Multicast Voice Transmission over Vehicular Ad Hoc Networks: Issues and Challenges

Paolo Bucciol, Federico Ridolfo and Juan Carlos De Martin

Abstract— In this paper we analyze the challenges and issues of transmitting multicast voice streams in a Vehicular Ad hoc Network (VANET). The considered scenario focuses on the multicast transmission of multimedia signals (MP3 streams) between a static and a mobile node. We examine various technical solutions to guarantee vehicle-to-infrastructure connectivity, such as software enhancements and directional antennas. To overcome transmission problems, directions on how to build an optimized client-server streaming software suite are given. The proposed suite is tested with various safety messages in a real-world network testbed. Results of the experiments are presented in terms of both network layer (such as packet loss rate) and application layer (Mean Opinion Score) metrics. Results show how the specific context determines the distribution of error patterns. To improve the performance of the transmission by reducing the average burst error length, block interleaving techniques are taken into consideration. The presented results show that a reasoned choice of the hardware and software parameters enables the transmission of multicast vocal messages by means of the standard 802.11b protocol.

Index Terms — Interleaving, Multicast, Multimedia, VANET

I. INTRODUCTION

NOWADAYS, Intelligent Transportation Systems (ITS) base inter-vehicle communication on highly evolved wireless solutions. Several protocols for wireless communication among vehicles have been proposed lately, such as WAVE (802.11p) or its ancestor DSRC [1], [2]. However, 802.11p is expected to become a standard in 2009, and deployment of compatible hardware will take some time. On the other hand, price of portable, 802.11-enabled devices is always lowering, and many car manufacturers already ship vehicles with built-in 802.11 wireless platforms. Thus, in our work we will adopt the 802.11b Wireless Local Area Network Standard.

By means of a wireless interface, vehicles can establish communication links in infrastructured mode (with an external infrastructure statically configured) or in ad hoc mode (dynamically, with both mobile and non-mobile nodes). Vehicles connecting to each other in ad hoc fashion create a particular Mobile Ad hoc NET (MANET), called Vehicular Ad hoc Network (VANET). Among potential applications are, for instance, safety applications, information systems and value-added services such as voice/audio streaming. In all those applications, where information is usually directed to more than one user, multicast communication is very important.

The rest of this paper is organized as follows. In Section II the simulation testbed which has been set up to validate the proposed techniques is presented. Section III presents the most relevant issues and challenges raised by the proposed scenario, while Section IV describes the solutions which have been foreseen basing on the constraints dictated by the specific context. Section V discusses the results, and Section VI draws the conclusions.

II. SIMULATION TESTBED

A. The scenario

We start by considering a typical urban scenario. More specifically, we focus on an intersection controlled by a traffic light. Our scenario is illustrated in Figure 1. The server is a multicast transmitter and is placed next to the intersection. Different antennas (gain, range) are used for the transmission, as explained in Section IV.D. Data is transmitted using the ad hoc mode of the 802.11b standard. In the receiver device, an 802.11b-enabled handheld device (see Section Table I), a client software has to be installed.

![Figure 1. The considered scenario.](image-url)
The performed experiments consist in passing through the intersection with the car and examine how changing a given set of parameters can influence the transmission. The parameter set includes the transmitter antenna, the car speed, the handheld device and the transmitted message.

B. Transmitted data

The voice data which has been transmitted during the experiments consists of several MP3 files containing various safety messages. All messages have been encoded at 192Kbit/s using the LAME MPEG-1 Layer 3 encoder. In order to minimize the bias given by the selection of a specific tuple of language, gender and message content, ten different generic safety messages (such as “traffic light warning”) have been recorded in four different combinations of language (English, Spanish and Italian) and gender (Male or Female), for a total of forty different messages. All messages have the same duration of about two seconds, regardless of language, gender or message content.

A single test consists in driving through the selected intersection with one of the receivers described in Section Table I and recording all the successfully received (and played back) data, in order to be able to perform off-line processing. In case of directional antennas, the handheld device automatically stops receiving packets after crossing the intersection1. In case of omni-directional antennas, the playback is stopped as soon as we move away from the intersection and the SINR drops below the 15dB value, under which the playback results annoying [25].

III. ISSUES AND CHALLENGES

Vehicular networks have intrinsic characteristics, such as interference and multipath effects, which make the problem of data transmission even more challenging than ordinary wireless networks [9],[12]. Among the key factors which most influence the performance of VANET applications are node mobility [4],[10] and the target scenario[16],[18]. Well-known solutions in the field of multimedia streaming, such as dynamic retry limit adaptation, show their limit in the presence of concurrent transmissions, high node density and mobility [11]. Other approaches, such as use of dynamic routing algorithms [7] and IPv6 [5], seem to be promising in the context of multi-hop communication, but are not suited to single-hop communication.

It has been shown that quality of the communication links in ad hoc networks is impaired by a high number of concurrent transmissions [8]. When the same information is of potential interest to many network nodes, such as public utility information or streaming services, multicast techniques are then foreseen in order to avoid saturation of the links due to many one-to-one transmission.

Although transmission of data flows over VANETs is already a challenging issue by itself [6],[19], transmission of multimedia data is even more critical [15],[20], due to hard timing constraints and high bitrates. In critical traffic conditions, the standard 802.11b infrastructured mode is unable to guarantee acceptable QoS [18], and multimedia applications can be successfully run only over ad hoc topologies [9]. However, multicast applications can still use the standard 802.11 Medium Access Control layer, if the information is broadcasted at moderate bitrate (multimedia transmissions are then limited to voice, audio or low quality video) [3]. In this context, directional antennas can be employed to improve the network scalability and the quality of the received stream [13],[14].

IV. ENHANCEMENTS AND CONSTRAINTS

In this Section we examine the various software and hardware enhancements needed to perform multicast transmission of multimedia data in a VANET environment. We also address the constraints required by the specific context. First, we analyze the structure of a client-server software suite to be used in VANET environment. Then, we introduce the reader to the technique of packet interleaving, whose goal is to reduce the length of error bursts when the communication channels are bursty (as in our case, as it will be explained later in this paper). Next, we proceed by analyzing two typical handheld receivers which could be used in this scenario for their size, autonomy and portability. Finally, we detail the characteristics of the antennas selected for the transmission.

A. Software

In order to perform real-time audio/voice multimedia streaming, a client-server software suite compliant with the standard RTP/UDP/IP protocol stack has to be implemented. A commonly used audio/voice codec such as the MPEG-1 Layer 3 (MP3) codec fulfills the needs of low-complexity and low-latency encoding and decoding. In addition to the MP3 encoder, at server side a multicast transmitter has to be implemented too. Typical behaviour of such transmitter includes parsing a MP3 file in input with usual bitrate (ranging from 32 to 192 Kbit/s), encapsulating one MP3 frame for each data packet and then sending them in real time. In this context, then, the terms MP3 frame and packet are equivalent. To further reduce the complexity of the server, the encoder and the transmitter can be included in a single application.

For what concerns the client software, it has to be designed taking into account the fact that receivers commonly used for mobile communications are small and portable. Their drawbacks include, for instance, little computational power and reduce autonomy. Moreover, most of the computational power is used to perform high priority task such as wireless connection management, stream decoding and payback. Such tasks are computationally intensive due to hardware limitations, such as lack of Floating Point Unit (FPU) in many devices and basic Direct Memory Access (DMA) engine which relies on hard use of the CPU also for the playback on the sound subsystem.

1 Some packets can still be received, though, due to secondary lobes or reflections.
Keeping in mind those (hard) constraints, enhancements to the client software shall be targeted to design simple and low-complexity error concealment techniques, packet reordering techniques and glitch reduction techniques. G McIntosh reduction and error concealment have already been considered in [25]; in this work we will use the frame copy error concealment technique and 15dB as the SINR threshold to play back the received stream. For what concerns packet reordering, the delivery of data on multipath wireless channels, such as urban VANETs, foresees the inclusion of a packet reordering algorithm in the client device. Such algorithms are usually based on ordered information coming from the sender, such as RTP sequence numbers. The algorithm implemented in the WiSafety project [22] uses a dynamic buffer to store received packets and send the right packet to the decoder. Simple wait algorithms, in which the decoder waits until the packet with the right RTP sequence number is received, perform poorly, since they cause glitches and other annoying sound artifacts in the playback. However, a more efficient wait algorithm could wait for delayed packets without impairing the quality of the transmission by stretching the playback time of the received packets already present in the buffer. A potential issue of this approach is related to the connectionless time of the communication channel, which in an urban VANET is relatively long compared to the packet decoding time. The client-server software suite used during the experiments, available as both source code and pre-compiled images for handheld devices at [22], has been developed in the context of the LScube project [23].

B. Packet interleaving

To improve the quality of the transmission, the adoption of error protection techniques is foreseen. Since multicast transmission does not employ a posteriori packet retransmission techniques, such as Automatic Retransmission reQuest, a priori and scalable error protection techniques such as Forward Error Correction (FEC) are foreseen to reduce the packet error rate. However, it has been shown in [28] that correlated packet losses (bursts), from which 802.11 communication links intrinsically suffer [26], reduce the effectiveness of FEC. Performance of systems can actually be impaired by high average burst error length [29].

In Figure 2, which refers to our experimental testbed, the distribution of transmission errors as a function of burst error length is plotted. In the case of the 8dBi antenna, more than 50% of packet losses are caused by error bursts longer more than 10 packets long. This effect is mitigated by the use of higher gain antennas.

Moreover, packet redundancy does not help in improving the performance, since it just injects even more overhead into the network. The use of a packet interleaver is then foreseen to drastically reduce the average length of packet error bursts [27]. A simple and low-complexity block interleaver is represented in Figure 3. A generic block interleaver reads the data in $d$ rows (depth of the interleaver), until $s$ columns (size of the interleaver) have been filled. The data is then transmitted in columns, following the numbering order shown between parenthesis in the example. The packet which suffers from the highest transmission delay is the one on the upper right, in this case packet 3, which is transmitted in $7^{th}$ position. In general, the maximum transmission delay $id$ of a generic interleaver of size $nxd$ can be computed as

$$id = d \cdot (s-1) + 1 - V_{1,s},$$

where $V_{1,s}$ is the original number of the element in the $(1,s)$ position (3 in the example). Thus, the interleaver delay $id$ of the example interleaver is equal to 4.

As a generic rule-of-thumb, a block interleaver of size $sxd$ allows to split error bursts sized up to $s$ packets into isolated losses. The drawback of this simple, although efficient, mechanism, is the increased transmission delay. In our case, each packet contains a MP3 frame, which is composed by 1152 PCM samples. At a sampling frequency of 44.1KHz, the time duration of each packet/frame is about 26.1ms. Following the ITU-T recommendation for voice communication [30], 150ms is the maximum overall delay allowed for high quality real-time vocal communication, while delays greater than 250ms make the playback annoying for the end user. In our case (single-hop multicast transmission of low-birate data), the transmission delay is almost negligible. Assuming that no other delays come from the processing and coding/decoding tasks (streaming of stored data and real-time decoding and playback), the computation of the $s$ and $d$ parameters of a block interleaver which fulfills the ITU-T recommendations is straightforward.

C. Receivers

The performance of our transmission suite has been tested with two different handheld receivers. The first receiver, an HP iPAQ h5550, has been configured with a Linux Familiar 0.8.2 distribution, based on the version 2.4 of the Linux kernel.

![Figure 2. Packet losses as a function of burst error length.](image)

![Figure 3. Example of a block interleaver.](image)
TABLE I

<table>
<thead>
<tr>
<th>Device n.</th>
<th>Characteristics</th>
<th>Gain (dBi)</th>
<th>Beam Width</th>
<th>Front to Back Ratio (dB)</th>
<th>Price (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>24</td>
<td>12°</td>
<td>&gt;30</td>
<td>180</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>18</td>
<td>30°</td>
<td>26</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>8</td>
<td>omnibl</td>
<td>-</td>
<td>15</td>
</tr>
</tbody>
</table>

The h5550’s CPU is an Intel PXA255, running at 400MHz and not equipped with WPU. The second receiver, a Sharp Zaurus SL-C 3200, has been configured with the OpenZaurus 3.5.4.1 distribution, based on the Linux 2.6 kernel. The SL-C 3200’s CPU is an Intel PXA270, equipped with integrated FPU and running at 400MHz.

Both devices are provided with an 802.11b-compliant wireless card. An Atheros wireless chipset is integrated in the h5550, while the Zaurus has been equipped with a D-Link DCF-660W Compact Flash card. We haven’t employed external antennas at the receiver side, since we were interested in setting up an easy-to-replicate environment, without the need of setting up anything more than a car cradle on the windscreen of the vehicle.

D. Antennas

Details of the antennas used for transmission are shown in Table I. We consider one omni-directional antenna with 8dBi gain, and two directional antennas, with different characteristics: one highly directional grid antenna with 24dBi gain and one panel antenna with 18dBi gain.

V. RESULTS

To validate the proposed framework, real experiments have been conducted in the city of Torino, Italy. The experiments have been conducted in two different scenarios: a day scenario and a night scenario. During the day, the wireless link suffers from the presence of traffic jams, intense traffic and cars parked at the side of the streets. During the night, traffic density is lower, car speed is more constant and less cars are parked along the streets.

A. Network layer results

In the experiments, three different car speeds have been considered, namely 40, 50 and 70 Km/h (25, 31 and 43 Mph). All the three antennas described in Section IV.D have been tested during the experiments. Both handholds described in Section Table I have been used as receivers, but no appreciable performance gap among them was observed. The tested combinations of antenna gain, car speed and scenario are presented in Table II.

For each combination, a minimum of 8 transmission tests for each vocal message (detailed in Section II.B) have been performed. Some combinations, like [8dBi, 70Km/h, day scenario], have not been tested since they wouldn’t have added useful information. As shown in Figure 4, during the day high gain antennas perform consistently better than lower gain antennas.

The only test conducted with the 8dBi antenna in the day scenario suffers from an unaccept able packet loss rate which exceeds 20%. Higher transmission gain (up to 24 dBi) has then to be employed in order to obtain acceptable packet loss rates. On the other hand, in the night scenario, as shown in Figure 5, almost negligible packet loss rates are obtained when moderate antenna gain (18dBi) is employed, thus avoiding the adoption of more costly antennas such as the 24dBi parabolic antenna.

However, the use of highly directional antennas helps in reducing the number of duplicated packets. Vehicular networks intrinsically suffer from the problem of duplicates, caused mainly by multipath effects. This is even more clear in the urban scenario, where duplicate packets sometimes exceed the 10% of received packets. The fact that the percentage of duplicates varies highly from day to night, leads to a clear connection between the duplicates and the number of cars (running and parked). To mitigate this effect, especially during the day, wide beam antennas should be avoided. As shown in Figure 4 and Figure 5, the wider the antenna beam, the stronger the possibility of interferences.
B. Application layer results

The ITU-T implementation of the PESQ algorithm [21] has been used to test the performance of the proposed framework. For each test, the MP3 stream received by the client has been decoded and stored for off-line processing. To evaluate the quality of the stream, it has been decoded and resampled at 16KHz, mono, 16 bits per sample, in order to fulfill the requirement of the PESQ quality test. The resulting score has then been mapped to a MOS scale following the ITU-T P.862.1 recommendation [24], whose values range from 1 (worst) to 5 (best).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Language</th>
<th>Average MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Italian</td>
<td>3.05</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>3.08</td>
</tr>
<tr>
<td>Female</td>
<td>Spanish</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>3.14</td>
</tr>
</tbody>
</table>

The low variance obtained for all the points of the graphs presented hereafter (<0.4 in MOS scale for each set of [car speed, antenna gain, scenario]) makes us confident on the reliability of the presented results.

The overall results are presented in Figure 6 (day scenario) and Figure 7 (night scenario). Unacceptable results, in terms of MOS scale, have score lower than 2.5. It is then confirmed what anticipated previously, that is, during the day it is necessary to employ high gain antennas in order to obtain a satisfactory quality, since with the 8dBi antenna acceptable quality cannot be achieved. Such antenna leads to unacceptable performance also in the night scenario, while very good quality is in this case achieved by the 18dBi antenna (due to the very low packet loss rates experienced). Table III shows the average MOS for each combination of language and gender. It is worth noting that messages pronounced by a female speaker experience a higher average MOS, and this is almost independent from the specific language.

C. Residual burst error length with block interleaving

In our experiments, we have considered block interleavers which introduce a delay of less than 250ms, according to the ITU-T recommendation G.114. All the results presented in this section refer to the average of the performance achieved in day and night scenario, respectively. Figure 8 and shows how the burst error length can be reduced by employing various block interleaving matrices which induce different delays. The starting burst error length corresponds to the value on the extreme left of the curve and is equal to 2.83 packets for the day scenario and 2.33 packets for the night scenario.

It comes to the eye that the trends of the two curves are different. Since the night scenario (bottom one) starts from a moderate average burst error length, it saturates when the interleaver delay is about 100ms. It implies that the performance which can be achieved by means of interleavers for high-quality real time communication (<150ms) cannot be further improved by using more complex interleavers which introduce even more delay. On the other side, the average burst error length of the day scenario decreases monotonically up to a delay of more than 200ms. It implies that it is not possible to achieve high-quality real time communication and obtain the best achievable performance from the interleaving technique at the same time.

VI. CONCLUSIONS

In this paper we analyzed the performance of multicast transmission of vocal messages in a Vehicular Ad hoc
Network (VANET). We focused on the transmission of MP3 streams in a typical urban context. We examined different solutions for guaranteeing vehicle-to-infrastructure connectivity, ranging from software enhancements to the use of omni-directional or directional antennas. To make performance evaluation of the proposed solutions, we developed an optimized client-server streaming software suite and tested it in a real-world network testbed. Different vocal messages, recorded in various combinations of language and gender, have been transmitted. Results showed that hardware and software must be chosen accurately to successfully transmit multicast vocal messages with acceptable quality, and such choice strongly depends on the specific scenario (such as the presence of traffic). Results also showed that the distribution of packet errors can be modelled once the context is known, and appropriate interleaving techniques can be used to guarantee high-quality real time communication almost free from long error bursts. Finally, we showed that the perceived quality depends on the speaker’s gender, more than the speaker’s language.

ACKNOWLEDGMENT

The authors wish to thank Maria Teresa Medina Quintana and Dr. Fabio De Vito from Politecnico di Torino, Italy, for their precious help during the experiments.

REFERENCES

[23] LScube Project. URL: http://live.polito.it/.