

# Efficient Inter-Vehicle Data Dissemination

Fei Ye, Sumit Roy, Haobing Wang

Dept. of Electrical Engineering

University of Washington

Seattle, WA 98195

{fye, sroy, hbwang89}@u.washington.edu

**Abstract**—Data services for in-vehicle consumption are expected to become a primary driver in the development of future vehicular networks. Due to download rate limitations of present wide-area cellular connectivity such as 3G (the likely ‘pipe’ to/from vehicles), direct peer-to-peer data sharing among vehicles can supplement vertical downloading. This paper studies the inter-vehicle data dissemination problem in a WAVE/802.11p vehicular ad hoc network, using network coding (NC). The gains from such NC, relative to a baseline scheme of random transmission, is estimated using analysis and simulations.

**Index Terms**—Data dissemination, vehicle-to-vehicle, ad hoc network, network coding, coupon collection problem.

## I. INTRODUCTION

Recent research, development and standardization advances in vehicular ad hoc networks (VANETs) [1]–[4] have motivated increasing interest in various data services for in-vehicle consumption, such as ‘commerce- and entertainment-on-the-wheel’. These include a variety of push/pull applications: local information (e.g., traffic notification, map updates) sent to vehicles; or timely data from the Internet (e.g., parking, notifications of events/attractions etc.). Vertical data downloads to users in vehicles can occur by existing wide-area cellular infrastructure (3G/4G) and/or proposed new roadside infrastructure based on short-range Dedicated Short Range Communications (DSRC) links. However, both these approaches have their own challenges: the modest data rate of present 3G links and the cost of large data downloads to the end-user under current data pricing regimes on one hand, and the intermittent hot-spot type roadside coverage envisaged with DSRC on the other (although DSRC can support upto peak channel data rate of 27 Mbps). In other words, at any time, only a few vehicles may possess content that is potentially desired by many. This leads to the following premise for content dissemination based on ad-hoc vehicular networking (see Fig. 1); using peer-to-peer modes for content distribution using vehicle-to-vehicle (V2V) ad hoc communications is both time and cost efficient.

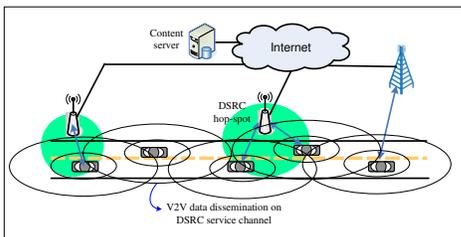


Figure 1. Inter-vehicle data dissemination supplements direct downloading.

For simplicity, we assume that a reference vehicle successfully downloads a large file via vertical network access, which it may then proceed to share with other vehicles in geographic vicinity (and possibly members of a previously established affinity group) through V2V mode. For the purposes of lightweight data exchange, packets in WAVE mode are transmitted as 1-hop broadcast with no feedback, as defined in 802.11p [1]. The efficiency of such broadcast communication among vehicles is studied in [5] and [6]. In order to improve latency, WAVE does provide priority mechanisms for propagation of emergency warning messages via multi-hop relaying [7], [8]; however, these forwarding mechanisms are not suitable for sharing large amount of non-prioritized content.

Due to the absence of feedback in WAVE broadcasts, hop-by-hop forwarding with Automatic Repeat reQuest (ARQ) is infeasible. Random broadcast - broadcasting a random packet from sender’s buffer at a transmission opportunity - is a baseline approach to disseminate data using V2V ad hoc communications. However, it’s inefficiency is illustrated in Fig. 2. Two vehicles  $N_2$  and  $N_3$  within transmission range of the source are waiting to receive two packets  $X_1$  and  $X_2$  from  $N_1$ . At this time,  $N_2$  has only  $X_2$ , while  $N_3$  has only  $X_1$  from prior transmissions. Even with lossless wireless links,  $N_1$  needs two (more) broadcast transmissions to complete the transfer *if packets are sent individually*. However, it is evident that when  $N_1$  is broadcasting  $X_1$ ,  $N_3$  does not benefit from overhearing this transmission; similarly for  $N_2$  when  $N_1$  is broadcasting  $X_2$ . Network coding leverages the above unexploited redundancies due to broadcast nature of wireless for more efficient data dissemination. Node  $N_1$  only needs to broadcast  $Y = X_1 \oplus X_2$ , then both  $N_2$  and  $N_3$  can both extract the desired packet after one broadcast via  $Y \oplus X_1$  ( $Y \oplus X_2$ ), respectively.

$X_1$  or  $X_2$  or  $X_1 \oplus X_2$  ?

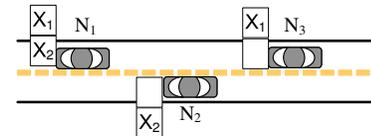


Figure 2. Network coding helps to achieve maximum data dissemination efficiency in a 3-node example.

Successful application of linear network coding [9] to a realistic VANET scenario requires going beyond the simple scenario of Fig. 2 and confronting non-trivial challenges such

as: Any multi-hop scenario between a source-destination pair requires prior route formation; this is not supported in WAVE broadcast.

This paper studies a canonical inter-vehicle data dissemination problem - 'pushing' content to vehicles in the +x direction from a source at one end of the platoon (Fig. 3), in the presence of *link unreliability*.<sup>1</sup> All participating vehicles learn their index numbers in this platoon by ordering their geography locations, i.e., from its own GPS devices and neighbors' beacon messages. We ignore any multi-access protocol aspects of the problem, by assuming an ideal scheduler; i.e., a single (source) node acts as the role of the disseminator at any given time and all other nodes in the forward direction are potential receivers with reception probabilities  $Q_1 > Q_2 > Q_3 > \dots > 0$ , starting from the nearest recipient. The instant broadcasting (source) node keeps transmitting till the *immediate next* node has collected the full data set. The broadcasting role is then handed-off to the first recipient at no MAC overhead in this ideal scheduler<sup>2</sup>. Namely, the *immediate next* node (relative to the current) assumes the role of the disseminator, once it has obtained the full data-set. Our contributions are two-fold:

- We mathematically characterize the completion time of a single dissemination step - pushing data set from instant source node  $k$  to node  $k+1$  - with and without network coding.
- We investigate via simulations the gains from network coding aided scheme relative to 1) random broadcast, and 2) a perfect feedback scheme, for more general scenarios.

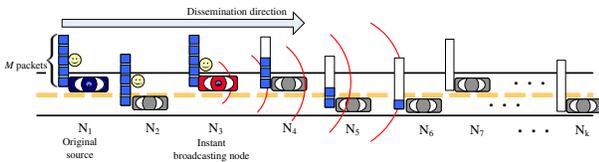


Figure 3. Data dissemination in a 1-D VANET.

## II. RELATED WORK

Peer-to-peer collaboration via vehicular communication is explored in [10], [11] for content sharing. SPAWN [10] extends wired Internet peer-to-peer content distribution (a.k.a. file swarming) protocols (e.g., BitTorrent) to VANET. However, the assumption of reasonable transport layer bandwidth (either TCP or UDP over multi-hop wireless links) in [10] does not fit into the WAVE broadcast paradigm. Similar observation was made in [11], for CodeTorrent, a network coding aided file swarming protocol with no underlying routing support.

[11]–[13] all study linear network coding in VANET settings. Our work and [11] are similar in that both apply network coding to MAC broadcasting. However, in [11],

<sup>1</sup>If the original source node is in the middle of a platoon, this can be treated as 2 data dissemination processes towards +x(-x) directions. Appropriate multi-access interference needs to be considered in this two-sided case.

<sup>2</sup>In an 802.11p MAC, a potential broadcasting node  $u$  can prioritize its first few coded packets using a much smaller contention window. The current disseminator terminates its broadcast upon receiving  $u$ 's transmission of the same data set. Then  $u$  will restore its contention window back to normal.

source nodes send out a file description and other nodes have to keep requesting/pulling neighbors for packets, while in our work, file is pushed to all nodes in a linear platoon in contrast. In the relaying protocol devised by [13] - CodeOn, nodes share 'content reception status', and then choose a transmission backoff delay that is inversely proportional to the amount of useful content they possess. As a result, the node with the most useful content gets to relay at the end of relay selection process. Symbol level network coding is conducted at the access point and at all relay nodes. While [13] explores the design via simulation, our work provides *analytical results* for V2V data dissemination under a simple scheduling model. [12] analyzes the multicast throughput of network coding aided content distribution in a linear VANET. Their result is based on ideal reception and collision models, in which only isolated packet transmission within a fixed distance to the receiver is received with probability 1, otherwise 0. We consider probabilistic links.

## III. SYSTEM MODEL

### A. Network Coding Operations

Linear Network Coding (LNC) is a block code conducted over finite field  $\mathbb{F}_q$ . Every packet consisting of  $\log(q)$ -bit symbols  $x_i$ , can be treated as a vector in  $\mathbb{F}_q$ . During any transmission opportunity, source broadcasts a following linear combination of all  $M$  packets it has stored,  $x_1, x_2, \dots, x_M$ :

$$y_i = \sum_{k=1}^M \gamma_k x_k, \quad (1)$$

where  $\gamma_k$ 's are random network coding coefficients selected uniformly in  $\mathbb{F}_q$ . Any received coded packet at a receiver can be expressed as a linear combination of the  $M$  original packets in the network, regardless of the actual sender of that packet. Specifically, if a node receives  $K$  coded packets  $y_1, \dots, y_K$ , these can be written in the following vector product form:

$$Y_{K \times 1} = \mathbf{A}_{K \times M} X_{M \times 1} \quad (2)$$

where  $\mathbf{A}_{K \times M}$  is the coefficient matrix containing all coding coefficients. In the case  $K \geq M$ , and matrix  $\mathbf{A}$  is full rank, and receiver can recover the data  $X = \mathbf{A}^{-1}Y$ .

For decoding, clearly the receiver needs to know the network coding coefficients  $\mathbf{A}$  used during all transmissions. We allow the transmitter to embed these coefficients in each packet. For instance, linear network coding of  $M$  packets over finite field  $\mathbb{F}_q$  needs  $M \log_2(q)$ -bit long coefficients in each packet, which represents 3.3% overhead when  $M = 100$ ,  $q = 16$ ,  $L = 1500$  Bytes. A newly received coded packet is called *innovative* if its coding coefficient increases the rank of  $\mathbf{A}$ , otherwise the packet is discarded. It has been shown in [14] that if NC coefficients are randomly chosen from  $\mathbb{F}_q$ , each received coded packet is innovative with high probability; hence, a node can decode the complete data set when it accumulates  $M$  packets with high probability.

## B. Network Model

We model the vehicle platoon as a static one-dimensional lattice network with equal inter-vehicle spacing. Initially, the source node has  $M$  packets denoted by  $x_1 \dots x_M$  that are to be sent to all other vehicles in the forward direction (see Fig. 3). Time is slotted so that each packet transmission consumes exactly one time slot. Each broadcast transmission is received by nodes in the forward direction with probabilities  $Q_1 > Q_2 > \dots > 0$ , starting from the first recipient relative to the disseminator. This reception probability sequence is a function of communication and topology parameters, i.e., transmit power, inter-vehicle spacing, channel path loss, etc., and is considered in simulations in Sec. V.

## IV. THE GAIN OF NETWORK CODING - ANALYZING A SINGLE DISSEMINATION STEP

According to the ideal scheduling rule, the broadcasting role is handed-off node by node to the forward direction, and so is the complete data set. A *single dissemination step* is the duration from a node becomes disseminator to the moment when the role is handed-off to its immediate neighbor. Assume node  $k$  (which has all  $M$  packets) gains the broadcasting role at the beginning of a step, its immediate neighbor - node  $k+1$  - has already received partial data set from prior broadcasts. Conditioned on the size of this partial data set at node  $k+1$ , we mathematically derive the probability mass function (PMF) of the completion time of a single step, with and without network coding. In both cases, reception probability at the receiver is  $Q_1$ ,  $0 < Q_1 < 1$ .

**Random scheme (RND):** In RND scheme, the broadcasting node keeps transmitting one random packet per slot till its immediate neighbor has all packets. Let  $T_{RND}$  denote the completion time for  $k$  to fill up  $(k+1)$ 's buffer, which already contains  $c$  distinct packets. Due to link failure, not every broadcast is heard by  $k+1$ . For a packet heard by  $k+1$ , the probability of it being new is  $\frac{M-c-j+1}{M}$ ,  $\forall j \in \{1, \dots, M-c\}$ , which decreases quickly as the buffer approaches full. Calculating  $T_{RND}$ 's PMF is essentially a generalization of classic *coupon collection problem (CCP)* studied in [15]. Conditioned on  $c$ ,  $T_{RND}$ 's PMF is given in the following theorem:

**Theorem 1.** *Initially, source has  $M$  packets, receiver has a subset of  $c$  distinct packets. The wireless link in between has a success probability of  $Q_1$ . No reverse feedback.  $T_{RND}$ , the completion time for source to fill up receiver's buffer by randomly sending one packet per transmission opportunity, has the following conditional probability mass function:*

$$\Pr(T_{RND} = n|c) = \sum_{k=M-c-1}^{n-1} \sum_{j=0}^{M-c-1} (-1)^j \binom{M-c-1}{j} \left(1 - \frac{j+1}{M-c}\right)^k \binom{n-1}{k} \bar{Q}_1^{k+1} (1 - \bar{Q}_1)^{n-k-1},$$

$$\forall n \geq M-c, \quad (3)$$

where  $\bar{Q}_1 = Q_1 \left(1 - \frac{c}{M}\right)$ .

*Proof:* See Appendix A. ■

In the special case when  $c = 0$  and  $Q_1 = 1$ , (3) reduces to the stopping time  $T_{CCP}$  in classic coupon collection problem (Eq. (3) in [15]).

$T_{RND}$ 's PMF (3) is plotted versus simulation result in Fig. 4.  $T_{RND}$  can also be represented by the sum of  $M-c$  independent, geometrically distributed random variables with means  $Q_1 \frac{j}{M}$ ,  $\forall j \in \{M-c, \dots, 2, 1\}$ . Using the independence between random trials, the expected value  $E[T_{RND}|c]$  is simply the sum of the means:

$$E[T_{RND}|c] = \frac{1}{Q_1} \sum_{j=1}^{M-c} \frac{M}{j}. \quad (4)$$

**Network coding aided scheme (NC):** In the network coding aided case, a receiving node stores all packets and keeps track of the rank of coefficient matrix  $\mathbf{A}$  (see (2)). When a new packet is successfully received, its coefficient is extracted and appended to  $\mathbf{A}$  as a new row. Then the receiver calculates  $\text{rank}(\mathbf{A})$ . If  $\mathbf{A}$  reaches full rank, the receiver can successfully recover the original data set. Otherwise, this receiver waits for more coded packets.

Let  $D_v$  denote the dimension of the subspace spanned by the received coded packets at receiver  $v$ . It is easy to see that  $D_v = \text{rank}(A_v)$ . Hence, "rank" and "subspace dimension" are used interchangeably throughout this paper. When  $v$  receives a coded packet from source  $u$ ,  $D_v$  increases if and only if the newly received packet is orthogonal to  $v$ 's subspace  $L_v$ . The following lemma shows the probability of dimension increase.

**Lemma 1.** *The probability that node  $v$  increases its dimension by one after receiving a broadcast transmission is:*

$$P_{D_v} = 1 - q^{-(M-D_v)}, \quad (5)$$

where  $q$  is the finite field size, and  $D_v$  is receiver  $v$ 's subspace dimension before reception.

*Proof:* Receiver  $v$ 's subspace  $L_v$  is a subset of disseminator  $u$ 's subspace, which is of dimension  $M$ . The rank increment probability is derived by considering the event that a received packet is orthogonal to  $L_v$ . ■

Let  $T_{NC}$  denote the completion time for the NC scheme. For  $j \in \{1, 2, \dots, M-c\}$ , let  $T_{NC}^j$  denote the number of transmissions needed for receiver's rank to go from  $c+j-1$  to  $c+j$ . Let  $P_{c+j-1}$  denote the probability that a received packet is innovative given the receiver's rank is  $c+j-1$ .  $T_{NC}^j$  has geometric distribution with parameter  $Q_1 P_{c+j-1}$ . Then,  $T_{NC}$  is the sum of  $M-c$  independent, geometrically distributed random variables. Conditioned on  $c$ , its mean value is

$$E[T_{NC}|c] = \frac{1}{Q_1} \sum_{j=1}^{M-c} \frac{1}{1 - q^{-j}}. \quad (6)$$

Using results from [14], we have the upper bound on the mean completion time:

$$E[T_{NC}|c] \leq \frac{1}{Q_1} \min \left( \frac{M-c}{1-q^{-1}}, M-c+1 + \frac{1-q^{-M+c+1}}{q-1} \right). \quad (7)$$

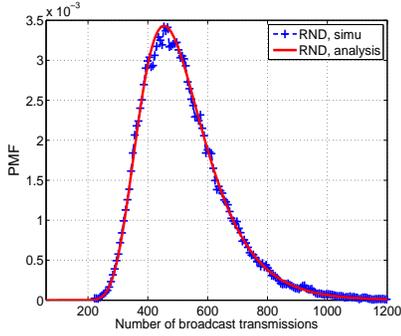


Figure 4. PMF of the random scheme's completion time  $T_{RND}$ . Analysis curve is eq. (3). For all 3 figures,  $M=100$ ,  $c=40$ ,  $Q_1=0.9$ .

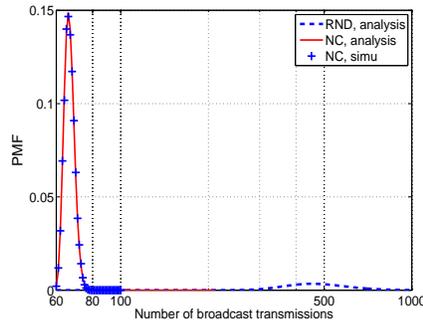


Figure 5. PMF of network coding scheme's completion time  $T_{NC}$ . Analysis curve is eq. (8).

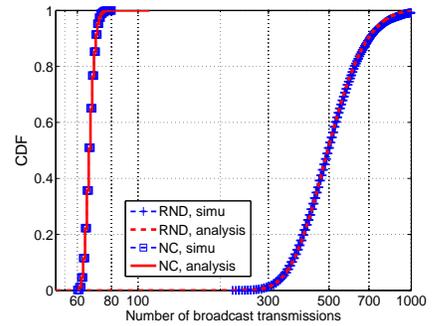


Figure 6. CDF of  $T_{NC}$  and  $T_{RND}$ . Analytical CDF curves are computed from PMFs (eq. (8) and (3)).

More loosely,  $E[T_{NC}|c] \leq \frac{1}{Q_1}(M - c + 2)$ ,  $\forall q \geq 2$ . On average, only  $\frac{2}{Q_1}$  excessive transmissions are incurred, regardless of the finite field size and data set size. If we adopt a reasonably large finite field in network coding operations, every received coded packet is innovative ( $P_{D_v} \approx 1$  for all the values of  $D_v$ ). For example,  $P_0 > \dots > P_{98} = 0.9998 > P_{99} = 0.984$  when  $q = 64$ ,  $M = 100$ . Conditioned on  $c$ ,  $T_{NC}$  is then approximated by a negative binomial random variable with the following PMF:

$$\Pr(T_{NC} = n|c) \approx \binom{n-1}{n-M+c} Q_1^{M-c} (1-Q_1)^{n-M+c}, \quad n \geq M-c, \quad (8)$$

This PMF is plotted versus simulation result in Fig. 5. Conditioned on the same  $c$ , PMF of  $T_{NC}$  is much more concentrated. The comparison of their cumulative distribution functions (CDF) in Fig. 6 shows that  $E[T_{NC}|c]$  is an order of magnitude smaller than  $E[T_{RND}|c]$ . The NC scheme works as if perfect feedback is available without actually sending any feedback in this two node case.

## V. SIMULATION OF A 1-D LATTICE NETWORK

We have derived the PMF of the completion time in a single dissemination step, conditioned on  $c$  - the number of distinct packets possessed by the first recipient at the beginning of each step. In fact, this partial data set is accumulated from multiple sources with different link success probabilities. Hence,  $c$  is a function of reception probability sequence  $Q_2, Q_3, \dots, Q_N, \dots$ , which in turn depends on topology and communication parameters, such as inter-vehicle spacing, transmit power, wireless channel, etc.

Simulation is used to study a more general case with consideration of multiple receivers. We simulate a sufficiently long 1-D network, where the source data (at the origin) is to be propagated to all nodes. The wireless link exhibits Rayleigh fading. The reception probability at the  $k$ -th recipient, relative to current disseminator, is

$$Q_k = \exp(-z n_0 p^{-1} (kd)^\alpha), \quad (9)$$

where  $p$  is transmit power,  $d$  is inter-vehicle spacing,  $\alpha$  is the path loss exponent,  $n_0$  is noise power, and  $z$  is the decoding

signal-to-interference-and-noise-ratio (SINR) threshold. Only current disseminator broadcasts in a slot, hence, no multi-access interference at receivers. Parameter values used in simulation are summarized in Table I. We compare three schemes - random (RND), network coding (NC) aided, and perfect feedback. Perfect feedback scheme assumes feedbacks with no communication overhead, hence, is the upper bound of non-network-coding solutions. With feedback, the instant disseminator only repeats packets that are not yet received by the first recipient.

Table II shows (1) the average number of packets that a disseminator broadcasts in order to push all the content to the immediate next node, (2) the average number of total and useful packets that a node receives. In scheme RND, a node sends and receives 4 times more packets, simply because the majority are duplicated copies. Perfect feedback eliminates most redundancy, and brings the non-NC scheme's performance close to the NC scheme's. Still, the NC scheme outperforms about 8%. Perfect feedback only guarantees that each transmitted packet is innovative to the first receiver, while network coding operations make (almost) all overhead transmissions at all receivers innovative. Fig. 7 shows the number of packets that a node receives from different neighbors when it reaches full buffer. Similarity between NC and perfect feedback schemes indicates that network coding in this VANET setting is trading computational complexity and storage overhead (which are not captured in this paper) for network feedback gain.

Table I  
PARAMETERS USED IN DATA DISSEMINATION SIMULATION

Platoon size	500 vehicles	$M$	100 pks
$p$	1e-5 Watt	$d$	100 m
$\alpha$	3	$q$	128
$L$	1024 Bytes	$z_{QPSK}$	8 dB
$R_{QPSK}$	6 Mbps		

## VI. CONCLUSION

This paper explores peer-to-peer content distribution using V2V ad hoc communications to supplement direct vehicular data download from cellular and road side DSRC infrastructures. Due to the absence of feedback in WAVE broadcasts,

Table II  
RECEIVED AND TRANSMITTED PACKETS PER NODE

Schemes	pks rcvd	rcvd non-innovative pks	pks sent
NC	100	<0.01	39.4
RND	524.9	424.9	207.3
Perfect feedback	108.1	8.1	42.5

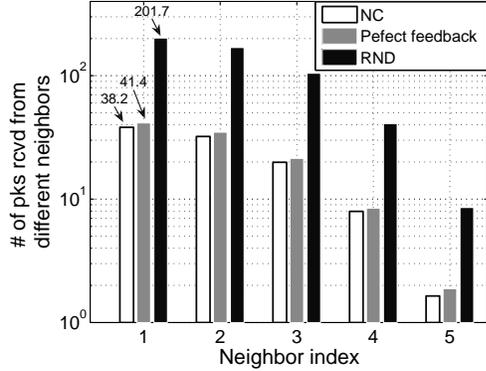


Figure 7. Number of packets that a node receives from different neighbors. Y-axis is in logarithmic scale.

random and network coding are two practical approaches for data dissemination, without incurring systematic modifications. We derive PMFs of the completion time of each dissemination step - pushing data set from current disseminator  $k$  to the first recipient node  $k+1$  - using random broadcast and with network coding, conditioned on the amount of distinct packets at  $k+1$  at the beginning of a step. With network coding, this conditional single step completion time is an order of magnitude less than that from a random broadcast scheme. Simulation results in a 1-D lattice VANET show the gains of network coding, relative to 1) random broadcast, and 2) an ideal scheme with perfect feedback, in a multi-hop vehicular network with Rayleigh fading wireless links.

## REFERENCES

- [1] "IEEE P802.11p/D5.0, IEEE 802.11 Amendment 7: Wireless Access in Vehicular Environments," Nov. 2008.
- [2] "IEEE Std 1609 family, IEEE Trial-Use Standard for Wireless Access in Vehicular Environments(WAVE)," Nov. 2006.
- [3] D. Jiang and L. Delgrossi, "IEEE 802.11 p: Towards an international standard for wireless access in vehicular environments," in *Proc. of IEEE Vehicular Tech. Conf.*, 2008, pp. 2036–2040.
- [4] J. Zhu and S. Roy, "MAC for dedicated short range communications in intelligent transport system," *IEEE Communication Magazine*, vol. 41, no. 12, pp. 60–67, Dec. 2003.
- [5] M. Torrent-Moreno, D. Jiang, and H. Hartenstein, "Broadcast reception rates and effects of priority access in 802.11-based vehicular ad-hoc networks," in *Proc. of the 1st ACM VANET workshop*, 2004, pp. 10–18.
- [6] F. Ye, R. Yim, S. Roy, and J. Zhang, "Efficiency and reliability of one-hop broadcasting in vehicular ad hoc networks," *IEEE J. on Selected Areas in Communications*, vol. 29, no. 1, pp. 151–160, 2011.
- [7] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *Communications Magazine, IEEE*, vol. 44, no. 1, pp. 74–82, 2006.
- [8] F. Ye, R. Yim, J. Guo, J. Zhang, and S. Roy, "Prioritized broadcast contention control in VANET," in *Proc. of IEEE Int'l Conf. on Communications (ICC)*, 2010, pp. 1–5.
- [9] S. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

- [10] A. Nandan, S. Das, G. Pau, M. Gerla, and M. Sanadidi, "Co-operative Downloading in Vehicular Ad-Hoc Wireless Networks," in *Proc. of the 2nd Annual Conf. on Wireless On-demand Network Systems and Services*, 2005, pp. 32–41.
- [11] U. Lee, J. Park, J. Yeh, G. Pau, and M. Gerla, "Code torrent: content distribution using network coding in vanet," in *Proc. of the 1st int'l workshop on Decentralized resource sharing in mobile computing and networking*, 2006, pp. 1–5.
- [12] M. Johnson, L. De Nardis, and K. Ramchandran, "Collaborative content distribution for vehicular ad hoc networks," in *Proc. of Allerton Conf. on Communication, Control, and Computing*, 2006.
- [13] M. Li, Z. Yang, and W. Lou, "Codeon: Cooperative popular content distribution for vehicular networks using symbol level network coding," *IEEE J. on Selected Areas in Communications*, vol. 29, no. 1, pp. 223–235, 2011.
- [14] D. Lucani, M. Médard, and M. Stojanovic, "Random linear network coding for time-division duplexing: field size considerations," in *Proc. of IEEE Global Telecommunications Conf.*, 2009, pp. 1–6.
- [15] H. Rubin and J. Zidek, "A waiting time distribution arising from the coupon collector's problem," 1965.

## APPENDIX

### A. Proof of Theorem 1 - $T_{RND}$ 's conditional PMF

We first state the following result from classic coupon collection problem without proof: Initially, source node has  $M$  packets, receiver has none. The source randomly sends a packet at each transmission opportunity, and the reception at receiver is error free.  $U_{M,n}$ , the number of missing packets at the receiver after  $n$  trials, has the following PMF:

$$\Pr(U_{M,n} = j) = \binom{M}{j} \sum_{k=0}^{M-j} (-1)^k \binom{M-j}{k} \left(1 - \frac{j+k}{M}\right)^n, \quad j \in \{\max\{M-n, 0\}, \dots, M-1\}. \quad (10)$$

In theorem 1, only  $M-c$  packets need to be filled, and the wireless link is unreliable. In a trial, if the receiver successfully receive a packet, and this packet is not among the existing  $c$  packets, we call this a *valid* trial. Its probability is simply  $\bar{Q}_1 = Q_1 \left(1 - \frac{c}{M}\right)$ .

Let  $\bar{U}_{M,n}$  be the number of missing packets after  $n$  trials. We first choose  $k$  trials out of  $n$  to be valid, then apply (10) after substituting total packet population  $M-c$  for  $M$ , and trial number  $k$  for  $n$ .  $\bar{U}_{M,n}$ 's PMF is

$$\Pr(\bar{U}_{M,n} = j|c) = \sum_{k=M-c-j}^n \Pr(U_{M-c,k} = j) \Pr(B(n, \bar{Q}_1) = k), \quad (11)$$

where  $B(n, \bar{Q}_1)$  is a binomial random variable with success probability  $\bar{Q}_1$ .

To derive the PMF of the completion time -  $T_{RND}$ , we consider the event that a receiver collects all but one packets after  $n-1$  trials, and receive the exact missing packet in the  $n$ -th trial. That is,

$$\Pr(T_{RND} = n|c) = \frac{\bar{Q}_1}{M-c} \cdot \Pr(\bar{U}_{M,n-1} = 1|c). \quad (12)$$

The result follows after substitution of (11) into the above expression. ■