www.ietdl.org

Published in IET Communications
Received on 13th July 2013
Revised on 18th December 2013
Accepted on 7th January 2014
doi: 10.1049/iet-com.2013.0611

ISSN 1751-8628

Performance analysis of network coding-based content distribution in vehicular *ad-hoc* networks

Mohammad H. Amerimehr, Farid Ashtiani

Advanced Communications Research Institute (ACRI), Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran
E-mail: ashtianimt@sharif.edu

Abstract: The authors investigate the content distribution among the vehicles of a cluster in a vehicular *ad-hoc* network, exploiting network coding. The vehicles collaborate to disseminate the coded data packets, received from a roadside info-station based on IEEE 802.11 medium access control (MAC) protocol. Two types of network coding are considered: random linear network coding (RLNC) over a large finite field and random XORed network coding (RXNC). An analytical model is proposed to address the effect of random access MAC as well as the correlation among received coded packets on the performance of content distribution. First, a *p*-persistent carrier sense multiple access approximation for IEEE 802.11 MAC is adopted, to derive the expected amount of time necessary to deliver the whole information (a data file) to the vehicles, that is, content distribution delay, based on RLNC. Second, the content distribution delay for RXNC is investigated. In this respect, the authors determine the probability of a newly received packet to be innovative for RXNC and characterise the probability that all nodes succeed to retrieve the data file, that is, content delivery ratio. Finally, the success of content distribution process for erasure channels is assessed.

1 Introduction

One of the promising applications in vehicular *ad-hoc* networks (VANETs) is content distribution among vehicles. The examples include disseminating multimedia files, software updates, advertising or emergency videos, and so on [1, 2]. A key challenge for content distribution is designing an effective algorithm for data distribution that provides high level of quality of service (QoS). For example, disseminating an emergency video over the vehicles in a delay-efficient manner is very important [3]. To improve the performance of data communication networks, network coding has emerged as an effective alternative to traditional routing [4]. Network coding generalises the traditional store-and-forward paradigm to store-code-forward, where each node stores the incoming packets and sends out a combination of stored data. Besides increasing the throughput and improving robustness in the presence of packet loss, network coding is also able to improve QoS by reducing data transferring delay [5].

One of the implementation methods for network coding is through random linear combination of the incoming packets to form an outgoing packet, that is, random linear network coding (RLNC) [6]. RLNC does not rely on network topology information and can be implemented easily in a distributed fashion [7]. It is also robust against network topology changes. In RLNC, selecting random coefficients from a sufficiently large field provides a significant throughput gain, since coded packets are innovative with high probability [8]. However, a simple implementation method for RLNC is through random XORed network coding (RXNC), that is, the random coefficients are binary numbers. In this case the overhead due to including coefficients in the packet header is sufficiently reduced at the cost of considerable probability of correlation among coded packets.

Recently, network coding has attracted much interest for content distribution in vehicular networks. It has been shown that network coding facilitates the data dissemination by eliminating the neighbour discovery process involved in traditional content distribution methods [9]. Moreover, adopting network coding-based schemes expedites the content distribution process which is suitable for multimedia applications.

In this paper, we provide a theoretical performance analysis of network coding-based content distribution in a vehicular network adopting IEEE 802.11 medium access control (MAC) protocol. This problem is challenging as discussed in the following. The nodes contend to acquire the channel and transmit their packets successfully. However, the nodes that succeed to transmit their packets do not take part in the future contentions. Therefore, the status of nodes changes during the content distribution process. Furthermore, packet collision incurred by 802.11 MAC protocol brings about packet drop at the destination. Moreover, the correlation among received coded packets affects retrieving the data file. More precisely, a newly received packet may not be innovative (i.e. it is a linear combination of the set of previously received packets) and should be distinguished from an innovative packet in the analysis. To the best of our knowledge, these factors are not addressed together in the previous analytical works. The main contribution of this paper can be summarised as follows:

(1) Adopting a p-persistent carrier sense multiple access (CSMA) approximation for IEEE 802.11 MAC, we derive the expected delay for content distribution based on RLNC over a large finite field. The results show that network coding provides a significant gain over traditional routing.
(2) We extend our analysis of (1) to a simple version of RLNC, that is, random XORed network coding (RXNC). To this end, we derive the probability of a newly received packet to be innovative for RXNC, as a function of the number of innovative received packets. Taking the correlation among packets into account, we determine the probability that all nodes succeed to retrieve the data file (i. e. content delivery ratio (CDR)) as well as the expected content distribution delay (CDD).
(3) We evaluate the impact of channel impairment on the success of content distribution process for RLNC by characterising the ratio of nodes that are able to retrieve the data file.

The reminder of our paper is organised as follows: A summary of related works is presented in Section 2. Section 3 describes the network scenario and our assumptions. Section 4 provides the analysis of CDR. The analysis of the expected CDD is developed in Section 5. We address the impact of packet erasure on the content distribution process in Section 6. We present the numerical results and simulations in order to show the performance and accuracy of our analytical approach in Section 7. Finally, Section 8 summarises the main results and discusses possible extensions of the proposed schemes.

2 Related works

Network coding is a promising technique for distribution of a common file among a set of nodes. For live multimedia streaming in VANETs, [10] proposes a symbol-level network coding-based method, CODEPLAY, to meet the criteria of live multimedia streaming such as short-time end-to-end delay for all receivers. Using the benefit of both network coding and wireless symbol-level diversity, [11] devises a novel scheme, CODEON, to distribute a popular content among vehicles in a lossy wireless network. Each vehicle receives a segment of a common interested file from a roadside access point and cooperates with other vehicles to distribute the file among them. Ahmed and Kanhere [12] suggested network coding-based cooperative content distribution, VANETCODE, to enhance cooperative downloading in VANETs. The access point sends a random linear combination of the segments to each vehicle. The vehicles, then, collaborate to distribute the received coded packets and retrieve the data file. The authors show that the proposed scheme eliminates the need for neighbour discovery and content selection which is necessary in CODEON and CODEPLAY. Using the simulations, they also show that VANETCODE outperforms the routing-based content distribution. Johnson *et al.* [13] addressed content distribution in multi-hop *ad-hoc* networks and proposed a distributed protocol using RLNC. They adopt a scheduled MAC to avoid collision. The authors of [14, 15] also propose network coding-based protocols for file sharing in VANETs to improve robustness against topology changes and intermittent connectivity. Lee *et al.* [16] investigated the impact of resource constraints (i.e. computation and memory) on the performance of the network coding-based content distribution. Information dissemination via multiple relay nodes is addressed in [17], where the relay nodes apply RLNC to spread the received information to other nodes. In this work, the expected decoding delay based on an ideal channel without collision or error is derived. Most of the prior works prove the benefit of network coding-based content distribution by simulations, and the analytical works in this area are limited.

Random access method used in VANETs (e.g. IEEE 802.11p protocol) or wireless *ad hoc* networks results in packet collisions which should be considered in network performance analysis. Reference [18] considers a broadcasting scheme based on network coding for wireless *ad hoc* networks. They quantify the impacts of IEEE 802.11 on the performance of network coding by evaluating the network throughput. In their analytical model, they assume that during data broadcasting, the probability of packet transmission is constant at each node. Moreover, it is assumed that each received packet contains new information. In the simulation results, however, they relax these assumptions and take the correlation among time varying number of active nodes into account, as a node may temporarily not take part in packet transmission if it does not have innovative packets to send to its neighbours. Yan *et al.* [19] addresses mobile content distribution in a VANET exploiting symbol-level network coding. They develop an analytical framework to evaluate the expected throughput from a roadside access point to each vehicle at a certain location. They consider realistic factors such as vehicle distribution and mobility pattern, channel fading and packet collision in their model. To simplify the analysis, they assume all nodes are saturated (i.e. each node always has a packet to send) and the received packets are innovative.

Ye *et al.* [20] study the inter-vehicle data dissemination using network coding. They consider a single source one-directional dissemination scenario in a tandem multi-hop network and provide the steady-state dissemination velocity of a file within the network. For the special case of three nodes, they also derive the probability mass function (PMF) of dissemination completing time by considering the correlation among the set of coded packets received at each node. In their work they assume collision-free transmissions among nodes. Firooz and Roy [9] proposed a network coding-based dissemination algorithm over an *ad hoc* network with a general topology. In this analytical framework, the authors consider packet reception failure due to fading and noise but adopt an interference-free scheduler to eliminate packet collision. Firooz and Roy [21] addressed distributing a common file to two vehicles in a VANET. Similar to scenario considered in our paper, the proposed content distribution in [21] comprises two phases. During the first phase, that is, roadside access point to vehicle phase, the access point sends coded packets to each vehicle. At the second phase, that is, vehicle to vehicle phase, the vehicles exchange the coded packets to retrieve the whole data. The authors in [21] derive an upper bound for the completion time, using collision-free MAC.

In our analytical framework, we consider three important issues affecting the network coding-based content distribution: (a) innovation probability of received coded packets, that is, the probability that a received packet has

doi: 10.1049/iet-com.2013.0611

new information and contributes to retrieve the data file, (b) non-saturated network, that is, the number of active nodes changes during the content distribution (the node that has not yet sent its packet is referred to as an active node) and (c) considering IEEE 802.11-based random access MAC protocol which results in packet loss due to collision. Therefore, our work takes a step forward to assess network coding benefits from a theoretical point of view. Existing analytical works in this area do not address these factors altogether. To obtain a more realistic model, we incorporate all these factors in our analytical model.

3 Network scenario

We consider content distribution in a highway among a cluster of N vehicles moving in the same direction (Fig. 1). The vehicles are interested in a common data file. A set of info-stations are installed at the side of the highway at regular intervals to send data to the intended vehicles. The high speed of vehicles limits the duration a vehicle is in the coverage of an info-station. Hence, an info-station can only send a part of the data file to each vehicle during the connection period. Assuming a typical scenario that the distance between two adjacent info-stations is about 2 km, the transmission range of each info-station is about 200 m, and the vehicle speed is about 25 m/s. A vehicle is able to establish a connection with an info-station only in a short period of time (depending on the length of the cluster and the number of vehicles in a cluster). The intermittent connection with the info-stations motivates the vehicles to collaborate with each other to acquire the whole data in the intervals that they do not have access to an info-station. This process is repeated whenever a cluster of vehicles passes a new info-station. The info-stations are connected with each other through the internet connection. Hence, using a set of info-stations, a large file (e.g. an advertisement video) can be delivered to the vehicles based on the cooperative content distribution. It is worth noting that broadcasting data from an info-station to a cluster of vehicles is not generally practical, because the length of cluster may be comparable to the transmission range of the info-station. Hence, the interval that the cluster lies in the range of an info-station is limited due to high velocity of vehicles. However, the relative velocities of vehicles (moving in the same direction) are usually much smaller than the absolute velocities. Thus, the vehicles can collaborate with each other in a larger time interval.

In the routing-based scenario, neighbouring nodes possibly contain the same data blocks. As a result, broadcasting a block is not beneficial to all neighbours. Hence, an efficient routing-based content distribution like SPAWN [22] relies on cooperating with neighbours and requesting for the missing blocks. Alternatively, we propose a network coding-based method similar to the method introduced in [12]. The network coding-based scenario eliminates the need for neighbour discovery and content selection because each coded block is very likely to contain innovative information for all neighbours. This property as well as the broadcast nature of wireless medium leads to efficient content distribution algorithms.

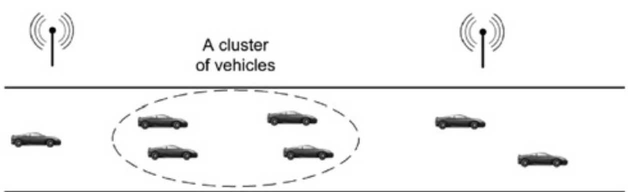

Fig. 1 *Content distribution among a cluster of N vehicles*

Although several info-stations may be involved in a complete content distribution process, we only focus on one period of content distribution, which comprises transferring a file from an info-station to the intended vehicles and cooperation among vehicles in order to retrieve the whole data. The info-station splits the data file into G pieces, called data generations. Each data generation is further split into M blocks (hereafter, referred to data packets, $M \le N$). In the exchange phase, the packets are transmitted such that data generations are retrieved sequentially. As an example, data generation may play the role of different frames of a video or different levels of details of an image such that each data generation enhances the quality of the image retrieved in the preceding generation. The motivation behind this double decomposition is explained later in this section. The number of blocks depends on the vehicles' velocity. Since the total size of data packets received by each vehicle is $1/M$ of the size of data file, higher velocity reduces the connection time with an info-station to receive data and thus increases M.

The content distribution process comprises two phases: receive phase and exchange phase. In the first phase (receive phase), the info-station sends a set of G packets (related to G generations, distinctly) to each vehicle passing its coverage area. The type of packets depends on the content distribution scheme. We address the following three schemes in this paper.

(1) *RLNC over a large finite field:* Each transmitted packet is a random linear combination of the data packets belonged to the same generation. The coding coefficients are selected from a large finite field.
(2) *Random XORed network coding (RXNC):* The info-station randomly XORs data packets of the same generation to generate and transmit a new coded packet to each vehicle (the coding coefficients are selected from the binary field).
(3) *Routing:* The info-station sends randomly one of the data packets to each vehicle without performing network coding.

Fig. 2 illustrates the receive phase of content distribution, where the info-station sends a coded packet (for each generation) to vehicles in its coverage area. It is worth noting that in both network coding schemes, the coding coefficients are incorporated in the packet header. If the coding coefficients are selected from $GF(2^q)$, the overhead is Mq bits. RXNC is an RLNC implementation method that

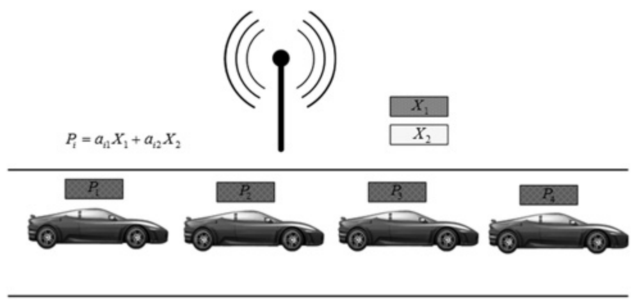

Fig. 2 *Receive phase of content distribution*
The info-station sends a coded packet to each vehicle

reduces the overhead significantly compared to RLNC over a large finite field, and simplifies computational complexity. On the other hand, RLNC over a large finite field increases the innovation probability among a set of coded packets which facilitates retrieving the data file at the exchange phase. More precisely, in RLNC, the probability that M data packets can be retrieved from N coded packets is approximately equal to $1-(1/q^{-(N-M+1)})$, provided that q is sufficiently large [21].

In the second phase (exchange phase), the vehicles exchange their packets received in the first phase. This process will end when a stopping condition occurs, which will be expressed later. We assume that the receive phase has been performed perfectly so that each vehicle has a packet at the end of this phase. We focus on the exchange phase in this paper.

We suppose a single-hop network of N vehicles, where each vehicle is in the transmission range of the others. The vehicles are moving in the same direction and with similar speeds such that the network topology is constant during the exchange phase. A classical collision model is used in our scenario, thus a transmission succeeds if there is no other simultaneous transmission in the interference range of the receiver [23]. We assume that all nodes are synchronised using a GPS device. The time is divided into slots. Each data packet transmission prolongs L slots.

We adopt IEEE 802.11 medium access protocol that is the base of IEEE 802.11p MAC protocol. Since IEEE 802.11 does not support ACK mechanism in broadcast mode [24], we propose a pseudo-broadcast method to ensure delivery of coded packets. In this method, the transmitting node sets a specific node as destination address in the packet header and unicasts the packet to that node. The other neighbours will overhear the channel to receive the transmitted packet. Only the intended node is responsible to send back an ACK after packet reception. We assume that the transmission of ACK is error-free. The packet will be retransmitted if the transmitter does not receive the ACK. Assuming collision as the main reason of packet loss, if the destination node receives the packet successfully, so do the other nodes.

It is worth noting that the content distribution in the exchange phase is comprised of distributing a set of data generations among the network. At the beginning of exchanging a data generation, all nodes are active and compete to acquire the wireless channel. In this state, each node selects its destination randomly. Once a node succeeds in sending its packet and receives its ACK, it becomes inactive in the sense that it does not compete with other nodes in the future. Hereafter, the active nodes select the last inactive node, rather than a random node, as the destination. The exchange phase of this generation will end when either all nodes become inactive or the 'stop message' is sent by one of the nodes. The stop message is sent via the ACK packet and indicates that the node succeeds to retrieve all data packets. The motivation behind sending ACK by the last inactive node is to ensure all nodes in the network have received the set of packets collected by that inactive node. Hence, if that inactive node can retrieve the data file, so do the others. Note that in this scheme, a node will not start transmitting data packets from a new generation until exchanging the previous generation is finished. This contributes to shorten the time of retrieving a data generation by reducing the collision probability. This is important in multimedia applications that the users benefit from retrieving fragments of data constantly, rather than waiting to acquire the entire data.

Table 1 Description of symbols

Symbol	Description
N	number of vehicles in a cluster
M	number of blocks (packets) in a data generation
G	number of data generations
n	number of active nodes
k	number of successful transmission
δ	slot size
ε	erasure probability
L	length of data in the number of slots
D	length of DIFS in the number of slots
L'	length of DIFS plus data in the number of slots
p	probability of packet transmission in p-persistent CSMA method
CW	contention window size
$P_{\text{SEP, RT}}$	the probability of successful exchange phase for routing
$P_{\text{SEP, NC}}$	the probability of successful exchange phase for RXNC
P_1	probability of successful reception for an overhearing node
P_{SCD}	successful content delivery probability

With respect to IEEE 802.11 standard, the size of typical data packets is in the order of 1 kB [25]. Consequently, in order to transmit a large file, it should be split into several small parts. On the other hand, employing network coding among a large number of elements will increase the overhead of network coding (incorporating coding coefficients in the file header) as well as decoding delay. This is another motivation to employ network coding among data packets of the same generation. By adjusting the number of data generation, the number of data packets at each generation (M) can keep small enough to reduce the coding overhead and decoding delay. For routing-based scenario, reducing the number of blocks results in smaller CDD, since each data generation can be retrieved using only M data packets.

In our analytical approach, MAC protocol is modelled as a p-persistent CSMA, which is shown to be a good approximation for IEEE 802.11 MAC protocol if the parameter p is carefully selected [26]. In this model, the exponential back-off process is approximated by a geometric distribution. When a node decides to transmit a packet, senses the channel at the beginning of each time slot and starts to send the packet with probability p if the channel is free. If the channel is busy, it attempts to transmit with the same probability in a next free time slot. A collision occurs if more than one node transmit at the same slot. Unsuccessful nodes retry to send packet in the next free slots, with the same probability. The parameter p is selected such that the average back-off interval coincides with the window-based back-off mechanism, that is, $p = (2/1 + E(\text{CW}))$, where $E(\text{CW})$ is the average contention window size [26]. The list of symbols used in our analysis is summarised in Table 1.

4 Analysis of CDR

In this section, we address the effect of correlation among the transmitted packets on the success of content distribution process which is quantified by a metric called CDR. In this section, we focus on exchange phase of data packets from the same generation. Hereafter, using the term exchange phase denotes the exchange phase of a single data generation. A successful exchange phase is defined as the

IET Commun., 2014, Vol. 8, Iss. 9, pp. 1447–1458
doi: 10.1049/iet-com.2013.0611

one that all nodes succeed to retrieve the data generation, using the received packets, at the end of the exchange phase. We define CDR as the ratio of the number of successful exchange phases to the total number of exchange phases. If RLNC is performed over a large finite field, condition $N \geq M$ ensures that all nodes can retrieve the data generation after M successful transmissions. Hence, CDR in this case is equal to one. However, for RXNC, since the received packets are generally not independent, the exchange phase may need more than M successful transmissions. Therefore, the nodes may not be able to decode the packets even after N successful transmissions, that is, at the end of exchange phase. If the set of N coded packets contains M innovative packets, the exchange phase is successful. Hence, CDR equals $p_n(M;N, 0)$, where $p_n(j;i, k)$ denotes the conditional probability of j innovative packets among next i successfully transmitted ones, provided that k innovative packets have been received so far. $p_n(j;i, k)$ is derived from the following lemma:

Lemma 1: $p_n(j;i, k)$ is given by

$$p_n(j;\, i,\, k) = \frac{R(j;\, i,\, k)}{(2^M - 1)^i} \tag{1}$$

where $R(j;i, k)$, defined in Appendix 1, is given by the following recursive equation (see (2))

Proof: See Appendix 1. □

In the case of RLNC over a large finite field, a successful exchange phase is comprised of M successful transmissions. In contrast, for RXNC the number of successful transmissions leading to a successful exchange phase varies between M and N.

The probability that a successful exchange phase is comprised of l successful transmissions is given by

$$P_{\mathrm{SEP,NC}}(l) = \frac{p_n(M-1;\, l-1,\, 0)\, p_n(1;\, 1,\, M-1)}{p_n(M;\, N,\, 0)}; \quad M \leq l \leq N \tag{3}$$

Note that to finish the exchange phase after l successful transmissions, the rank of first $l-1$ received coded packets should be $M-1$. Similarly, the CDR and probability of successful exchange phase for routing, $P_{\mathrm{SEP,RT}}$, are derived, using the following lemma:

Lemma 2: Let $p'_n(j;i, k)$ denotes the conditional probability of j innovative packets among next i successfully transmitted ones, provided that k innovative packets have been received so far. $p'_n(j;i, k)$ is given by

$$p'_n(j;\, i,\, k) = \frac{R'(j;\, i,\, k)}{M^i} \tag{4}$$

where $R'(j;i, k)$ is given by the following recursive equation (see (5))

Proof: See Appendix 2. □

Hence, for routing the CDR is equal to $p'_n(M;N, 0)$. Moreover, we obtain (see (6))

5 Analysis of expected CDD

The total time of the content distribution process is affected by the MAC protocol as well as the innovation probability of the received packets. In this section, we provide an analytical approach to compute the average time of the process. We focus on the exchange phase of data among the same generation. We define the CDD as the total time it takes to finish the exchange phase. To analyse the expected CDD, we are interested in the derivation of expected time interval between two consecutive successful transmissions, defined as virtual transmission time (VTT). A VTT generally comprises successful slots as well as possibly idle and collision slots. Successful and collision slots denote slots in which successful and unsuccessful transmissions occur, respectively. Let $T_{\mathrm{VTT}}(n)$ denotes the expected time of VTT, provided that there are n active nodes in the system. It is obtained by the following expression derived in [27] as the average time between two consecutive successful transmissions

$$T_{\mathrm{VTT}}(n) = \frac{L' - (L'-1)(1-p)^n}{np(1-np)^{n-1}}\,\delta \tag{7}$$

where $L' = L + D$. D, δ denote the length of DIFS in the number of time slots and size of a time slot, respectively. Also, n is the number of active nodes. It is worth

$$R(j;\, i,\, k) = \begin{cases} (2^{k+j}-1)R(j;\, i-1,\, k) + (2^M - 2^{k+j-1})R(j-1;\, i-1,\, k); & i, j > 1 \\ (2^k - 1)R(0;\, i-1,\, k); & i > 1, j = 0 \\ 2^k - 1; & j = 0, i = 1 \\ 2^M - 2^k; & j = 1, i = 1 \end{cases} \tag{2}$$

$$R'(j;\, i,\, k) = \begin{cases} (k+j)R'(j;\, i-1,\, k) + (M-k-j)R'(j-1;\, i-1,\, k); & i, j > 1 \\ kR'(0;\, i-1,\, k); & i > 1, j = 0 \\ k; & j = 0, i = 1 \\ M-k; & j = 1, i = 1 \end{cases} \tag{5}$$

$$P_{\mathrm{SEP,RT}}(l) = \frac{p'_n(M-1;\, l-1,\, 0)\, p'_n(1;\, 1,\, M-1)}{p'_n(M;\, N,\, 0)}; \quad M \leq l \leq N \tag{6}$$

mentioning that in our scenario, when nodes have not transmitted their packets, they are saturated. Once a node sends its packet successfully, it does not participate in the future competitions. Thus, the number of saturated nodes (active nodes) changes during the exchange phase.

5.1 RLNC over a large finite field

If network coding is done over a large finite field, the encoded packets are innovative with high probability. As a result, the data transferring is finished when M nodes (vehicles) successfully send their packets to the others, because each node will have M independent coded packets and can retrieve all data packets.

Now, we observe the system at the end of a successful transmission, that is, after a VTT period. The number of active nodes decreases by one after a VTT period. Let $T(n, k)$ denotes the expected CDD when there are n active nodes and k innovative coded packets are necessary to retrieve the data generation (i.e. $M - k$ innovative packets have been received so far). Clearly, $T(n, k)$ is calculated by the following equation

$$T(n, k) = T_{\text{VTT}}(n) + T(n-1, k-1) \tag{8}$$

Note that during the content distribution process, the number of active nodes varies from N to $N - M+1$. Hence, the expected CDD is equal to $\sum_{i=N-M+1}^{N} T_{\text{VTT}}(i)$. Using (7), the expected CDD is derived by the following theorem.

Theorem 1: The mean value of CDD adopting network coding over a large finite field is given by

$$T_1 = \sum_{i=N-M+1}^{N} \frac{L' - (L'-1)(1-p)^i}{ip(1-ip)^{i-1}}\delta \tag{9}$$

5.2 Random XORed network coding

In this section, we assume that the info-station selects some of M data packets randomly. In other words, whenever it wants to send a new packet, XORs randomly selected packets and sends the coded packet to the vehicle. In contrast to coding over a large finite field, the randomly XORed packets may not be all innovative. In other words, it is not sufficient to send M packets successfully to complete data transferring, because some packets may not be innovative. We can represent the system at the end of each successful transmission with a two-dimensional Markov chain, where (n, k) represents the number of active nodes and the number of successfully received innovative packets, respectively. We define $T(n, k;M)$ as the expected time it takes to complete data transferring of M data packets, starting from state (n, k).

Theorem 2: The expected value of CDD when RXNC is applied is given by $T(N, 0;M)$, which is calculated recursively as (see (10))

Proof: Given that the system is in the state (n, k), after a successful packet transmission with average duration equal to $T_{\text{VTT}}(n)$, two events may occur: (a) The transmitted packet is linearly independent of k innovative packets that have been received so far with probability $p_n(1;1, k)$. Hence the mean residual time of CDD is $T(n-1, k+1;M)$. (b) The transmitted packet is not innovative, with probability $p_n(0;1, k)$. Consequently, the mean residual time of CDD is $T(n-1, k;M)$. Using (7) completes the proof. □

It is worth noting that computing $T(n, k;M)$ is initiated by calculating $T(1, 0;M)$ and $T(1, 1;M)$, directly from (10). The other terms are computed recursively.

5.3 Routing

We assume that whenever a vehicle passes the info-session, it receives one of the M data packets, randomly. In the exchange phase, the vehicles exchange the received packets in a similar way mentioned in previous sections. Note that some vehicles may have the same data packets. As a result, it may not be sufficient to exchange M packets. Similar to the analysis of RXNC, we provide a mathematical analysis to compute the transferring delay.

The mean of CDD is given as follows (see (11))

6 Extension to packet erasure channels

Up to now, we have assumed that a packet may be lost only due to collision. In this section, we incorporate the channel impairments in our model and address how it affects the success of the exchange phase. For simplicity, we assume i.i.d. packet erasure channels, that is, a transmitted packet may be corrupted at each destination independently with probability ε. Hence, a transmitted packet may be failed to be received at a destination by either concurrent packet transmissions (collision) or channel erasure.

The pseudo-broadcast method discussed in Section 3 ensures the packet failure detection at the intended neighbour. However, the reception status of the other nodes are ignored. Consequently, during the content distribution process, some nodes may not be able to receive all transmitted packets. Hence, at the end of the exchange phase only a fraction of nodes will retrieve all data packets. To investigate the effect of packet erasure on our content

$$T(n, k; M) = \begin{cases} \dfrac{L' - (L'-1)(1-p)^n}{np(1-p)^{n-1}}\delta + \sum_{j=0}^{1} p_n(j; 1, k)T(n-1, k+j, M); & n > 0, k < M \\ 0; & \text{otherwise} \end{cases} \tag{10}$$

$$T(n, k; M) = \begin{cases} \dfrac{L' - (L'-1)(1-p)^n}{np(1-p)^{n-1}}\delta + \sum_{j=0}^{1} p'_n(j; 1, k)T(n-1, k+j; M); & n > 0, k < M \\ 0; & \text{otherwise} \end{cases} \tag{11}$$

IET Commun., 2014, Vol. 8, Iss. 9, pp. 1447–1458
doi: 10.1049/iet-com.2013.0611

distribution algorithm, we define a performance metric, successful content delivery probability, P_{SCD} as the probability that all nodes are able to retrieve all data packets at the end of the content distribution period. For the sake of simplicity, we only consider the case of RLNC over a large finite field. The probability of packet reception by the tagged destination (i.e. the node that is determined to send the ACK) equals one, as we assume that the transmitter continues to retransmit the failed packet until the tagged node successfully receives the packet. Now, we compute the probability of successful reception of a packet destined to a tagged node at an overhearing node. We observe the system after each VTT, referred to as an observation epoch, such that the packet is sent without collision. Hence, the packet is received by each overhearing node independently with probability $1-\varepsilon$ at each observation epoch. The successful reception at an overhearing node occurs by one of the two following events: (i) At the first observation epoch, the packet is received by the overhearing node. (ii) The packet transmissions to both overhearing node and tagged node are failed in a set of consecutive epochs and the overhearing node receives the packet in the next epoch (irrespective of the reception status of the tagged node). Consequently, we obtain the probability of successful reception for an overhearing node as

$$P_1 = (1-\varepsilon) + \varepsilon^2(1-\varepsilon) + \varepsilon^4(1-\varepsilon) + \cdots = \frac{1}{1+\varepsilon} \quad (12)$$

The probability of successful packet reception by a specific node is determined by the probability that the node is either a tagged or an overhearing node, which in turn depends on the detail of content distribution process. Clearly, a lower bound for the probability of successful reception at a specific node equals P_1. The probability that a node is able to retrieve the data generation depends on whether it has contributed in the exchange phase or not (i.e. if it has sent its packet or not). Let P_A and P_B denote the former and the latter probabilities, respectively. A non-contributing node can acquire one more packet compared to a contributing one, taking its own packet into account.

To compute P_{SCD}, we focus on the M coded packets that are successfully received by intended destinations (referred to as tagged coded packets). More precisely, due to channel erasure, a contributing node may retransmit its coded packet several times in order to receive an ACK. For each contributing node, the last transmitted coded packet which is followed by an ACK is referred to as a tagged coded packet. A contributing node is able to retrieve the data generation by receiving all the tagged coded packets. A non-contributing node, however, requires at least $M-1$ coded packets. Note that if the two following conditions are satisfied, the exchange phase will finish successfully: (i) Each contributing node has received all $M-1$ tagged coded packets at the end of exchange phase; (ii) Each non-contributing node has received at least $M-1$ tagged coded packets at the end of exchange phase. Packet loss may result in a case that more than M coded packets are received by some nodes. In other words, some nodes may receive non-tagged packets beside the tagged ones and use these packets to retrieve the data generation. Hence, the aforementioned conditions for retrieving the data generation are sufficient (not necessary). Consequently, lower bounds for P_A and P_B can be obtained as

$$P_A \geq P_1^{M-1} \quad (13)$$

$$P_B \geq \binom{M}{M-1} P_1^{M-1}(1-P_1) + P_1^M \quad (14)$$

Note that if $\varepsilon \ll 1$, the probability of receiving non-tagged packets are negligible and we can only consider the tagged packets. Thus, at the end of an exchange phase, there are M contributing and $N-M$ non-contributing nodes. A lower bound for P_{SCD} is derived as

$$P_{\text{SCD}} \geq P_1^{M(M-1)} \left[\binom{M}{M-1} P_1^{M-1}(1-P_1) + P_1^M \right]^{N-M} \quad (15)$$

Now, we assess how the packet erasure affects the expected CDD. To compute $T(n, k)$ we note that at the end of a VTT, one node succeeds to transmit its packet without collision. The packet is received by each node with probability $1-\varepsilon$. If the packet is received by the intended destination, the transmitting node becomes inactive. Thus, the remaining time to finish the exchange phase is $T(n-1, k-1)$. Otherwise, the remaining time is $T(n, k)$. Hence, we can obtain the following equation

$$T(n, k) = T_{\text{VTT}}(n) + (1-\varepsilon)T(n-1, k-1) + \varepsilon T(n, k) \quad (16)$$

Thus, we infer

$$T(n, k) = \frac{T_{\text{VTT}}(n)}{1-\varepsilon} + T(n-1, k-1) \quad (17)$$

Using (7) and (17), we obtain

$$T_2 = \sum_{i=N-M+1}^{N} \frac{L' - (L'-1)(1-p)^i}{(1-\varepsilon)ip(1-ip)^{i-1}} \delta \quad (18)$$

where T_2 is the mean of CDD. Comparing (18) and (9) shows that CDD does not vary significantly, assuming that $\varepsilon \ll 1$.

It is worth noting that in RXNC and traditional routing, the probability of file retrieval depends on the details of the received packets as well as the contribution of the node in content distribution. The number of contributing nodes is also variable. Hence, obtaining accurate bounds for probability of successful content distribution is very difficult in these scenarios. Nevertheless, for the case of RXNC and traditional routing, to assess the effect of channel erasure on the success of content distribution, we have added simulation results in Section 7.

Table 2 Simulation parameters

Parameters	Values
distance between two info-stations	2 km
transmission range	200 m
file size	12 MB
data generation size	600 B
MAC protocol	IEEE 802.11
transmission rate	12 Mbps
slot time	20 μs

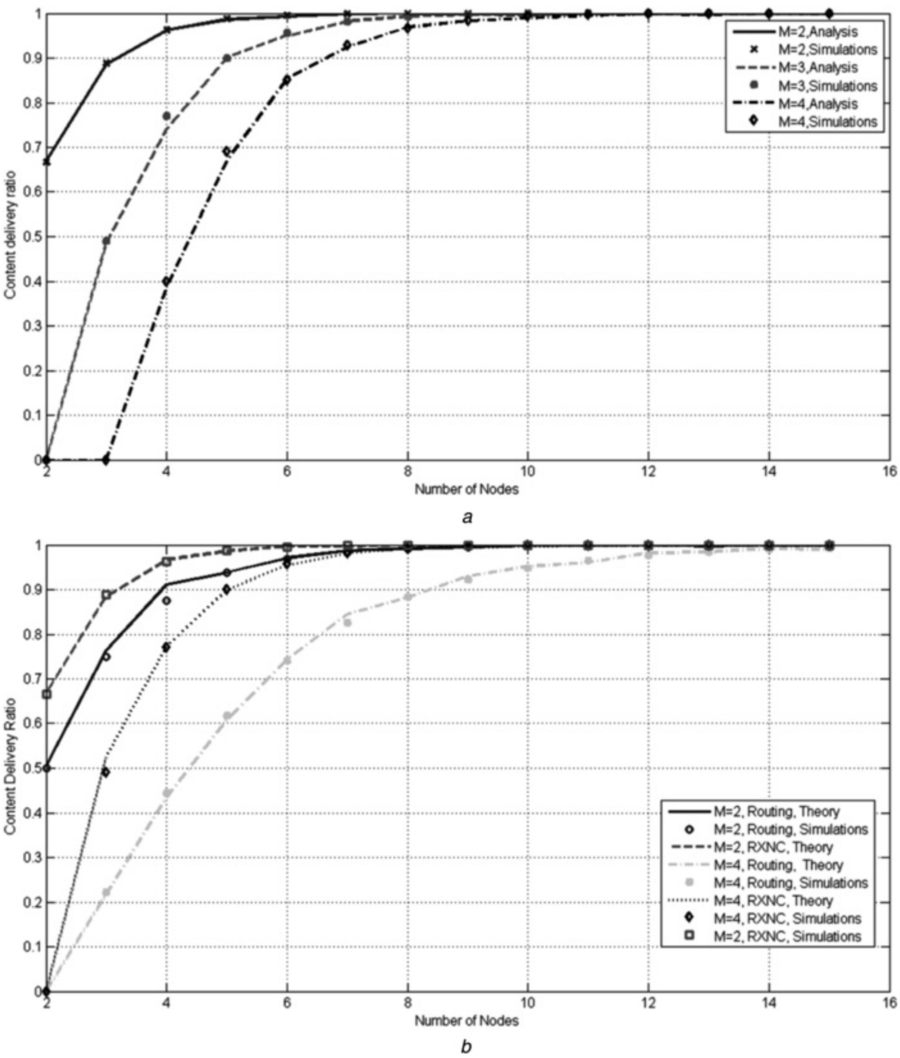

Fig. 3 *Content delivery ratio against number of nodes*

a RXNC scheme

b Comparison of RXNC and routing schemes

7 Numerical results

In this section, we illustrate the efficiency of our theoretical framework and provide simulation results to validate our analysis. The parameters used in the numerical evaluation are shown in Table 2. In order to evaluate the accuracy of our analytical model for 802.11 MAC, we adopt 802.11 MAC protocol with binary exponential back-off in our simulation. The contention window size is initiated by CW_{min} and doubled each time a collision occurs. After a successful transmission, the contention window size is reset to CW_{min}. Our simulations are done in a MATLAB environment.

Fig. 3*a* shows the effect of velocity of vehicles on content delivery ratio. As the velocity increases, the access point has less time to communicate with moving vehicles. Thus, a data generation should be split to smaller parts, resulting in a larger M. For the case of RXNC, to achieve a CDR higher than 0.98, the number of nodes should be at least equal to 5, 7 and 9 for $M = 2$, 3 and 4, respectively. Hence, as the velocity increases, a larger cluster size (i.e. N) is required to achieve a successful content distribution for RXNC. Note that for RLNC over a large finite field M nodes are sufficient. Thus, there is a compromise between complexity of network coding and cluster size. Fig. 3*b* compares the content delivery ratio for RXNC and routing for $M = 2$, 4. The difference is larger for $M = 4$. Consequently, the gain of network coding improves in higher velocity. For example, to achieve the ratio higher than 0.98, the cluster size should exceed 12 and 7 for routing and RXNC, respectively.

The PMF of a successful exchange phase, in the case of RXNC, is depicted in Fig. 4*a* for $M = 2$, 4. Note that in both cases, the maximum probability is obtained for $l = M$, where l denotes the number of successful transmissions. Fig. 4*b* illustrates the PMF of a successful exchange phase for routing. Compared to RXNC, the PMF for routing is more flat. Moreover, the maximum may not be equal to M, as in RXNC.

CDD is computed using the proposed analytical method and the results are compared with the simulations. Fig. 5 compares the average CDD for RLNC over a large finite field, RXNC and routing. Network coding provides significant performance improvement over traditional routing; however, RXNC exhibits 30% additional delay compared to RLNC over a large finite field. Hence, there is a compromise between CDD and network coding complexity. Network coding provides significant gain when the number of blocks, M, is much less than the number of nodes, N. However, the coding gain decreases as M

doi: 10.1049/iet-com.2013.0611

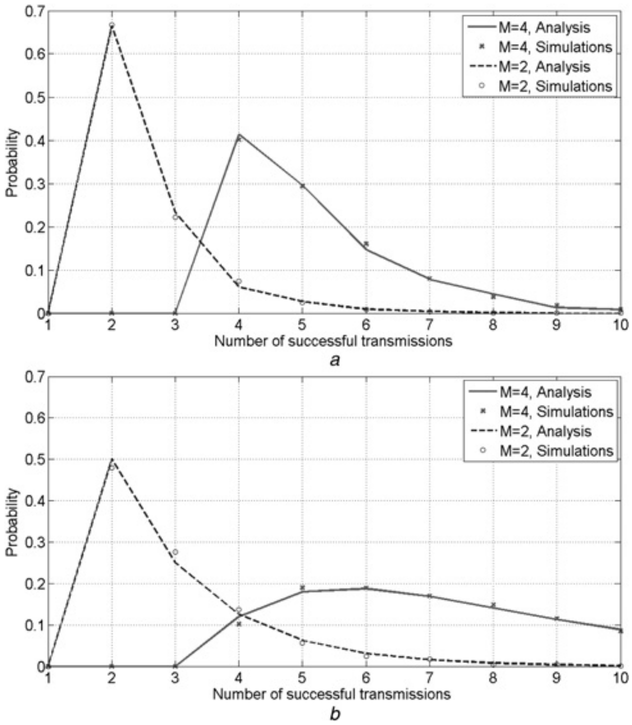

Fig. 4 *PMF of a successful exchange phase, N = 10*

a RXNC scheme
b Routing scheme

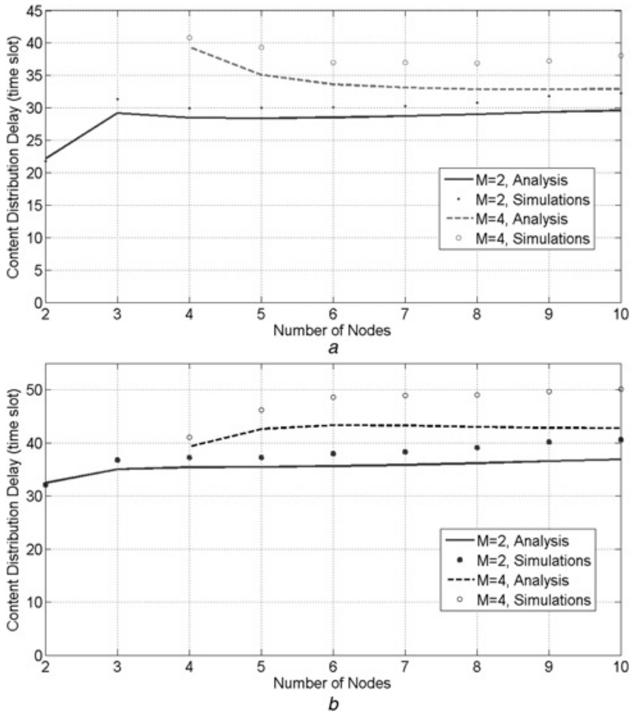

Fig. 6 *Expected CDD*

a RLNC over a large finite field
b RXNC, $CW_{min} = 16$

approaches *N*. The effect of vehicles' velocity on average CDD in RLNC over a large finite field and RXNC have been investigated in Figs. 6*a* and *b*, respectively. In both cases, the analytical results follow the trend of simulation results. However, the approximated modelling of 802.11 MAC protocol results in a difference between simulations and analysis that varies between 1 and 17%. It is worth mentioning that higher velocity (larger *M*) results in more delay. Hence, in order to achieve the same CDD, the cluster size should be reduced in higher velocities. Fig. 7*a* illustrates the effect of packet erasure probability on the content delivery success, when RLNC over a large finite field is applied. If the packet erasure probability is very low (below 0.001), the file recovery is almost always successful. The channel degradation reduces the ability of nodes to retrieve the data generation. For example, when the packet erasure probability is 0.01, only 90% of nodes succeed to fully retrieve the data generation. Note that the analytical bound (15) is very close to its actual value.

Fig. 7*b* compares the success of content distribution for RXNC and traditional routing methods. Although the performance of both the methods degrades as the packet erasure probability increases, RXNC-based method outperforms the routing-based method significantly. The lower performance of the latter is due to high correlation among the received packets. Moreover, comparing Figs. 7*a* and *b* shows that the performance of RLNC-based and RXNC-based methods are very similar.

In our analytical model, we have assumed that the cluster of vehicles is unchanged during the exchange phase. To assess

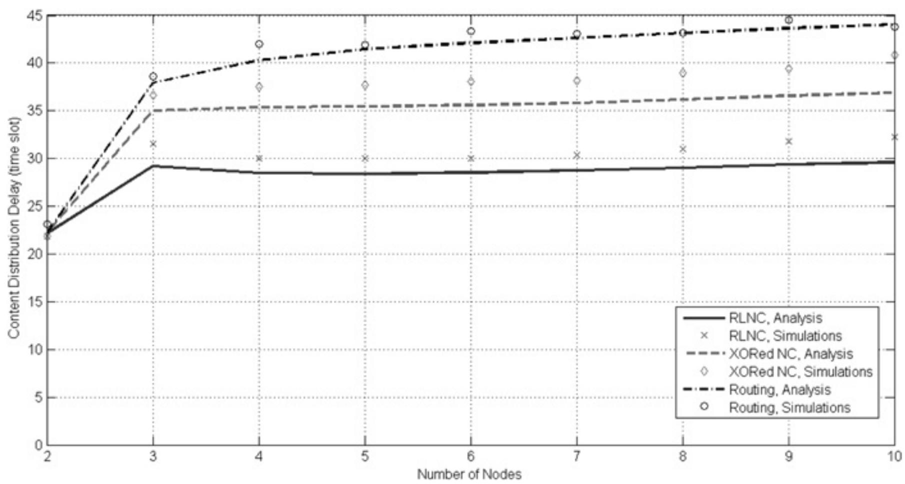

Fig. 5 *Comparison of CDD among different schemes: $CW_{min} = 16$, M = 2*

doi: 10.1049/iet-com.2013.0611

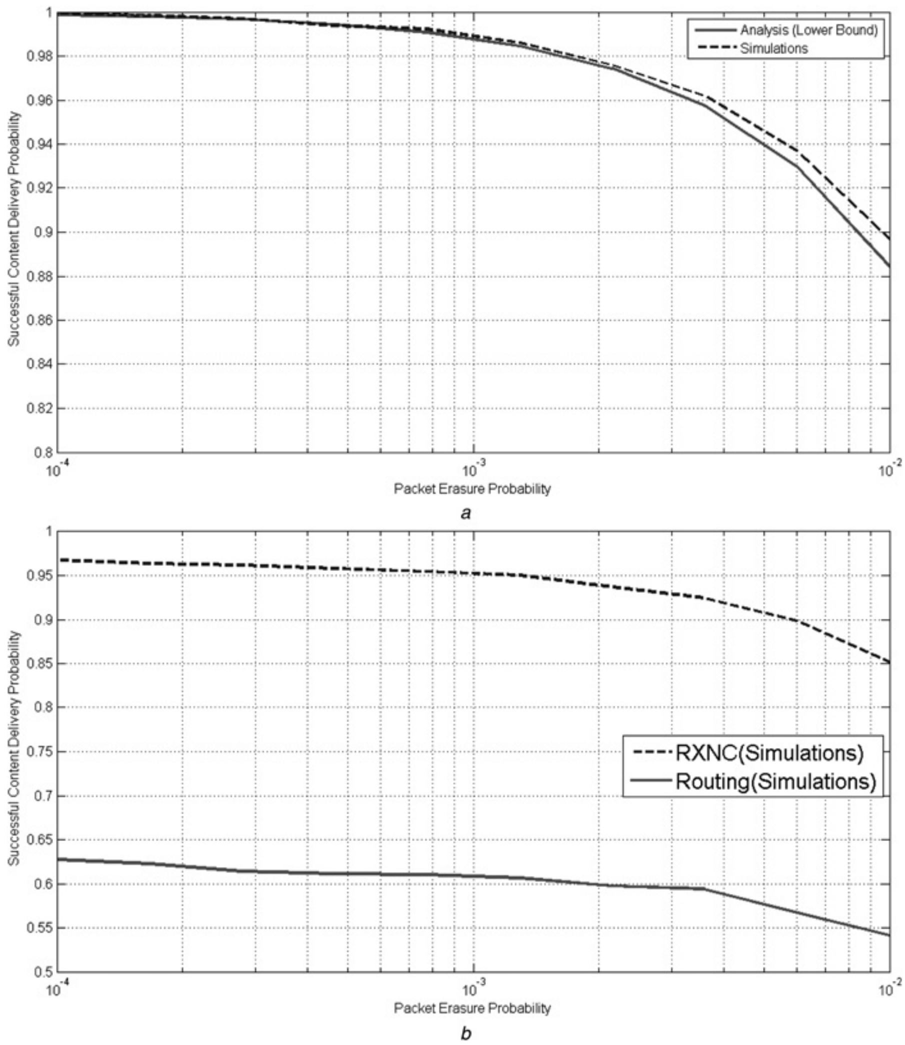

Fig. 7 *Successful content delivery probability against packet erasure probability, N = 8, M = 4*

a RLNC over a large finite field
b RXNC and routing

the validity of this assumption, we considered three scenarios in Fig. 8. In scenario A, the vehicles' velocities vary in the range 15–20 m/s. In scenarios B and C, the range of velocities is $15-25$ and $15-30$ m/s, respectively. In each scenario, the size of cluster is 100 m at the start of exchange phase. Hence, during the exchange phase, due to difference between the velocities, the distance between the vehicles is changed which affects the network topology (the network is not single-hop anymore, if some vehicles exit the transmission range of the others) as well as the performance of the content distribution. We adopt IEEE 802.11 MAC in the simulation. Fig. 8*a* shows that the CDR of RLNC and RXNC are acceptable for the three scenarios. However, in routing, the CDR is degraded considerably. Thus, RLNC and RXNC can tolerate wider variety of velocities than routing, since each vehicle can help the other vehicles in its transmission range to retrieve the data. However, for routing, the vehicles will be able to retrieve the whole data provided that they receive some specific packets. Hence, if the connection to specific vehicles are failed, it cannot retrieve the data generation. Fig. 8*b* demonstrates the CDD for the three scenarios. As expected, scenario C exhibits the largest CDD, since the connection failure is high due to fast topology changing during the exchange phase.

8 Conclusions

In this paper, we analysed a two-phase network coding-based content distribution in a single-hop VANET. In the first phase, the info-station divides the data file into data blocks and transmits coded data packets, constructed from the data blocks, to vehicles passing the info-station. In the second phase, the vehicles collaborate to exchange the received coded data packets based on an IEEE 802.11 MAC to retrieve the data file. We employed two network coding schemes: RLNC over a large finite field and RXNC. We devised a theoretical framework to compute the CDD for both network coding schemes and compared the results with routing-based content distribution. Our results showed that using RLNC over a large finite field is superior to RXNC in terms of expected CDD, although it introduces additional overhead in packet headers. However, both network coding schemes expedite the content distribution process, compared to routing-based method. It was recognised that the expected CDD is a function of number of vehicles as well as the number of data blocks. We also quantified the effect of correlation among received coded packets on content distribution success by characterising content distribution ratio as the probability that vehicles are

doi: 10.1049/iet-com.2013.0611

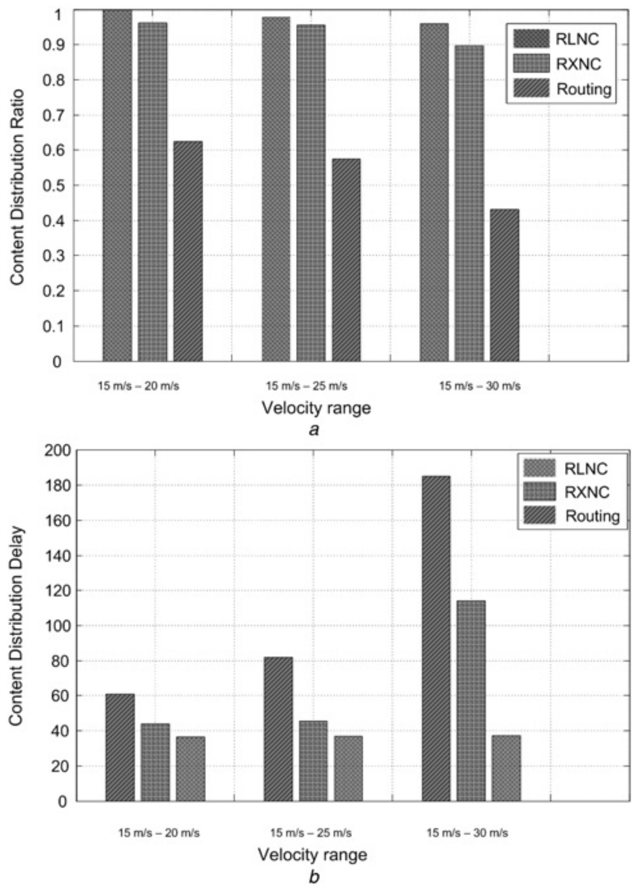

Fig. 8 *Effect of vehicles' velocity range on the performance of the content distribution methods: N = 8, M = 4*

a CDR
b CDD

able to retrieve the data generation at the end of the exchange phase. Finally, we addressed the impact of channel non-idealities on the success of content distribution process, for RLNC over a large finite field. To assess our analysis, we also performed the simulations. The comparison between simulation and theoretical results confirmed the accuracy of our theoretical framework.

Considering the vehicles speed and mobility patterns in the analytical model as well as investigating their impact on the performance analysis is one of our future works. Moreover, extending the analytical results in Section 6 to the case of RXNC method is also interesting. Furthermore, we addressed IEEE 802.11 MAC in our paper; hence an interesting future work is to provide an analytical model for content distribution based on IEEE 802.11p protocol.

9 References

1 Fiore, M., Barcelo-Ordinas, J.: 'Cooperative download in urban vehicular networks'. IEEE Sixth Int. Conf. Mobile Ad hoc and Sensor Systems, MA, October 2009
2 Park, J.-S., Lee, U., Oh, S.Y., Gerla, M., Lun, D.S.: 'Emergency related video streaming in VANET using network coding'. Third Int. Workshop on Vehicular Ad Hoc Networks, 2006, pp. 102–103
3 Park, J.-S., Lee, U., Gerla, M.: 'Vehicular communications: emergency video streams and network coding', *J. Internet Services Appl.*, 2010, **1**, (1), pp. 57–68
4 Ahlswede, R., Cai, N., Li, S.-Y., Yeung, R.: 'Network information flow', *IEEE Trans. Inf. Theory*, 2000, **46**, (4), pp. 1204–1216
5 Xuan, Y., Lea, C.-T.: 'Network-coding multicast networks with QoS guarantees', *IEEE/ACM Trans. Netw.*, 2011, **19**, (1), pp. 265–274
6 Ho, T., Medard, M., Koetter, R., *et al.*: 'A random linear network coding approach to multicast', *IEEE Trans. Inf. Theory*, 2006, **52**, (10), pp. 4413–4430
7 Katti, S., Rahul, H., Hu, W., Katabi, D., Medard, M., Crowcroft, J.: 'XORs in the air: practical wireless network coding', *IEEE/ACM Trans. Netw.*, 2008, **16**, (3), pp. 497–510
8 Eryilmaz, A., Ozdaglar, A., Medard, M., Ahmed, E.: 'On the delay and throughput gains of coding in unreliable networks', *IEEE Trans. Inf. Theory*, 2008, **54**, (12), pp. 5511–5524
9 Firooz, M.H., Roy, S.: 'Data dissemination in wireless networks with network coding', *IEEE Commun. Lett.*, 2013, **17**, (5), pp. 944–947
10 Yang, Z., Li, M., Lou, W.: 'Codeplay: Live multimedia streaming in VANETs using symbol-level network coding'. 18th IEEE Int. Conf. Network Protocols (ICNP), October 2010, pp. 223–232
11 Li, M., Yang, Z., Lou, W.: 'CODEON: Cooperative popular content distribution for vehicular networks using symbol level network coding', *IEEE J. Sel. Areas Commun.*, 2011, **29**, (1), pp. 223–235
12 Ahmed, S., Kanhere, S.S.: 'VANETCODE: network coding to enhance cooperative downloading in vehicular ad hoc networks'. Int. Conf. Wireless Communications and Mobile Computing, ser. IWCMC '06, 2006, pp. 527–532
13 Johnson, M., Chen, M., Ramchandran, K.: 'On the role of network coding in uncoordinated multihop content distribution over ad hoc networks'. Int. Symp. Network Coding (NetCod), July 2011, pp. 1–6
14 Nicrovanachumrus, P., Choochaisri, S., Intanagonwiwat, C.: 'Efficiency and speed improvement for network coded file distribution with adaptive request'. IEEE 13th Int. Conf. Communication Technology (ICCT), September 2011, pp. 944–949
15 Lee, U., Park, J.-S., Yeh, J., Pau, G., Gerla, M.: 'Code torrent: content distribution using network coding in VANET'. First Int. Workshop on Decentralized Resource Sharing in Mobile Computing and Networking, ser. MobiShare '06, 2006, pp. 1–5
16 Lee, S.-H., Lee, U., Lee, K.-W., Gerla, M.: 'Content distribution in VANETs using network coding: the effect of disk I/O and processing O/H'. Fifth Annual IEEE Communications Society Conf. Sensor, Mesh and Ad Hoc Communications and Networks, June 2008, pp. 117–125
17 Beraldi, R., Alnuweiri, H.: 'Delay tolerant podcasting with network coding'. Wireless Days (WD), IFIP, October 2010, pp. 1–5
18 Asterjadhi, A., Fasolo, E., Rossi, M., Widmer, J., Zorzi, M.: 'Toward network coding-based protocols for data broadcasting in wireless ad hoc networks', *IEEE Trans. Wirel. Commun.*, 2010, **9**, (2), pp. 662–673
19 Yan, Q., Li, M., Yang, Z., Lou, W., Zhai, H.: 'Throughput analysis of cooperative mobile content distribution in vehicular network using symbol level network coding', *IEEE J. Sel. Areas Commun.*, 2012, **30**, (2), pp. 484–492
20 Ye, F., Roy, S., Wang, H.: 'Efficient data dissemination in vehicular ad hoc networks', *IEEE J. Sel. Areas Commun.*, 2012, **30**, (4), pp. 769–779
21 Firooz, M., Roy, S.: 'Collaborative downloading in VANET using network coding'. IEEE Int. Conf. Communications (ICC), 2012, pp. 4584–4588
22 Das, S., Nandan, A., Pau, G.: 'SPAWN: a swarming protocol for vehicular ad-hoc wireless networks'. First ACM Int. Workshop on Vehicular Ad Hoc Networks, 2004, pp. 93–94
23 Gupta, P., Kumar, P.: 'The capacity of wireless networks', *IEEE Trans. Inf. Theory*, 2000, **46**, (2), pp. 388–404
24 'IEEE Standard for Information Technology – Telecommunications and information exchange between systems – local and metropolitan area networks – specific requirements – Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications', ANSI/IEEE Std 802.11, 1999 Edition (R2003), 2003, pp. i–513
25 Bianchi, G.: 'Performance analysis of the IEEE 802.11 distributed coordination function', *IEEE J. Sel. Areas Commun.*, 2000, **18**, (3), pp. 535–547
26 Cali, F., Conti, M., Gregori, E.: 'Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit', *IEEE/ACM Trans. Netw.*, 2000, **8**, (6), pp. 785–799
27 Yedavalli, K., Krishnamachari, B.: 'Enhancement of the IEEE 802.15.4 MAC protocol for scalable data collection in dense sensor networks'. Sixth Int. Symp. Ad Hoc, and Wireless Networks and Workshops, April 2008, pp. 152–161

10 Appendix

10.1 Appendix 1: Proof of Lemma 1

We can represent a coded-packet with a vector composed of binary elements and of length M. Let Γ be the set of these vectors (we exclude the all-zero vector). Let $R(j;i, k)$ denotes the ways we can select i vectors $V_1, \ldots, V_i \in \Gamma$, such that exactly j vectors are independent from a specific k linearly independent vectors. It is equivalent to the number of binary matrices of dimension $(i+k) \times M$ and rank $j+k$ with given k independent rows. Clearly, by this definition, p_n can be calculated from (1). Equation (2) is derived based on the fact that the matrix can be constructed by one of the two following ways: (a) Take a $(i+k-1) \times M$ matrix with rank $j+k$ and add a dependent row. (b) Take a $(i+k-1) \times M$ matrix with rank $j+k-1$ and add an independent row. We can select a non-zero vector linearly dependent with the given $j+k$ independent vectors in

$$\binom{k+j}{1} + \binom{k+j}{2} + \cdots + \binom{k+j}{k+j} = 2^{k+j} - 1 \quad (19)$$

ways. The number of non-zero vector linearly independent with the given $j+k-1$ independent vectors will be

$$2^M - 1 - \binom{k+j-1}{1} - \binom{k+j-1}{2} - \cdots - \binom{k+j-1}{k+j-1} = 2^M - 2^{k+j-1} \quad (20)$$

Hence, (2) immediately infers.

10.2 Appendix 2: Proof of Lemma 2

The proof is very similar to the previous section. Let $R'(j;i, k)$ denotes the ways we can select i data packets such that exactly j packets are new, provided that we already have k distinct data packets. These packets are chosen in one of the following ways: (a) Choose $i-1$ packets such that j packets are new (i.e. no packet among k distinct ones) and add a repetitive packet. (b) Choose $i-1$ packets such that $j-1$ packets are new and add a new packet. Then, (5) is derived.

IET Commun., 2014, Vol. 8, Iss. 9, pp. 1447–1458
doi: 10.1049/iet-com.2013.0611