Overhead-free congestion control and data dissemination for 802.11p VANETs

Omar Chakroun, Soumaya Cherkaoui *

INTERLAB Research Laboratory, Université de Sherbrooke, Sherbrooke, Canada

**Abstract**

Direct radio-based vehicle-to-vehicle communication can be used to prevent accidents by letting vehicles exchange information about their state, intentions, and the road conditions. Although it is now a standard, the IEEE 802.11p protocol for vehicular ad hoc networks (VANETs) has known shortcomings. Dissemination is not reliable over distances higher than 300 m and congestion in the communication channel can lead to very low rates of safety messages delivery. Multi-hop routing and access channel techniques are well known approaches that were separately investigated to improve network effectiveness. However, achieving low end-to-end latency with multi-hop techniques is usually at the cost of lower data delivery rates, which in turn causes problems for the effectiveness of safety services embedded in vehicles. In this paper, we introduce a new dissemination and congestion avoidance scheme for safety messages over IEEE 802.11p VANETs. In order to ensure good delivery rates beyond 300 m, the approach propagates information over two hops while avoiding the resulting congestion by using a fully distributed asymmetrical transmit power adjustment technique. The scheme uses two time-dependent optimization-under-constraint processes to elect the best vehicle to act as a relay for data forwarding. The scheme can estimate the probability of reception rate (PRR), and adjust the forwarding distance to meet the minimum requirements of PRR and delivery distance to fit specific safety application requirements. The proposed solution, unlike previous dissemination techniques, works simultaneously on reducing congestion due to multi-hop relaying and on ensuring low end-to-end delay. Simulation results confirm the effectiveness of the proposed adaptation and relaying scheme and its advantageous network performance compared to others, under various traffic constraints.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Vehicular communications play an important role in implementing next generation active safety applications. Vehicular Ad hoc Networks (VANETs) leverage communicating devices introduced in vehicles so that these can exchange useful information. This information exchange aims to extend vehicles and drivers perception by constructing a global awareness of the surrounding environment and other vehicles intentions. Data communication among vehicles ensures high information reach which gives the driver more reaction time to avoid hazardous situations. In high speed environments such as highways and freeways, the time for a vehicle to decelerate down to a safe speed or to a zero speed is longer. Since the driver reaction time is constant and usually ranges between 1.5 and 3 s, it is important to convey information about the hazardous events as soon as possible and ideally in a few milliseconds. There are two types of safety messages. The first one is event driven, where messages notify about hazardous situations. The timely delivery of such messages might be life critical, and therefore these messages are particularly delay sensitive. The delivery of such messages up to a certain distance for the emitter is also required to ensure a timely reaction from vehicles potentially affected by the event. The second type of messages is routine messages, which contain information about the speed, location, heading of the sender, etc. These messages are periodically broadcasted to immediate neighboring vehicles so that they can update their information on the surrounding environment. These messages also are delay sensitive, but their periodic broadcast can cause channel congestion.

Multiple routine (beacon) messages per second are needed to provide the required accuracy for safety applications to operate correctly. The key issue related to such beaconing activity is how to ensure the fair trade-off between data availability and freshness without causing channel congestion. Adding to this issue, is the fact that the channel access technique used in the 802.11 systems,
i.e. DCF, is an asynchronous one and slatterns wireless medium resources.

Pursuing effective networking solutions for VANETS, a number of challenges have to be addressed. Knowing that the main dissemination technique in VANETs is broadcast, how to ensure that continuous broadcasting will not affect network performances (i.e. delivery rates and delays) especially close to the channel saturation threshold? Also, while 802.11p was designed to reach a dissemination distance of 1000 m, the actual maximum achievable range with 802.11p barely reaches 300 m. Therefore, how to meet the dissemination distance needed by a broad range of safety applications without causing congestion?

Efforts to tackle the above challenges have generally been targeted towards a single goal among three: (1) designing new channel access schemes to avoid transmission collisions, (2) designing congestion control schemes to prevent network performance degradation in dense scenarios, and (3) proposing multi-hop dissemination schemes to propagate messages over the network. The first and second solutions generally fail to extend messages delivery distance, while the third kind of solutions usually further exacerbates network congestion. In this paper, we address these challenges simultaneously as optimization problems that have to be solved simultaneously to form an efficient safety messages dissemination scheme for 802.11p/VANETs.

Our work complements the approaches in [15,16] and proposes MORS, an efficient, overhead-free approach for multi-hop safety messages dissemination over 802.11p with congestion control. The scheme uses local measurements with no need for control messages exchange and therefore does not induce an overhead. It integrates a dissemination metric, based on an approximation of the expected reliability and communication range usage. The approach aims to improve the packet reception rate (PRR) and reduce the End-to-End (E2E) message dissemination delay. For this, we use multiple optimization-under-constraints processes that characterize each optimization problem: 1) the communication range choice to minimize congestion and 2) the optimal farthest relay designation according to the current network state in order to maximize the probability of reception.

The remainder of this paper is structured as follows; Section 2 discusses existing schemes for data dissemination and power/range adaption. Section 3 introduces the problem formulation. Section 4 details MORS, and its different operating phases. Section 5 presents the theoretical system analysis. Section 6 gives the simulations result obtained with realistic scenarios. Finally, Section 7 concludes the paper.

2. Related works

Reliable message dissemination over 802.11p networks is hindered by a number of problems, the most notorious of which is the broadcasting storm that causes congestion in channels already experiencing multipath, high nodes velocity, and a wide dynamic range of signal strengths. Another problem which can add to congestion in accessing the communication medium is that beacon messages will carry security-related overhead leading to messages having a large size up to 800 bytes [12]. This will result in payload being disseminated on multiple messages over an already congested channel.

In order to enhance message broadcasting reliability, and extend the message delivery distance, some authors worked on new channel access schemes or congestion control methods, while others focused on multi-hop routing schemes.

A number of access schemes and congestion control schemes have been proposed in the literature whose main goal is to ensure message delivery with the best achievable Quality of Service (QoS) [10–12] based on link state [13,14] or disconnections number [9, 17,18]. In previous works [15,16,21], we introduced VDA, a time-synchronized access over 802.11p which ensures contention-free and fairness in channel access between vehicles even in dense scenarios. However VDA only insures fairness in a two-hop vicinity of participating vehicles and non-participating contending vehicles will experience packet losses.

To account for the 1000 m dissemination barrier, messages have to be forwarded for multiples hops. For this purpose, some works focused on Broadcast schemes, others on Unicast or Geocast multi-hop schemes.

In the broadcast schemes category, Smart Broadcast (SB) [24] uses a distance based forwarding technique electing the farthest node in the communication range of the emitter as a relay based on blackburst. Korkmaz et al. proposed two designs; Urban Multi-hop Broadcast (UMB) [2] which uses a continuous message exchange to calculate distances between communicating nodes and elect the farthest one as a relay. The other design is Ad hoc Multi-hop Broadcast (AMB) which is an improvement of UMB electing the closest node to an intersection as a relay to a particular section of the road. Another scheme, Fast Broadcast (FB) [4] uses greedy forwarding which adapts the waiting time before rebroadcasting by giving the farthest vehicle in the communication range a higher priority to relay the message. Reliable and Efficient Alarm Message Routing in VANETs (REAR) [5] considers PRR as a main metric to guaranty, but does not offer any bound on data forwarding delays. RMOSGP [12] and GVGrid [12] respectively rely on categorizing communicating vehicles based on their speed, heading, and the number of sub-sequent link disconnections. However designs leveraging control messages exchange, such as the latter will induce an overhead which can have negative impact on the network performance.

In the unicast-based schemes category, Naumov et al. introduced Connectivity Aware Routing (CAR) [9] which by pre-establishing the dissemination path guarantees lower delays. It uses HELLO messages exchange from the source to the destination and on the reverse path to construct a routing route similarly to ADDV. DV-CAST [6] ensures high messages delivery by electing the less loaded links all over the routing path to construct a route. Moreno et al. in [8] proposed a highly dynamic transmission power adaptation scheme which guarantees a fair channel access between vehicles. It involves exchanging control messages containing status information such as network density and neighbors’ number. Again, the use of control messages induces an overhead which can impact network performance. The MHVB scheme introduced by Tatsuaki et al. [7] tunes the beacons messaging frequency depending on the number of nodes in the communication range to avoid network congestion. However, it does not offer any guaran-
tee neither on the rate of successfully delivered messages nor on the delivery delays.

In the Geocast-based schemes category, Position-Based Adaptive Broadcast (PAB) [3] integrates a design to overcome network disconnections by implementing a store-and-forward scheme used in case of links breakage. It uses position and speed information to construct a global routing map in the network. Such a scheme guarantees messages delivery but not the delay. DTSG [23] introduces another approach of geo-casting, called time-stable as it acts on the time when messages are geo-casted. DTSG integrates the idea of helping vehicles navigate in the opposite direction. However, geo-casting needs a continuous exchange of control messages containing information for mechanisms such as cluster formation, etc. Another scheme, ROVER [22] is a simple geo-casting technique based on zones definitions (zone of forwarding and zone of relevance). It integrates a neighbors’ discovery technique based on messages exchange and groups’ definition, which at the end leads to extra overhead compromising network performances. Besides, a lost discovery message in their design means that a part of
the zones will not be aware of the event. Ayaida et al. in [26] presented a highly interesting concept combining routing protocol with location-based services to reduce the signaling overhead for a hybrid and hierarchical geographic routing protocol. In such schemes, the messages are forwarded using the last receiver updated position. Afterward, a location request is triggered in the vicinity of that “old” position information. This reduces the overhead geographically since and end-to-end signaling is not needed. Jerbi et al. in [27] introduced an intersection-based geographical routing scheme which combines a clustering-like technique for cell formation and a store-and-forward technique. In this scheme, the route is discovered while sending data messages (in a hop basis) and can be applied for a fixed or moving infrastructure (electing a moving vehicle as a relay). For the aforementioned two techniques, the resulting delay while interesting for service applications, is not suitable for safety messages since it exceeds 1 s. In addition to the misused network resources, geo-casting techniques generally need spatial relevance and do not ensure any constraint on delivery delay.

3. Problem formulation and motivation

We propose an overhead free dissemination scheme in which no overhead is caused by specific messages exchange to enhance communication, i.e. Blackburst or similar messages. In this scheme, a judicious message forwarder selection technique is based on a multi-metric. We promote a “unicast” technique over one hop to avoid channel congestion caused by the broadcasting storm and thus elect only one forwarder. We use the fact that on-board antennas are omnidirectional and that nodes can be configured to detect and decode all messages even if these are not addressed to them, i.e. similar to the promiscuous mode in wired networks. In our scheme, the only node that can rebroadcast a message is the one selected using the multi-metric. The address of this node will be tagged on the message to be disseminated. Again all this operation is performed without extra messages as overhead. The scheme also uses a congestion avoidance technique based on an asymmetric power adjustment leveraging real-time communication density estimation.

A first part of the problem description can be modeled as a trade-off between maximizing the range of dissemination and keeping communication density under certain threshold to preserve networking performances. The second part of the problem can be modeled as maximizing an objective function which is used to choose the best relay upon possible relays in the emitter vicinity. A possible writing form of the problem can be expressed as in (P1), where $U_i$ designates the objective function governing the relay choice of node $n_i$, $C_{R_i}$ the current communication range of node $n_i$, $C_{D_i}$ and $C_{D_{th}}$ designate respectively the communication density measured locally at node $n_i$ and the communication density threshold.

$$\text{Max } \{C_{R_i}, \text{Max } U_i\}$$

(P1)

For each selected power corresponding to the maximum achievable range considering the current network load, the problem is to designate the best relay in the achievable range that maximizes the objective function. As the same behavior will be processed at each hop, we promoted an optimization process in which every hop corresponds to a cycle and in which the valuation function is the whole maximization process (P1). Ideally, an optimal policy would continuously, at each sensing phase, assign a power to each node and select the best fitting corresponding relay by solving the (P1) problem to optimality. However, the exact (P1) solution requires a global knowledge of all feasible relay choices, power assignment and a centralized algorithm to solve a mixed integer non-linear problem (NP-hard generally) such as (P1) on a sensing phase basis. This is unpractical for real-time decision making and goes against the objective of reducing network load, but it provides the general idea for a distributed algorithm that seeks to achieve local optimality based on locally collected information. That is why local optimality and a problem decomposition approach, as shown in the section below, has been promoted to maintain an acceptable complexity and processing time.

Hereafter, an enhanced design scheme will be theoretically discussed based on two distinct optimization sub-processes: (1) a power adjustment technique to select the right emitting power with respect to the induced communication density (CD) and (2) a data dissemination scheme based on local measurements and on a newly designed multi-metric that takes into account: (a) distance estimation between emitting and next relay, (b) an approximation of the link reliability in terms of expected PRR.

4. Proposed approach: MORS

We present our proposed scheme, called Multi-metric Overhead-Free Routing Scheme (MORS). MORS is a multi-metric data dissemination scheme based on two primary metrics; (a) PRR; and (b) Distance (D) over communication range (CR) ratio (D/CR). These two metrics are locally measured and every node is supposed to have the ability to compute them.

MORS operates in two phases and is the combination of two schemes that are time-dependent on each other; (1) Fully Distributed Congestion Control (FD2C) which performs a range/power adaptation to guarantee the fastest on-hop message delivery according to the network state in terms of CD. (2) Unicast Multi-hop Data Dissemination (UM2D) whose main function is to perform the next-hop relay node election based on a compromise between envisioned PRR while maximizing D/CR. The use of PRR guarantees reliability of messages forwarding and D/CR ratio maximization to choose the forwarder guarantees less hops and reduces the overall dissemination delay.

In this work we assume that:

(a) All vehicles are equipped with 802.11p enabled communication devices compliant to the standard DSRC/802.11p, use the VDA channel access scheme [15] and their output power and receivers sensitivity are known.

(b) Signals are subject to the same attenuation in both directions of a particular link. Considering two communicating nodes A and B; if a received signal strength (RSS) attenuation measurement is performed at node B (respectively A), it will be the same as that measured at node A side (respectively B).

(c) The detection range is supposed to be twice the communication range.

(d) Without loss of generality, and for simplification purposes, we suppose that all nodes use the same message frequency and message size.

We use $CD_i$ definition as specified in Eq. (1) [1], which is computed at node $n_i$ based on transmission range value ($C_{R_i}$ in meters), message frequency ($\text{Msg} \_\text{frq}$ in Hz) and vehicles density ($\text{Veh} \_\text{density}$, in vehicles/km of road).

$$CD_i = \text{Msg} \_\text{frq} \times C_{R_i} \times \text{Veh} \_\text{density}_i$$

(1)

Table 1 summarizes the variables used for the optimization processes.

In the following we detail the two optimization sub-processes of MORS, (a) the adaptation phase with FD2C and (b) the dissemination phase with UM2D.
Table 1
Table of symbols for the optimization process.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CD_i$</td>
<td>Communication density perceived by node $n_i$.</td>
</tr>
<tr>
<td>$CD_{th}$</td>
<td>Communication density threshold, after which an adaptation is initiated.</td>
</tr>
<tr>
<td>$CR$</td>
<td>Maximum achievable communication range.</td>
</tr>
<tr>
<td>$Msg_{freq}$</td>
<td>Message frequency in Hz of node $n_i$.</td>
</tr>
<tr>
<td>$Veh_density$</td>
<td>Vehicle density in vehicle/km of road perceived by node $n_i$.</td>
</tr>
<tr>
<td>$N$</td>
<td>Set of vehicles (nodes) in a given area.</td>
</tr>
<tr>
<td>$N_i$</td>
<td>Set of vehicles (nodes) perceived by node $n_i$ in its vicinity for a given $CR_i$.</td>
</tr>
<tr>
<td>$CR_i$</td>
<td>Current communication range of node $n_i$.</td>
</tr>
<tr>
<td>Propag</td>
<td>Function of propagation model determining the communication range for any transmission power.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Propriety function between the two main metrics.</td>
</tr>
<tr>
<td>$Prt_{i,j} \in N_i$</td>
<td>Probability of reception rate function for messages emitted by node $n_i$ to node $n_j$.</td>
</tr>
<tr>
<td>$Pwr_{i}$</td>
<td>Discrete power value selected by the node $n_i$.</td>
</tr>
<tr>
<td>$\text{dist}_{i,j}$</td>
<td>Distance between node $n_i$ and node $n_j$.</td>
</tr>
<tr>
<td>$\text{dist}_{\text{cross}}$</td>
<td>Cross-over distance. Defined as the distance to switch from Friis to Two-ray model. [20] Depends on the transmission rate, chosen frequency and coding rate.</td>
</tr>
</tbody>
</table>

4.1. Fully Distributed Congestion Control (FD2C)

FD2C constitutes the first optimization sub-process. It performs a range/power adaptation to guarantee the fastest message delivery according to the network state in terms of $CD$. It is a fully distributed congestion control mechanism which controls the load by adjusting the transmitting power locally at each node. Each node estimates the current $CD$. By adjusting its power, a node limits its detection range which will, in turn, lower its perceived local $CD$. In other words, the bigger the transmission range of a node, the farther it will affect transmissions of other nodes with transmission collisions but the longer the reach of its sent messages. At the same time, the bigger the transmission range of a node, the bigger its detection range, and the more it will perceive traffic of others.

Given a set of nodes $N = \{n_1, n_2, \ldots, n_m\}$, each node estimates the local load by performing an evaluation of the $CD$ based on Eq. (1). This evaluation is performed using overhearing technique. Using this technique, a node can detect vehicles in its communication range and thus evaluate the overall local $CD$. Note that the range estimation is based on the used propagation model, designated here by function Propag. For each node $n_i \in N$, $CR_i \in (0, CR)$, the communication range will be a function of the emitting power. Using the conclusions made by Jiang et al. [1] that for a fixed message size, the network behavior and performances such as PRR for any technique given technique will depend only on $CD$, the process can be modeled by an optimization under constraint problem as illustrated below, where $Pwr_{i}$ represents node $n_i$ current transmission power, and $CD_i$ and $CD_{th}$ point to the measured $CD$ at node $n_i$ and the $CD$ threshold respectively. Note that $Pwr_{i}$ constitutes a discrete power value selected by $n_i$ depending on the equipment capabilities.

$$\text{Max} \left\{CR_i = \text{Propag}(Pwr_{i}) \right\} \quad \text{(P.2.1)}$$

The optimization process is promoted involving maximizing the one-hop transmission range (to keep a low end-to-end delay) while ensuring a certain $CD_{th}$ threshold is not attained. It is worth noting that, since an overhearing technique is introduced, this approach can be easily extended to support multiple messaging frequencies and message sizes.

4.2. Unicast Multi-hop Data Dissemination (UM2D)

The need for the multi-hop data forwarding is exacerbated by the fact that while 802.11p was designed to reach a dissemination distance of 1000 m, the actual maximum achievable range with 802.11p barely reaches 300 m. FD2C decides the maximum transmission range, but many nodes falling within that range can be chosen as relays. A metric is needed to discriminate among all candidate nodes. From previous works [6–9] it is clear that using a single metric to devise multi-hop dissemination solutions is insufficient and a combination of many metrics is needed. In our work, we use a newly designed multi-metric based on real-time measurements of distance to, and link quality to, neighboring nodes (candidate forwarding nodes). Link quality influences the PRR. The better the link quality the bigger is the expected PRR. Distance influences end-to-end delay. Since the biggest component of transmission delay is the channel access delay, the smallest the number of hops, the lower is the end-to-end delay. Therefore, the farthest the next relay chosen, the smallest is the expected number of hops to go farther and consequently the lower is the overall end-to-end dissemination delay.

Link state can be characterized in multiple manners; life duration, number of disconnections, and duration of disconnections are some of them. In our work, we chose link state in terms of estimated PRR over distance as illustrated in Eq. (2) [20] where the PRR is a function of the current communication range of node $n_i$ and the distance between an emitting node $n_i$ and a receiver node $n_j$ (a candidate relay).

$$Prt_{i,j} = \begin{cases} 
2^{-\gamma \frac{\text{dist}_{i,j}}{\text{dist}_{\text{cross}}}} & \text{dist}_{i,j} < \text{dist}_{\text{cross}} \\
2^{-\gamma \frac{\text{dist}_{i,j}}{\text{CR}_i}} & \text{dist}_{i,j} \geq \text{dist}_{\text{cross}} 
\end{cases} \quad \text{(2)}$$

where $\text{dist}_{\text{cross}}$ designates the cross-over distance [20] for a given transmission rate (message frequency $\times$ message size), $N_i$ is a set of vehicles (nodes) perceived by node $n_i$ in its vicinity for a given $CR_i$, $N_j = \{n_1, n_2, \ldots, n_m\}, 0 < \text{dist}_{i,j} < CR_i$ (note the $CD$ is the cardinality $m$ of $N_i$ per meter) and $\gamma$ is defined as in (3).

$$\gamma = \frac{1}{\text{dist}_{\text{cross}}^2} \quad \text{(3)}$$

A simpler writing form of Eq. (2) can be as follows

$$Prt_{i,j} = 2^{-3\beta (A)} (1 + 3(A)^2 + \frac{9}{2}(A)^4) \quad \text{(4)}$$

where $A$ is expressed as

$$A = \begin{cases} 
\frac{\text{dist}_{i,j}}{\text{CR}_i} & \text{dist}_{i,j} < \text{dist}_{\text{cross}} \\
\frac{\text{dist}_{\text{cross}}^2}{\text{CR}_i} & \text{dist}_{i,j} \geq \text{dist}_{\text{cross}} 
\end{cases} \quad \text{(5)}$$

The objective function we have chosen for the second part of (P1) is a combination of the PRR estimation as in Eq. (4), the ratio between the link length and the communication range, and $\beta$, $0 \leq \beta \leq 1$ a proportionality function between the two metrics so we can promote one over the other depending on the targeted safety application. The objective function in the second part of (P1) is rewritten as

$$U_{i,j}(\beta) = \beta \left( \frac{\text{dist}_{i,j}}{\text{CR}_i} \right) + (1 - \beta)Pr_{i,j}(A) \quad \text{(6)}$$

The constraint expressed on the distance ensures that none of the nodes that are out of the communication range already assigned by FD2C is selected as a relay. This maximization process is
Table 2
Example of local connectivity table (node a, CR\textsuperscript{a} = 100 meters).

<table>
<thead>
<tr>
<th>Link</th>
<th>a, b</th>
<th>a, c</th>
<th>a, d</th>
<th>a, e</th>
<th>a, f</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR\textsubscript{a}</td>
<td>50</td>
<td>70</td>
<td>60</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>PRR measurement</td>
<td>0.95</td>
<td>0.81</td>
<td>0.9</td>
<td>0.98</td>
<td>0.42</td>
</tr>
<tr>
<td>(U(\beta = 0.5))</td>
<td>0.72</td>
<td>0.75</td>
<td>0.73</td>
<td>0.65</td>
<td>0.71</td>
</tr>
<tr>
<td>(U(\beta = 0.3))</td>
<td>0.82</td>
<td>0.78</td>
<td>0.81</td>
<td>0.79</td>
<td>0.6</td>
</tr>
<tr>
<td>(U(\beta = 0.7))</td>
<td>0.63</td>
<td>0.73</td>
<td>0.69</td>
<td>0.50</td>
<td>0.82</td>
</tr>
</tbody>
</table>

performed on every hop for each available link. The resulting optimization sub-process in which we try to maximize the objective function \(U\) is as follows.

\[
\max_{\beta \in \mathbb{R}^+} \left\{ \left( \frac{\text{dist}_{i,j}}{\text{CR}_i} \right) + (1 - \beta) \text{Prr}_{i,j} \right\} \tag{P2.2}
\]

where:

\[
A = \begin{cases} 
\frac{\text{dist}_{i,j}}{\text{CR}_i}, & \text{dist}_{i,j} < \text{dist}_{\text{cross}} \\
\text{dist}_{\text{cross}} \times \frac{\text{CR}_i}{\text{dist}_{\text{cross}}}, & \text{dist}_{i,j} \geq \text{dist}_{\text{cross}} 
\end{cases}
\]

Then, the whole optimization process combination can be rewritten as an algorithm combining the two main sub-processes (P2.1) and (P2.2), while keeping the temporal link governing their execution and the use of the optimization approach on hop-by-hop basis. Note that the envisioned optimality is only local. This choice was made to prevent network congestion caused by information gathering to construct global network knowledge.

Algorithm 2.1 Power adjustment (FD2C).

1: \(\text{CD}_0\): CD threshold
2: \(\text{CD}_i\): CD measurement at node \(i\)
3: \(\text{CR}_i\): maximum transmission range for node \(i\)
4: \(\text{Pwr}_i\): emitting power at node \(i\)
5: For each node \(i\)
6: \(\text{If } (\text{CD}_i < \text{CD}_0)\)
7: \(\text{Pwr}_i \leftarrow \text{Raise}_\text{pwr}()\)
8: \(\text{Else if } (\text{CD}_i > \text{CD}_0)\)
9: \(\text{Pwr}_i \leftarrow \text{Reduce}_\text{pwr}()\)
10: \(\text{Else}\)
11: \(\text{Pwr}_i \leftarrow \text{Maintain}_\text{pwr}()\)
12: \(\text{End if}\)
13: \(\text{Return } \text{Pwr}_i, \text{CR}_i\)
14: \(\text{End for}\)

Algorithm 2.2 Relay selection (UM2D).

1: Given (\(\text{CR}_i\), \(\beta\))
2: \(j\): relay ID, \(k\): table index
3: \(\text{dist}_{i,j}\): distance of node \(j\) from node \(i\)
4: \(\beta\): proportionality function
5: Table\(\text{i}()\): connectivity table at node \(i\) side
6: \(U_i(\beta)\): objective function as in Eq. (6)
7: For each possible relay \(j\) within \(\text{CR}_i\) reach
8: evaluate \(U_{ij}\), evaluate \(\text{dist}_{ij}\)
9: Table\(\text{i}()\) = add_entry\(\text{(i, j, U_{ij})}\)
10: \(\text{End for}\)
11: \(\text{Return } f^* = \text{Table(index.of (Max(U_{ij})))}\)

Table 2 shows an example of the locally generated connectivity table at a node \(a\) while varying \(\beta\) parameter. The highlighted cells point to the optimal relay choice and give measurements of the expected PRR.

5. System design analysis

5.1. Algorithms complexity

Dissociating the main optimization problem (P1) into two sub-problems (P2.1) and (P2.2), while maintaining their temporal dependency, reduces the overall problem complexity in terms of execution and reduces its resolution latency. Let assume the existence of \(n\) nodes in the network and consider that simple instructions as comparisons do not induce more complexity. (P2.1) will run \(n\) times one for each node, which constitutes a complexity up to \(O(n)\). Assume that \(m < n\) are in the communication range of the node \(i\) with \(m < n\) which gives to (P2.2) a complexity up to \(O(m)\) with \(O(m) < O(n)\). Thus, the overall approach complexity can be expressed as \(O(n + m)\).

Let assume the case where we are using the problem definition as in (P1). In that case, a possible description of (P1) will be: for each node \(i\) and for each power level \(\text{Pwr}_i\), the (P2.2) algorithm will be executed. Thus, the overall problem complexity will be \(O(n \times m)\) (Fig. 1).

5.2. Probabilistic one hop delay analysis

The one hop delay is assumed to be the latency between the reception of a message and its dissemination to the relay node. Considering the range as a circular area, the number of vehicles in such area can be assumed to follow a Poisson distribution with density parameter \(\lambda\). Thus the probability to have \(k\) vehicles in the communication range is given by

\[
p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \tag{7}
\]

5.2.1. VDA functionalities establishment delay

Considering the fact that our work is based on the timesynchronized channel access technique called Vehicular Deterministic Access (VDA) [15], for each vehicle, which needs to communicate, time slots are reserved corresponding to its communication requirements. Thus, each time slot (TS) of duration \(T\) can experience one of the following states: (1) idle state in which this particular TS is not used, (2) collision state in which two or more vehicles choose the same TS to communicate, and (3) success state in which one and only one vehicle used that TS to exchange information. Knowing that each vehicle using VDA has to exchange VDAOp messages in broadcast mode to allocate its TS, the aforementioned state delays can be expressed as follows:

\[
T_{\text{idle}} = T \tag{8}
\]
\[
T_{\text{collision}} = TVDAOp + DIFS \tag{9}
\]
\[
T_{\text{success}} = TVDAOp + SIFS + TVDAOpreply \tag{10}
\]

where \(TVDAOp\) is the delay to send the VDAOp reservation message, \(TVDAOpreply\) is the delay to receive the acknowledgment, SIFS is the Short Interframe Space duration and DIFS the Distributed Inter-Frame Space duration in 802.11.
Let $p = \frac{1}{T}$ denote the probability that a node selects a random slot, where $nT$ denotes the number of available timeslots. Let $p_{idle}$, $p_{success}$, $p_{collision}$ denote respectively the probability that a TS experiences one of the aforementioned states, then the probabilities can be expressed as follows:

$$p_{idle} = e^{-\lambda p}$$  \hspace{1cm} (11)

$$p_{success} = \lambda pe^{-\lambda p}$$  \hspace{1cm} (12)

$$p_{collision} = 1 - p_{idle} - p_{success} = 1 - e^{-\lambda p}(1 + \lambda p)$$  \hspace{1cm} (13)

Let $T_{\text{contention}}$ denote the delay between the first VDAOP sent and the reception of acknowledgment messages from all the vehicles in the communication range, which in turn corresponds to the necessary delay for VDA to reach the steady state in which the vehicle has its TS reserved. Note that the occurrence of one of the states does not depend on the previous states and can be modeled by a geometric distribution. Hence, $T_{\text{contention}}$ can be expressed as in Eq. (14).

$$T_{\text{contention}} = \frac{1 - p_{success}}{p_{success}} \left( T_{idle} \frac{p_{idle}}{1 - p_{success}} + T_{collision} \frac{p_{collision}}{1 - p_{success}} \right)$$  \hspace{1cm} (14)

Let $T_{\text{retransmission}}$ denote the retransmission delay for VDAOP after $n$ unsuccessful contention rounds in the VDA reservation procedure which can be expressed as in Eq. (15).

$$T_{\text{retransmission}} = \frac{1 - p_{success}}{T.n.p_{success}} T_{\text{VDAOP}}$$  \hspace{1cm} (15)

Hence, the time delay to a vehicle using VDA to reach the steady state can be expressed as in Eq. (16).

$$T_{\text{delay_vda}} = T_{\text{contention}} + T_{\text{retransmission}} + T_{\text{success}}$$  \hspace{1cm} (16)

5.2.2. Power adjustment and relay selection delay

In the previous sections, we introduced a power adjustment scheme and a relay choice technique to enhance the network connectivity and reduce the overall dissemination delay in a dynamic manner which combined with the optimization approach constitute the main contributions in this work. This adaptive behavior introduces an extra delay necessary for the optimization functionalities. In our design, the power adjustment phase is triggered according to the communication density measurements made locally by each node.

The adaptation delay, $T_{\text{delay_adaptation}}$, which is the delay incurred at a node to perform the optimization operations, can be expressed as in Eq. (17), where $T_{\text{pwr_adj}}$ and $T_{\text{relay_choice}}$ are respectively the power adjustment delay and the relay choice delay.

$$T_{\text{delay_adaptation}} = T_{\text{pwr_adj}} + T_{\text{relay_choice}}$$  \hspace{1cm} (17)

Let us assume that we maintain the definition of communication density (CD) as expressed in Eq. (1) and assume that all vehicles use the same maximum communication range (CR) and the same messaging frequency (Msg.frq). Then CD, condition is reduced to a vehicle density threshold $\text{Veh\_density}_{th}$. $p_{NR}$ and $p_{OT}$ denote respectively the case where there is no available relay in the communication range and the case where the number of neighbor vehicles exceeds the maximum allowable for a reliable communication (saturation).

$$\text{Veh\_density}_{th} = \frac{CD_{th}}{\text{Msg\_freq\_CR}}$$  \hspace{1cm} (18)

$$p_{NR} = P(X = 0) = e^{-\lambda p}$$  \hspace{1cm} (19)

$$p_{OT} = P(X \geq k) = \sum_{i=k}^{\infty} \frac{e^{-\lambda p}}{i!} = 1 - \sum_{i=0}^{k-1} \frac{e^{-\lambda p} i^{i}}{i!}$$  \hspace{1cm} (20)

$$T_{\text{pwr_adj}} = T_{\text{pwr_adj}}(p_{\text{NR}} + p_{\text{OT}})$$  \hspace{1cm} (21)

$T_{\text{relay_choice}}$ is the time spent accessing to the node table and extracting the relay corresponding to the one which maximizes the objective function in (P2.2). This time can be neglected compared to the power adjustment delay. Note that $T_{\text{pwr_adj}}$ the adjustment delay depends on hardware characteristics and is usually given as a mean delay.

5.2.3. Message transmission delay

The message transmission delay $T_{\text{transmission}}$ here is defined as the necessary delay for a packet or message to be sent on the channel and received by the corresponding relay/sink node. After VDA functionalities establishment, this delay reduces to simply sending a packet of m bytes over a link which has the ability to send at a rate of r bits per second which in turn leads to a transmission delay estimation as follows where $L_x$ and $r$ denote respectively the packet size and the nominal link rate.

$$T_{\text{transmission}} = \frac{L_x}{r}$$  \hspace{1cm} (22)

Using the models expressed here, we can derive two major delay bounds: $T_{\text{min\_bound}}$ and $T_{\text{max\_bound}}$. The minimum delay that can be experienced can be modeled as the case where no adaptation is needed ($T_{\text{delay\_adaptation}} = 0$) and that the VDA procedure succeeds in the first round ($T_{\text{retransmission}} = 0$). In contrast, the maximum delay represents the case where VDA establishment exceeds the maximum allowed rounds inducing $n$ retransmission and with the need of at least one adaptation round. Thus, $T_{\text{min\_bound}}$ and $T_{\text{max\_bound}}$ can be expressed respectively as Eqs. (23) and (24).

$$T_{\text{min\_bound}} = T_{\text{contention}} + T_{\text{success}} + T_{\text{transmission}}$$  \hspace{1cm} (23)

$$T_{\text{max\_bound}} = T_{\text{contention}} + T_{\text{retransmission}} + T_{\text{success}} + T_{\text{delay\_adaptation}} + T_{\text{transmission}}$$  \hspace{1cm} (24)

6. Tests results and model validation

In this section, we present the simulations results conducted using NS-2 to compare and evaluate the effectiveness of our approach compared to schemes basic Vehicular Deterministic Access (VDA) and Distributed Coordination Function (DCF) of 802.11p. We implemented MORS scheme over Nakagami-m fading channel and compared its effectiveness while varying proportions in the objective function in different network conditions. Two main performance metrics are evaluated: (1) the packet forwarding delay and (2) the packet reception rate considering; (a) one hop relaying scheme and (b) multi-hop scheme.

6.1. Simulation parameters

We simulated an 8 lane highway (4 in each direction) with 10 vehicles per lane. We implemented FD2C and UM2D over VDA discussed in [15]. Simulations parameters are summarized in Table 3. Tests are presented based on two main sub-sections: (1) one hop results analysis and (2) applying the design to the multi-hop neighborhood. In this section, MORS means MORS adaptation and dissemination schemes implemented over VDA, VDA and DCF means respectively VDA access and DCF access schemes applied to two-hop neighborhood. Six power levels were implemented corresponding to ranges from 50 to 300 m with 50 m variation. We simulated multiple flows in each scenario emanating from vehicle and each of which containing two types of messages: (a) highly prioritized traffic that mimics emergency messages and (b) low prioritized traffic for routine messages on a periodic messaging basis. In the presented results we are only interested in mean delay.
and PRR values for emergency messages as they are the ones targeted in this work.

6.2. Results analysis

The first step in validating the proposed model and approach is to assess its effectiveness in a one-hop operation before looking into its multi-hop operation in the second step. Hereafter, results will be presented based on such a division.

6.2.1. One hop packet reception rate analysis

Packet reception rate is defined as the ratio between packets originating from a source vehicle and the total amount of successfully received packets by the surrounding vehicles in its communication range. As packets reception is of noticeable importance due to the conveyed information value and the impact of vehicles inter-distance on such metric, we checked the system behavior in multiple communication densities and presented such results while varying the considered communication range. Since our scheme needs to estimate the distance of nodes, and noticing that distance estimation based on Nakagami-m propagation model is not possible, we used the Two-Ray-Ground [19] model to calculate distances as indicated in the study in [25]. The work in [25] states that the communication range and the PRR when using 3.5 loss exponent with Two-Ray-Ground model will be comparable to the results when using Nakagami-m (m = 3).

Fig. 2 shows the impact of varying the chosen distance (communication range) over the maximum communication range on the PRR while taking into account transmission ranges up to 300 m. In this figure, we also compared the use of Nakagami-m vs. TwoRayGround. As expected, using the Nakagami-m model extends the messages reachability to distances over those when using the TwoRayGround model but has a more realistic impact on the PRR over all distances. Multiple β values were simulated and the system behavior model in terms of PRR was compared to theoretical Nakagami-m model to significant points where the selected vehicle meets the optimal distance choice using such β values. Conclusions can be made that for low transmission ranges (corresponding to low power usage), the measured PRR meets the Nakagami-m theoretical ones. For high power usage, the PRR measured only meets the theoretical one on distances up to approximately 70% of the maximum achievable range. This is due to the probability of collisions that raises especially those caused by the hidden node problem. Such a problem is persistent on vehicular networks since neither RTS/CTS nor ACK usage is allowed.

6.2.2. One hop delay analysis

The one-hop delay is the time latency between a successful message reception in the nth hop and its transmission to the n + 1th hop. In MORS, such metric includes: (1) channel reservation, (2) power adjustment latency and relay election, and (3) the messages transmission in itself. Fig. 3, hereafter, shows the impact of varying vehicles density on the one hop delay corresponding to multiple β values and compared to the theoretical maximum and minimum bounds calculated as in Section 5.

We notice that such delay is contained between the two bounds and only exceeds them in two cases for densities greater than 100 vehicles per km road. First, for β value equal to 0.7 which is due to retransmission caused by messages delivery failures when using long links combined with high vehicle density. These failures are caused by the channel unreliability in distances close to the maximum achievable range. And second, for β value equal to 0.5, the one hop delay drops under the minimum theoretical bound. This is due to the power adjustment scheme which triggers a power hop that increases the communication range and consequently reduces the delay. The maximum delay deviation per hop does not exceed 1 ms for steady state densities (45–90 vehicle per km road), but in contrast reaches 2 ms for extreme values corresponding to very low and very high vehicles densities.

Table 4 shows the mean delay deviation resulting from simulation compared to maximum and minimum theoretical delay bounds. For maximum delay bound, the mean deviation is about 0.5 ms which represents about 17.5% of the mean one hop delay. Meanwhile, the 0.8 ms minimum delay deviation represents about 29% of the mean one hop delay. This proves that the scheme has a tendency to come close to the maximum bounds barrier which is due to the fact that mobility introduces unpredictable density changes. In such situations, the probability that the scheme triggers a power adjustment as a countermeasure is higher which induces extra-delays.

6.2.3. End to end delay analysis

MORS performance in terms of end-to-end delay was studied and compared to VDA and DCF while varying β values, subjected to various communication densities. Fig. 4(A) shows that, since
Table 4
Maximum theoretical and simulation delay bound deviation function of vehicles density.

<table>
<thead>
<tr>
<th>Vehicles density (veh/km road)</th>
<th>Deviation % max</th>
<th>Deviation % min</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.00012613</td>
<td>-</td>
</tr>
<tr>
<td>30</td>
<td>0.00037273</td>
<td>0.000168628</td>
</tr>
<tr>
<td>45</td>
<td>0.00097356</td>
<td>0.000997323</td>
</tr>
<tr>
<td>60</td>
<td>0.00016019</td>
<td>0.000975331</td>
</tr>
<tr>
<td>75</td>
<td>0.000805443</td>
<td>0.00041155</td>
</tr>
<tr>
<td>90</td>
<td>0.00121231</td>
<td>0.00013237</td>
</tr>
<tr>
<td>100</td>
<td>0.00050073</td>
<td>0.00082216</td>
</tr>
</tbody>
</table>

Fig. 4. (A) End-to-end MORS mean delay compared to other solutions while varying CD. (B) End-to-end MORS mean delay varying β proportionality parameter and subjected to different CD.

MORS integrates a power adjustment scheme, it can overcome the disconnection problem that can happen in light load conditions. MORS can perform data dissemination in densities less than 25K CD (which corresponds to vehicles density down to 10 vehicles per km road) even if the latency is greater than expected. The extra delay is caused by the power adjustment latency, since the power will be increased gradually until the system finds at least one viable relay. In stable condition, VDA overcomes MORS performances, this is due to the relay selection latency while VDA uses a simplest greedy forwarding technique based on the farthest node election. Note that VDA operates only up to two-hop, however, MORS operates in a multi-hop manner up to 1 km distances (at least four hops, using the highest power). In medium and high loaded conditions, MORS outperforms VDA due to its capability to prevent congestion by reducing the number of possible collisions and electing the best available links for messages delivery. DCF is outperformed by VDA and MORS in both high and low load conditions and presents respectively about 46% and 48% excess E2E delay.

MORS clearly overcomes the geo-casting technique since it delivers the message to the destination in 1/10th of the time and without introducing overhead which is by the way necessary for ROVER functionalities since it serves to collect information on neighboring vehicles and on defining the ZOR and ZOF. In fact, ROVER does not seem to be useful for safety messages dissemination in VANETs since it takes over 300 ms for an end-to-end delivery which is three times more than the standard specification. Geo-casting needs a continuous exchange of control messages containing information such as positioning, cluster formation, speed, and heading. In addition to the misused network resources, generally geo-casting techniques need spatial relevance but do not ensure any constraint on delivery delay.

Fig. 4(B) shows the impact of varying β values on the E2E delay. On the one hand, when using high β values such as 0.7, the distance is sub-served over reliability. Is such a case, the dissemination delay will be reduced but links will be subjected to get broken or to multiple fluctuation that can cause packets losses. On the other hand, using lower β values, reduces the probability of packets collisions since it favors reliability, but raises the E2E delay since the system needs numerous hops to attend the 1 km dissemination barrier. In low densities, the adaptation scheme induces an extra-delay for all β values and its impact is inversely proportional to β. While, in medium densities, the system behavior is relatively stable, in high densities, the delay increases especially for extreme β values. This is due to the need of adjustment to overcome the network congestion.

6.2.4. Multi-hop packets reception rate analysis

Fig. 5(A) shows the packet reception rate in various communication densities for MORS, VDA, DCF and two chosen Geo-cast dissemination schemes, i.e. DTSG [23] and ROVER [22]. VDA and MORS outperform DCF as they enhance scheduling by allocating TS to communicating vehicles. MORS integrates an extra-enhancement since it introduces a power adjustment and a new parametric relay election technique. At low communication densities, MORS outperforms VDA only by 10% and reaches a PRR up to 97%, while in medium and high communication densities the performances gap is wider which proves the power adjustment and relay election combination effectiveness over standard VDA. This is due to MORS capability to avoid congestion, adapt the communication density and therefore avoid possible packet losses. This supports the previously presented remark that MORS is particularly efficient in medium and high communication densities.

We notice that ROVER presents a reverse behavior compared to other schemes. It performs better in high load conditions and reaches a reception probability up to 80%. On the other hand, DTSG is only ensuring 60% reliability and only at close range under 150 m. This supports the remark that MORS is particularly efficient in medium and high communication densities and does not induce and extra-overhead thanks to its locally measured metrics.

Fig. 5(B) shows the impact of varying β values on MORS performance. On the one hand, the use of lower β values makes MORS behave more efficiently in terms of PRR. This is due to the choice of closest nodes as relay, while the use of higher β values has the reverse effect by reducing the PRR since it elects the longest links to forward emergency messages. But, on the other hand the use of higher β reduces the overall dissemination delay since it reduces the hop count. Due to these links instability, the resultant PRR is lower when using high β values.

Fig. 5(C) represents a magnification of Fig. 5(B) when using β value equal to 0.1. It shows how high PRR levels can MORS reach in extreme conditions. This demonstrates that for PRR sensitive applications, such an approach can be proven efficient.
over PDR using measurement

Fig. 5. (A) Packet reception rate in different CD conditions. (B) Packet reception rate (mors) varying $\beta$ values subjected to different CD conditions. (C) Packet reception rate (mors/$\beta = 0.1$) while varying CD conditions.

6.2.5. Delay-PRR tradeoff metric

As MORS introduced a tradeoff between the distance over communication range usage and the PRR, we define a new metric to measure its performances and compare them to other approaches using other access schemes. Let us define PDR as the PRR over Delay Ratio which characterizes the forwarding scheme effectiveness function of the two main performance metrics in VANETs.

$$\text{PDR} = \frac{\text{PRR}}{\text{Delay}}$$

To maximize such a metric, a scheme has to reduce the dissemination delay or increase the PRR. However, these two metrics evolve inversely. Reducing the delay means increasing distance over communication range usage and that latter affects the PRR since longer links are less reliable and subjected to signal fluctuations. Fig. 6(A) hereafter presents the resulting PDR metric measurement for some $\beta$ values in MORS compared to VDA and DCF when these schemes use greedy forwarding, up to 1 km. Fig. 6(A) shows the evolution of the resulting PDR while varying the communication density. A global observation is that the PDR decreases when increasing the communication density.

MORS approach is proven to be efficient for medium and high densities in terms of ensured messages reliability and reduced delay. The best performance is ensured when taking into account $\beta$ value equal to 0.5 since the PDR metric ensures values over 50 for approximately all CD values, thus taking into account PRR and distance over range ratio equally proportional. Taking $\beta$ value equal to 0.7 ensures the same behavior as VDA for medium and high communication loads and a reduced gain of approximately 20% compared to equi-proportional metrics choice. Taking $\beta$ value equal to 0.3 ensures a more stable behavior over variable communication density and presents a reduced mean gain approximately the same as when using $\beta$ value equal to 0.7. DCF presents the least efficient approach since it slatters the network performances and only considers distance rather than both metrics which at the end reduces considerably the PDR that can be guaranteed.

Tests were conducted to figure out the percentage of time that UM2D selects the best relay compared to theoretical analysis. We tested every scenario over 10 times in which over 500 messages are exchanged and multiple relays over the dissemination path are elected, and we compared the results to the theoretically deduced ones. Fig. 6(B), hereafter, presents results of such experiment. We noticed that, in light communication densities, the relay selection scheme behaves greatly and elects the best relay 83% of the time. In contrast, in medium and high communication densities, its effi-
outperforms VDA and DCF in terms of end-to-end delay in high communication densities, and in terms of PRR in medium communication densities. MORS is effective for highly congested environments where high communication density results in high number of packets collisions. Additionally, the objective function used in MORS comprises an adaptive proportionality factor which allows it be adjusted to handle differently various type of emergency messages depending on their respective requirements in order to enhance the global scheme performances.

Acknowledgements

This work is part of CooPerCom, an international research project (Canada–France). The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the French Agence Nationale de la Recherche (ANR) for their support to this project.

References


