A Fountain Codes-based Data Dissemination Technique in Vehicular Ad-hoc Networks

Veronica Palma*, Elena Mammi†, Anna Maria Vegni*, and Alessandro Neri*

*Department of Applied Electronics, University of Roma Tre, Rome, Italy
Email: {vpalma1, amvegni, neri}@uniroma3.it
†Fondazione Ugo Bordoni, Rome, Italy
Email: emammi@fub.it

Abstract—In this paper we propose a data dissemination technique based on Fountain codes, that is particularly suitable for vehicular environments, where mobility and connectivity issues often occur. We investigate the reliability of a real-time service and data delivery for multicast transmission over lossy Vehicular Ad-hoc NETworks (VANETs).

The proposed technique represents the next work of a previous approach initially applied to traditional Mobile Ad-Hoc NETworks (MANETs). This technique combines application layer channel coding based on Luby Transform (LT) codes with multicast delivery in a wireless network exploiting PUMA (Protocol for Unified Multicasting through Announcements). The main goal is to achieve a real-time service with an high quality level in a lossy vehicular network environment. The rateless property, the ability of adapt the code on-the-fly, makes Fountain codes an attractive solution for data broadcast/multicast applications, like the well-known comfort applications.

In a VANET scenario characterized by dynamic topology conditions, packet losses and disconnections, the proposed approach leads to an improvement on arrival times of packets towards destination vehicles. Simulation results provide the effectiveness of our technique compared with traditional data dissemination approach.

I. INTRODUCTION

Vehicular Ad-hoc NETworks (VANETs) are a particular class of Mobile Ad-hoc NETworks (MANETs), where mobile nodes are vehicles moving at different speeds and forming dynamic network scenarios [1]. VANETs provide data communications among nearby vehicles via Vehicle-to-Vehicle (V2V) protocol, in the support of Internet access, as well as a variety of safety applications. In time-varying connectivity scenarios (i.e. sparse-traffic scenarios), packet delivery represents an important challenge mainly due to MAC layer collisions. Ongoing efforts are aimed at enabling inter-vehicle communications supported by network infrastructure, in order to provide seamless connectivity and efficient data propagation even in sparse-traffic scenarios. Intelligent Vehicular Ad-Hoc Networking (InVANET) has defined a smart novel way of using vehicular networking by integrating on multiple wireless technologies, such as 3G cellular systems, IEEE 802.11, and IEEE 802.16 e, for effective Vehicle-to-Infrastructure (V2I) communications [2].

V2V and V2I communication technology has been developed as part of the Vehicle Infrastructure Integration (VII) initiative [3]. The VII project considers the network infrastructure as composed by several Road Side Unit (RSU) systems, each of them equipped with a 5.9 GHz DSRC transceiver—for communications between vehicles and RSUs—, and a GPRS interface—to forward messages to the backbone networks—. In such heterogenous network environments, data dissemination and delivery are a challenge.

In this paper we propose a packet dissemination scheme based on Fountain codes, with the twofold purpose of both (i) improving throughput in the vehicular network, and (ii) keeping low end-to-end delays. The proposed approach aims vehicles to collect a minimum amount of out-of-turn packets, protected by a Fountain-based coding technique. In the decoding phase, vehicles are then able to reconstruct the original data information flow.

The feature to adapt the code on-the-fly provides Fountain codes as a reliable solution for data dissemination of multicast flows in networks with packet loss probability. Fountain codes’ main property is the generation of a continuous data flow, similar to the action of water falling from a spring into a collecting bucket, [4]. Once the bucket is full, the collecting process ends and further processing on decoding the content of the bucket will take place. Packet collection does not require the information about which droplets are falling into the bucket as long as the bucket is full. In this vision, the use of Fountain codes are an appealing, capacity approaching Application Layer-Forward Error Correction (AL-FEC) solution, particularly suitable for data transmissions over packet losses and time-vary connectivity networks, like VANETs.

Our proposed technique lets vehicles to benefit of opportunistic delivery of data, which is a typical feature of vehicular networks. Basically, we present a two-steps algorithm for efficient data delivery, such as (i) each packet flow is protected by an LT code and is sent to intermediate destination vehicles (i.e. relay nodes); (ii) once the relay vehicles have collected the minimum amount of packets, they forward messages to destination vehicles which can finally reconstruct the original information flow by mean of LT decoding.
The paper is structured as follows. In Section II we investigate previous related works on data dissemination protocols in VANETs, particularly those approaches based on Luby Transform (LT) codes. Section III introduces our proposed technique. Particularly, Subsection III-A and III-B describe the LT codes adopted in our work, and the coding/decoding processes, respectively. In Subsection III-C the main steps of our Fountain coding-based data dissemination algorithm are discussed, and then the main simulation results are shown in Section IV. The comparison of the proposed technique with a traditional related work provides its effectiveness in terms of low packet arrival time and bandwidth occupancy, as well as a recovery of lost packets whenever they occur. Finally conclusions are drawn in Section V.

II. RELATED WORK

Many authors have investigated data dissemination techniques for VANETs. Different schemes are based on particular vehicular communication protocols (i.e. V2V, V2I and I2V) and analyze how messages are propagating in VANETs (i.e. message propagation and end-to-end delivery). Generally, vehicular networks lack of connectivity due to quick disconnections, high mobility and rapidly changing network topology. VANETs suffer from a reliable data delivery specially in sparse-traffic or totally disconnected scenarios, where vehicle density is low or null, respectively [5]. In these scenarios, the amount of packets which can be successfully received by a vehicle depends on the traffic patterns and vehicle speed.

Data dissemination represents a challenge specially in commercial applications (e.g. Internet access, video-on-demand, advertising dissemination, and so on). In comfort applications where data flows are larger than ones in safety applications, message dissemination should be efficient in order to reconstruct a whole data flow from a limited number of received messages.

The potentiality of using Fountain coding for data dissemination in VANETs has been already exploited in [6], [7] and [8]. The main advantages lay on efficient and reliable vehicular communications even in high dynamic networks, since vehicles can reconstruct a whole data flow from a limited low number of received packets.

In [6] Fountain codes are used to encode packets which are disseminating among vehicles in the network. However this approach is limited to communications among vehicles via V2V protocol, and network infrastructure is not considered. In [7] the authors propose VANETCODE, a content distribution scheme based on network coding. This approach assumes the content as divided into smaller blocks which are linearly encoded by vehicles. The use of network coding in VANETs provides a rapid sharing of real-time messages, particularly suitable for comfort applications. However, the VANETCODE approach is limited to data delivery from infrastructure to vehicles (i.e. I2V), and messages are propagating in a bounded area, that is the RSU’s wireless network.

In [8] Cataldi et al. propose a I2V2V scheme, where vehicles can communicate both with network infrastructure (i.e. I2V) and other neighboring vehicles (i.e. V2V). This hybrid scheme provides a cooperating approach between vehicles, since messages are delivered from the infrastructure to a set of relay vehicles, and then directly to the destination vehicles. As a result, the method improves the speed of data delivery in an end-to-end connection (i.e. from a RSU to a destination vehicle). The technique in [8] is based on rateless codes which provides data reconstruction in a fast way with low overhead.

In this paper we relay on the hybrid scheme as proposed in [8], but assume a novel approach of application layer channel coding based on LT codes with multicast delivery with PUMA (Protocol for Unified Multicasting through Announcements). Our technique is an enhanced version of a previous work, suitable for general wireless networks like MANET [9]. We extend such approach, called as Fountain Code based AL-FEC, to vehicular networks based on I2V2V scheme. The choice of an hybrid communication protocol like I2V2V is mainly due to the need of exploiting network infrastructure to avoid disconnections in sparse-traffic scenarios. Moreover, the hybrid approach is particularly suitable for comfort applications which use network infrastructure for advertisements dissemination [10].

III. PROPOSED TECHNIQUE

In this section we describe our proposed technique for an efficient and low traffic load packet dissemination in VANETs. This method exploits the rateless property of Fountain codes by means of LT codes optimized for small message length. The following Subsections III-A addresses on a brief overview of LT codes used in our technique. Then, Subsection III-B, and III-C describe the proposed method, and the related algorithm for data dissemination in VANETs, respectively.

A. LT codes

Luby Transform (LT) codes represent the first practical realization of Fountain codes, [11]. They are sparse-graph codes developed for erasure channels. The ideal Fountain codes are rateless, this means that the coding rate is not fixed a priori. Therefore by the use of Fountain coding the encoding packets can be generated on-the-fly.

The most important characteristic of Fountain codes realizations is that the source data packets can be recovered from any subset of the received packets, given that enough packets are received. The number of packets to be received from a source is $N = (1 + \varepsilon)K$, where $K$ is the number of source data packets and $\varepsilon$ is the overhead. As well known for $K \to \infty$, $\varepsilon$ tends to 0; leveraging on this property, Fountain codes performances grow with high values of $K$. Nevertheless, in our work we consider an LT code that is optimized for small message length. Since the aim of the proposed work is the reduction of the transmission latency, but still achieving an high quality level in a lossy environments, we consider small size messages.

The key factor of the LT codes is the degree distribution, since the efficiency of the codes is strictly depending on this aspect. In our work, we consider an algorithm for iterative
optimization of the degree distribution by using an approach based on importance sampling [12]. The used distributions are based on Soliton Distribution, whose main feature is that the probability of degree-one is less than the probability for degree-two. Basically, the used degree distribution is the following:

\[ p_i = \begin{cases} \eta_1, & \text{for } i = 1, \\ \eta_2, & \text{for } i = 2, \\ \frac{1}{(i-1)!}, & \text{for } i = \{3, 4, \ldots, n\} \text{ and } i \neq 100. \end{cases} \]  

(1)

The distribution is then normalized and the optimized parameters are \( \eta_{\text{opt}} = (0.083, 0.487, 0.032) \). The choice of parameter values is very important, because a wrong choice could lead to poor performance of the code.

In the encoding phase each packet is generated from \( K \) data sources \( s = [s_1, s_2, \ldots, s_K] \) according to the following steps:

1) Each packet is encoded on the basis of the encoding degree \( d_n \), that is the number of source packets which generate the \( n \)-th packet. The encoding degree is randomly chosen from a degree distribution, depending on the number of data sources \( K \);

2) After choosing the number of source packets, the \( n \)-th packet is generated according to the bitwise sum modulo 2 of \( d_n \) packets.

In the decoding phase, a received packet is either completely uncertain packet or completely certain packet. The decoding is an iterative process, done by using information of which source packets are added together in a received packets.

The algorithm finds a \( t_n \) received packet that is related to only one source packet \( s_K \); if no packet is found, the decoding fails. Otherwise, \( s_K \) is set equal to \( t_n \) and is added to all vehicles \( t_n \) that are connected to \( s_K \); then all the connections related to source packet \( s_K \) are removed. Then, the decoding algorithm starts again iteratively.

The decoder task is to recover \( s \) from \( t = Gs \), where \( G \) is the matrix associated with the inter-vehicular connections. Both side of the transmission (i.e. relay and destination vehicles) know the matrix, even when it is randomly generated. A more detailed explanation of decoding process is given in the following Subsection III-B.

B. Coding and Decoding processes

It is important to note that the decoding phase has a high computational cost, because it requires the inversion of the matrix \( G \). As a consequence, the decoding time is not irrelevant. The matrix inversion is a necessary step when a non-systematic Fountain code is used.

In the proposed scenario there are multiple sources that cooperate to send the source information to the same destinations. Concerning this scenario, we propose to build the coding matrix in an efficient manner to allow a faster decoding. The basic idea is to partition the entire coding matrix in a number of \( M \times M \) blocks, where \( M \) is equal to the number of source nodes (i.e. relay vehicles). For instance, if there are \( M \) sources, the LT coding matrix \( G \) will be a block matrix composed by (i) \( M \) sub-matrices (i.e., \( B_m, 1 \leq m \leq M \)), (ii) \( (M - 1) \) sub-matrices (i.e., \( C_{m-1,m}, 2 \leq m \leq M \)), and (iii) \( (M - 1) \) sub-matrices (i.e., \( C_{m,m+1}, 1 \leq m \leq M - 1 \)), laying on the main, upper and lower diagonal of \( G \), respectively. The matrix \( G \) is expressed as follows 1:

\[
G = \begin{bmatrix}
B_1 & C_{1,2} & 0 & \ldots & 0 & 0 \\
C_{2,1} & B_2 & C_{2,3} & \ldots & 0 & 0 \\
0 & C_{32} & B_3 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & B_{M-1} & C_{M-1,M} \\
0 & 0 & 0 & \ldots & C_{M,M-1} & B_M
\end{bmatrix}
\]

(2)

The sub-matrices of the lower and upper diagonal implement the overlapping among adjacent sources. Note that a source data packet matches with the rows of the entire coding matrix. To ensure that the entire matrix \( G \) respects the degree distribution, the \( M \) sub-matrices \( D_m \) defined as follows:

\[
D_1 = \begin{bmatrix} B_1 \\ C_{2,1} \end{bmatrix}, D_m = \begin{bmatrix} C_{m-1,m} & B_m \\ C_{m+1,m} \end{bmatrix}, D_M = \begin{bmatrix} C_{M-1,M} \\ B_M \end{bmatrix}
\]

(3)

are randomly generated in accordance with the selected degree distribution.

The advantage of this technique is that the decoding phase can be performed independently considering each block, thus requiring much less time. This strategy leads to an improvement of the total overhead directly related to the size of the overlapping, i.e. \( M \). However, when the channel used to transmit the encoding packets (i.e. \( t \)-th source to the destinations) has an high value of Packet Loss Rate (PLR) and the destination receives a number of packets that is not enough to perform the decoding phase, it can be carried out by using the \( t \)-th block and its neighbor. This is possible only if there is overlapping between neighboring blocks.

C. Data Dissemination Algorithm

In vehicular networks data dissemination is often affected by vehicle disconnections due to high speed and low density scenarios. It follows that transmission delay and end-to-end latency are a big issue. The purpose of our technique is then to reduce the end-to-end latencies occurring in multicast services. Our method aims to decrease the transmission delay and, at the same time, the occupancy time of the available bandwidth.

In this subsection, we describe the main steps of our data dissemination algorithm.

\(^1\) Notice that there is no constraint on the position of sub-matrices \( B \) within the LT coding matrix \( G \), but all source data packets should have at least one connection with encoded packets.
Message dissemination can be achieved by exploiting both (i) I2V and (ii) V2V communication protocol. The first provides data transmission from RSUs to a set of vehicles — called relay vehicles — moving inside the infrastructure radio coverage. We assumed relay vehicles routing data messages into the network according to PUMA protocol, [13]. Relay vehicles collect the amount of data messages as long as they are inside the RSU’s radio range. Then, V2V communication protocol provides data delivery to destination vehicles driving in the opposite direction respect to relay vehicles. Messages coming from several relay vehicles are collected by destination vehicles which are able to reconstruct the whole data flow. Fig. 1 depicts the considered VANET scenario.

The proposed algorithm is comprised of two phases, characterized by (i) I2V and (ii) V2V communications. Each vehicle in the network should be connected to RSUs via I2V, or to neighboring vehicles via V2V, according to its Infrastructure Connectivity (IC) parameter, [14]. IC parameter gives information whether a vehicle is able to connect to an RSU (i.e. IC = 1), otherwise to neighboring nodes (i.e. IC = 0). IC parameter is then checked at the beginning of the algorithm.

The main phases of our algorithm are the following:

1) I2V phase: as long as each vehicle remains inside the radio range of an RSU, it collects and buffers packets. Due to high mobility issue, vehicles are not able to receive the complete data flow and route partial information towards neighboring vehicles. In this phase the use of PUMA routing protocol is exploited;

2) V2V phase: each destination vehicle receives portions of data flow coming from the relay vehicles, and then decodes the original flow once it has received at least $K(1 + \varepsilon)$ packets, i.e. the useful amount of packets for the decoding task. In this way, destination vehicles are able to recover the whole data flow, through a minimum amount of packets.

Fig. 2 shows the pseudo-code for the proposed data dissemination algorithm. As input the pseudo-code gets (i) the value of IC parameter for each vehicle in the network, and (ii) the $(K, \varepsilon)$ LT parameters. As output the algorithm returns the number of encoded packets $N_{pkt}$ which are received by the destination vehicles. Notice that the communication protocols (i.e. I2V and V2V) vary in each phase of the algorithm. The algorithm does not care if the decoding task has been successful; it only addresses on the reception of the minimum number of LT encoded packets that is necessary for the decoding phase.

IV. SIMULATION RESULTS

In this section we describe the main results of the proposed data dissemination algorithm. The effectiveness of Fountain coding applied to VANETs is proven in terms of packet arrival and decoding times reduction, compared with the traditional data dissemination technique.

We considered a vehicular network scenario, as well as depicted in Fig. 1. Vehicles are moving in clusters and forming interconnected blocks of vehicles, at speed of 10 m/s according to random waypoint model [15], while RSUs are fixed and transmit 64 bytes data packets to every vehicle crossing their radio coverage. Data communications are based on IEEE 802.11p standard.

The whole data flow is comprised of 3000 packets, that is a low number of packets with respect to those used in traditional LT coding [11]. We assumed sources (i.e. RSUs) encoding data flows using modified LT codes, particularly suitable for a reduced number of packets [12]. Once LT coding has been

2We define as traditional technique the approach that does not consider any LT coding on packet transmission.

3We assumed a small packet size in order to avoid high packet loss rates, as considered in [16].
done, RSUs multicast 3000 packets plus an overhead of almost 20% of useful packets to vehicles crossing the RSUs' radio coverage.

We assumed a set of relay vehicles (i.e. three relay vehicles in each RSU’s wireless network), traveling in the N (North) direction. LT encoded packets are transmitted (i) via radio from RSU to relay vehicles with I2V protocol, and then (ii) propagating multi-hop in the S (South) direction (i.e. opposite to relay vehicles’ travel direction). During the time spent inside the wireless network, relay vehicles are able to collect LT encoded packets via I2V communications. After collecting the amount of data packets available during the overall connection to the RSUs, the relay vehicles forward a portion of data information to neighboring opposite destination vehicles (i.e. D1 and D2) via V2V protocol 4.

In Table I, a detailed list of the parameters used in our simulations is reported. In our simulations we used NS-2 software [18], on a laptop Intel Core 2, CPU T5200 with a frequency clock of 1.66 Hz and 1 GB for RAM. The LT coding has been done with different values of η opt, (i.e. (0.083, 0.487, 0.032)), as well as assumed in [9].

Two scenarios have been considered to evaluate our method. The first one represents the ideal case where no packet loss is supposed to be. The second case is, instead, affected by different Packet Loss Rates (PLR), which are mainly due to MAC collisions, random losses and link failures. As a consequence, the high the PLR, the high the overhead in the simulated scenario.

In the first scenario, the arrival time of $K(1 + \varepsilon)$ packets is compared with traditional case, where all the vehicles forward packets without buffering messages. No PLR is introduced and simulations show the effectiveness of our technique in terms of a reduction of arrival times (i.e. 48 s with our technique, compared to 135 s with the traditional approach, as shown in Fig. 3).

In Fig. 4, the arrival time comparison is evaluated for different packet loss rates (i.e. 2, 4, 6, 8%). The total transmission time lasts 270 s and each phase of the simulation, i.e. 12V and V2V, lasts 135 s each one. However, in the simulation results only V2V communication has been considered since it represents the core phase of the proposed algorithm where it is noticeable a transmission delay reduction. Simulation results show that with our technique, transmission times are drastically reduced even when packet loss rates are introduced.

In addition, the use of LT codes allows to recover source data packets from any subset of the received packets, with the constraint that the minimum number of packets has been received, i.e. $N_{\text{pkt}}$. The decoding phase is then done by each destination vehicle.

In the case that a destination vehicle tries to invert the entire decoding matrix, the computational cost is quite high. However, if the sub-matrices strategy is adopted, the computational cost decreases. Indeed, as shown in Fig. 5 by the use of sub-matrices strategy the decoding time decreases (i.e., light black results), with respect to the decoding time needed for the inversion of the complete matrix (i.e., dark black results). The best (i.e. minimum value) and the worst (i.e. maximum value) cases of the computational cost of the proposed strategy are depicted as well. By using the sub-matrices strategy, a destination vehicle can spend from 1/6 up to 1/2 of the time needed to decode the entire matrix, respectively in the best and worst cases.

Finally, notice that there is a difference between the use of sub-matrices and the uncorrelated matrices in the case of high packet loss rate. The sub-matrices strategy results more efficient, since each destination vehicle can exploit its neighbors in order to invert each sub-matrices, and this is

---

4We assumed the well-known bridging technique for V2V communications, which exploits connectivity from other neighboring vehicles in order to propagate messages in the highway [17].
possible due to the overlapping property of sub-matrices.

V. CONCLUSION

In this paper we introduced a novel data dissemination technique for vehicular networks. Our algorithm exploits the benefit of the most important characteristic of Fountain codes, i.e. the source data packets can be recovered from any subset of the received packets, if enough packets are received. A similar version of proposed idea has been previously proven to be effective in traditional MANETs. However, the main strengths are more suitable for vehicular environments, where random network topologies and disconnections often occur.

The main purpose of our data dissemination technique is the minimization of packet arrival time and bandwidth occupancy, as well as the transmitted stream be protected using LT coding. In addition this method maintains the recovery capability of Fountain code also in case of packets losses.

ACKNOWLEDGMENT

This work has been partially supported by Fondazione Ugo Bordoni and RadioLabs Consortium S.p.A.

REFERENCES