destination with the help of relay(s). The data size of each packet is B bits. We also define that $u = \log q$. Thus, the coded packet, which combines all the k original packets, can be represented as

$$Y = [Y_\alpha, Y_c] = [\{\alpha_1, \cdots, \alpha_k\}, \sum_{i=1}^{k} \alpha_i X_i], \tag{4}$$

where α_i is the i-th network coding coefficient of u bits. Y_α records the coefficients and Y_c is the combination of data. The corresponding structure of the packet is illustrated in Fig. 1.

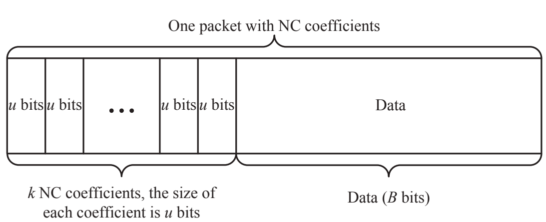


Fig. 1. The structure of packet with network coding coefficients and data

In order to guarantee the efficiency of each hop, the size of the NC coefficients size should be designed as small as possible. We define the B_{total} as the size of aggregate packet, which includes both data and the NC coefficients. The throughput loss is defined as follow:

$$C(n) = \frac{B_{total}}{B}. \tag{5}$$

In Fig. 1, the front ku bits are reserved for the network coding coefficients, and therefore the size of each packet is deterministic, i.e., $B_{total} = uk+B$ bits. Hence, the throughput loss can be represented as

$$C(n) = \Theta\left(\frac{ku}{B} + 1\right). \tag{6}$$

It should be noted that the throughput loss does not change during the transmission since the network coding operations are performed in the Galois field $\mathbb{F}_q$.

The throughput loss in ad hoc networks differs from that in some of the other networks (e.g. butterfly model [2]) in which the k and u can be negligible compared to the size of data. However, in ad hoc networks, for the purpose of order gain on goodput and delay, the number of combined packets must be sufficiently large, which makes the throughput loss an ineligible factor. Although we can increase the order of data size to reduce the throughput loss, this will also cause significant loss on delay since the packets cannot be decoded until $\Theta(k)$ different packets are received.

Decoding loss of network coding: The decoding loss is caused by decoding failure of RLNC. Since the network coding coefficients are randomly selected from Galois field $\mathbb{F}_q$, the destination may not decode the k original packets successfully. Thus, we must consider the probability that destination cannot successfully decode the k packets, which is treated as the decoding loss in this paper. Moreover, the feature of decoding loss is introduced in [18] as follow.

Lemma 1: (Tracey Ho. *et al.* [18]) When random network coding is employed, the probability that the random network code is valid is at least $(1 - 1/q)^\eta$, where η is the number of links with associated random coefficients and q is the size of coding Galois field.

C. Some Useful Concepts and Parameters

We list some definitions which will be used in this paper as follows.

Useful link: In order to calculate the decoding loss, we define the useful link as the link with associated random coefficients, i.e., the link belonging to the path from the source to the destination. For example, in one hop, a node combines G received packets and the packet it already has with associated random coefficients. The number of useful links for this hop is $G+1$ if the new packet or its combinations with other packets will be received by the destination. Instead, if the new packet or its combination with other packets is not finally received by the destination, these links are not useful links.

Generation size of network coding: When network coding is employed, a group of packets are combined together according to network coding. The generation size of network coding is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

TABLE I
NOTATIONS AND DEFINITIONS

Notation	Definition
n	The total number of nodes in the network.
k	The number of packets in one network coding group. (Generation size)
B	The size of data in each packet in bits.
q	The size of the Galois field $\mathbb{F}_q$ for network coding coefficient.
u	q measured in bits.
W	The transmission bandwidth.
$T_g(n)$	The per-node throughput without considering the overhead of network coding.
$D_g(n)$	The average delay without considering the overhead of network coding.
$P(n)$	The probability of successful decoding.
$C(n)$	The per-hop throughput loss.
η	The average number of useful links in random i.i.d. mobility model.
ζ	The average number of useful links in random walk model.
$T(n)$	The per-node goodput when considering the overhead of network coding.
$D(n)$	The average delay when considering the overhead of network coding.

defined as the number of packets in the group, which is the same for each group.

In Table I, we list some essential notations and definitions that will be used later.

D. Network Performance Metrics

Definition of goodput: For a given scheme, we define the goodput as the maximum achievable data transmission rate. In *t* time-slots, we assume that there are $M(i,t)$ successfully decoded data bits transmitted from node *i* to its destination node *j*. Firstly, the *long term goodput* of this S-D pair is defined by $\lambda_i(n)$ as

$$\lambda_i(n) = \liminf_{t\to\infty} \frac{1}{t} M(i,t). \tag{7}$$

Then the *goodput* of this model for a given scheme is defined as $T(n)$, which is the maximum value satisfying

$$\lim_{n\to\infty} \mathbb{P}(\lambda_i(n) \geq T(n) \text{ for all } i) = 1. \tag{8}$$

It should be noted that the NC coefficients are not considered in $M(i,t)$ since no data is conveyed by them. Moreover, we study the random networks instead of arbitrary networks, and therefore the transport throughput, which is defined as the rate timing the distance, is not appropriate in our networks since it is just defined for arbitrary networks [16].

Definition of delay: For a given scheme, we define the delay as the average time from the time that the source begins to output a group (network coding group in Table I) of packets to the time that the corresponding destination successfully decodes the data conveyed by them. Moreover, for wireless networks, we assume that the operation time spent in coding/decoding is negligible compared to the transmission time.

Definition of delay/goodput tradeoff: The delay/goodput tradeoff is defined as follow

$$\frac{D(n)}{T(n)}, \tag{9}$$

which is important performance for a transmission scheme and widely studied in [5] [8].

III. Analysis for Static Networks

In this section, the static network model is introduced at first, which includes the network topology and transmission model. Moreover, we propose the transmission scheme for this model, and the corresponding goodput and delay are also analyzed based on the consideration of throughput loss and decoding loss.

A. Network Topology

We focus on the networks which consist of n randomly and evenly distributed static nodes in a unit square area. These nodes are randomly grouped into source-destination (S-D) pairs.

B. Transmission Model

In this paper, we adopt the protocol model [16], which is a simplified version of physical model since it ignores the long distance interference and transmission. Moreover, it is indicated in [16] that the physical model can be treated as the protocol model on scaling laws when the transmission is allowed if the *Signal to Interference Noise Ratio* (SINR) is larger than a given threshold. In this model, As introduced in [16], when node i transmits to a node j, denoting the distance between them as $d(i,j)$, the transmission is successfully if both of the following conditions are satisfied:

- $d(i,j) < r(n)$.
- $d(j,k) \geq (1+\Delta)r(n)$ for every other node k simultaneously transmitting over the same channel as i, where $\Delta > 0$ is a constant factor which depends on the acceptable interference.

We also set the finite bandwidth as *W* bits/sec. In this model, $r(n) = \Theta\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)$, which guarantees that the network is connected[16]. Furthermore, we divide the network into $\frac{4}{r^2(n)}$ square cells, and K^2-TDMA is employed, which allows each K^2 adjacent cells to be active with a round-robin fashion. K is constant and should satisfy the protocol model. Thus, K is defined as $K = \lceil 1 + (1+\Delta)\sqrt{2} \rceil$.

C. Transmission Scheme for Static Networks

In this model, three kinds of nodes are involved, i.e., source node, relay node and destination node. Each node in the network may act as one or some of the three roles. Moreover, the PNC is adopted in this scheme since the network topology is fixed. This can be realized and calculated in a computing center of this network.

Network coding based relay scheme:. In the static networks with PNC, G in (1) and (2) satisfies $G = \Theta(1)$, and the corresponding CSI is pre-known at each node before transmission.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

Since the network is so large that each node hardly knows and stores all of the network coding coefficients, we only focus on the case that each node does not know the network coding coefficients of others. Moreover, the case that each node knows all of the network coding coefficients will be discussed at the end of this section.

For the transmission scheme, the k packets (as a generation group) are transmitted in a digraph with the minimum total Euclidean distance between connected nodes. In the static networks, the 'when-to-stop' signal is transmitted by hand-shaking, and the whole unicast session will not stop until $(1+\varepsilon)k$ different packets arriving at destination node, where $\varepsilon > 0$ is a constant. There are three steps for the transmission scheme:

- *Step 1*: The source node combines the k original packets and generates $(1+\varepsilon)k$ packets according to network coding. Afterwards, it transmits the packets to $(1+\varepsilon)k$ nearest nodes (relays) as multi-unicast.
- *Step 2*: All the relay nodes in one cell are separated into some groups, and each group includes G nodes. The nodes belonging to the same group transmit packets to the next cell simultaneously. Afterwards, the nodes in the next cell employ PNC which has been introduced above. A simple example is illustrated in Fig. 2. The step 2 is finished when all the packets are transmitted to the nearest cells around the destination cell.
- *Step 3*: All the packets in the nearest cells around the destination cell will be transmitted to the destination node as 'many-to-one' transmission, which is called 'converge-cast'.

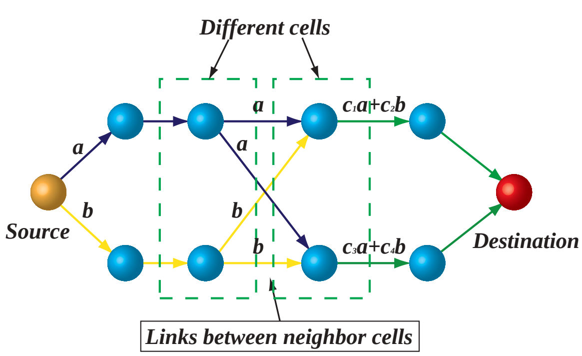


Fig. 2. Different links of one session can be active between neighbor cells simultaneously without interference

Remark 1: As described in [16], $\Theta\left(\frac{\sqrt{n}}{\sqrt{\log n}}\right)$ nodes are needed to transmit one packet and there are n nodes in the network. Therefore, we assume $k = O(\sqrt{n\log n})$ in our paper. For the case $k = \omega(\sqrt{n\log n})$, we can divide the k packets into $\frac{k}{\sqrt{n\log n}}$ groups to complete the transmission.

D. The Performance of Static Networks

The decoding loss is derived in the following Lemma 2. Moreover, the throughput and delay without considering the throughput loss and decoding loss are derived in Lemma 3. The throughput in Lemma 3 is the average number of packets that can be sent from the source to the destination in one time-slot, which is different from the goodput. In fact, these packets may include the network coding coefficients and linearly correlated packets. Therefore, it is the gross throughput, which is greater than the data goodput $T(n)$ in (8). Similarly, the delay in Lemma 3 does not consider the transmission of network coding coefficients and linearly correlated packets, and therefore it is smaller than the real delay $D(n)$ defined in Section II.D.

Lemma 2: Considering a unicast static networks with PNC, there must be a solution of network code in field $\mathbb{F}_q$ to ensure the network is feasible for any constant q and $q > 1$.

The proof can be found in [18].

Lemma 3: If the throughput loss and decoding loss are not considered, the gross throughput is $T_g(n) = \Theta\left(\frac{\min\{k,G\}W}{\sqrt{n}\log n}\right)$ bits/sec and the corresponding delay is $D_g(n) = \Theta\left(\frac{Bk\sqrt{n\log n}}{\min\{k,G\}W}\right)$ sec.

Proof: We assume that there are $(1+\varepsilon)k$ packets which are transmitted from one source, the total transmitted data bits for this session are $\Theta(kB)$ bits. If $k < \frac{G}{1+\varepsilon}$, all packets are transmitted through $\Theta\left(\frac{\sqrt{n}}{\sqrt{\log n}}\right)$ cells. The hops between neighbor cells ($(1+\varepsilon)k$ hops) can be active simultaneously, and it will take $\frac{B}{W}$ sec for them. The first and third steps are ignored here compared to the second step because k is too small. There are total $\Theta\left(\frac{kn\sqrt{n}}{\sqrt{\log n}}\right)$ hops in the network. Hence, by employing TDMA, the delay for this condition can be represented as

$$D_g(n) = \Theta\left(\frac{B}{W}\cdot\frac{kn\sqrt{n}}{\sqrt{\log n}}\cdot\frac{\log n}{kn}\right) = \Theta\left(\frac{B\sqrt{n\log n}}{W}\right) \text{ sec.} \tag{10}$$

Since there is no replica in order sense, the gross throughput for this condition is

$$T_g(n) = \Theta\left(\frac{kW}{\sqrt{n\log n}}\right) \text{ bits/sec.} \tag{11}$$

If $k \geq \frac{G}{1+\varepsilon}$ and $k = O(\sqrt{n\log n})$ as in Remark 1, the step 1 and 3 will take $\Theta\left(\frac{k^{\frac{3}{2}}}{\sqrt{\log n}}\right)$ hops. Thus, there are $\Theta\left(\frac{k^{\frac{3}{2}}}{\sqrt{\log n}} + \frac{k\sqrt{n}}{\sqrt{\log n}}\right)$ hops for one session. As discussed above, the delay for this condition is

$$\begin{aligned} D_g(n) &= \Theta\left(\frac{B}{W}\cdot\left[\frac{nk^{\frac{3}{2}}}{\sqrt{\log n}} + \frac{nk\sqrt{n}}{\sqrt{\log n}}\right]\cdot\frac{\log n}{Gn}\right) \\ &= \Theta\left(\frac{kB\sqrt{n\log n}}{GW}\right) \text{ sec.} \end{aligned} \tag{12}$$

And the gross throughput for this condition can be represented as

$$T_g(n) = \Theta\left(\frac{GW}{\sqrt{n\log n}}\right) \text{ bits/sec.} \tag{13}$$

■

Theorem 1: In random static networks, when considering throughput loss and decoding loss, the data goodput is provided in (14) and delay is provided in (15), where $k = O(\sqrt{n\log n})$ as in Remark 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

Proof: Firstly, we analyze the goodput performance. It can be derived based on $T_g(n)$, throughput loss $C(n)$ and probability of successful decoding $P(n)$. According to the scheme for static networks, there are k packets combined as a group, and therefore the throughput loss satisfies $C(n) = \Theta\left(\frac{ku}{B}+1\right)$. Moreover, according to Lemma 2, the size of field $\mathbb{F}_q$ can be constant, e.g., $u = 1$. Consequently, considering Lemma 2 and Lemma 3, the total goodput is

$$T(n) = \frac{T_g(n)P(n)}{C(n)} = \begin{cases} \Theta\left(\frac{kWB}{k\sqrt{n\log n}+B\sqrt{n\log n}}\right) \text{bits/sec} & \text{if } k < \frac{G}{1+\varepsilon}, \\ \Theta\left(\frac{GWB}{k\sqrt{n\log n}+B\sqrt{n\log n}}\right) \text{bits/sec} & \text{if } k \geq \frac{G}{1+\varepsilon}, \end{cases} \tag{14}$$

where $k = O(\sqrt{n\log n})$ as in Remark 1. The total delay is derived based on $D_g(n)$, $C(n)$ and $P(n)$, and it is represented as

$$D(n) = \frac{D_g(n)C(n)}{P(n)} = \begin{cases} \Theta\left(\frac{k\sqrt{n\log n}+B\sqrt{n\log n}}{W}\right) \text{sec} & \text{if } k < \frac{G}{1+\varepsilon}, \\ \Theta\left(\frac{k^2\sqrt{n\log n}+kB\sqrt{n\log n}}{GW}\right) \text{sec} & \text{if } k \geq \frac{G}{1+\varepsilon}, \end{cases} \tag{15}$$

where $k = O(\sqrt{n\log n})$ as in Remark 1. ■

Now we consider the case that each node knows other network coding coefficients. The Y_α in (4) becomes useless since it is known by the related nodes. Therefore, the transmitted packet can be represented as

$$Y = \sum_{i=1}^{k} \alpha_i X_i. \tag{16}$$

Thus, the size of Y is B bits, and there is no throughput loss. As a result, the goodput and delay for this condition is the same as in Lemma 3.

IV. Analysis for Mobile Networks

In this section, we will introduce the mobile network models, which include network topology, mobility models and transmission model. Furthermore, the transmission schemes are also proposed for both of the mobility models, and the corresponding goodput and delay are derived in the similar way as in Section III.

A. Network Topology

We study the mobile networks which consist of n mobile nodes in a unit square area. These nodes are randomly grouped into source-destination (S-D) pairs.

B. Mobility Models

In mobile networks, the total region (the unit square) is divided into $m = \Theta(n)$ square cells instead of $\Theta(\frac{n}{\log n})$, where $m < n$. Our work can be applied to many mobility models, and we mainly concentrate on the random i.i.d. mobility model and random walk model, which are defined as follows.

Random i.i.d. mobility model: In random i.i.d. mobility model, each node will be in a randomly selected cell independently and identically in the next time-slot, which means that each node will be in any cell with the same probability. As a result, the network topology changes drastically in every time-slot, and the network behavior cannot be predicted. This model is also adopted in[5], [20].

Random walk model: In random walk model, if a node is in cell $(i, j) \in \{1, 2, ..., \sqrt{m}\}^2$ at the current time-slot, it will be in the five cells: (i, j), $(i+1, j)$, $(i-1, j)$, $(i, j+1)$, $(i, j-1)$ with the same probability in the next time-slot. In this model, the speed of each node is limited, and the model can therefore better describe the realistic node mobility than random i.i.d. mobility model. Previous literatures adopting the random walk mobility model include [5], [20] and [21].

C. Transmission Model

We also adopt the protocol model and the K^2-TDMA scheme for mobile networks. Moreover, the transmission range is represented as $r(n) = \frac{\sqrt{c_0}}{\sqrt{n}}$ where c_0 is constant and $nr^2(n) > 2$. Therefore, for an arbitrary cell s, the number of nodes in it is

$$N(s) = \sum_{i=1}^{n} \mathbf{1}_{i \text{ is in cell } s}, \tag{17}$$

where $\mathbf{1}_{i \text{ is in cell } s}$ are independent i.i.d. Bernoullian random variables. Moreover, the mean of $\mathbf{1}_{i \text{ is in cell } s}$ is $\frac{c_0}{n}$. Therefore, according to the Chernoff bound, if $c_0 > 12\ln 4$, the $N(s)$ satisfies

$$\begin{aligned} &Pr\left\{\frac{c_0}{2} \leq N(s) \leq \frac{3c_0}{2}\right\} \\ &= 1 - Pr\left\{N(s) < \frac{c_0}{2}\right\} - Pr\left\{N(s) > \frac{3c_0}{2}\right\} \\ &= 1 - e^{-\frac{c_0}{8}} - e^{-\frac{c_0}{12}} > \frac{1}{2}, \end{aligned} \tag{18}$$

is constant. Thus, there are constant number of nodes in one cell with constant probability. In addition, the bandwidth is also assumed to be *W* bits/sec.

D. Transmission Schemes for Mobile Networks

The following schemes are applicable to both random i.i.d. mobility model and random walk model. Firstly, we define 3 kinds of transmissions: source to relay (S-R), relay to relay (R-R) and relay to destination (R-D). In addition, source to destination transmission also belongs to R-D. When the relay receives a new packet, it combines the packet it has with the one it receives by randomly selected coefficients, and then generates a new packet. Simultaneous transmission in one cell is not allowed since the receiver is hard to obtain multiple CSIs from different transmitters at the same time. Hence, we employ the random linear network coding for mobile models. Specifically, we adopt two schemes:

- *2-hop relay scheme*, R-R transmission is not allowed in this scheme. All the packets are transmitted from source to destination by at most 2 hops. The probability that either S-R or R-D is selected is $1/2$. Each packet will be deleted t_d seconds after its generation, where $t_d = \Omega(D(n))$ is decided based on the delay of the networks. When S-R transmission is selected, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

source will randomly select a node in the same cell and transmit a combined packet of k original packets with coefficients which are randomly selected from $\mathbb{F}_q$. When R-D transmission is selected, the relay will transmit the corresponding packet to the destination and then delete this packet. The destination decodes the network coding when it receives $(1+\varepsilon)k$ different packets, where ε is a positive constant.

- *Flooding scheme*, All the three transmissions are allowed, and the probability that one of them is selected is $1/3$. The delete time t_d for this scheme is also $t_d = \Omega(D(n))$. The R-D transmission is the same as in 2-hop relay scheme. When S-R transmission is selected, the source will transmit a combined packet of k original packets with network coding coefficients to all the nodes within the same cell. When R-R transmission is selected, one node will be randomly selected, and it will transmit one of its packets to the other nodes in the same cell equiprobably. After receiving the packet, each node in this cell combines the packet with the same session packet it has (if there is one) as in (3), where $G = 1$. Afterwards, it will store the combination in its memory and delete the received packet as well as the old packet of this session. The decoding time of destination is the same as in 2-hop scheme.

Remark 2: The k of 2-hop relay scheme satisfies $k = O(n)$ because there is no gain when $k = \omega(n)$ and we can separate k packets into $\frac{k}{n}$ groups to change the problem into the condition that $k = O(n)$.

Remark 3: The k of flooding scheme satisfies $k = O(\log n)$ in random i.i.d. mobility model and $k = O(\sqrt{n})$ in random walk model, because there is no gain when k is greater. Therefore, we can separate k packets into $\frac{k}{\log n}$ groups in random i.i.d. mobility model and $\frac{k}{\sqrt{n}}$ groups in random walk model.

In order to explain the Remark 3, we firstly focus on the flooding scheme for random i.i.d. mobility model. The delay of flooding scheme for random i.i.d. mobility model is proved to be $\Theta(\log n)$ [5] without network coding. Moreover, the delay will be greater than $\Theta(\log n)$ if $k = \omega(\log n)$ when network coding is adopted, i.e., the delay becomes $\Theta(k)$, which is shown in the proof of Theorem 5 in [5]. Based on this proof, the destination will receive k combined packets in $\Theta(k)$ time-slots for each $\frac{1}{n}$ phases. Therefore, the gross throughput does not increase with k, however, the corresponding delay grows with k. Thus, it is meaningless to analyze the case $k = \omega(\log n)$ for flooding scheme in random i.i.d. mobility model. Similarly, since the delay of flooding scheme in random walk model is $\Theta(\sqrt{n})$ [5], we do not focus on the case $k = \omega(\sqrt{n})$ for the same reason.

E. The Performance of Random I.I.D. Mobility Model

Firstly, we give the following lemma which will be utilized in our analysis.

Lemma 4: Assuming that there are *m* cells in the networks, *b* nodes are randomly located in them. The expectation of the number of cells which hold at least 1 node is given as follows:

$$\mathbb{E}(N) = \begin{cases} b - o(b) & \text{if } b = o(m), \\ \Theta(m) & \text{if } b = \Omega(m). \end{cases} \tag{19}$$

Proof: The result for the condition $b = \Omega(m)$ can be found in Lemma 4.6 in [27]. For the case $b = o(m)$, the number of cells that there are more than one node in it is

$$\begin{aligned} & m\left(1-\left(1-\frac{1}{m}\right)^b-\frac{b}{m}\left(1-\frac{1}{m}\right)^{b-1}\right) \\ =&\Theta\left(m\left(1-1+\frac{b}{m}-\frac{b}{m}-\frac{b^2}{m^2}\right)\right) \\ =&\Theta\left(\frac{b^2}{m}\right) \\ =&o(b). \end{aligned} \tag{20}$$

Thus, $\mathbb{E}(N) = b - o(b)$ for this case. ■

Afterwards, we analyze the goodput and delay performance with the consideration of throughput loss and decoding loss for random i.i.d. mobility model.

Lemma 5: Considering one unicast session in the networks with random i.i.d. mobility model and network coding, we denote the number of links with associated random coefficients as η. The η_{2-hop} is of order $\Theta(k)$ for 2-hop relay scheme, and $\eta_{flooding} = O(\min\{2^{(1+\varepsilon)k}\log n, n\log n\})$ for flooding scheme.

Proof: Firstly, we consider the 2-hop relay scheme. The destination receives $(1+\varepsilon)k$ combined packets from $(1+\varepsilon)k$ relays (may include the source). Hence, the unicast session is composed by source, destination and these relays. We ignore the other relays because the destination does not receive any packet from them. Thus, the number of links with associated random coefficients is $\Theta(k)$ for both the first the second hops. As a result, we have $\eta_{2-hop} = \Theta(k)$.

Then we focus on the flooding scheme. For this scheme, packet will arrive at the destination in $\Theta(\log n)$ hops in average [5]. Considering one unicast session, it can be divided into three steps: the first step is from the beginning to the time that there are $\frac{n}{\log\log n}$ nodes holding packets of this session; the second step is from the end of the first step to the time that there are $\frac{1}{r^2(n)}$ nodes holding packets; the rest of the session is the third step. The numbers of useful links for the three steps are η_1, η_2 and η_3, respectively.

For the first step, according to Lemma 4, two nodes who hold packet will meet each other with probability $o(1)$. Moreover, when a packet is transmitted to a node that does not hold packet, the number of useful link is 1 for this hop. Thus, the number of useful link for each hop in the first step is 1 with probability 1. Besides, it must be noticed that some of the packets held by relays at the end of the first step will not be transmitted to the destination. For example, if a packet x_i is hold by node i at the end of the first step, it may be combined with other packets in the second or third steps. However, it is possible that none of its combinations are transmitted to the destination when the destination receives $(1+\varepsilon)k$ packets. Therefore, such kind of packets and nodes will not be considered when calculating η_1 and η_2. We define $\varphi_j (j = 1, 2)$ as the number of nodes who satisfy: 1) the node holds packet of this session at the end of step j, 2) the packet

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

or its combinations will be transmitted to the destination in the third step. Hence, the η_1 can be bounded by $\eta_1 = O(\varphi_1 \log n)$ because there are $O(\log n)$ time-slots for step 1.

Afterwards, for the second step, since there may be $\Theta(n)$ nodes holding packets, the probability that two nodes (who hold packet) meet each other may be $\Theta(1)$. In our scheme, when a node receives a packet and it already has one packet from the same source, it will generate a new packet by combing them together. Therefore, there are 2 useful links in this hop, and the η_2 can be bounded by $\eta_2 = O(\varphi_2 2^{\log \log n})$ because there are $O(\log \log n)$ time-slots for this step [5].

At last, for the third step, since there are $\frac{1}{r^2(n)} = \Theta(n)$ nodes holding packets, the destination will receive one packet in constant time-slots. Thus, the third step lasts for $\Theta(k)$ time-slots, and each hop has at most 2 useful links. Considering the i-th received packet by the destination, we denote $L(i)$ as the number of links (step 3 only) it transmitted in. Therefore, η_3 must satisfy $L((1+\varepsilon)k) < \eta_3 < \sum_{j=1}^{(1+\varepsilon)k} L(j)$. Then we derive the $L(i)$. Since the destination receives the i-th received packet at time-slot $\Theta(i)$ from the beginning of the third step, the $L(i)$ can be bounded by $L(i) = O(2^{ci})$ where c is a constant. Therefore the upper bound of η_3 is as follow

$$\eta_3^{upper-bound} = \sum_{j=1}^{(1+\varepsilon)k} L(j) = \Theta(L((1+\varepsilon)k)). \tag{21}$$

Thus, $\eta_3 = O(L((1+\varepsilon)k)) = O(2^{c(1+\varepsilon)k})$ and $\varphi_2 = O(2^{c(1+\varepsilon)k})$.

Consequently, notice that φ_1, φ_2 are no greater than n and $\varphi_1 = O(\varphi_2 2^{\log \log n})$, the total number of useful links for flooding scheme is upper-bounded as

$$\eta_{flooding} = \eta_1 + \eta_2 + \eta_3 = O(\min\{2^{c(1+\varepsilon)k} \log^2 n, n \log n\}). \tag{22}$$

■

In the following lemma, we derive the gross throughput and delay, which do not consider the impact of the extra bits of network coding coefficients and linearly correlated packets.

Lemma 6: In random i.i.d. mobility model, if the throughput loss and decoding loss are not considered, the gross throughput of 2-hop relay scheme in mobile networks with network coding is $T_g(n) = \Theta\left(\frac{W\sqrt{k}}{\sqrt{n}}\right)$ bits/sec and the corresponding delay is $D_g(n) = \Theta\left(\frac{B\sqrt{kn}}{W}\right)$ sec. The gross throughput of flooding scheme in mobile networks with network coding is $T_g(n) = \Theta\left(\frac{Wk}{n \log n}\right)$ bits/sec and the corresponding delay is $D_g(n) = \Theta\left(\frac{B \log n}{W}\right)$ sec.

Proof: For 2-hop relay scheme, the destination starts to receive packets and the source starts to transmit packets at the beginning. We assume that a unicast session is finished after transmitting t time-slots (which means that the transmission for this session is active in the t time-slots). If there are N nodes holding packets of this unicast session, the destination will receive a packet with probability $\frac{N}{n}$. Therefore, the expectation of the total number of received packets is

$$\Theta\left(\sum_{N=1}^{t} \frac{N}{n}\right) = \Theta(\frac{t^2}{n}), \tag{23}$$

which equals to $\Theta(k)$ because the destination needs $(1+\varepsilon)k$ packets to decode. Hence, $t = \Theta(\sqrt{kn})$, and the transmission is finished when $\Theta(\sqrt{kn})$ nodes hold packets for one unicast session. For each session there are $\Theta(\sqrt{kn})$ hops, and there are n unicast sessions. Thus, by employing TDMA, the delay for 2-hop relay scheme is $\Theta\left(\frac{B\sqrt{kn}}{W}\right)$ sec. Moreover, the replicas in this network are of order $\Theta\left(\frac{\sqrt{nk}}{k}\right) = \Theta\left(\sqrt{\frac{n}{k}}\right)$. Hence, the throughput for 2-hop relay scheme is $\Theta\left(\frac{W\sqrt{k}}{\sqrt{n}}\right)$ bits/sec.

Afterwards, we analyze the flooding scheme. If there are N nodes holding different combined packets, the destination will take at least $\Theta\left(\log \frac{n}{N}\right)$ time-slots in average to receive one packet. Afterwards, there will be $\Theta\left(NC^{\log \frac{n}{N}}\right) = \Theta(n)$ nodes holding packets of this session. C here is the average number of nodes in one cell. As a result, it will take $\Theta(\log n)$ time-slots to transmit packets to n nodes and $\Theta(k)$ time-slots to receive $(1+\varepsilon)k$ packets from them. Since $k = O(\log n)$, by employing TDMA, the delay for flooding scheme is $D(n) = \Theta\left(\frac{B \log n}{W}\right)$ sec. Since $(1+\varepsilon)k$ packets are received in $\Theta(\log n)$ time-slots and this happens once for $\Theta\left(\frac{1}{n}\right)$ phase, the throughput is $\Theta\left(\frac{Wk}{n \log n}\right)$ bits/sec. ■

According to above results, we can obtain the goodput $T(n)$ and corresponding delay $D(n)$ for i.i.d. mobility model, which are demonstrated in the following theorem.

Theorem 2: In random mobile networks with random i.i.d. mobility model, when considering throughput loss and decoding loss, the goodput of 2-hop relay scheme is as in (24) and delay is as in (25), where $k = O(n)$ as in Remark 2. The goodput of flooding scheme is as in (26) and delay is as in (27), where $k = O(\log n)$ as in Remark 3.

Proof: The probability of successful decoding is shown in Lemma 1 for mobile networks. We prove the theorem in the same way as in Theorem 1. Since k packets in one group will be combined together, the throughput loss is $\Theta\left(\frac{ku}{B} + 1\right)$. Thus, in 2-hop relay scheme, the total goodput is

$$T(n) = \Theta\left((1 - 1/q)^{\gamma k} \frac{WB\sqrt{k}}{u\sqrt{n}k + B\sqrt{n}}\right) \text{bits/sec}, \tag{24}$$

and the total delay is

$$D(n) = \Theta\left((1 - 1/q)^{-\gamma k} \frac{uk\sqrt{kn} + B\sqrt{kn}}{W}\right) \text{sec}, \tag{25}$$

where $k = O(n)$ and γ is a constant. In flooding scheme, the total goodput is

$$T(n) = \Theta\left((1 - \frac{1}{q})^{\eta_{flooding}} \frac{WBk}{nku \log n + Bn \log n}\right) \text{bits/sec}, \tag{26}$$

and the total delay is

$$D(n) = \Theta\left((1 - \frac{1}{q})^{-\eta_{flooding}} \frac{ku \log n + B \log n}{W}\right) \text{sec}, \tag{27}$$

where $\eta_{flooding} = O(\min\{2^{(1+\varepsilon)k} \log^2 n, n \log n\})$, $k = O(\log n)$. ■

F. The Performance of Random Walk Model

The performance of the random walk model can be derived in the similar way as in i.i.d. mobility model. Firstly, we give

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

the following lemma which will be utilized in our analysis.

Lemma 7: In the random walk model, for source i, if it has been active for m_0 time-slots, the expectation of the number of nodes that meet source i is of order $\Theta(m_0)$, where $m_0 \leq n$.

Proof: We define $N(i, j, m_0)$ as the number of times that node i meets j within m_0 time-slots. Therefore, the number of nodes that meet source i at time-slot m_0 equals to $N = \sum_{j=1}^{n} \min\{N(i, j, m_0), 1\}$. According to [5], N satisfies

$$N = \frac{m_0 n}{\tau}, \tag{28}$$

where τ is the random variable representing the inter-meeting time between i and j, which satisfies $\mathbb{E}\{\tau\} = \Theta(n)$ and $Var\{\tau\} = \Theta(n^2 \log n)$ [5]. If $m_0 \leq n$, the mean of N can be lower-bounded as

$$\begin{aligned} \mathbb{E}\left\{\sum_{j=1}^{n} N(i, j, m_0)\right\} &= \mathbb{E}\left\{\frac{nm_0}{\tau}\right\} \\ &\geq \frac{nm_0}{\mathbb{E}\{\tau\}} \\ &= \Theta(m_0). \end{aligned} \tag{29}$$

Moreover, since $N \leq m_0$, it is obvious that $\mathbb{E}\{N\} = \Theta(m_0)$. Furthermore, recall that for a random variable X, we have $Var\{f(X)\} \approx (f'(\mathbb{E}\{X\}))^2 Var\{X\}$, and therefore $Var\{N\}$ can be further derived as

$$\begin{aligned} Var\{N\} &= Var\left\{\frac{nm_0}{\tau}\right\} \\ &= \Theta\left(\frac{nm_0^2 Var\{\tau\}}{\mathbb{E}^4\{\tau\}}\right) \\ &= \Theta\left(\frac{m_0^2 \log n}{n}\right). \end{aligned} \tag{30}$$

Thus, according to the Chebyshev inequality, for any constant $0 < c < 1$,

$$\Pr\{N < (1-c)m_0\} \leq \frac{Var\{N\}}{c^2 \mathbb{E}^2\{N\}} = \Theta\left(\frac{\log n}{n}\right) \to 0, \tag{31}$$

which proves this lemma. ∎

Afterwards, we analyze the performance for random walk model.

Lemma 8: Considering one unicast session in the network with random walk model and network coding, we represent the number of links with associated random coefficients as ζ. The ζ_{2-hop} is of order $\Theta(k)$ for 2-hop relay scheme, and $\zeta_{flooding} = O(kn)$ for flooding scheme.

Proof: Firstly, we consider the 2-hop relay scheme. Similar to the random i.i.d. mobility model, the destination receives $(1+\varepsilon)k$ packets from $(1+\varepsilon)k$ relays (may include source). The number of the useful links for S-R and R-D are both $(1+\varepsilon)k$. Thus, $\zeta_{2-hop} = \Theta(k)$ for 2-hop relay scheme.

Afterwards, we consider the flooding scheme. The unicast session is divided into two parts. Considering a sub-network of size $(1+\varepsilon)k \times (1+\varepsilon)k$ cells centered at destination, the first part is from the beginning to the time when there is at least one node holding packet in each cell in this sub-network. The second part is the rest of the session. We define the number of useful links for the two parts as ζ_1 and ζ_2, respectively.

For ζ_2, the i-th $(1 \leq i \leq (1+\varepsilon)k)$ received packet is related with at most $i \times i$ cells around destination. Therefore, ζ_2 can be bounded by

$$\zeta_2 = O\left(\sum_{j=1}^{(1+\varepsilon)k-1} j^2\right) = O(k^3). \tag{32}$$

For ζ_1, we consider two ranges. One is the source range, which is defined as the sub-network of size $\Theta(j) \times \Theta(j)$ at time-slot j around the source node. Another range is the destination range, which is defined as the sub-network of size $\Theta(t_{\zeta_1}+(1+\varepsilon)k-j) \times \Theta(t_{\zeta_1}+(1+\varepsilon)k-j)$ at time-slot j around the destination node, where $t_{\zeta_1} = \Theta(\sqrt{n})$ is the beginning time of part 2. The source range represents the region that the source packets can arrive at time-slot j. Moreover, the destination range represents the range of the packets may be received by the destination after j time-slots. Thus, the overlap of these two ranges is where the useful links in at time-slot j. Hence, the ζ_1 is bounded by the summation of links in the overlap of these two ranges from time-slot 1 to t_{ζ_1}, which can be calculated as

$$\zeta_1 = O\left(\sum_{j=1}^{t_{\zeta_1}} \min\{(j, t_{\zeta_1} + (1+\varepsilon)k - j\}(1+\varepsilon)k\right) = O(kn). \tag{33}$$

Consequently, the total number of useful links for the flooding scheme of random walk model can be upper-bounded as

$$\zeta_{flooding} = \zeta_1 + \zeta_2 = O(kn). \tag{34}$$

∎

We analyze the gross throughput and the corresponding delay in the following lemma, which do not consider the impact of the extra bits of network coding coefficients and linearly correlated packets.

Lemma 9: In random walk model, if the throughput loss and decoding loss are not considered, the gross throughput of 2-hop relay scheme with network coding is $T_g(n) = \tilde{\Theta}\left(\frac{Wk}{n}\right)$ bits/sec and the corresponding delay is $D_g(n) = \tilde{\Theta}\left(\frac{Bn}{W}\right)$ sec. The gross throughput of flooding scheme in mobile networks with network coding is $T_g(n) = \Theta\left(\frac{Wk}{n\sqrt{n}}\right)$ bits/sec and the corresponding delay is $D_g(n) = \Theta\left(\frac{B\sqrt{n}}{W}\right)$ sec.

Proof: Firstly, we consider the 2-hop relay scheme in random walk model. It is proved in [28] that if there is an region of size $c_n \times c_n$ centered at the source and there are k_n nodes starting at the source at time-slot 0, the probability of the event that one or more of the k_n nodes ever exit the area is $o(1)$ in t_0 time-slots when $\frac{c_n^2}{t_n} \geq \frac{8 \log n}{n}$ in this model. We assume that c_n is half of the distance between source and destination, which is a constant in average. Therefore, if $t_n \leq \frac{c_n^2 n}{8 \log n}$, the nodes which hold source i's packets can meet the destination with probability 0 when n goes to infinity. Consequently, the delay satisfies $D_g(n) = \Omega\left(\frac{Bn}{\log nW}\right)$ sec. On the other hand, $D_g(n) = O\left(\frac{Bn \log n}{W}\right)$ sec if $k = O(n)$ because $\Theta(n)$ nodes will hold the packets from the source after $\Theta(n \log n)$ time-slots [5] and the destination will receive them in $O(n \log n)$ time-slots. Therefore, we obtain that $D_g(n) = \tilde{\Theta}\left(\frac{Bn}{W}\right)$ sec.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

For the throughput of 2-hop scheme, the replicas must be calculated. Based on Lemma 7, the number of nodes which hold packet from the source is $\Theta(D_g(n))$. However, the destination only receives $(1+\varepsilon)k$ packets, which means that the replicas are $\Theta\left(\frac{D_g(n)}{k}\right)$. Consequently, the network transmits kn packets by $\Theta\left(D_g(n)n\right)$ hops, and the per-node throughput of the network is $T_g(n)=\tilde{\Theta}\left(\frac{Wk}{n}\right)$ bits/sec.

For the flooding case, owing to the low speed of each node, no packet will be received by the destination node within $\Theta(\sqrt{n})$ time-slots. Moreover, it is proved that there will be at least one node holding packet in each cell at time-slots $\Theta(\sqrt{n})$ [5]. Therefore, the delay for it is $D_g(n) = \Theta\left(\frac{B(\sqrt{n}+k)}{W}\right)=\Theta\left(\frac{B\sqrt{n}}{W}\right)$ sec according to Lemma 7. The transmission happens once for $\Theta\left(\frac{1}{n}\right)$ phases. Therefore, the throughput for flooding scheme is $T_g(n)\Theta\left(\frac{Wk}{n\sqrt{n}}\right)$ bits/sec. ■

According to above results for random walk mobility model, the goodput $T(n)$ and the corresponding delay can be derived in the following theorem.

Theorem 3: In random mobile networks with random walk model, when considering throughput loss and decoding loss, the goodput of 2-hop relay scheme is as in (35) and delay is as in (36), where $k=O(n)$ as in Remark 2. The goodput of flooding scheme is as in (37) and delay is as in (38), where $k=O(\sqrt{n})$ as in Remark 3

Proof: The probability of successfully decoding is shown in Lemma 1 for mobile networks. We prove the theorem in the same way as in Theorem 1. Moreover, since k packets are combined as a network coding group, the throughput loss is $\Theta\left(\frac{ku}{B}+1\right)$. Hence, in 2-hop relay scheme, the total goodput is

$$T(n)=\tilde{\Theta}\left((1-1/q)^{\varrho k}\frac{WBk}{(uk+B)n}\right)\text{bits/sec,} \tag{35}$$

and the total delay is

$$D(n)=\tilde{\Theta}\left((1-1/q)^{-\varrho k}\frac{(uk+B)n}{W}\right)\text{sec,} \tag{36}$$

where $k=O(n)$ and ϱ is a constant. In flooding scheme, the total goodput is

$$T(n)=\Theta\left((1-\frac{1}{q})^{\zeta_{flooding}}\frac{WBk}{kun\sqrt{n}+Bn\sqrt{n}}\right)\text{bits/sec,} \tag{37}$$

and the total delay is

$$D(n)=\Theta\left((1-\frac{1}{q})^{-\zeta_{flooding}}\frac{ku\sqrt{n}+B\sqrt{n}}{W}\right)\text{sec,} \tag{38}$$

where $k=O(\sqrt{n})$ and $\zeta_{flooding}=O(kn)$. ■

V. Discussions

In this session, we discuss the results above and optimize the delay/goodput tradeoff and goodput for both static and mobile networks. The corresponding optimal data size *B*, generation size *k* and network coding field $\mathbb{F}_q$ are also derived. Besides, we compare the results with no NC case. The data size for network without NC is shown in Remark 4. Since there is no network coding in this case, the throughput of it can be treated as the goodput. Additionally, it should be noted that the units for *B*, $T(n)$, $D(n)$ and the tradeoff are *bits*, *bits/sec*, *sec* and *sec*^{2}*/bits*, respectively. For the sake of brevity, we do not list the units in the results during our discussion.

Remark 4: In the network without network coding, the increment of data size will lead to an enlargement of delay, while the goodput remains the same. Consequently, data size should be as small as possible. The smallest data size is $B=\Theta(1)$ in static networks, because all the transmission paths are fixed. In mobile networks, the destination ID must be conveyed in the packet. As a result, the data size is at least $B=\Theta(\log n)$.

A. Discussion on Static Networks

When the nodes know all of the related network coding coefficients in static networks, there is nether throughput loss nor decoding loss, and the goodput gain and delay/goodput tradeoff improvement are the same, i.e., $\Theta(\min\{k,G\})$. However, it will take too much memory to record all the related coefficients. Hence, we do not discuss this case and only concentrate on the case that nodes only know its own coefficients. The delay/goodput tradeoff of static networks is obtained from (14) and (15) as

$$\begin{cases}\Theta\left(\frac{\left(k\sqrt{n\log n}+B\sqrt{n\log n}\right)^2}{kBW^2}\right) & \text{if } k<\frac{G}{1+\varepsilon},\\ \Theta\left(\frac{k\left(k\sqrt{n\log n}+B\sqrt{n\log n}\right)^2}{BW^2G^2}\right) & \text{if } k\geq\frac{G}{1+\varepsilon}.\end{cases} \tag{39}$$

It is easy to prove that if $k=1$, the tradeoff is only $\Theta\left(\frac{n\log n}{W^2}\right)$ when $B=\Theta(1)$, which is the optimal case. Thus, the delay/goodput tradeoff cannot be improved by employing network coding. The reason is that when k is greater, the throughput loss of network coding becomes larger, and therefore the delay and tradeoff grow too. Moreover, if we optimize the goodput, it will be $\Theta\left(\frac{WG}{\sqrt{n\log n}}\right)$ when $B=\Theta(G)$ and $k=\Theta(G)$. Due to simultaneously transmission, it is reasonable that there is the gain of $\Theta(G)$ comparing with the tradeoff of no network coding case, i.e., $\Theta\left(\frac{W}{\sqrt{n\log n}}\right)$. Hence, there is only constant gain when $G=\Theta(1)$.

B. Discussion on Mobile Networks with Random I.I.D. Mobility Model

For random i.i.d. mobility model, we firstly consider the 2-hop relay scheme. The corresponding delay/goodput tradeoff is obtained from (24) and (25) as

$$\Theta\left((1-1/q)^{-2\gamma k}\frac{n(uk+B)^2}{W^2B}\right), \tag{40}$$

where $k=O(n)$. Therefore, the optimal data size is $B=\Theta(uk)$, and therefore the tradeoff is $\Theta\left((1-1/q)^{-2\gamma k}\frac{nuk}{W^2}\right)$. The tradeoff will be optimal when $u=\Theta(\log k)$. As in Remark 4, *B* is at least $\Theta(\log n)$. Hence, $k=\Theta(\frac{\log n}{\log\log n})$ and $u=\Theta(\log\log n)$. In that condition, the tradeoff is $\Theta\left(\frac{n\log n}{W^2}\right)$. The result shows that there is no order gain on tradeoff in 2-hop relay scheme. If we optimize the goodput, it becomes $\Theta(W)$ for the configuration $k=\Theta(n)$, $u=\Theta(\log n)$ and $B=\Theta\left(n\log n\right)$. Hence, there is no order gain on goodput by employing network coding comparing with no replicas case. However, when compared to the replicas case, whose goodput is $\Theta\left(\frac{W}{\sqrt{n}}\right)$, the goodput improvement is of order $\Theta\left(\sqrt{n}\right)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

Afterwards, we consider the flooding scheme. The flooding scheme delay/goodput tradeoff of random i.i.d. mobility model is obtained from (26) and (27) as

$$\Theta\left((1-\frac{1}{q})^{-2\eta_{flooding}}\frac{n(ku\log n+B\log n)^2}{W^2Bk}\right), \tag{41}$$

where $k = O(\log n)$. Therefore, the optimal data size is $B = \Omega(ku)$. For any $k = o(\log n)$, according to the expression of $\eta_{flooding}$, the relation between k and u satisfies $u = O(k)$. Therefore, in order to optimize the tradeoff, the configuration can be $B = \Theta(\log n)$, $k = \Theta(\sqrt{\log n})$ and $u = \Theta(\sqrt{\log n})$, and the optimal tradeoff becomes $\Theta\left(\frac{n\log^{\frac{5}{2}} n}{W^2}\right)$. The tradeoff of flooding scheme without network coding in [23] is $\Theta\left(\frac{n\log^3 n}{W^2}\right)$ when the data size is $B = \Theta(\log n)$ as in Remark 4. Hence, there is a small gain $\Theta\left(\sqrt{\log n}\right)$ on tradeoff by employing network coding since there are too many replicas in no network coding case. In addition, since $\eta_{flooding} = \Theta(n\log n)$ when $k = \Theta(\log n)$, the optimal goodput in (26) is $\Theta\left(\frac{W}{n}\right)$ for the case $B = \Theta(\log^2 n)$ and $u = \Theta(\log n)$. The goodput gain in that condition is $\Theta(\log n)$.

C. Discussion on Mobile Networks with Random Walk Model

In random walk model, we firstly consider the 2-hop relay scheme. The delay/goodput tradeoff is obtained from (35) and (36) as

$$\tilde{\Theta}\left((1-1/q)^{-2\varrho k}\frac{n^2(uk+B)^2}{W^2Bk}\right), \tag{42}$$

where $k = O(n)$. The tradeoff will be optimal when $B = \Theta(uk)$ and $u = \Theta(\log k)$. Thus, in order to obtain optimal tradeoff, as in Remark 4, $B = \Theta(\log n)$, $k = \Theta(\frac{\log n}{\log\log n})$ and $u = \Theta(\log\log n)$ which are the same as it in 2-hop random i.i.d. mobility model. The optimal tradeoff is $\tilde{\Theta}\left(\frac{n^2}{W^2}\right)$. There is still no order gain on tradeoff. When we optimize the goodput, the optimal configuration of network coding is $k = \Theta(n)$, $u = \Theta(\log n)$ and $B = \Theta(n\log n)$. The corresponding optimal goodput is $T(n) = \tilde{\Theta}(W)$. Therefore, there is no order gain on goodput by employing network coding comparing with the no replicas scheme. However, when compared to the replicas case, the goodput gain is of order $\tilde{\Theta}(n)$. We further find that in both 2-hop random i.i.d. mobility model and 2-hop random walk model, the structures of unicast session are the same. Hence, the optimal configurations for these models are almost the same.

For the flooding scheme, the delay/goodput tradeoff of random walk model is obtained from (37) and (38) as

$$\Theta\left((1-\frac{1}{q})^{-2\zeta_{flooding}}\frac{n^2(ku+B)^2}{W^2Bk}\right), \tag{43}$$

where $k = O(\sqrt{n})$. Thus, the optimal data size is $B = \Theta(ku)$. When $k = O(\sqrt{n})$, there must be $u = O(\log kn) = O(\log n)$. Moreover, since the number of hops from source to destination is at least $\Theta(\sqrt{n})$, we can obtain that $\zeta_{flooding} = \Omega(\sqrt{n})$, and therefore $u = \Theta(\log n)$. Hence, the optimal tradeoff becomes $\Theta\left(\frac{n^2\log n}{W^2}\right)$. Comparing with the tradeoff of flooding scheme without network coding $\Theta\left(\frac{n^2\log n}{W^2}\right)$ in [5], when the data size is $B = \Theta(\log n)$, there is still not improvement. Moreover, the optimal goodput is $\Theta\left(\frac{W}{n}\right)$ when $k = \Theta(\sqrt{n})$, $B = \Theta(\sqrt{n}\log n)$ and $u = \Theta(\log n)$. There is order gain of $\Theta(\sqrt{n})$ on goodput. Furthermore, the corresponding tradeoff is $\Theta\left(\frac{n^2\log n}{W^2}\right)$ which is optimal. Thus, the optimal tradeoff and goodput can be achieved simultaneously in random walk model with flooding scheme.

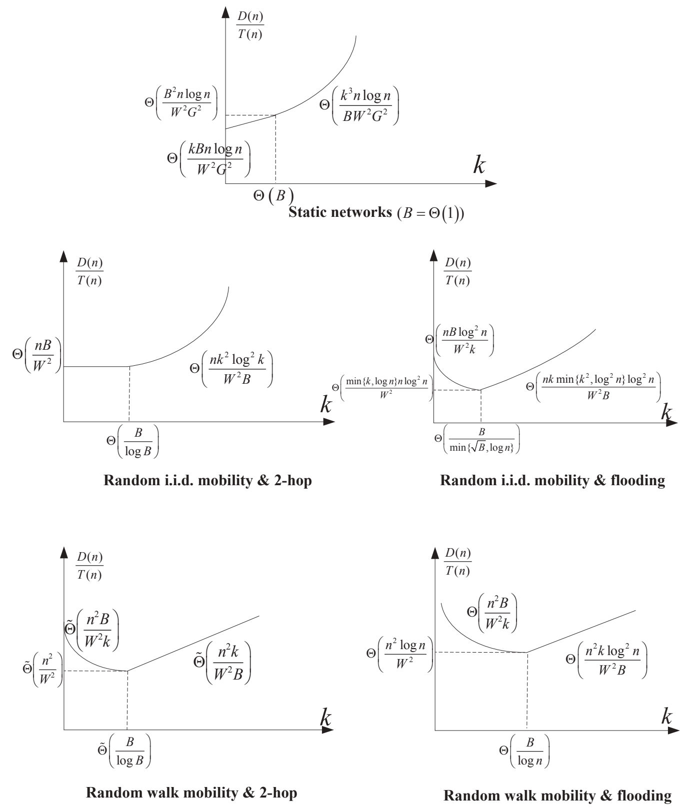


Fig. 3. The relation between tradeoff and generation size for each model and scheme

TABLE II
THE IMPROVEMENTS OF DELAY/GOODPUT AND GOODPUT PERFORMANCE BY EMPLOYING NETWORK CODING

Model	Tradeoff gain	Goodput gain
Static	$\Theta(1)$	$\Theta(1)$
I.i.d. mobility 2-hop	$\Theta(1)$	$\Theta(\sqrt{n})$
I.i.d. mobility multi-hop	$\Theta(\sqrt{\log n})$	$\Theta(\log n)$
Random walk 2-hop	$\tilde{\Theta}(1)$	$\tilde{\Theta}(n)$
Random walk multi-hop	$\Theta(1)$	$\Theta(\sqrt{n})$

The relations between the delay/goodput tradeoff and k are illustrated in Fig. 3 for each model and scheme. Furthermore, the improvements of network coding are listed in Table II[2] and all of the corresponding optimal results are shown in Table III[3].

[2]Since network coding schemes are redundancy-based, the performance of network coding schemes is compared with the redundancy-based schemes without network coding in this table.

[3]In Subsection V.C, it is proved that the optimal tradeoff and goodput can be achieved at the same time in random walk model with flooding scheme. Therefore, we note the case 2 and 3 together here.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

TABLE III
THE CONFIGURATIONS AND PERFORMANCE OF NETWORK CODING FOR DIFFERENT MODELS AND SCHEMES. CASE 1: NETWORK CODING IS NOT EMPLOYED, WHERE (a) IS NO REPLICAS CASE AND (b) IS REPLICAS CASE. CASE 2: NETWORK CODING IS EMPLOYED TO OPTIMIZE THE DELAY/GOODPUT TRADEOFF. CASE 3: NETWORK CODING IS EMPLOYED TO OPTIMIZE THE GOODPUT.

Model	Case	Tradeoff(sec^2/bits)	Goodput(bits/sec)	Delay(sec)	B(bits)	k	u(bits)
Static	1	$\Theta\left(\frac{n\log n}{W^2}\right)$	$\Theta\left(\frac{W}{\sqrt{n\log n}}\right)$	$\Theta\left(\frac{\sqrt{n}\log n}{W}\right)$	$\Theta(1)$	-	-
	2	$\Theta\left(\frac{n\log n}{W^2}\right)$	$\Theta\left(\frac{W}{\sqrt{n\log n}}\right)$	$\Theta\left(\frac{\sqrt{n}\log n}{W}\right)$	$\Theta(1)$	1	1
	3	$\Theta\left(\frac{n\log n}{W^2}\right)$	$\Theta\left(\frac{WG}{\sqrt{n\log n}}\right)$	$\Theta\left(\frac{G\sqrt{n}\log n}{W}\right)$	$\Theta(G)$	$\Theta(G)$	$\Theta(1)$
I.i.d. mobility 2-hop	$1(a)$	$\Theta\left(\frac{n\log n}{W^2}\right)$	$\Theta(W)$	$\Theta\left(\frac{n\log n}{W}\right)$	$\Theta(\log n)$	-	-
	$1(b)$	$\Theta\left(\frac{n\log n}{W^2}\right)$	$\Theta(\frac{W}{\sqrt{n}})$	$\Theta\left(\frac{\sqrt{n}\log n}{W}\right)$	$\Theta(\log n)$	-	-
	2	$\Theta\left(\frac{n\log n}{W^2}\right)$	$\Theta\left(\frac{W\sqrt{\log n}}{\sqrt{n\log\log n}}\right)$	$\Theta\left(\frac{\sqrt{n}\log^{\frac{3}{2}} n}{W\sqrt{\log\log n}}\right)$	$\Theta(\log n)$	$\Theta(\frac{\log n}{\log\log n})$	$\Theta(\log\log n)$
	3	$\Theta\left(\frac{n^2\log n}{W^2}\right)$	$\Theta(W)$	$\Theta\left(\frac{n^2\log n}{W}\right)$	$\Theta\left(n\log n\right)$	$\Theta(n)$	$\Theta(\log n)$
I.i.d. mobility flooding	1	$\Theta\left(\frac{n\log^3 n}{W^2}\right)$	$\Theta\left(\frac{W}{n\log n}\right)$	$\Theta\left(\frac{\log^2 n}{W}\right)$	$\Theta(\log n)$	-	-
	2	$\Theta\left(\frac{n\log^{\frac{5}{2}} n}{W^2}\right)$	$\Theta\left(\frac{W}{n\sqrt{\log n}}\right)$	$\Theta\left(\frac{\log^2 n}{W}\right)$	$\Theta(\log n)$	$\Theta(\sqrt{\log n})$	$\Theta(\sqrt{\log n})$
	3	$\Theta\left(\frac{n\log^3 n}{W^2}\right)$	$\Theta\left(\frac{W}{n}\right)$	$\Theta\left(\frac{\log^3 n}{W}\right)$	$\Theta(\log^2 n)$	$\Theta(\log n)$	$\Theta(\log n)$
Random walk 2-hop	$1(a)$	$\Theta\left(\frac{n\log^2 n}{W^2}\right)$	$\Theta\left(W\right)$	$\Theta\left(\frac{n\log^2 n}{W}\right)$	$\Theta(\log n)$	-	-
	$1(b)$	$\tilde{\Theta}\left(\frac{n^2}{W^2}\right)$	$\tilde{\Theta}\left(\frac{W}{n}\right)$	$\tilde{\Theta}\left(\frac{n}{W}\right)$	$\Theta(\log n)$	-	-
	2	$\tilde{\Theta}\left(\frac{n^2}{W^2}\right)$	$\tilde{\Theta}\left(\frac{W}{n}\right)$	$\tilde{\Theta}\left(\frac{n}{W}\right)$	$\Theta(\log n)$	$\Theta(\frac{\log n}{\log\log n})$	$\Theta(\log\log n)$
	3	$\tilde{\Theta}\left(\frac{n^2}{W^2}\right)$	$\tilde{\Theta}\left(W\right)$	$\Theta\left(\frac{n^2}{W}\right)$	$\tilde{\Theta}(n\log n)$	$\Theta(n)$	$\Theta(\log n)$
Random walk flooding	1	$\Theta\left(\frac{n^2\log n}{W^2}\right)$	$\Theta\left(\frac{W}{n\sqrt{n}}\right)$	$\Theta\left(\frac{\sqrt{n}\log n}{W}\right)$	$\Theta(\log n)$	-	-
	2,3	$\Theta\left(\frac{n^2\log n}{W^2}\right)$	$\Theta\left(\frac{W}{n}\right)$	$\Theta\left(\frac{n\log n}{W}\right)$	$\Theta(\sqrt{n}\log n)$	$\Theta(\sqrt{n})$	$\Theta(\log n)$

VI. CONCLUSION

In this paper, we analyze the network coding configuration in both static and mobile ad hoc networks so as to optimize the delay/goodput tradeoff and the goodput with the consideration of the throughput loss and decoding loss of network coding. These results reveal the impact of network scale on network coding system, which is not studied in previous works. Moreover, we also compare the performance with the corresponding networks without network coding. The results indicate that network coding provides improvement on goodput in mobile networks, but no gain on delay/goodput tradeoff in all of the proposed models and schemes except for the flooding scheme in i.i.d. mobility model.

VII. ACKNOWLEDGEMENT

This work is supported by NSF China (No. 61325012, 61271219), China Ministry of Education Doctor Program (No. 20130073110025), Shanghai Basic Research Key Project (No.11JC1405100), Shanghai International Cooperation Project (No. 13510711300) and the open research fund of Zhejiang Provincial Key Lab of Data Storage and Transmission Technology, Hangzhou Dianzi University (No.201306).

REFERENCES

[1] T. Berger, "Multiterminal source coding," *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, Aug. 1977.

[2] R. Ahlswede, N. cai, S. Li, and R.Yeung, "Network information flow," *Information Theory, IEEE Transaction on* , vol. 46, no. 4, pp. 1204-1216, Jul. 2000.

[3] R. Ahlswede, N. cai, S. Li, and R.Yeung, "Linear network coding," *Information Theory, IEEE Transaction on* , vol. 49, no. 2, pp. 371-381, Feb. 2003.

[4] J. Liu, D. Goeckel, et al., "Bounds on the gain of network coding and broadcasting in wireless networks," *Proc. of IEEE INFOCOM 2007*, Anchorage, Alaska, USA, May 2007.

[5] C. Zhang, X. Zhu and Y. Fang, "On the improvement of scaling laws for large-scale MANETs with network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 5, pp. 662-672, Jun. 2009.

[6] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Joint Scheduling and Resource Allocation in Uplink OFDM Systems for Broadband Wireless Access Networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 226 - 234, Feb. 2009.

[7] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink Scheduling and Resource Allocation for OFDM Systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 288 - 296, Jan. 2009.

[8] C. Zhang, Y. Fang and X. Zhu, "Throughput-delay tradeoffs in large-scale MANETs with network coding," *proc. of IEEE INFOCOM 2009*, Rio de Janeiro, Brazil, Apr. 2009.

[9] Q. Yan, M. Li, Z.Yang, et al., "Throughput analysis of cooperative mobile content distribution in vehicular network using symbol level network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 2, pp. 484-492, Feb. 2012.

[10] S. Karande, Z. Wang, H. Sadjadpour, J. Garcia-Luna-Aceves, "Multicast throughput order of network coding in wireless ad-hoc networks," *Communications, IEEE Transactions on*, vol. 59, no. 2, pp. 497-506, Feb. 2011.

[11] K. Lu, S. Fu, Y. Qian, and H. Chen "On capacity of random wireless networks with physical-layer network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 5, pp. 763-772, Jun. 2009.

[12] Z. Yang, M. Li, W. Lou, "Codeplay: live multimedia streaming in VANETs using symbol-level network coding," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 8, pp. 3006-3013, Aug. 2012.

[13] S. Sharma, Y. Shi, et al., "Is network coding always good for cooperative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2333536, IEEE Transactions on Vehicular Technology

communications?" *Proc. of IEEE INFOCOM 2010*, San Diego, CA, USA, Mar. 2010.

[14] A. Asterjadhi, E. Fasolo, et al., "Toward network coding-based protocols for data broadcasting in wireless ad hoc networks," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 662-673, Feb. 2010.

[15] D. Niu and B. Li, "Analyzing the resilience-complexity tradeoff of network coding in dynamic P2P networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 22, no. 11, pp. 1842-1850, Nov. 2011.

[16] P. Gupta and P. Kumar, "The capacity of wireless networks," *Information Theory, IEEE Transactions On*, vol. 46, no. 2, pp. 338-404, Mar. 2000.

[17] M. Riemensberger, Y. Sagduyu, M. Honig, W. Utschick, "Training overhead for decoding random linear network codes in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 5, pp. 729-737, Jun. 2008.

[18] T. Ho, M. Médard, R. Koetter, et al., "A random linear network coding approach to multicast," *Information Theory, IEEE Transaction on*, vol. 52, no. 10, pp. 4413 - 4430, Oct. 2006.

[19] U. Lee, S. Lee ; K. Lee ; M. Gerla, "Understanding processing overheads of network coding-based content distribution in VANETs." *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 11, pp. 2304-2318, Nov. 2013.

[20] L. Fu, Y. Qin, X. Wang, X. Liu, "Throughput and delay analysis for convergecast with MIMO in wireless networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 4, pp. 768-775, Apr. 2011.

[21] W. Huang, X. Wang, "Throughput and delay scaling of general cognitive networks," *Proc. of IEEE INFOCOM 2011*, Shanghai, China, Apr. 2011.

[22] B.Ribeiro, P. Wang, et al., "Sampling directed graphs with random walks," *Proc. of IEEE INFOCOM 2012*, Orlando, USA, Mar. 2012.

[23] M. Neely and E. Modiano, "Capacity and delay tradeoffs for ad-hoc mobile networks," *Information Theory, IEEE Transaction on*, vol. 51, no. 6, pp. 1917-1937, Jun. 2005.

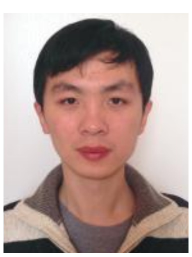
[24] S. Zhang, S. Liew, and P. Lam, "Hot topic: physical-layer network coding," *MOBICOM 2006*, Los Angeles, CA, USA, Sep. 2006.

[25] T. Koike-Akino, P. Popovski, V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 5, pp. 773-787, Sep. 2009.

[26] R. Koetter M. Médard, "An algebraic approach to network coding," *Networking, IEEE/ACM Transaction on*, vol. 11, no. 5, pp. 782-795, Jun. 2003.

[27] C. Hu, X. Wang, D. Nie and J. Zhao, "Multicast scaling laws with hierarchical cooperation," *Proc. of IEEE INFOCOM 2010*, San Diego, CA, USA, Mar. 2010.

[28] X. Lin, G. Sharma, R. Mazumdar, N. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with Brownian mobility," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2777-2784, Jun. 2006.



Yi Qin received the B.S. and M.S. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2012, respectively. He is currently working toward the Ph.D. degree with Shanghai Jiao Tong University, Shanghai, China.

His research interests include D2D communication, D2D discovery, and Ad Hoc network scaling laws study.

Feng Yang received his Ph.D degree in Information and Communication from ShangHai JiaoTong University. Since 2008 , he has been on the faculty at ShangHai JiaoTong University,where he is an associate professor of electronic information. His research interest include wireless video communication, multihop communication. He take part in the programme of Beyond 3G Wireless Communication Testing System and in charge of system design. By now he is the PI of some national projects, including the National High Technology Research and Development Program of China (863 Program), National Natural Science Foundation of China.

Xiaohua Tian received his B.E. and M.E. degrees in communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2003 and 2006, respectively. He received the Ph.D. degree in the Department of Electrical and Computer Engineering (ECE), Illinois Institute of Technology (IIT), Chicago, in Dec. 2010. He is currently an Assistant Professor in Department of Electronic Engineering of Shanghai Jiao Tong University, China. He won the Highest Standards of Academic Achievement 2011 of IIT and Fieldhouse Research Fellowship 2009 of IIT, which is awarded to only one student over IIT each year. His research interests include application-oriented networking, Internet of Things and wireless networks. He serves as the guest editor of International Journal of Sensor Networks, publicity co-chair of WASA 2012. He also serves as the Technical Program Committee member for Wireless Networking Symposium, Ad Hoc and Sensor Networks Symposium of IEEE GLOBECOM 2013, Ad Hoc and Sensor Networks Symposium of IEEE ICC 2013, Wireless Networking Symposium of IEEE GLOBECOM 2012, Communications QoS, Reliability, and Modeling Symposium (CQRM) of GLOBECOM 2011, and WASA 2011.

Xinbing Wang received the B.S. degree (with hons.) from the Department of Automation, Shanghai Jiaotong University, Shanghai, China, in 1998, and the M.S. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2001. He received the Ph.D. degree, major in the Department of electrical and Computer Engineering, minor in the Department of Mathematics, North Carolina State University, Raleigh, in 2006. Currently, he is a professor in the Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai, China. Dr. Wang has been an associate editor for IEEE/ACM Transactions on Networking and IEEE Transactions on Mobile Computing, and the member of the Technical Program Committees of several conferences including ACM MobiCom 2012, ACM MobiHoc 2012-2014, IEEE INFOCOM 2009-2014.

Hanwen Luo received B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1977. He was the Vice Director of the Shanghai Institute of Wireless Communications Technology and the Vice Director of the Institute of Wireless Communication Technology, Shanghai Jiao Tong University, Shanghai, China. He was the leading Specialist with the China 863 High-Tech Program on beyond third-generation wireless communication systems and with the China 973 High-Tech Program on the research of military equipment.

He is currently a Full Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. His research interests include mobile and personal communications.

Haiquan Wang (M03) received the M.S. degree in Nankai University, China, in 1989, and the Ph.D. degree in Kyoto University, Japan, in 1997, both in mathematics, and the Ph.D. degree in University of Delaware, Newark, in 2005, in electrical engineering. From 1997 to 1998, he was a Postdoctoral Researcher with the Department of Mathematics, Kyoto University. From 1998 to 2001, he was a Lecturer (part-time) with Ritsumei University, Japan. From 2001 to 2002, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Delaware, Newark. From 2005 to 2008, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He joined the College of Communications Engineering, Hangzhou Dianzi University, Hangzhou, China, in July 2008 as a faculty member. His research interests include wireless communications, digital signal processing, and information theory.

Mohsen Guizani (S'85-M'89-SM'99-F'09) is currently a Professor and the Associate Vice President for Graduate Studies at Qatar University, Doha, Qatar. He was the Chair of the Computer Science Department at Western Michigan University from 2002 to 2006 and Chair of the Computer Science Department at University of West Florida from 1999 to 2002. He also served in academic positions at the University of Missouri-Kansas City, University of Colorado-Boulder, Syracuse University and Kuwait University. He received his B.S. (with distinction) and M.S. degrees in Electrical Engineering; M.S. and Ph.D. degrees in Computer Engineering in 1984, 1986, 1987, and 1990, respectively, from Syracuse University, Syracuse, New York. His research interests include Computer Networks, Wireless Communications and Mobile Computing, and Optical Networking. He currently serves on the editorial boards of six technical Journals and the Founder and EIC of "Wireless Communications and Mobile Computing" Journal published by John Wiley (http://www.interscience.wiley.com/jpages/1530-8669/). He is the author of eight books and more than 300 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE Journals and Magazines. He also served as member, Chair, and General Chair of a number of conferences. Dr. Guizani served as the Chair of IEEE Communications Society Wireless Technical Committee (WTC) and Chair of TAOS Technical Committee. He was an IEEE Computer Society Distinguished Lecturer from 2003 to 2005. Dr. Guizani is an IEEE Fellow and a Senior member of ACM.