XOR Network Coding for Data Mule Delay Tolerant Networks

Qiankun Su, Katia Jaffrès-Runser, Gentian Jakllari, Charly Poulliat

XOR Network Coding for Data Mule Delay Tolerant Networks

(Invited Paper)

Qiankun Su, Katia Jaffrès-Runser, Gentian Jakllari and Charly Poulliat
University of Toulouse
IRIT-INPT/ENSEEIHT
{qiankun.su, kjr, jakllari, charly.poulliat}@enseeiht.fr

***Abstract*—We propose a simple yet efficient scalable scheme for improving the performance of Delay Tolerant Networks (DTNs) with data mules by using XOR network coding. We carry out a theoretical analysis based on a model abstracted from the Village Communication Networks (VCNs), beginning with two villages and then extending to N villages. We also examine how the delivery probability is affected by the different overlapping intervals of two data mules. The theoretical analysis indicates that the maximum delivery probability increases by 50% and our simulation results illustrate this point, showing that the overhead ratio and average delay are reduced as well. Finally, our scheme is applied to a real network, the Toulouse public transportation network. We analyze the dataset, calculate the overlapping intervals of inter-vehicles and the amount of data that transit vehicles can exchange in one day, showing a 54.4% improvement in throughput.**

***Index Terms*—Delay tolerant networks, XOR, data mules, network coding, village communication networks**

I. Introduction

In the face of the emerging need for anytime, anywhere, anyplace communications, computer networks are expected to operate in a variety of, and sometimes very challenging, environments. These challenges stem from high delay, such as inter-planetary networks, harsh propagation environments, node mobility, lack of infrastructure, such as rural and remote areas, and military battlefields, limited power, such as mobile sensor networks, etc. DTNs, opportunistic networks or wireless sensor networks constitute typical challenged network examples [1].

An interesting example of DTNs is VCN as introduced in [2]. In this case, there is no decent communication infrastructure between remote villages and cities of a country. Communication between the inhabitants of these villages is made possible thanks to so-called data mules. These mules are for instance vehicles (e.g. buses, cars, etc.) carrying wireless transceivers and a given storage capacity. Each time a data mule visits a village or a city, it uploads the data dedicated to its inhabitants on a central server and downloads the data to be transferred to other remote destinations from this server. The mule carries the data from one village to another one, possibly fetching new data to a city connected to Internet. In such a scenario, messages take time to be transferred which depends on the mules trajectory. This communication paradigm is reminiscent of legacy mail and thus, only asynchronous communications of non delay-sensitive data is possible. It is possible in such a scenario to leverage the network created by bus lines to provide a communication infrastructure at low cost for developing countries. Real implementations of VCNs have already been tested in the last decade. DakNet [2], developed by MIT Media Lab researchers, provides low-cost digital communication for remote villages in India and Cambodia by using public buses as data mules. The UMass DieselNet [3] is running on 35 buses in Amherst.

To deploy such a VCN, buses have to be equipped with wireless communication capabilities using a standard wireless LAN interface (e.g. IEEE802.11) to connect to the access points located at bus stops in the villages or the city. To reduce end-to-end communication delay, it is important for buses to meet for some time at the same location to exchange data using this available wireless LAN. This scenario benefits end-to-end delay, but it reduces significantly the WLAN throughput available for each bus. Indeed, the bus stop access point bandwidth is shared by the buses and thus, the size of the data to be transferred at the bus station is limited. This paper proposes a simple yet efficient solution to overcome this limitation. Here, we do not provide a new protocol for VCNs, but we show how to efficiently leverage inter-session XOR network coding at the contention points of the network to significantly gain in throughput.

Network coding has been shown to significantly improve the throughput in regular networks, especially in wireless networks owing to the broadcast nature of wireless medium [4]. Relay nodes linearly combine sets of messages they have received to create a new network coded messages that can be forwarded to different next-hop nodes. Different combinations can be created by relay nodes by choosing different encoding coefficients. Typically, a coded message m_c is obtained as the weighted sum of a set of K messages $m_{i,i\in[1..K]}$ following $m_c = \sum_{i=1}^{K} \alpha_i m_i$, where $\alpha_i, i = 1 \cdots K$, are randomly chosen scalar elements from a finite field $\mathbb{F}_q$. In this paper, we consider basic XOR-network coding where messages are simply xor-ed together $m_c^{xor} = \oplus_{i=1}^{K} m_i$.

The destination nodes extract the desired message from a set of combined (i.e. coded) messages by solving the linear system created by these coded messages and their coefficients. If only messages that belong to the same session are network coded together, we refer to intra-session network coding. One session is defined here by the sequence of messages sent by a source node to one (unicast) or more destination nodes (multicast). In this paper, we propose a solution where *inter-session* XOR network coding is performed. In this case, messages that belong to different sessions are network coded together by the relay nodes.

In this paper, we model first a basic village to village communication scenario where two sessions are network coded together at the city bus stop. We draft the main lines of an inter-session XOR network coding exchange protocol to analyse the performance gain compared to a basic store-and-forward mechanism. Next, we extend this protocol to handle N village-to-village communications crossing at the same city bus stop. The performance analysis of this scenario enables us to calculate the gain that can be achieved in a multi-session VCN compared to basic store-and-forward. Finally, we extend the analysis to calculate the gain such a mechanism could bring if it would be deployed in a real public transportation network. The example of the bus network of the city of Toulouse is used to calculate theoretically the gain achieved by inter-session XOR network coding.

This paper is organized as follows. Section II summarizes the main related works. The system model and theoretical analysis are presented in Section III and Section IV, respectively. Main results for the Toulouse bus lines are given in Section V. Section VI shows how our scheme benefits a real network. Finally, Section VII concludes this work.

II. Related work

In DTNs, popular ad hoc routing protocols such as AODV (Ad hoc On-Demand Distance Vector) or DSR (Dynamic Source Routing) fail to route messages since there is no end-to-end path available at all times. Thus, specific protocols leveraging the store-carry-and-forward principle [5] have been proposed to handle delay tolerant communications. The baseline protocol is the epidemic protocol [6]. This protocol is similar to a viral infection process where nodes continuously replicate messages to newly discovered contacts until messages are delivered. This protocol provides the best delivery probability, but at the cost of flooding the whole network with unnecessary message copies. Dedicated protocols such as the MaxProp protocol [3] have been designed for VCNs. Their purpose is to route each message from one village to another one using the features of the underlying DTN network. The purpose of this paper is not to provide another protocol for VCNs, but to show how to efficiently leverage inter-session XOR network coding at the contention points of the network to significantly gain in throughput.

Network coding was first proposed in the pioneer paper of Ahlswede et al. [7]. In [4], the authors developed COPE to improve the throughput of wireless mesh networks by using XOR network coding. Unlike traditional approaches in which intermediate nodes simply act as store and forward routers, relay nodes in COPE combine packets from different sessions using the XOR operator and then forward the XOR-ed packet to the next-hop in a single transmission. However, this scheme is not effective for networks consisting of sparse nodes such as DTN for lack of overhearing neighbours' transmission [8]. Some works have implemented intra-session network coding [9], leveraging the erasure codes for reliability of a single flow. We focus here on inter-session network coding to gain in throughput for all the flows of the bus network.

Recent works have presented intra- and inter-session network coding solutions for delay tolerant networks [8], [10]–[13]. Most of these works have focused on people-centric DTNs where several unicast communications between mobile pairs of nodes exist. Gains in throughput have been shown most of the time for nodes moving with homogeneous mobility models and when there are high resource constraints (e.g. limited bandwidth and buffer). But this is often at the cost of an increased computation complexity or an important signalling load. For instance, in [13], the authors show how challenging it is to apply inter-session random network coding to people-centric DTNs without information on the underlying network topology. In the proposed solution, authors exploit properties of the social graph (given by the SimBet protocol) to improve the routing performance but at the cost of having all nodes memorizing and updating social graph structure information. Our work proposes a simple implementation of XOR network coding reminiscent of the solution proposed by Kreishah et al. in [10]. In this work, pairwise inter-session network coding is performed to provide efficient decentralized rate control. In our case, we perform pairwise inter-session XOR network coding between the two sessions shared by two remote villages to reduce the limitation created by the limited bandwidth available at the bus stops.

III. System model

A. VCN scenarios

In this paper, we consider a VCN where a city interconnects several villages thanks to data mules that are materialized by public transportation buses that move back and forth between the villages and the city. Buses are equipped with a wireless communication interface that connects to base stations located in the city and the villages. Storage is available at buses and at base stations. In our analysis, we assume that storage capacity is unbounded. Each village creates a message every Δ unit time, with Δ the message creation period.

In our model, we assume that the final destination of each bus is the city, and that the bus moves back and forth between the city and one or more remote villages. When a bus connects to a village base station, it uploads the data intended for the village inhabitants and downloads the data to be transferred outbound. When it connects to the city base station, it uploads the data coming from the villages crossed and downloads the data for its way back to the villages. The city base station routes the messages between the different bus lines.

The first scenario investigated in this paper is the basic *village-to-village* or 2-village communication depicted on Fig. 1. In this scenario, we assume two remote villages are connected to the city by two buses. The city base station has

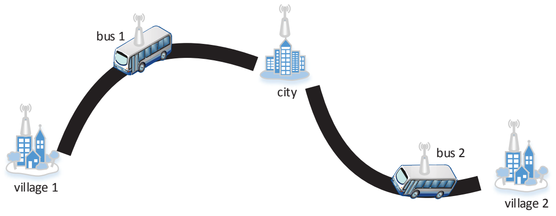

Fig. 1: A village-to-village communication network

TABLE I: Notations and definitions

Term	Description
t_v	bus waiting time in a village
t_b	bus travel time
t_c	bus waiting time in the city
T	Round-trip time of a bus, equivalent to $(2t_b + t_v + t_c)$
Δ	message creation period at villages, $\Delta > 0$
L	number of messages carried by one bus, equivalent to T/Δ
N	number of delivered messages *without* network coding
N_{nc}	number of delivered messages *with* network coding
P	delivery probability *without* network coding, defined by $N/(L \cdot N_v)$
P_{nc}	delivery probability *with* network coding, defined by $N_{nc}/(L \cdot N_v)$
G_p	delivery prob. gain with network coding, defined by $(P_{nc} - P)$
N_v	number of remote villages
t_1	waiting time that line1 stays alone at the station
t_{12}	overlapping intervals that both line1 and line2 stay at the station
t_2	waiting time that line2 stays alone at the station
r	ratio of overlapping intervals, defined by $t_{12}/(t_1 + t_{12} + t_2)$
R	bit rate or transmission rate of base station
D	amount of data two buses are able to exchange *without* network coding
D_{nc}	amount of data two buses are able to exchange *with* network coding
G_t	throughput gain achieved by network coding, defined by $(D_{nc} - D)/D$

two message queues, storing the messages to be downloaded by bus1 or bus2 respectively.

In the second scenario, we extend the previous scenario to N_v villages interconnected by the city. In this scenario, all buses arriving at the city upload their data to the base station and download the data intended to the villages on their way back. The city base station has N_v message queues. It is a one-to-one correspondence among message queues, buses and villages. The data uploaded by a bus is stored in the corresponding queue in the light of its destination address.

Main notations are summarized in Table I, which includes symbols and definitions used in the rest of the paper. The whole analysis is performed for an homogeneous setting: all buses stay for the same duration t_v at the village and t_c at the city. They have as well the same travel time t_b. Notations are illustrated in Fig. 2 for the 2-village scenario.

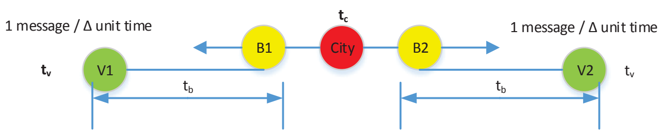

Fig. 2: Village-to-village communication network model

B. End-to-end communication delay

In both scenarios, we are interested in minimizing the end-to-end communication delay between the two communicating villages. The end-to-end communication delay between two villages is the sum of $i)$ the time the two buses stay at the villages, $ii)$ the travel time of the two buses and, $iii)$ the duration the data is being stored at the city. The duration the data is being stored at the city is minimized if the two buses arrive and leave at the same date. In this case, the time the data is being stored is equal to the duration t_c a bus stays at the city.

In Section IV, we first model scenarios 1 and 2 with the assumption that buses arrive and leave at the same date. Next, this assumption is relaxed, and we look at the case where the times the buses stay at the city overlap, but not completely.

C. Medium access control

To be fair, we assume a medium access control at the base station that divides the bandwidth equally between the contending nodes. Thus, any fair medium access control mechanism (e.g. Carrier Sense Multiple Access (CSMA), Time Division Multiple Access (TDMA), etc.) can be implemented in practice. For the 2-village scenario, when the two buses are connected to the base station, there are three contending nodes: the two buses and the access point that has to download the data to the buses.

Since three contending nodes access the network with a fair MAC protocol, each one of them gets a third of the communication bandwidth. If we assure a basic symmetric communication where both buses have to upload and download the same quantity of data D, the base station has to transfer $2D$ data units to the two buses while each buses have to upload D data units. There is a clear contention problem where the base station needs twice as much bandwidth as the buses.

Of course, the problem scales with the same imbalance for the N_v villages scenario. If the N_v buses are connected during the same interval t_c to the base station, the base station needs twice more bandwidth than the whole set of buses together (i.e. $2N_v * D$). And the base station gets only a $1/(N_v + 1)$ share of bandwidth at the MAC layer.

D. XOR network coding operations

It is possible to mitigate this unbalanced bandwidth demand by introducing XOR network coding at the city base station, and thus have the base station only require a fair share of the available bandwidth. The bus1-city-bus2 network can be interpreted as two-way relay network where the benefit of XOR network coding is well known.

Fig. 3 illustrates how network benefits from XOR network coding. Without network coding, it takes 4 transmissions to exchange the two packets $P1$ and $P2$ between bus B1 and bus B2, whereas it only takes 3 transmissions with network coding. After receiving $P1$ and $P2$, the base station broadcasts in a single transmission the network coded packet $P3 = P1 \oplus P2$. Both buses can extract the new packet by xor-ing with the

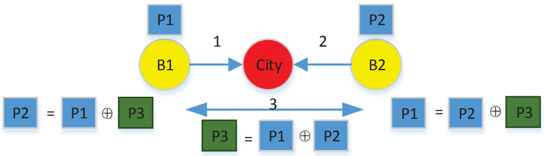

Fig. 3: Inter-session XOR network coding benefit

packet they have previously sent. For instance, B1 obtains $P2$ by calculating $P3 \oplus P1$. With XOR network coding, the base station needs a single transmission instead of two, providing the most efficient use of the underlying fair communication MAC protocol.

As we will show in the following section, an important improvement can be achieved for the N_v village communication network using XOR network coding as well. The N_v bus lines are paired together, creating $\lfloor N_v/2 \rfloor$ two-village communication networks. Thus, for each village pair, one base station communication is spared compared to the no network coding case.

A possible implementation of this mechanism is to have the base station do a transmission of the xor-ed packet of randomly selected village pair with nonempty message queue. The base station reads the two head-of-line messages corresponding to the queues of the two villages composing the pair, combines them with a simple XOR operation and transmits them using the underlying fair MAC protocol. The two buses simply obtain the new message to carry to the remote villages performing the xor operation mentioned earlier. If failing to find a pair of villages with nonempty queue (no xor probability), the transmission is degenerated into the basic store-and-forward one.

IV. Theoretical Analysis

In this section, we analytically derive the gain in delivery probability G_p when performing our pairwise inter-session XOR network coding compared to the no-coding forwarding at the city base station. The formal definition of G_p is given in Table I. In the following, we derive for various scenarios its expression as a function of the message creation period Δ, i.e. of the network load. Here, all villages create messages with the same period Δ. The scenarios of 2-village and N_v-village are investigated assuming first that all buses arrive and leave the base station at the same time. Next, this assumption is relaxed, and the times buses are connected to the base station only overlap partly.

A. Two-village scenario

In the case that all nodes can make full use of the available bandwidth, the maximum number of messages each bus can transfer is $t_c/3$ with a rate of $R = 1$ message per time unit. The base station transfers $t_c/3$ messages to two buses without network coding, but the number is doubled ($2 \cdot t_c/3$) with network coding.

The delivery probability without network coding P, with network coding P_{nc} and the gain G_p is calculated for different message creation periods by:

- For $\Delta < 3 \cdot T/t_c$: In this case, more messages are created per bus line than the base station can serve at best with network coding, and no protocol can drain all messages. In other words, the number of messages carried by one bus, denoted L, exceeds $t_c/3$.

Thus, we have $P = \Delta \cdot \frac{t_c}{6 \cdot T}$, $P_{nc} = \Delta \cdot \frac{t_c}{3T}$ and $G_p = \Delta \cdot \frac{t_c}{6 \cdot T}$.

- For $\Delta \in [3 \cdot T/t_c, 4 \cdot T/t_c)$: In this case, the base station can only drain all created messages if it uses network coding. Here, we have $L \leq t_c/3$ (permitting complete message drain with network coding) and $L > t_c/4$ (non complete message drain without network coding). Following, we have a coding probability of $P_{nc} = 1$ and

$$P = \frac{N}{2 \cdot L} = \frac{t_c - 2 \cdot L}{2 \cdot L} = \frac{t_c}{2 \cdot T} \cdot \Delta - 1 \tag{1}$$

$$G_p = P_{nc} - P = 2 - \frac{t_c}{2 \cdot T} \cdot \Delta \tag{2}$$

The gain reaches its maximum of $G_p = 0.5$ at $\Delta = 3 \cdot T/t_c$.

- For $\Delta \geq 4 \cdot T/t_c$: The base station can drain all messages with and without network coding as $L \leq t_c/4$. Thus $P = 1$, $P_{nc} = 1$ and $G_p = 0$.

B. N_v-village scenario

The previous analysis is extended to N_v villages for the pairwise inter-session XOR network coding protocol described earlier. As N_v buses enter the city, (N_v+1) nodes compete for the channel with an access probability of $1/(N_v+1)$. For this scenario, the delivery probabilities P and P_{nc} and the gain in probability G_p is calculated as a function of Δ.

In this scenario, the maximum number of messages a bus can upload is $t_c/(N_v + 1)$ with a rate of $R = 1$ message per time unit. If *no network coding* is applied, the maximum number of messages that N_v buses can download from the base station is $t_c/(N_v + 1)$. If pairwise *network coding* is applied, this number is doubled since for any two uploads, there is only one download from the base station.

- For $\Delta \leq (N_v+1) \cdot T/t_c$: The number of packets created L is strictly higher than $t_c/(N_v+1)$ and thus the base station can not drain all packets whichever protocol is being used. In this case, we have:

$$P = \frac{t_c/(N_v+1)}{N_v \cdot L} = \frac{t_c \cdot \Delta}{N_v \cdot (N_v+1) \cdot T} \tag{3}$$

$$P_{nc} = \frac{2t_c/(N_v+1)}{N_v \cdot L} = \frac{2 \cdot t_c \cdot \Delta}{N_v \cdot (N_v+1) \cdot T} \tag{4}$$

$$G_p = P_{nc} - P = \frac{t_c \cdot \Delta}{N_v \cdot (N_v+1) \cdot T} \tag{5}$$

- For $\Delta \in \left(\frac{(N_v+1) \cdot T}{t_c}, \frac{3 \cdot N_v \cdot T}{2 \cdot t_c}\right)$: In this case, we have $L < \frac{t_c}{N_v+1}$ and $L > \frac{2 \cdot t_c}{3 \cdot N_v}$. Network coding can still not drain all

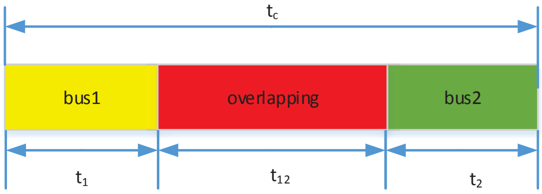

Fig. 4: The overlapping waiting time at the station

messages since it can at most transmit $\frac{2 \cdot t_c}{3 \cdot N_v}$, but it gets more efficient than in the previous case. Here, we have:

$$P = \frac{t_c - L \cdot N_v}{L \cdot N_v} = \frac{t_c \cdot \Delta}{N_v \cdot T} - 1 \quad (6)$$

$$P_{nc} = \frac{(t_c - L \cdot N_v) \cdot 2}{L \cdot N_v} = \frac{2 \cdot t_c \cdot \Delta}{N_v \cdot T} - 2 \quad (7)$$

$$G_p = P_{nc} - P = \frac{t_c \cdot \Delta}{N_v \cdot T} - 1 \quad (8)$$

• For $\Delta \in \left[\frac{3 \cdot N_v \cdot T}{2 \cdot t_c}, \frac{2 \cdot N_v \cdot T}{t_c}\right)$: In this case, network coding can drain all messages but regular forwarding cannot. Indeed, $L \leq \frac{2 \cdot t_c}{3 \cdot N_v}$ but $L > \frac{t_c}{2 \cdot N_v}$, the maximum number of messages regular forwarding can manage. Here, $P_{nc} = 1$ and:

$$P = \frac{t_c - L \cdot N_v}{L \cdot N_v} = \frac{t_c \cdot \Delta}{N_v \cdot T} - 1 \quad (9)$$

$$G_p = P_{nc} - P = 2 - \frac{t_c \cdot \Delta}{T \cdot N_v} \quad (10)$$

The gain reaches the maximum ($\frac{1}{2}$), while $\Delta = \frac{3 \cdot N_v \cdot T}{2 \cdot t_c}$ where the city node is *just* able to drain all created messages with network coding.

• For $\Delta \geq \frac{2 \cdot N_v \cdot T}{t_c}$: The base station drains all created messages even without using network coding. Thus $P = 1$, $P_{nc} = 1$ and $G_p = 0$.

C. *Different overlapping intervals*

The above discussion is based on the assumption that buses are perfectly synchronized: they arrive and leave at the same time at the city. However, it does not reflect real networks and therefore, we examine what happens when intervals overlap partially.

Fig. 4 represents the main notations used to represent the partially overlapping intervals of two buses. Here, t_c is divided into three parts: interval t_{12} where both buses are connected to the base station, and two evenly spaced intervals where either bus1 or bus2 is connected. It is possible to derive the expressions of P, P_{nc} and G_p as a function of t_{12}. For the sake of conciseness, only the case where the base station cannot drain messages with network coding for the 2-village scenario is given with the assumption of $t_1 = t_2$:

$$P = \frac{\frac{t_c - t_{12}}{2} + \frac{t_{12}}{3}}{2 \cdot L} = \frac{3 \cdot t_c - t_{12}}{12 \cdot T} \cdot \Delta \quad (11)$$

$$P_{nc} = \frac{\frac{t_c - t_{12}}{2} + \frac{t_{12}}{3} \cdot 2}{2 \cdot L} = \frac{3 \cdot t_c + t_{12}}{12 \cdot T} \cdot \Delta \quad (12)$$

TABLE II: Model and simulation parameters

Model Parameter	Value	Simulation Parameter	Value
t_v	100s	simulation duration	43200s
t_b	100s	time-to-live	6000s
t_c	100s	buffer size	infinite
N_v	2, 4	-	-

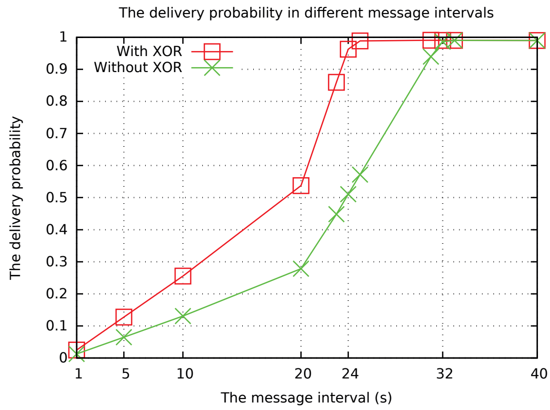

Fig. 6: The delivery probability for $N_v = 4$

$$G_p = P_{nc} - P = \frac{t_{12}}{6 \cdot T} \cdot \Delta \quad (13)$$

As we can see, the more overlapping intervals, the more improvements achieved by network coding. The other cases are plotted in Section V, Fig. 7.

V. Results

The previous theoretical analysis has been conducted for the model parameter presented in Table II ('unit time' is equivalent to one second). Theoretical results are compared in the following to simulations performed with The ONE simulator [14].

A. *Two-village scenario*

Fig. 5-(left) represents both theoretical and simulation-based delivery probabilities as a function of the message creation period Δ. Simulation results are consistent with the theoretical ones and in this case the maximum gain is achieved at $\Delta = 12$. In addition, our simulation results also show that both the overhead ratio (i.e. the ratio of the number of transmissions to the number of messages delivered) and the average latency reduce sharply with our scheme (cf. in Fig. 5-(middle) and (right) respectively).

B. *N_v-village scenario*

As we can see from the theoretical and simulation results shown in Fig. 6 for $N_v = 4$, the maximum gain is still of around 50% while $\Delta = 24$.

C. *Different overlapping intervals*

To plot Fig. 7, we have set the message creation interval to 12 to see how t_{12} affects the delivery probability. Not surprisingly, the increase of the overlapping interval increases the

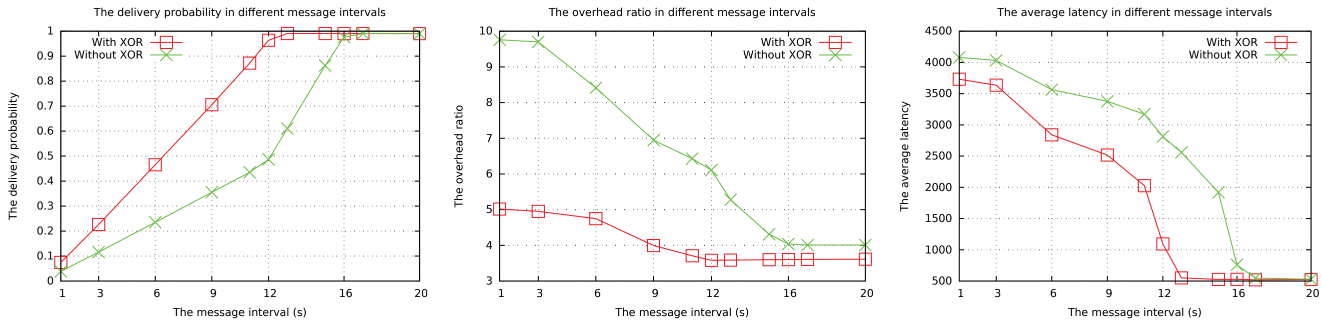

Fig. 5: Delivery probability (left), overhead ratio (middle) and average latency (right) for $N_v = 2$

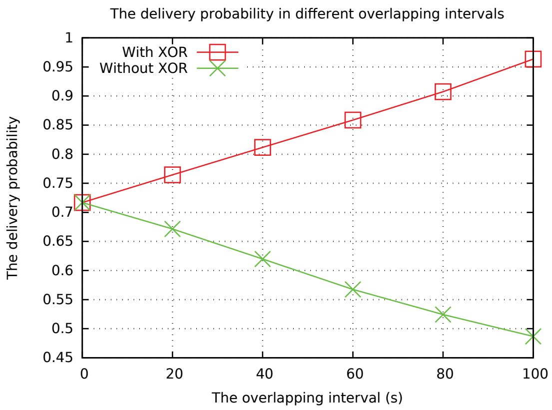

Fig. 7: The delivery probability in different overlapping intervals ($N_v = 2$)

delivery probability of network coding while the one of non-network coding drops rapidly because of channel competition.

VI. Discussion on Expanding to Real Networks

Motivated by the discussion in Section IV-C, we are interested in computation the gain of our scheme for a real public transportation network. Therefore, based on the Toulouse open data for the Tisseo bus network [1], we have calculated the overlapping intervals of any two buses and the amount of data two buses can exchange in one day. From these values, we can derive the throughput gains that can be achieved with our pairwise inter-session XOR network coding solution.

We only consider the origin and destination stops since the waiting time of transit vehicles at stops is not recorded in this dataset. The waiting time can be calculated by pairing buses with the assumption that transit vehicles are scheduled in 'FIFO' (first arrival, first departure). In the case that several buses of the same line parked in a station for a certain period of time, the waiting time of this line is simplified as from the arrival time of the first bus to the departure time of the last bus for the reason that the base station can only establish a connection with one of them at the same time.

With the overlapping intervals of a pair of buses $(line1, line2)$, the amount of data that two buses can exchange in one day is be calculated with:

$$D = [min(t_1, t_2) + \frac{1}{3} \cdot t_{12}] \cdot R \tag{14}$$

$$D_{nc} = [min(t_1, t_2) + \frac{2}{3} \cdot t_{12}] \cdot R \tag{15}$$

Table III shows the throughput improvements achieved by network coding in descending order on the condition of $R = 100Mb/s$ and the weekday service type [2]. Take the 8$^{\text{th}}$ row $(10s, 2s)$ as an example, the line $10s$ and $2s$ make a cross at *Cours Dillon*. Within one day, the overlapping intervals of them are 1 hour 3 minutes and therefore the ratio of the overlapping intervals is 82.89%. The amount of data that those two lines can exchange is $16.5GB$ *without* network coding but $32.25GB$ *with* network coding, an improvement of 95.45%.

VII. Conclusion

In this paper, we proposed a simple yet efficient scalable scheme for improving the performance of DTNs with data mules by using XOR network coding. We showed that our scheme improves the throughput and delivery probability while also reducing the overhead ratio and average latency.

References

[1] M. R. Schurgot, C. Comaniciu, and K. Jaffres-Runser, "Beyond traditional dtn routing: social networks for opportunistic communication," *IEEE Communications Magazine, Volume 50, Number 7, July 2012*, vol. 50, no. 7, pp. 155–162, 2013.

[2] A. S. Pentland, R. Fletcher, and A. Hasson, "Daknet: Rethinking connectivity in developing nations," *Computer*, vol. 37, no. 1, pp. 78–83, 2004.

[3] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "Maxprop: Routing for vehicle-based disruption-tolerant networks." in *INFOCOM*, vol. 6, 2006, pp. 1–11.

[4] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking (ToN)*, vol. 16, no. 3, pp. 497–510, 2008.

[5] N. Chakchouk, "A survey on opportunistic routing in wireless communication networks," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.

[6] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," Technical Report CS-200006, Duke University, Tech. Rep., 2000.

[1] https://data.toulouse-metropole.fr/explore/dataset/tisseo-gtfs/

[2] Four service types: Weekday, Saturday, Sunday and Holiday

TABLE III: The throughput improvements achieved by network coding

line1	**line2**	t_1	t_{12}	t_2	r**(%)**	D**(GB)**	D_{nc}**(GB)**	G_t**(%)**	**station**
10s	38s	0:00:00	1:15:00	0:56:00	57.25	18.75	37.5	100.00	Cours Dillon
2s	38s	0:00:00	1:04:00	1:07:00	48.85	16.0	32.0	100.00	Cours Dillon
2s	78s	0:17:00	0:34:00	0:00:00	66.67	8.5	17.0	100.00	Université Paul Sabatier
2s	81s	0:00:00	0:51:00	1:37:00	34.46	12.75	25.5	100.00	Université Paul Sabatier
78s	81s	0:00:00	0:34:00	1:54:00	22.97	8.5	17.0	100.00	Université Paul Sabatier
79s	81s	0:00:00	1:08:00	1:20:00	45.95	17.0	34.0	100.00	Université Paul Sabatier
T1	T2	12:37:00	6:04:00	0:00:00	32.47	91.0	182.0	100.00	Palais de Justice
10s	2s	0:12:00	1:03:00	0:01:00	82.89	16.5	32.25	95.45	Cours Dillon
2	VILLE	4:29:00	6:54:00	0:13:00	59.48	113.25	216.75	91.39	Cours Dillon
112	62	0:28:00	8:04:00	2:50:00	70.97	142.0	263.0	85.21	Ramonville
111	112	2:48:00	8:03:00	0:29:00	71.03	142.5	263.25	84.74	Ramonville
112	37	0:31:00	8:01:00	3:40:00	65.71	143.5	263.75	83.80	Ramonville
26	36	0:45:00	10:16:00	1:58:00	79.08	187.75	341.75	82.02	Borderouge
2	82	10:44:00	1:39:00	0:07:00	13.20	30.0	54.75	82.50	Université Paul Sabatier
202	62	1:22:00	11:03:00	0:50:00	83.40	203.25	369.0	81.55	Castanet-Tolosan
...	...	...	...	...	...	...	...	...	...
In total:		-	-	-	38.12	26891.75	41425.0	54.04	-

[7] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204–1216, 2000.

[8] S. Ahmed and S. S. Kanhere, "Hubcode: hub-based forwarding using network coding in delay tolerant networks," *Wireless Communication and Mobile Computation*, vol. 13, no. 9, May 2011.

[9] A. Petz, C.-L. Fok, C. Julien, B. Walker, and C. Ardi, "Network coded routing in delay tolerant networks: An experience report," in *Proceedings of the 3rd Extreme Conference on Communication: The Amazon Expedition*. ACM, 2011, p. 4.

[10] A. Khreishah, W. Chih-Chun, and N. B. Shroff, "Rate control with pairwise intersession network coding," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 816–829, June 2010.

[11] M. Heindlmaier, D. Lun, D. Traskov, and M. Medard, "Wireless inter-session network coding - an approach using virtual multicasts," in *IEEE International Conference on Communication*, 2011, pp. 1–5.

[12] X. Zhang, G. Neglia, J. Kurose, D. Towsley, and H. Wang, "Benefits of network coding for unicast application in disruption-tolerant networks," *Networking, IEEE/ACM Transactions on*, vol. 21, no. 5, pp. 1407–1420, 2013.

[13] N. Shrestha and L. Sassatelli, "Inter-session network coding in delay tolerant mobile social networks: An empirical study," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*. IEEE, 2015, pp. 1–6.

[14] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the 2nd international conference on simulation tools and techniques*, 2009, p. 55.