



Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Overview paper

Supporting real-time traffic in multihop vehicle-to-infrastructure networks

Gökhan Korkmaz*, Eylem Ekici, Füsün Özgüner

Department of Electrical and Computer Engineering, The Ohio State University, 2015 Neil Avenue, 205 Drees Lab., Columbus, OH 43210-1272, USA

ARTICLE INFO

Article history:

Received 31 March 2008

Received in revised form 29 January 2009

Accepted 4 May 2009

Available online xxxx

Keywords:

Wireless networks

Vehicular communication

Real-time

Quality of Service

Mobile Internet

IEEE 802.11

ABSTRACT

In this paper, a new Controlled Vehicular Internet Access protocol with QoS support (CVIA-QoS) is introduced. CVIA-QoS employs fixed gateways along the road which perform periodic admission control and scheduling decisions for the packet traffic in their service area. The CVIA-QoS protocol is based on Controlled Vehicular Internet Access (CVIA) protocol that was designed only for the best-effort traffic. The most important contribution of the CVIA-QoS protocol is providing delay bounded throughput guarantees for soft real-time traffic, which is an important challenge especially for a mobile multihop network. After the demands of the soft real-time traffic is met, CVIA-QoS supports the best-effort traffic with the remaining bandwidth. Simulation results confirm that in CVIA-QoS protocol, the real-time throughput is not affected from the best-effort load and its delay is much smaller than CVIA delay when both the real-time and best-effort load exist in the channel. It has been observed that, unlike, CVIA-QoS, IEEE 802.11e with multi-hopping suffers from lower throughput and high number of real-time packet drops.

© 2009 Published by Elsevier Ltd.

1. Introduction

Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication systems have the potential for enabling new safety applications, and extending existing ones to vehicles and their passengers. In October 1999, Federal Communications Commission (FCC) allocated Dedicated Short-Range Communication (DSRC) spectrum at 5.9 GHz to “increase traveler safety, reduce fuel consumption and pollution, and continue to advance the nation’s economy (FCC, 1999)”. In 2004, a new FCC ruling (FCC, 2004) defined six service and one control channels for DSRC. The US Department of Transportation (USDOT) sponsored standard process under ASTM that voted to base DSRC on IEEE 802.11a (Lee Armstrong, 2008). However, the suitability of IEEE 802.11a for vehicular networks is still an important research question. The results of the simulations using IEEE 802.11a protocol show that it is worth doing further research on communication designs that might utilize the channel more efficiently even for single hop broadcast (Xu et al., 2004). A survey presenting and discussing different wireless communication standards as well as proprietary solutions that are intended especially for a high-speed network can be found in Bilstrup (2007). Finally, according to experiments in Pinart and Barona (2008), using only 3G standards based solutions show poor QoS performance for vehicular communications.

Although most of the research in DSRC targets only safety applications, the convergence of Internet and vehicular networks enables important applications such as web browsing, voice and video streaming, remote vehicle diagnostics, mobile office, gaming, and real-time navigation information. These applications can attract great demand and have important impact on the market penetration rate of V2V and V2I systems. As a result, the safety applications will also benefit from this high market penetration rate since they are shown to be effective only above a certain penetration rate threshold (Tsugawa

* Corresponding author. Tel.: +1 408 221 8893.

E-mail addresses: korkmazg@ece.osu.edu (G. Korkmaz), ekici@ece.osu.edu (E. Ekici), ozguner@ece.osu.edu (F. Özgüner).

and Kato, 2003). However, it is shown that non-safety use of DSRC may have to be severely restricted during peak hours traffic to insure that automotive safety is not compromised (Wang and Hassan, 2008).

In recent years, several research efforts target V2V and V2I systems (Hinsberger et al., 2007). Among these, the FleetNet project (FLEETNET, 2008) and its follow-up project Network on Wheels (NoW) Festag et al. (2008) investigate the integration of the Internet and vehicular networks. This integration requires mobility support, efficient communication, discovery of services, and support of legacy applications. FleetNet uses an IPv6 based addressing solution to address the vehicles. The proposed architecture contains stationary Internet gateways (IGW) along the road with two interfaces connecting vehicular networks to the Internet (Bechler et al., 2003). Vehicles communicate with distant IGWs via multi-hopping. This architecture is useful not only to connect vehicles to other networks but also to connect isolated vehicle groups to each other (Tchouto et al., 2003; Baccelli et al., 2007). In Tsukada et al. (2008), a routing experiment is setup with simultaneously active multiple egress interfaces for each vehicle where a packet can be routed through internet or mobile ad-hoc network to improve network delay and bandwidth. When there are multiple APs (access points) exist in a network, hand-off is another study topic where specialized IEEE 802.11 hand-off algorithms are needed for the vehicular network (Giannoulis et al., 2008). However, in these proposals, there are no specific solutions describing how to solve the data movement problem over multiple hops. In Korkmaz et al. (2006), we introduced a fair and high-throughput internet access protocol, CVIA, for multihop vehicular networks. Although the CVIA protocol provides fairness and high throughput for vehicular Internet access, it is not suitable for applications like video and voice streaming which require throughput and delay guarantees.

In this paper, a new Controlled Vehicular Internet Access protocol with QoS support (CVIA-QoS) is proposed. The CVIA-QoS protocol is a cross-layer solution for vehicular multihop networks spanning MAC and routing functions. CVIA-QoS protocol uses admission control for soft real-time traffic to provide delay bounded throughput guarantees. To achieve this goal, fast and slow packet propagation methods are defined for real-time and best-effort traffic, respectively. The real-time sessions send registration requests to the gateways that perform scheduling and admission control decisions periodically. After the demands of the soft real-time traffic is met, the remaining bandwidth is allocated to the best-effort traffic. CVIA-QoS protocol provides much lower and bounded delay to soft real-time traffic when compared with original CVIA protocol. Furthermore, once admitted by a gateway, the throughput level of the real-time sessions is not affected by the increase in the best-effort traffic.

2. Assumptions and internet access model

We consider a vehicular network which accesses the Internet through fixed Internet gateways along the road. These gateways have two interfaces: a wireless interface for the vehicular network and another interface connected to the Internet. Although the wireless interface of these gateways has a limited wireless coverage, their range can be increased with multi-hop communication. As a result, a gateway can communicate with a vehicle at a distance several times longer than its physical transmission range. The range of a gateway where it provides Internet access service is called the *virtual transmission radius*. We assume that gateways send periodic service announcements to indicate the availability of the service in their service area. We also assume that the uplink and the downlink packets are transmitted over two frequency separated channels.

Vehicles are assumed to be equipped with GPS devices used for time synchronization and obtaining vehicle positions. Vehicle positions obtained via GPS are exchanged among one-hop neighbors. When a vehicle enters the virtual transmission radius of a gateway, it registers itself with the gateway. In this work, each vehicle is assumed to have one access point to the wireless medium. All devices inside the vehicle (e.g., laptops, PDAs, vehicle's onboard computer, sensors, etc.) use this interface to communicate with the outside world, creating an aggregated traffic going in and out of the vehicle. The devices inside the vehicle establish a mobile network. We assume that mobility management is handled by the proposed network mobility solutions in the literature Perera et al. (2004).

3. IEEE 802.11e

In this section, we outline IEEE 802.11e since in vehicular network, Enhanced Distributed Coordination Function (EDCF) of IEEE 802.11e is the first candidate for Internet access with QoS. IEEE 802.11e standard is the extension of the legacy IEEE 802.11 standard for Quality of Service enhancements in LAN applications. Unlike IEEE 802.11b, IEEE 802.11a, and IEEE 802.11g, the main enhancement in IEEE 802.11e is proposed for the Media Access Control (MAC) layer.

In IEEE 802.11e, a station with high priority traffic waits less before it sends its packet that gives the high priority traffic a higher chance of being sent than low priority traffic. To wait less, interframe space and maximum contention window are chosen shorter for the high priority traffic in IEEE 802.11e.

3.1. Interframe space

In legacy IEEE 802.11, all nodes wait at least DIFS amount of time before sending a new packet. The channel must be idle for DIFS time before transmission. IEEE 802.11e defines different interframe spaces for different traffic classes which are called Arbitration Interframe Spaces (AIFS):

Table 1
Parameters of IEEE 802.11e.

Access category	CW_{min}	CW_{max}	$AIFSN_{AC}$
0	15	1023	7
1	15	1023	3
2	7	15	2
3	3	7	2

$$AIFS = SIFS + AIFSN_{AC} \cdot aSlotTime, \quad (1)$$

where $SIFS$ and $aSlotTime$ are PHY parameters of IEEE 802.11e; and $AIFSN_{AC}$ depends on the access category (AC) as shown in Table 1. Note that when $AIFSN_{AC} = 2$, AIFS is equal to DIFS. Since the traffic class with the shorter AIFS accesses the channel sooner, shorter AIFS gives higher priority.

3.2. Contention window (CW)

As part of collision avoidance mechanism, if a station detects a busy channel before starting a transmission, it performs a backoff procedure. According to this procedure, the station has to keep sensing the channel for an additional random time after detecting the channel being idle for DIFS. This random time is in terms of slot time and it is chosen uniformly from a Contention Window ($CW = [0, CW_{min}]$). If no ACK is received for the packet, CW is doubled until it reaches CW_{max} . CW_{min} and CW_{max} parameters of different access categories can be found in Table 1. In IEEE 802.11e, traffic classes are assigned different CW s and each class runs its own backoff procedure. As a result, if two stations belonging to different traffic classes, start their backoff at the same time, the class with higher priority is more likely to access the channel before the low priority one because it has higher chance of selecting a lower backoff time.

3.3. Problems

Although IEEE 802.11e gives high priority traffic a higher chance of being sent, the protocol does not guarantee either throughput or delay. Best-effort packets access the channel in a random fashion, and even if their CW window is larger, a station with a best-effort packet can still access the channel before high priority packets. It means that the traffic class differentiation is not absolute.

IEEE 802.11e provides service differentiation by assigning a shorter CW window for high priority packets. When the number of nodes generating high priority traffic is low, this short CW is sufficient to avoid collisions. However, short CW causes a serious problem in multihop vehicular networks where high number of vehicles exist in the transmission range and many

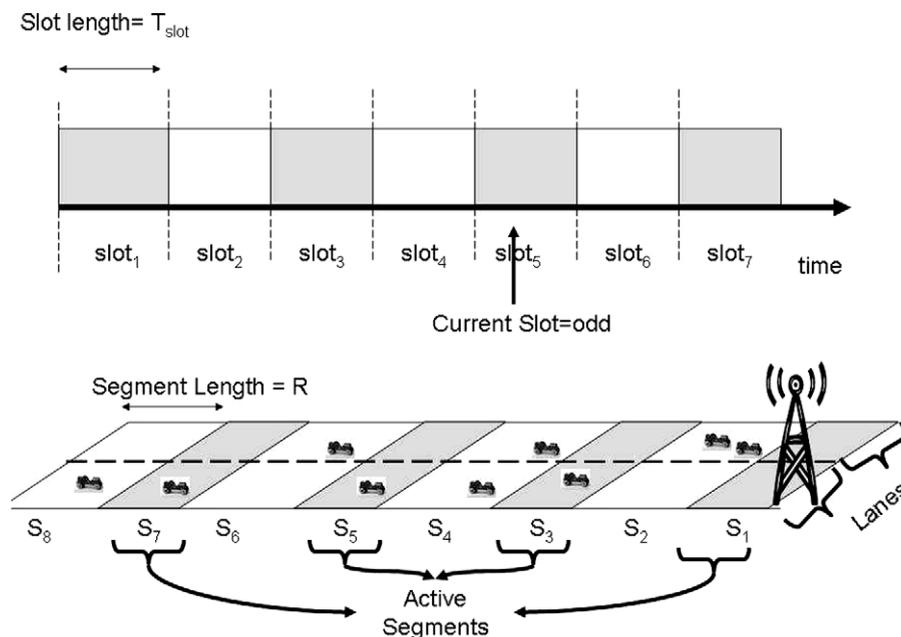


Fig. 1. Slots and segments.

vehicles are hidden to the communicating vehicle pairs. Consequently, there can be hundreds of vehicles which can interfere with the communication leading to a high number of collisions and low throughput for IEEE 802.11e.

4. Controlled Internet Access Protocol (Korkmaz et al., 2006)

In this section, we present the CVIA protocol on which our proposed CVIA-QoS protocol is based. CVIA is a cross-layer communication protocol for vehicular internet access along highways. The protocol functions span MAC and network layers. The objective of the protocol is *to increase the end-to-end throughput while achieving fairness in bandwidth usage between road segments* for the best-effort traffic. The CVIA protocol aims to solve two main problems of IEEE 802.11 protocol in multi-hopping along a highway: low throughput and starvation of packets originating from vehicles far away from gateways.

The CVIA protocol forms temporary single hop clusters to mitigate the hidden node problem as it is unlikely for a vehicle to be a *hidden node* for a transmission between two vehicles in a single hop vehicle cluster. To form these clusters, the time is divided into slots and the service areas of the gateways are divided into segments as shown in Fig. 1. In each time slot, the CVIA protocol allows transmissions only in the segments which are far away from each other, i.e., the transmission in one segment does not interfere with the transmissions in the other active segments. To allow transmissions occur in all segments, the active segments are switched periodically and a new set of segments becomes active in every time slot. Since the active segments do not interfere with each other, there is not a theoretically limit for the number of hops. However, increasing the number of hops increases the delay in the network.

The packet traffic is controlled by designated vehicles in segments called the temporary routers (*TR*) as shown in Fig. 2. In each time slot, the temporary router which is responsible for the outbound traffic receives packets originating from other segments and local packets from its own segment. At the end of the time slot, all packets are moved out to the next segment together without contention.

4.1. Definitions

- (i) *Communication range R*: The physical transmission range of the vehicles as well as the gateway.
- (ii) *Virtual transmission radius (VTR)*: The radius of the service area of the gateway where it provides Internet access service.
- (iii) *Segment $i(S_i)$* : Fixed section of VTR of length R . The segment closest to the gateway is denoted by S_1 .
- (iv) *VTR length N* : The number of segments in VTR of a gateway.
- (v) *Time slot $j(TS_j)$* : Time duration of length T_{slot} .
- (vi) *Neighboring segment S_{i+}* : The neighboring segment in the direction of packet dissemination. S_{i+} is the neighboring segment of S_i closer to the gateway in the uplink channel and the neighboring segment of S_i farther away from the gateway in the downlink channel.
- (vii) *Neighboring segment S_{i-}* : The neighboring segment in the opposite direction of the packet dissemination.
- (viii) *Interference parameter (r)*: $r = \left\lceil \frac{\text{interference range}}{R} \right\rceil + 1$.
- (ix) *Active segment*: Segments where vehicle communication is allowed to occur in a time slot. S_i is active in TS_j if $(i \bmod r)$ equals to $(j \bmod r)$. Note that there are $r - 1$ inactive segments between two active segments according to the active segment definition. In Fig. 1, when the current time slot is T_5 and $r = 2$, segments S_1, S_3, S_5 , and S_7 become active. In this example, when segment $i(S_i)$ is active, its two neighboring segments (S_{i-}, S_{i+}) become inactive. In the next time slot (TS_{j+1}), all segments change states where inactive segments become active and active segments become inactive.

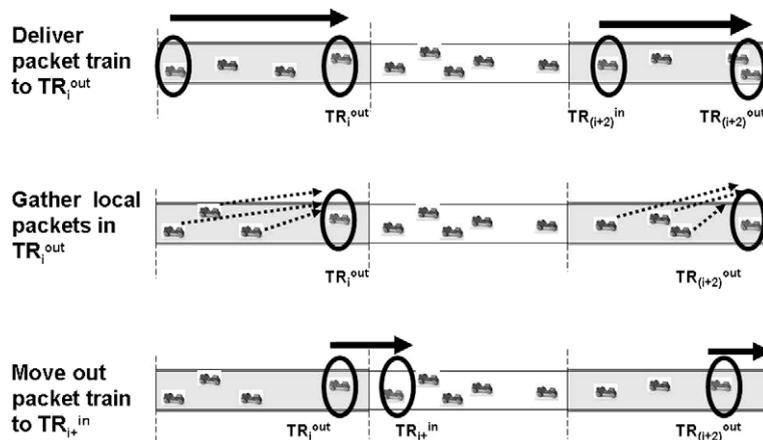


Fig. 2. Packet movement types.

- (x) *Outbound temporary router* (TR_i^{out}): In active segments, the packets are gathered in vehicles closest to the segment border in the direction of packet dissemination. The vehicle where the relayed packets are collected is called *outbound temporary router*.
- (xi) *Inbound temporary router* (TR_i^{in}): At the end of an active time slot, TR_{i-}^{out} moves the packets to S_i . In S_i , the vehicle which receives all packets from S_{i-} is called *inbound temporary router*.
- (xii) *Packet train*: It is a method where several packets are sent with only one RTS/CTS handshake among the same source and destination pair.

4.2. Packet movement types

CVIA employs two vehicles as temporary routers in each active segment i : TR_i^{out} and TR_i^{in} . All packets entering a segment go through TR_i^{in} and all packets leaving the segment go through TR_i^{out} . We choose TR_i^{in} as the closest vehicle to S_{i-} and TR_i^{out} as the closest vehicle to S_{i+} . The CVIA protocol uses three types of packet movements as shown in Fig. 2:

1. TR_i^{in} delivers the packet train originating from other segments to TR_i^{out} .
2. Local packets of the segment is gathered by TR_i^{out} .
3. TR_i^{out} creates a new train using packets received in (1) and (2), and sends it to TR_{i+}^{in} in S_{i+} .

4.3. Protocol phases

4.3.1. Inactive phase

Vehicles in inactive segments do not access the channel. Inactive and active segments are determined using the slot number i and segment number j , i.e. S_i is active when $(i \bmod r) = (j \bmod r)$. Vehicles compute i and j as follows: $i = \lfloor \frac{\Delta d}{R} \rfloor$, and $j = \lfloor \frac{\Delta t}{T_{slot}} \rfloor$, where Δd is the distance of the vehicle to the gateway and Δt is the time passed since an absolute reference point, e.g., 12:00 pm. Note that vehicles obtain their positions and synchronize their clocks using GPS. The vehicles learn the positions of gateways from a digital road map database or the service announcement packets broadcast periodically by gateways.

4.3.2. Temporary router selection phase

At the beginning of each time slot, next temporary routers are selected by TR^{in} s. Since the topology of the vehicular network changes fast, new temporary routers must be selected periodically. The selected routers are called $TR_{i,next}^{out}$ and $TR_{i,next}^{in}$ until they become active. $TR_{i,next}^{out}$ and $TR_{i,next}^{in}$ must stay inside the segment for an amount of time called the *router lifetime*; therefore, they should be away from the segment borders by a certain distance. The details of the temporary router selection phase is presented in Korkmaz et al. (2006).

4.3.3. Intra-segment packet train movement phase

TR_i^{in} starts delivering the packet train received from S_{i-} in the previous slots to TR_i^{out} . To avoid contention, TR_i^{in} has the highest access priority and waits only SIFS duration before accessing the channel.

4.3.4. Local packet gathering phase

After the intra-segment packet train movement ends, the channel becomes idle for DIFS duration and vehicles access the channel using a contention based channel access scheme. In this local packet gathering phase, vehicles employ the DCF method of the IEEE 802.11 standard. Each node has the same priority and they directly send their packets to TR_i^{out} . To decrease the number of collisions, each vehicle starts this packet gathering phase with a random backoff counter.

4.3.5. Inter-segment packet train movement phase

TR_i^{out} has the highest channel access priority among the vehicles in the LPG phase. When time left in the active slot is just enough to move out packets in the queues, TR_i^{out} accesses the channel and ends the LPG phase. The new packet train is moved to TR_{i+}^{in} before the end of T_{S_j} . Note that all the packets pass through S_1 . Therefore, each TR_i^{out} can send out at most $(C/N) \cdot (N - i + 1)$ packets where C is the capacity of the protocol. These packets are buffered in TR_{i+}^{in} until S_{i+} becomes active.

5. CVIA-QoS

In Korkmaz et al. (2006), it is shown through simulations that CVIA has up to 80% higher end-to-end throughput capacity when compared with the IEEE 802.11 protocol. In addition, the CVIA protocol distributes the throughput fairly among segments while the vehicles far away from the gateway suffer from starvation under the IEEE 802.11 protocol. Although CVIA is a successful communication protocol for the best-effort traffic, it is not suitable for real-time applications. In CVIA, the packets experience high delays that is directly proportional to the number of segments in the virtual transmission range as packets move only one segment in each time slot. In addition, due to lack of an admission control mechanism, the protocol cannot provide guaranteed throughput to real-time sessions.

Because of the shortcomings of the existing protocols, a new protocol, CVIA-QoS, is designed to provide throughput guarantees and fixed delay bound to soft real-time applications like voice and video streaming in linear vehicular networks. The

best-effort traffic is handled with the remaining bandwidth after allocating resources for real-time traffic. In CVIA-QoS, one time slot is divided into two periods, namely high priority period (HPP) and low priority period (LPP). Unlike the CVIA protocol, packets admitted to HPP is delivered to the gateway in *one time slot*. Furthermore, an admission control mechanism is introduced where admission decisions are made by the gateways and executed by the temporary routers.

5.1. Phases in the CVIA-QoS protocol

In CVIA-QoS, time is divided into slots of T_{slot} . Each time slot is composed of high priority and low priority periods as shown in Fig. 3. At the beginning of the HPP, sessions that request service in HPP send registration packets to outbound temporary routers. Once these requests are collected in all segments, they are delivered to the gateway. To grant end-to-end throughput guarantees to sessions, the gateway uses the information about the new session requests, buffer status of the temporary routers and the information about already active sessions. The session admission decisions and time allocations for different phases in the time slot are sent to temporary routers. Temporary routers poll the vehicles that are granted access. After the polling phase, the collected packets from all segments are propagated to the gateway in packet trains in one time slot. Since the maximum length of LPP is T_{slot} , the network delay of the packets admitted to HPP are bounded by T_{slot} . In the remaining time, the original CVIA protocol is used to serve the best-effort traffic in LPP.

In this section, the phases inside the HPP and LPP periods are explained in detail. Moreover, computing the length of the HPP phase is discussed in Section 5.2.

5.1.1. High priority period

5.1.1.1. Registration phase. Vehicles with new real-time sessions and vehicles with active sessions that enter a new segment send short registration packets to TR^{out} s using IEEE 802.11 protocol without RTS/CTS handshakes. Inside the registration packets, vehicles send:

1. Source ID.
2. Session ID.
3. Session Lifetime (SLT).
4. Minimum acceptable throughput.
5. Maximum throughput.

Gateways use the *minimum acceptable throughput* to grant admission to the session. However, if there is more BW available in a slot, more packets from the session up to *maximum throughput* can also be scheduled. Session Lifetime is the expected active time of a session. The gateway may grant a *session time* less than SLT. However, sessions can piggyback new requests in their data packets if they need more time in a contention free manner.

Vehicles start the phase with preset random backoff counters to decrease the probability of packet collisions at the beginning. If a registration packet is received successfully by TR^{out} , an ACK packet is sent to the vehicle. To mitigate the hidden node problem, fixed length sub-phases are defined in the registration phase. In each sub-phase, a new set of segments become active to send registration packets. Segment i becomes active in sub-phase k if $i \bmod r = k \bmod r$. There are $r - 1 = \lfloor \frac{\text{interferencerange}}{R} \rfloor$ inactive segments between two active segments. As a result, after r sub-phases, all segments will have become active once. Registration sub-phases are depicted in Fig. 4 for $r = 2$. In this figure, registration packets are sent in parallel in active segments and registration phase is composed of two sub-phases.

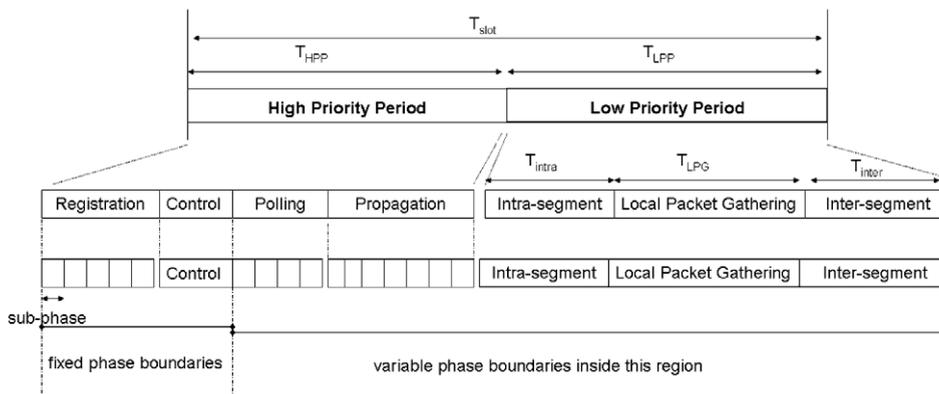


Fig. 3. Phases in the CVIA-QoS protocol.

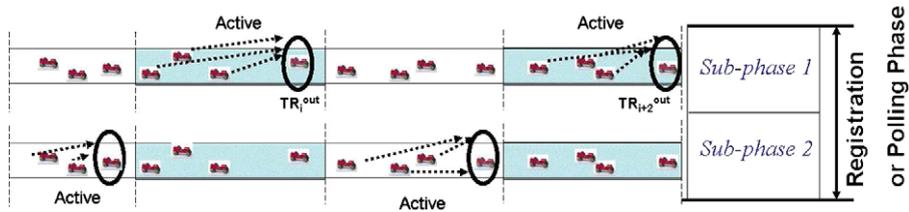


Fig. 4. Sub-phases of registration and polling phases. Note that activity rules of the segments are same for both registration and polling phases.

5.1.2. Control phase

After the Registration Phase, a control packet train is initiated by the furthest TR^{out} . Control trains are forwarded by only TR^{out} s and TR^{in} s during their propagation as shown in Fig. 5. Each TR^{out} includes the new high priority session requests and each TR^{in} includes the number of low priority packets in its buffer. Based on this new information and the information about the existing active sessions, the gateway makes scheduling and admission control decisions as described in Section 5.3. In the second control train, the schedules are sent from the gateway to all TRs. In this control train originating from the gateway, the lengths of all phases and sub-phases until the end of the current time slot are announced. The fixed length and variable length phases are shown in Fig. 3.

5.1.3. Polling phase

Based on the schedule received from the gateway, TR^{out} s poll the vehicles with active sessions. In response to polling, vehicles send their data packets. Acknowledgments to data packets are piggy-backed in the polling packets by TR^{out} s. Similar to the Registration Phase, r polling sub-phases are employed to mitigate the hidden node problem in the polling phase. The length of the polling phase depends admitted real-time packets, the number of segments in the virtual transmission range, and r . Polling sub-phases are shown in Fig. 4 for $r = 2$ which is the same figure used to depict the registration sub-phases. Since the number of packets scheduled in each segment can be different, the active segment which needs the longest time (highest number of packets) determines the length of the current sub-phase. As a result, sub-phase lengths are not fixed in polling phase unlike registration phase.

5.1.4. Propagation phase

In the propagation phase, high priority packets collected by TR^{out} s are moved towards the gateway in fast packet trains. The packet trains are initiated in parallel by segments $\{S_i, S_{i+r}, S_{i+2r}, \dots, S_k\}$, where $1 \leq i \leq r$ and $k \leq N$. This phase is also composed of several sub-phases. Segment i becomes active in sub-phase m if $i \bmod r = m \bmod r$.

When S_i becomes active, the fast train originating from S_{i-} is moved to TR_i^{out} . Then, all local packets collected in the segment are attached to the packet train. At the end of each sub-phase, TR_i^{out} sends the packet train to TR_{i+} in the next segment. With this approach, all polled packets are delivered to the gateway in one time slot. As a result, network delay experienced by real-time packets admitted to HPP is bounded by T_{slot} . As in polling phase, the length of the propagation phase depends on admitted real-time traffic, the number of segments in the virtual transmission range and r .

Fig. 6 depicts the packet movement in propagation sub-phases. After the first sub-phase, packets from two segments merge in one train and start moving together in parallel with other high priority trains. Since it takes N steps for the packets from the furthest away segment to reach to the gateway, there are N sub-phases in each propagation phase.

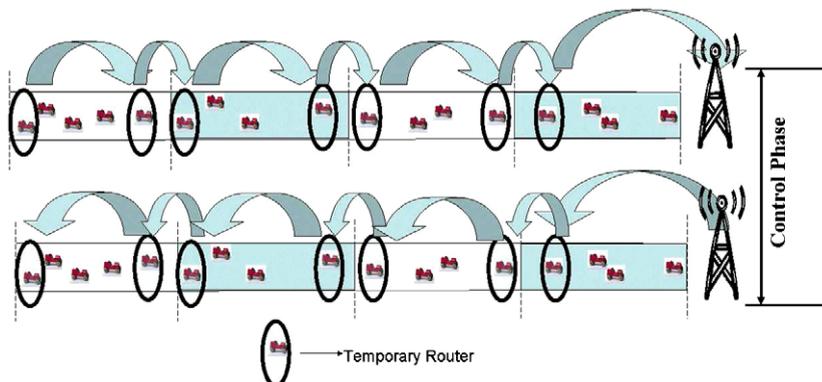


Fig. 5. Control phase: control packet visits all the temporary routers. Encircled vehicles are temporary routers.

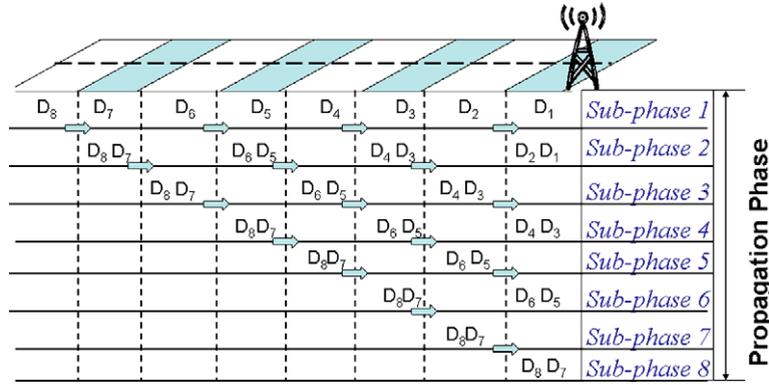


Fig. 6. Data movement in propagation sub-phases where D_i represents packets originating from segment i .

A further optional improvement in propagation phase is using block-ACKs where instead of acknowledging each packet, one packet from the destination can acknowledge multiple data packets. Note that block-ACK mechanism is already proposed for IEEE 802.11e and 802.11n; However, this mechanism is helpful when there are a lot of packets in one source. Since CVIA and CVIA-QoS protocols gather packets to one point and then move them together in packet trains, they are more suitable and can get more benefits from block-ACKs.

5.1.5. Low priority period

The phases in this period are identical to the packet movement phases of the original CVIA protocol presented in Section 4.3, namely intra-segment packet train movement phase, local packet gathering phase, inter-segment packet train movement phase. Note that the packet train utilizing the low priority time period is referred to *slow packet train* since the packets can propagate one segment in each time slot (T_{slot}) and the packets are buffered in TR^{in} s.

5.2. Length of the high priority period

The number of high priority packets that can be served in one time slot is directly proportional to the length of the high priority period. However, longer HPP results in longer delays. Thus, the maximum length of HPP is an important design parameter affecting both the throughput and delay.

In a system where HPP can be extended to cover the full slot, starvation of best-effort packets becomes a problem. To avoid starvation, a maximum ratio of the length of the HPP phase to T_{slot} is defined as $ratio_{HPPmax}$. As a result, once the maximum HPP length ($T_{HPP,max}$) is decided, the slot length can be chosen as

$$T_{slot} = T_{HPP,max} \cdot \frac{1}{ratio_{HPP,max}} \tag{2}$$

The length of HPP is equal to the sum of the lengths of four phases: $T_{HPP} = T_{register} + T_{control} + T_{polling} + T_{propagation}$, where $T_{register}$, $T_{control}$, $T_{polling}$, and $T_{propagation}$ are the lengths of the registration, control, polling, and propagation phases, respectively.

5.2.1. $T_{register}$

$T_{register} = r \cdot t_{register}$, where $t_{register}$ is the length of registration sub-phase.

In the registration phase, vehicles access the channel using IEEE 802.11 protocol without RTS/CTS handshake. Since the neighboring segments are inactive during this phase, we assume that hidden terminals do not exist. To compute the saturation throughput of IEEE 802.11 protocol, Bianchi presents a Markov model where back-off window of each station is modeled by a bidimensional process (Bianchi, 2000). Using the Bianchi’s analysis, when the number of vehicles in each segment is 20 and the minimum contention window size is 64, almost 78% of the registration phase can be used for successful packet transmission. Note that this is a worst case estimate for our system since it is highly unlikely that all vehicles have a new session initiation request at the same time. In addition, Bianchi’s estimate assumes that each vehicle has another packet to send immediately after it sends the current packet. However, in the registration phase, each vehicle has only one packet to send.

Fig. 7 show the number of new session initiations in a second that can be received on the average from each segment. Using this figure, when we allocate $t_{register} = 2.5$ ms, approximately 69 new session initiations can be received every second.

5.2.2. $T_{control}$

$T_{control} = (T_{cp} \cdot 2N) \cdot 2$, where T_{cp} is the length of the control packet. This control packet visits both TRs in each segment; therefore, it is multiplied by $2N$. The resulting time should be doubled since in a second control train, the schedules are sent

from the gateway to all TRs. For $N = 4$, 1 ms is sufficient for the control phase to deliver all information to and then from the gateway.

5.2.3. $T_{polling}$

$T_{polling} = \sum_{i=1}^r t_{polling}^i$, where $t_{polling}^i$ is the length of the polling sub-phase i . Similar to registration operation, polling is also done in parallel. The lengths of the registration sub-phases are fixed and equal. On the other, the lengths of the polling sub-phases depend on the number of packets to be polled in segments, which are scheduled by the gateway. As a result,

$$t_{polling}^i = (SIFS + T_{poll} + SIFS + T_{DATA}) \cdot N_{max}^i, \quad (3)$$

$$N_{max}^i = \max(N_m^i), \quad m \in \text{active segments in sub-phase } i, \quad (4)$$

where T_{poll} and T_{DATA} are the length of the polling, and data packets, respectively. N_r^i is the number of scheduled packets in segment r and N_{max}^i is the maximum number of packets that is scheduled from a segment among all active ones. Note that the exact length of $T_{polling}$ depends on the distribution of scheduled packets among segments.

5.2.4. $T_{propagation}$

$T_{propagation} = \sum_{i=1}^N t_{propagation}^i$, where $t_{propagation}^i$ is the length of the polling sub-phase i . Similar to polling phase, the lengths of the propagation sub-phases depend on the number of packets in segments that are changing as the packets propagate. The important feature of this phase is that packets move in packet trains.

$$t_{propagation}^i = T_{to} + (SIFS + T_{DATA} + SIFS + T_{ACK}) \cdot Npro_{max}^i, \quad (5)$$

$$T_{to} = DIFS + T_{RTS} + SIFS + T_{CTS}, \quad (6)$$

$$Npro_{max}^i = \max(Npro_m^i), \quad m \in \text{active segments in sub-phase } i, \quad (7)$$

where T_{RTS} and T_{CTS} are the lengths of RTS and CTS packets, respectively. $Npro_m^i$ is the number of propagation packets in segment m and $Npro_{max}^i$ is the maximum number of packets that is in a segment among all active ones.

5.3. Scheduling and admission control

The gateways are responsible for accepting session initiation requests and determining the lengths of the phases and sub-phases in a time slot. As shown in Fig. 3, the length of the registration and control sub-phases are fixed in each frame. Therefore, timing of polling and propagation sub-phases are scheduled relative to the end of the control phase. Gateways use the HPP phase length calculations presented in Section 5.2 for admission control and determining the lengths of polling and propagation sub-phases. The sessions are not admitted if the following first condition does not hold:

$$\text{Condition I : } T_{HPP}^{new} < T_{slot} \cdot \text{ratio}_{HPPmax} = T_{HPP,max}, \quad (8)$$

where T_{HPP}^{new} is the length of HPP after scheduling the new session.

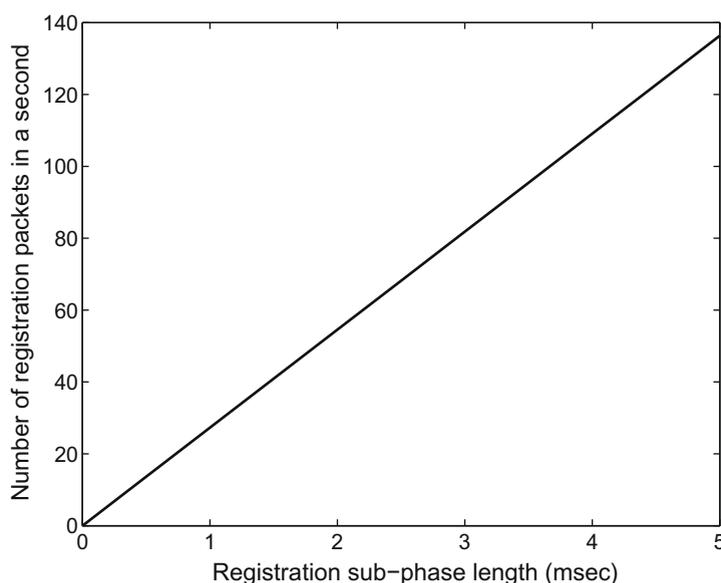


Fig. 7. Number of session registration packets in a second, $T_{slot} = 100$ ms.

In addition to maximum length bound on HPP, the CVIA protocol introduces a bound on the maximum length of LPP. The maximum ratio of the LPP to slot time is called $Ratio_{max}$. This bound also implies a bound on the minimum length of HPP which is $1 - Ratio_{max}$. Increasing the length of the LPP increases the number of new low priority packets accepted to the network. These packets are buffered in TR^{in} s; Therefore, the longer LPP, the higher the number of packets buffered in the network. Since the temporary routers are mobile and they change segments, packets cannot be buffered in temporary routers longer than the routers' lifetime beyond which they may leave their segments. Thus, in the CVIA-QoS, packets buffered in the network are scheduled before accepting new real-time session requests. The buffered packets in the network decrease the response time of the protocol when a bursty real-time demand is encountered.

The response time is the price paid for not dropping any packets. As a result, low priority packets stored in TR buffers impose a second condition, which can be more strict than condition in Eq. (8) for admission control:

$$\text{Condition II : } T_{HPP}^{new} < T_{slot} - T_{LPP,buffered}, \quad (9)$$

where $T_{LPP,buffered}$ is the time it takes to schedule buffered low priority packets in LPP.

As a result, the scheduling algorithm has four main steps:

1. *Schedule already accepted high priority traffic:* Packets of already accepted sessions are scheduled to HPP. With scheduling already accepted traffic before accepting new requests, CVIA-QoS protocol satisfies the minimum throughput guarantees given to real-time sessions in the previous slots.
2. *Schedule low priority packets in the buffers of TR^{in} s:* since the CVIA-QoS protocol does not drop any packets once admitted to the network, the gateways allocate time for the buffered best-effort packets before the temporary routers leave the segment. Buffered low priority packets that are admitted in the previous slots are scheduled to LPP. To achieve this schedule, a minimum time is allocated in intra-segment (T_{intra}) and inter-segment (T_{inter}) packet movement phases. In step 5, if there is still BW available for new best-effort packets, T_{inter} phase is extended.
3. *Schedule packets from new sessions:* After scheduling the packets of the existing sessions and buffered low priority packets, the gateway attempts to schedule new sessions if the following condition holds:

$$T_{HPP}^{new} < \min(T_{slot} - T_{LPP,buffered}, T_{slot} \cdot ratio_{HPPmax}). \quad (10)$$

If the $T_{HPP}/T_{slot} < 1 - ratio_{max}$ even after scheduling the new real-time packets, some of the best-effort packets already scheduled to LPP are rescheduled to HPP until HPP reaches its minimum length= $T_{slot} \cdot (1 - Ratio_{max})$.

4. *Schedule new best-effort packets :* If there is still some remaining time in the current time slot, it is dedicated to new best-effort packets. To gather these new packets, T_{LPG} amount of time is allocated to Local Packet Gathering Phase in LPP. In addition, enough time to move those packets to the next segment is added to T_{inter} whose details can be found in Korkmaz et al. (2006). Note that if the real-time traffic is very high, no new best-effort packet is admitted to the network by setting $T_{LPG} = 0$ if the real-time demand exceeds the capacity of the time slot.

6. Performance evaluation

The performance of CVIA-QoS is assessed through simulations. In the simulation scenarios, simulated vehicles move on a linear highway segment with two lanes, one for each direction of traffic flow. Vehicles randomly enter the service area with exponentially distributed inter-arrival times. On the average, there are 34 vehicles/km per lane. Each vehicle is assigned a speed from a Gaussian distribution with a mean of 90 km/h and standard deviation 5 km/h at the beginning of the simulation. The assigned speeds do not change during simulation. The common parameters of the simulations are: transmission range = 350 m, data rate = 27 Mbps, payload = 2000 or 500 bytes, base protocol = 802.11a (DSRC), Maximum number of packet retries = 10, interference range to transmission range ratio = 1. The protocol parameters are: $T_{slot} = 100$ ms, $N = 4$ or 8, and $r = 2$. In IEEE 802.11e, $AC = 0$ and $AC = 4$ are used for the best-effort and real-time packets, respectively. Other parameters of the MAC layer and the physical layer are taken from the ASTM E2213-02 standard document (DSRC, 2002).

To examine the effects of packet length and virtual transmission radius on the performance of the network, four different scenarios are simulated and analyzed for various packet loads. The parameters for these scenarios are as follows: Scenario I: payload = 2000 bytes, $VTR = 4R$ ($N = 4$), Scenario II: payload = 2000 bytes, $VTR = 8R$ ($N = 4$), Scenario III: payload = 500 bytes, $VTR = 4R$ ($N = 4$) and Scenario IV: payload = 500 bytes, $VTR = 8R$ ($N = 8$).

6.1. Behavior of CVIA-QoS

In this section, the behavior of the CVIA-QoS protocol is examined for its HPP phase length, transient response delay, and throughput capacity. In Section 6.2, the performance of the CVIA-QoS protocol is compared with CVIA and IEEE 802.11e protocols in terms of throughput and delay.

6.1.1. Length of HPP

Fig. 8 shows the length of the high priority period when the real-time packet demand is increased in the system. The generated real-time packets are distributed randomly to each vehicle and than the curves are obtained using the HPP length

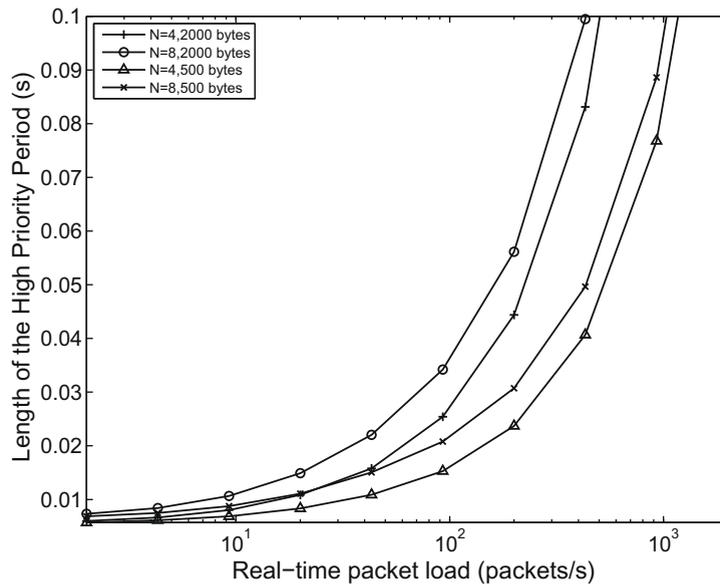


Fig. 8. Length of HPP.

calculation given in Section 5.2. HPP length is an important parameter which determines the length of T_{slot} . If a shorter slot length is desired to decrease delay, average real-time throughput corresponding to T_{slot} can be found using this figure.

When real-time demand is high, the length of the HPP can grow until it reaches $T_{slot} = 0.1$ s. For a given real-time load, HPP reaches its final length after a transient period. Fig. 8 shows the average length of HPP at steady state. For larger N , the length of the HPP phase is longer because it takes longer to deliver packets originating from the furthest segment. It can be deduced that increasing the service area of the gateway increases the overhead of CVIA-QoS.

6.1.2. Transient response

In our simulations, maximum ratio of HPP to T_{slot} is taken as 1. In this case, HPP length is extended until the longest possible length. This is the worst case condition for the transient response. When real-time demand increases in the network, the CVIA-QoS protocol responds by first shrinking and then eliminating the Local Packet Gathering Phase, which prevents new best-effort packets from entering the network. However, even when no new best-effort packet is accepted, it takes time to empty buffers and the real-time packets can only use the full throughput capacity after some delay. The speed of this response depends on the maximum allowable number of packets in the TR buffers and the number of segments in the virtual

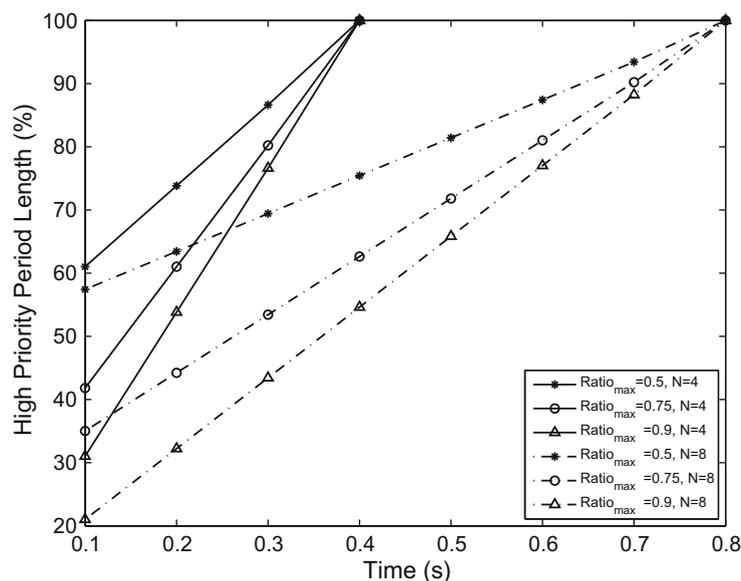


Fig. 9. Transient HPP length which can be dedicated to real-time traffic.

transmission radius of the gateway. Fig. 9 shows the transient behavior of the length of the HPP that can be dedicated to real-time traffic. When the curve reaches 100%, it means that all of the bandwidth is dedicated to real-time packets.

In this analysis, the buffers of TR^{in} s are assumed to be full with best-effort packets before real-time traffic increases. At time 0, the real-time demand is increased to the total protocol throughput capacity. Fig. 9 shows the HPP length percentage for Scenario II and Scenario IV for different $Ratio_{max}$ and N values after the increase in real-time load. It takes $N - 1$ time slots to empty the buffers of the furthest TR. Hence, it takes longer to serve real-time traffic with full capacity when N is 8 in Scenario IV. Although response time to reach full capacity depends on N, the bandwidth that can be dedicated to real-time traffic before reaching the maximum capacity depends on $Ratio_{max}$ value. In all Scenarios, when $Ratio_{max}$ is higher, the length of HPP that can be dedicated to real-time traffic becomes lower.

6.1.3. Throughput ratio of CVIA-QoS to CVIA

In this scenario, we consider a system at saturation where all vehicles have new best-effort packets to send. Fig. 10 shows the total throughput capacity of CVIA-QoS and CVIA when the real-time load is increased in the system for different N and

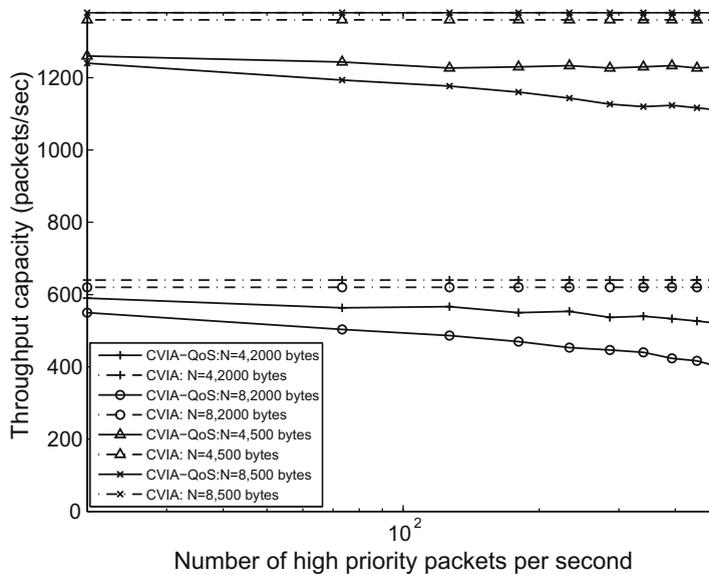


Fig. 10. Throughput capacity.

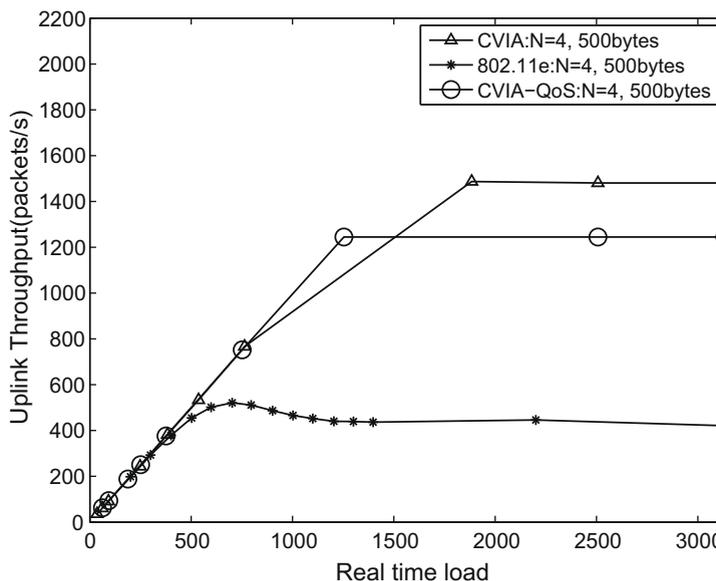


Fig. 11. Real-time throughput when there is only real-time packet demand.

payload values. Throughput capacity of CVIA-QoS is always smaller than CVIA since CVIA-QoS protocol allocates some portion of its throughput to provide delay bounded throughput guarantees to real-time traffic. Any real-time packet admitted to the system reserves bandwidth in segments on its path within the same time slot. As a result, when the N is high, the capacity of the system decreases for both payloads as shown in Fig. 10. When the real-time load is increased, more packets are served in the HPP period leading to a decrease in the overall capacity of the system. As a result, since handling best-effort traffic in HPP would result in decrease in total throughput, best-effort traffic is carried in LPP using CVIA guidelines.

6.2. Comparison of CVIA-QoS with CVIA and IEEE 802.11e

6.2.1. Throughput when no best-effort demand exist in the system

When no best-effort packet exists in the system, the throughput is only composed of real-time packets. As shown in Fig. 11, the throughput curves of all three protocols increase linearly when the packet load is low. When the load is increased further, the CVIA and CVIA-QoS protocols reach saturation points that are approximately three times the maximum

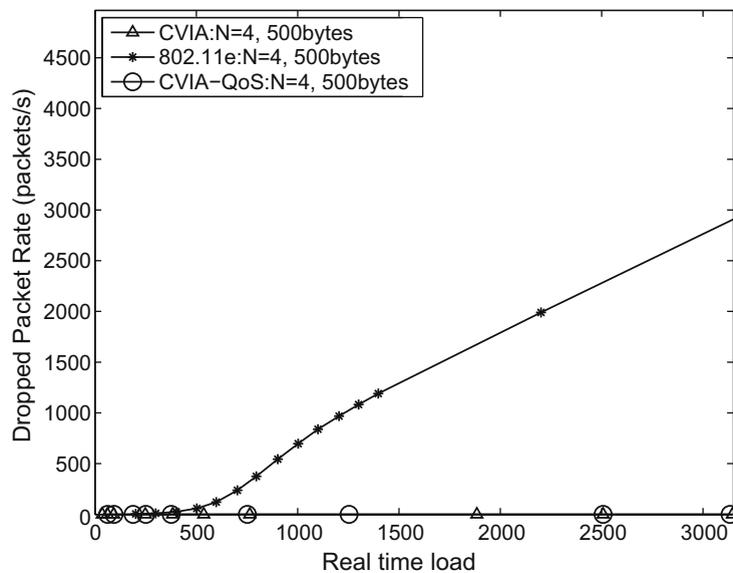


Fig. 12. Packet failure rate when there is only real-time packet demand.

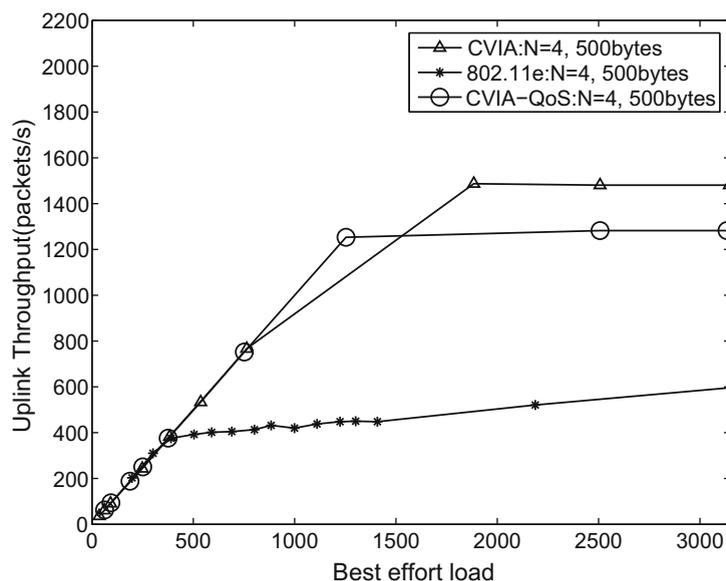


Fig. 13. Best effort throughput when there is only best-effort packet demand.

throughput of the IEEE 802.11e protocol. This figure shows that both CVIA and CVIA-QoS protocols use the channel more efficiently than IEEE 802.11e protocol. This low IEEE 802.11e throughput is caused by small contention window of the high priority packets which gives priority over low priority packets. However, when the real-time load is high, high priority packets start to collide among each other. As a result, many packets are dropped in IEEE 802.11e as shown in Fig. 12. This high number of packet failures decreases the throughput of the IEEE 802.11e protocol.

6.2.2. Throughput when only best-effort demand exist in the system

Fig. 13 shows the throughput curves of all three protocols when only the best-effort packets exist in the system. In this scenario CVIA-QoS protocol operates very similar to the CVIA protocol, where, except the registration and control phases, the entire slot is dedicated to low priority traffic. As a result of this overhead, the throughput of CVIA-QoS becomes lower than CVIA. The throughput of the 802.11e protocol is lower than the CVIA and CVIA-QoS. However, in this scenario, the number of packet failures are lower than the real-time load scenario as shown in Fig. 14.

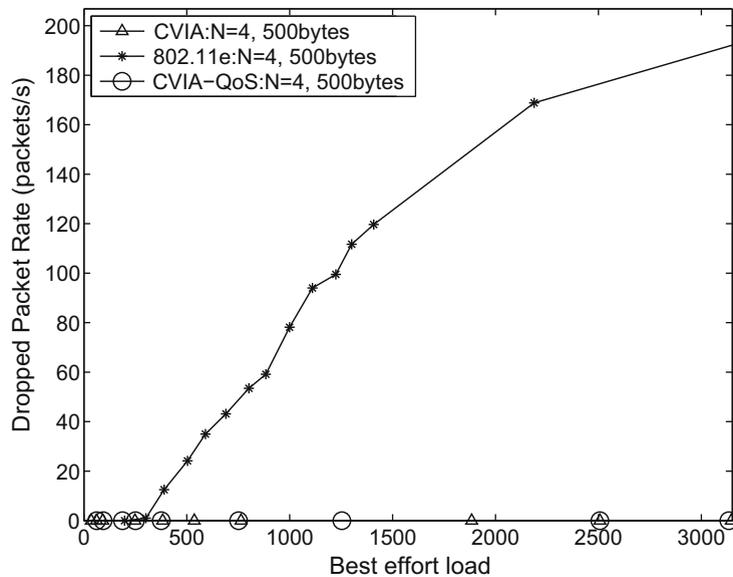


Fig. 14. Packet failure rate when there is only best-effort packet demand.

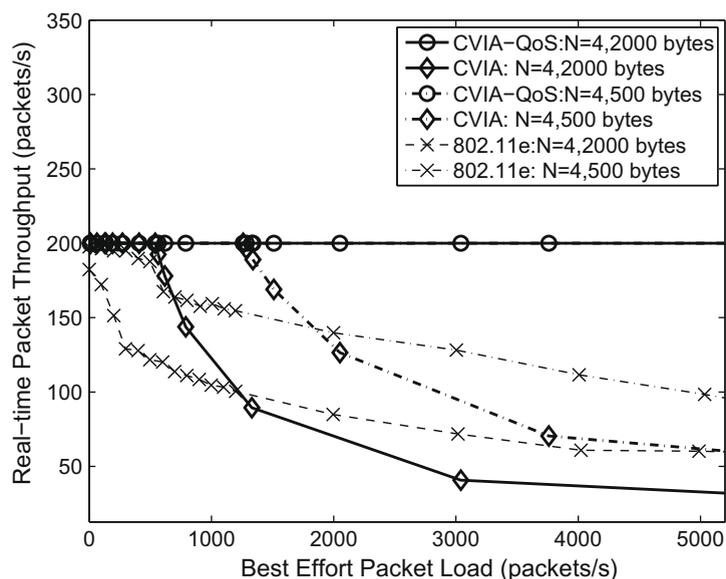


Fig. 15. Real-time throughput in CVIA, CVIA-QoS and 802.11e protocols (real-time demand is 200 packets/s).

6.2.3. Effect of best-effort load on real-time throughput

Up to this point, we have observed throughput of the protocols under only real-time or only best-effort load. In real operation, both real-time and best-effort load exist in the system at the same time, where the best-effort load should not affect the real-time throughput. Fig. 15 shows the effect of the increase in the best-effort load on the real-time throughput. In all scenarios, real-time demand is 200 packets/s and it remains constant. In CVIA and IEEE 802.11e, real-time and best-effort packets contend for the same channel. Therefore, when best-effort load increases, the share of real-time packets in the total end-to-end throughput decreases. On the other hand, since the CVIA-QoS protocol employs admission control and polls the real-time packets without any contention, the throughput of the real-time traffic is not affected by the background best-effort load. Creating a separate channel for the real-time packets is one of the most important features of the CVIA-QoS protocol that makes it superior when compared with the CVIA protocol.

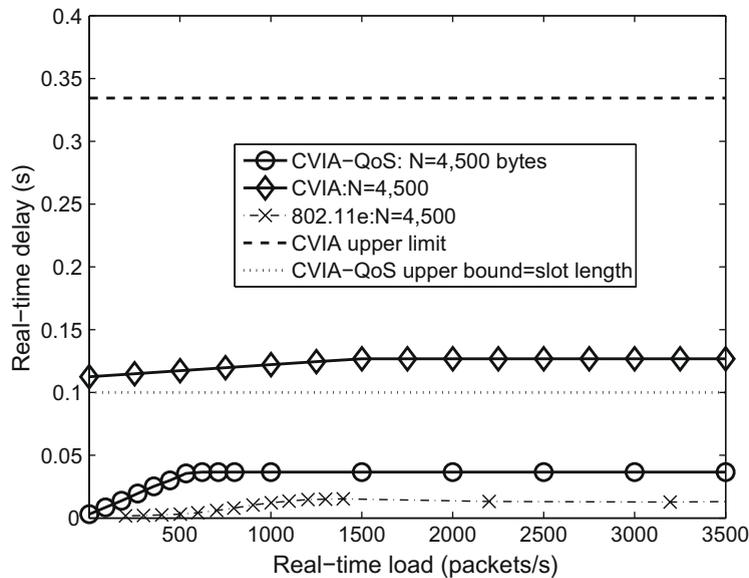


Fig. 16. 500 bytes payload, no best-effort load.

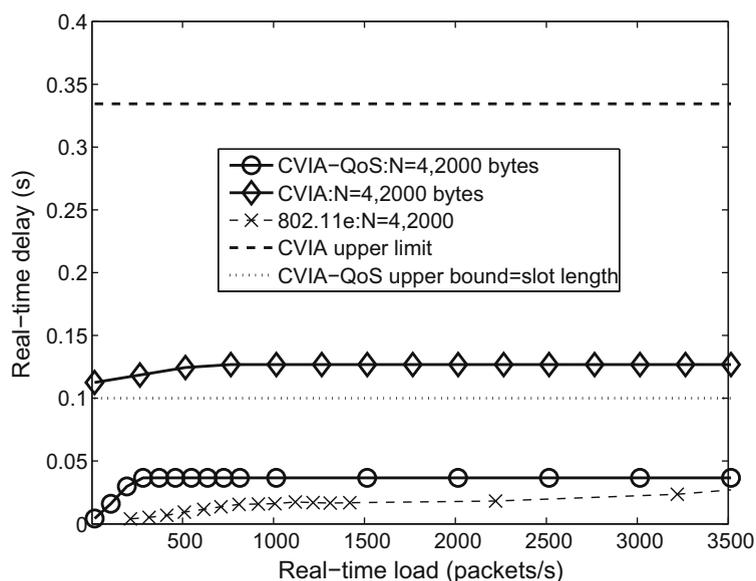


Fig. 17. 2000 bytes payload, no best-effort load.

6.2.4. Delay

In this section, the network delay is defined as the time elapsed between the instant the packet is transmitted from the source and the reception time of the packet by the final destination. Queueing delays in the relay nodes are also included in the delay measurement. In CVIA-QoS, packets admitted to high priority period are delivered to the gateway in a fast train reaching to the gateway in one T_{slot} . Thus, T_{slot} is the upper delay limit for the CVIA-QoS protocol and it is plotted in Figs. 16–19. On the other hand, under the CVIA protocol, all packets use the slow train which can move only one segment in each time slot. Especially when N is high, high priority packets originating from the furthest segments experience higher delays. All figures also show the average delay experienced by the high priority packets admitted to the network from the furthest segment in CVIA protocol as CVIA upper limit.

Figs. 16 and 17 show the network delay of real-time packets when there is only real-time load in the channel. Payload of these figures are 500 bytes and 2000 bytes, respectively. In these figures, the average CVIA delay increases with load and becomes almost constant. The delay of the CVIA-QoS protocol is always lower than the average CVIA delay. Moreover,

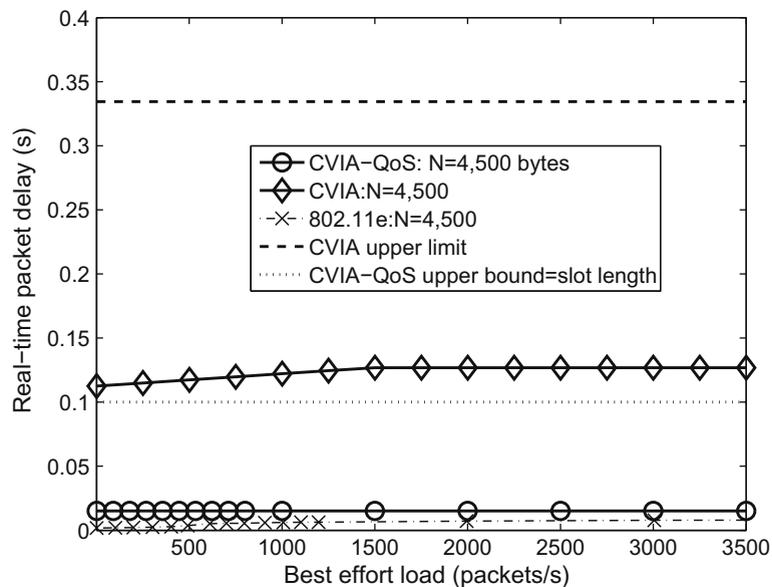


Fig. 18. 500 bytes payload, effect of best-effort load when real-time load is 200 packet/s.

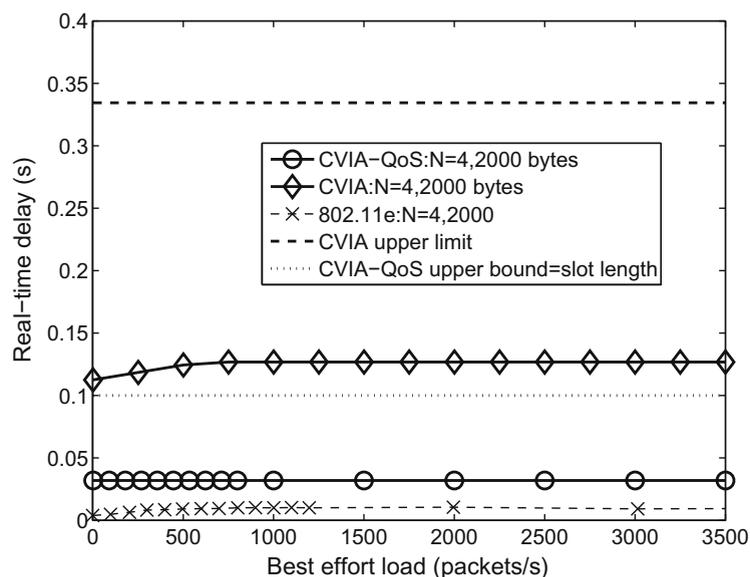


Fig. 19. 2000 bytes payload, effect of best-effort load when real-time load is 200 packet/s.

CVIA-QoS delay is much lower than CVIA upper limit. Note that if the number of segments were larger, the CVIA delay would be much larger while CVIA-QoS delay would not be affected. Although the delay of the IEEE 802.11e packets is the lowest in these figures, it is because of the high failure rate of IEEE 802.11e as shown in Fig. 12. Since the maximum contention window for high priority packets is small, these packets do not stay long in the queues of the relay nodes, and most of them collide with other packets. Figs. 18 and 19 show the network delay of real-time packets for the mixed load scenario where the real-time load is constant and 200 packets/s. Note that in CVIA-QoS, similar to the throughput performance, real-time delay is not affected by the best-effort load and it is almost six times less than the average and 17 times less than the upper limit of the CVIA protocol.

7. Conclusions and future work

In this paper, a new channel access and routing strategy for vehicular Internet access along highways with QoS support is introduced. The CVIA-QoS protocol use periodic admission control and scheduling for soft real-time traffic to provide delay bounded throughput guarantees. After the demands of the soft real-time traffic is met, CVIA-QoS supports the best-effort traffic by allocating the remaining bandwidth.

Our simulation results confirm that in CVIA-QoS protocol, the real-time throughput is not affected from the best-effort load and its delay is much smaller than CVIA delay when both the real-time and best effort is carried in the network. The throughput of IEEE 802.11e is much lower than both CVIA-QoS and CVIA. In addition, IEEE 802.11e suffers from high number of real-time packet failures at high loads.

In our future work, the CVIA-QoS protocol will be improved by allowing coordination of multiple gateways and multi-segment reservation strategies. The packet traffic can be diverted from busy gateways to the gateways experiencing light load to improve performance. In a linear highway without intersections, this can be achieved by varying the virtual service area of the gateways based on the load. However, if intersections exist in the packet trains' paths which are shared by more than two gateways, a scheduling must be done for the packets of multiple gateways.

References

- ASTM, 2002. ASTM E2213-02 Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems 5 GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications.
- Baccelli, E., Clausen, T., Jacquet, P., Nguyen, D., 2007. Integrating VANETs in the internet core with OSPF: the MPR-OSPF approach. In: International Conference on ITS Telecommunications (ITST).
- Bechler, M., Franz, W., Wolf, L., 2003. Mobile internet access in FleetNet. In: 13. Fachtagung Kommunikation in Verteilten Systemen (KiVS 2003).
- Bianchi, G., 2000. Performance analysis of the 802.11 distributed coordination function. IEEE JSAC 18 (3), 535–547.
- Bilstrup, K., 2007. A survey regarding wireless communication standards intended for a high-speed vehicle environment. Tech. Rep. IDE0712, Univ. Halmstad.
- FCC, October 1999. FCC 99-305. FCC Report and Order.
- FCC, February 2004. FCC 03-024. FCC Report and Order.
- Festag, A., Noecker, G., Strassberger, M., Libke, A., Bochow, B., Torrent-Moreno, M., Schnauffer, S., Eigner, R., Catrinescu, C., Kunisch, J., 2008. (NoW) Network on wheels: project objectives, technology and achievements. In: 5th International Workshop on Intelligent Transportation (WIT), pp. 211–216.
- FLEETNET, 2008. FleetNet homepage. <<http://www.et2.tu-harburg.de/fleetnet/english/vision.html>> [accessed 12.01.2008].
- Giannoulis, A., Fiore, M., Knightly, E.W., 2008. Supporting vehicular mobility in urban multi-hop wireless networks. In: MobiSys '08: Proceeding of the 6th International Conference on Mobile Systems, Applications, and Services. ACM, New York, NY, USA, pp. 54–66.
- Hinsberger, A., Wieker, H., Riegelhuth, G., Zurlinden, H., 2007. Benefits and technology of an intelligent roadside unit system for vehicle to infrastructure and infrastructure to centre communication. In: 14th World Congress on Intelligent Transport Systems.
- Korkmaz, G., Ekici, E., Ozguner, F., 2006. A cross-layer multi-hop data delivery protocol with fairness guarantees for vehicular networks. IEEE Transactions on Vehicular Technology 55 (3), 865–875.
- Lee Armstrong, 2008. Dedicated Short Range Communications (DSRC) Homepage [online]. <<http://www.leearmstrong.com/dsrc/dsrchomeset.htm>> [accessed 12.01.2008].
- Perera, E., Sivaraman, V., Seneviratne, A., 2004. Survey on network mobility support. Mobile Computing and Communications Review 8 (2), 7–19.
- Pinart, C., Barona, I., Alba, C., 2008. Experimental measurement of vehicular QoS. In: 7th European Congress and Exhibition on Intelligent Transport Systems and Services (ITS 2008).
- Tchouto, J.-J., Kutzner, K., Bochow, B., Luckenbach, T., Bechler, M., Wolf, L., 2003. Connecting vehicle scatternets by internet-connected gateways. In: Multiradio Multimedia Communications.
- Tsugawa, S., Kato, S., 2003. Evaluation of incident information transmission on highways over inter-vehicle communications. In: Intelligent Vehicles Symposium, 2003, pp. 12–16.
- Tsukada, M., Mehani, O., Ernst, T., March 2008. Simultaneous usage of NEMO and MANET for vehicular communication. In: WEDEV 2008: 1st Workshop on Experimental Evaluation and Deployment Experiences on Vehicular Networks in conjunction with TRIDENTCOM 2008, Innsbruck, Austria, March 18, 2008.
- Wang, Z., Hassan, M., 2008. How much of DSRC is available for non-safety use? In: VANET'08.
- Xu, Q., Mak, T., Ko, J., Sengupta, R., October 2004. Vehicle-vehicle safety messaging in DSRC. In: 1st ACM Workshop on Vehicular Ad-hoc Networks, pp. 19–28.