Proc. 8th Conference on ITS Telecommunications (ITST), pp. 199-203, Phuket, Thailand, October 2008

Congestion Control for Safety Messages in VANETs: Concepts and Framework

W. Zhang, A. Festag, R. Baldessari, and L. Le
NEC Laboratories Europe, Kurfürsten-Anlage 36, 69115 Heidelberg, Germany
email: zhang|festag|baldessari|le@nw.neclab.eu

***Abstract*— Congestion control in packet data networks has been extensively studied. However, most algorithms are not directly applicable to safety messages in Vehicular Ad Hoc Networks (VANETs), which have stringent requirements on delay, reliability and dissemination area. Although certain algorithms have been investigated, there is still a lack of understanding on the related concepts, and there is no systematic analysis on different mechanisms in the overall context so far. This paper explores related concepts and their dependencies, makes a systematic analyse and presents a framework on congestion control for safety messages in VANETs.**

I. INTRODUCTION

Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications, also referred to as V2X communications, are the enabling technology for road safety applications. Vehicles and road side units equipped with short-range wireless communication devices based on IEEE 802.11p communicate with each other in self-organized networks called VANETs. In principle, V2X communications combined with vehicular on-board and road-side sensors may support road safety by two means: broadcasting periodic safety messages and disseminating event-driven safety messages [1] [2]. These safety messages typically need to be delivered within a geographical area with certain reliability and delay limit. The periodic messages, also called beacons, carry vehicles' status information such as positions and speeds (Fig. 1). Beacons can be generated at the application layer or at the network layer, and are used by neighbouring vehicles to become aware of their surrounding and to avoid potential dangers. Event-driven safety messages are generated when an abnormal condition or an imminent danger is detected, and disseminated within a certain area with high priority (Fig. 2). Critical event-driven messages usually have strong reliability and delay requirements.

It is well known that vehicular communication environments are characterized by highly mobile vehicles, extremely frequent topology changes and a great variation in the number of vehicles in a certain region. To meet the specific requirements of V2X communications in such environments, geographical routing is applied. Geographical routing assumes that vehicles acquire information about their own positions (i.e. geodetic coordinates) via GPS or other positioning systems. If a vehicle intends to send data to a known target geographic location, it chooses another vehicle as message relay, which is located in the direction towards the target position. The same procedure is executed by every vehicle on a multi-hop path until the destination is reached. Results from extensive network simulations and measurements have indicated that geographical routing has good performance in realistic environments [3]. Basically, geographical routing comprises three forwarding schemes: Geographical Unicast provides packet delivery between two nodes via multiple wireless hops; Geographical Broadcast (geocast) distributes data packets to nodes located in a certain geographical region; Topologically-Scoped Broadcast provides re-broadcasting of a data packet from a source to all nodes in the n-hop neighbourhood. Fig. 1b shows a typical scenario of geocast, which disseminates information about an accident to a geographical area. These forwarding schemes are chosen as the core networking protocol for vehicular communications by the Car-to-Car Communication Consortium (C2C-CC), the major European industry consortium for vehicular communications [2].

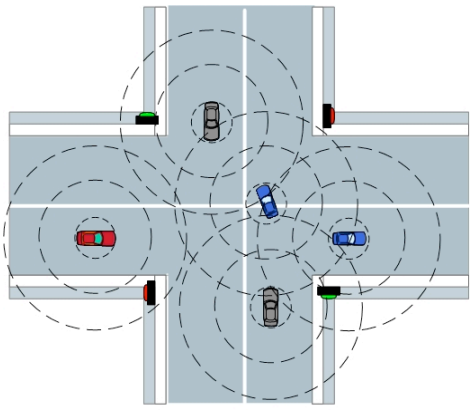

Fig. 1 Periodic beacons sent at certain frequency

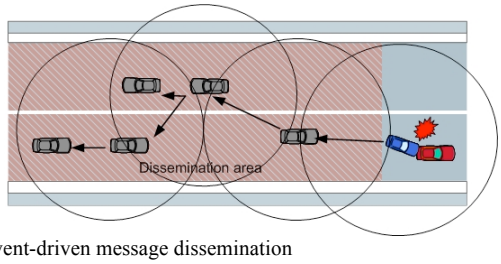

Fig. 2 Event-driven message dissemination

In Europe 30 MHz spectrum in the range from 5.875 to

5.905 GHz is being allocated for critical road safety and traffic efficiency applications in 2008. It is likely that two types of channels will co-exist, namely, one Control Channel (CCH) and one or more Service Channels (SCHs). The CCH will be used for critical road safety applications and beacons, while SCHs will be used for safety and traffic efficiency applications. It is generally believed that a multiple-receiver concept as well as a time synchronous system based on IEEE 802.11p/P1609 system will work, and both ideas will be considered in the evaluation of channel usage schemes.

If a large number of vehicles send beacons at a high frequency or event-driven messages are broadcast multiple times (typical scenarios are illustrated in Fig. 1), the communication channel will easily get congested. It is very important to keep the CCH free from congestion in order to ensure timely and reliable delivery of safety messages. Although congestion control in data networks has been extensively studied, most algorithms do not address safety messages in VANETs. Ongoing discussions indicate that congestion control in VANETs is a cross-layer issue. Although certain algorithms have been investigated, there is still a lack of understanding on the related concepts, and there is no systematic analysis on different mechanisms in the overall context so far. This paper explores related concepts and their dependencies, makes a systematic analyse and presents a framework on congestion control for safety messages in VANETs. This paper is organized as follows. We first explore related concepts on congestion and congestion control. Following this, we present a framework for congestion control by classifying different congestion mechanisms and proposing architecture. Finally, we summarize the paper and point out open questions and future work.

II. CONCEPTS OF CONGESTION CONTROL IN VANETS

A. Congestion and congestion control

Congestion in data networks could be considered as "The condition where the offered load (demand) from the user to the network is approaching or exceeds, the network design limits for guaranteeing the Quality of Service (QoS) specified in the traffic contract."[3] Typical reasons of congestion in data networks are the overbooking of resources or failure in the network. From this definition, two concepts are important to understand congestion, i.e. network load and QoS. Observing data networks, these two concepts are normally well defined. To deal with congestion in data networks, in principle, three steps are taken in sequence [4]:

- monitor networks and detect congestion;
- pass congestion information to protocol instances;
- adjust system operation to cope with congestion.

In practice, most congestion control mechanisms are based on the end-to-end approach at the transport layer. The most prominent protocol is TCP, and a number of variants of TCP have been proposed to cope with congestion in different contexts [6] [7]. The common feature of these protocols is their reactive behaviour, i.e. they take measures in response to some feedback from the network, and behave as closed-loop control. In principle, an open-loop approach may also be applied to cope with congestion. It aims to reduce congestion by designing good protocols in order to prevent congestion from happening in the first place. Thus, it is also called congestion prevention. Mechanisms in this category share the commonality that they take proactive actions without relying on network status. They cover different policies at the transport, network and data link layer such as retransmission, acknowledgement, flow control, admission control, routing algorithm, etc.

We observe that the predominant communication paradigm for safety messages is broadcast, and unicast has very limited usage often confined to one-hop communications. Since critical safety messages have stringent reliability and delay requirements, they must be disseminated with minimum delay. With new connection scenarios and application requirements, congestion control for safety messages in VANETs also requires corresponding new mechanisms. The end-to-end TCP-like congestion control scheme does not fit the requirement anymore. We believe it is necessary that each node takes both the following two approaches in a distributed way:

- Reactive approach: reduce network load in response to locally obtained feedback from the network.
- Proactive approach: reduce network load irrespective of the network load in order to prevent congestion from happening in the first place

B. Network capacity and load

In the typical scenario illustrated in Fig. 1, a vehicle may be located in the transmission ranges of many neighbouring vehicles. With the CSMA/CD medium access scheme used by IEEE 802.11 systems, a vehicle can only send packets if it senses the channel to be free. For such ad hoc networks, the metrics bit-meters/s has been proposed to describe network capacity [8]. However, it may not be a proper measure for safety messages in VANETs since the throughput is not of primary importance. We propose to use the channel busy time seen by a node as the metrics of the network load. It indicates the percentage of time a channel is sensed busy, and it is location dependent and locally obtainable.

In the following, we derive a simple model to describe the channel busy time in VANETs. For the purpose of our discussion, we assume a one-dimensional scenario where all vehicles share the same lane, and each node transmits packets with an average packet transmission rate λ (measured in packets per unit time), a packet size τ (measured in transmission time) and transmit power P_t. Given these parameters, around a node the percentage of time the wireless medium will be busy due to its own transmission is $\lambda\tau$. For example, if there is only one node and its packet arrival rate is λ = 1/s with packet transmission duration τ = 0.1s, the channel busy time will be $\lambda\tau$ = 1*0.1 =

0.1, which means the channel is busy for 10% of the time. Similarly, the channel busy time caused by transmissions of N similar nodes will be the sum of the load from them. For instance, given the parameters from the previous example, if there are 5 nodes located within the sensing range of each other, the total channel busy time will be 0.1*5 = 0.5. The theoretical maximum value of the channel busy time is 100%, which means the channel is always sensed busy. Simulations have shown that with the channel busy time of 40%, the packet loss probability is considerably higher than the case with 20% load [9]. The reason is that increased load leads to more collisions.

The sensing range of a node depends on the transmit power of a packet and the propagation environment. Assuming a packet is transmitted with power P_t, the range within which the packet will be sensed is proportional to P_t1/n, where the value n is the path-loss factor with typical values between 2 to 5 in vehicular environments. We denote the sensing range as $C \cdot P_t1/n$, where C may be considered as a constant for this simple scenario. Given node density β (measured in number of nodes per unit distance) for the one-dimensional scenario, the average channel busy time L caused by all nodes within the sensing range of each other can be formulated as

$$L = C \cdot P_t1/n \cdot \beta \cdot \lambda \cdot \tau . \quad (1)$$

We note that C is a constant and β is a given parameter in equation (1). Thus, in order to reduce the network load, we identify three approaches:

- reduce the packet transmission rate λ;
- reduce the packet transmission duration τ;
- reduce the transmit power P_t .

Actually, these three approaches to reduce network load are not limited to the simple model presented above, but also hold for more complex models. We will refer to these three approaches in the discussion on the framework of congestion control in VANETs.

C. QoS and network performance

There are different approaches to define QoS [10]. However, related work on safety messages on VANETs is rather limited. The VSC project has defined some parameters and presented preliminary communication requirements for selected safety applications [2]. Observing the special requirements of safety messages, we define three parameters for the network performance:

- Dissemination area: the geographical area of interests including its shape and size.
- Latency: the maximum delay of delivering a message within a geographical area.
- Reliability: the minimum probability of receiving a message.

Fig. 3 indicates that the requirements on network performance of various safety applications may vary from each other. In principle, it is very difficult for the network to meet the requirements of all safety messages. It is essential to ensure the dissemination of the most important and relevant messages. Such messages are typically critical emergency messages, which have very stringent reliability and latency requirements (e.g. application B in Fig. 3), and they should be treated with high priority. We propose to categorize different types of safety messages into several classes based on the network performance parameters and priority. With classification, packets from different applications but in the same class may be treated in the network in a similar way. For example, messages with different priorities may be mapped to different IEEE 802.11p priority classes.

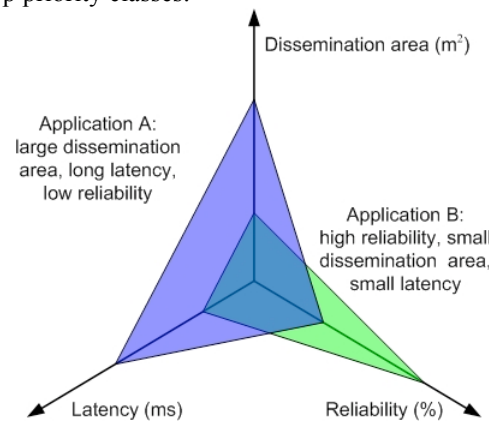

Fig. 3. Network performance parameters with two exemplary safety applications

Actually, there are correlations between these parameters. Usually, to cover a large dissemination area requires that packets be transmitted via multiple hops, thus, will have long latency. Reliability may be improved by increasing the number of retransmissions, which will lead to long latency, or be improved by limiting the dissemination only to the vehicles nearby, which will result in reduced dissemination area. Moreover, the values of the parameters could also be situation-dependent. For instance, if the density of vehicles on a road is high, a safety message will mostly be relevant to vehicles in the near vicinity. Thus, the required dissemination area may be small. All these make it difficult to specify the values of network performance parameters.

III. Framework of Congestion Control in VANETs

From the discussion in the previous section, we may group congestion control into three major categories, i.e. (A) packet transmission rate control, (B) packet transmission duration control and (C) the transmit power control. With this categorization, we may explore more possibilities to cope with congestion and find commonalities among different approaches, thus apply existing experience and knowhow. We have to bear in mind that congestion control is a cross-layer issue. Its functions reside at different layers and there are also interactions between different layers. To put the mechanisms into the communication protocol stack, we propose architecture illustrated in Fig. 4. based on the communication architecture proposed by the C2C-CC. Different from the TCP/IP protocol stack, the transport layer has only very trivial functions for safety messages, thus it is omitted in the architecture. The Information Connector in

the architecture provides an interface between layers for cross-layer signalling. In the following, we briefly discuss different mechanisms in these three categories and highlight some research areas. It is not intend to give an exhaustive list of all possible mechanisms. The number of each mechanism is corresponds the same number in the Fig. 4.

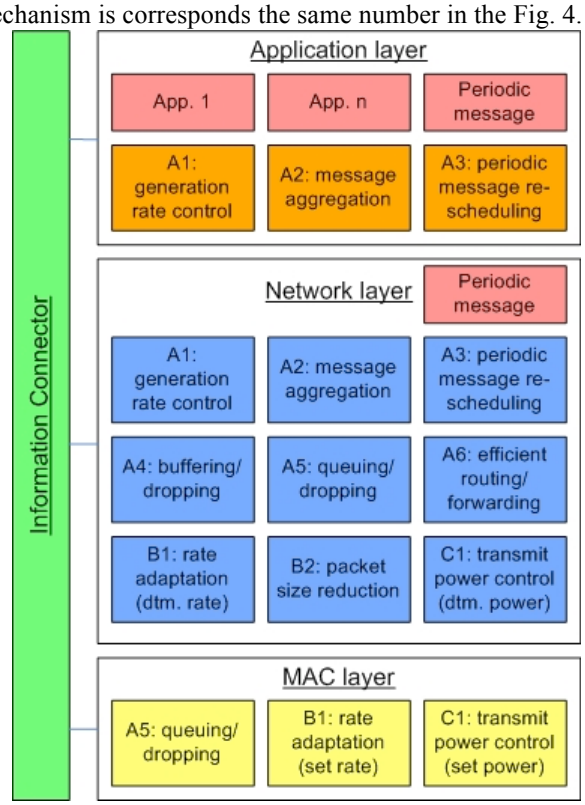

Fig. 4. Congestion control architecture

Packet transmission rate control refers to controlling the rate of packets injected into the network either from a source node or from a forwarding node. There are a number of approaches to realize it. The most straightforward one is to control packet generation rates from source nodes (A1). Message aggregation is a special feature in VANETs (A2). The basic idea is to aggregate the content from multiple messages into a single packet in order to reduce the number of packets. One example is to aggregate received messages with local information at the application layer. This is also called information-centric forwarding, where the forwarding decision is made based on application semantics [2]. Closely related with packet aggregation is the rescheduling of periodic messages (A3). If a periodic message is aggregated with an event-driven message, the next periodic message could be rescheduled at a time later than the original scheduled time. Packet transmission rate control could also be applied to buffering and queuing (A4, A5), where buffering refers to protocol specific mechanisms in VANETs such as buffering packet when there is no next hop. Efficient dropping strategies are needed for buffering and queuing in order to reduce the number of packets to be transmitted. In addition, efficient routing and packet forwarding algorithms (A6) will also reduce the transmission rate, and they are usually part of a routing or packet forwarding protocol. One example is to control the number of packet retransmissions for geocast in order to reach the required reliability with minimum overhead [11].

Packet transmission duration control means controlling the transmission time of a packet on the air. If multiple data rates are supported, increasing the data rate may reduce the transmission time (B1). This is especially useful for large packets because the synchronization time at the physical layer remain unchanged for different data rates. However, a higher data rate usually requires a higher transmit power from a sender, thus may cause a higher interference to other nodes. There exists trade-off between reduced transmission time and increased interference caused to other nodes. Another possibility to reduce the packet transmission time is to reduce the packet size (B2). Although most safety messages have small payload, the overhead introduced by secure routing could be much larger than the message itself [12]. Therefore, reducing such overhead may considerably reduce the data size. Other possible approaches to reduce data size include using efficient coding methods for geographical positions of vehicles [13].

Transmit power control (C1) for unicast is used to adopt the minimum power needed to transmit a packet from a transmitter to a receiver, which is based on the estimation of the radio propagation loss between the transmitter and receiver. The power control function at the network layer uses a certain algorithm to determine the transmit power of a packet, and the power is then set for the packet at the MAC layer before passed to the physical layer. In comparison with using the full power, it reduces harmful interferences caused to other nodes and allows more simultaneous transmissions [14]. Unlike unicast, broadcast is addressed to a group of receivers. Simulations have shown that the reception probability of event-driven messages can be ensured by controlling the transmit power of periodic messages [9].

As mentioned in Section II A, the traditional end-to-end TCP-like congestion control is not suitable, and new mechanisms are needed. We observe that some key areas on the congestion control for safety message are worthy of extensive research.

First, well defined network performance parameters are essential for designing congestion control algorithms, in that an algorithm should not compromise the required network performance. As mentioned in Section II C, the values related to the network performance such as the dissemination range, reliability and delay in VANETs need to be specified for each type of safety application. To meet these requirements, support from safety application designers is needed.

Second, considering the current state-of-the-art technology of IEEE 802.11, there is still no standard mechanism to obtain the channel load information from the wireless interface. Some parameters may be easily obtained, such as the number of neighbouring nodes, the received data rate, etc. But they do not represent the real channel load. With such constraint, congestion control algorithms have to

find appropriate means to monitor channel load and apply appropriate congestion control algorithms accordingly.

Third, more work is needed to understand pros and cons of each individual congestion control mechanism, especially considering the safety application requirements. For example, little research work has been reported so far on packet generation rate control. Although rate control itself has been extensively studies, how to do it in a distributed way in order to meet the requirements of safety applications is still an open issue. Another exam is that little research work on transmit power control for geocast has been reported so far.

Last but not least, methods are needed to apply the most appropriate mechanisms for congestion control. This is an advanced approach which requires the aforementioned work as a basis. Each mechanism has certain limitation and its usage may affect safety applications from a certain different aspect. For example, applications may require a minimum packet generation rate, thus will set a hard limit on rate generation control. In this case, transmit power may be applied. Furthermore, there are also interactions between different congestion control mechanisms. For instance, if several packets having different requirements on transmit power are aggregated into a single packet, it is still an open question how to set the power for the aggregated packet.

IV. Conclusion and Future Work

In this paper, we explore related concepts of congestion and congestion control of safety messages in VANETs. Considering the special features in VANETs, we suggest that each node locally applies both the reactive and proactive approach in a distributed way. We propose to use channel busy time as metrics for network load and define three parameters for the network performance of safety messages. Congestion control algorithms should reduce the channel busy time in order to meet the requirements of the network performance. We present a framework for congestion control by classifying different mechanisms and reveal areas of future research which include to define network performance parameters, to cope with state-of-the-art technology hardware limit, to understand pros and cons of each individual congestion control mechanism and to choose the most appropriate congestion control mechanisms.

Acknowledgment

This work is partly funded by the EC ICT Collaborative Project GeoNet under FP7 with the Grant Agreement Number 216269.

References

[1] Car-to-Car Communication Consortium, “CAR 2 CAR Communication Consortium Manifesto,” Version 1.1, Aug. 2007, available at http://www.car-to-car.org/.

[2] Vehicle Safety Communications Project, “Final Report,” Apr. 2006, available at http://www-nrd.nhtsa.dot.gov/.

[3] A. Festag, H. Füßler, H. Hartenstein, A. Sarma, and R. Schmitz, “FleetNet: Bringing Car-to-Car Communication into the Real World,” Proc. 10th ITS World Congress and Exhibition, Nagoyoa, Japan, Nov. 2004.

[4] D. L. Spohn, “Data Network Design,” Second edition, McGraw-Hill, 1997.

[5] A. S. Tanenbaum, “Computer Networks,” Third edition, Prentice-Hall, 1996.

[6] J. Widmer, R. Denda, and M. Mauve, “A Survey on TCP-Friendly Congestion Control,” IEEE Network, vol. 15, no. 3, May 2001, pp. 28-37.

[7] C. Lochert, B. Scheuermann, M. Mauve, “A Survey on Congestion Control for Mobile Ad Hoc Networks,” Wireless Communications and Mobile Computing, vol. 7, issue 5, Apr. 2007, pp. 655 – 676.

[8] P. Gupta, P.R. Kumar, “The Capacity of Wireless Networks,” IEEE Transactions on Information Theory, vol. 46, issue 2, Mar.2000, pp. 388-404.

[9] M. Torrent-Moreno, P. Santi, H. Hartenstein, “Distributed Fair Transmit Power Adjustment for Vehicular Ad Hoc Networks,” Proc. SECON, Reston, VA, USA, Sept. 2006.

[10] J. Gozdecki, A. Jajszczyk, R. Stankiewicz, “Quality of Service Terminology in IP Networks,” IEEE Communications Magazine, vol. 41, no. 3, Mar. 003, pp. 153-159.

[11] M. Torrent-Moreno, “Inter-Vehicle Communications: Assessing Information Dissemination under Safety Constraints,” Proc. WONS, Obergurgl, Austria, Jan. 007.

[12] C. Harsch, A. Festag, and P. Papadimitratos, “Secure Position-Based Routing for VANETs,” Proc. IEEE VTC Fall, Baltimore, MD, USA, Oct. 007.

[13] J. Haerri, F. Filali, C. Bonnet, “Rethinking the Overhead of Geo-localization Information for Vehicular Communications,” Proc. IEEE WiVeC'07, Sept.- Oct. 2007, Baltimore, MD, USA.

[14] A. Festag, R. Baldessari, H. Wang. “On Power-Aware Greedy Forwarding in Highway Scenarios,” Proc. WIT, Hamburg, Germany, Mar. 2007, pp. 31-36