Throughput Analysis of Cooperative Mobile Content Distribution in Vehicular Network using Symbol Level Network Coding

Qiben Yan, *Student Member, IEEE,* Ming Li, *Member, IEEE,* Zhengyu Yang, *Member, IEEE,* Wenjing Lou, *Senior Member, IEEE,* Hongqiang Zhai, *Member, IEEE*

***Abstract*—This paper presents a theoretical study of the throughput of mobile content distribution (MCD) in vehicular ad hoc networks (VANETs). Since VANET is well-known for its fast-changing topology and adverse wireless channel environments, various protocols have been proposed in the literature to enhance the performance of MCD in a vehicular environment, using packet-level network coding (PLNC) and symbol-level network coding (SLNC). However, there still lacks a fundamental understanding of the limits of MCD protocols using network coding in VANETs. In this paper, we develop a theoretical model to compute the achievable throughput of cooperative MCD in VANETs using SLNC. By considering a one-dimensional road topology, the expected achievable throughput for a vehicle at a certain distance from the content source is derived, for both using PLNC and SLNC. Our proposed model is unique since it captures the effects of multiple practical factors, including vehicle distribution and mobility pattern, channel fading and packet collisions. Through numerical results, we demonstrate how the achievable throughput decreases with the distance of a vehicle from the source, and evaluate the throughput gain of SLNC over PLNC under different vehicle distributions.**

***Index Terms*—Mobile Content Distribution, VANET, Symbol-Level Network Coding, Achievable Throughput.**

I. Introduction

MOBILE content distribution (MCD) is a promising service in VANETs, where multimedia contents are distributed from one or more fixed access points (APs) to the vehicles driving through an area of interest (AoI). Examples of MCD services include: live video broadcast of road traffic and weather conditions; periodic broadcast of multimedia advertisements from local businesses; updates of the GPS map about a city. However, the provisioning of MCD in VANETs meets several challenges. On one hand, multimedia contents containing audio and video usually require *high downloading rate* (or high end-to-end throughput), which demands high network bandwidth. On the other hand, wireless is a well-known lossy medium with limited bandwidth, in which channel fading and various interference dramatically reduce the throughput. The high mobility of VANETs, leading to fast and unpredictable topological changes, will further exacerbate the frequent packet losses and collisions.

Network coding (NC) is a common technique adopted in MCD as an effective approach to improving the bandwidth efficiency and simplifying the protocol design. NC basically breaks the store-and-forward packet forwarding paradigm by allowing intermediate nodes to combine received packets. Since each coded packet could benefit multiple receivers simultaneously, the bandwidth efficiency can be improved. More recently, *symbol-level network coding (SLNC)* has been proposed [1]. By combining packets at symbol level, SLNC allows a node to recover correctly received symbols from erroneous packets. Hence, SLNC provides increased successful packet reception rate due to its better error tolerance.

However, there is a lack of theoretical foundation and understanding on the performance limits such as achievable throughput by SLNC, especially in realistic conditions such as high mobility MCD scenario in VANETs. In this paper, we endeavor towards bridging the above gap by first establishing a theoretical framework to analyze the achievable throughput of SLNC in general wireless networks, and then apply our model to study MCD in a highway VANET scenario. We consider a line-shape road topology where the vehicles are positioned according to some distribution, and they cooperatively broadcast continuous contents coming from the AP to all the vehicles inside an AoI using SLNC. We are interested in answering two questions: 1) Regarding realistic issues of channel fading, interference and node distribution, how does the achievable throughput at a node changes with its distance from the AP? 2) Given a specific vehicle mobility pattern, what is the downloading volume a vehicle can obtain from the AP during the time period it drives through the AoI?

The main contributions of this paper can be summarized as follows:

(1) We propose an analytical model to compute the achievable throughput of single-flow unicast and multicast using SLNC in general wireless networks, based on network flow and queueing theory.

(2) We apply the above model to derive the expected achievable throughput of cooperative MCD system at a certain distance from the source using PLNC or SLNC in VANETs. Then, the expected downloading volume of a vehicle can be obtained by aggregating the expected achievable throughput over its distance to the source with a certain vehicular mobility pattern, which captures the dynamic property of VANET.

(3) We demonstrate numerical results from our model. Our findings provide insights on optimized choices for cooperative MCD system design in VANETs. We also show the throughput gain of SLNC over PLNC under different channel conditions and vehicle distributions.

II. Related Work

A. Capacity Scaling Law of Wireless Networks with Network Coding

The previous works focused on deriving the information-theoretic capacity of wireless network with NC [2]–[5]. Another similar approach reveals the capacity of NC in a wireless network via the "asymptotic throughput". [6] and [7] studied a dense network model considering interference and noise to obtain its asymptotical throughput. In contrast, in this paper we consider an extended network model with constant node density, which is more realistic for the VANETs.

Lun et al. [8] proposed a theoretical model to compute the exact capacity region of random linear network coding in both wired and wireless networks. This is the most related work in terms of technical approach, but we have several key differences: 1) we model the achievable throughput of SLNC instead of PLNC; 2) we provide a method to derive the achievable throughput for MCD in VANETs under practical conditions including channel fading, medium contention, symbol correlation and vehicle distribution; 3) from our results, we give several insights on the design of cooperative MCD systems in VANETs.

B. Achievable Throughput of MCD in VANETs

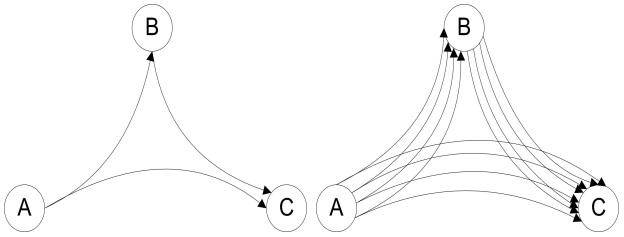
Recently, SLNC is introduced into cooperative MCD to further improve the downloading performance [9], [10]. However, there still lacks an analytical model to compute the achievable throughput of SLNC in MCD system and to quantify the gain of it compared with PLNC. There are only a few works studying the capacity of unicast or multicast in a VANET setting. The asymptotic transport capacity of VANETs was studied in [11], and they derived the achievable throughput in VANETs without considering interference and vehicle distribution. In [12], Johnson et al. considered a similar scenario to ours, and derived the achievable throughput without taking into account channel fading or vehicle distribution; as a result, their achievable throughput has no connection with the source-destination distance and node density.

III. Achievable Throughput of Symbol-level Network Coding in Wireless Network

In this section, we first put forward a system model to analyze SLNC in generic wireless network. Then based on this model, we propose a theoretical framework using queueing theory and flow network to obtain the achievable throughput of SLNC.

A. System model for SLNC in Wireless Network

Unlike PLNC, SLNC operates in the granularity of symbols. The readers may refer to [1] for more detailed description of SLNC coding operation. In this section, we will formulate a system model for analyzing SLNC in generic wireless network. Before we elicit the system model, first of all, we must have a scheduling strategy to determine how packets are injected onto each arc. However, the packet scheduling problem is a difficult problem with many existing works [13]–[15] contributing to address this problem. In our generic framework, we assume a scheduling scheme is pre-defined. The packet transmission in our model works as follows: each node stores packets that it receives, then new packets are generated for injection by SLNC whenever the node obtains a transmission opportunity through the scheduling scheme. Consequently, each node potentially always has packets stored in its memory, and injection can happen whenever it gets a chance.



(a) Hypergraph (for modeling PLNC) (b) Multi-Hypergraph (for modeling SLNC)

Fig. 1: Graph models for PLNC and SLNC

Wireless network is modeled as a directed hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{A})$ [8], where $\mathcal{N}$ is the set of nodes, $\mathcal{A}$ is the set of hyperarcs. A hyperarc (i, J) represents a lossy broadcast link, where i is a node from $\mathcal{N}$, J is a non-empty subset of $\mathcal{N}$. Some injected packets on (i, J) are received by a set K which is a subset of J. We define the average rate at which packets are injected to the hyperarc (i, J) and is received by exactly $K \subset J$ as z_{iJK}. If the corresponding packet counting process of arrival packets during time τ is $A_{iJK}(\tau)$, the average rate $z_{iJK} = lim_{\tau\to\infty} A_{iJK}(\tau)/\tau$.

We believe the above hypergraph model is an accurate abstraction of packet-level wireless network, thus is a suitable model for analyzing PLNC. On the other hand, SLNC encodes the information at symbol level, which makes the hypergraph model inappropriate without capturing the reception status of each symbol in the packet. In order to fully capture SLNC performance, we convert the hypergraph model into a $multi-hypergraph$ $\mathcal{H}^M = (\mathcal{N}, \mathcal{A}^M)$, where M is the number of symbols contained in each packet. In multi-hypergraph, as in Fig. 1(b), there are M corresponding hyperarcs $\{a_1, a_2, ..., a_M\}$ in the hyperarc sets $\mathcal{A}^M$ all with the same starting node i and end sets J. We call these M related hyperarcs as $multi-hyperarcs$. Conceptually, hyperarc a_m corresponds to the transmission of the mth coded symbol. By virtual of multi-hypergraph, the process of one packet injected to one hyperarc is decomposed into M injections of symbols to M hyperarcs.

Before delving into the achievable throughput of SLNC with the multi-hypergraph model, we present the definition of achievable throughput:

Definition 1 (Achievable throughput): The feasible flow rate a source node can forward to its destinations over a long term.

B. Achievable Throughput for SLNC in Wireless Network

In this section, we give our general results for achievable throughput of SLNC starting from special cases. We first consider a two-link tandem network, which is a multigraph shown in Fig. 2. The tandem network with PLNC has been studied in [8], in which the propagation of innovative packets through a node follows the propagation of jobs through a

Fig. 2: two-link tandem network. Left: packet level queueing network; right:symbol level multi-queueing network.

single-server queueing station, which can be analyzed using fluid approximation for discrete-flow networks. The achievable throughput with PLNC is then proven to be determined by the average packet arrival rate on each arc.

On the other hand, the tandem network with SLNC has M arcs between two nodes (See Fig. 2). We virtually maintain one queue for each symbol position at the queueing station of each node, so that each node maintains M multiple queues. Then, the propagation of innovative symbols through arc (i, j) can be regarded as the propagation of jobs through a multiple single-server queueing system consisting of M single-server queueing stations. In the following, we demonstrate that the single-server queueing station at each symbol position works the same way as the single-server queueing station of PLNC.

As we mentioned above, [8] shows that the propagation of packets with innovative information can be modeled as a single-server queueing station, which results in the achievable throughput determined by the average packet arrival rate on each arc, if the average rate exists. So if the arrival process for each symbol has a same average rate, the similar relationship between achievable throughput and average arrival rate still holds for each symbol.

Switch from the perspective of packets to symbols, the arrival of one received packet on arc (i, j) can be translated into the arrival of all received symbols in the packet on the same arc. On one hand, if these symbols are independently received, it is straightforward that the arrival processes of symbols are independent and identical distributed (i.i.d.). Assume symbols' reception probability is P_{sym}, packets' reception probability is P_{pkt} and average packet arrival rate is z_{ij}, then the average symbol arrival rate at the mth symbol position is:

$$z_{ij}^{(m)} = \frac{P_{sym}}{P_{pkt}} z_{ij}, \tag{1}$$

which apparently exists and is the same for each symbol. But on the other hand, if these symbols are received with correlated errors, by modeling the correlated channel states as Markov chains [16], the arrival processes of packets or symbols still have average arrival rates according to [8]. Furthermore, by scrutinizing a long term, each symbol position will undergo a same fading process on average, through which all the symbols will gradually attain the same average symbol reception probability $\overline{P}_{sym}$, so that each of them must have the same average symbol arrival rate. Therefore, the symbol-level achievable throughput turns out to be determined by average symbol arrival rate at each link.

Assume the mth symbol injection rate on arc (i, j) is $r_{ij}^{(m)}$, which is the same as packet injection rate r_{ij}, the average arrival rate of mth received symbol is thus $z_{ij}^{(m)} = (1-\varepsilon_{ij}^{(m)})r_{ij}^{(m)}$, where $\varepsilon_{ij}^{(m)}$ is the symbol loss rate. Therefore, the achievable throughput of mth symbol in two-link tandem network is:

$$R^{(m)} \leq min(z_{12}^{(m)}, z_{23}^{(m)}), \tag{2}$$

which is determined by average symbol arrival rate on each link. As the packet arrival rate on arc (i, j) with PLNC is $z_{ij} = (1 - \varepsilon_{ij})r_{ij}$, thus the difference between achievable throughput of SLNC and PLNC is determined by the loss rates' ratio: $\frac{1-\varepsilon_{ij}^{(m)}}{1-\varepsilon_{ij}}$.

In practice, the symbol loss rate is smaller than packet loss rate. Consider if the symbol losses are independent with each other, the packet loss rate is then $\varepsilon_{ij} = 1 - (1 - \varepsilon_{ij}^{(m)})^M$. Therefore, we conclude that the gain of achievable throughput of SLNC over PLNC comes from symbol-level diversity, which is mainly due to the increased successful transmission rate at symbol-level.

The result of two-link tandem network can be extended to L-link tandem network, thus the achievable throughput of mth symbol in L-link tandem network is given by:

$$R^{(m)} \leq min_{1\leq i\leq L}\{z_{i(i+1)}^{(m)}\}. \tag{3}$$

Finally, we further extend the results into symbol-level wireless networks, in which the multi-hypergraph can then be separated into M independently conceptual hypergraph for each symbol. In mth hypergraph $\mathcal{A}_m^M$, the mth symbol injected on hyperarc (i, J) is received by exactly $K \in J$. The average arrival rate of mth symbol is $z_{iJK}^{(m)}$, then the achievable throughput of mth symbol from source s to destination t is given by:

$$R_t^{(m)} \leq min_{\mathcal{Q}\in\mathfrak{Q}^{(m)}(s,t)}\{\sum_{(i,J)\in\Gamma_+(\mathcal{Q})}\sum_{K\not\subset\mathcal{Q}} z_{iJK}^{(m)}\}, \tag{4}$$

where $\mathfrak{Q}^m(s, t)$ is the set of all cuts between s and t on the mth conceptual hypergraph, and $\Gamma_+(\mathcal{Q})$ denotes the set of forward hyperarcs of the cut $\mathcal{Q}$, i.e. $\Gamma_+(\mathcal{Q}) := \{(i, J) \in \mathcal{A}_m^M | i \in \mathcal{Q}, J\backslash\mathcal{Q} \neq \emptyset\}$. Evidently, the achievable throughput of each symbol is determined by the minimal cut of wireless hypergraph.

Finally, the overall achievable throughput from source s to destination t at symbol-level is given by:$R_t = \sum_{m=1}^{M}\{R_t^{(m)}\}$, which is measured in number of symbols per second, from which we can easily obtain achievable throughput in bytes or bits per second.

IV. Throughput analysis of cooperative MCD system using SLNC

In this section, we apply the above model to compute the achievable throughput of cooperative MCD system using SLNC in VANETs.

A. System Model of Cooperative MCD System

We consider the following single-flow multicast MCD service architecture with a line-shaped road topology. The content providers such as Public Transport Authorities disseminate safety information or commercial Ads to a roadside AP, which is the only data source in the network. The AP then broadcasts these files to the vehicles inside a geographical AoI. The vehicles inside AP's coverage will actively download all the overheard packets and store them in the buffer. We adopt 802.11 MAC protocol as the underlying medium access control scheme. Once the vehicle gets a chance to transmit, it will generate a coded packet using SLNC and transmit it to its neighbors far away from the AP, so that all the vehicles

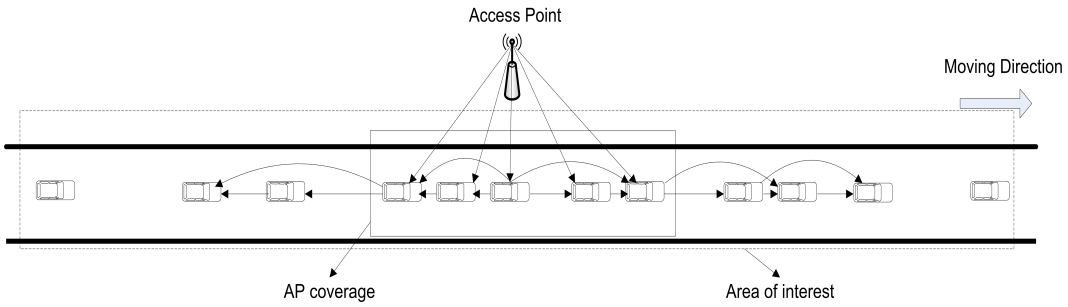


Fig. 3: The architecture for cooperative MCD. AP owns contents to broadcast to vehicles; vehicles distribute their received contents cooperatively. The arrows denote information flows.

will continuously transmit and receive coded packets in order to recover the complete files. The cooperative MCD service architecture is illustrated in Fig. 3.

To make the problem tractable, we make some important assumptions to the vehicles' neighborhoods. First, we assume the road is long enough such that the neighborhoods of a vehicle from both sides are extending to infinite, with which boundary effect is eliminated. Second, we assume all the vehicles on the road have homogeneous neighborhood distribution, i.e. the positions of neighboring vehicles follow a same node distribution with identical but tunable vehicle density across time and space.

Then we give some definitions to some frequently used phrases in this paper:

Definition 2 (Link average arrival rate): The average reception rate that a node can achieve over a particular link (arc) in a long term;

Definition 3 (Distance-limited achievable throughput): The expected achievable throughput of a node with a certain distance away from the source in a network under some specific node distribution;

Definition 4 (Expected downloading volume): The expected total amount of information that a vehicle can accumulate when driving through an AoI.

It is well known that the network with 802.11 MAC scheme potentially exhibits an unstable behavior [17], which enforces us to study *saturation throughput*. Therefore, we focus on the achievable throughput in the scenario when all the nodes including source and relays potentially always have packets stored in their buffers and send coded packets whenever is possible, which represents a stable condition.

Based on the achievable symbol throughput of SLNC indicated in Eq. (4), we take three steps to approach the achievable throughput of cooperative MCD system: first, under a specific VANET topology, we compute per-link average symbol arrival rate by "communication density"; second, we address the max-flow min-cut problem to get the maximal divergence out of the source over all capacity-achievable flows, and acquire distance-limited achievable throughput by averaging among all topological instances; third, by integrating vehicle mobility, expected downloading volume is obtained via aggregating distance-limited achievable throughput of a measured vehicle inside an AoI.

B. Symbol-Level Link Average Arrival Rate

Considering the realistic issues of channel fading, interference and node distribution, the concept of "communication density" [18], described as $\delta p d_{cr}$, where δ is vehicle density, p is message generation rate and d_{cr} is transmission range, is a promising metric to measure the node reception probability of a line-shaped broadcast system with 802.11 MAC protocol under practical channel conditions. By taking advantage of communication density which already takes channel variety and packet collision into account, the reception rate of node A contributed by source S through the arc (S, A), considering interference from its neighboring nodes, can be represented by a closed-form formula.

1) Link Average Arrival Rate Considering Channel Fading: The link average arrival rate considered in this section is in symbol-level, which can also be denoted as average symbol reception rate. One important factor impacting the reception rate is channel randomness which is brought by channel fading. In Rayleigh fading channels, we assume the fading process is consistent over the transmission of one physical symbol. Thus, the pdf of *physical symbol*[1] SNR γ_{ps} is given by [19]:

$$f(\gamma_{ps}) = \frac{1}{\bar{\gamma}_{ps}} e^{-\frac{\gamma_{ps}}{\bar{\gamma}_{ps}}}, \tag{5}$$

where $\bar{\gamma}_{ps}$ is the average physical symbol SNR. Consider a simple path-loss model for signal propagation, the receiving power is decayed with distance as $d^{-\alpha}$, where d is the distance, α is the path-loss factor. The average physical symbol SNR can then be written as $\bar{\gamma}_{ps} = \frac{E_{ps}}{d^{\alpha} N_0}$, in which E_{ps} is the transmission power per physical symbol, N_0 denotes the Gaussian noise spectral density. If we model the loss process as a SNR threshold model with threshold γ_{th}, the probability that the physical symbol is correctly received at a distance d from the source is given by:

$$P_{succ,ps}(d) = P(\gamma(d) \geq \gamma_{th}) = e^{-\frac{\gamma_{th}}{\bar{\gamma}_{ps}}} = e^{-\rho d^{\alpha}}, \tag{6}$$

where $\rho = \frac{\gamma_{th} N_0}{E_{ps}}$.

We assume a symbol in SLNC contains μ physical symbols. Consider two distinct types of Rayleigh fading channel for performance comparison: channel A is characterized by flat slow fading with no line-of-sight (LoS) path, while channel B is characterized by fast fading. These two types of channels indicate two extreme cases in real vehicular network, while the real channel condition is in between that of channel A and channel B, affected by variable vehicle densities and surrounding environments. With channel A, the error rates of different physical symbols are fully correlated within the packet because of the larger channel coherence time in slow

[1]It denotes the symbol generated by modulation schemes.

fading scenario. Then the probability that a symbol in SLNC is correctly received at a distance d from the source in this case is give by:

$$P_{succ,ns\text{-}corr}(d) = P_{succ,ps}(d) = e^{-\rho d^{\alpha}}. \quad (7)$$

In contrast, the error rates of various physical symbols are independent with each other under channel B. Then the probability that a symbol is correctly received at a distance d from the source in this case is give by:

$$P_{succ,ns\text{-}ind}(d) = (P_{succ,ps}(d))^{\mu} = (e^{-\rho d^{\alpha}})^{\mu}. \quad (8)$$

Finally, the link average arrival rate of mth symbol on arc (i, j) considering Rayleigh fading can be written as:

$$z_{fad}^{(m)}(i,j) = r_{i,j}^{(m)} \cdot P_{succ\text{-}fad}^{(m)}(d_{ij}), \quad (9)$$

where $r_{i,j}^{(m)}$ is the injection rate of mth symbol, d_{ij} is the length of arc (i, j), $P_{succ\text{-}fad}^{(m)}(d_{ij})$ can be derived from Eq. (7) and Eq. (8) for either fully-correlated symbol error case or independent symbol error case.

2) Link Average Arrival Rate Considering Collision: In our cooperative MCD system, another important factor we need to explore is interference caused by simultaneous transmissions, which is mostly induced by the well known "hidden terminal" problem. Even worse, the saturated transmission will exacerbate the packet collision to cause more packets to drop, degrading the throughput performance.

Recall the assumption that all the vehicles on the road have homogeneous neighborhood distribution. As a reasonable extension, we assume the probability of message transmission in a slot is constant and independent of nodes. Furthermore, we assume any packet collision will surely result in packet reception error, then the corresponding packet reception probability at a distance d is [18]:

$$P_{succ}(d) = p \cdot (1 - p_{fad}) \cdot (1 - p_{col}), \quad (10)$$

where p is message transmission probability in one slot, p_{fad} is packet error probability caused by channel fading, and p_{col} is packet error probability caused by packet collision. Each of the above three parameters should be evaluated to determine packet reception probability. Next, we will evaluate p_{col} and p under two channel conditions respectively, as p_{fad} has already been addressed in the previous section.

The packet collision probability in 802.11 MAC protocol is determined by both carrier-sensing technique and hidden node interference. The carrier sense probability at distance d from the source is given by $\chi(d)$. Under the slow fading Rayleigh channel A, given a SNR threshold for carrier sensing, the probability of the source's carrier being "sensed" at distance d can be given by $\chi(d) = e^{-\chi d^{\alpha}}$ with $\chi > 0$, and χ depends on the packet transmission power and sensing threshold, which is closely tied to the carrier sense range d_{cs}. Then, the probability of sensing a new transmission during one time slot is given by [18]:

$$r = 1 - e^{-\sqrt{\pi}\delta p d_{cs}}, \quad (11)$$

where δ is the node density, p is the message transmission probability in one slot. However, computing the carrier sense probability becomes more involved under the fast fading Rayleigh channel B with independent symbol errors. The following theorem characterizes the carrier sense probability under fast fading channel:

***Theorem** 1:* Assume the number of physical symbols used for carrier sensing is N, the probability of sensing a new transmission during one time slot under fast fading channel with independent symbol errors is given by:

$$r = 1 - exp(-\frac{\sqrt{\pi}\delta p d_{cs}}{\sqrt{N}}\{1 + \sum_{n=1}^{N-1}[\frac{(n-\frac{1}{2})(n-\frac{3}{2})\cdots(\frac{1}{2})}{n!}]\}). \quad (12)$$

Proof: See Appendix A for proof. ■

Taking both the carrier sense capability and hidden-node collision into considering, the probability of packet collision occurred at the node is presented as [18]:

$$p_{col} = 1 - e^{-(2S_{msg}-3)\delta p d_{cr}}, \quad (13)$$

where S_{msg} is the message size in slots, and d_{cr} is the communication range.

As mentioned above, we are considering saturation condition, so that all the nodes potentially always have packets to transmit, resulting in an infinite message generation rate at every node. Therefore, according to [18], the message transmission probability in one slot p is computed by:

$$\frac{1}{p} \cong T_0 + \frac{(W-1)}{2} T_1, \quad (14)$$

where W is the backoff window size, T_0 is the average amount of time a node spends in its own transmission including the extra waiting time due to sensible overlapping transmissions and T_1 is the average time spent to decrease the backoff counter by 1. Both of T_0 and T_1 are given in [18], which are expressed in terms of r, and r is in turn stated in terms of p together with some model parameters (See Eq. (11) and (12)). Hence, we can calculate the message transmission probability in one slot p by numerical method from Eq. (11), (12) and (14).

Note that the message mentioned above is in unit of time slots. After deriving message transmission probability p, the data transmission rate at each node is given by:

$$R_{tr} = p \cdot S_{msg} \cdot \frac{L_{pkt}}{\tau_s}, \quad (15)$$

where S_{msg} is the message size in slots, L_{pkt} is the payload length in one time slot, τ_s is the time slot duration. R_{tr} can be measured as bits per second or bytes per second.

The above formulations are all in granularity of packets. As for symbol-level reception rate, the same form of Eq. (10) still holds. For one thing, the symbol error rate caused by channel fading is provided in the previous section. For another, symbol collision probability can be derived based on packet collision probability of Eq. (13) as follows. For simplicity of analysis, we assume the packet collision is uniformly occurred among a whole packet and the symbols after the occurrence of packet collisions are all discarded. Therefore, in the average case, half of the symbols in one collided packet will be devastated over a long term, i.e. the mth symbol collision probability is halved:

$$p_{col}^{(m)} = p_{col}/2. \quad (16)$$

Combine the results of Eq. (10), (13) and (16), the symbol-level link average arrival rate on arc (i, j), which is equal to

the mth symbol's average reception rate $R_{rec}^{(m)}(d_{ij})$ at distance d_{ij}, is expressed as follows:

$$z_{i,j}^{(m)} = R_{rec}^{(m)}(d_{ij}) = R_{tr} \cdot P_{\textit{succ-fad}}^{(m)}(d_{ij})[1 - \frac{1}{2}(1 - e^{-(2S_{msg}-3)\delta pd_{cr}})]. \quad (17)$$

So up to now, we have attained symbol-level link average arrival rate for each link using 802.11 MAC protocol. Next, with a specific network topology, according to Eq. (4), we are going to find the max-flow min-cut of this multicast network with each link weighted by its link average arrival rate.

C. Distance-Limited Achievable Throughput of Cooperative MCD System

We begin with a deterministic network topological instance $\mathcal{D}_i$ generated by a certain node distribution, for example, Poisson distribution. $\mathcal{D}_i$ can be viewed as a graph containing randomly distributed nodes. We imagine a virtual node t located on graph $\mathcal{D}_i$, who has a distance d away from the source s. We are concerning about the achievable throughput at note t. In order to derive it, we further define the hypergraph between source s and t as $\mathcal{G}_{st}^{(i)}(d) \subset \mathcal{D}_i$. The max-flow from s to t is determined by $\mathcal{G}_{st}^{(i)}(d)$ from the network flow's perspective. Integrating the results of previous section, we then employ Ford-Fulkerson algorithm to search for max-flow in the hypergraph $\mathcal{G}_{st}^{(i)}(d)$ with each link weighted by its link average arrival rate, which leads to the achievable throughput at distance d from source in this specific topological instance.

After that, we average over all these topological instances to obtain distance-limited achievable throughput $DLAT^{(m)}(d)$ at the mth symbol position. Hence, the overall distance-limited achievable throughput at symbol-level is $DLAT(d) = \sum_{m=1}^{M} DLAT^{(m)}(d)$.

D. Expected Downloading Volume with Mobility

According to the definition, the downloading volume can be computed for each specific moving VANET instance, so that the expected downloading volume is obtained by calculating the expectation of the downloading volume among all possible moving VANET instances, which is hard to derive. Alternatively, we divide the moving VANET instances into static topology snapshots, each of which lasts for a short time interval. Because of the linearity of expectation, we calculate distance-limited achievable throughput at the vehicle's instant position during each short time interval. Then we aggregate distance-limited achievable throughput over time and distance to get the expected downloading volume with a certain mobility pattern.

Each time interval should be much larger than the maximal end-to-end transmission delay in order to be counted as "a long term" for calculating achievable throughput, and also it should be small enough to keep the network topology relatively static. As an example, in realistic wireless environment, we assume a meaningful AoI has a range of $1200m$. Then, we assume the message transmission probability in each slot is $p = 0.05$, thus the average time spent for channel contention for each hop is $1/p = 20$ slots. And further assume the message transmission air time including the MAC overhead is 55 slots[2] and one slot time is $20us$, then, the maximal end-to-end transmission delay is approximately $7.5ms$. For example, we choose the short time interval as $1s \gg 7.5ms$, then ideally more than 133 transmissions will occur inside each AoI, which is sufficient enough to maintain a proper network flow for a long term to reach achievable throughput. On the other hand, with the vehicle velocity as fast as $30m/s$, the vehicle movement of $30m$ in one second interval will not alter the whole network topology. Therefore, we can regard each network instance as a static snapshot by dividing the time period into one second interval.

TABLE I: Parameters

Parameters	Values
$Number\ of\ topology\ instances$	10000
$Data\ transmission\ rate$	12 Mbps
$Message\ size$	15 slots
$CW\ (Backoff\ window\ size)$	15 slots
$Slot\ time\ \tau_s$	20 us
$Data\ transmission\ range$	250 meters
$Carrier\ sensing\ range$	400 meters
$Path\ loss\ factor$	2
$Length\ of\ AoI$	1.2 km
μ	4

Combining the results of $DLAT(d)$ for each snapshot, we can obtain the expected downloading volume EDV_s of vehicle s as follows:

$$EDV_s = \sum_{k=0}^{\lfloor \frac{T_{out}^s - T_{in}^s}{\Delta t} \rfloor} \Delta t \cdot DLAT(|d(k\Delta t)|), \quad (18)$$

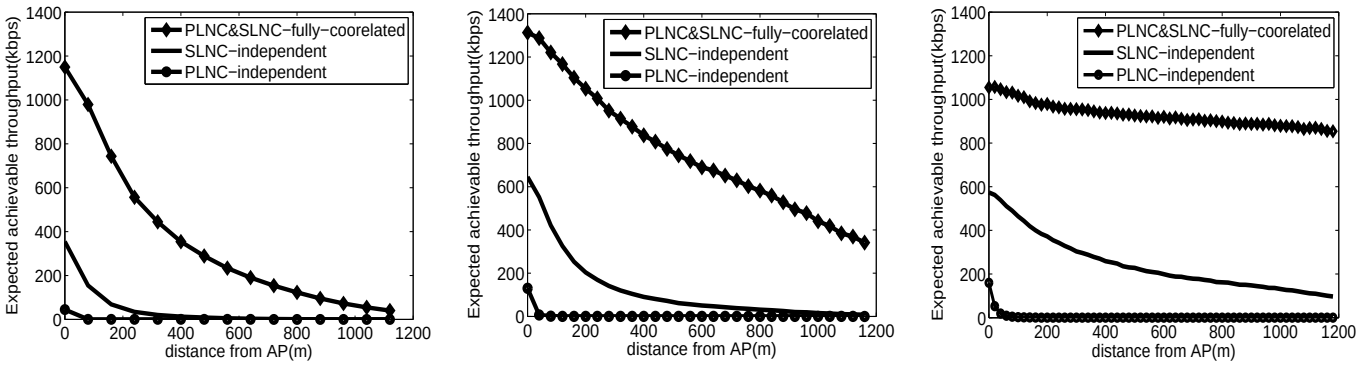
where T_{in}^s (T_{out}^s) is the time when vehicle s enters (departs) the AoI, which is determined by vehicle velocity, vehicle density and the length of AoI. Δt is the pre-defined short time interval and $|d(k\Delta t)|$ is the distance between the vehicle and AP at time $k\Delta t$, which gradually diminishes when approaching the AP and increases when departing the AP.

V. Numerical Evaluation and Discussion

In this section, we demonstrate numerical results from our model. The systematic parameters used in the numerical evaluation is shown in Table. I.

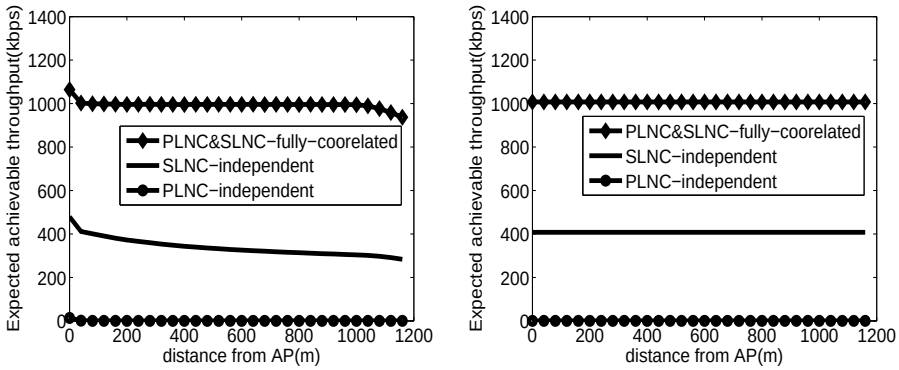
First, we generate various network topological instances with node inter-distance following exponential distribution. Two different channel conditions are considered: one is fast fading channel with independent symbol errors and the other is slow fading channel with fully-correlated symbol errors. In Fig. 4(a)-(c), we show the achievable throughput of different vehicle density variances from our models. Note that each sub-figure has three throughput curves, for packet error rate and symbol error rate at both fast fading and slow fading channel respectively. From the numerical results denoted in each figure, we notice that SLNC has a performance gain over PLNC under fast fading channel, while in the slow fading case, PLNC and SLNC appear to have the same achievable throughput. This result confirms that the benefit of SLNC over PLNC comes from symbol-level diversity, which will diminish with the increasing of the correlation degree among different

[2]We assume packet payload is 1500 bytes, other 802.11 overhead is 150 bytes including preamble, header and ACK, and the slot time is 20 us. Let the data transmission rate be 12 Mbps, the air transmission time is $\frac{(1500+150)\times 8}{12} = 1100us$, which equals to 55 slots.



(a) Average inter-distance is 80m (b) Average inter-distance is 40m (c) Average inter-distance is 20m

Fig. 4: Average achievable throughput of MCD in vehicular network with exponential inter-distance distribution



(a) Uniform inter-distance distribution with 40m average inter-distance (b) Regular inter-distance with 40m inter-distance

Fig. 5: Comparison of different inter-distance distribution

symbol errors. Therefore, the gain of SLNC over PLNC is upper bounded by that of the fast fading case with independent symbol errors.

Now, let us focus on the influence of different vehicle density variances among three sub-figures. Comparing the expected achievable throughput in the case with fully-correlated symbol errors, we notice some instructive insights: when the vehicle density increases, the expected achievable throughput at nearer hops from the source first increases due to better connectivity (Fig. 4(a) – Fig. 4(b)), and then decreases due to heavier interference (Fig. 4(b) – Fig. 4(c)). Moreover, the slope of these performance curves becomes smaller with the growing vehicle density, implying that the achievable throughput decreases slower with distances. The reason for the above results originates from the fact that the high vehicle density guarantees a better overall network connectivity, which induces a relatively stable achievable throughput performance over a long distance. Consequently, we reveal an interesting interplay between network connectivity and interference level, which will in turn provide insights on optimization of MCD system design in VANETs. In short, for different vehicle density scenarios, both network connectivity and inter-node interference should be jointly considered to determine offered source data rate and AP deployment.

Furthermore, we study achievable throughput performance under two different node distributions in Fig. 5. One is uniform distribution and the other is regular network with equal inter-distance. Both of these two distributions have the same expected inter-distance as 40 meters. Fig. 5(a) and Fig. 5(b) show that under fast fading channel, SLNC has a much higher gain over PLNC, while the expected achievable throughput of PLNC is approaching zero. As a special case, the throughput performance of a regular network always appears as a straight line because of the deterministic topology and homogeneous node density.

Finally, from Fig. 6 and Fig. 7, we show the expected downloading volume with different average inter-distances. The x-axis in the figure denotes the starting distance of a vehicle from AP. As the movement of vehicles follows the routines of first moving forward AP and then turning away from AP, we define the distance that a vehicle passes through as a doubled starting distance. Without embedding traffic model, Fig. 6 shows the expected downloading volume with fixed velocity of $20m/s$, in which case, the vehicle velocity has no relationship with vehicle density. The vehicle in the sparsest density case accumulates the lowest downloading volume, due to the low achievable throughput near AP and fastest drop of achievable throughput with the increasing distance, illustrated in Fig. 4(a). However, the median density case shows the highest downloading volume until its starting distance exceeds a point near 650 meters. Beyond that transition point, the downloading volume of heaviest density case overtakes that of median density case. The same interplay between network connectivity and interference level, demonstrated in Fig. 4, still

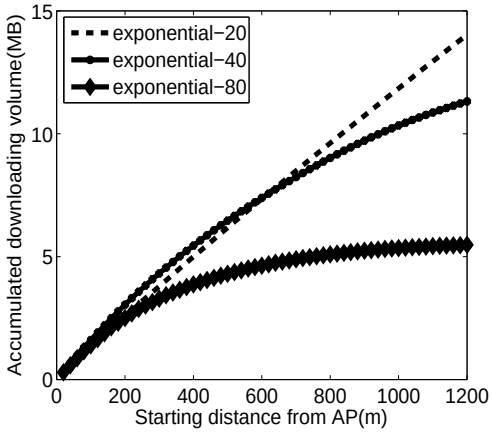


Fig. 6: The expected downloading volume of different average inter-distanc

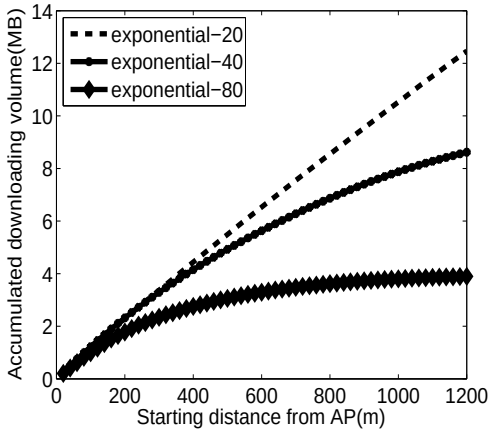


Fig. 7: The expected downloading volume of different average inter-distances with traffic model

holds to explain this phenomenon: due to a higher interference level, a small volume is accumulated around the AP in the heaviest density case. However, the downloading volume steadily rise up with the expansion of AoI, owning to its stable overall connectivity.

In the following, an empirical traffic model is employed to capture the relationship between vehicle velocity and vehicle density [20]:

$$v(m/s) = -0.15\delta + 30, \tag{19}$$

where $v(m/s)$ is the vehicle velocity, δ represents the number of vehicles per kilometer or vehicle density. Fig. 7 shows that the vehicles in the heaviest density network obtain the highest expected downloading volume, and the downloading volume fall down with the decrease of vehicle density. The traffic model of Eq. (19) indicates that a higher vehicle density brings about a lower vehicle velocity, which prolongs the time period spending in the AoI, and correspondingly induces higher downloading volume by the vehicles. Hence, in highest density case, the vehicle accumulates more contents due to its extended downloading time in the AoI.

VI. Conclusion

In this paper, we have done an analytical analysis for the achievable throughput of cooperative MCD in VANETs using SLNC. We first propose a generic model to compute the achievable throughput of SLNC in wireless networks. We show that as long as the packet arrivals conform to an arrival process having average rates, the difference in achievable throughput between SLNC and PLNC is determined by symbol-level diversity. We then propose a method to analyze the expected achievable throughput of cooperative MCD system with SLNC in VANETs, by involving realistic factors such as channel fading, interference and vehicle distributions. Furthermore, by considering vehicular mobility pattern, we compute the expected downloading volume of a vehicle passing through an AoI. Through numerical results, we reveal the impacts of using PLNC & SLNC under different channel fading levels, vehicle densities and vehicle distributions to the throughput limits of MCD, which provide valuable insights for optimized deployment of APs and the cooperative MCD system design.

Appendix A
Proof of Theorem 1

Under the fast fading Rayleigh channel with independent symbol errors, we want to calculate the probability of sensing one new transmission during one time slot. Carrier sense happens at the beginning of each time slot. We assume the number of physical symbols for carrier sensing within one slot is N, then the pdf of average SNR of N independent physical symbols is given by:

$$f(\gamma_{avg}) = \frac{(N\lambda)^N \gamma_{avg}^{N-1} e^{(N\lambda)\gamma_{avg}}}{(N-1)!}, \tag{20}$$

where $\lambda = \frac{d^2 N_0}{E_{ps}}$. Then, the probability that a transmitted carrier being sensed at distance x from the source is given by:

$$\chi(x) = P(\gamma_{avg} \geq \eta_{th}) = \sum_{n=1}^{N} e^{-(N\lambda)\eta_{th}} \cdot ((N\lambda)\eta_{th})^n / n!, \tag{21}$$

where η_{th} is the power threshold for sensing a transmitted carrier. Therefore, the probability of sensing at least one new transmission at a certain node n_0 within a slot is given by:

$$p_{sense} = 1 - \prod_{i=1}^{N} (1 - p_i \chi(x_i)), \tag{22}$$

where p_i is the transmission probability of node n_i, and x_i is the distance between n_i and n_0. We assume all the nodes have the same p_i, whose value is p. By averaging over different node positions, we get the average probability of sensing a new transmission within one slot as follows:

$$r = 1 - exp(-\delta p \int_{-\infty}^{\infty} \chi(|x|) dx), \tag{23}$$

where

$$\int_{-\infty}^{\infty} \chi(|x|) dx = \sqrt{\pi/\rho} + \sum_{n=1}^{N-1} [\frac{(n-\frac{1}{2})(n-\frac{3}{2}) \cdots (\frac{1}{2})}{n!} \cdot \sqrt{\pi/\rho}], \tag{24}$$

and $\rho = N \cdot \frac{\eta_{th}}{E_{ps}}$.

Therefore, the average carrier sense probability within one slot can be simplified as:

$$r = 1 - exp(-\frac{\sqrt{\pi}\delta p d_{cs}}{\sqrt{N}} \{1 + \sum_{n=1}^{N-1} [\frac{(n-\frac{1}{2})(n-\frac{3}{2}) \cdots (\frac{1}{2})}{n!}]\}), \tag{25}$$

where d_{cs} is the carrier sense range.

REFERENCES

[1] S. Katti, D. Katabi, H. Balakrishnan, and M. Medard, "Symbol-level network coding for wireless mesh networks," in *SIGCOMM '08*, 2008, pp. 401–412.

[2] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Transaction on Infomation Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[3] J. Liu, D. Goeckel, and D. Towsley, "The throughput order of ad hoc networks employing network coding and broadcasting," in *IEEE MILCOM 2006*, 2006.

[4] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the gain of network coding and broadcasting in wireless networks," in *INFOCOM 2007, IEEE*, May 2007, pp. 724–732.

[5] S. Karande and Z. Wang, "On the multicast throughput capacity of network coding in wireless ad-hoc networks," in *FOWANC'09*, May 2009, pp. 21–27.

[6] A. Ramamoorthy, J. Shi, and R. Wesel, "On the capacity of network coding for random networks," *Information Theory, IEEE Transactions on*, vol. 51, no. 8, pp. 2878 –2885, 2005.

[7] Z. Kong, S. Aly, E. Soljanin, E. Yeh, and A. Klappenecker, "Network coding capacity of random wireless networks under a signal-to-interference-and-noise-ratio model," in *Allerton '07*, 2007.

[8] D. S. Lun, M. Mdard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," *Physical Communication*, vol. 1, no. 1, pp. 3 – 20, 2008.

[9] Z. Yang, M. Li, and W. Lou, "Codeplay: Live multimedia streaming in vanets using symbol-level network coding," in *Network Protocols, 2010. (ICNP '2010) Proceedings. Eighteenth International Conference on*, November 2010.

[10] M. Li, Z. Yang, and W. Lou, "Codeon: Cooperative popular content distribution for vehicular networks using symbol level network coding," *IEEE Journal on Selected Areas in Communications(JSAC)*, vol. 29, no. 1, pp. 223 –235, January 2011.

[11] M. Nekoui, A. Eslami, and H. Pishro-Nik, "Scaling laws for distance limited communications in vehicular ad hoc networks," in *Communications, 2008. ICC '08. IEEE International Conference on*, May 2008, pp. 2253 –2257.

[12] M. Johnson, L. D. Nardis, and K. Ramch, "Collaborative content distribution for vehicular ad hoc networks," in *Allerton Conference on Communication, Control, and Computing*, September 2006.

[13] D. Lun, N. Ratnakar, M. Medard, R. Koetter, D. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum-cost multicast over coded packet networks," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2608 –2623, 2006.

[14] Y. Wu, M. Chiang, and S.-Y. Kung, "Distributed utility maximization for network coding based multicasting: A critical cut approach," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on*, 2006, pp. 1 – 6.

[15] Y. Wu, P. Chou, and S.-Y. Kung, "Minimum-energy multicast in mobile ad hoc networks using network coding," *Communications, IEEE Transactions on*, vol. 53, no. 11, pp. 1906 – 1918, 2005.

[16] H. S. Wang and N. Moayeri, "Finite-state markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, February 1995.

[17] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 3, pp. 535–547, 2000.

[18] D. Jiang, Q. Chen, and L. Delgrossi, "Communication density: A channel load metric for vehicular communications research," in *IEEE MASS 2007*, October 2007, pp. 1–8.

[19] J. G. Proakis, "Digital communications," August 2000.

[20] J. M. D. Castillo and F. Benitez, "On the functional form of the speed-density relationship-i: general theory," *Transportation Research Part B: Methodological*, vol. 29, pp. 373–389, October 1995.