Computer Networks 55 (2011) 2443–2464

Contents lists available at ScienceDirect

Computer Networks

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

Opportunistic broadcast of event-driven warning messages in Vehicular Ad Hoc Networks with lossy links

Ming Li [a,*], Kai Zeng [b], Wenjing Lou [a]

[a] Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA, United States
[b] Department of Computer Science, University of California, Davis, CA, United States

ARTICLE INFO

Article history:
Received 23 August 2010
Received in revised form 7 April 2011
Accepted 13 April 2011
Available online 23 April 2011
Responsible Editor: E. Ekici

Keywords:
Vehicular Ad Hoc Networks
Broadcast
Opportunistic routing
Lossy links
Reliability

ABSTRACT

Multi-hop broadcast is a key technique to disseminate time-sensitive event-driven safety warning messages (WMs) in Vehicular Ad hoc Networks (VANETs). Due to the lossy nature of the vehicular wireless environment and the fact that the implementation of broadcast at the link layer uses unreliable transmissions (i.e., lack of positive ACKs), highly reliable, scalable, and fast multi-hop broadcast protocol is particularly difficult to design in VANETs with lossy links. Schemes that use redundant network layer broadcasts have been proposed. However, the tradeoff between reception reliability and transmission count in such schemes needs to be carefully considered.

In this paper (The preliminary version of this paper appeared in [1], IEEE MASS 2009.,) we propose an opportunistic broadcast protocol (OppCast) that aims at simultaneously achieving high WM packet reception ratio (PRR) and fast multi-hop message propagation while minimizing the number of transmissions. A double-phase broadcast strategy is proposed to achieve fast message propagation in one phase and to ensure high PRR in the other. OppCast exploits opportunistic forwarding in each transmission to enhance the WM reception reliability and to reduce the hop delay, and to carry out reliable and efficient broadcast coordination, we propose the use of explicit broadcast acknowledgements (BACKs) which effectively reduces the number of redundant transmissions. OppCast is also extended to handle sparse and disconnected VANETs, where the protocol adaptively switches between fast opportunistic forwarding and the store-carry-and-forward paradigm. Extensive simulation results show that, compared with existing competing protocols, OppCast achieves close to 100% PRR and faster dissemination rate under a wide range of vehicular traffic densities, while using significantly smaller number of transmissions.

1. Introduction

Communication in Vehicular Ad Hoc Networks (VANETs) has been an active research area in recent years. VANET is a multi-hop mobile network designed to provide a wide range of road applications such as safety warning [2,3], congestion avoidance or mobile infotainment [4]. One of the most important applications of VANET is the broadcast of event-driven emergency warning messages (WMs) like accident and hazard warning. For example, after two vehicles collided with each other on a highway, or traffic congestion happens because of heavy rain or snow, the upcoming vehicles need to be notified immediately. In both cases, the WMs should be disseminated out with short delay to vehicles that are up to several kilometers away, not only to prevent more possible accidents, but also to enable the vehicles to make a detour as early as possible to avoid congestion. Some example applications of event-driven warning message along with their communication requirements are given in [5]. According to Chen

* Corresponding author. Tel.: +1 508 831 5649.
E-mail addresses: mingli@wpi.edu, mingli@ece.wpi.edu (M. Li), kzeng@cs.ucdavis.edu (K. Zeng), wjlou@ece.wpi.edu (W. Lou).

1389-1286/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2011.04.005

et al. [5] and the Dedicated Short Range Communication [6] (DSRC), the typical one-hop broadcast delay requirement for many event-driven messages varies from 100 to 500 ms within an one-hop communication range from 200 to 300 m, while the typical delay of periodical safety messages is smaller than 100 ms. In situations where the one-hop communication range of a vehicle does not reach the intended distance of a warning message, multi-hop broadcast is necessary to disseminate those time-sensitive warning messages through VANET. For the delay requirement of multi-hop broadcast WMs, it is natural to extend those of single-hop WMs, i.e., multiply the delay requirement of the single-hop WM by the number of hops needed. For example, an emergency brake light message should be delivered to vehicles 1000 m away within 400 ms.

There are three main performance goals in WM broadcast. (1) High reliability, which is usually measured as the percentage of vehicles that received the warning message. (2) Fast dissemination, that is the warning messages should be delivered to the vehicles with short end-to-end delay. (3) High scalability, which means the WM's propagation should incur small transmission overhead (especially when the network is dense), since unnecessary transmissions waste precious bandwidth resource in VANET. However, in real VANETs these goals are hard to achieve simultaneously. The major challenge comes from the lossy wireless transmissions [7,8], which undermine the reliability of one-hop broadcast. According to studies on the DSRC [9], the one-hop broadcast reception rate is low, due to both channel fading and packet collisions caused by hidden terminals. Also, there is no channel resource reservation mechanism in 802.11 for broadcast, which could incur severe packet collisions in a dense network with congested channels. Unlike unicast, in VANET it is difficult to let every vehicle acknowledge the reception of each broadcast message, mainly due to the ACK implosion problem [10]. Therefore, there is hardly a reception guarantee for one-hop link layer broadcast.[1]

Since it often incurs high complexity to enhance the reliability of broadcast from link-layer, most previous works have focused on broadcast strategies that use redundant network layer retransmissions. The blind flooding leads to the well-known *broadcast storm* problem [12] where packet collisions could arise due to uncoordinated simultaneous rebroadcasts. Various schemes were proposed to mitigate this problem, such as probability-based schemes [13] and timer-based schemes [14–17]. Although these schemes enjoy high reliability when the channel load is moderate, the amount of redundant transmissions becomes prohibitively large under dense network or heavy data traffic. This drawback heavily degrades the broadcast performance, and limits their scalability to be deployed in real VANETs.

In this paper, we enhance the reliability of WM dissemination in VANETs from both the network and link layers. From the link layer, we exploit opportunistic reception and forwarding of WMs to combat the lossy links, and use a explicit broadcast acknowledgement at each hop as reception feedback to ensure the penetration of a WM to the whole interested region. From the network layer, we propose to use controlled redundant rebroadcasts to ensure the reception of a WM by vehicles within localized parts of the network. We also propose distributed and optimized parameter selection methods that minimizes the transmission count. Our protocols can achieve very close to 100% packet reception ratio (PRR) under a variety of vehicular scenarios. Our main contributions can be summarized as follows.

First, we propose the *opportunistic broadcast protocol* (OppCast), a fully distributed protocol that simultaneously achieves high reliability and fast WM propagation while incurring small transmission overhead. The broadcast scheme consists of two types of broadcast phases, where one phase quickly propagates a WM using relatively long hops, and the other phase uses additional make-up transmissions between the long hops to ensure a certain PRR. The designs of both phases are optimized to minimize the total number of transmissions.

Second, we propose a distributed *opportunistic broadcast coordination function* (OBCF), a reliable and efficient MAC-layer broadcast primitive for the recipients of a single broadcast to select the "best" relay nodes in a localized manner. OBCF exploits the idea of opportunistic forwarding to enhance the reception reliability and reduce the hop delay in each single transmission; by transmitting a long-range, short broadcast acknowledgement (BACK) before each rebroadcast, OBCF effectively minimizes the redundant transmissions and alleviates the hidden terminal problem in a lossy environment.

Third, we propose OppCast-Ext that copes with network partitions in sparse VANETs. To maintain both high WM reception reliability and small end-to-end delay, OppCast-Ext switches adaptively between the fast propagation mode and the store-carry-and-forward mode, where the first one is used within continuous vehicle platoons and the second is used between platoons. The optimal switching condition between the two modes are characterized via both theoretical analysis and simulations.

Finally, we carry out extensive NS-2 simulations to evaluate the performance of OppCast and OppCast-Ext. Results show that OppCast and OppCast-Ext outperform several state-of-the-art protocols by achieving close to 100% PRR and faster dissemination under a wide range of scenarios, while the transmission overhead is much smaller. The tradeoff between reliability, end-to-end delay and transmission overhead is characterized. To the best of our knowledge, this is the first work that thoroughly studies this tradeoff under a realistic physical layer model in VANETs.

The rest of the paper is organized as follows. Section 2 presents the related work, and in Section 3 we give the problem statement. Followed by that is an overview of our OppCast scheme in Section 4, and then the main design part is presented in Section 5. Section 6 presents the theoretical analysis on parameter optimization, and Section 7 is the performance evaluation. Section 8 concludes the paper.

[1] Although recent technique [11] can solve this problem in wireless networks with fixed topology, it is less suitable for VANETs with dynamic topology.

2. Related work

Earlier works on broadcast protocols in VANETs mostly assumed the ideal propagation model, i.e., whether a packet reception is successful merely depends on a fixed transmission range. Examples include the UMB [18] and SB [19]. They designate the furthest node that receives a packet as the relay node to maximize the one-hop progress [18] or minimize the one-hop delay [19].

However, as shown by empirical studies, the wireless channel in VANETs is far from perfect [20]. Instead, channel fading is the primary challenge and has a major impact on broadcast reception rates [9,8]. The suggested realistic propagation model on the highway is the *Nakagami model* [7], where the packet reception probability of single broadcast decreases with the distance, which aggravates the "broadcast storm" problem. The *probability-based methods* [12,13] simply let each node rebroadcast a packet with some probability. However, the probabilistic forwarding decision still results in redundant rebroadcasts and does not solve the broadcast storm problem.

2.1. Opportunistic forwarding

A promising way to deal with lossy links in multi-hop wireless networks is *opportunistic forwarding* [21,22]. It exploits the "spatial diversity" enabled by the broadcast nature of the wireless medium, so that the probability of at least one node forwards a transmitted packet is greatly increased. In this way, each transmission is useful with high probability, and the forwarding delay and the number of transmissions can be reduced.

The concept of opportunistic forwarding has recently been applied to event-driven emergency message broadcast in VANETs. Existing works include contention-based dissemination (CBD) [16], emergency message dissemination for vehicular environments (EMDV) [3], contention-based forwarding [8], OB-VAN [23] and location-based flooding [14]. In these works, in order to maximize the "hop progress" of the message in its each rebroadcast, the common idea is to employ a contention process where the farthest node in the message dissemination direction is opportunistically selected. This is often achieved by letting each node that receives the same rebroadcast packet set up a contention timer, in which the backoff time is inversely proportional to the distance of that node to the sender. As a result, the nodes located nearer to the sender backoff longer times and quit contention whenever they hear rebroadcasts (or ACK signals) from a node that has larger hop progress.

However, under the presence of lossy wireless links, the previous schemes could incur unnecessary duplicate rebroadcasts (especially under congested channel) [3], which takes up the previous bandwidth of the VANET and may adversely affect the performance of other messaging services such as periodical safety beacon messages. This is because the rebroadcast of a vehicle located farthest in the WM dissemination direction may not be heard by other contending nodes due to channel fading. Although these duplicate rebroadcasts can enhance the reliability of WM reception, currently the "duplicate level" is often controlled by adjusting the threshold count of received duplicate messages for each contending node to decide whether to quit the contention. However, it is very difficult to select the appropriate parameters so that a desired WM reception reliability level is achieved, without causing more redundant rebroadcasts.

In addition, the contention processes could result in undesirably large end-to-end WM propagation delay. For nodes to decide whether to participate in contention, there is always some "contention range" chosen by each forwarder upon its rebroadcast. If this range is too large, even the farthest node would wait for a long time before rebroadcast since the contention count-down time tends to be prolonged. Towards solving this problem, in EMDV [3] Torrent-Moreno et al. proposed to first designate the farthest node (who would rebroadcast without delay) within a "forwarding range" that is shorter than the communication range, and if that node does not receive the rebroadcast then other nodes continue participating in normal contention processes like that in CBD [16]. However, the issue of how to select an optimal forwarding range has not been addressed.

In summary, the inefficiency of the existing schemes are resulted from two aspects. (1) Nodes make forwarding decisions based on heuristic guesses of whether neighbors have received the same packet or not. The reliability requirement (PRR) of the WM propagation is not considered, and no optimizations have been made so far. (2) In the coordination mechanisms, rebroadcast messages are employed as "implicit acknowledgements", which always subject to channel fading and collisions. Thus it is hard to effectively suppress unnecessary redundant rebroadcasts. In this paper, we solve these two problems accordingly, by (1) explicitly considering the reliability (PRR) as one of the relay node selection criteria, in that we minimize the number of rebroadcasts to satisfy a given PRR requirement; (2) designing a more reliable and efficient broadcast coordination mechanism, where a short, explicit "broadcast acknowledgement" (BACK) is send out (at the base rate) before each WM's rebroadcast, which has larger communication range than the normal WMs. The BACK not only effectively suppresses redundant rebroadcasts, but also clears the channel for the rebroadcast. With BACK, the contention (relay selection) processes can be optimized for higher reliability and lower end-to-end delay.

2.2. Reliability in VANET broadcast

Elbatt et al. [24] studied one-hop periodic broadcast in cooperative collision warning applications. They characterized the tradeoffs between the packet inter-reception latency, application broadcast rate and transmission range. For multi-hop WM broadcast, Resta et al. [25] analyzed theoretically the tradeoff between vehicles' probability to receive a WM within time t and the link level reliability. A comprehensive survey of broadcast protocols in vehicular networks can be found in [5]. Our work differs from the above in that we cast insight on the application-level tradeoff between packet reception performances and the overall transmission count, under a realistic channel model.

2.3. Scalability in VANET broadcast

In [26], Scheuermann et al. proposed a scalability criterion for data aggregation in information dissemination of general purpose information in VANETs, where the targeted communication paradigm is continuous transmission of measurement data from multiple sources to multiple destinations. Example applications include cooperative traffic information management or decentralized parking guidance systems. Their main result is that, any suitable aggregation scheme must reduce the bandwidth at which information about an area at distance d is provided to the cars asymptotically faster than $1/d^2$.

However, in this paper we are focusing on emergency warning message dissemination, which is different from "general purpose" warning messages. The "scalability" definition in our paper and many other papers [27,13] refers to the overall transmission count incurred in the IR per source packet. If this transmission count is high, it could adversely affect the performance of the WM broadcast protocol by increasing the chance of collisions. Thus in this paper, we reduce the transmission count via broadcast suppression techniques (i.e, OBCF).

2.4. Partitioned VANETs

VANETs turn out to be partitioned (or disconnected) sometimes, especially under sparse vehicular traffic, which falls into the delay-tolerant network (DTN) paradigm. Ilias and Cecilia [28] proposed an opportunistic event dissemination protocol that employs cache and periodic replay mechanisms to keep a message alive in an area. In [29], the authors proposed a routing protocol, which uses local routing in connected clusters and store-carry-forward at cluster boundaries in order to reduce latency and overhead. In this paper, OppCast is extended into a DTN-compatible broadcast scheme, where the protocol adaptively switches between normal dissemination and store-carry-and-forward modes based on vehicles' local traffic densities. Recently and independently from our work, Tonguz et al. proposed DV-CAST [27], which extended the work in [13] to handle network disconnections. The rebroadcast decisions in DV-CAST are also made based on local topology information; however, there are no reliability (or, broadcast success rate) guarantees in DV-CAST. In [30], Ros et al. proposed an acknowledged parameterless broadcast protocol (AckPBSM) which can handle a variety of vehicular scenarios including intermittent connectivity. A connected dominating set (CDS) is established based on local topology information indicated by periodic beacon messages to reduce redundant rebroadcasts, and acknowledgements of received packets are piggybacked in beacons to ensure the reception reliability. However, the length of beacon message will grow linearly with the number of generated WMs, which may cause a scalability issue. Meanwhile, Ros et al. [30] do not study the tradeoff between reliability level and overhead (transmission count).

Different from all the previous works, in our OppCast extension we focus on finding the optimal switching traffic density, beyond which unnecessary redundant transmissions will be incurred, and below which the desired PRR requirement cannot be fulfilled.

3. Problem statement

3.1. Model and assumptions

In this paper, we present our event-driven warning message (WM) broadcast protocol using a highway scenario. Fig. 1 shows the system model, which is a line-topology highway that may have multiple lanes. The VANET consists of vehicles equipped with on board units (OBUs) that can communicate with each other. Suppose a safety-related event (e.g. an accident) happens somewhere, where the source vehicle's OBU begins to disseminate WMs towards the interested region (IR) via broadcast. The IR is defined as the road segment of length $\mathscr{L}$ along the message dissemination direction, which can either include only the co-directional lanes or lanes in both directions. The source is called the *origin of IR*, and the other boundary of IR is called the *end of IR*. Since the width of the highway is far less than the length of IR, for simplicity we model the vehicles to be located in one-dimension.

We assume vehicles are GPS-capable and each vehicle obtains its location in real-time. This is also assumed by many other works in the literature [31,32,18,16,3,13,27].

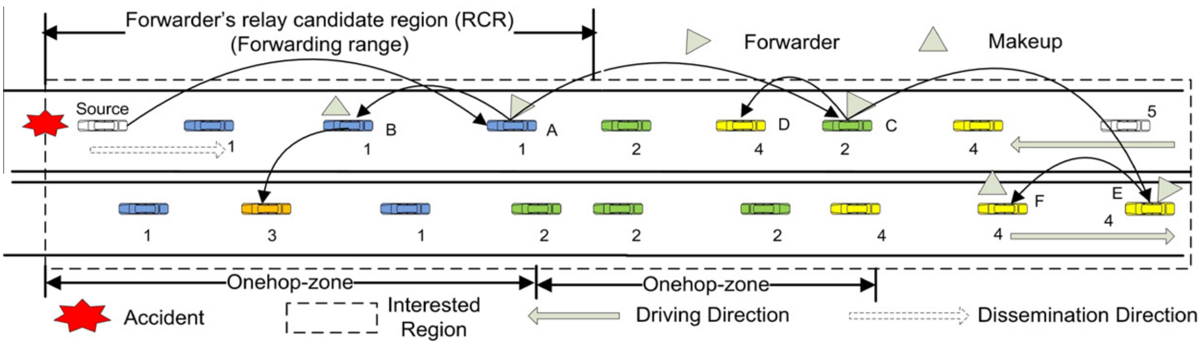

Fig. 1. VANET model and overview of the broadcast scheme. The number i near each vehicle indicates that the vehicle receives the WM upon the ith (re)broadcast.

When GPS is not available (e.g. in tunnels), vehicles use complementary methods to estimate their locations. For example, this can be done by combining the vehicle speed measured by the speedometer with the GPS map. Also, vehicles are aware of the existence and locations of all neighboring vehicles, as they broadcast one-hop beacon messages every 100 ms [6]. These beacons are routine safety messages, and warning messages are event-driven. They share the control channel [33].

The *network of interest* can be modeled as an undirected graph $G(V,E)$, where V is the set of nodes within the IR. We adopt a probabilistic radio propagation model, where the probability that a node v at a distance x from a node u receives a broadcast packet directly from u is expressed as a (decreasing) function with x, which is denoted as $P_r(x)$. This function accounts for channel fading; it can either be derived from a propagation model [34] or measured from practice. In reality, the VANET channel condition may vary with time and space, and there exist algorithms to adaptively adjust the radio model parameters in order to better cope with the impact of channel variations. This is orthogonal to the problem studied in this paper, and interested readers please refer to [35–37]. In this paper, we assume identical node transmission power, and each bidirectional link $l = (u,v) \in E$ is associated with a packet reception probability $P_r(d(u,v))$, where $d(u,v)$ is the distance between u and v. In addition, the packet reception at each vehicle is assumed to be independent.

3.2. Objectives

It is essential in VANET to let every vehicle receive an emergency warning message for critical applications. For the multi-hop WM dissemination considered in this paper, since we are also interested in investigating the tradeoff between PRR and transmission count, we aim at providing reliable broadcast service in terms of ensuring the *packet reception ratio* (PRR) of the network of interest to be larger than a threshold P_{th} ($PRR \geqslant P_{th}, P_{th} \in (0,1)$). This is called the *network PRR requirement*. The PRR requirement can vary arbitrarily; thus we believe this is a more general approach to WM broadcast reliability than previous works.

Definition 1 (Packet reception ratio *PRR*). Given a network G and a source s, *PRR* is defined as the percentage of nodes in IR that receive a WM originated from s.

In the meantime, it is also very important for vehicles to be warned in a timely manner. Since the broadcast reception delay ($t_{v,m}$) of WM m at each vehicle v is related to v's distance to the source (d_v), we define the individual dissemination rate as $d_v/t_{v,m}$. The *dissemination rate* is then defined as the individual dissemination rates averaged among all WMs sent and vehicles in the IR. Therefore, the second goal is to reach high dissemination rate. The PRR and dissemination rate capture the application level performances. Note that, to capture the worst case performance, we also define the broadcast delay of a WM to be the maximum end-to-end delay for the last vehicle in the IR to receive it.

Finally, in WM broadcast it is desirable to minimize the transmission overhead, which is defined as the expected total number of transmissions incurred for each WM. Unnecessary transmissions take up bandwidth, increase the channel access delay and the chance of packet collision. This, in turn, degrades the broadcast performance.

4. Overview of OppCast

The OppCast consists of two types of broadcast phases: *fast-forward-dissemination* (FFD) and *makeup-for-reliability* (MFR). Intuitively, the FFD phase uses relatively long hops to advance the WM towards the end of IR for the purpose of fast propagation. The FFD phase is realized via relaying the WM by a series of *forwarder nodes* that lie successively along the message dissemination direction, where each next hop forwarder node's distance to the previous one is relatively large. These forwarder nodes thus divide the IR into several *one-hop zones*. Due to lossy links and the independent reception assumption, however, vehicles within these one-hop zones may not all receive the packet upon one relay node's transmission. Thus we use additional make-up transmissions that constitute MFR phases to ensure the PRR of the network. In particular, in the MFR phase of each one-hop zone, a minimal set of *makeup nodes* are successively selected until the accumulated packet reception probability of each node within that one-hop zone is larger than a pre-defined P_{th}. In order to satisfy the PRR requirement of the whole network, the reliability of each hop's forwarder node selection is ensured by a retransmission mechanism. By deriving the optimal parameter that controls the length of the one-hop zones, the PRR requirement is satisfied with the least transmission overhead. Note that, we refer to both forwarder and makeup nodes as relay nodes.

The concept of opportunistic forwarding is exploited by OppCast in every transmission to enhance the WM reception reliability and minimize the broadcast latency. Each relay node's (re)broadcast triggers the selection of next forwarder or/and makeup nodes, where the selection of each type of relay is associated with a different relay candidate region (RCR). In the RCR, each node is a potential candidate of the next relay node; due to the broadcast nature of the wireless medium, this greatly enhances the probability that at least one relay candidate receives and rebroadcasts the WM, especially when the network is dense. In the FFD phase, each node (forwarder candidate) that received a (re)broadcast from a previous forwarder contends to become a forwarder in a distributed manner. To maximize the hop progress, each forwarder candidate computes a backoff delay that is inversely proportional with the distance from it to the previous forwarder. The one with the largest hop progress will rebroadcast first and become the forwarder. In order for the forwarder to reliably and efficiently suppress other forwarder candidates, we propose to use an *explicit broadcast acknowledgement* (BACK) message before each (re)broadcast. The BACK is a short message with longer range than ordinary event-driven WMs, since it is sent at the base rate while WMs are sent at a higher rate. In this way, the previous (re)broadcast is acknowledged, and redundant rebroadcasts are avoided. For each MFR phase, the above contending mechanism is also used to select the makeup nodes, where the selection priority is set as how much additional reception reliability can each makeup candidate bring to its neighbors.

The BACK is a key component in OppCast, and in order to realize the above concepts, we design an opportunistic broadcast coordination function (OBCF) as the underlying MAC protocol used in each broadcast transmission. Specifically, we use the BACK as a way to clear the channel for the subsequent rebroadcast (similar to the function of clear-to-send (CTS) in IEEE 802.11 unicast), which can suppress most of the hidden terminals. Furthermore, we enhance the backoff delay function in previous works, to reduce the hop delay and the possibility of packet collisions. As a result, it is ensured with high probability that in each RCR of each transmission, only one relay is selected.

Due to the use of BACKs, we are able to carry out optimizations on relay selection. To minimize the total number of incurred transmissions for each event-driven WM, we carefully consider the tradeoff between WM dissemination rate and the transmission count. Central to this tradeoff is the length of the RCR for selecting forwarders, namely *forwarding range* (FR). We found the optimal FR, given the vehicle traffic density and the PRR requirement.

In addition, we extend OppCast to handle the partitioned, sparse VANETs. The vehicles in the opposite road of the source are employed as data mules only if a forwarder indicates that its local traffic density is smaller than a certain threshold. Therefore, the protocol adaptively switches between the normal (fast propagation) mode and store-carry-and-forward mode according to the local traffic densities.

Next we use a simple example to illustrate the broadcast process of OppCast in a well-connected VANET (the normal mode). Fig. 1 shows a bi-directional highway with the WM source in the upper lane. After the 1st broadcast by the source, the vehicles with number 1 receive the WM. Those vehicles inside the forwarding range of the source start the forwarder contention process, and node A that is farthest to the source becomes the forwarder and sends a BACK before it actually rebroadcasts. After A's rebroadcast (2nd), node B contends and becomes the makeup for the first one-hop zone between source and A (since B derives that the minimum packet reception probability for nodes in that zone is smaller than P_{th}), and node C is selected as the next hop forwarder. Node B's 3rd rebroadcast actually covers the rest of nodes in the first one-hop zone that had not received the WM, while node C's 4th rebroadcast forms the second one-hop zone between A and C. Now all the nodes in the first and second one-hop zones compute the packet reception probability for others, and found that to be larger than P_{th}. Therefore no makeup nodes are further selected. In the mean time, the WM is being propagated further along the dissemination direction. A high-level protocol flow-chart of OppCast is given in Fig. 2.

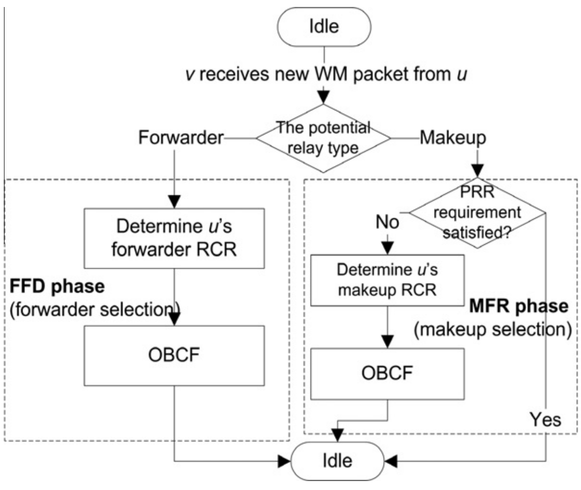

Fig. 2. The high-level flow chart of OppCast (at node v when receiving a WM from node u).

5. OppCast: main design

In this section, we describe the components of OppCast. We begin by introducing the FFD and MFR phases from the high level, and then present the underlying OBCF MAC protocol used in relay selection for each transmission. The main notations in this paper are summarized in Table 1.

5.1. Fast forward dissemination

The goal of the FFD phase is to achieve fast WM propagation by using relatively long forwarding hops. Immediately after each forwarder F rebroadcasts, all nodes within the *forwarder RCR* that receive the WM are *candidate forwarders* and participate in the contention process to become the next hop forwarder. For F, its forwarder RCR is a road segment from F towards the end of IR, whose length is called the *forwarding range* (FR). The priority of the candidate forwarders increases with the hop-progress (their distance to F), in order to maximize the dissemination rate. Ideally, the candidate forwarder that is farthest to the previous forwarder should become the next hop forwarder. Each candidate forwarder sets a backoff timer that

Table 1
Summary of main notations.

RCR	Relay candidate region
FR	The forwarding range (length of forwarder's RCR)
IR	The interested region
PRR	Packet reception ratio
F	A forwarder
$G(V,E)$	The modeled graph of the VANET in IR
P_{th}	The required PRR threshold
$d(u,v)$	Distance between two nodes u and v
$P_r(x)$	Packet reception probability at a distance x
$\xi(v)$	Accumulated packet reception probability (APRP) at node v
$E[NT]$	Expected total number of transmissions
$M_{\ell,\lambda}$	A makeup node at the ℓth level and λth branch
$Z_{\ell,\mu}$	A subzone of ℓth level and number μ
$W_{\ell,\mu}$	The middle point used to select makeup node $M_{\ell,\mu}$
Δt_v	The backoff delay of node v for relay contention
ρ	Vehicle density

is inversely proportional to the hop progress, and a BACK is sent out immediately after this timer expires to suppress others. To ensure that only *one* forwarder is selected for each hop, the BACK itself should be reliable and not prone to collision. Consequently, an one-hop zone is formed between two successive forwarders, whose length is upper bounded by FR; and the index of the one-hop zone increases one-by-one. Note that, in order to propagate the WM all through the IR, retransmissions are adopted by each forwarder to ensure that a next hop forwarder is selected. The details of the forwarder selection mechanism are presented in the OBCF in Section 5.3.

Apart from the OBCF, the other key design issue here is how to choose an appropriate value for the parameter FR. Intuitively, the larger the FR is, the faster a WM can be disseminated. However, this may result in a larger transmission count. Recall that we adopt a probabilistic propagation model, if the actual hop progress is too large the expected percentage of nodes within the one-hop zone that receives the WM will be lower than the desired PRR. Thus more makeup nodes within the one-hop zone will be needed to help rebroadcast the WM, and more transmissions could in turn slow down the overall dissemination rate. On the other hand, if the FR is too small, there will be many redundant rebroadcasts since the transmissions of two successive forwarders overlap with each other. Therefore, we will focus on minimizing the expected total number of transmissions $E[NT]$. Since this involves knowledge of the MFR phase, we defer the derivations to Section 6.

5.2. Makeup for reliability

If a node u receives a WM from the kth forwarder and u is located in the one-hop zone created by the $k-1$th and kth forwarder, u will run through the distributed makeup selection process. A key concept in the MFR phase is the accumulated packet reception probability (APRP) of a node, which captures the idea that for any node u located in an one-hop zone, the probability of receiving a WM packet m increases as m is consecutively rebroadcasted multiple times by the relay nodes near u (the concrete definition of APRP will become clear in the following). The objective of the MFR phase is to ensure that the APRPs of all the nodes in each one-hop zone are larger than the given threshold P_{th}. If all the nodes' APRPs are larger than P_{th} in each one-hop zone, the PRR requirement of the network can be ensured. To minimize the number of makeups in each one-hop zone, the idea is to give the highest priority to a node whose rebroadcast can maximize the minimum APRP of all the nodes in the one-hop zone.

We first illustrate the intuition of the makeup selection algorithm using the scenario in Fig. 3. For the time being, assume for simplicity that the left forwarder F_L is the source of WM packet m. After the one-hop zone $Z_{0,0}$ is formed, we already have the left and right forwarders rebroadcasted. Since the packet reception probability decreases with distance, the middle of $Z_{0,0}$ is covered with the least APRP. Intuitively, selecting a node $M_{1,0}$ in the middle (or nearest to the middle) is most helpful to increase the minimum APRP for other nodes in $Z_{0,0}$. After $M_{1,0}$

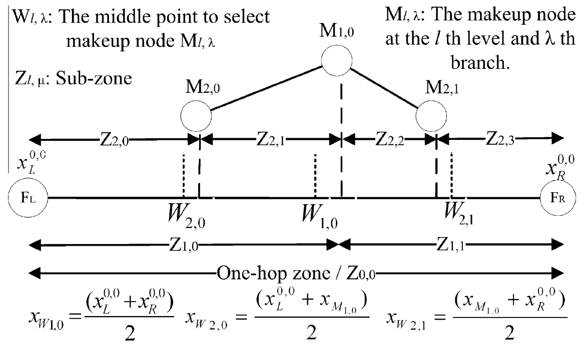

Fig. 3. A makeup selection tree, maximum level = 2.

broadcasts, it divides $Z_{0,0}$ into two sub-zones. Similarly, the middle points of these sub-zones have the least APRP, and again new makeups closest to the middle points are selected. This process is continued until the minimum APRP of all nodes in all sub-zones are larger than P_{th}.

Before we give a more rigorous treatment, we introduce some notations. The makeups form a binary tree, which is indexed by level ℓ and branch λ. A makeup is denoted as $M_{\ell,\lambda}$, $\lambda \in [0,\ldots,2^{\ell-1}-1]$. The depth of the tree is bounded by a maximum level.[2] At level ℓ, the makeups split the one-hop zone into 2^ℓ sub-zones, denoted as $Z_{\ell,\mu}$ (its number $\mu \in [0,\ldots,2^\ell-1]$). Each ℓth level sub-zone $Z_{\ell,\mu}$ is defined by scanning the one-hop zone from left to right, and assigning two consecutive relay nodes at level $l \leqslant \ell$ as its left and right boundaries $(x_L^{\ell,\mu}, x_R^{\ell,\mu})$. The one-hop zone is regarded as $Z_{0,0}$, which is bounded by $x_L^{0,0}$ and $x_R^{0,0}$, coordinates of the left and right forwarders (F_L and F_R). The right forwarder is regarded as the 0th level makeup.

Next we show how the APRP is defined and evaluated. Since a node u can hardly receive a rebroadcast packet from relays far away from it, we only consider contributions of rebroadcasts from specific nearby relays that are in the same one-hop zone with u. For a particular WM packet m and node u, we define a set of *locally visited nodes* by m, which consists of relay nodes on the tree branch leading to $u : \{F_L, F_R, M_{1,\lambda_1}, \ldots, M_{\ell,\lambda_\ell}\}$, where ℓ is the level of the newest sub-zone containing u. For example, if u is located in $Z_{2,1}$ which happens to be the newest sub-zone containing u, the locally visited nodes are $\{F_L, F_R, M_{1,0}, M_{2,0}\}$. This is a conservative estimation of APRP, since we have neglected the contributions of rebroadcasts from (possible) relays on sibling branches and those in other one-hop zones.

Upon receiving a WM m, each node u locally estimates the APRP of each neighbor node v within $Z_{\ell,\mu}$ iteratively, based on the locally visited nodes:

$$\begin{aligned} \xi_0(v) &= 1 - (1 - P_r(d(F_L, v)))(1 - P_r(d(F_R, v))), \\ \xi_\ell(v) &= 1 - \left(1 - P_r(d(M_{\ell,\lambda_\ell}, v))\right)(1 - \xi_{\ell-1}(v)), \end{aligned} \tag{1}$$

where $\xi_\ell(\,)$ denotes the ℓth iteration. If the minimum APRP: $\min_{v \in Z_{\ell,\mu}} \xi_\ell(v) \geqslant P_{\text{th}}$, then u knows the APRP requirement is satisfied. Otherwise, if u is in the makeup relay candidate

[2] This will not cause broadcast storm since the maximum level needed is small and bounded, and OBCF greatly reduces packet collisions.

region of M_{ℓ,λ_ℓ}, it becomes a makeup candidate and starts the OBCF according to its priority. The makeup RCR of M_{ℓ,λ_ℓ} is simply the sub-zones created by M_{ℓ,λ_ℓ}. For example, in Fig. 3, makeup RCR of F_R is $Z_{0,0}$, and that of $M_{1,0}$ consists of two sub-zones $Z_{1,0}$ and $Z_{1,1}$.

The priority of u reflects the minimum APRP of nodes in $Z_{\ell,\mu}$ after u rebroadcasts m, which is denoted as $\xi^*|u$. This can be calculated by doing another iteration on Eq. (1). For mathematical convenience, let us define the *APRP function*: $\Phi_{\ell,\mu}(x),\ x \in [x_L^{\ell,\mu}, x_R^{\ell,\mu}]$ over each sub-zone as a function of location coordinate x, which can be regarded as the APRP at location x given the rebroadcasts of the locally visited nodes $\{F_L, F_R, M_{1,\lambda_1}, \ldots, M_{\ell,\lambda_\ell}\}$. It is easy to see that for each node v, $\Phi_{\ell,\mu}(x_v) = \xi_\ell(v)$. Next, we claim that the priority of u decreases as the distance of u to the middle of the sub-zone which u is located in increases.

Proposition 1. *Function $\Phi_{\ell,\mu}(x)$ is concave. If it is symmetric w.r.t the middle point $W_{\ell+1,\mu}$ of sub-zone $Z_{\ell,\mu}$, then for any sequence of nodes $\{i_0, i_1, \ldots, i_n\}$ within $Z_{\ell,\mu}$ such that $d(i_0, W_{\ell+1,\mu}) < d(i_1, W_{\ell+1,\mu}) < \cdots < d(i_n, W_{\ell+1,\mu})$, we have*

$$\xi^*|W_{\ell+1,\mu} > \xi^*|i_0 > \cdots > \xi^*|i_n.$$

Proof. See Appendix A. □

Note that, the above optimality is derived under the assumption that $\Phi_{\ell,\mu}$ is a symmetric function. In reality, $\Phi_{0,0}$ is strictly symmetrical; with the level of broadcast increases, $\Phi_{\ell,\mu}(x)$ deviates from being symmetrical gradually because of the impact of the broadcasts of other lower-level relay nodes. However, the deviation degree is small if the level is small [38]. In practice, to satisfy 99% PRR, it is usually enough for the maximum level to be 2.

Remark. Although the makeup selection algorithm in the MFR phase may seem complex, it should not be interpreted as an overkill for the whole scheme. Rather, it is a necessary component in OppCast to achieve the desired PRR using a minimal number of rebroadcasts. We present it in the above way in order to include the general case. Indeed, the only parameter in the algorithm is the maximum makeup level, and the algorithm at each node is simple enough. For the reception probability $P_r(x)$, one can use a empirical model suitable for the VANET such as the one in [34]. Plus, it incurs lower transmission count than the straightforward strategy where each node contends to become a makeup node using a random priority. Moreover, nodes can make transmission decisions in an on-demand fashion to satisfy a desired PRR requirement, which is a feature not possessed by existing schemes in the literature.

5.3. Broadcast coordination in OppCast

In the following, we present OBCF, which is the underlying mechanism for selection of both forwarder and makeups. A broadcast coordination process (or contention process) is started immediately after each rebroadcast by a relay node. The primary goal is to let the relay candidates agree on the actual relay nodes in a distributed way, and for the selected relays to perform collision-free broadcast. The OBCF consists of the following components: (1) a process for the relay candidates to contend for the relaying opportunity; (2) a resource reservation mechanism to avoid collision and suppress hidden terminals; (3) a retransmission mechanism to prevent the WM from dying out. Its process is generally described as follows.

(i) When a node v receives a WM m for the first time (from node u), if v is in the RCR of u, v becomes a relay candidate. Then v sets a broadcast backoff timer (BBT) for m and calculates m's backoff delay. Also, v sets a *self allocation vector* (SAV) at MAC layer. The SAV suspends the transmission of other types of packets from v itself until there are no ongoing OBCF processes. This design provides packet-level priority access for WMs, since the WMs are safety-critical and have the highest priority in VANETs.

(ii) If v senses a busy signal from the physical layer, v will pause all its BBTs that are still counting down, in order to prevent collision and to keep its BBTs on the same page with that of other nodes. When the physical layer indicates idle again, v will resume all the paused BBTs.

(iii) If the BBT for m expires without receiving a broadcast acknowledgement (BACK), node v becomes a relay node for packet m, and sends a short MAC-layer BACK *at the base rate* to suppress other candidates. After the BACK has been transmitted, and after a short inter-frame space (SIFS), the WM is sent immediately at the data rate (higher than the base rate). The BBTs for other WMs are also paused during transmission.

(iv) If v receives a BACK for packet m from another node w before its own BBT expires, and if w is a relay node that contends for the same relaying opportunity with v, v will cancel the BBT. After that, v clears the SAV, and sets a *network allocation vector* (NAV) to reserve the time period for the WM that follows the BACK to suppress hidden terminals. Also, v pauses all of its own BBTs.

(v) The OBCF process for m finishes when the NAV for m expires, or v finishes broadcast of m as a relay node. The SAV of v is cleared only if there are no OBCF processes going on, or when a NAV is set.

(vi) Each source or forwarder F sets an additional, recurring *retransmission timer* (different from BBT) after transmitting m for the first time. This timer expires after every period of *MAX_WAIT_TIME* (the maximum delay of receiving a BACK from a forwarder, which is adaptively set; see below). This timer is only canceled when F receives a BACK that acknowledges the reception of m from a forwarder or makeup belonging to an one-hop zone with higher index than F. Otherwise, whenever this timer expires, F retransmits m until the maximum allowed number of retransmissions *MAX_RETX* is reached.

The time line of events are illustrated in Fig. 4.

A key element of OBCF is the delay function in BBT. A higher priority implies a smaller backoff delay. Observe

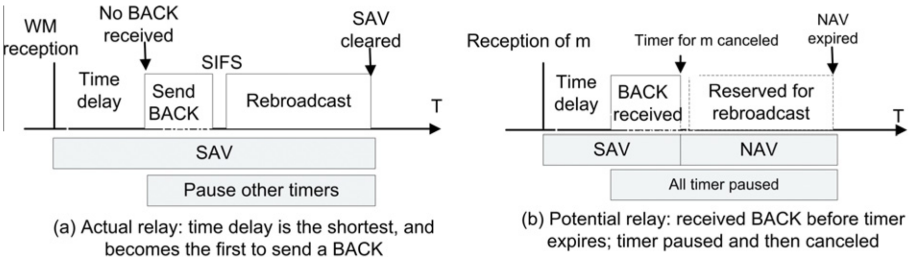

Fig. 4. Time domain illustration of OBCF.

that, for both types of relays, the RCR is a road segment, and the priority of nodes in a RCR increases/decreases monotonically from one end to the other. In the FFD phase, a node with a larger hop progress should have a higher priority in rebroadcast. In the MFR phase, a node closer to the center of a sub-zone (which is the boundary of RCR) should rebroadcast earlier. So in both cases the delay can be expressed by a function of the distance.

A straightforward way to set the timer is to let the delay be inversely proportional to the distance from the sender. However, two or more nodes that happen to be very close in space cannot be distinguished by this method. Thus, we propose an *enhanced slotted delay function*, where the RCR is divided into multiple equal-length spatial segments. The length of a segment adapts to the local vehicle density, which results in one node per segment on average. Each spatial segment corresponds to a time slot, where the central time of each slot increases linearly with the segment number, while a random jitter is used to separate potentially multiple nodes in the same segment to prevent collision. In this way, even if two nodes were very close spatially and were in the same segment, when one of them transmits first, the other node can hear it. Also, a guard time is placed between adjacent time slots that provides a minimum difference between backoff times for nodes in adjacent segments, in order to enforce the priorities of nodes in different segments. Although within the same segment, the nodes' priorities are not strictly followed, the impact of this reduces with the increase of traffic density since the segment length decreases.

Let x_I denote the boundary of RCR towards which the delay should increase, and x_D denote the boundary towards which the delay should decrease. For a node v located in u's RCR, v's backoff delay Δt_v used for its BBT is calculated as:

$$S_v = \left\lfloor \frac{|x_v - x_D|}{L} \right\rfloor, \quad L = 1000/\rho, \quad x_v \in [x_I, x_D], \tag{2}$$

$$\Delta t_v = \begin{cases} [S_v \cdot (T+\delta) + T \cdot Rand(0,1)], & x_v \in [x_I, x_D]; \\ \infty, & \text{otherwise}, \end{cases} \tag{3}$$

where S_v is the segment number of v, L is the segment length (ρ is the vehicle density in # of vehicles/km which can be estimated distributively), T is the maximum delay range of nodes in a segment, δ is a guard time which is used to separate two neighboring segments. Note that, by the above construction, $\Delta\, t_v$ is always small for an actual forwarder. This is because when the network is sparse, though the actual forwarder may not be close to the boundary of the RCR, each segment is long and the forwarder's segment number is small. When the network is dense, it is more probable that the forwarder locates in the first few segments.

In the above, the segment length is $L = 1000/\rho$. On average, in the distance of L, there will be 1 vehicle. The actual number of vehicles in each segment depends on the vehicle distribution; but since ρ will be locally estimated by each relay node (see Section 6.1.2), our method can effectively ensure a small number of vehicles in each segment.

To set the parameter *MAX_WAIT_TIME* in the retransmission timer, we estimate the upper bound on the time delay that a forwarder receives a BACK. Choose a d_{max} to be larger than the maximum possible RCR length, then according to Eq. (2), $S_v = \lfloor \frac{d_{max}}{L} \rfloor$. When no BACK are received from nodes in the segments with numbers smaller than S_v except the one equals to S_v, we have *MAX_WAIT_TIME* $\approx (T+\delta)\cdot S_v$. In this paper, we set $d_{max} = 1000$, thus *MAX_WAIT_TIME* $= (T+\delta)\cdot\rho$.

The OBCF has several advantages. First, the redundant transmissions are eliminated more effectively. Because the BACK is transmitted at the base rate, for which the threshold of received signal to interference and noise ratio (SINR) is lower at a receiver than using the data rate, it can be received by most of the relay candidates. Second, the one-hop delay is small. This is because (1) in OBCF a node pauses its timers when its detects a busy channel which prevents a collision. (2) In the relatively rare case that a node is in a nearby segment with a relay but does not hear the latter's BACK: since the BACK is very short (its transmission takes around 80 μs when the payload length is 14 bytes), choosing $T+\delta$ to be larger than the BACK transmission time can prevent most of the BACK collisions from happening. In our simulations, we found that $T = 80$ μs and $\delta = 20$ μs are good values. Third, BACK is also used to suppress the hidden terminals. As illustrated in Fig. 5, the transmission range of BACK is larger than twice that of the WM, which means most of the hidden terminals to the WMs are avoided.

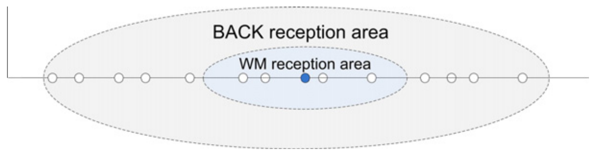

Fig. 5. BACK suppresses hidden terminals.

5.4. Extension to disconnected VANET

When the VANET is sparse, for example at nights or when the penetration rate is low, the VANET tends to be disconnected [39] and the WM cannot be propagated through the whole IR. Thus we need additional mechanisms to ensure the reliable reception of the WM by vehicles in the interested region. To this end, we extend OppCast to handle this situation, and still assume the WM is disseminating from west to the east.

Like many routing schemes for disconnected VANETs [32,39], for WM broadcast in this paper we also take advantage of bi-directional mobility, i.e., utilize the vehicles driving in opposite direction to store-carry-and-forward the WM packets. While the idea is simple, for OppCast, we need to carefully design the protocol so that the three previously proposed objectives are still satisfied. Thus, several questions arise: (1) who and when to use the store-carry-and-forward method? (2) how to ensure the WM reception reliability in the network when store-carry-and-forward is adopted? (3) how should the parameters in the protocol be adjusted under sparse VANETs? (4) can we still preserve the advantages of using BACKs?

To answer these questions, we allow the protocol to switch adaptively between the normal dissemination mode (as previous described) and the store-carry-and-forward mode, according to local network topology. The straightforward solution is that, in "connected" parts such as vehicle clusters a message should propagate fast until it reaches the end of a cluster; while the end vehicle in a cluster performs store-carry-forward to enhance reliability (PRR); and this is also the idea in existing schemes [29,27]. However, it can hardly guarantee the desired level of reliability, because this strategy neglects the differences in the local traffic densities. The normal mode is adopted even if there are only a few (>0) vehicles around a relay node, where the relay's rebroadcast may not be heard by any vehicles in the message direction, which results in a stop of message propagation.

Thus in this paper, in addition to using the straightforward store-carry-forward condition, we propose to employ local traffic density as a decision factor, i.e., a relay node u switches to store-carry-forward mode whenever its local traffic density ($\hat{\rho}(u)$) is smaller than a threshold. To implement this function, a *carry_flag* is used which is set to 1 if $\hat{\rho}(u) < \rho_{\text{th}}$, and will be set back to 0 if u's rebroadcast is received by vehicles in the message direction. Node u learns about others' message receptions via BACKs sent before their rebroadcasts. The decision tree for OppCast-Ext is depicted in Fig. 6. Note that, vehicles within the IR will only follow the diagram when they receive a new packet, while vehicles outside the IR do not ignore duplicate packets. The reason for this is similar to that in [27].

For each node v, its locally estimated traffic density is computed as:

$$\widehat{\rho(v)} = \frac{\text{Number of neighboring vehicles within range } CR \text{ in the message direction}}{CR}, \quad (4)$$

where CR is the *communication range* of the WM, defined equivalently (for the same transmission power) under the two-ray ground propagation model. We only count the neighbors in the message direction, since the FFD of WM is mainly impacted by the density of those vehicles.

In the following, we illustrate the detailed forwarding rules using a simple example.

(i) We assume the source vehicle is west-bound in the following. For any forwarder (or source) F driving towards west, upon receiving a WM m for the first time, if its locally estimated traffic density ($\widehat{\rho(F)}$) is smaller than a *threshold density* ρ_{th} or there are no vehicles in the message direction (east), it will carry the packet and set a *carry_flag* = 1. (For example, the node A in Fig. 7(a)). Only if F receives a BACK from a relay in the east will it set *carry_flag* = 0. If F has not received such BACK after it retransmits m *MAX_RETX*

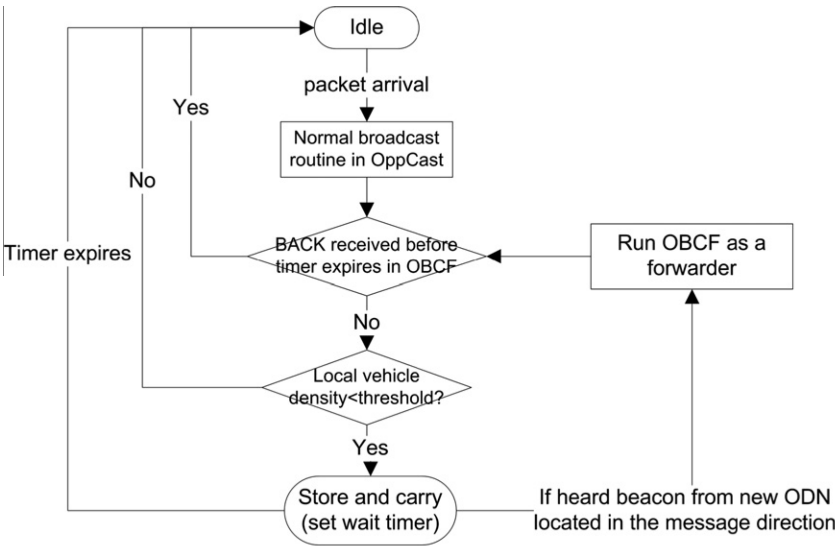

Fig. 6. The decision tree for OppCast-Ext.

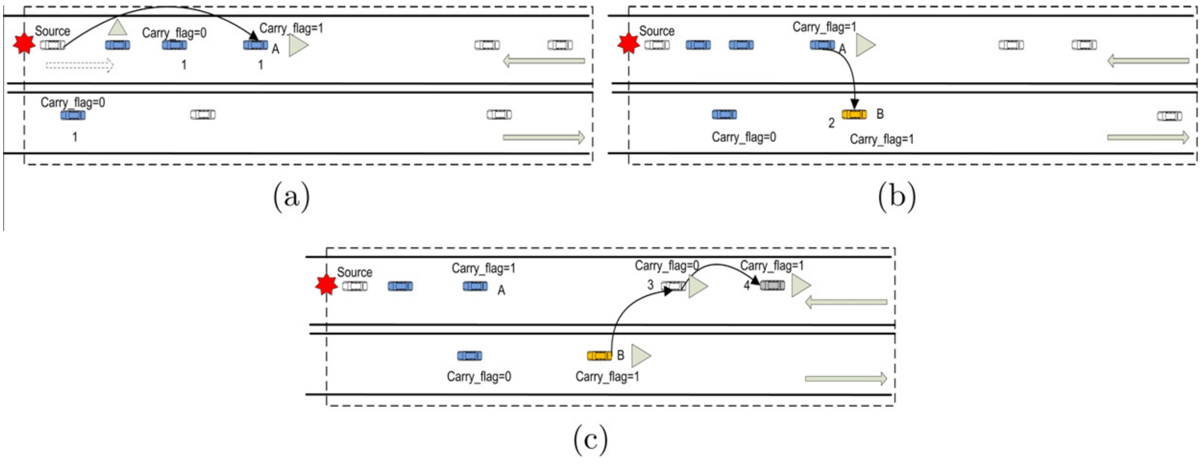

Fig. 7. Dissemination process of a WM in the OppCast-extension for the sparse VANET. Legends are the same with Fig. 1.

times, it will store the packet. Later when F receives a beacon message from a vehicle driving opposite and also located in the east of F, if *carry_flag* = 1, a rebroadcast of m will be triggered and exactly the same OBCF is invoked (Fig. 7(b)). Note that, the other node is very close to F since the beacon rate is usually high (e.g., 10 beacons/s), thus the single-hop reliability is very high in this case.

(ii) For any vehicle V driving towards east, upon receiving WM m from a forwarder F, if F's local traffic density $\widehat{\rho(F)} < \rho_{\rm th}$ (contained in m), V will temporarily set *carry_flag* = 1. No matter if *carry_flag* = 1 or 0, if the number of neighbors in the east is larger than 0, V participates in relay selection as usual. This is to let the WM propagate to the end of a "cluster" as fast as it could and to reduce the total transmission count. If *carry_flag* = 1, only when V later receives a BACK from another relay in the east and also driving towards east, V will set *carry_flag* = 0. Otherwise, if sometime later V receives a beacon message from a node in the east but driving towards the west, FFD will be triggered at V according to OBCF (Fig. 7(c)). If V is the head of a cluster driving east, V will always carry the WM, and thus reliability is ensured with high probability.

In the above way, the WM can be still disseminated with high speed in connected clusters, where "connected" should be interpreted in probabilistic sense, under opportunistic forwarding and retransmissions. For "disconnected" parts, the opposite directional vehicles act as data mules. Here we need to determine a suitable ρ_{th}. If ρ_{th} is too small, a WM may not be able to get through the network which affects the PRR; if ρ_{th} is too large, there will be redundant rebroadcasts by data mules. Therefore, we find out the optimal ρ_{th} in Section 6.

5.5. Implementation issues

We first give the definition of "neighbors" adopted in the implementation. For a node v to become a neighbor of node u, u needs to receive at least one beacon message from v recently within a reasonable amount of time, such as 5 times the beacon broadcast interval.

In addition, we need to ensure that the receivers of each (re)broadcast agree on an unified RCR to determine whether to participate in contention processes. Therefore, we include the locally calculated (optimal) forwarding range (see Section 6) of the sender in every broadcast packet from a forwarder/source, and each node that receives the packet uses the received value in the packet instead of its own. For the additional makeups to decide their RCRs, they simply rely on the locations of previous visited nodes piggybacked in the rebroadcast packets.

Next, we present the format of the WM packet header in Fig. 8. The first $2\ldots 19$ bytes are used for FFD phases, and $20\ldots 4k+29$ bytes are used for MFR phases, where k is the maximum level of makeups in a one-hop zone. The fields x_{left} and x_{right} stand for the boundary locations of the current one-hop zone or sub-zone. The list of locations of visited nodes in current one-hop zone include the left and right forwarders, and visited makeups; it is used in MFR to calculate the APRP. Although when $k = 2$ the header length is 37B, it is relatively small compared to the message length, which is usually more than 200B.

In the BACK messages, header includes few information: the one-hop zone index, makeup level, and the sub-zone index. The one-hop zone index increases by one when the WM traverses another forwarder, and upon receiving a BACK, a node can determine whether to cancel its own rebroadcast. In practice, in order to enhance the reliability of BACK to suppress redundant transmissions, an additional rule is used. The BACK from a forwarder with one-hop zone index $k + i$, $i \geqslant 0$ suppresses the forwarder candidates in the kth one-hop zone, and a higher level makeup suppresses lower level ones in the same one-hop zone.

For interested regions that contain multiple road segments, OppCast can be easily extended to handle that. For example, at an intersection, we can let the vehicle closest to the intersection act as a temporary source for the WM it has received, and each WM needs to include from

Fig. 8. The WM packet header format.

which road segment and which direction it came, in order to avoid loops. If the VANET is sparse and at some time there is no vehicle close to a road intersection, each vehicle at the end of a cluster/group can store-carry the WMs and rebroadcast them when it reaches the intersection. Within each road segment, the same OppCast protocol is run.

5.5.1. Protocol exceptions

OppCast is tolerant with the exception that a BACK of a forwarder is not received by some nodes within an one-hop zone. Assume some node u within an one-hop zone rebroadcasts after the primary forwarder v, there is a chance for some other node w within the next one-hop zone to receive the WM from u rather than from v. The nodes that have already received v's rebroadcast will ignore that of u's since they are in the IR. Then, w who is not aware of the v's broadcast, will participate in the contention process started by u's transmission. According to OBCF, during the timer countdown, w will listen to the channel and avoid collision with possible rebroadcasts from candidate forwarders in its one-hop zone. That is, multiple FFD phases can indeed coexist for the same WM.

However, it is not easy for a node to distinguish which FFD phase it is in, since node w above is not aware of v's broadcast (the other, "primary" FFD phase). Thus, some nodes will mistakenly estimate the accumulated packet reception probability within the "secondary" one-hop zone, and induce further unnecessary makeup selections which should not have happened, since it might already been covered by the primary one-hop zone. With those less well coordinated redundant transmissions, packet collisions may ultimately happen which degrades the reliability. Therefore, it is still essential in OppCast to ensuring the reliability of BACK, for the sake of preventing the transmission count from becoming uncontrolled.

6. Parameter optimization

In this section, we optimize two parameters in OppCast: the forwarding range FR and threshold density ρ_{th}.

6.1. Optimize the forwarding range

As mentioned in Section 5.1, the goal here is to minimize the expected total number of transmissions $E[NT]$. Stated formally, we want to find the FR,

$$Min\ E[NT], \tag{5}$$

$$s.t.\quad \forall v \in G, \quad PRR \geqslant P_{\text{th}}. \tag{6}$$

The optimization is targeted at the well-connected case (vehicle density is larger than ρ_{th}), whereas it is straightforward to show that under the strategies of the FFD and MFR phases, the constraint is satisfied with high probability.

Thus, we first compute the expected number of transmissions ($E[NT]$), based on which the optimal FR is derived. We first introduce the centralized solution to find the optimal FR, and then propose a distributed, locally optimized version. The centralized solution takes as input the average vehicle density ρ of IR, and approximates the $E[NT]$. Since $E[NT]$ has no closed form expression, the optimal FR that minimizes $E[NT]$ is sought out by sampling and searching.

Let us consider a one-dimensional VANET where the IR length $\mathscr{L}$ is sufficiently large. Assume there are no redundant transmissions and no packet collisions. Further, we assume there are enough relay candidates so that the PRP requirement can always be satisfied. Finally, the Rayleigh fading model is used for pairwise PRP function: $P_r(d) = exp\left(-\frac{P_{rxth}}{P_{ref}} d^2\right)$, where P_{rxth} is the reception threshold power, P_{ref} is the reference receive power at distance 1m by free space propagation model.

The total number of transmissions NT can be expressed as:

$$NT = \sum_{i=1}^{X} (M_i + \omega_i), \tag{7}$$

$$\sum_{i=1}^{X} Y_i = \mathscr{L}, \tag{8}$$

where X is the number of one-hop zones, and M_i is the number of makeups in the ith one-hop zone. Y_i is the length of the ith one-hop zone, and $M = M(Y)$ is a single-variable function of Y, while Y is related to both FR and ρ. ω is the number of retransmissions made by the ith forwarder. Since Y_i are i.i.d. random variables, X is also a random variable.

Therefore, we have

$$\begin{aligned}\mathscr{L} &= E\left[\sum_{i=1}^{X} Y_i\right] = E_X\left[E_{Y|X}\left[\sum_{i=1}^{X} Y_i|X\right]\right] \\ &= E[Y_1|X=1]P(X=1) + E[Y_1+Y_2|X=2]P(X=2) + \cdots\end{aligned} \tag{9}$$

For an approximation, we neglect the dependence between X and Y_i (i.e. $E\left[\sum_{i=1}^{X} Y_i|X=i\right] \approx E\left[\sum_{i=1}^{X} Y_i\right] = i \cdot E[Y]$), then

$$\mathscr{L} = E[X] \cdot E[Y], \tag{10}$$

and thus

$$E[NT] = E[X](E[M] + E[\omega]), \tag{11}$$

where $E[X]$ is the average number of one-hop zones, $E[M]$ is the average number of makeups in each one-hop zone, $E[Y]$

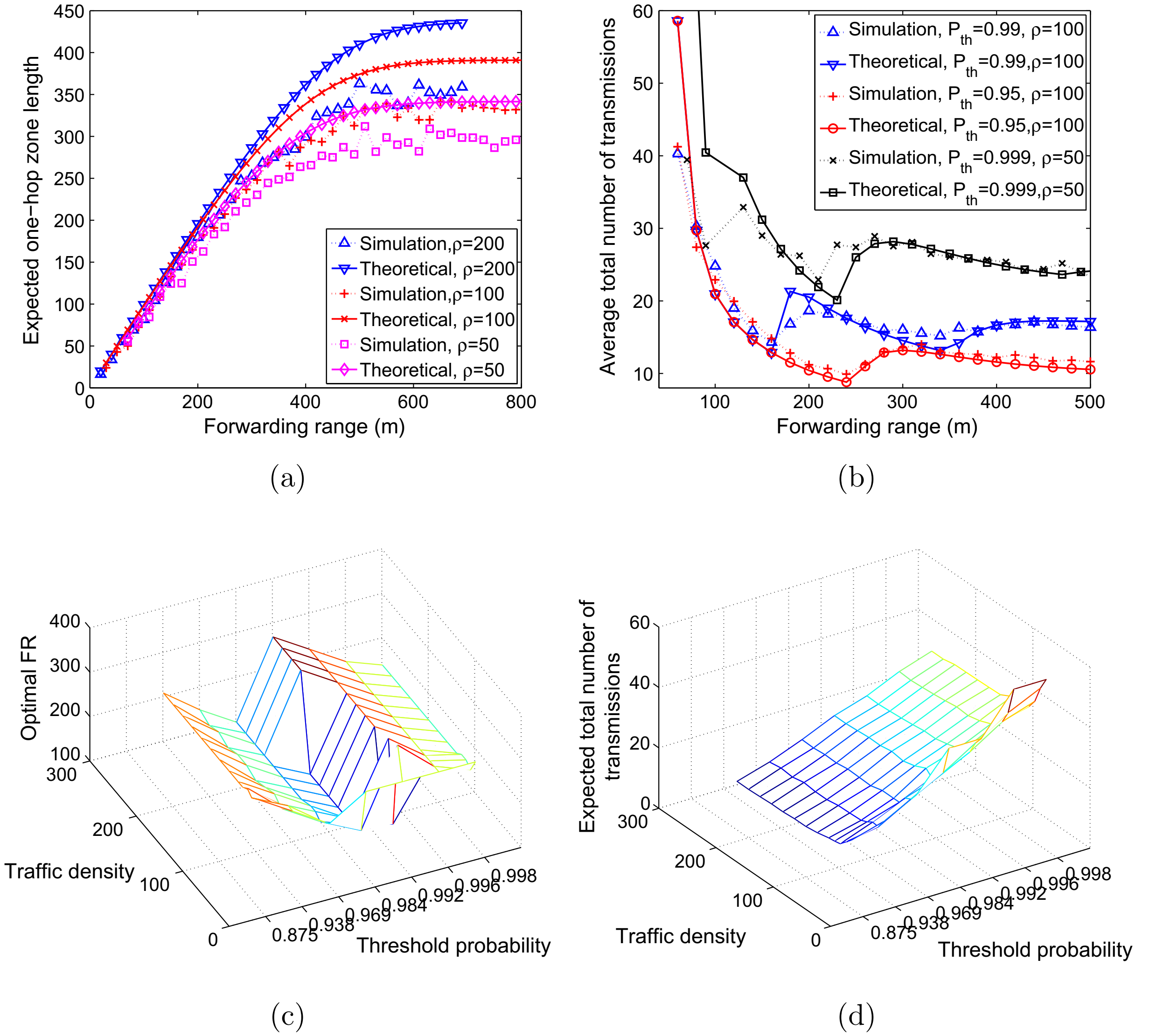

Fig. 9. (a) Numerical validation of $E[Y]$. (b) Numerical validation of $E[NT]$, $\mathscr{L} = 2$ km. (c) Optimal forwarding range, $\mathscr{L} = 5$ km. (d) Expected total number of transmissions, $\mathscr{L} = 5$ km. $CR = 250$ m in all figures.

is the average one-hop zone length, $E[\omega]$ is the expected retransmission count of each forwarder.

We then approximate $E[Y]$ and $E[M]$ by fixing the inter-space between successive vehicles to[3] $L = 1000/\rho$.

$$E[Y] = \sum_{k=1}^{N} kL \cdot P_F(k, L), \quad N = \left\lfloor \frac{FR}{L} \right\rfloor, \tag{12}$$

where $P_F(k,L) = P_r(kL)\prod_{j=k+1}^{N}(1 - P_r(jL))$. From Fig. 9(a), we can see the above equation yields a good approximation to the average one-hop zone length. Similarly,

$$E[M] = \sum_{k=1}^{N} M(kL) \cdot P_F(k, L), \tag{13}$$

where $M(kL)$ is the number of makeups needed in an one-hop zone of length kL, under the ideal case where each makeup locates in the middle of its parent's sub-zone.

For each forwarder, the expected number of retransmissions to be made is:

$$E[\omega]' = \frac{1}{1 - \prod_{i=1}^{\lfloor \frac{FR}{L} \rfloor}[1 - P_r(iL)]}, \tag{14}$$

and $E[\omega] = min\{E[\omega]', MAX\_RETX\}$.

Finally, the expected total number of transmissions is obtained by Eq. (11). $E[NT]$ is a function of both FR and ρ; however, it has no closed form solution. Under a fixed ρ, the optimal FR that minimizes $E[NT]$ can be obtained by searching FR from L to R_c (e.g. 500 m), by setting the sampling interval to a small enough value, e.g., 10 m.

[3] The uniform distribution of vehicle positions is adopted in performance evaluation.

6.1.1. Theoretical insights

First, we carry out simulations to verify the above results. An idealized version of the protocol (referred to as IDEAL) is implemented in NS2, where the BACK can be reliably received by all nodes in the network. The averaged vehicle density in IR is adopted as an input in IDEAL.

Fig. 9(b) compares the theoretical value of $E[NT]$ to the average number of transmissions in IDEAL. The theoretical values are close to the simulated values for all the vehicle densities and P_{th} shown, and the same for the optimal points of *FR*.

Interestingly, the optimal *FR* also exhibits an opportunistic behavior, depending on the required PRR and vehicle density. In Fig. 9(c), the optimal *FR* increases and decreases recurringly as the P_{th} increases. The reason is twofold. (1) On the one hand, using some particular "longer hops" reduces the number of transmissions. Note that the $E[NT] \sim E[Y]$ function has multiple local minimas.[4] Using a farther minimal point not only reduces the number of hops, but also contributes to the APRP of the other nodes. (2) On the other hand, the longer a hop is, the less possible it is for a WM to reach that far. The $E[Y]$ is upper-bounded when ρ is fixed.

For example, when $CR = 250$ m, $E[Y]$ is always less than 350 m. For $P_{th} = 0.95$, the first two local minimal points are $E[Y] = 220$ and 450, which implies the *FR* corresponding to the first one is optimal. Therefore, the above results indicate that when the network is well-connected, *the best strategy is try to opportunistically use long hops, but only when that long hop is feasible to reach statistically*.

Finally, in Fig. 9(d) the average total number of retransmissions increase linearly when the threshold probability increases inverse exponentially towards 1. This reveals the intrinsic tradeoff between the desired WM reception ratio and transmission count, which is a helpful result for WM broadcast applications in VANETs.

6.1.2. Distributed algorithm

In OppCast, a distributed algorithm is used to set the *FR* since global vehicle density information is not available. Each node calculates its own optimal *FR* based on the local vehicle density estimated from its direct neighbor nodes' locations. A node is considered to be a neighbor as long as a beacon is heard from it within 1 s. Each forwarder estimates its local traffic density $\widehat{\rho(v)}$, based on which a local optimal *FR* is derived and included in every rebroadcast packet, and every vehicle that receives it use the same *FR* included in the packet. In addition, an upper limit (e.g., 1000 m) is imposed on the *FR* to prevent the hop delay from being too large.

Note that, in the OppCast extension to the sparse VANET, due to short optimal *FR* at some traffic densities, while the vehicles received the WM may all be located outside the *FR*, there is a small chance of forwarder shortage. To deal with this situation, we let each forwarder increase its *FR* by 2× after it retransmits the same packet another time, until reaching the limit.

[4] Because as $E[Y]$ increases beyond these local minima points, number of makeups per hop will first increase and then remain fixed.

6.1.3. Discussion

Our optimization is carried out using the equal interdistance vehicle distribution model (regularly distributed). However, through Fig. 9(a) and (b), it can be seen that the results is not so sensitive to the uniform vehicle distribution, which is a common mobility model adopted in most previous works. In fact, the variations in vehicle densities in the regular and uniform models are both small. In reality, some vehicles may travel in platoons; one may wonder if such mobility pattern would affect the effectiveness of optimization. Let us imagine a well-connected platoon of length 1km; the vehicles in it can often be regarded as nearly regularly distributed. Since the distributed algorithm is based on local traffic density estimation, and the range of "locality" is really restricted to $CR = 250$ m (Eq. (4)), the algorithm is expected to work well for vehicles within the platoon. Near the boundaries of the platoon, if the local density falls below ρ_{th}, the vehicle density experiences large variation; store-carry-forward will be used where the optimization does not play an important role. If the local density is larger than ρ_{th}, the variations in density is relatively small and our algorithm still applies well.

6.2. Optimal threshold density

Above a certain threshold density, the network is connected and the PRR requirement can be satisfied, but a higher threshold is unnecessary which incurs redundant transmissions. Thus, we first calculate the probability that the VANET is connected, which means successive forwarders can be selected to propagate the WM towards the end of IR. Still using the simplified model in the above, for a given ρ, the probability that a forwarder is selected for one hop equals

$$P_F = \sum_{k=1}^{N} P_F(k, L), \quad N = \left\lfloor \frac{FR_{opt}}{L} \right\rfloor, \tag{15}$$

where FR_{opt} is the optimal FR. The expected number of hops equals $\frac{\mathscr{L}}{E[Y]}$, where $E[Y]$ is computed from Eq. (12), by substituting *FR* with FR_{opt}. Thus

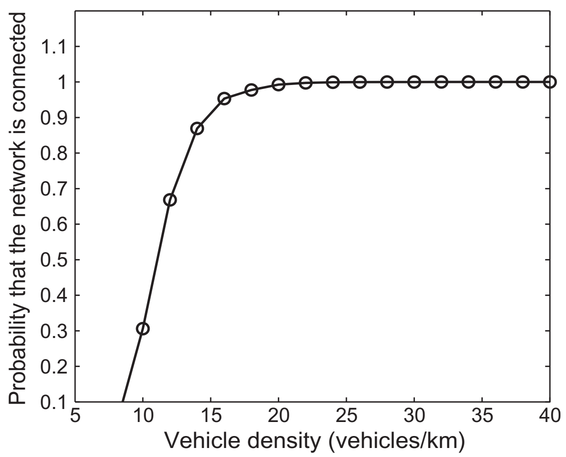

Fig. 10. Probability that the VANET is connected, $CR = 250$ m, $\mathscr{L} = 5$ km.

$$P_{connect} = P_F^{\frac{\mathcal{L}}{E[Y]}}. \tag{16}$$

The result is shown in Fig. 10. It can be seen that when P_{th} = 0.95, the optimal ρ_{th} is between 15 and 20.

7. Performance evaluation

In this section, we evaluate the performance of OppCast. The compared protocols are as follows.

- Slotted p-persistence broadcast [13] (*Slotted-p*). Upon receiving a packet from j, a node i rebroadcasts the packet with a fixed probability q after the backoff delay T_{ij}, if it receives the WM packet for the first time and has not received any duplicates during the delay. Otherwise, it drops the packet. The delay-distance function is slotted and linear. *Slotted-p* is shown to be the best among the probability-based protocols [13]. We set τ = 5 ms, N_S = 5, q = 0.5 (the settings used in [13]) and the forwarding range R = CR = 250 m in the simulations.
- Contention based dissemination (CBD) [16], a typical broadcast protocol also based on opportunistic forwarding. It does not differentiate between relay nodes, and uses WM as implicit ACKs. A node in the forwarding range will set a backoff timer upon receiving a WM for the first time; it cancels the timer only if it receives duplicate WMs during the backoff process, otherwise it rebroadcasts. The delay-distance function is continuous and linear. We set the maximum backoff delay to be 10 ms, which is below the value (50 ms) adopted by [16] (since our CR = 250 which is smaller than the one used in [16], and the channel tends to be less congested). Also, we set R = CR = 250 m.

Meanwhile, the IDEAL protocol is also compared, which can be regarded as a lower-bound to the transmission overhead since it has no collisions and redundant transmissions. The proposed protocols are named by appending the threshold PRR to the protocol type, e.g., for OppCast95, P_{th} = 95%.

7.1. Simulation setup

OppCast and its extension (OppCast-Ext) is implemented in NS-2.33 [40], which supports probabilistic propagation models. The parameters are summarized in Table 2. The other PHY and MAC layer parameters follow the default settings of IEEE 802.11p. The Rayleigh fading model is used, which is a special case of the Nakagami model (m = 1).

For the vehicle mobility, we use the USC VANET mobility generator [41] to generate the movement patterns. Vehicles are placed uniformly at random in the road area; when a vehicle hits the freeway's boundary it randomly selects the other end as its new entry point of the map, which removes the boundary effect. Also, the initial velocities of the cars are chosen uniformly from 20 to 30 m/s, and they can accelerate with acceleration taken randomly from −2 to 2 m/s^2. No overtaking is allowed currently.

Table 2
Parameter settings.

Maximum time slot length, guard time	80 µs, 20 µs
CR for WM and BACK	250 m, 628 m
Transmission rates for WM and BACK	12 Mbps, 3 Mbps
Tx power, CSThresh, Noise floor	10, −96, −98dBm
WM, Beacon and BACK length	292, 72, 14 bytes
MAX_RETX	3
Vehicle distribution	Uniformly random
Range of global vehicle density	5–200 cars/km
Vehicle speed	Randomly sampled from 72–108 km/h
Road length, IR length	6 km, 4–5 km (2 lanes/direction)
Maximum makeup level	2

For the data traffic model, every vehicle periodically generates 10 beacon messages every second for routine safety applications. The beacons have the same *CR* and transmission rates with WMs. Furthermore, each (west bound) vehicle located between 1 km and 2 km randomly generates urgent event-driven WMs at an average messaging rate of r packets/s, according to the Poisson arrival model.

Each WM's IR is chosen to be the road segment between its source and the east end of the highway. Three sets of experiments are conducted. In the first one, the traffic density ρ ranges from 20 to 200 vehicles/km while the messaging rate r is fixed to 0.1 packets/s. For the second, ρ is fixed to 80 while r is varied from 0.01 to 10 packets/s, and for the third, ρ changes from 5 to 60. In the third case, some vehicles in the opposite driving direction will definitely not receive the WM due to disconnection, so we set the IR to include only the two co-directional lanes. Each simulation run lasts for 10–200 s, and a random scenario is generated where vehicles are uniformly distributed. Fig. 11 shows the results where each point is averaged from 5 repetitive runs with different topologies generated using different seeds, and the error bars indicate 95% confidence interval.

7.2. Results for OppCast without extension

7.2.1. WM reception ratio

We first fix r = 0.1, and change ρ. In Fig. 11(a), when ρ = 60 ∼ 200, OppCast99 maintains average PRR of above 99%, and that of OppCast95 is higher than 98%. This shows OppCast indeed satisfies the PRR requirement when the network is well connected. The average PRR turns out to be higher than the thresholds, because the PRR requirement is taken as a minimum requirement in each MFR phase. When the network is sparse, i.e., ρ = 20 ∼ 50, the PRR of OppCast protocols is still higher than 90%, which is much higher than Slotted-p and CBD. The advantage is primarily because of the FFD phase trying to guarantee the forwarders span the whole network. The PRR in this case is lower than required, since there may not be enough relay nodes due to network partition.

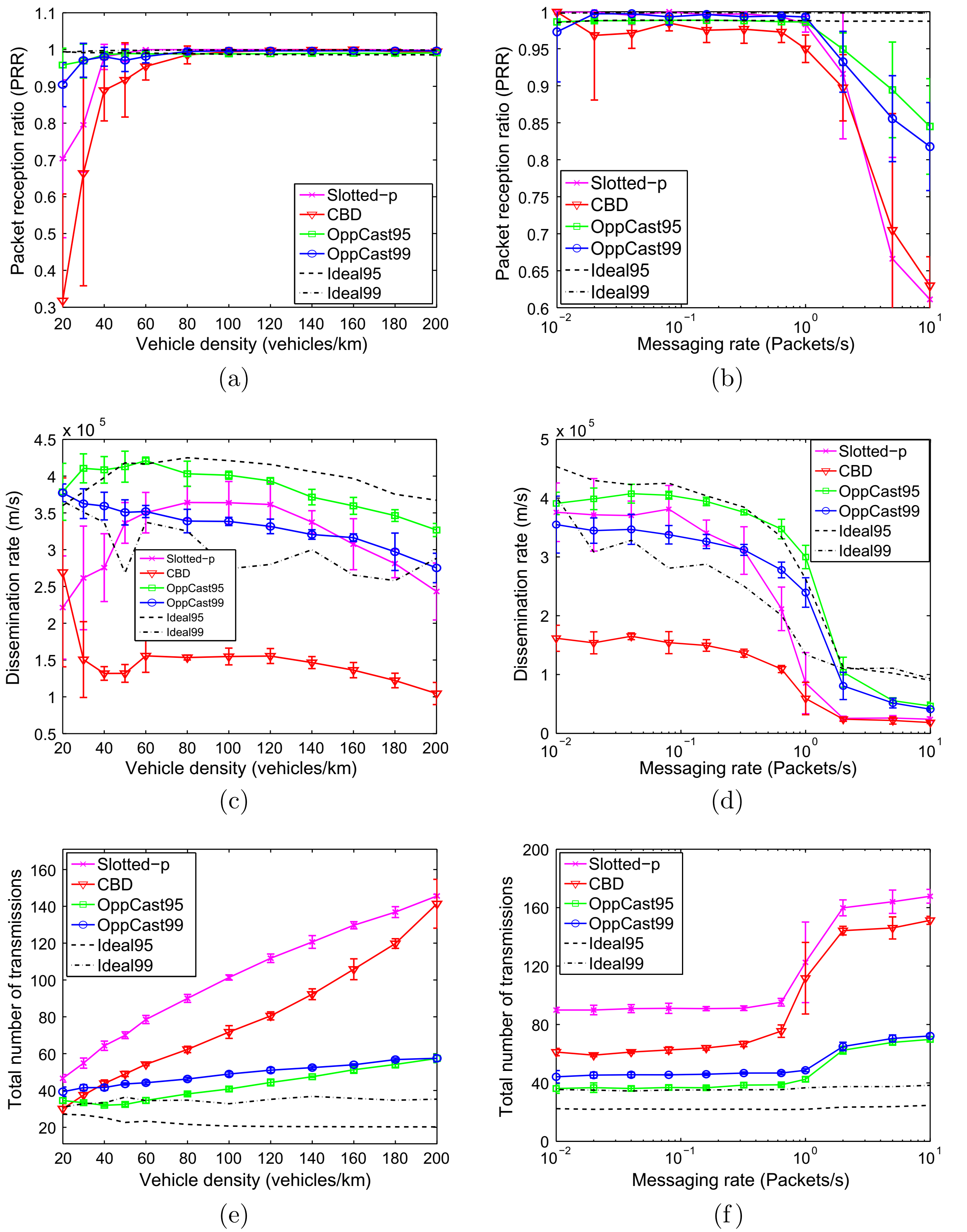

Fig. 11. Simulation results (OppCast without extension). (a), (c) and (e): fix $r = 0.1$, change ρ. (b), (d) and (f): fix $\rho = 80$, change r.

Fig. 11(b) shows the PRR results of the second experiment. It can be seen that the PRR requirement in OppCast is always satisfied when r is small to moderate. The decrease of PRR only happens when message generation is very dense, i.e., $r > 1$. However, the PRR of OppCast is still much higher than Slotted-p and CBD in this case, while OppCast introduces much less overhead. Similar results can be observed for the dissemination rate. This shows OppCast is more scalable, i.e., more capable of handling saturated message traffic situations than other protocols.

7.2.2. Dissemination rate and delay

From Fig. 11(c), it can be seen that the dissemination rate of OppCast95 is the highest except for IDEAL95, for all the vehicle densities shown. Similar results is shown in Fig. 11(d), where OppCast95's dissemination rate is still among the highest for all the messaging rates. This can be mainly attributed to the opportunistic forwarding concept adopted in OBCF, which always utilize the farthest forwarder candidate so that the one-hop delay is minimized. On the other hand, for OppCast99, although the achieved reliability is a little higher than OppCast95, the dissemination rate is smaller. It turns out that the reduced dissemination rate is a cost to enhance the WM reception ratio in OppCast.

To further investigate the dynamics of WM dissemination in OppCast and see why it performs better, we show in Fig. 12 the end-to-end delay results of each vehicle in the IR correlated with its distance to the source, for a typical WM disseminated in VANET with traffic density equal to 80 vehicles/km (well-connected). Remarkably, the last vehicle in the IR receives the WM within about 12 ms, which is much less than that of the CBD and slotted-p. Furthermore, the delay-distance curve increases smoothly showing that there is little gap between reception times of successive rebroadcasts (read from the y-axis). This shows the effectiveness of the carefully designed coordination mechanism for relay selection (OBCF), where average hop-delays in the order of 10–100 μs can indeed be achieved. Moreover, although we used makeup nodes so that nodes that missed the WM in the FFD phase can receive it later, the introduced delay variance is negligible.

In contrast, other protocols do not enjoy the same level of fast propagation. For CBD, there are obvious time gaps between consecutive rebroadcasts, which is partially due

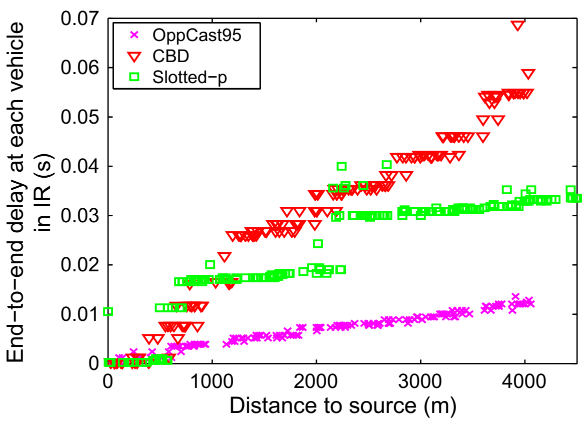

Fig. 12. The end-to-end delay of each vehicle in the IR vs. distance to the source, $\rho = 80$, $r = 0.1$.

to its large backoff delay in relay contention processes. However, the maximum backoff delay (10 ms) is already much less than the adopted value in [16]. By studying Fig. 12 in more details, one can observe that some time gaps are relatively long, and many next hop relays are located near their previous hops. This suggests that during relay contention process in CBD, due to channel fading and poor coordination, packet collisions happen more frequently, resulting in sub-optimal relays being selected. While in OppCast, the FFD phase is employed to propagate the WM towards the end of IR in the first place. To guarantee this we use the BACK, by which the channel is cleared before each rebroadcast, and candidate relays in each newly traversed one-hop zone suspend their counting down timers during BACK to give priority to the forwarding WM. Slotted-p is somewhat different, in that the continuous propagation periods are longer than that of CBD, however the gaps are even larger. The former is naturally due to zero delay for relays outside of the contention region; but the latter indicates that packet collisions are even worse. This is mainly because Slotted-p is still controlled flooding; though using a coarse-grained slotted timer function, it cannot completely eliminate the broadcast storm. The above results show that, redundant transmissions indeed undermine broadcast performance to a large extent, and the explicit BACK mechanism in OBCF is effective and necessary.

7.2.3. Transmission overhead

The transmission overhead is evaluated by the total number of WM packet rebroadcasts incurred per WM sent by the source. Since the length of a BACK is quite small compared with a WM packet, we neglect the overhead caused by BACKs. In Fig. 11(e), as vehicle density increases to 200, the total number of transmissions incurred by OppCast95 and OppCast99 is about 40% of that of CBD. More importantly, the overhead increases slower with respect to vehicle density than in Slotted-p and CBD, because the relay selection mechanisms are optimized, and the OBCF is effective in reducing redundant transmissions and packet collisions under the presence of lossy links. In CBD, because of channel fading the rebroadcast of relays cannot be heard by many other relay candidates, which leads to large amount of redundant transmissions. On the other hand, in OppCast, using BACK we can exert more fine-control over the selection of makeups, which turns out to be less than 3 per one-hop zone. The above indicates that the high reliability and fast dissemination are achieved in a resource-efficient way in OppCast.

7.2.4. The tradeoffs

The OppCast95 achieves competitively high PRR and the highest dissemination rate using the smallest number of transmissions. The OppCast99 achieves higher PRR than OppCast95 in most scenarios, but uses more transmissions and leads to slower dissemination. Since we take into multiple objectives in designing OppCast, this reflects the fundamental tradeoff between them: to achieve higher reliability, more transmissions are needed, which in turn causes larger broadcast latency. Furthermore, when the PRR is already close to 1, a marginal gain in PRR would demand noticeably more transmissions, and will result in a big decrease in the dissemination rate, as is in the case of OppCast99. *Thus, using a lower PRR goal, such as 95% is better than 99% in this sense*.

On the other hand, the Slotted-p exploits a different tradeoff: use aggressive rebroadcasts to achieve high reliability and relatively high dissemination rate. However, this is not very resource-efficient, since it consumes a much larger portion of the VANET bandwidth. Also, too many transmissions adversely affect the dissemination rate, as one can see from Figs. 11 and 12.

Note that, in our comparisons, we have not extended the Slotted-p and CBD to allow a forwarder perform multiple retransmissions as is the case in OppCast. This could be done to enhance their PRRs under the disconnected case; however, the gain is very small when the network is well-connected, and it results in even more transmission overhead.

7.2.5. How reliable is the BACK?

Next, we investigate deeper about the reliability of BACK in OppCast, and discuss how the broadcast performance will be affected by BACK. Ideally, BACK should achieve three goals: (1) acknowledge the transfer of relaying opportunity and suppress all redundant relay candidates; (2) inform the previous forwarder to cancel retransmission; (3) suppress hidden terminals to reserve the channel for WM broadcast. In IDEAL protocol, all these goals are achieved perfectly. But in reality, BACK is still subject to losses. This comes from either fading, or collisions between BACK and its hidden terminals. Consequently, there may exist redundant relay nodes, redundant rebroadcasts or WM collisions.

In Fig. 11, the performance degradation of OppCast w.r.t. IDEAL is also shown. When message traffic is dense ($r > 1$), the PRR in OppCast is lower than that of IDEAL. Since PRR is the primary goal in OppCast, when a BACK is not heard by a relay node, they tend to use more rebroadcasts to guarantee PRR. We then show the reliability of BACK by showing the total number of (re)broadcasts for each WM, which is broken down into number of relays and retransmissions by forwarders in Table 3.

For both OppCast95 and OppCast99, when $\rho = 120$ the number of relays are 1/3 more than their IDEAL counterparts (optimal), which consists major part of the redundant (re)broadcasts, but is acceptable. The number of retransmissions in OppCast95 increases faster with ρ, since OppCast99 uses more makeups that send BACKs to cancel forwarders' retransmissions. The redundant transmissions lead to PRR over-provisioning in OppCast.

In order to reduce the redundant (re)broadcasts, the BACK has to be reliably received by more nodes in the network. However, using longer communication range for BACK is not necessarily better, which will cause the exposed terminal problem. This can be seen from the lower dissemination rate of IDEAL when the channel load (ρr) is low to moderate (Fig. 11(c) and (d)). We believe that, to balance the goals of suppressing hidden terminals and avoiding exposed terminals, it is a good choice to set BACK's range to be around twice of WM's CR. The intuitive explanation is that, if the BACK's range is smaller than twice of WM's CR, there will be hidden terminals that cannot be suppressed; on the other hand, if the BACK's range is larger than twice of WM's CR, the nodes within the area from twice of WM's CR to the BACK's range will become exposed terminals.

Table 3
Average number of relays and retransmissions.

ρ	IDEAL95		OppCast95		IDEAL99		OppCast99	
60	24.0	1.5	27.5	6.8	39	1.5	41.5	3.8
120	22.0	1.3	31.6	13.0	36	1.3	49.5	4.0

7.3. Performance evaluation of OppCast with extension

In the following, we study the performance of OppCast-Ext and compare it with an existing distributed broadcast protocol, DV-CAST [27], which was proposed by Tonguz et al. to handle both the broadcast storm problem and disconnected network in highway VANETs in a seamless manner. The idea of DV-CAST is to employ a broadcast suppression technique when the network is dense, and to use a store-carry-and-forward strategy when the network is sparse. In DV-CAST, vehicles use only local connectivity information to make forwarding/carry decisions. There are three routing parameters to be locally determined by each vehicle: a destination flag (DFlg) which indicates whether the vehicle is inside the region of interest (ROI, or IR); message direction connectivity (MDC) which decides whether a vehicle is the last one within a cluster/group, and opposite direction connectivity (ODC) which tells if the vehicle is connected to any vehicle moving in the opposite direction. Vehicles whose DFlg = 1 will ignore duplicate packets to reduce unnecessary transmissions, while vehicles with DFlg = 0 act as "data mules" that store-carry-forward the same message more than once.

7.3.1. Simulation setup

We have implemented and simulated both OppCast-Ext and DV-CAST in NS-2.33. DV-CAST is implemented with the weighted-p persistence algorithm [13] for broadcast suppression (as recommended by [27]). The simulation scenario is, each vehicle located between 1 km and 2 km in the two west-bound lanes generates event-driven WMs according to a poisson process of average rate $r = 0.1$ (in the first 10 s), which are to be disseminated to all the following vehicles within the simulation highway area. We show the results of PRR, dissemination rate, broadcast delay[5] and total transmission count in Fig. 13. Each data point shows the average and error bars indicate the 95% percentile of 10–20 runs (with different topologies). To simulate the sparse VANET and the transition from sparse to dense regime, the vehicle density changes from 5 vehicles/km to 60 vehicles/km.

[5] The broadcast delay in each run is taken as the maximum end-to-end delay of the last vehicle receiving each WM in the IR, averaged among all the WMs sent in that run. It can be regarded as a worst case performance metric. We have included broadcast delay in the disconnected scenario since the dissemination rate cannot be measured very accurately here (vehicles may move a non-negligible distance before the arrival of a WM).

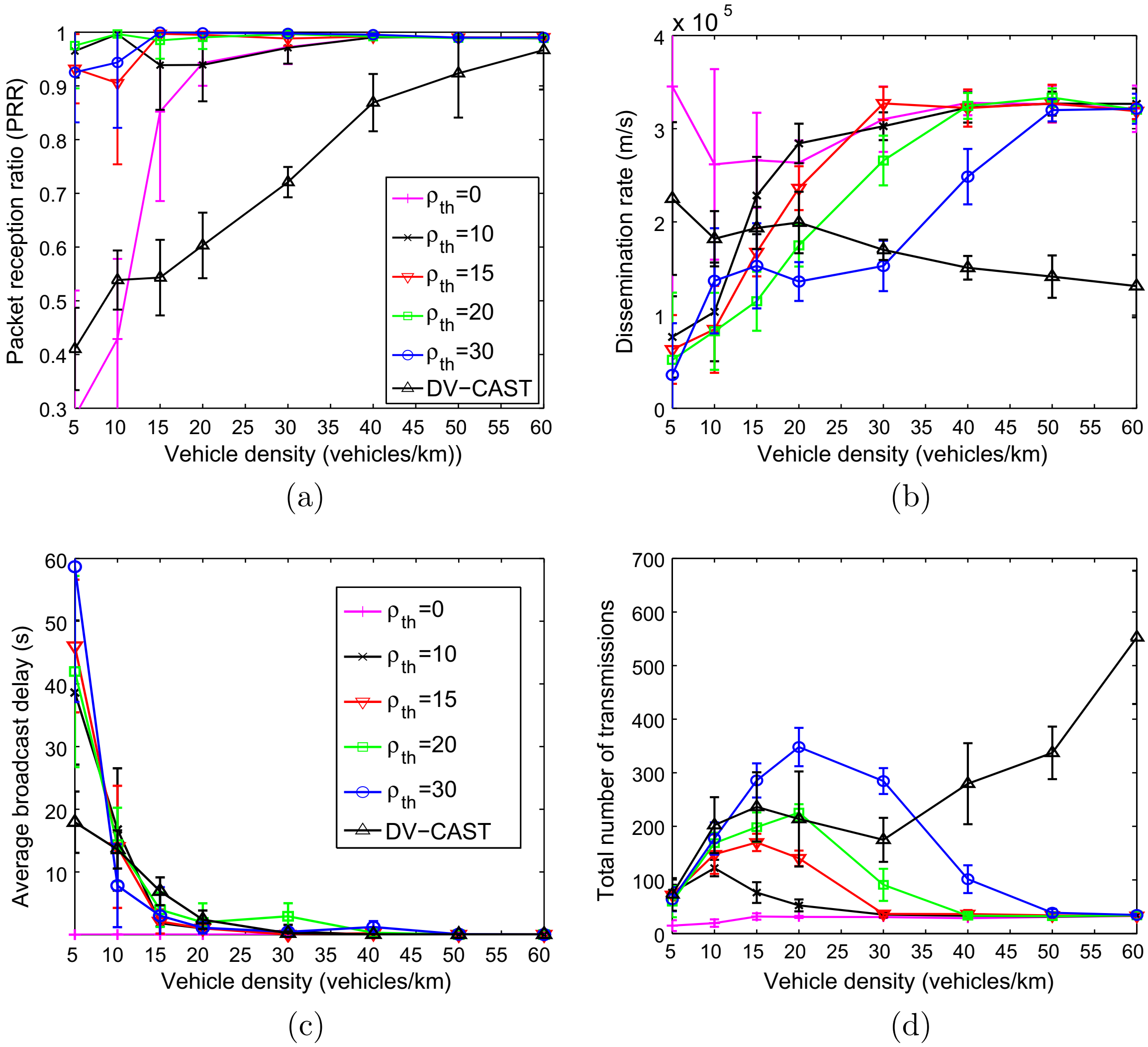

Fig. 13. Simulation results for OppCast95-Ext under sparse VANET and comparison with DV-CAST, $P_{th} = 0.95$, $r = 0.1$. The case $\rho_{th} = 0$ corresponds to OppCast without extension. The legends in all the sub-figures are the same with (a) (Each point is averaged from 10 to 20 repetitive runs with randomly generated vehicular mobility).

7.3.2. Simulation results

For OppCast-Ext, the effects of using different threshold densities ρ_{th} are clear from the figures. When the ρ_{th} increases from 0 to 30, the achieved PRR tends to increase for all the densities, the dissemination rate decreases, and the transmission overhead increases. These results are in line with the intuition, since the higher the ρ_{th}, the more often data mules are used and more redundant transmissions are incurred, which also makes the channel more saturated and thus increases the channel access delay. The redundant transmissions increase dramatically with the ρ_{th}, since we do not allow a data mule to cancel its carrying status for WMs, in order to ensure the PRR.

In addition, jointly considering the PRR requirements ($PRR \geqslant P_{th} = 0.95$) and the transmission overhead, one can find that the optimal threshold density equals 20 from Fig. 13. This matches well with our theoretical result derived in Section 5.2, where the optimal ρ_{th} is between 15 and 20.

As also shown by the simulation results, the PRR of DV-CAST is much lower than OppCast-Ext when the vehicle density is low, while the total transmission count incurred per source packet in DV-CAST is higher than that of OppCast-Ext when $\rho \geqslant 40$. These can be explained as follows. First, in DV-CAST there are no sufficient mechanisms to ensure the reliable reception of a WM packet from both the network layer and the link layer when the network connectivity is low. From the network layer, in DV-CAST a node decides to store-carry-forward based on local MDC/ODC status which are 0/1 indicators. A vehicle does not store-carry-forward a WM when MDC = 1 (as long as there is one neighbor vehicle in the message direction). However, *this strategy has not taken into account the probability of packet loss due to channel fading*. In reality, the message reception cannot be characterized deterministically using a fixed "transmission range". Although there are some "neighbor vehicles" in the message direction, there is a non-negligible chance that none of them could receive a packet in a single broadcast. In fact, the "connectivity" from the network level should be defined as the probability that a message can penetrate the whole IR, considering the actual message reception probability in each hop. The relationship between the network connectivity and the global vehicle density has already been shown in Fig. 10. It is not hard to see that as long as the vehicle density is smaller than the connectivity threshold we should let vehicles store-carry-forward the WMs. Furthermore, if the inter-vehicle distances follow the same distribution across the IR (such as exponential distribution), we can use the local vehicle density as the store-carry-forward indicator since the local density yields a good estimation of the local densities in other parts of the network. When the inter-vehicle distance distribution changes dramatically over a few KMs, using the local vehicle density will lead to a lower-bound of the connection probability, which ensures the message delivery. To this end, in OppCast-Ext we explicitly employ the local vehicle density information as a more fine-grained connectivity indicator, and vehicles decide whether to store-carry-forward based on comparing their local vehicle densities in the message direction with a threshold ρ_{th}. As we can see from Fig. 13, the PRR becomes almost 1 for all the densities when $\rho_{th} \geqslant 20$, which coincides with the result in Fig. 10. This store-carry-and-forward approach is different from and more reliable than that of DV-CAST which is a representative one.

On the other hand, from the link layer, DV-CAST does not guarantee the reception of a WM by a next hop vehicle. This is corroborated by the lower PRR of DV-CAST than OppCast when $\rho_{th} = 0$ where no store-carry-forward is made. In DV-CAST when message direction connectivity (MDC) equals 0 and opposite direction connectivity (ODC) equals 1, a vehicle that is a destination (DFlg = 1) will only rebroadcast once and go back to IDLE state, assuming that a vehicle in the opposite direction will receive the packet and carry it later on. However, this is insufficient as the wireless link in VANET is lossy – the opposite direction vehicle(s) may not receive the packet at all and the packet will die out. Moreover, in the broadcast suppression mechanisms in DV-CAST (including weighted-p, slotted-p, slotted-1 persistence), multiple vehicles may try to broadcast at the same time, which will cause packet collisions and further decrease the broadcast success rate. Although, in those broadcast suppression mechanisms, a node will rebroadcast with probability one after it has not heard any rebroadcasts within a time window to prevent the message from dying out, this applies mainly to well-connected scenarios. When the network is disconnected, broadcast suppression is not applicable.

In contrast, *throughout the design of OppCast and OppCast-Ext, we aim at providing reliable WM broadcast in VANET with lossy links from both the network and link layers.* Correspondingly, from the network layer OppCast uses the relay selection in MFR phase to ensure a certain PRR of each WM; from the link layer we introduce the opportunistic broadcast coordination function (OBCF) based on the concept of opportunistic forwarding, which uses an explicit broadcast acknowledgement (BACK) mechanism along with a forwarder retransmission mechanism to ensure the per-hop broadcast reliability. The whole idea of opportunistic forwarding is to leverage space diversity, as the node that receives a packet and has the maximum hop progress will rebroadcast it and effectively suppress other candidate nodes. This process avoids packet collision, since (1) in the broadcast backoff function in OBCF every node's backoff delay is distance-related and is unique; (2) the BACK is broadcast at the lowest rate and can be received reliably by most of the candidate nodes. Furthermore, a vehicle will retransmit at most *MAX_NUM_RETX* times if it does not hear a BACK from another vehicle in the message direction.

In addition, from the scalability point of view, when the network is well-connected, DV-CAST uses message rebroadcasts as implicit acknowledgements to suppress redundant rebroadcasts. However, implicit acknowledgements could be frequently lost, which decreases the effectiveness of broadcast suppression. This is why the transmission count in DV-CAST is higher than OppCast. For $\rho \leqslant 30$, the transmission count of DV-CAST is not the highest, which is because its PRR under these cases is low.

As to the broadcast delay, from Fig. 13(c) we can see that the delay decreases with the increase of vehicle density, and it is evident that there is a transition from the sparse to dense traffic regime which happens at around $\rho = 20$ vehicles/km. The transition is seamless. For both OppCast-Ext and DV-CAST, the average delay when $\rho > 20$ is smaller than 500 ms, except the ones in OppCast-Ext with $\rho_{th} = 20$ at $\rho = 20, 30$, and $\rho_{th} = 30$ at $\rho = 40$. The exceptions are possibly due to the small probability that the message does not penetrate the whole IR in the FFD phase, and must be delivered via store-carry-forward by data mules.

Finally, to reveal the WM dissemination dynamics under sparse VANET, we show the end-to-end delay v.s. distance graph in Fig. 14. The plateaus indicates the connected parts in the network. We can see that OppCast

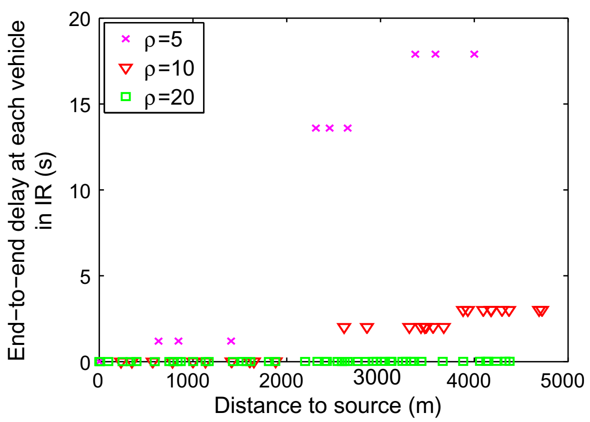

Fig. 14. The end-to-end delay of vehicles under sparse VANET, OppCast95-Ext, $\rho_{th} = 20$.

still allows a WM to propagate very fast within connected platoons, while always relay the WM to the next platoon successfully. This is ensured by the retransmission mechanism by the forwarders.

7.3.3. Discussion of robustness to imprecise neighbor knowledge

We note that the local vehicle density information exploited by OppCast and OppCast-Ext does not have to be very accurate. Because if the vehicle density is underestimated (which is common since beacons may not be received), in OppCast-Ext a vehicle will switch to store-carry-forward, which would only increase the reliability of WM reception. For OppCast, an evidence of the robustness of the algorithm to the vehicle density can be found in Fig. 11, where the IDEAL protocol uses global vehicle density for computing the optimized FR, while OppCast uses a distributed algorithm adopting locally measured vehicle density instead. The simulated PRR of the two protocols are almost equal when the network is well-connected, and the differences in transmission counts are small too.

On the other hand, in OppCast and OppCast-Ext, each vehicle is assumed to broadcast beacons at a rate of 10 messages/second (which is also required by DSRC). In OppCast the local vehicle density is measured by only considering vehicles within the nominal communication range (250 m); while the network topology almost does not change over (say, 0.5) seconds, this means a vehicle's beacon will have a high probability to be received after 5 repetitive broadcasts, and the local vehicle density can be measured reliably.

8. Conclusion

In this paper, we propose a fully-distributed opportunistic broadcast protocol (OppCast) for multi-hop dissemination of event-driven warning messages in VANETs with lossy links. Aiming at achieving high WM reception reliability and fast dissemination in a resource-efficient way, we propose a double-phase broadcast method in which fast propagation is ensured by one phase, and the desired reliability level is ensured by the other. The concept of opportunistic forwarding is exploited at each hop to enhance reception reliability and provide small hop delay. As a key idea in OppCast, we use explicit broadcast acknowledgements (BACK) in rebroadcast contention so that the optimal relays can always be selected with a high chance, and the undesired redundant rebroadcasts are dramatically reduced. Through extensive simulations we show that, compared with state-of-the-art protocols, OppCast achieves higher WM packet reception ratio, higher dissemination rate using lower amount of transmissions. More importantly, the BACK is shown to be a more reliable and effective approach than implicit acknowledgements adopted in previous works. In addition, we extend OppCast to handle disconnected VANET scenarios, in which the optimal threshold density to switch between normal dissemination and store-carry-and-forward scheme is characterized. Our results reveal the intrinsic tradeoff and intricate interplay between WM reception reliability, dissemination rate and overhead, and we believe it will provide valuable guidelines to VANET designers.

Acknowledgements

This work was supported in part by the US National Science Foundation under Grants CNS-0746977, CNS-0716306, and CNS-0831628. And we thank the anonymous reviewers for their helpful comments.

Appendix A. Proof of Proposition 1

Proof. The concavity of $\Phi_{\ell,\mu}(x)$ is straightforward. Let $\Phi_{\ell+1}(x,x_M)$, x, $x_M \in Z_{\ell,\mu}$ denote the $\ell+1$th level APRP given a node at x_M broadcasts, where $Z_{\ell,\mu}$ consists of $Z_{\ell+1,2\mu}$ and $Z_{\ell+1,2\mu+1}$. We use $W_{\ell+1,\mu}$ to represent a middle point and its coordinate interchangeably. It can be seen from the properties of concave and symmetric functions that $W_{\ell+1,\mu}$ is the minimum point of $\Phi_{\ell,\mu}(x)$. Then $\Phi_{\ell+1}(x, W_{\ell+1,\mu})$, $x \in Z_{\ell,\mu}$ is also symmetric w.r.t $W_{\ell+1,\mu}$:

$$\begin{aligned}
&\Phi_{\ell+1}(2W_{\ell+1,\mu}-x, W_{\ell+1,\mu}) \\
&\quad = 1-\big(1-P_r(|W_{\ell+1,\mu}-(2W_{\ell+1,\mu}-x)|)\big) \\
&\qquad \times \big(1-\Phi_{\ell,\mu}(2W_{\ell+1,\mu}-x)\big) \\
&\quad = 1-\big(1-P_r(|x-W_{\ell+1,\mu}|)\big)\big(1-\Phi_{\ell,\mu}(x)\big) \\
&\quad = \Phi_{\ell+1}(x, W_{\ell+1,\mu})
\end{aligned} \tag{A.1}$$

So there are two minimal points, x_L^* and x_R^* in $\left[x_L^{\ell+1,2\mu}, x_R^{\ell+1,2\mu}\right]$ and $\left[x_L^{\ell+1,2\mu+1}, x_R^{\ell+1,2\mu+1}\right]$ respectively, which are both equal to the minimum value of $\Phi_{\ell+1}(x, W_{\ell+1,\mu})$ in $Z_{\ell,\mu}$. In the following, we pick a point $x_{i_0} > W_{\ell+1,\mu}$ from the sequence of nodes within $Z_{\ell,\mu}$. First, we show that the minimum value of $\Phi_{\ell+1}(x, x_{i_0})$ is smaller than that of $\Phi_{\ell+1}(x, W_{\ell+1,\mu})$. At point x_L^*, we have $\Phi_{\ell+1}(x_L^*, x_{i_0}) < \Phi_{\ell+1}(x_L^*, W_{\ell+1,\mu})$:

$$\begin{aligned}
&\Phi_{\ell+1}(x_L^*, x_{i_0}) - \Phi_{\ell+1}(x_L^*, W_{\ell+1,\mu}) \\
&\quad = \big(P_r(|x_{i_0}-x_L^*|) - P_r(|W_{\ell+1,\mu}-x_L^*|)\big)\big(1-\Phi_{\ell,\mu}(x_L^*)\big) < 0,
\end{aligned} \tag{A.2}$$

since $P_r(x)$ is monotonically decreasing and $x_{i_0} > W_{\ell+1,\mu}$. Therefore, $\xi^*|W_{\ell+1,\mu} > \xi^*|i_0$. Similarly, for $x_{i_0} < W_{\ell+1,\mu}$, $\Phi_{\ell+1}(x_R^*, x_{i_0}) < \Phi_{\ell+1}(x_R^*, W_{\ell+1,\mu})$. Immediately, for any two nodes $\{i_0, i_1, \ldots, i_n\}$ within $Z_{\ell,\mu}$ such that $d(i_0, W_{\ell+1,\mu}) < d(i_1, W_{\ell+1,\mu}) < \cdots < d(i_n, W_{\ell+1,\mu})$, $\xi^*|W_{\ell+1,\mu} > \xi^*|i_0 > \cdots > \xi^*|i_n$. □

References

[1] M. Li, W. Lou, K. Zeng, Oppcast: Opportunistic broadcast of warning messages in vanets with unreliable links, in: IEEE MASS '09, 2009, pp. 534–543.

[2] Q. Xu, T. Mak, J. Ko, R. Sengupta, Vehicle-to-vehicle safety messaging in dsrc, in: VANET '04, 2004, pp. 19–28.

[3] M. Torrent-Moreno, J. Mittag, P. Santi, H. Hartenstein, Vehicle-to-vehicle communication: Fair transmit power control for safety-critical information, IEEE Transactions on Vehicular Technology 58 (7) (2009) 3684–3703.

[4] M. Li, Z. Yang, W. Lou, Codeon: cooperative popular content distribution for vehicular networks using symbol level network coding, selected areas in communications, IEEE Journal on 29 (1) (2011) 223–235, doi:10.1109/JSAC.2011.110121.

[5] R. Chen, W.-L. Jin, A. Regan, Broadcasting safety information in vehicular networks: issues and approaches, IEEE Network 24 (1) (2010) 20–25.
[6] DSRC and IEEE 1609 standard family, <http://www.standards.its.dot.gov/Documents/advisories/dsrc_advisory.htm>.
[7] M. Torrent-Moreno, M. Killat, H. Hartenstein, The challenges of robust inter-vehicle communications, IEEE VTC 1 (2005) 319–323.
[8] M. Torrent-Moreno, F. Schmidt-Eisenlohr, H. Fussler, H. Hartenstein, Effects of a realistic channel model on packet forwarding in vehicular ad hoc networks, in: IEEE WCNC, 2006, pp. 385–391.
[9] M. Torrent-Moreno, D. Jiang, H. Hartenstein, Broadcast reception rates and effects of priority access in 802.11-based vehicular ad-hoc networks, in: VANET '04, ACM, 2004.
[10] K. Tang, M. Gerla, Mac layer broadcast support in 802.11 wireless networks, in: MILCOM 2000, vol. 1, 2000, pp. 544–548.
[11] A. Dutta, D. Saha, D. Grunwald, D. Sicker, Smack: a smart acknowledgment scheme for broadcast messages in wireless networks, SIGCOMM Comput. Commun. Rev. 39 (4) (2009) 15–26.
[12] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, J.-P. Sheu, The broadcast storm problem in a mobile ad hoc network, in: IEEE/ACM MobiCom, 1999, pp. 151–162.
[13] N. Wisitpongphan, O. Tonguz, J. Parikh, P. Mudalige, F. Bai, V. Sadekar, Broadcast storm mitigation techniques in vehicular ad hoc networks, IEEE Wireless Communications 14 (6) (2007) 84–94.
[14] S. Oh, J. Kang, M. Gruteser, Location-based flooding techniques for vehicular emergency messaging, in: IEEE MobiQuitous, 2006, pp. 1–9.
[15] R. Mangharam, R. Rajkumar, M. Hamilton, P. Mudalige, F. Bai, Bounded-latency alerts in vehicular networks, in: MoVE, 2007, pp. 55–60.
[16] M. T-Moreno, Inter-vehicle communications: Assessing information dissemination under safety constraints, in: WONS, IEEE, 2007, pp. 59–64.
[17] H. Alshaer, E. Horlait, Optimized adaptive broadcast scheme for inter-vehicle communication, in: IEEE VTC, 2005, pp. 2840–2844.
[18] G. Korkmaz, E. Ekici, Urban multi-hop broadcast protocol for inter-vehicle communication systems, in: Proceedings of VANET '04, ACM, 2004, pp. 76–85.
[19] E. Fasolo, A. Zanella, M. Zorzi, An effective broadcast scheme for alert message propagation in vehicular ad hoc networks, IEEE ICC 9 (2006) 3960–3965.
[20] V. Taliwal, D. Jiang, H. Mangold, C. Chen, R. Sengupta, Empirical determination of channel characteristics for dsrc vehicle-to-vehicle communication, in: VANET '04, ACM, 2004.
[21] M. Zorzi, R.R. Rao, Geographic random forwarding (geraf) for ad hoc and sensor networks: Multihop performance, IEEE Transactions on Mobile Computing 2 (4) (2003) 337–348.
[22] S. Biswas, R. Morris, Exor: opportunistic multi-hop routing for wireless networks, SIGCOMM Comput. Commun. Rev. 35 (4) (2005) 133–144.
[23] B. Blaszczyszyn, A. Laouiti, P. Muhlethaler, Y. Toor, Opportunistic broadcast in vanets (ob-van) using active signaling for relays selection, in: ITS Telecommunications, 2008, 2008, pp. 384–389.
[24] T. ElBatt, S.K. Goel, G. Holland, H. Krishnan, J. Parikh, Cooperative collision warning using dedicated short range wireless communications, in: VANET '06, ACM, 2006, pp. 1–9.
[25] G. Resta, P. Santi, J. Simon, Analysis of multi-hop emergency message propagation in vehicular ad hoc networks, in: ACM MobiHoc, New York, NY, USA, 2007, pp. 140–149.
[26] B. Scheuermann, C. Lochert, J. Rybicki, M. Mauve, A Fundamental Scalability Criterion for Data Aggregation in VANETs, in: MobiCom 2009: Proceedings of the Fifteenth ACM SIGMOBILE International Conference on Mobile Computing and Networking, 2009.
[27] O. Tonguz, N. Wisitpongphan, F. Bai, Dv-cast: a distributed vehicular broadcast protocol for vehicular ad hoc networks, IEEE Wireless Communications 17 (2) (2010) 47–57.
[28] L. Ilias, M. Cecilia, Opportunistic spatio-temporal dissemination system for vehicular networks, in: MobiOpp '07, ACM, 2007, pp. 39–46.
[29] M. Abuelela, S. Olariu, I. Stojmenovic, Opera: Opportunistic packet relaying in disconnected vehicular ad hoc networks, in: IEEE MASS '08, 2008, pp. 285–294.
[30] F. Ros, P. Ruiz, I. Stojmenovic, Reliable and efficient broadcasting in vehicular ad hoc networks, in: Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th, 2009, pp. 1 –5.
[31] J. C-Varghese, W. Chen, O. Altintas, S. Cai, Survey of routing protocols for inter-vehicle communications, in: MobiQuitous, IEEE, 2006, pp. 1–5.
[32] T. Nadeem, P. Shankar, L. Iftode, A comparative study of data dissemination models for vanets, in: Third Annual International Conference on Mobile and Ubiquitous Systems: Networking Services, 2006, pp. 1–10.
[33] D. Jiang, V. Taliwal, A. Meier, W. Holfelder, R. Herrtwich, Design of 5.9 GHz dsrc-based vehicular safety communication, IEEE Wireless Communications 13 (5) (2006) 36–43.
[34] M. Killat, H. Hartenstein, An empirical model for probability of packet reception in vehicular ad hoc networks, EURASIP J. Wirel. Commun. Netw. (2009) 1–12.
[35] L. Yang, J. Guo, Y. Wu, Channel adaptive one hop broadcasting for vanets, in: 11th International IEEE Conference on Intelligent Transportation Systems, 2008, 2008, pp. 369 –374.
[36] H. Hartenstein, K. Laberteaux, A tutorial survey on vehicular ad hoc networks, Communications Magazine, IEEE 46 (6) (2008) 164–171.
[37] M. Ghassan et al., An adaptive channel model for VBLAST in vehicular networks, EURASIP Journal on Wireless Communications and Networking (2009).
[38] M. Li, W. Lou, Opportunistic broadcast of emergency messages in vehicular ad hoc networks with unreliable links, in: QShine '08, 2008, pp. 1–7.
[39] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, O. Tonguz, Routing in sparse vehicular ad hoc wireless networks, IEEE JSAC 25 (8) (2007) 1538–1556.
[40] Ns2, <http://www.isi.edu/nsnam/ns>.
[41] USC, The usc mobility generator tool, <http://nile.cise.ufl.edu/important/software.htm>.

Ming Li received his B.E. and M.E. degrees in Electronic and Information Engineering both from Beihang University, China, in 2005 and 2008, respectively. He is currently a Ph.D. student in the Electrical and Computer Engineering department at Worcester Polytechnic Institute. His current research interests include: wireless ad hoc and sensor networks, with emphases on protocol design and network security; and security & privacy in E-healthcare systems.

Kai Zeng received Ph.D. degree in the Electrical and Computer Engineering at Worcester Polytechnic Institute in 2008. He obtained B.E. degree in Communication Engineering and M.E. degree in Communication and Information System both from Huazhong University of Science and Technology, China, in 2001 and 2004, respectively. He is currently a postdoctoral researcher at Computer Science Department of University of California, Davis. His research interests are in wireless networking and network security.

Wenjing Lou earned a B.E. and an M.E. in Computer Science and Engineering at Xi'an Jiaotong University in China, an MASc in Computer Communications at the Nanyang Technological University in Singapore, and a PhD in Electrical and Computer Engineering at the University of Florida. She joined the Electrical and Computer Engineering department at Worcester Polytechnic Institute as an assistant professor in 2003, where she is now an associate professor. Her current research interests are in the areas of ad hoc, sensor, and mesh networks, with emphases on network security and routing issues. She was a recipient of the U.S. National Science Foundation Faculty Early Career Development (CAREER) award in 2008.