Marion Berbineau Magnus Jonsson
Jean-Marie Bonnin Soumaya Cherkaoui
Marina Aguado Cristina Rico-Garcia
Hassan Ghannoum Rashid Mehmood
Alexey Vinel (Eds.)

LNCS 7865

Communication Technologies for Vehicles

5th International Workshop, Nets4Cars/Nets4Trains 2013
Villeneuve d'Ascq, France, May 2013
Proceedings

NETS 4 Cars Trains

Springer

Lecture Notes in Computer Science 7865

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Marion Berbineau Magnus Jonsson
Jean-Marie Bonnin Soumaya Cherkaoui
Marina Aguado Cristina Rico-Garcia
Hassan Ghannoum Rashid Mehmood
Alexey Vinel (Eds.)

Communication Technologies for Vehicles

5th International Workshop, Nets4Cars/Nets4Trains 2013
Villeneuve d'Ascq, France, May 14-15, 2013
Proceedings

Springer

Volume Editors

Marion Berbineau, IFSTTAR, LEOST, Villeneuve d'Ascq, France
E-mail: marion.berbineau@ifsttar.fr

Magnus Jonsson, Halmstad University, Sweden
E-mail: magnus.jonsson@hh.se

Jean-Marie Bonnin, Telecom Bretagne, Cesson Sévigné, France
E-mail: jm.bonnin@telecom-bretagne.eu

Soumaya Cherkaoui, Sherbrooke University, Canada
E-mail: soumaya.cherkaoui@usherbrooke.ca

Marina Aguado, University of the Basque Country, Bilbao, Spain
Email: marina.aguado@ehu.es

Cristina Rico-Garcia
German Aerospace Center, Oberpfaffenhofen-Wessling, Germany
E-mail: cristina.ricogarcia@dlr.de

Hassan Ghannoum, SNCF, Paris, France
E-mail: hassan.ghannoum@sncf.fr

Rashid Mehmood, Huddersfield University, UK
E-mail: r.mehmood@gmail.com

Alexey Vinel, Tampere University of Technology, Finland
E-mail: vinel@ieee.org

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-37973-4 e-ISBN 978-3-642-37974-1
DOI 10.1007/978-3-642-37974-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013935956

CR Subject Classification (1998): C.2, C.3, C.4, I.6

LNCS Sublibrary: SL 5 – Computer Communication Networks
and Telecommunications

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Communication Technologies for Vehicles Workshop series provides an international forum on the latest technologies and research in the field of intra- and inter-vehicles communications and is organized annually to present original research results in all areas related to physical layer, communication protocols and standards, mobility and traffic models, experimental and field operational testing, and performance analysis.

First launched by Tsutomu Tsuboi, Alexey Vinel, and Fei Liu in Saint Petersburg, Russia (2009), Nets4Cars-Nets4Trains workshops have been held in Newcastle-upon-Tyne, UK (2010), Oberpfaffenhofen, Germany (2011), and Vilnius, Lithuania (2012). These proceedings contain the papers presented at the 5th International Workshop on Communication Technologies for vehicles (Nets4Cars and Nets4Trains 2013), which had dedicated tracks for road-and-rail-based approaches and took place in Villeneuve d'Ascq, at IFSTTAR, France, in May 2013, with the technical support of GRRT (Groupement regional recherché Transport) and CISIT (International Campus on Safety and Intermodality in Transportation).

Our call for papers resulted in 24 submissions. Each of them was assigned to the Technical Program Committee members and 17 submissions were accepted for publications (12 for the road track and 5 for the rail track). Each accepted paper got at least two independent reviews. In addition, three invited papers were accepted. The order of the papers in these proceedings corresponds to the workshop program.

We extend a sincere "thank you" to all the authors who submitted the results of their recent work, to all the members of our hard-working comprehensive Technical Program Committee, as well as the thoughtful external reviewers. Also, we extend a special "thank you" to Yann Cocheril for the preparation of the proceedings and the website. We invite all the experts in the field to join us in Offenburg, Germany, for Nets4Cars-Nets4Trains 2014.

May 2013

Marion Berbineau
Magnus Jonsson
Jean-Marie Bonnin
Soumaya Cherkaoui
Marina Aguado
Cristina Rico-Garcia
Hassan Ghannoum
Rashid Mehmood
Alexey Vinel

Organization

Workshop Organizers

General Co-chairs

Berbineau Marion	IFSTTAR, France
Vinel Alexey	Tampere University of Technology, Finland

TPC Co-chairs (Nets4trains)

Aguado Marina	University of the Basque Country, Spain
Ghanoum Hassan	SNCF, France
Rico-Garcia Cristina	DLR, Germany

TPC Co-chairs (Nets4cars)

Bonnin Jean-Marie	Telecom Bretagne, France
Cherkaoui Soumaya	Sherbrook University, Canada
Jonsson Magnus	Halmstad University, Sweden

Steering Committee

Li Xu	State University of New York, USA
Molinaro Antonella	UNIRC, Italy
Rodrigues Joel	University of Beira Interior, Portugal
Sikora Axel	University of Applied Sciences Offenburg, Germany
Strang Thomas	DLR, Germany
Tsuboi Tsutomu	Hamamatsu Agency for Innovation, Japan
Zhang Yan	Simula Research Lab, Norway

Technical Program Committee

Adín Iñigo	CEIT, Spain
Altintas Onur	Toyota InfoTechnology Center, Japan
Aniss Hasnaa	IFSTTAR, France
Belghith Abdelfettah	ENSIT - HANA Research Group, Tunisia
Belimpasakis Petros	Bang & Olufsen, Germany
Boeglen Hervé	University of Haute Alsace, France
Boillot Florence	IFSTTAR, France
Bondi André	Siemens Corporate Research, USA
Boucadair Mohamed	Orange, France
Buburuzan Teodor	Volkswagen, Germany
Caleffi Marcello	UNINA, Italy

Campolo Claudia	UNIRC, Italy
Cerqueira Eduardo	UFPA, Brazil
Curado Marilia	University of Coimbra, Portugal
Cocheril Yann	IFSTTAR, France
Daher Robil	University of Rostock, Germany
Dayoub Iyad	IEMN-DOAE, France
Delot Thierry	UVHC, LAMIH, France
Elfaouzi Nour-eddin	IFSTTAR, France
Ernst Thierry	INRIA, France
Festag Andreas	Nec Labs, Germany
Filali Fethi	QMIC, Qatar
Gantsou Dhavy	UVHC, LAMIH, France
Garcia Francisco	Agilent Labs, UK
Geller Benoît	ENSTA Paris tech, France
Gozalvez Javier	UMH, Spain
Gransart Christophe	IFSTTAR, France
Gusikhin Oleg	Ford Research & Advanced Engineering, USA
Härri Jérôme	EURECOM, France
Haziza Nathalie	THALES, France
Heijenk Geert	University of Twente, The Netherlands
Hilt Benoit	University of Haute Alsace, France
Imran Muhammad Ali	University of Surrey, UK
Kandeepan Sithamparanathan	RMIT University, Melbourne, Australia
Kassab Mohamed	IFSTTAR, France
Knopp Raymond	EURECOM, France
Koucheryavy Yevgeni	Tampere University of Technology, Finland
Laouiti Anis	Telecom SudParis, France
Lehner Andreas	DLR, Germany
Lüeddecke Katrin	DLR, Germany
Marais Juliette	IFSTTAR, France
Mendizabal Jaizki	CEIT, Spain
Mottier David	MERCE, France
Park Byungkyu (Brian)	University of Virginia, USA
Peyret François	IFSTTAR, France
Pignaton Edison	University of Brasilia, Brazil
Renaudin Valérie	IFSTTAR, France
Roullet Laurent	Alcatel-Lucent, France
Ruiz Sandoval Elia	Berner & Mattner Systemtechnik GmbH, Germany
Santi Paolo	National Research Council, Italy
Senouci Sidi Mohammed	University of Bourgogne, France
Simoens Sébastien	ALSTOM, France
Simon Eric	IEMN-TELICE, France

Sondi Patrick	IFSTTAR, France
Vazão Teresa	Instituto Superior Técnico, Portugal
Vétillard Jean-Noël	ALSTOM, France
Wahl Martine	IFSTTAR, France
Wetterwald Michelle	EURECOM, France
Wisitpongphan Nawaporn	Carnegie Mellon University, USA

Additional Reviewers

Owezarski Philippe	LAAS-CNRS, France
Sambo Yusuf	University of Surrey, UK

Hosting Institution

IFSTTAR, COSYS-LEOST, Villeneuve d'Ascq, France

Organizing Committee

Bourbotte Daniel	IFSTTAR, France
Cocheril Yann	IFSTTAR, France
Davoust Corinne	IFSTTAR, France
Delsinne, Bernard	IFSTTAR, France
Kassab Mohamed	IFSTTAR, France
Masson Émilie	IFSTTAR, France
Saint-Saens Isabelle	IFSTTAR, France
Schellaert Valérie	IFSTTAR, France
Sondi Patrick	IFSTTAR, France

Sponsoring Institutions

IFSTTAR, France
Tampere University of Technology, Finland
GRRT, France
CISIT, France
I-Trans, France

Table of Contents

Invited Papers

Road Track

Rail Track

V2V Communication Channels: State of Knowledge, New Results, and What's Next

David W. Matolak

Department of Electrical Engineering, University of South Carolina, Columbia, SC
matolak@sc.edu

Abstract. This paper surveys the field of vehicle-to-vehicle (V2V) communication channels. Motivated by intelligent transportation systems and vehicular safety, V2V research has proliferated in recent years. We provide a short description of V2V communication systems, and the importance of key channel parameters. This is followed by a discussion of basic channel characteristics—the channel impulse response and channel transfer function, and their statistical description—and how V2V channels differ from the more familiar cellular radio channel. Modeling of the V2V channel is covered by a review of the literature on V2V channels, addressing path loss, delay spread, and Doppler spread. We describe the two most popular methods for modeling V2V channels, tapped-delay line models and geometry-based models, then briefly discuss multiple-antenna channels and the crucial V2V channel characteristic of non-stationarity. A potential channel classification scheme for V2V channels is given, and some recent results on the channel within parking garages, and on sloped terrain, are provided. We end the paper with a short discussion of what may come next in this vibrant field.

Keywords: vehicle to vehicle communications, propagation.

1 Introduction

Vehicle to vehicle (V2V), vehicle to infrastructure (V2I) and vehicle to roadside (V2R) communication systems research has grown tremendously in recent years, e.g., [1], [2]. From initial studies approximately ten years ago [3], and early work on protocols [4], research has burgeoned in the past five years to several hundred papers. Conference papers on inter-vehicle network simulations alone number over one hundred in the period 2009-2012 [5], and a recent survey on V2X[1] channels alone also had a reference list of over one hundred citations [6].

Although there were some earlier papers on channel characteristics and performance aspects of V2V communication systems, e.g., [7]-[10], those were somewhat isolated studies, whereas in the past five to seven years, interest in V2X communications has gained momentum among governments, industries, and academia across the

[1] We use abbreviation V2X to denote any/all among V2V, V2I, V2R, and use the individual abbreviations when we intend to be specific.

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 1–21, 2013.

world. The primary context for this is V2X as part of intelligent transportation systems (ITS) [11]. The scope of ITS is broader than V2X, as it encompasses railway, maritime, and aeronautical transportation systems as well, but it was not until government and industry announced programs (and allocated spectrum) that focused academic research began to grow. Motivations for ITS include increased system efficiency ("green" transportation), reduced transportation delays, economic growth, passenger entertainment, and most importantly, safety. This is most acute in V2X, as automobile traffic accidents still claim thousands of lives each year in large developed nations, e.g., [12].

Even limiting ourselves to only the V2X communications field, the area's breadth is substantial [13], hence as in other complex systems such as cellular radio, investigators employ reductionism and separate the system design into multiple parts, each of which can be addressed separately. The ISO communications protocol stack [14] is the most common and convenient method for such separation. In that protocol stack, the lowest two layers—the physical (PHY) and medium access control (MAC)—have arguably seen the most attention by V2V researchers. The PHY layer, or more specifically one key component of the PHY layer, the V2V communication channel, is the focus of this paper.

Since research began on the V2V channel (possibly [7]), it has been recognized that the V2V channel is distinct from that of many typical communication system channels. The closest comparison may be to the cellular channel, and the main distinguishing features of the V2V channel in comparison to that channel are that in V2V channels, (i) antenna heights of both transmitter (Tx) and receiver (Rx) are low, and (ii) both Tx and Rx are mobile [15]. As a consequence of these differences, V2V channels can have the line of sight (LOS) between Tx and Rx obstructed more frequently, and "scattering" is often non-isotropic. With both Tx and Rx moving, channel variation rates can also be larger than in cellular, or in other words, the V2V channel (modeled as random) is statistically stationary for a shorter time period than in cellular. In addition, due to multiple scattering or rapid time variation, in some cases amplitude fading may be more severe than in the most common (Rayleigh) cellular fading model [16].

In any complicated propagation environment, for digital communications purposes, there are several channel parameters that are most important to quantify for their effect on system performance. These include path loss, delay dispersion, and Doppler spread [17]. Path loss is also known as transmission loss or attenuation, and represents the power loss between the Tx and Rx antennas. For a given value of transmit power, path loss determines link range. Delay dispersion quantifies the extent of the V2V channel impulse response (CIR), most often using the root-mean-square (RMS) delay spread (DS) σ_τ, although other measures such as the delay window W_x or delay interval I_Y are also used [18]. Delay spread is reciprocally related to frequency selectivity, quantified as the coherence bandwidth B_{coh}, hence for wideband signals these (~equivalent) parameters can be used to select equalizer structure for single-carrier signals or subcarrier bandwidth and cyclic prefix in multicarrier systems. The Doppler spread f_D quantifies the channel's spreading effect upon a transmitted tone in the frequency domain due to motion; this is reciprocally related to the channel's coherence time t_{coh}, which is roughly a measure of how long in the time domain the channel's statistics remain approximately constant. These dispersion and coherence parameters are all based upon the classic analysis by Bello on randomly time-varying channels [19]. Worth pointing out is that these parameters are, for any given setting, single numbers only, hence we can not expect to glean from them all that is required for a complete description of the channel. These parameters are summary measures derived from multi-dimensional correlation functions of the time-varying CIR $h(\tau,t)$ and its Fourier transform, the time-varying channel transfer function (CTF) $H(f,t)$. The use of these single channel parameters in characterizing rapidly temporally- (and spatially-) varying channels must be done with care, hence some authors also gather statistics of "instantaneous" CIRs and CTFs instead of statistics derived from averages of these responses [20].

In characterizing the cellular channel, researchers recognized the need to classify the channel into multiple categories based upon the local physical environment. The same holds true for the V2V channel, and as research has progressed, the V2V channel classes themselves have evolved. The earliest V2V channel work used channel classifications similar to those used for cellular, i.e., urban, suburban, and rural. Yet as V2V applications were studied further, additional environments, e.g., street intersections [21] and tunnels [22], were deemed sufficiently important to warrant their own channel characterization. Sub-classes that account for vehicle density and street geometry are also employed. These unique classifications reveal additional distinction for the V2V channel, which will be of importance in any standardization efforts.

In this paper we summarize the body of work on V2V channels. For some context, Section 2 provides a short discussion of planned V2V communication system features, cites summary papers published on V2V channels, and provides some basic V2V channel descriptions. Section 3 contains more detail on the various V2V channel model types that have been proposed and studied in the literature. We also include a draft version of our V2V channel classification scheme. Section 4 contains some examples of new results for specific V2V settings: the parking garage and sloped terrain. In Section 5 we provide our thoughts on "what's next" in the characterization of V2V channels, and Section 6 is the conclusion.

2 State of Knowledge

2.1 V2V Communication Systems

Since the allocation of frequency spectrum for V2X applications in the 5 GHz band (5.9 GHz in the US and 5.7 GHz in Europe), most studies of V2X communication systems have focused on this region of spectrum, although other bands may be used in other parts of the world (e.g., the 700-MHz band in Japan, [23]). Measurements have also been taken in the 900 MHz [9] and 2.4 GHz bands [24], [25]. In all cases, V2V settings are established roadways in urban, suburban, or rural environments.

The current air interface or communication waveform is defined in the IEEE 802.11p V2V standard [26], a modification to the 802.11a wireless local area network standard. The V2V scheme is also termed Dedicated Short Range Communication (DSRC), and along with the associated MAC-layer IEEE 1609 standards, the group of standards is often termed Wireless Access for Vehicular Environments (WAVE) [27]. Since based upon the 802.11a standard, the waveform is a multicarrier orthogonal frequency-division multiple access (OFDMA) signal that employs various signaling alphabets, and multiple user access is controlled by time-division within 10 MHz channels, yielding data rates from approximately 3 Mbps to 27 Mbps [28]. Packet duration is a key parameter, and this depends upon the packet type. With a typical transmit power level of up to approximately 29 dBm [29], DSRC link range is planned to be less than 1 km.

2.2 Basic V2V Channel Characteristics

Our survey here aims to include representative references across the literature, but will not be exhaustive. As mentioned, the V2V channel will be short range (~few meters to 1 km), employs low-height antennas, and has both Tx and Rx mobile. These features have led investigators to pursue multiple types of studies, from experiments and measurements over a range of environments to specific detailed studies of important elements in the V2V link that affect channel characteristics, e.g., antennas [30].

Several summary papers have also been published on V2V channels. This includes [6], [15], and [31]-[35]. In most cases, what is being characterized is the V2V CIR, or its equivalent, the CTF. The multipath V2V CIR can be expressed as follows:

$$h^{(c)}(\tau,t) = \sum_{k=0}^{L(t)-1} z_k(t)\alpha_k(t)\exp\{\,j[\omega_{D,k}(t)(t-\tau_k(t)) - \omega_c(t)\tau_k(t)]\}\delta[\tau-\tau_k(t)] \qquad (1)$$

where $h(\tau,t)$ is defined as the response of the channel at time t to an impulse input at time $t\text{-}\tau$ (Bello's input delay spread function [9]). This CIR is a sum of $L(t)$ "discrete impulses" via the Dirac deltas, which while an approximation—interpreted as the channel imposing specific discrete attenuations, phase shifts, and delays upon any signal transmitted—is very good for signal bandwidths of tens of MHz or more. Variable α_k in (1) is the k^{th} path amplitude, and the argument of the exponential is ϕ_k, the k^{th} resolved phase, with τ_k the k^{th} path delay. Carrier frequency is $f_c=\omega_c/(2\pi)$, and the k^{th} resolved Doppler frequency is $f_{D,k}= \omega_{D,k}/(2\pi)$ where (assuming LOS or "single-bounce" reflections) $f_{D,k}(t)=v(t)f_c\cos[\theta_k(t)]/c$, where $v(t)$ is relative velocity between Tx and Rx, $\theta_k(t)$ is the aggregate phase angle of all components arriving in the k^{th} delay "bin," and c is the speed of light. The delay bin width is approximately equal to the reciprocal of the signal bandwidth, e.g., for a 10 MHz signal, bin width is 100 nanoseconds—components separated in delay by an amount smaller than the bin width are "unresolvable." The k^{th} resolved component in (1) thus often consists of multiple terms ("subcomponents") from potentially different spatial angles $\theta_{k,i}$. We do not address "spatial" channels, e.g., [36], [37], in detail here, but will briefly discuss the use of multiple antennas in a subsequent section; for this case, (1) can represent the CIR between any given Tx-Rx antenna pair.

Note that the form of (1) is generalized from that typically seen in texts, e.g., [38], in that it allows for a channel class (superscript "*c*" on *h*), a time-varying number of transmission paths $L(t)$, and a "persistence process" $z(t)$ that accounts for the finite "lifetime" of propagation paths. Any explicit V2V channel models may not make use of all these features. Also note that the amplitudes (α's) and phases represent aggregate channel effects. This is illustrated conceptually in Fig. 1, where the path numbered "0" is the direct, possibly LOS path, paths "1-4" are single-reflection paths, and path "5" is a double reflection, so the parameters of (1) must effectively model the cumulative effect of all transmissions, reflections, absorptions, etc.

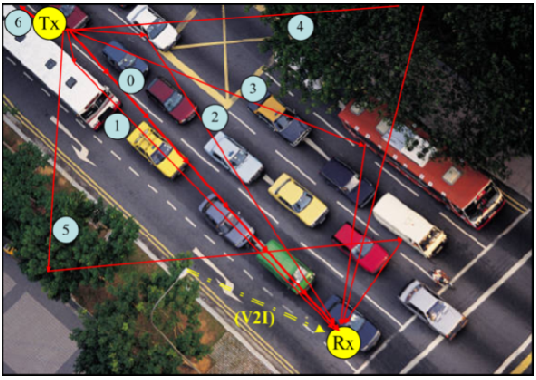


Fig. 1. Example V2V channel, urban area

The CTF corresponding to (1) is

$$H^{(c)}(f,t) = \sum_{k=0}^{L(t)-1} z_k(t)\alpha_k(t)e^{j2\pi f_{D,k}(t)(t-\tau_k(t))}e^{-j2\pi f_c\tau_k(t)}e^{-j2\pi f\tau_k(t)} \qquad (2)$$

All exponentials in (2) are functions of time; the last exponential expresses the time-varying frequency dependence. The second exponential can change significantly with small delay ($\tau_k(t)$) changes when f_c is large, e.g., nanosecond delay changes (corresponding to ~30 cm path length changes) cause more than 2π shifts when $f_c \geq 1$ GHz. This second term typically dominates the small scale fading variation (spatially, on the order of a wavelength λ), since $f_c >> f_{D,k}$. For example, for a 5.9 GHz V2V carrier frequency, if relative velocity is 63 m/s (roughly 140 miles/hour), $f_{D,max}$=1239 Hz<<f_c.

The CIR in (1) evinces variation in both delay τ and time t. This type of CIR is implementable in analysis, simulations, or hardware in terms of the popular tapped-delay line (TDL) model. The TDL is a linear, finite IR filter, which for the V2V case should have time-varying tap weights, and potentially a time-varying number of taps $L(t)$. The TDL structure is shown in Fig. 2, where the k^{th} transmitted symbol is denoted s_k, and the k^{th} channel output sample r_k.

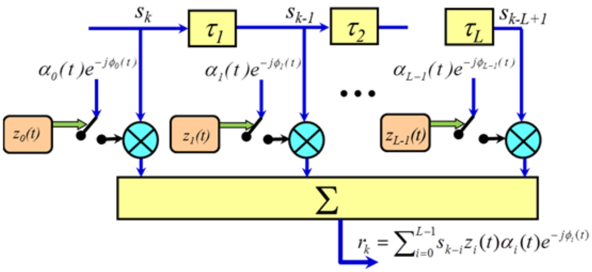


Fig. 2. Time-varying tapped-delay line model of CIR

When either the CIR or CTF is measured or determined via computer simulations, the received power can be found by summing the multipath component (MPC) amplitudes (α_k^2 in (1) and (2)), and with a known transmit power, antenna gains, and RF cable losses, channel path loss can be computed. The function obtained as α_k^2 versus delay is called the power delay profile (PDP). Delay-domain statistics such as the RMS-DS σ_τ are computed over a local area large enough to enable a sufficient number of PDPs, but small enough to ensure stationarity—choosing such an area is hence a balance between these competing constraints. Some researchers have tried to quantify the so-called "stationarity interval" or "stationarity distance" [39], and this is environment- or "channel-class-," and frequency-specific. Alternatively one may collect statistics on the instantaneous RMS-DS [20]. Similar comments pertain to estimation of RMS Doppler spreads.

To obtain the summary parameters (σ_τ and f_D) analytically one must compute correlation functions of (1) and/or (2) which requires ensemble averages. In practice only time averages can be obtained, although some authors make assumptions regarding distributions (e.g., of amplitudes and phases) to obtain computable expressions for the correlation functions. As with measurements of time averages, such assumptions will pertain only over suitably defined stationarity regions. Such computations can be done numerically in geometry-based models, which we discuss subsequently. These correlation functions, when reduced to functions of one variable by virtue of the wide-sense stationary (WSS) and/or uncorrelated scattering (US) assumptions, are described in detail in [19], [38], and [40], to which the interested reader is referred for background.

No well established theoretical models exist for the V2V channel PDP $P_h(\tau)$ or its transform, the spaced-frequency correlation function $P_H(\Delta f)$, although there has been a nice analytical derivation for the spaced-time autocorrelation function $R_H(\Delta t)$, from which the coherence time can be estimated [41] given some geometric parameters. The lack of an analytical PDP is true for the cellular channel as well, simply because a wide range of PDP shapes with a range of delay spreads is known to exist in practice. Common shapes for average PDPs include the uniform, and the more realistic exponential decay.

Based on this discussion, statistically *non*-stationary (NS) channel modeling is seeing sustained interest for the V2V setting. We discuss NS considerations in the following section in the context of explicit models for the V2V channel.

3 V2V Channel Modeling

If we begin from the analogy to cellular radio channels, what is required for a useful V2V channel model is three component models for (i) path loss, (ii) shadowing, and (iii) MPC fading. Based upon the physical causes of these components, the rates of time-variation of these effects increase as we go from (i) to (ii) to (iii). Thus in the TDL model of Fig. 2, we could account for shadowing by multiplying the TDL input (or output) by a more slowly varying shadow random process, and account for path loss by an even more slowly varying multiplying factor. For most V2V communications of a few packets in length, path loss would be well modeled as a constant. We discuss these three effects next, with a focus on empirical models—those based upon measurements. Deterministic (analytical) and "mixed" models are described in subsequent sections.

Recent works on path loss include [42] and [43], wherein the well-known 2-ray model [17] is suggested for LOS conditions, particularly in rural settings, and the also common log-distance model [44] is used for obstructed-LOS conditions in suburban and urban areas. The log-distance model takes the form

$$L(d) = A_0 + 10nlog_{10}(d/d_0) + X \tag{3}$$

where $L(d)$ is the path loss in dB at distance d, A_0 is the line intercept at reference distance $d=d_0$, n is the path loss exponent, and X is a Gaussian random variable that in NLOS cases represents shadowing, but for LOS cases simply represents the deviation of measurements (or simulation results) from the linear fit of the first two terms. Results in [43] indicate urban and suburban values of $n<2$, the free-space value. One explanation for this is some "waveguiding" effects due to buildings, but this seems unlikely, particularly in suburban areas where buildings (houses) lining streets are set back from the street and building density is moderate. Reference [45] employed a ray-tracing based computer simulation to estimate urban path loss, with modest improvements over a 2-ray (or "few-ray") model, with only limited comparison with measurements. This model's complexity in comparison to the 2-ray or log-distance approach reduces its appeal. In [46] the authors confirmed 2-ray behavior with measurements and simulations for LOS conditions on a vacant street, and in [47] the authors conducted measurements for various antenna heights in urban areas, and produced an "augmented" log-distance model that accounts for both Tx and Rx antenna height, for both LOS and NLOS conditions. Path loss exponents of approximately 3 for the LOS case and 4 for the NLOS case were found. The authors of [48] also employed ray tracing to estimate path loss in urban intersections, and found good agreement with measurements. This result is an empirical intersection-specific model that is a function of the logarithm of the product of two distances: the distance of each car from the intersection.

Models for MPC fading are also typically subsumed into models for PDPs, and these sometimes include statistics on Doppler as well. References [49]-[57] are a representative sample from the literature. In [49] the authors present TDL models based upon measurements in a number of environments including highways and cities. Both small scale (Rician or Rayleigh) MPC fading and asymmetric Doppler spectra were provided. Measured Doppler spectra were noted to be time-varying, implying NS conditions. The representative environments in [49] are not comprehensive, and Doppler spectra were fit to a limited number of shapes to enable use of a commercial fading emulator, but these explicit models could be useful components of a more comprehensive V2V channel model "set." Reference [50] is an earlier and less complete version of [49]. In [51] the authors present statistics on instantaneous RMS-DS, and also include a model for MPC non-stationarity in the form of the persistence process ($z_k(t)$ in (1) and (2)). Results for two levels of automobile traffic density were provided for the highway settings, and again, this reference provides explicit TDL models. Example coherence bandwidths were also provided, with additional results for multiple values of channel bandwidth presented by the authors in [52]. Results in [53] include both path loss and example PDPs, and statistics on RMS-DS, from which TDL models could be derived for rural, urban, and highway settings. In [54] the authors provide summary statistics on coherence (or "stationarity") distance, and small scale fading amplitude parameters. Although no explicit TDL models were provided, these authors found instances of severe fading (characterized by fading statistics worse than the Rayleigh model [16]), similar to findings in [51]. Although the addressed V2V settings are not comprehensive, the models cited in [49] and [51]-[54] are some of the most mature to date. Reference [55] compared RMS-DS results of several independent measurement campaigns, and [56] collected data on both RMS-DS and Doppler spreads. Similarly, the authors of [57] presented RMS-DS results for several V2V settings, and also provided models for small-scale (Rician K-factor) and shadowing (Suzuki log-normal standard deviation) distributions.

This brief discussion reveals that to date there are several models (TDL) that can be employed for V2V analysis and simulations. Table 1 shows values for RMS-DS from these references, illustrating a fairly wide range of results for the various expressway and urban settings.

Table 1. Root-Mean-Square Delay Spread Values σ_τ for V2V Settings

Reference	Frequency Band (GHz)	σ_τ (nanosec)	Comments
[49]	5.9	40.2 67.6 35.7	Expressway oncoming Expressway same direction, w/wall Urban canyon, oncoming
[51]	5.1	20 48 200 125.8 1328	Expressway median, low traffic density Expressway median, high traffic density Expressway 90th percentile value Urban mean Urban maximum
[53]	5.9	41 47	Expressway mean, std. dev. of σ_τ=14 ns Urban mean, std. dev. of σ_τ=22 ns
[55]	5.2	15	Doppler > $2v/\lambda$ found
[56]	5.7	54 247.7 17	Expressway mean, oncoming Expressway maximum, oncoming Urban mean, convoy
[57]	5.3	165 2083 373 2100	Expressway mean Expressway maximum Urban mean Urban maximum

Given models for path loss and small-scale fading CIRs, what remains to complete our three-component V2V channel model is a model for V2V shadowing. This is conspicuously absent in the literature, with but one reference citing a few measured results [58]. In this reference, a typical shadowing value of approximately 10 dB was reported when a large truck was placed between the Tx and Rx on two passenger cars. In the following subsections we describe additional types of V2V channel models and features. The division according to our section headings is not always sharp, as some models contain aspects that fit within more than one category.

3.1 Geometry-Based Models

The use of geometry-based models has seen much recent attention. This classification is also not without subtlety, as the original geometry-based models were deterministic (ray tracing), but have evolved to allow stochastic elements. We briefly describe these deterministic models in the following section. With stochastic elements, geometry-based models are now often termed geometry-based stochastic channel models (GBSCMs). References [10] and [59]-[67] constitute our selection of this class. These models represent the CIR as a sum of complex exponentials, and can hence be termed sum-of-sinusoids (SoS) models [17].

Reference [10] is an older paper that extends the work in [7] and [8] to three dimensions. The authors derive expressions for Doppler spectra and autocorrelations, which contain integrals that must be evaluated numerically, but the expressions are quite generic, allowing for arbitrary antenna gain patterns and angle-of-arrival probability density functions (pdfs). In [59], the authors presume Rayleigh fading a' priori, and develop a simulation to efficiently emulate fading that agrees with the so-called "reference model" in [7]. In [60], another detailed analytical simulation model was presented. This model assumes two cylindrical "rings" of reflecting objects surrounding Tx and Rx, and another "reference model" that assumes rich scattering and central limit theorem (CLT) applicability, and implementations via the SoS approach. (This model also incorporates multiple-input/multiple output (MIMO) modeling with antenna arrays at both Tx and Rx vehicles.) The complexity of this model is relatively high, as a number (1-2 dozen) of geometric parameters must be specified; values for some of these are drawn from user-selectable random distributions. The authors validate their SoS implementation against the theoretical (WSS) autocorrelations and Doppler spectra. Reference [61] also employs the "2 ring" scattering model and, similar to [60], assumes that the CLT and WSSUS apply. Correlation functions and Doppler spectra for a narrowband V2V channel were derived. The work in [62] extends the work in [60] to the MIMO case. In [63] the authors add an ellipse to the 2-ring geometry and again find expressions for correlation functions and Doppler spectra. The authors of [64] propose a complexity reduction method for GBSCMs by introducing a composite function to represent all diffuse scattering. This composite function is a sum of weighted prolate spheroidal sequences (whose complexity may not be universally viewed as trivial), with weights presumably determined by measurement data. The authors of [65] model reflections along a street symmetrically lined with buildings to develop a time-frequency narrowband geometric model. Reference [66] is yet another variation, with multiple scattering rings about both Tx and Rx. Our final citation for this class of channels is [67], which is distinguished from most of the others by virtue of detailed specifications of distributions of geometric parameters derived from measurements; prior cited GBSCM references do this to a far more limited extent.

Once configured, the complicated GBSCMs have the ability to replicate a range of V2V scattering conditions, and given "tuning" via measurement data, can yield realistic results. In fact, in [35] we were able to show that by proper parameter settings, we could obtain MPC persistence process results from a modified version of the GBSCM in [67] that were in excellent agreement with independent measurement results from [51]. As noted in [35] though, the most serious drawback of GBSCMs in comparison to TDL models is run time: GBSCM models can require run times of a factor from 10 to 180 times larger than that of the TDL to generate the same number of model CIRs. In addition, the measurement based model of [67] pertains to the highway setting only, hence additional development is required for realistic GBSCMs for other V2V settings.

3.2 Deterministic Models and MIMO Models

Deterministic models typically employ ray tracing. The ray-tracing technique has been used widely for a number of environments outside V2V, and once configured for their specific setting, fairly accurate results can be obtained. References [68]-[70] are example V2V ray tracing results. Without "randomization" of the geometric parameters, these models are site-specific, and retain a complexity that is usually even larger than that of GBSCMs.

The use of multiple-input/multiple output (MIMO) processing in wireless systems via multiple antennas at both Tx and Rx has also grown prolifically over the past decade [37], and use of MIMO in V2V communications is no exception. Several references that employ MIMO in analysis and/or measurements have already been cited in regard to GBSCMs, and as noted, the MIMO channel is a generalization of the single antenna channel that can be described by a matrix of CIRs with entries of the form of (1) corresponding to specific Tx-Rx antenna pairs. A CIR for this model from [67] is

$$h(\tau;t) = \sum_{k=0}^{L-1} \alpha_k(t) e^{j2\pi d_k(t)} \delta(t-\tau_k(t)) \delta(\theta_R - \theta_{R,k}(t)) \delta(\theta_T - \theta_{T,k}(t)) g_R(\theta_R) g_T(\theta_T) \tag{4}$$

where α is again MPC amplitude, $d_k(t)$ is the distance from Tx to reflector to Rx, θ_R is angle of arrival, θ_T angle of departure, and g_R and g_T are antenna gain patterns at Rx and Tx, respectively. Based upon measurements, [67] found that the fading MPC amplitudes "appeared and disappeared" over time, just as in the "birth-death" TDL persistence processes models.

In [71] the authors showed example received power and CIR results for the intersection setting, illustrating that MIMO beamforming can be advantageous. Reference [72] provided results showing increased data rate with another 2 by 2 MIMO system, wherein with the albeit idealized assumption of isotropic scattering, achievable data rate was still shown to decrease with an increase in velocity. The authors of [73] quantified performance gains in throughput with a SIMO configuration.

3.3 Non-stationarity

As noted in [33], the WSSUS assumptions are so common that many investigators either explicitly or tacitly assume these conditions apply. Careful analysis and measurements of the V2V channel have though revealed that the validity of the WSSUS assumptions is not guaranteed.

The formulation in (1) yields time-varying CIRs and their corresponding statistics via MPC persistence processes $z_k(t)$. These binary random processes take values in the set {0,1}, and account for finite "lifetimes" of MPCs. Although the 0/1 alphabet yields abruptness in what should be gradual (but rapid) transitions, this can be easily solved by interpolation. Results in [51] and [74] show that if non-stationarity is not taken into account in V2V channel modeling, system performance results do indeed change. Physical explanations for this medium-scale or "mesoscale" [75] persistence effect are rapid obstruction from nearby vehicles, and delay "drift," where the MPC absolute delay changes over time so that MPCs move from one delay bin to another. The WINNER channel models (not V2V) in [76] provide several modeling methods to represent this type of medium-scale behavior, via 2D spatial filtering, turning entire clusters of multipath components on/off, etc. In [77] we investigated use of higher-order Markov models for multipath persistence, and found that model order need not be larger than two. We also found that NS Markov chains composed of a concatenation of distinct 1st-order chains (i.e., a non-homogeneous chain) provided no substantial improvement in terms of agreement with measured data.

A more formal approach makes use of time-frequency analysis [78] to describe the CIR or CTF. An excellent framework for this appears in [79], in which a "local scattering function" (LSF) is derived as useful for representing the channel correlation functions within a stationarity region in the time and frequency domains. The complexity of such a representation warrants further scrutiny before its application to V2V channel modeling. In [39] the authors quantified stationarity intervals in a MIMO system in several environments via a correlation matrix distance. These intervals were modeled as being lognormal in distribution, with means on the order of 3 m (~50 wavelengths at 5.3 GHz). Small-scale fading fit the Weibull distribution, and both the Weibull shape parameter and the large-scale fading (~shadowing) standard deviation were fit to an extreme value distribution, but in the end, [39] provided no easily implementable or explicit V2V channel model. Reference [80] addresses the LSF directly, and provides a multitaper estimator for it. In [81] the authors measured and developed models for MPC lifetimes (~persistence) at urban intersections. Components with lifetimes both above and below a threshold of 50 ms were found to be lognormally distributed, with short-lived MPCs large in number and long-lived contributions very small in number. Reference [82] is another work on measurements and estimation of the LSF; stationarity times were found to be as small as 23 ms on highways for Tx and Rx driving in opposite directions, but were approximately 1.5 seconds for same-direction driving. All this work can contribute to the development of NS V2V channel models, but thus far no "integration" of all findings has been conducted.

3.4 A V2V Channel Classification Scheme

As noted, V2V channels may not fit within the popular cellular classes of urban, suburban, and rural. At minimum, vehicular traffic density subclasses could be employed within these environments. Additional settings specific to V2V communications include tunnels, street intersections, and possibly overpasses/underpasses, slopes, and traffic circles. Although numerous authors have addressed many of these settings, there has been no effort to systematically and comprehensively classify V2V channels. An exemplary model for such a classification may be that of the COST 259 effort [83], [84], wherein a hierarchy of multiple cell types contains multiple radio environments, within which are multiple propagation scenarios.

An example V2V channel classification diagram is shown in Fig. 3. This classification is still in "draft form," and some elements may later be added, omitted, combined, or re-defined. For reasons of length, we do not describe this classification in detail. The hierarchy consists of three levels, with some additional detail on the local settings listed at the bottom as "setting features."

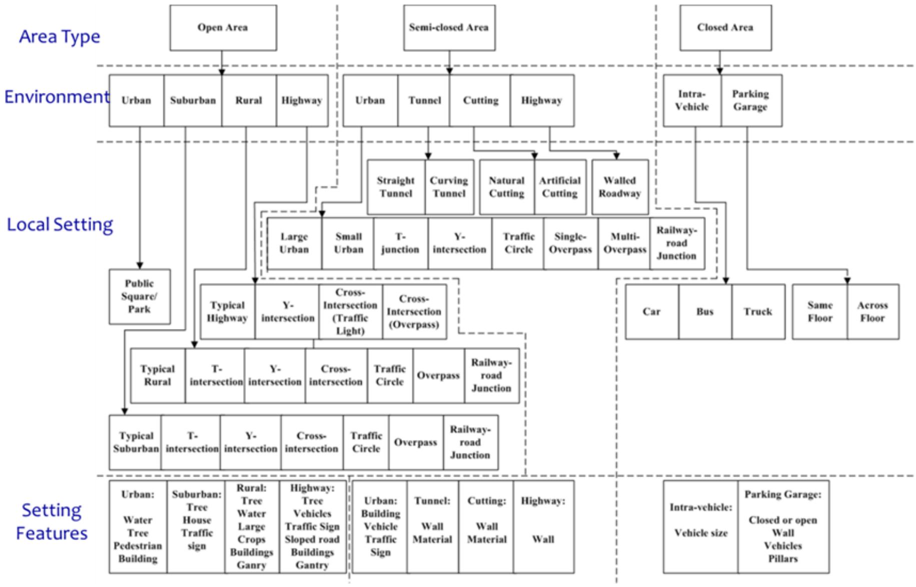


Fig. 3. Potential V2V channel classification scheme

4 Example New Results

4.1 Parking Garages

Here we extend the V2V channel classes to another environment most common in urban areas: the parking garage. These garages are usually multi-floored structures built from concrete and steel, and may be fully enclosed, underground or above-ground, or may have sides that are open to the outside. They are hence distinct from most indoor environments previously studied. To our knowledge, only a few investigations of parking garage channel characteristics have been reported. Reference [85] considered ultrawideband (3.28-5.03 GHz) propagation in underground parking garages for short distances (up to 20 m), for LOS conditions only. For this completely enclosed structure, a log-distance path loss model was developed, yielding a path loss exponent n=1.5. The authors also characterized RMS-DS, and found values less than 30 ns. In [86] the author reported on path loss only for two parking garages at a frequency of 1.8 GHz, for distances up to approximately 70 m. Antenna heights were 1 m and 1.2 m (lower than that for typical V2V mountings on car roofs). The popular 2-ray model [17] was used to fit measured results. Reference [87] measured indoor-to-outdoor propagation losses for multiple building types, including parking garages, over a wide frequency range: 800 MHz-8 GHz.

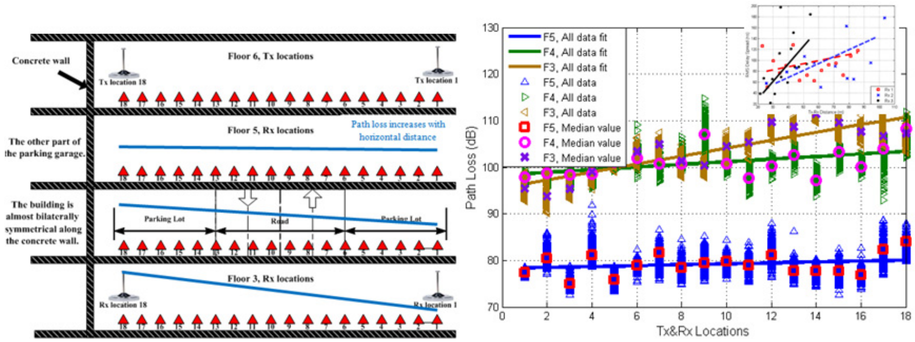


Fig. 4. Garage floor plan for floor-to-floor measurements (left) and floor-to-floor path loss vs. Tx & Rx locations for all three floors (right); inset on right is *same-floor* RMS-DS vs. Tx-Rx distance for three distinct Rx locations within the garage.

We recently obtained some results for parking garage channels in the 5 GHz band [88]. Using a 50-MHz spread spectrum signal received with a stepped correlator receiver, results for propagation path loss and RMS-DS for Tx and Rx on the same floor, and on different floors were obtained. Same-floor measurements sample both LOS and NLOS conditions; floor-to-floor results are all NLOS. The floor-to-floor setting and path loss results vs. location on the floors are shown in Fig. 4; the inset in the path loss (right-side) figure shows example same-floor RMS-DS results vs. distance for three distinct locations. The floor-to-floor path loss results show a dependence both on vertical distance and horizontal distance from the open garage sides. Path loss appears to be well modeled as a sum of free-space, floor attenuation factors, and a factor proportional to the distance away from the garage's open side. Depending upon location and the density of parked cars, average RMS-DS can vary from 30 ns to 190 ns for same-floor conditions. Additional results and analysis will appear in [88].

4.2 Path Loss on a Slope

Although path loss has been measured for many situations—including V2V channels—one setting that we have found absent a thorough study is the path loss obtained when vehicles are at opposite ends of a slope. Although early work on path loss covered different types of terrain in several frequency bands, no explicit model applicable to V2V path loss over an intervening slope is apparent. Even the classic paper by Okumura, et. al. [89], does not pertain here since those results were for cellular conditions with tall base stations.

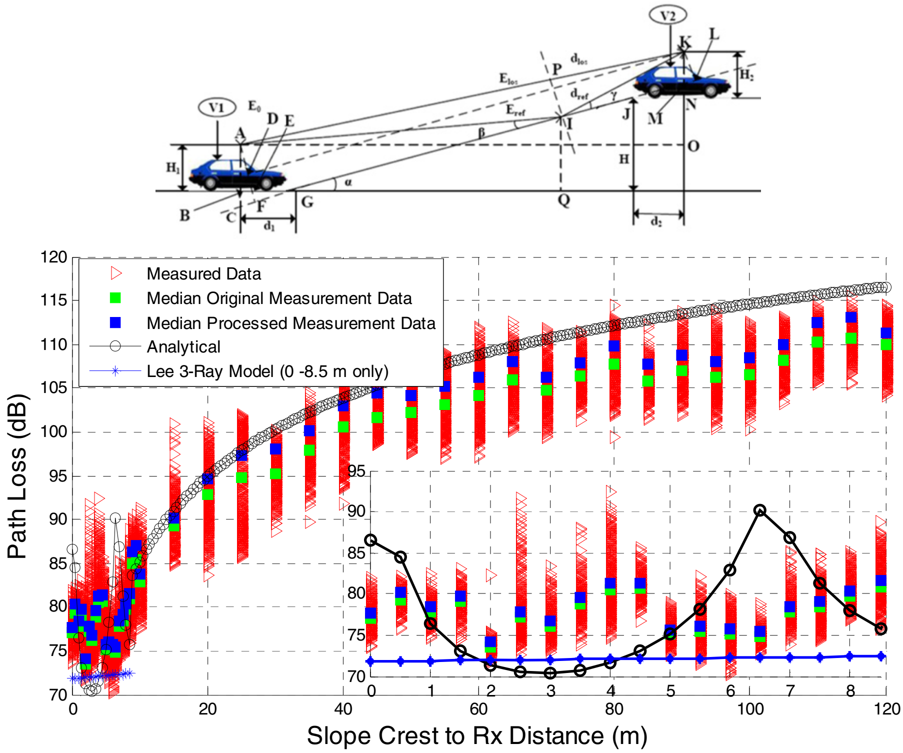


Fig. 5. Slope path loss geometry (top), and analytical & measured results for slope path loss vs. distance (bottom)

Hence we have generalized the familiar 2-ray model analysis to apply to a slope [90]. The geometry is shown in Fig. 5, and analytical and measurement results are shown below the slope diagram. The measurement data was collected at a slope near the University of South Carolina campus, in Columbia, SC, via our 5 GHz channel sounding equipment, on 29th December 2012. The measurements were on a section of Bull Street, between Heyward St. (slope bottom) and Whaley St. (slope top). The slope height is 3.49 m, and has a 22.5° tilt angle. Two observations of note are as follows: at shorter distances, the slope path loss fluctuates by approximately 5 dB, and is larger than that predicted by Lee's 3-ray model [91]; and, for larger distances, the slope is well modeled by a knife-edge diffraction even out to 120 m from the slope crest. Analysis will appear in [90].

5 What's Next

Based upon the prior discussion, one thing that needs attention is vehicular shadowing. Although some measured results and models based upon them may have tacitly included shadowing, this phenomenon can be significant enough to warrant explicit modeling of its own. The two-dimensional geometry based models are at present incapable of modeling diffraction, so measured results for shadowing depths caused by trucks and buses and other large vehicles of various sizes should be collected for a range of distances between Tx, shadowing vehicle, and Rx, in order to develop a V2V shadowing model.

In both [92] and [93], V2V channel time variation was cited as being critical to assess for ensuring reliable V2V communications, particularly with the current DSRC standard. Alternatives to the DSRC standard do exist, e.g., LTE [94], but the fact remains that rapid time variation can be a significant channel impairment. This leads to the phenomenon of channel non-stationarity, and how it should best be modeled. Research has shown that incorporating non-stationarity is essential for obtaining realistic communication system performance results, but the question of how to both accurately and efficiently model non-stationarity is an open question. Geometry based models can incorporate NS behavior "naturally," but are complex, and at present can not account for shadowing (the addition of which would further complicate them). Tapped delay line models can be augmented to incorporate NS behavior, via MPC persistence, or some other method of "growing" and "decaying" MPCs, but the best method for this approach is yet to be determined. Both model types require expansion to allow for "transitions" between different types of V2V channel environments.

Finally, for the purposes of repeatability in analysis and especially communication system simulations for radio development and testing, standardized V2V channel models should be developed. As with other complex systems such as cellular [76], a comprehensive *set* of models will be required.

As noted in [95] in reference to the assessment of V2V performance limits, "One aspect would be better channel models…." The modeling of V2V channels will continue to flourish over the next few years.

6 Conclusion

In this paper we first motivated the study of the V2V communication channel for future applications. After providing a short description of the important channel characteristics, we conducted a review of the literature. This review covered results on measurements and analysis for multiple V2V channel types. The extent of the review, although not exhaustive, demonstrates the importance of this topic. We described conventional tapped delay line models, and newer more complex geometry-based models, and for both indicated the need to account for the channel's changing statistics. This led to a brief discussion of work on non-stationary V2V channel modeling. A potential channel classification scheme was also described. Some new results for V2V channels in parking garages and on a slope were also provided to emphasize that work in the area of V2V channels is not completed. We ended the paper with a discussion on remaining issues, and the most important of these are good models for vehicular shadowing, appropriate incorporation of non-stationarity, and the development of standardized V2V channel models.

Acknowledgment. The author would like to thank his graduate student Ruoyu Sun, and visiting scholar Pengyu Liu, for their assistance with the parking garage and slope path loss results.

References

1. IEEE Vehicular Technology Magazine, Special Issue on V2V Communications 2(4) (December 2007)
2. Karagiannis, G., Altintas, O., Ekici, E., Heijenk, G., Jarupan, B., Lin, K., Weil, T.: Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards, and Solutions. IEEE Comm. Surveys & Tutorials 13(4), 584–616 (2011)
3. Zhu, J., Roy, S.: MAC for Dedicated Short Range Communications in Intelligent Transportation. IEEE Comm. Mag. 41(12), 60–67 (2003)
4. Biswas, S., Tatchikou, R., Dion, F.: Vehicle-to-Vehicle Wireless Communication Protocols for Enhancing Highway Traffic Safety. IEEE Comm. Mag. 44(1), 74–82 (2006)
5. Joerer, S., Sommer, C., Dressler, F.: Toward Reproducibility and Comparability of IVC Simulation Studies: A Literature Survey. IEEE Comm. Mag. 50(10), 82–88 (2012)
6. Molisch, A.F., Karedal, J., Tufvesson, F., Paier, A., Bernado, L., Zemen, T., Klemp, O., Czink, N.: Vehicular Channel Characterization and Its Implication for Wireless System Design and Performance. Proc. IEEE 99(7), 1189–1212 (2011)
7. Akki, A.S., Haber, F.: A Statistical Model of Mobile-to-Mobile Land Communication Channel. IEEE Trans. Veh. Tech. VT-35(1), 2–7 (1986)
8. Akki, A.S.: Statistical Properties of Mobile-to-Mobile Land Communication Channels. IEEE Trans. Veh. Tech. 43(4), 826–831 (1994)
9. Davis, J.S., Linnartz, J.P.M.G.: Measurements of Vehicle-to-Vehicle Propagation. In: Proc. Asilomar Conference, Monterey, CA, October 31-November 1 (1994)
10. Vatalaro, F., Forcella, A.: Doppler Spectrum in Mobile-to-Mobile Communications in the Presence of Three-Dimensional Multipath Scattering. IEEE Trans. Veh. Tech. 46(1), 213–219 (1997)
11. ITS project (February 2013), `http://www.its.dot.gov/index.htm`
12. National Highway Traffic Safety Administration, US DOT (February 2013), `http://www-fars.nhtsa.dot.gov/Main/index.aspx`
13. Faezipour, M., Nourani, M., Saeed, A., Addepalli, S.: Progress and Challenges in Intelligent Vehicular Networks. Proc. ACM 55(2), 90–100 (2012)
14. International Standards Organization (ISO), (February 2013), `http://www.iso.org`
15. Matolak, D.W.: Channel Modeling for Vehicle-to-Vehicle Communications. IEEE Comm. Mag. 46(5), 76–83 (2008)

16. Matolak, D.W., Frolik, J.: Worse-than-Rayleigh Fading: Experimental Results and Theoretical Models. IEEE Comm. Mag. 49(4), 140–146 (2011)
17. Parsons, J.D.: The Mobile Radio Propagation Channel. John Wiley & Sons, New York (2000)
18. ITU document ITU-R P.1407-1, Multipath Propagation and Parameterization of its Characteristics (1999-2003)
19. Bello, P.: Characterization of Random Time-Variant Linear Channels. IEEE Trans. Comm. 11, 360–393 (1963)
20. Molisch, A.F., Steinbauer, M.: Condensed Parameters for Characterizing Wideband Mobile Radio Channels. Int. Journ. Wireless Information Networks 6(3), 133–154 (1999)
21. Schumacher, H., Tchouankem, H., Nuckelt, J., Kuerner, T., Zinchenko, T., Leschke, A., Wolf, L.: Vehicle-to-Vehicle 802.11p Performance Measurements at Urban Intersections. In: Proc. IEEE ICC, Workshop on Intelligent Vehicular Networking, Ottawa, ON, CA, June 10-15 (2012)
22. Bernado, L., Roma, A., Paier, A., Zemen, T., Czink, N., Karedal, J., Thiel, A., Tufvesson, F., Molisch, A.F., Mecklenbrauker, C.F.: In-Tunnel Vehicular Radio Channel Characterization. In: Proc. IEEE Spring VTC, Budapest, Hungary, May 15-18 (2011)
23. Chen, S., Wyglinski, A.M., Pagadarai, S., Vuyyuru, R., Altintas, O.: Feasibility Analysis of Vehicular Dynamic Spectrum Access via Queueing Theory Model. IEEE Comm. Mag. 49(11), 156–163 (2011)
24. Acosta-Marum, G., Ingram, M.A.: A BER-Based Partitioned Model for a 2.4 GHz Vehicle-to-Vehicle Expressway Channel. Wireless Pers. Comm. 37, 421–433 (2006)
25. Konstantinou, K., Kang, S., Tzaras, C.: A Measurement Based Model for Mobile-to-Mobile UMTS Links. In: Proc. IEEE Veh. Tech. Conf., Singapore, May 11-14, pp. 529–533 (2008)
26. Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems—5 GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications, ASTM e2213-03 (September 2003)
27. Uzcategui, R.A., Acosta-Marum, G.: WAVE: A Tutorial. IEEE Comm. Mag. 47(5), 126–133 (2009)
28. Cheng, L., Henty, B.E., Cooper, R., Stancil, D.D.: A Measurement Study of Time-Scaled 802.11a Waveforms over the Mobile-to-Mobile Vehicular Channel at 5.9 GHz. IEEE Comm. Mag. 46(5), 84–91 (2008)
29. Gallagher, B., Akatsuka, H.: Wireless Communications for Vehicle Safety: Radio Link Performance and Wireless Connectivity Methods. IEEE Veh. Tech. Mag. 1, 4–16 (2006)
30. Schack, M., Kornek, D., Slottke, E., Kuerner, T.: Analysis of Channel Parameters for Different Antenna Configurations in Vehicular Environments. In: Proc. IEEE Fall Veh. Tech. Conf., Ottawa, ON, Canada, September 6-9 (2010)
31. Molisch, A.F., Tufvesson, F., Karedal, J., Mecklenbrauker, C.F.: A Survey on Vehicle-to-Vehicle Propagation Channels. IEEE Wireless Comm. Mag. (12), 12–22 (2009)
32. Wang, C.X., Cheng, X., Laurenson, D.I.: Vehicle-to-Vehicle Channel Modeling & Measurements: Recent Advances & Future Challenges. IEEE Comm. Mag. 47(11), 96–103 (2009)
33. Matolak, D.W., Wu, Q.: Vehicle-To-Vehicle Channels: Are We Done Yet? In: Proc. Globecom 2009 Workshop on Networking Intelligent Vehicles and Infrastructures, Honolulu, HI, USA (December 4, 2009)

34. Matolak, D.W.: Radio Channel Modeling for Vehicle-to-Vehicle/Road Communications. In: Zhou, M.-T., Zhang, Y., Yang, L.T. (eds.) Wireless Technologies for Intelligent Transportation Systems, ch. 2. Nova Science Publishers (March 2010)
35. Matolak, D.W., Wu, Q.: Channel Models for V2V Communications: A Comparison of Different Approaches. In: Proc. European Conf. on Antennas & Propagation, Rome, Italy, April 11-15 (2011)
36. Boche, H., Bourdoux, A., Fonollosa, J.R., Kaiser, T., Molisch, A., Utschick, W.Q.: Smart Antennas: State of the Art. IEEE Vehicular Tech. Magazine 1(1), 8–17 (2006)
37. Gesbert, D., Bolcskei, H., Gore, D.A., Paulraj, A.J.: Outdoor MIMO Wireless Channels: Models and Performance Prediction. IEEE Trans. Comm. 50(12), 1926–1934 (2002)
38. Stuber, G.L.: Principles of Mobile Communication, 2nd edn. Kluwer Academic Pub., Boston (2001)
39. Renaudin, O., Kolmonen, V.-M., Vainikainen, P., Oestges, C.: Non-Stationary Narrowband MIMO Inter-Vehicle Channel Characterization in the 5 GHz Band. IEEE Trans. Veh. Tech. 59(4), 2007–2015 (2010)
40. Matolak, D.W.: Channel Modeling for Vehicle-to-Vehicle Communications and Networking. In: Santos, R.A., Licea, V.R. (eds.) Wireless Technologies in Vehicular Ad Hoc Networks: Present and Future Challenges. IGI Global Publishing (2012)
41. Zheng, Y.R.: A Non-Isotropic Model for Mobile to Mobile Fading Channel Simulations. In: Proc. MILCOM 2006, Washington, DC (2006)
42. Cheng, L., Henty, B.E., Stancil, D.D., Bai, F., Mudalige, P.: Mobile Vehicle-to-Vehicle Narrow-band Channel Measurement and Characterization of the 5.9 GHz Dedicated Short Range Communication (DSRC) Frequency Band. IEEE Journ. Selected Areas Comm. 25(8), 1501–1516 (2007)
43. Karedal, J., Czink, N., Paier, A., Tufvesson, F., Molisch, A.F.: Path Loss Modeling for Vehicle-to-Vehicle Communications. IEEE Trans. Veh. Tech. 60(1), 323–328 (2011)
44. Perez Fontan, F., Espineira, P.M.: Modeling the Wireless Propagation Channel: a Simulation Approach with Matlab. John Wiley & Sons, West Sussex (2008)
45. Tabatabaei, S.A.H., Fleury, M., Qadri, N.N., Ghanbari, M.: Improving Propagation Modeling in Urban Environments for Vehicular Ad Hoc Networks. IEEE Trans. Intelligent Transp. Sys. 12(3), 705–716 (2011)
46. Ito, Y., Taga, T., Muramatsu, J., Suzuki, N.: Prediction of Line of Sight Propagation Loss in Inter-Vehicle Communication Environments. In: Proc. IEEE Int. Symp. Pers., Indoor, & Mobile Radio Comm., PIMRC, Athens, Greece, September 3-7 (2007)
47. Konstantinou, K., Kang, S., Tzaras, C.: A Measurement-Based Model for Mobile-to-Mobile UMTS Links. In: Proc. IEEE Spring Veh. Tech. Conf., Singapore, May 11-13 (2008)
48. Schack, M., Nuckelt, J., Geise, R., Thiele, L., Kuerner, T.: Comparison of Path Loss Measurements and Predictions at Urban Crossroads for C2C Communications. In: Proc. European Conf. on Antennas & Propagation, Rome, Italy, April 11-15 (2011)
49. Acosta-Marum, G., Ingram, M.A.: Six Time- and Frequency-Selective Empirical Channel Models for Vehicular Wireless LANs. IEEE Vehicular Technology Mag. 2(4), 4–11 (2007)
50. Acosta-Marum, G., Ingram, M.A.: Doubly Selective Vehicle-to-Vehicle Channel Measurements and Modeling at 5.9 GHz. In: Proc. Int. Symp. Wireless Pers. Multimedia Comm., San Diego, CA, September 17-20 (2006)
51. Sen, I., Matolak, D.W.: Vehicle-Vehicle Channel Models for the 5 GHz Band. IEEE Trans. Intelligent Transp. Systems 9(2), 235–245 (2008)

52. Matolak, D.W., Wu, Q., Sen, I.: 5 GHz Band Vehicle-to-Vehicle Channels: Models for Multiple Values of Channel Bandwidth. IEEE Trans. Vehicular Tech. 59(5), 2620–2625 (2010)
53. Kunisch, J., Pamp, J.: Wideband Car-to-Car Radio Channel Measurements and Model at 5.9 GHz. In: Proc. IEEE Fall Veh. Tech. Conf., Calgary, AB, Canada, September 21-24 (2008)
54. Renaudin, O., Kolmonen, V.-M., Vainikainen, P., Oestges, C.: Wideband Measurement-Based Modeling of Inter-Vehicle Channels in the 5 GHz Band. In: Proc. European Conf. on Antennas & Propagation, Rome, Italy, April 11-15 (2011)
55. Paier, A., Karedal, J., Czink, N., Dumard, C., Zemen, T., Tufvesson, F., Mecklenbrauker, C.F.: Comparison of Lund 2007 Vehicular Channel Measurements with the IEEE 802.11p Model. COST 2100 TD(08) 436, Wroclaw, Poland, February 6-8 (2008)
56. Paschalidis, P., Wisotzki, M., Kortke, A., Peter, M., Keusgen, W.: Wideband Car-to-Car MIMO Radio Channel Measurements at 5.7 GHz and Issues Concerning Application-Oriented Systems. In: Proc. 1st IEEE Veh. Tech. Society Wireless Access in Veh. Env (WAVE) Conf., Dearborn, MI, December 8-9 (2008)
57. Renaudin, O., Kolmonen, V.-M., Vainikainen, P., Oestges, C.: Wideband MIMO Car-to-Car Radio Channel Measurements at 5.3 GHz. In: Proc. IEEE Fall Veh. Tech. Conf., Calgary, AB, Canada, September 21-24 (2008)
58. Boban, M., Vinhoza, T.T.V., Ferreira, M., Barros, J., Tonguz, O.K.: Impact of Vehicles as Obstacles in Vehicular Ad Hoc Networks. IEEE Journ. Sel. Areas Comm. 29(1), 15–28 (2011)
59. Patel, C.S., Stuber, G.L., Pratt, T.G.: Simulation of Rayleigh-Faded Mobile-to-Mobile Communication Channels. IEEE Trans. Comm. 53(11), 1876–1884 (2005)
60. Zajic, A.G., Stuber, G.L.: Three-Dimensional Modeling, Simulation, and Capacity Analysis of Space-Time Correlated Mobile-to-Mobile Channels. IEEE Trans. Veh. Tech. 57(4), 2042–2054 (2008)
61. Patzold, M., Hogstad, B.O., Youssef, N.: Modeling, Analysis, and Simulation of MIMO Mobile-to-Mobile Fading Channels. IEEE Trans. Wireless Comm. 7(2), 510–520 (2008)
62. Zajic, A.G., Stuber, G.L., Pratt, T.G., Nguyen, S.: Wideband MIMO Mobile-to-Mobile Channels: Geometry-based Statistical Modeling with Experimental Verification. IEEE Trans. Veh. Tech. 58(2), 517–534 (2009)
63. Cheng, X., Wang, C.-X., Laurenson, D.I., Salous, S., Vasilakos, A.V.: An Adaptive Geometry-Based Stochastic Model for Non-Isotropic MIMO Mobile-to-Mobile Channels. IEEE Trans. Wireless Comm. 8(9), 4824–4835 (2009)
64. Czink, N., Kaltenberger, F., Zhou, Y., Bernado, L., Zemen, T., Yin, X.: Low-Complexity Geometry-Based Modeling of Diffuse Scattering. In: Proc. European Conf. on Antennas & Propagation, Barcelona, Spain, April 12-16 (2010)
65. Chelli, A., Patzold, M.: A Non-Stationary MIMO Vehicle-to-Vehicle Channel Model Derived from the Geometrical Street Model. In: Proc. IEEE Fall Veh. Tech. Conf., September 5-8 (2011)
66. Yoo, S., Lee, J., Kim, K.: Modeling and Characteristics of Mobile-to-Mobile Wideband MIMO Channel Based on the Geometrical Multi-Radii Two-Rings with Specified Frequency Selectivity. In: Proc. European Conf. on Antennas & Propagation, Prague, Czech Republic, March 26-30 (2012)
67. Karedal, J., Tufvesson, F., Czink, N., Paier, A., Dumard, C., Zemen, T., Mecklenbrauker, C.F., Molisch, A.F.: A Geometry-Based Stochastic MIMO Model for Vehicle-to-Vehicle Communications. IEEE Trans. Wireless Comm. 8(7), 3646–3657 (2009)

68. Maurer, J., Fugen, T., Schafer, T., Wiesbeck, W.: A New Inter-Vehicle Communications (IVC) Channel Model. In: Proc. IEEE Veh. Tech. Conf., vol. 1, pp. 9–13 (September 2004)
69. Maurer, J., Schafer, T.M., Wiesbeck, W.: A Realistic Description of the Environment for Inter-Vehicle Wave Propagation Modeling. In: Proc. IEEE Vehicular Tech. Conf., Atlantic City, NJ, October 7-11, pp. 1437–1441 (2001)
70. Maurer, J., Schafer, T.M., Wiesbeck, W.: Physical Layer Simulations of IEEE 802.11a for Vehicle-Vehicle Communications. In: Proc. IEEE Vehicular Tech. Conf., Dallas, TX, September 25-28 (2005)
71. Paschalidis, P., Mahler, K., Kortke, A., Wisotzki, M., Peter, M., Keusgen, W.: 2 X 2 MIMO Measurements of the Wideband Car-to-Car Channel at 5.7 GHz on Urban Street Intersections. In: Proc. IEEE Fall VTC, San Francisco, CA, September 5-8 (2011)
72. Ohlmer, E., Fettweis, G., Plettemeier, D.: MIMO System Design and Field Tests for Terminals with Confined Space—Impact on Automotive Communication. In: Proc. European Conf. on Antennas & Propagation, Rome, Italy, April 11-15 (2011)
73. Nuckelt, J., Kuerner, T.: MRC Performance Benefit in V2V Communication Systems in Urban Traffic Scenarios. In: Proc. European Conf. on Antennas & Propagation, Prague, Czech Republic, March 26-30 (2012)
74. Wang, B., Sen, I., Matolak, D.W.: Performance Evaluation of 802.16e in Vehicle to Vehicle Channels. In: Proc. IEEE Fall VTC, Baltimore, MD, October 1-3 (2007)
75. Calcev, G., et al.: A Wideband Spatial Channel Model for System-Wide Simulations. IEEE Trans. Veh. Tech. 56, 389–403 (2007)
76. WINNER II interim channel models, D1.1.1VI.1, world wide website (February 2013), `https://www.ist-winner.org/WINNER2-Deliverables/D1.1.1.pdf`
77. Matolak, D.W., Wu, Q.: Markov Models for Vehicle-to-Vehicle Channel Multipath Persistence Processes. In: Proc. 1st IEEE Veh. Tech. Society Wireless Access in Veh. Env. (WAVE) Conf., Dearborn, MI, December 8-9 (2008)
78. Cohen, L.: Time-Frequency Analysis. Prentice-Hall, Upper Saddle River (1995)
79. Matz, G.: On Non-WSSUS Wireless Fading Channels. IEEE Trans. Wireless Comm. 4(5), 2465–2478 (2005)
80. Bernado, L., Zemen, T., Paier, A., Karedal, J., Fleury, B.H.: Parameterization of the Local Scattering Function Estimator for Vehicular-to Vehicular Channels. In: Proc. IEEE Fall Veh. Tech. Conf., Anchorage, AK, September 20-23 (2009)
81. Paschalidis, P., Mahler, K., Kortke, A., Peter, M., Keusgen, W.: Statistical Evaluation of Multipath Component Lifetime in the Car-to-Car Channel at Urban Street Intersections Based on Geometrical Tracking. In: Proc. IEEE Spring Veh. Tech. Conf., Yokohama, Japan, May 6-9 (2012)
82. Paier, A., Zemen, T., Bernado, L., Matz, G., Karedal, J., Czink, N., Dumard, C., Tufvesson, F., Molisch, A.F., Mecklenbrauker, C.F.: Non-WSSUS Channel Characterization in Highway and Urban Scenarios at 5.2 GHz Using the Local Scattering Function. In: Proc. Int. Workshop on Smart Antennas, Helsinki, Finland, February 26-27 (2008)
83. Molisch, A.F., Asplund, H., Heddergott, R., Steinbauer, M., Zwick, T.: The COST259 Directional Channel Model—Part I: Overview and Methodology. IEEE Trans. Wireless Comm. 5(12), 3421–3433 (2006)
84. Asplund, H., Glazunov, A.A., Molisch, A.F., Pedersen, K.I., Steinbauer, M.: The COST259 Directional Channel Model—Part II: Macrocells. IEEE Trans. Wireless Comm. 5(12), 3434–3450 (2006)
85. Lee, J.-Y.: UWB Channel Modeling in Roadway and Indoor Parking Environments. IEEE Tran. Vehicular Tech. 59(7), 3171–3180 (2010)

86. Phaiboon, S.: Propagation Path Loss Models for Parking Buildings. In: 5th Int. Conf. on Information, Communications and Signal Processing, Bangkok, Thailand, pp. 1348–1351 (2005)
87. Okamoto, H., Kitao, K., Ichitsubo, S.: Outdoor-to-Indoor Propagation Loss Prediction in 800-MHz to 8-GHz Band for an Urban Area. IEEE Trans. Vehicular Tech. 58(3), 1059–1067 (2009)
88. Matolak, D.W., Sun, R., Liu, P.: Parking Garage Channel Characteristics at 5 GHz for V2V Applications. In: IEEE Fall Veh. Tech. Conf. (February 2013) (submitted)
89. Okumura, Y., Ohmori, E., Kawano, T., Fukuda, K.: Field strength and its variability in VHF and UHF land mobile radio service. Rev. Electr. Communications Lab. 16, 825–873 (1968)
90. Liu, P., Matolak, D.W., Ai, B., Sun, R.: Path Loss Modeling for Communication on a Slope. IEEE Trans. Veh. Tech. (March 2013) (submitted)
91. Ohira, T., Hirai, T., Tomisato, S., Hata, M.: A Study of Mobile Path Loss Estimation Models for a Sloping Terrain Area in Cellular Systems. In: Asia Pacific Conf. on Communications, Jeju Island, Korea, October 15-17 (2012)
92. Alexander, P., Haley, D., Grant, A.: Cooperative Intelligent Transport Systems: 5.9 GHz Field Trials. Proc. IEEE 99(7), 1215–1235 (2011)
93. Fernandez, J.A., Borries, K., Cheng, L., Vijaya Kumar, B.V.K., Stancil, D.D., Bai, F.: Performance of the 802.11p Physical Layer in Vehicle-to-Vehicle Environments. IEEE Trans. Veh. Tech. 61(1), 3–14 (2012)
94. Vinel, A.: 3GPP LTE Versus IEEE 802.11p/WAVE: Which Technology is Able to Support Cooperative Vehicular Safety Applications. IEEE Wireless Comm. Letters 1(2), 125–128 (2012)
95. Dressler, F., Kargl, F., Ott, J., Tonguz, O.K., Wischhof, L.: Research Challenges in Intervehicular Communications: Lessons of the 2010 Dagstuhl Seminar. IEEE Comm. Mag. 49(5), 158–164 (2011)

Internet Onboard: Technical Analysis

Hassan Ghannoum and David Sanz

SNCF, Innovation & Research Department
40 Avenue des Terroirs de France
75012, Paris, France
{hassan.ghannoum,david.sanz}@sncf.fr

Abstract. This paper gives an overview of wireless technologies solutions that could be used in order to deliver onboard Internet Access Service to train passengers, high speed trains included. These solutions include satellite and terrestrial technologies. The paper discusses the different solutions and presents their advantages, drawbacks and their constraints of implementation as well. Finally, some conclusions and perspectives of future research and development challenges are given.

Keywords: Internet Access Service, satellite technologies, cellular technologies, train-to-trackside communications.

1 Introduction

During the last years, people are becoming more and more used to have Internet access, any time, using different devices (Smartphone, tablet, laptop…) and anywhere, at home, in the office, in public areas… and on the move as well. Trains should not be an exception. That's why the last few years witnessed a constant growth of Internet commercial services onboard trains, launched by many railway operators.

The first solution for train passengers to get Internet connectivity could be through a direct access via their mobile operator network. However, this solution requires appropriate management and capacity of the cells and sufficient radio coverage of the railway lines as well. Besides, the train structure acts like a Faraday cavity and the radio signal attenuation is often higher than 15 dB, reaching sometimes 30 dB, in particular in case of heat protective coated glasses. To overcome this problem and to improve the radio link performance, repeaters could be installed on the train. However, many repeaters may be needed for each train and the costs of the equipment, its installation and its maintenance could be then prohibitive. In addition, the performance of these repeaters is strongly dependent on the radio coverage quality: measures carried out onboard TGV trains by 3 different mobile operators in France showed that the amplification of a weak signal to noise ratio leads to a "bad quality" amplified signal.

The other solution to have Internet connectivity onboard trains could be via an onboard wireless LAN network (typically WiFi) set up by the railway operator. To do this, various communications solutions can be foreseen including cellular solutions,

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 22–30, 2013.

WLAN or WMAN-based solutions, hybrid terrestrial/satellite solutions and two-way satellite solutions. In this paper, we will focus on this scenario and present a technical analysis of the different train-to-ground communications solutions involved.

2 Train-to-Ground Communications

In order to provide broadband internet connectivity services to train passengers, many railway operators have been experimenting, testing, developing, and some of them launching commercial services using train-to-ground communications. These train-to-ground communications include satellite and/or terrestrial solutions. In this section, these solutions will be presented and discussed.

2.1 Satellite Solutions

Satellite communication technologies could be an interesting candidate to deliver train-to-ground connectivity [1] due to its several advantages:

- Easy coverage of wide geographic areas (already available)
- Real broadband connectivity suitable to connect and aggregate the traffic of a large number of mobile terminals
- Support of high speed mobility of the terminals
- Low CAPEX as compared to terrestrial train-to-trackside solutions (no infrastructure needed along the track)
- Flexible/adaptable and easy scalable solutions (to different configurations and conditions)
- Can be easily integrated with alternative terrestrial solutions for full coverage in out-of-visibility areas (tunnels, stations, …)

However, in order to implement satellite train-to-ground connectivity, some important constraints have to be taken into account:

- The satellite to be used has to be in line of sight of the train antenna in order to achieve broadband connectivity. Any obstacle passing across this line of sight (catenary, bridge, high building…) will generate fading/erasures in the received/transmitted signal.
- Since the communication link requires a high gain antenna, its beam-width will be very thin (between 1 and 3 degrees are generally considered). The more the beam-width will be thin, the more accurate the pointing and tracking antenna system will have to be. In addition, a running train is subjected to many kinds of movements: accelerations/decelerations, curves, transversal shifts… Since the pointing of the antenna to the right satellite must be very accurate in order to avoid interfering with other satellites, a high performance pointing and tracking antenna system will be needed. To counteract train attitude, the antenna has to be steerable towards any azimuth direction (360°) and towards a range of elevation angles that will depend on the kind of satellite(s)

(defiling, geo-stationary, constellation of satellites), on its orbital longitude position (when considering geo-stationary satellites) and on the geographic area to be covered (latitude and longitude range).

- In general, GEO (Geostationary Earth Orbit) satellites are especially attractive because they use a geo-synchronous orbit positioned at 36 000 Km form the surface of the earth, at the Equator level, in order to be seen as a fixed point in the sky. In addition, one single satellite can cover 30% of the earth's surface. That's why most of the telecommunications and broadcasting satellites in the world are geo-stationary ones. However, these satellites are not the most suitable for train-to-ground connectivity due to the following reasons:
    - ° GEO satellites introduce an important propagation delay in comparison to LEO (Low Earth Orbit) and MEO (Medium Earth Orbit) ones. This important propagation delay can be a problem for highly interactive applications and some modifications and optimizations are required in order to accelerate the TCP/IP flow.
    - ° Since GEO satellites are positioned over the Equator, northern latitudes require small elevation angles which mean less satellite availability since smaller buildings; mountains… will be seen as additional obstacles.
    - ° The bandwidth price is very high (more than 1.5 M€ in Europe for a 36 MHz transponder).

  Despite those drawbacks, all satellite-based Train-to-Ground connectivity solutions in the market are based on the use of GEO satellites. This could be due to the fact that using GEO satellites guarantees to have a large choice of products, constructors and satellite operators and also the availability of in-orbit unused bandwidth capacity.

- Out-of-visibility areas like tunnels, urban areas and stations are a very important issue since the connectivity can be missed for some minutes. To overcome this problem a "gap-filler" solution can be used. Two main kinds of "gap-filling" solutions can be foreseen:
    - ° Satellite repeaters: one antenna is installed on the ground in order to recover the satellite signal and redistribute it into the non-visibility area. This solution requires the deployment of infrastructure along the tracks. The agreements of the owner of the railway infrastructure and the regulatory telecommunications bodies are needed.
    - ° Vertical handover: switching to another technology, like for example WiFi, when the train enters the non-visibility zone.
- Finally, the railway domain specific constraints must be of course taken into account: electromagnetic compatibility with other existing systems, installation and maintenance, enough space for antenna installation…

In 2003, SNCF started studying the use of satellite technologies in order to deliver Internet access onboard its high speed trains. In that period, this was the most suitable solution in France. SNCF has conducted many experimentation campaigns since 2003 and the first SNCF commercial service "BoxTGV" delivering Internet services to passengers' onboard TGV high speed trains has been launched in December 2010 along the French East high speed line [2]. This service is based on a 2-way Ku-band satellite main link with WiFi radio coverage for non-satellite-visibility areas (tunnels and stations). The "BoxTGV" antenna system and architecture are shown on Figures 1 and 2 respectively.

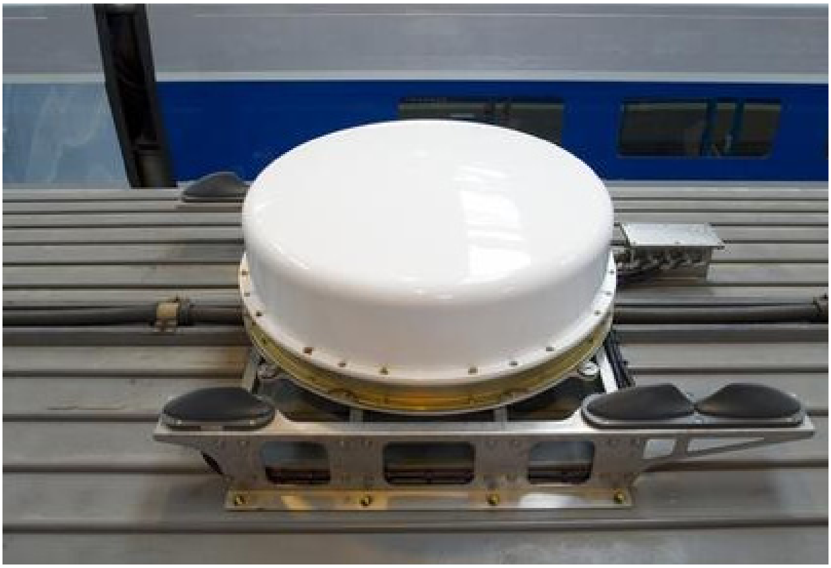


Fig. 1. "BoxTGV" antenna system

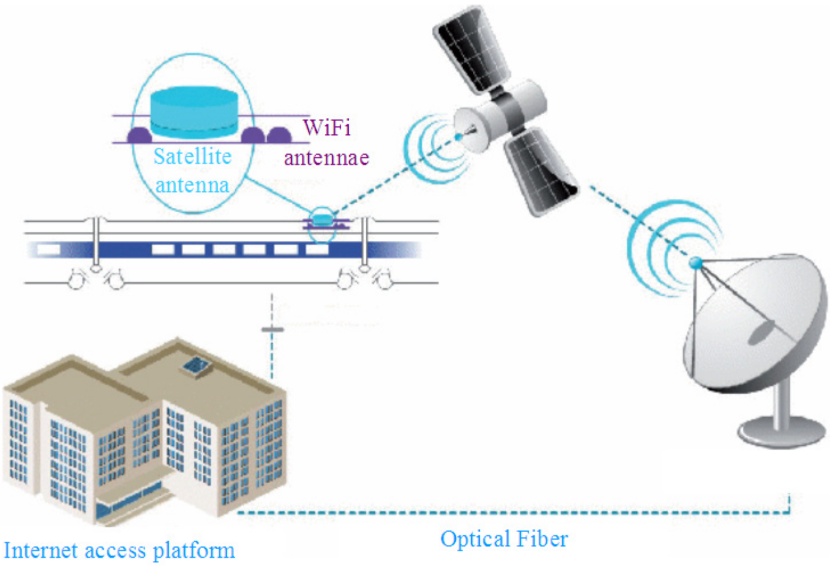


Fig. 2. "BoxTGV" architecture

Another satellite systems issue is operational expenditure (OPEX) related to satellite capacity, which can be seen as a rare resource. Since the available aggregated data throughput for the customers will depend directly on the contracted satellite capacity, the regular costs generated must be taken into account in the business model of the foreseen service. However, customers' needs grow with time and will ask for more and more satellite bandwidth for the service, thus generating important additional operational costs. If having more customers using the service can generate more income, having the same number of customers asking for mote throughput will not generate more income, thus representing a serious problem for the business model of the service. The new generation of Ka Band satellites [3] could be the answer to this issue since their larger capacity allows the reduction of the satellite bandwidth cost. Another advantage of Ka Band satellites is that the higher frequencies used for the transmissions allow using smaller antennae, what can be a major issue in function of the considered rolling stock. However, using Ka Band satellites has also some drawbacks:

- Ku band equipments are not compatible, new antenna and transmission/reception equipments must be developed for the strong railway domain constraints. Pre-existent Ku equipments will need to be changed.
- Atmospheric attenuation in case of bad weather conditions (mostly fog, rain and snow), will be much higher, bringing deep fading effects and persistent attenuations to the signal that can reach 15 dB in the worst cases.
- Existent Ka Band satellites use (typically) 250 to 500 km cellular coverage in order to be able to reuse frequency bands geographically (to optimize satellite capacity). Dynamic frequency allocation and horizontal handover techniques will be needed in order to assure a seamless connectivity for trains passing from one cell to another. The operation of the global network will be then much more complex.

2.2 Terrestrial Solutions

Terrestrial technologies that can be used in order to deliver Internet access to trains passengers can classified in two categories: train-to-trackside communications and cellular communications.

2.2.1 Cellular Solutions

Cellular solutions could be a very interesting candidate in order to provide reliable two-way train-to-ground communications. These solutions are typically based on one onboard mobile router connected to multiple antennae that have to be installed on the roof of the train. The mobile router can generally accommodate multiple SIM modules to allow aggregation of data on different available mobile operator networks or to allow cross border operation. Cellular solutions can be seen as very attractive compared to satellite solutions where the bandwidth price is very high and the needed onboard antenna is much larger. However, cellular solutions require sufficient radio coverage and an appropriate capacity of the cells along the railway lines. Besides, these technologies are subject to licensing and the operation and performance of the service will depend on the coverage and operation modes decided by mobile networks operators.

Nomad Digital [4] and Icomera [5] are the leading providers of railway train-to-ground connectivity based on cellular solutions.

The following figure shows an example of an onboard system architecture based on cellular solutions with two cellular antennae installed on the roof of the train and a WiFi network to deliver passengers' service.

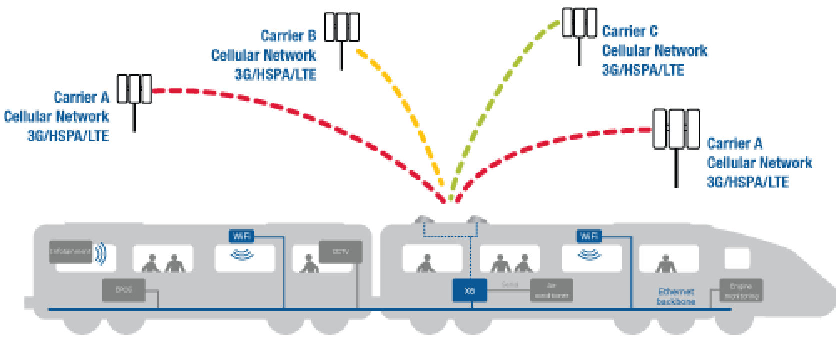


Fig. 3. Cellular solution onboard architecture [5]

One of the main advantages of cellular solutions is their ability to be easily upgraded to new technologies. Current 3G mobile routers can easily evolve to be able to run on future LTE [6] networks. In addition, geographical coverage is progressively improved by mobile operators, making this kind of solutions more and more interesting.

2.2.2 Train-to-Trackside Communications

Current satellite and cellular solutions are able to deliver connectivity speeds of a few Mbps to a single train. These throughputs are comfortable for the delivery of Internet access to some tens of passengers, but needed data-rates are growing very fast along with users' demand. The number of passengers willing to connect will grow very rapidly in the next years, in particular due to the very large success of smartphones. To meet this future growing demand for this kind of services onboard, mainly in terms of bandwidth, train-to-trackside communication solutions can be used. However, this kind of solutions relies on long and complex deployments of infrastructures along the track which are in general very expensive. In order to reduce deployment costs, it is essential to try to minimize the number of sites required to ensure radio coverage of the high speed railway network. The radio coverage is thus one of the main features to be taken into account when choosing a terrestrial communication technology. Another key feature to be considered is data rate offered by that technology, since the objective is to provide the highest possible throughputs for high speed train passengers. These two features, radio coverage and data rate, are closely related. Therefore, the more suitable technologies will be the ones offering the best trade off between range and throughput. Obviously, other features have to be studied as well, including reliability, security and need for a licensed radio spectrum. Among the various terrestrial technologies, 3G and 4G may be considered. But these technologies are subject to licensing and the operation and performance of the service will depend on the coverage and operation modes decided by mobile networks operators. Proprietary technologies like Telefunken Racoms [7], Flarion or Reicom [8] may be considered as well. These technologies deliver good performance and, in some cases, have been specifically designed for railway applications. Their major drawbacks are the dependence on a single manufacturer and the deployment costs, which are considerably higher than those of standard technologies like WiFi for example. Among the various terrestrial technologies, WiFi IEEE802.11 is a very interesting candidate. Indeed, it's an unlicensed, well known and inexpensive technology delivering very good performance even at high mobile speeds. Previous WiFi-a/b/g standards were studied few years ago by SNCF and Orange Labs and the tests carried out in early 2005 showed that WiFi is one of the more suitable terrestrial technologies that can be used for train-ground high data rate communications. The more recent generation of WiFi standards is WiFi "n" [9], which offers a better performance trade-off in terms of coverage and throughputs thanks to the use of state of the art techniques on antennae and signal processing, at the physical and MAC layers. This standard has been studied by SNCF and Orange Labs through an experimentation campaign between Q4 2009 and Q1 2010, using some of the first 802.11n products in the market in order to implement broadband train-to-ground connectivity. The next figure shows the 2-sites architecture that has been implemented for the experimentation campaign. The results of these preliminary tests were very interesting with throughputs up to some tens of Mbps over 5 km coverage area which is close to the average distance between 2 consecutive GSM-R sites in France [10].

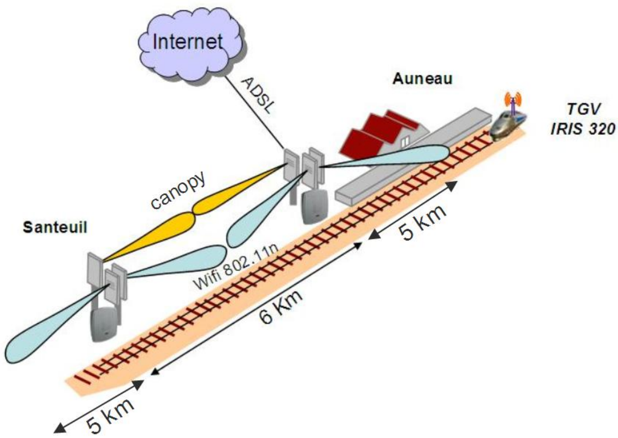


Fig. 4. WiFi'n' train-to-trackside communication architecture

Laser technology could also be used for train-to-ground communications. This technology has been tested by the Railway Technical Research Institute in Japan with the collaboration of JR-West and Keio University [10]. The obtained transmission rate was of approximately 700 Mbps on the TCP layer between the ground and the train running at a speed of approximately 130 km/h. However, the major drawbacks of such solution are the small transmission distance (300 or 400 meters), the maintenance of such system and its high dependency to environment conditions.

An interesting radio technology that should be considered as a perspective for train-to-trackside communication is CR (Cognitive Radio) technology. CR technology is an emerging intelligent radio technology that can observe its environment, detect spectrum gaps and adapt its parameters to the environment in order to guarantee an optimal radio link performance. SNCF is currently participating to a French national research (Agence Nationale de la Recherche) project named "CORRIDOR" that aims to study the use of CR technology for Railway applications, especially in high-speed context. After the definition of the railway user needs carried out in 2012, in 2013 the project will focus on the development of the intelligent onboard terminals and of the intelligent trackside infrastructure equipments, before demonstrating the implemented system onboard a high speed train in France in 2014. The three service scenarios considered by the project are:

- Very high data rate Internet access for passengers
- Train-to-ground transmission of real time video flows from onboard CCTV systems
- Reliable low data rate link for railway safety-related applications.

3 Conclusions and Perspectives

Onboard internet access has become a standard of comfort for trains' passengers. To meet this need, railway operators could use different solutions for train-to-ground communications. These solutions include satellite, cellular or train-to-trackside solutions. This paper gives an overview of these different solutions and presents their main advantages and drawbacks.

SNCF deployed in 2010 its first commercial Internet service for passengers based on satellite train-to-ground solution as it was the most suitable solution in France. This service will need to evolve to meet the foreseen evolution in passengers demand. In this context, cellular and train-to-trackside solutions could be considered.

References

1. Sanz, D.: Satellite Technologies for Broadband Internet Access Onboard High Speed Trains, WCRR 2006, Montréal (2006)
2. `https://boxtgv.voyages-sncf.com/`
3. `http://www.eutelsat.com/satellites/EUTELSAT-KA-SAT-9A.html`
4. `http://www.nomadrail.com/`
5. `http://www.icomera.com/`

6. http://www.3gpp.org/LTE
7. http://www.tfk-racoms.com/en
8. http://www.reicom.it/
9. http://standards.ieee.org/getieee802/download/802.11n-2009.pdf
10. Ghannoum, H., Sanz, D., Villeforceix, B., Philippe, H., Mercier, P.: De-livering broadband internet access for high speed trains passengers using the new WiFi standard 802.11n for train-to-ground communications. In: 9th World Congress on Railway Research, Lille (2011)
11. Nakagawa, S., Matsubara, H., Nakamura, K., Tatsui, D., Haruyama, S., Teraoka, F.: Broadband Telecommunication System for Railways Using Laser Technology. In: 9th World Congress on Railway Research, Lille (2011)
12. CORRIDOR project, http://corridor.ifsttar.fr/

A QoS-Based Multi-user Scheduler Applied to Railway Radio-Communications*

Nicolas Gresset, Jonathan Letessier, and Hervé Bonneville

Mitsubishi Electric R&D Centre Europe, Rennes, France

Abstract. This paper investigates what role could play the 3GPP LTE in railway radio-communications. We study a first case of shared infrastructure where the train control traffic is subcontracted to a public land mobile network operator, and a second case of dedicated infrastructure where the LTE-like network only supports the railway communications. We highlight the multi-user scheduler as the key enabler for managing a heterogeneity of sensitive and best effort traffic, and propose a new scheduler that maximizes the best effort traffic throughput while guarantying the sensitive traffic quality of service.

1 Introduction

The history of safety in trains started with the train era. At first, track side signals have been designed and deployed to limit train speed or to require the train to stop; the train driver has to obey to the signal and act accordingly. Soon, the necessity emerged to enforce the control, by designing a system making the train automatically stop in case the driver missed the signal or did not obey to it. The system evolves to in-cabin report of speed and stop signaling. It is however a coarse speed control, with only a few levels of speed limits. Hard blocks definition and enforcement allows controlling distance between trains, to prevent collisions.

With the increased train speed, track-side indications become incompatible with human view and reaction time, leading to the necessity of a better in-cabin signaling report. Hard blocks limit the train density a track is able to handle securely, although the always increasing need of sustainable transports implies higher railways densities. Turning to moving blocks with a high performance safety control provided by an Automatic Train Control (ATC) system (e.g., [6]) becomes a necessity. For example, the US Congress has requested the implementation of PTC (Positive Train Control) systems, as a minimum ATC targeting security issues, by end of 2015 for heavy rail (freight and passenger services) [4]. The CBTC (Communication-Based Train Control) is a more complete version of PTC enhancing security, capacity and sustainability, and is described in the IEEE Std 1474.1-1999 [5].

* This work was performed within project SYSTUF, which is subsidized by the French ministry of Industry in the framework of the AMI ITS program.

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 31–45, 2013.

A wireless communication between train and ground allows a permanent and fine control. The wireless system can be based on a proprietary radio layer, on a wireless-LAN standard such as IEEE 802.11, or be inherited from cellular area such as GSM-R used in European ERTMS. In addition to vital applications such as CBTC, train operators have identified new needs in their effort to enhance public transports [1], such as Closed Circuit TeleVision (CCTV) and on-board public Internet access. The Quality of Service (QoS) and capacity requirements of these new services will require new wireless telecommunication technologies. This paper investigates the possible role of LTE-like systems in future railway radio-communications.

In Section 2, we compare the advantages and drawbacks of several communication infrastructures for ATC. We also present the typical services of ATC. In Section 3, we make a brief overview of the LTE OFDMA multiple access technique and of scheduling techniques, and propose in Section 4 a throughput maximizing QoS-aware scheduler taking into account the heterogeneity of traffic. In Section 5, we evaluate the impact of transmitting train control traffic in a shared and dedicated infrastructure. Finally, we conclude by giving the pros and cons of shared vs dedicated LTE infrastructures.

2 Automatic Train Control

2.1 Telecommunication Infrastructure for Automatic Train Control: Shared vs Dedicated

The communication infrastructure for train control is usually dedicated, i.e., owned and operated by the train operators. Indeed, they have the will to keep full control and responsibility of the security and safety of the passengers and trains. However, the communication infrastructure represents a significant cost in terms of deployment, operation and maintenance.

The 3GPP LTE is expected to become a worldwide widespread wireless communication toolbox, which could be used outside the Public Land Mobile Network (PLMN) domain of application. Indeed, it fulfills all the needs of modern communication: first by the flexibility and robustness of the LTE physical (PHY) (see, e.g., [7]) and Media Access Control (MAC) layers; and also by its ability to offer in a secure fashion a wide variety of services and applications through an IP-based core network. Thus, a dedicated LTE-like macro-cellular system is a good candidate for the evolution of railway radio-communications. A dedicated bandwidth might be necessary in that case for ensuring a good quality of service for the cell-edge users. A compromise exists between performance, CAPEX, OPEX, and band licensing when comparing such systems to proprietary deployments based, e.g., on 802.11 technologies on ISM bands.

In order to reduce CAPEX and OPEX, it could be envisaged to open the train operator infrastructure to other Public Safety Communications (PSC) services, offering for example an access to police forces in underground stations. In another example, the USA recently imposed LTE for US global PSC[8], which confirms the trend of non specific radio technologies. In a more futuristic approach, the available resource could also be shared with a PLMN operator, but the train control traffic must be prioritized in order to fulfill QoS metrics specific to railway communications. Those metrics are described in Section 2.2 and prioritization is performed via multi-user scheduling, as described in Section 3.

In yet another approach, the communication part of the train control could be subcontracted to a PLMN operator. Thus, this would reduce the costs, but raises the issues of quality of service, prioritization of subcontracted services, and responsibility. Indeed, the LTE system capacity could reach its limit if several critical services, such as train control, intelligent transportation systems, smartgrids communications, PSCs, machine type communications (MTCs), are managed by the same PLMN. It becomes the responsibility of the PLMN operator or governmental entities to prioritize such services, which is a rupture with the current way to manage ATCs.

In order to evaluate the pros and cons of shared vs dedicated LTE infrastructure for train communications in terms of performance, we define in next Section the QoS requirements of various train communication applications.

2.2 Services and QoS Metrics for Railway Communication Services

Today's railway communication systems mainly address vital operational services, such as CBTC. In a near future, new services will be required [2]. In a category of applications classified as non-vital operational, we can find CCTV. It can be used to monitor passengers (on-board monitoring), but also to monitor tracks by reporting in advance any obstruction like a car stuck in a level crossing. CCTV is also used to monitor train doors while stopping at a station. Train maintenance, including real-time monitoring from ground and alarm reports, or on-board system configuration before train departure, fall also in the non-vital category, along with on-board passenger information. Multimedia services as Internet access to passengers and on-board television are considered in a third category named comfort applications.

These different types of traffic are very heterogeneous in nature and requirements. CBTC traffic typically requires a bandwidth of 100 kbps, but is a vital application and as such shall be treated with the highest priority. A CCTV camera generates a few - say 1 or 2 - Mbps of data. Applications like train configuration or on-board television program upload behave as bulk file transfers, requiring a high throughput in a rather short period of time. Internet accesses offered to passengers is best-effort and it is admitted today that 20 to 30 simultaneous users can share 1 Mbps of bandwidth [3]. However, Internet usage is evolving rapidly and the acceptable level of Internet access service is likely to be more bandwidth-demanding in the future. Voice services can be split into the operational category for communications dedicated to the train crew or into the comfort class when service is offered to passengers, and should be treated with different priorities. Usually, these different applications are supported by specific communication systems being able to merge them into a unique one would be a definite advantage. Today's communication technologies should be able to support the required bandwidth. They shall be capable of providing the specific QoS required by the various railway services in order to be accepted by train operators. In this paper, we only consider CBTC and CCTV services as prioritized services, and assume that a maximum latency of 50ms is required at the MAC layer. This latency is defined as the time interval between the packet arrival at the base station and its transmission.

3 Multi-user Scheduling in LTE Systems

3.1 An Overview of OFDMA in LTE

In the 3GPP LTE cellular network, the multiple access technique in downlink is based on OFDMA, the time resource being divided into frames comprising ten sub-frames, and the frequency resource within a sub-frame is divided in PRBs (Physical Resource Block). One of the main advantages of the OFDMA technique with respect to more static multiple access techniques such as TDMA, FDMA or CDMA is the ability to dynamically allocate resources to the users according to their QoS requirements and wireless channel quality. The user scheduling decision is at least partially centralized, the 3GPP LTE standard being built under the assumption that the scheduling operation is performed at the base station (eNB) of each cell. A limited amount of information is exchanged between neighboring cells in order to apply inter-cell interference coordination (ICIC) on a long term basis. In future deployments, such as for the 3GPP LTE-Advanced system and beyond, larger coordination and centralization are expected in order to further improve the system performance, in particular at the scheduler level.

The per-cell scheduler collects channel state information (CSI) on the frequency and time selective channel, which is obtained by an uplink feedback of downlink measurements. We consider that the CSI is quantized at the PRB level. One User Equipment's (UE) CSI can be classified as short-term when the channel varies sufficiently slowly in the time domain for the scheduler to have a relevant knowledge of the UE's instantaneous channel realization. When the channel fluctuations in the time domain are quicker than the feedback delay from the UE to the eNB, the scheduler must take the decision based on a long term CSI which usually relates to the Signal to Interference plus Noise Ratio (SINR) averaged over time. As a result of ICIC, the SINR can be frequency selective.

When the UEs are not co-located, they observe frequency selective channels which are different one from each others. This allows in the best case, when the scheduler relies on short-term CSI, to allocate the PRBs providing the best performance to each UE without any conflict. As a result, transmitting the data of each user only on the best sub-carriers of the OFDM channel exploits the frequency diversity. Such an opportunistic frequency scheduling operation is also called a multi-user diversity technique. Unfortunately, in high traffic load, conflicts occurs and must be solved by the scheduler.

3.2 QoS-Aware Scheduling in OFDMA Systems

In practice, the resource allocation process is highly connected with the packet fragmentation and retransmission at the radio link control (RLC) layer and with the link adaptation that determines the transmission parameters at the physical layer (PHY). In this paper, we simplify the traffic classification with only two types: Best Effort Traffic (BET) and Sensitive Traffic (ST).

The BET is not associated to any QoS metric such as the latency or the packet error rate. The applications that generate BET are typically web browsing, ftp, file sharing, e-mails and social networks. It is also often considered that the buffers at the RLC are always full, which means that the scheduling decision can be PHY-oriented: the most spectrally efficient resource and transmission scheme are used and the amount of data that can be transmitted is pulled out from the RLC buffers accordingly. When the UEs are at low speed, the short-term CSI feedback can be used in order to take benefit from the multi-user diversity. When the UEs are moving (generally above 20km/h), the scheduler relies on the long-term SINR for deciding on the resource allocation and link adaptation. Three standard schedulers, among a large variety [9][10], are usually considered for PHY-oriented decision on full buffer BET:

- The *round robin* scheduler multiplexes the users on the PRBs. Resource fairness is guaranteed between users, but the multi-user diversity is not exploited which degrades the spectral efficiency.
- The *max-SINR* scheduler allocates a PRB to the user with the best SINR. Resource fairness is not obtained, and the multi-user diversity is exploited, which maximizes the spectral efficiency but not the users throughput.
- The *proportional fair* scheduler makes a compromise between spectral efficiency and fairness by selecting, for each PRB, the user for which the ratio between the instantaneous transmission rate over the average transmission rate is the highest.

The ST is associated with latency, jitter and error rate constraints. We assume that the traffic is periodic, i.e., that the scheduling decision is packet-oriented. In other words, the packets must be sent within a given time window and are pushed out from the RLC to the scheduler. The applications that generate ST are typically VoIP, online gaming and video streaming; train control would be one of them. The resource is usually allocated to ST users in a round robin fashion.

The simplest way to manage such an heterogeneity of traffic is to assign different priority levels to the packets according to their traffic type, and perform the resource allocation starting from the highest priority packets. Unfortunately this does not allow for optimizing the QoS metrics for all users. Indeed, by using a scheduler that first allocates resource for ST users, the degrees of freedom for scheduling BET users are reduced. In extreme cases, ST users are allocated to the highest capacity resource of BET users, which decreases the system spectral efficiency. Another choice of resource might be more optimal and still satisfying the ST QoS constraints. On the contrary, by first allocating resource to BET users without any checking might not allow to fulfill the error rate or latency requirements of ST users.

4 A QoS-Aware Multi-user Scheduler Maximizing Throughput

In this Section, we propose a multi-user frequency scheduler that jointly ensures latency constraints of the ST users and throughput maximization for the BET users. We consider that ST packets are pushed out from the RLC and scheduled in a packet oriented fashion while BET packets are pulled out and scheduled in a PHY-oriented fashion. The main principle of this scheduler is to allocate the best resource to the BET packets first and schedule the ST packets as late as possible in the decision process (not in the time domain) while guaranteeing their QoS with a checking function.

4.1 General Algorithm

The scheduler has a buffer of the frequency/time table of the resource allocation. This table stores, at a given time, the packet index that must be transmitted on a given PRB. It is organized into sub-frames, aligned with the transmission structure defined at the PHY layer. The scheduler has an internal timer synchronized with the PHY layer frame rate. When the time has come to provide the next frame to the PHY layer for transmission, a buffer shifting is performed and a new empty set of PRBs is inserted at the end of the buffer. We assume that the so-called RLC-packets are provided by the RLC buffers, and segmented into packets of one PRB as presented in Section 4.2. We also assume that when one RLC-packet is pushed out from the RLC, it is tagged with a delay constraint related to the latency requirements of the application.

When a new packet belonging to a ST user arrives at the input of the scheduler, it is stored in a ST buffer and a *PRB tagging* function updates a PRB tag table according to the ST QoS. The PRB tagging is updated after each new allocation or after any new buffer shifting after a frame transmission to the PHY layer.

In parallel, before applying a BET scheduling step on one free PRB of the buffer (e.g., with *Round robin*, *max-SINR*, *Proportional fair*), a *Resource Checking* is performed for the buffered ST packets based on the PRB tag tables. It allocates resource for the ST packets only when necessary, i.e., when any new BET allocation would result in resource shortage for ST packets scheduling.

The PRB tagging allows for defining which PRBs that are not already allocated in the scheduler table satisfy the QoS constraint, such as latency, jitter and performance metric. The performance metric can for example be a capacity threshold, or an average error rate being a function of the SINR and packet payload. For each ST packet, the PRB tagging function creates or updates the packet's PRB tag table which stores the indexes of the subset of PRBs satisfying the QoS constraint. The packet is dropped when its table is empty.

The PRB tag tables are illustrated in Fig. 1, where three ST packets are considered. The capacity metrics C_1, C_2 and C_3 of the three packets are frequency selective and are assumed constant in time for this example. The frequency allocation is assumed possible when the capacity metric is above a performance threshold. The three considered ST packets have latency constraints of 4, 2 and 6 sub-frames, respectively. Thus, the PRBs for the ST packet P_i are tagged when C_i is larger than the threshold and the sub-frame index in the scheduler buffer satisfies the latency constraint.

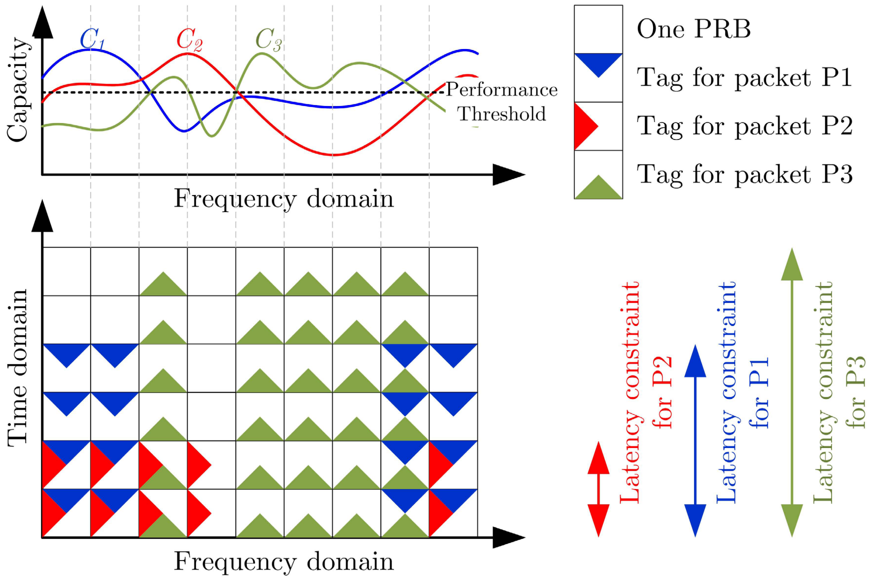


Fig. 1. PRB Tag function for three ST packets

When the future frame is sent from the MAC to the PHY layer for transmission, the PRB tag tables are updated accordingly: the sent PRB indexes are removed from the tables, the PRB indexes are shifted by the number of PRBs by frame, the PRB tagging is performed for the new PRBs inserted at the end of the buffer and the delay constraints associated to the ST packets in the ST buffer are decreased by a time frame duration.

The Resource Checking prevents any resource shortage for ST packets after a future BET packet allocation.

The Algorithm 1 illustrates the resource checking before a BET allocation for a given PRB. Let us assume that N packets are pre-allocated to the given PRB, and their corresponding tag tables $T_{1\leq i\leq N}$ are known. The algorithm first applies

```
Data: N, T_{1≤i≤N} PRB tag tables
for K ← 1 to N do
    for ω ∈ Ω(K) do
        n_t = card (⋃_{i∈ω} T_i);
        if n_t = K then
            call K-allocation(ω);
            return 1;
        end
    end
end
return 0;
```

Algorithm 1: Resource Checking algorithm

a loop on an increasing $1 \leq K \leq N$ value. Let $\Omega(K)$ be a set of cardinality $\binom{K}{N}$ comprising all possible index vectors of cardinality K. The algorithm checks, for all possible vectors $\omega \in \Omega(K)$ of K indexes, that the cardinality of the union of the K considered tag tables $T_{i\in\omega}$ is larger than K. If not, the K identified ST packets at stake must be allocated on the K identified PRBs, the *K-allocation* function is called, and the *Resource Checking* is restarted. Whatever PRB is allocated to a BET packet, if the condition is satisfied for all K values and sets of K PRBs, another allocation of the ST packets candidate on this PRB is possible. Thus, the scheduling can process a PRB allocation for the BET traffic while guarantying the QoS of ST traffic.

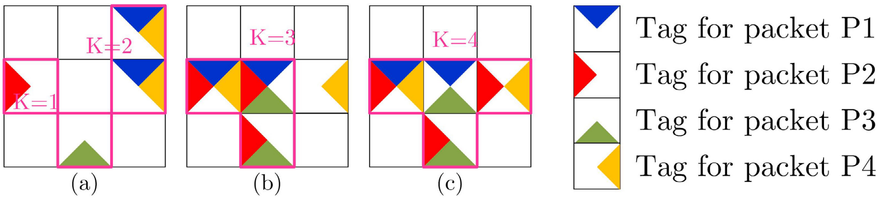


Fig. 2. Examples of resource checking leading to the call of the K-Allocation function

Fig. 2 illustrates three toy examples of the PRB tag tables of four ST packets leading to the call of the *K-allocation* function. In Fig. 2-(a), the *Resource Checking* step with $K = 1$ leads to allocate PRBs for packets 2 and 3. The resource checking for packets 1 and 4 leads to the call of the *K-allocation* with $K = 2$. In Fig. 2-(b), the *K-allocation* function is called after the *Resource Checking* for $K = 3$ and the considered PRBs are allocated for packets 1, 2 and 3. After update of the tables, the packet 4 is allocated after the *Resource Checking* for $K = 1$. In Fig. 2-(c), the *K-allocation* function is called after the *Resource Checking* for $K = 4$ and allocates the 4 PRBs to the 4 packets.

The K-Allocation function assumes that the union of K candidate PRB tag tables contains exactly K PRBs, but that each table has a cardinality lower or equal to K. At least one solution of a possible allocation is found by computing the number m_i of packets that can be allocated to the i-th PRB in the union of the K PRB tag tables. The algorithm loops on the PRBs of the union in the increasing order of m_i. For each PRB i, it allocates a ST packet, the table of which has the lowest cardinality among those containing the i-th PRB, and updates the PRB tag tables and m_i values accordingly. At the end of the algorithm, all K packets are allocated to the K PRBs.

4.2 Packet Segmentation and Optimized Pre-allocation

In the previous Section, we have proposed a scheduler relying on tag tables for ST packets and checking the resource before any PRB allocation for a BET packet. The complexity and performance of the scheduler depends on the packet segmentation into one PRB-length packets, and on the pre-allocation that builds the tag tables.

The packet segmentation divides the ST RLC-packet pushed out from the RLC buffer, and that must be scheduled within the time-frame of the scheduler buffer, into packets of one PRB. The segmentation relies on an allocation metric depending on the expected performance of the transmission at the PHY layer.

When the users are moving at low speed, we consider that the CSI feedback allows to know the channel at the transmitter (closed loop). In that case, we consider the channel capacity as the allocation metric for scheduling. We consider in this paper a packet segmentation based on the payload resulting from a pre-allocation to PRBs in a round robin fashion. Each time a PRB payload (number of bits it can carry) is computed from the channel capacity, it is subtracted from the RLC-packet payload until all RLC-packets are segmented.

For each user moving at high speed, only the SINR is known at the transmitter (open loop). In that case, we consider the average rate obtained from the outage probability as the allocation metric: for ST traffic, we compute the average rate for each SINR such that the outage probability reaches a target QoS of 1e-2; for BET traffic, we find for each SINR the rate R maximizing $(1 - P_o(SINR))R$ where $P_o(SINR)$ is the outage probability of the channel for a given SINR. In this case, all PRBs can carry an identical number of bits, and the ST RLC-packet payload is equally distributed among PRBs.

The optimized pre-allocation limits the cardinality of the tag tables. Indeed, the complexity of the *resource checking* step can become intractable when the cardinality of the union of the tag tables grows. In order to reduce this complexity, we can limit the number of possible pre-allocation and not consider all PRBs that satisfy the QoS constraint for the packet. As a remark, it is sufficient to have one pre-allocation per tag table to satisfy the QoS constraint for the ST packet. Thus, reducing the number of pre-allocation decreases the flexibility in the allocation of the BET packets, and reduces the multi-user diversity for this traffic type. The trade-off between the performance of the BET users and the complexity of the scheduler must be optimized. We propose to limit at maximum the number of possible allocation per tag table, but to maximize the cardinality of the union of the tag tables associated to one PRB. In other words, all users sharing a PRB pre-allocation should not share other PRB pre-allocations. In order to reach this condition, we first apply the RLC-packet segmentation into one PRB-length packets as described above, which defines a payload per packet. Then, we consider that each packet must be pre-allocated to M PRBs. The PRBs are selected sequentially, and for each PRB, we randomly choose one packet, the payload of which is supported by the PRB, if any. We then decrease the number of pre-allocation of this packet by one unit, until all packets are pre-allocated M times. In extreme cases, one packet can be allocated several times to the same PRB. By using this pre-allocation strategy, we manage to have large degrees of freedom in the allocation of BET users. Furthermore, simulation results also show that the *resource checking* has quasi optimal performance even when we limit the search to small values of K (e.g., $K < 3$) and to K equal to the cardinality of the union of the tag tables associated to the considered PRB.

5 System Level Simulations

In this Section we consider the downlink of a macro-cellular LTE system at 2GHz. We consider a uniform and random deployment of UEs in a static multi-cell system-level simulator following the 3GPP case 3 parameters [11]. The small-scale Rayleigh channels are the ITU 6-path Typical Urban channel model. In all simulations, ideal channel estimation and measurements feedback are assumed. We consider a 10MHz bandwidth divided into 50 PRBs of 12 sub-carriers, with a sub-carrier spacing of 15 kHz. In time domain, the smallest unit is the slot composed of 7 OFDM symbols. One sub-frame of 1 ms has to 2 slots. The channel is considered quasi-static for non-moving UEs. We make 1000 snapshots of UEs positions, comprising N_{BET} UEs with a BET traffic and N_{ST} UEs with a ST traffic (trains in our case). The BET traffic satisfies the full buffer assumption, and the ST traffic has a MAC latency constraint of 50ms. We consider three possible data rates: 100kbps for CBTC-like traffic, 1Mbps for CCTV, and 4Mbps for CCTV with several video streams.

First, we will consider a shared infrastructure where $N_{BET} = 20$ regular PLMN LTE users per cell, for each snapshot, use a BET service. We assume that several trains lie in the cell, and use ST services.

In Fig. 3, we first consider the case with no ST traffic as the reference performance. Then, we consider that two trains at low speed receive downlink communications with a throughput of 4Mbps each. We consider here an extreme case in order to illustrate the gains obtained by the proposed scheduler. In some of the snapshots, the 10MHz LTE capacity is not sufficient to support the ST throughput. This happens when the trains are located at the cell edge and require a too large number of resource. In average, 7% of the ST packets are dropped and BET

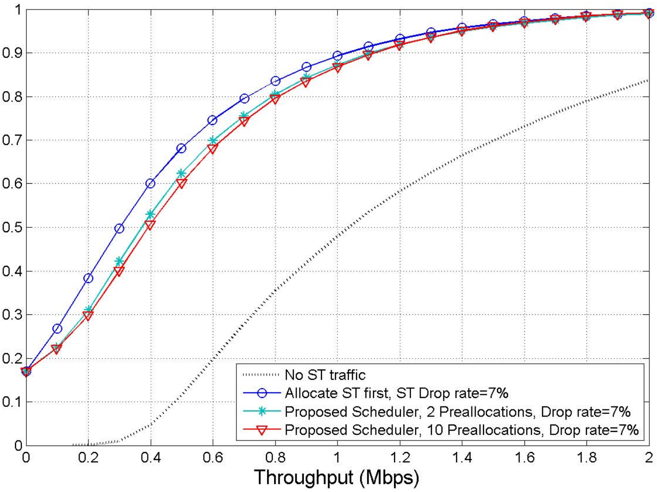


Fig. 3. Comparison of the c.d.f of the BET users throughput, 20 BET users per cell, 2 low-speed ST users with 4Mbps traffic

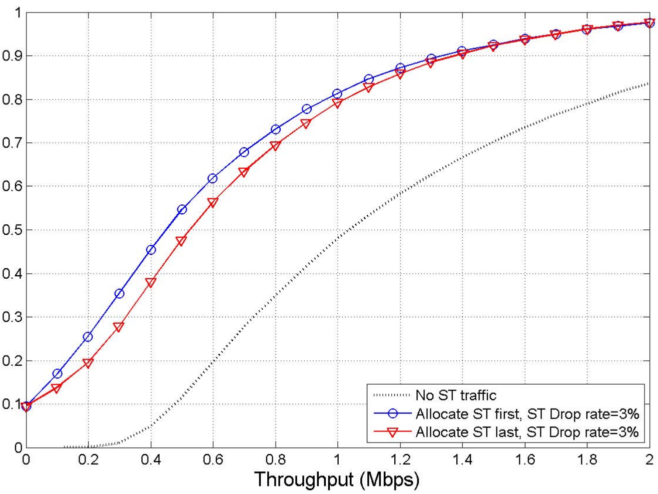


Fig. 4. Comparison of the c.d.f of the BET users throughput, 20 BET users per cell, 2 high-speed ST users with 1Mbps traffic

users are not served in 18% of the cases. When using the state of the art priority-based scheduling, a large loss is observed on the BET users. When maximizing the throughput with the proposed QoS-aware scheduler, the BET throughput is improved with respect to the state of the art. This illustrates that, by allocating the ST packets only when necessary, the BET scheduler has more degrees of

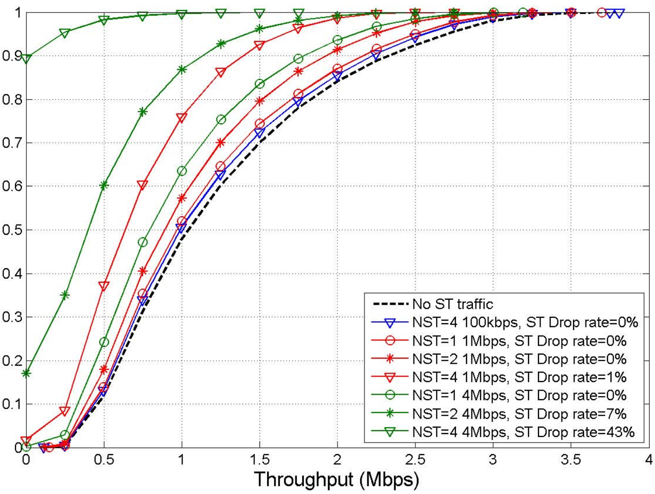


Fig. 5. Comparison of the c.d.f of the BET users throughput, 20 BET users per cell, and low-speed ST users

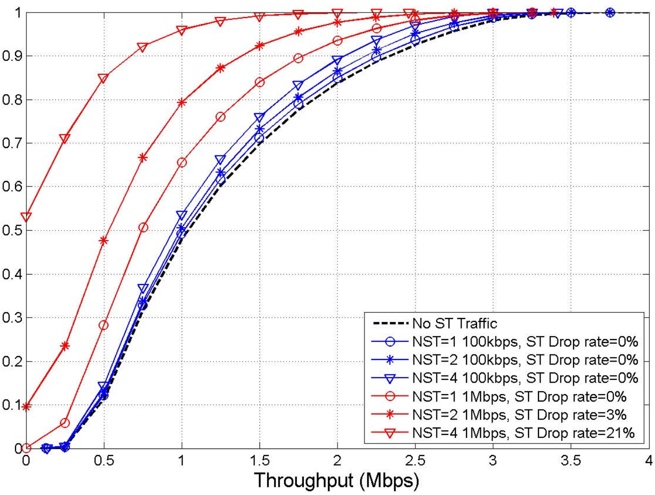


Fig. 6. Comparison of the c.d.f of the BET users throughput, 20 BET users per cell, and high-speed ST users

freedom. We consider two cases with $M = 2$ and $M = 10$ pre-allocations per ST packet. We observe that two pre-allocations are sufficient to reach the maximum throughput constrained by the ST QoS, which drastically limits the proposed scheduler complexity. Finally, we also see from the optimized BET throughput that the main factor of degradation of the BET users is the resource loss, and not the multi-user diversity loss. In Fig. 4, we consider that two trains at high speed receive downlink communications with a throughput of 1Mbps each. In

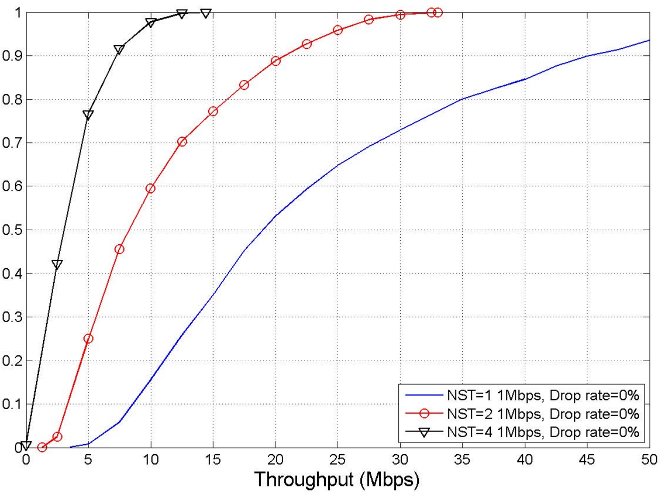


Fig. 7. Comparison of the c.d.f of the on-board BET throughput, with low-speed ST users

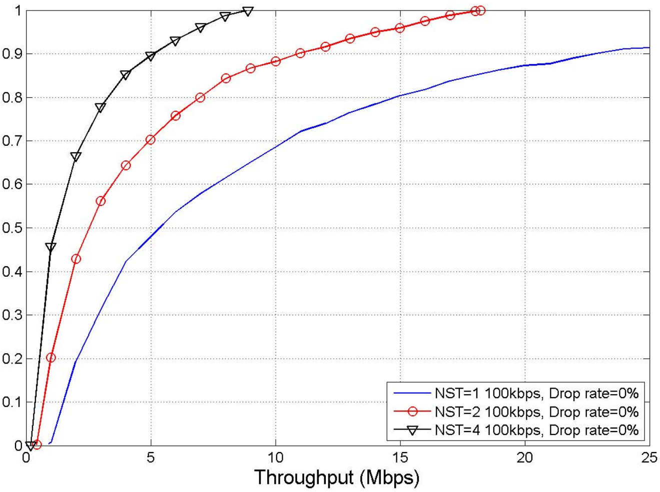


Fig. 8. Comparison of the c.d.f of the on-board BET throughput, with high-speed ST users

average, 3% of the ST packets are dropped and BET users are not served in 10% of the cases. When the allocation metric for the ST traffic is based on the wideband SINR, no PRB is better than another in the latency window. The QoS-aware scheduler is simplified to an allocation of all ST packets that fills the buffer as late as possible in the decision process. As for the low-speed case, a gain of 20% on the BET throughput is observed with respect to the state of art, and the resource taken for ST traffic has the main impact on the loss. We have shown that the proposed QoS-aware scheduler optimizes the BET throughput, with a reduced complexity when used along with an optimized packet segmentation and pre-allocation. By using such an optimized scheduler, we guarantee that the conclusions taken on the impact of the train communications to the PLMN users are relevant. For an even better case study, ICIC techniques should be taken into account. It has to be noted that in many low ST load scenarii, the proposed scheduler shows no gain with respect to the state of the art.

In Fig. 5, we consider 1, 2 or 4 low-speed trains with ST throughput of 100kbps CBTC, 1Mbps CCTV or 4Mbps (high quality or several) HQ-CCTV. First, the CBTC at 100kbps has no impact on the BET throughput for four trains, and this is also the case for one and two trains in the cell which are not plotted on the figure. When the trains use CCTV at 1Mbps, the impact on the BET traffic becomes non negligible with two trains and very important with four trains. The results clearly show that the 10MHz LTE capacity cannot support HQ-CCTV at 4Mbps. In Fig. 6, we consider 1, 2 or 4 high-speed trains with ST throughput of 100kbps (CBTC) and 1Mbps (CCTV). We can see that the impact of the CBTC traffic, even for trains at high speed which consumes more resource, is limited. However, the impact of CCTV is non negligible, even for one train. In conclusion of Fig. 5 and Fig. 6, a 10MHz LTE system can support CBTC with a guaranteed QoS for up to four trains with a negligible impact on the PLMN LTE users. When the train approaches the stations at lower speed and requires CCTV services, the impact on the PLMN LTE users becomes large.

We now consider a dedicated LTE network, where the only users are trains. We assume that the leftover resource is used for non critical traffic, such as on-board Internet access, and place a BET user in each train.

In Fig. 7, we consider that several low-speed trains are using CCTV, and observe the cdf of the throughput for the on-board BET user. The ST packets are never dropped, which shows that the QoS is guaranteed. The on-board BET throughput suffers from the resource sharing between several trains. From the assumption that twenty users can share a 1Mbps Internet access with a good experience, when four low speed trains are considered, an average throughput of 5Mbps can be shared between hundred users per train. In Fig. 8, we consider that several high-speed trains are using CCTV, and observe the cdf of the throughput for the on-board BET user. Since the BET users are also high-speed, we use the wideband SINR-based allocation metric. The BET throughput is drastically reduced with respect to the low-speed case, a loss factor of around four is due to the non-exploitation of the channel knowledge at the transmitter (open loop). When four low speed trains are considered, an average throughput of 1Mbps can be shared between only twenty users per train with a correct Internet experience. In conclusion of Fig. 7 and Fig. 8, a dedicated 10MHz LTE system can support CBTC and CCTV services with a guaranteed QoS. Non critical traffic can also be provided on-board, with a significant throughput at low speed and a reduced yet correct one at high speed.

6 Conclusions

In this paper, we have shown that the CBTC service can be provided with guaranteed QoS and no impact on the PLMN users of a shared infrastructure. The CCTV or other high throughput services, even when only activated at low speed, have a strong impact on the PLMN users' throughput. In a dedicated infrastructure scenario, we have shown that the sensitive traffic's QoS is always guaranteed, and that additional throughput can be offered on-board, especially at low speed. Thus, the dedicated LTE infrastructure is a potential candidate for the future needs of railway communications, and should be compared with proprietary deployments in terms of deployment costs, OAM, and performance.

References

1. LTE/SAE - The Future Railway Mobile Radio System? Long-Term Visions on Railway Mobile Radio Technologies. Technical Report, UIC (September 2009)
2. SYStéme de Transport Urbains du Futur (SYSTUF), Project description (2012)
3. Barbu, G.: E-TRAIN: Broadband communication with moving trains. Technical Report, Guidelines for application and development, UIC (2010)
4. Congressional Research Service, Positive Train Control (PTC): Overview and Policy Issues (July 2012)
5. IEEE, IEEE Standard for Communications-Based Train Control (CBTC) Performance and Functional Requirements, IEEE Std 1474.1-1999 (1999)
6. Dong, H., Ning, B., Cai, B., Hou, Z.: Automatic Train Control System Development and Simulation for High-Speed Railways. IEEE Circuits and Systems Magazine 10(2), 6–18 (2010)
7. Sesia, Toufik, Baker: LTE - The UMTS Long Term Evolution: From Theory to Practice, 2nd edn. Wiley (2011)
8. Federal Communications Commission, FCC takes action to advance nationwide broadband communications for America's first responders (January 2011)
9. Jalali, A., Padovani, R., Pankaj, R.: Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In: IEEE VTC, Spring, Tokyo (2000)
10. Liang, X., Wang, A., Zhou, S., Yao, Y.: A dynamic resource scheduling algorithm for OFDM system. In: The 9th Asia-Pacific Conference on Communications (2003)
11. 3GPP, Further advancements for E-UTRA physical layer aspects, 3GPP TSG-RAN-E-UTRA, TR 36.814 (March 2010)

Survey on Context-Aware Publish/Subscribe Systems for VANET

Mickaël Royer, Alain Pirovano, and Fabien Garcia

ENAC - French Civil Aviation University

Abstract. The publish / subscribe scheme is an efficient paradigm for communication widely used in wired networks. However the use of this paradigm in the context of mobile networks such as VANETs is still an open research topic. In particular, the communication needs of mobile nodes often depend on the node state, location and/or surroundings. Consequently, extensions to the P/S paradigm have been proposed that introduce the concept of *node context*. This paper propose a survey on these context-aware publish / subscribe solutions for VANETs. We will particularly focus our discussion on two example applications drawn from the context of civil aviation and will argue that Aeronautical Ad hoc Networks (AANETs) are a subgroup of VANETs.

Keywords: Vehicular Ad Hoc Network (VANET), Aeronautical Ad Hoc Network (AANET), Publish / Subscribe communication systems.

1 Introduction

Publish/Subscribe scheme is a communication paradigm which differs in many ways from classical communication schemes where addresses are used to identify the recipients. These addresses can target a single receiver (unicast communication) or multiple receivers (multicast communication). On the contrary, Publish/Subscribe systems allow *event* distribution from *publisher* (event producer) to *subscriber* (event consumer) without the use of any explicit address. Instead, the event distribution is based on declared subscribers interests. Those interests can be expressed in different ways depending on the subscription model used.

In the *topic based* subscription model, clients subscribe to topics characterized by a name. Events are then published on topics and the P/S system is responsible for the forwarding of events to clients who subscribed to this topic. The topic based subscription is quite simple but it's also limited since we can only filter events from a single attribute, i.e.: the topic name.

On the other side, in the *content based* subscription model, clients base their subscriptions on the content of events. So, when an event is published, the P/S system has to forward it to the clients whose subscriptions match the event content. The filter syntax "attribute – operator – pattern" can be used to define this subscription as described in [1].

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 46–58, 2013.

For example, persons holding stock of a company are generally interested in the evolution of its price. With to a topic based P/S system, we can easily think of a solution where these clients subscribe to a `'stock topic'` and receive notifications each time the price of the stock has changed. With the use of a content based P/S system, we can be much more efficient in filtering events. For example, a subscription can allow to receive notification only when price pass through a threshold.

P/S systems offer many advantages over traditional communication. In particular, they offer communications decoupled in space (subscribers do not need to know publishers and vice-versa) and potentially in time if the P/S system is able to store events for clients which are temporally disconnected (as in delay tolerant networks [2]).

In the past decade, many solutions (see for example [3] and [4]) have been proposed by the research community and industrials to build P/S system over wired networks. However, as emphasized in [5], those solutions are ill-adapted to the emerging category of networks called Mobile Ad hoc Networks (MANET). MANETs are defined as self-configuring networks of mobile nodes connected by wireless links without fixed infrastructure. In this category of networks, one sub-category will be dealt with in this paper, the Vehicular Ad hoc Networks (VANET) whose nodes are vehicles. In both MANETs and VANETs, disconnections between nodes are a common occurrence and the wired network centred solutions for P/S systems are not likely to take handle them satisfactorily. Accordingly, new solutions have been proposed to offer P/S communication systems optimized for VANET. The aim of this article is to provide a survey of these solutions.

This paper is organized as follow, in section 2 we present important characteristics of a VANET and argue that AANET (Aeronautical Ad hoc Networks, i.e.: VANETs composed of aircraft) are a subgroup of VANETs. In section 3, we illustrate possible applications of P/S systems over VANETs. We then use those applications to present the expected features of P/S systems for VANETs and the main challenges associated to their definition in section 4. Finally, we describe existing solutions of P/S systems adapted to VANET and evaluate their suitability to our purpose and constraints in section 5, before concluding this paper in section 6.

2 VANETs Characteristics and AANETs

A VANET can be defined as a self-organizing communication network composed of mobile vehicles using wireless communication links. VANETs are a subcategory of MANETs with some specificities. The most important is the high mobility of their nodes which leads to more frequent partitioning of the network [6]. Although, VANETs are associated most of the time to car, we argue in this section that other vehicle types can be used, like civil aircraft for example.

Indeed, lately, new communication means are studied to fulfil airlines and passenger communication requirements. In this context, AANETs (Aeronautical

Ad hoc NETworks, i.e.: MANETs with aircraft as nodes) represent an attractive solution. The feasibility of such networks has been demonstrated for instance in [7]. This category of network can be seen as a subcategory of VANETs as emphasised in [8]: "MANETs that span planes, trains, automobiles, and robots are called vehicle ad hoc networks (VANETs)". As illustrated in table 1, we list the main characteristics of VANETs based on descriptions found in related articles and show that AANETs share most of the characteristics of VANETs with cars. Some characteristics of AANET are obvious like *high mobility of nodes* or *high probability of network partition*. Some others need more explanation. In the same way as cars have movements constrained by road and traffic, civil aircraft have to follow trajectory imposed by air traffic controller. Furthermore, AANETs can also be subject to channel congestion problem. Indeed, some area, like close to airports, can have a hight density of aircraft, which may cause interference problem as explained in [9].

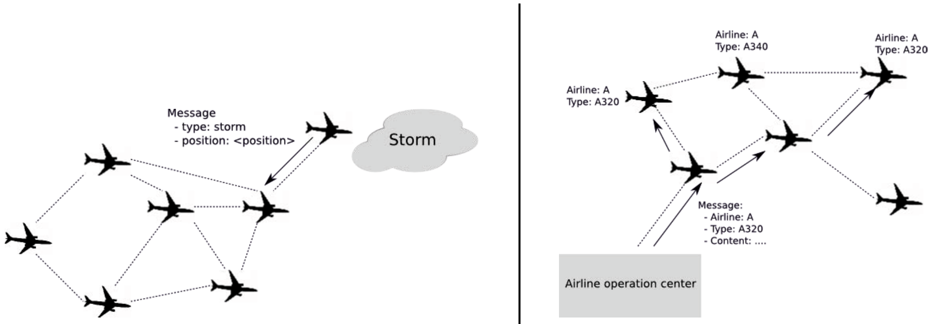
Table 1. VANET / AANET characteristics comparison

VANET Characteristics	Ref	Relevant for AANET
High mobility of nodes	[6], [10], [11], [12], [13]	Yes
High probability of network partition	[6], [10], [12]	Yes
Predictability of the movements of nodes	[6], [11], [12], [13]	Yes
Very large scale	[6]	Yes
Velocity restricted by speed limits and road traffic	[11]	No
Channel congestion risk due to a high vehicular traffic density	[14]	Yes
No energy restriction	[11], [12]	Yes
Type of application : geographical, hard delay constraint	[12]	(Yes)
Interaction with on-board sensors	[12]	Yes

As we will see in the next section, many applications of the P/S communication paradigm can be applied to AANETs, especially for the ATC (Air Traffic Control) or AOC (Airline Operational Communication) domains. This is the reason why this specific subgroup of VANETs has been targeted in the examples introduced in the next section. However, these examples can be easily transposed to VANETs with car (ex: cab company communication toward all of their cabs or announcement of a car accident)

3 Application Examples in AANET

To better understand the specific needs of AANET Publish / Subscribe applications, this section introduces two such example applications in the context of civil aviation. The first example is the dissemination of weather situation update to aircraft located at a specific geographic area. The second example is the dissemination of content to all aircraft of an airline company.



(a) An aircraft sends a message about a dangerous storm

(b) an airline sends a message to a part of its fleet

Fig. 1. Example of P/S applications in AANETs

3.1 Weather Situation Update

For safety reasons, weather informations are crucial during a flight. Some phenomena like storm or wind shear are very dangerous for aircraft and so must be transmitted from ground weather stations or from an aircraft discovering a new situation to all aircraft that might encounter it. Today, most of these communications, in particular in oceanic area, use satellite links which are more expensive than classical air-group communication infrastructure like VDL (VHF Data Link) or HFDL (HF Data Link) network, as illustrated in [15].

A new solution for those communications can be found with the use of a P/S system over an AANET composed of commercial aircraft and ground stations. In this case, publishers would be weather stations or aircraft which have knowledge about dangerous weather situations and subscribers would be aircraft which whose trajectory will get close to the concerned geographical area during the period of the weather phenomena. In order to cover this communication need, events must contain at least the following attributes: type of phenomenon, geographic area of weather phenomena and lifetime (since the weather informations are necessarily associated to a time period).

Similarly, in order to subscribe an aircraft need to know its geographical position (for example with a GPS receiver) and, if possible, its short term trajectory. Indeed, an aircraft will be only interested in weather information near its actual or short term position. A subscription will consequently follow the following principle : "I want to receive all weather informations regarding the geographical area `<position>` which will happen in the time period `<time>` (deduced from current position and trajectory)."

Consequently, as illustrated in figure 1a, an event will be disseminated to subscribers located near the geographical area of the event, or converging to that point in the near future.

3.2 Message to All Aircraft of an Airline

Another important class of communication in civil aviation consists in the communication between airlines operational center and their aircraft (this type of communication is called Airline Operational Communication or AOC). During a flight, many informations must be exchanged between these two actors (e.g.: maintenance information, crew communications, etc.). Nowadays, this kind of communication mainly uses ACARS (Aircraft Communication Addressing and Reporting System). ACARS is a network deployed in the 80s by airlines to allow data communication with their aircraft. Depending on the aircraft position, it can use VDL (VHF Data Link) network or SATCOM means. Although robust, this network has very limited resources (bandwidth in the amount of tens of Kbits/s, [16]).

As in the previous example, a new solution can be proposed with the use of a P/S system over an AANET (which offers better performance than ACARS, as explained in [7]) composed of aircraft and airline operational center. In this example, as illustrated in figure 1b, publishers will be airline operational center and subscribers will be aircraft. To be properly filtered, events and subscriptions must contain at least the airline name as attribute. Furthermore, the advantage of a P/S system over a classical multicast communication would be in the filtering capabilities it offers. Indeed, more focused communications can easily be achieved with events being transmitted only to a specific type of aircraft or to aircraft bound for a given airport.

4 Expected Features of P/S Systems for VANET

The examples of the previous section show that traditional P/S system can not provide an optimal answer to the VANET content-based communication research question. In the first example, with a classical P/S solution, planes can subscribe to `'dangerous weather situation'`. However, this subscription will be too large and will lead to the reception of useless events since planes are interested specifically by `'dangerous weather situation` *`close to their position or route`*`'`. To accommodate this kind of subscription, a new concept is needed: the *node context*. Furthermore, one of the VANETs features is the high probability of network partitioning leading to frequent disconnections. A consequence of these disconnections is that some events (potentially critical) will not be received. To avoid this situation, a P/S system designed for VANETs needs to allow *the persistence of events in the network*. These two features are detailed in this section.

4.1 Node Context

A node context is a list of attributes which complements subscriptions, and is used to more precisely filter received events. This better filtering allows a reduction in the use of network resources which are scarce in VANETs. Several definitions exist for node context. Simple definitions reduce context to the node location ("location context", see [17], [18] or [19]). Some other definitions like [20] are more generic and define the node context has a combination of static attributes and dynamic attributes, either defined in an absolute way or defined in relation with other nodes. For example, the velocity of a node can be considered as a dynamic context attribute. An absolute definition of this attribute will be `'velocity = 50 mph'` while a relative definition will be `'velocity superior to the velocity of the node in front of me'`. In relation to the examples discussed above, a node context would be defined as the node location, changing over time. We think that static attributes are not mandatory as they can easily be included in the subscriptions. For the second example introduced above, the company name of an aircraft can be seen as a static context attribute. In a content-based P/S system, this context attribute could easily be replaced by an `'airline name'` attribute correctly filled in subscriptions and published events.

Node context should not be mandatory to dispatch events. Indeed, in the example in section 3.2 the node location is not relevant for the communication (i.e.: an airline want to send a message to all aircraft, regardless of their position).

4.2 Persistent Events

As explain in [21], the persistence of events is the ability for a P/S system to store events during its lifetime. Thus, subscribers interested by an event but not connected when this event has been sent are able to retrieve it later.

Traditional P/S systems for fixed infrastructure network do not necessarily implement this feature. However, this feature is essential for VANETs where disconnections of node are frequent. To save network resources, the persistence can be handled with the introduction of a lifetime associated to each event. Indeed, most of the time, an event is of interest only for a given time period and can be deleted afterwards. Some solutions, like[20], also use a lifetime associated to each subscription. We do not think this is mandatory since P/S API generally include an UNSUBSCRIBE primitive to remove a subscription from the system. To add persistence to a P/S system, two family of solutions exist. Events can be retransmitted periodically by publishers, this is costly in network bandwidth though, especially for a VANET. Another solution would be to store events in the network during its lifetime. Events can be hold by publishers (the simplest way) or by any node in the network (such as a broker for a P/S system based on overlay network, see next section).

5 Existing Solutions

Recently, technical solutions have been proposed to offer Publish/Subscribe communication systems adapted to VANETs. They can be classified in three categories: geographic routing based solutions, proximity routing based solutions and overlay network based solutions. In this section, the technical principles are detailed for each of these groups of solutions. Advantages and disadvantages are then identified, particularly in the light of the expected features of a P/S systems detailed in the previous section (node context and persistence of events). The two AANET based applications introduced in section 3 will be used as illustrations. Figure 2 show the situation considered, which includes two communications from a same publisher. The first communication correspond to a weather situation update sent toward a geographical area where 4 planes are present. The second communication is a message sent to all plane of an airline (grey planes in the figure).

Fig. 2. Communication example in an AANET

5.1 Geographic Routing Based Solutions

The solutions presented in this section use geographic routing to dispatch events from the publisher to the subscribers. In order for those to work correctly, two conditions must be met: network nodes must be equipped with a positioning system (e.g.: a GPS receiver) and all events must be associated with a geographical area. Many examples of applications respecting these assumptions can be introduced, starting with the example from section 3.1.

PCBD. Persistent Content-Based Dissemination [22] is a solution that was defined for VANETs composed of both cars and info-stations (i.e.: fixed infrastructure nodes). This solution uses the following mechanisms: first, the publisher sends an event to a geographical area using a geographical routing protocol, this event is received by all nodes (cars and info-stations alike) in the area; then this event is stored in an info-station if there is one available, or else in vehicles. To choose the most appropriate vehicles, navigation systems can be used to determine the vehicles that will remain for the longest time in the area. Each vehicle periodically advertises its planned route and its interests (classical subscription mechanism, with in addition a vehicle context, the planned route). Info-stations or vehicles receiving subscriptions try to match them with their stored events. Events with positive matching are then sent to the corresponding subscriber.

Context-Aware PS for MANET. As for PCBD, this solution, introduced in [20], uses a geographical routing protocol in order to disseminate an event in the geographical area associated. There are, however, two significant differences. First, the notion of context is explicitly defined and the classical publish / subscribe API has been extended to use it efficiently. Secondly, nodes interested by events for a specific zone do not have to wait to be in the target area to send their subscription request since they also use a geographical routing protocol.

Application Example. As illustrated in figure 3, geographic routing based solutions fit well with applications where subscribers are grouped in an area, but they are not adapted to case where subscribers are dispersed in the network. In this last case it would equivalent to broadcast the message since the target geographical area covers the entire network.

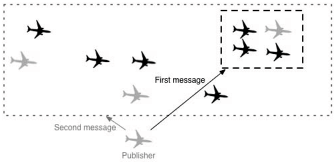


Fig. 3. Application example with Geographic routing based solution

5.2 Proximity Routing Based Solutions

Solutions detailed in this section use the idea that most nodes interested by an event are close to the publisher. From this assumption, events are broadcast within a restricted geographical range around the publisher.

STEAM. Scalable Timed Events And Mobility, presented in [18], is a content based publish/subscribe system designed for Wireless Ad-Hoc Network. In order to achieve scalability, this solution introduce the concept of proximity filters in addition to topic and content filters. This new filter introduces a range of dissemination around the publisher for the events. Each event is then sent only to nodes within this range of the publisher. This is motivated by the applications envisioned for this system (e.g.: a traffic light application where traffic lights produce events to indicate light changes to vehicle near them) and by the need to achieve good scalability on Ad Hoc networks with limited resources.

PERHAVO/LPSS. Location-based Publish-Subscribe Service, defined in [23], uses the notion of proximity to disseminate events and subscription to a geographical range around the publisher and the subscriber respectively. To match a publication to a subscription one has to meet two conditions : the content match (as for any other content based P/S system) and the location match (i.e.: the subscriber and the publisher must both be located in the intersection of the publication and subscription spaces).

ALPS. Adaptive Location-based Publish/Subscribe [19] is a location-based Publish/Subscribe solution. As LPSS, this solution includes a notion of location context complementing content-based queries. This context is then used to complement the content match with a location match. There are two main differences with LPSS. First ALPS introduces three content match strategies: "(1) message-centric algorithms (MCAs), where publishers broadcast messages in the message range and subscribers are in charge of performing matches; (2) query-centric algorithms (QCAs), where subscribers broadcast queries in the query space and publishers perform matches and subsequently route messages; and (3) hybrid ones (Hybrid) where both messages and queries are broadcast within a restricted area. In this last strategy, intermediate nodes are in charge of performing matches and routing messages to subscribers." In this terminology, LPSS uses only message-centric algorithm. The second difference with LPSS is that ALPS offers persistence of messages in the network thanks to the introduction of a lifetime specified by the application for each event. To ensure this persistence, events are retransmitted periodically by publishers.

Application Example. All the solutions detailed in this section introduce and use the location context of node to dispatch an event. Some of them (like ALPS) also implements persistent messages, which is a critical feature for a P/S system adapted to VANET as discussed in section 4.2. However, as illustrated in figure 4, proximity routing based solutions, although they have the advantage of not using many network resources, are only suitable for cases where publishers and subscribers are close together. Consequently, this kind of solutions cannot be used for the applications introduced in section 3.

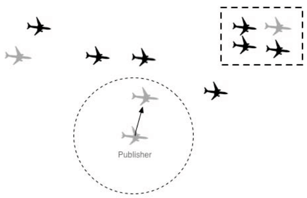


Fig. 4. Application example with Proximity routing based solution

5.3 Solutions Based on Overlay Network

The solutions that will be discussed in this section introduce a new component in the P/S architecture: the *broker*. Brokers are the entities responsible for dispatching events from publishers to subscribers and are interconnected thus forming an overlay network.

Most P/S middleware for wired-networks are overlay network based solutions ([3], [4]). Accordingly, some P/S systems try to adapt this concept in the context of MANETs. They are detailed in this section.

REDS / SPCF. REconfigurable Dispatching System [24] is a content based P/S framework adapted to mobile network topologies like those of VANETs. REDS is a modular solution which offers several protocol implementations to dispatch event and manage dynamic broker topology (as in [25]). In particular,

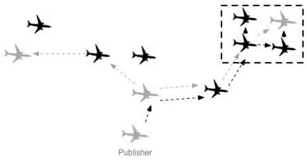


Fig. 5. Application example with solutions based on overlay network

REDS offers an implementation for a context-aware P/S system with the SPCF (Shortest Path Context Forwarding) protocol [26]. This solution introduces a new API to manage node context in subscriptions and publications. This context information is used in the routing process in the overlay network, with the introduction of a context table for each broker. SPCF uses this table with the more classical content table to make forwarding decisions in the broker network.

REBECA. Rebeca Event-Based Electronic Commerce Architecture [27] is another content Publish/Subscribe system based on a modular framework. Unlike REDS, REBECA has been primarily developed for fixed networks. An extension of REBECA as then been proposed in [28] to support mobility. This extension introduces the concept of logical and physical mobility for clients. Physical mobility results in the disconnection of a client from his broker and the connection to another broker. This phase can produce a loss of events and so has to be handled correctly by the P/S System. Logical mobility happens when a client moves without broker disconnection. In this case, the client may need to update its subscriptions (called "automated location awareness within a defined environment" in the article). Logical mobility can be seen as a restrictive context used to update a subscription. The mobility extension of REBECA proposes solutions to deal with these two mobility issues by using two features of REBECA: the publisher advertisement [21] and the routing algorithm based on "Learning by the Reverse Path" [29].

Application Example. As illustrated in figure 5 (where all *visible* nodes are considered to be brokers), solutions based on overlay network are suitable for the two application examples introduced in the section 3. However, the main disadvantage of these solutions is that significant network resources consumption goes toward maintaining the overlay network.

5.4 Synthesis

Table 2 provides a synthesis of P/S solutions according to several criteria: adaptability according VANET characteristics or distribution of recipients and available features for applications.

Table 2. Solution synthesis

Solution families		Geographic routing based solutions		Proximity routing based solutions			Overlay network based solutions	
Solutions		PCBD	Context-Aware PS for MANET	STEAM	LPSS	ALPS	REDS/SPCF	REBECCA
Network Characteristics	With infrastructure		X					X
	Without infrastructure	X	X	X	X	X	X	
Features	location context aware	X	X	X	X	X	X	X
	persistent events	X	X			X		X
Distribution of recipients	Located in a limited area	X	X	X	X	X	X	X
	Dispatched in all the network						X	X

6 Conclusion

In this article, we have performed a survey on current solutions for publish / subscribe systems in VANETs. We first presented the main characteristics of VANETs, and argued that Aeronautical Ad Hoc Networks were a subgroup of VANETs. We then presented two examples of application of a publish/subscribe system in the particular context of AANETs in civil aviation, stating that similar applications could be found in the more general context of the VANETs. Using these two applications, we have extracted two desirable features of P/S systems for VANETs, namely the node context filtering and the persistence of events in the network. Finally, we presented three classes of solutions for P/S systems in VANETs, geographic routing based solutions, proximity based solutions and overlay network based solutions. As we emphasized in each case, none of these solutions can tackle all the communication needs of the two example applications. Geographic routing and proximity routing based solutions are ill-adapted to the case where subscribers are disseminated throughout the network. To this regard overlay network based solutions seem the best suited, but the amount of overhead introduced by the maintenance of the overlay network might be a problem in resources limited VANETs.

A comprehensive evaluation of the existing solutions through simulations and the development of a specifically designed P/S system able to address the underlined limits shall then be our next step toward the definition of a scalable, flexible P/S system for VANETs.

References

1. Aguilera, M.K., Strom, R.E., Sturman, D.C., Astley, M., Chandra, T.D.: Matching events in a content-based subscription system. In: Proceedings of the Eighteenth Annual ACM Symposium on Principles of Distributed Computing, PODC 1999, pp. 53–61. ACM, New York (1999)
2. Zhang, Z.: Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges. IEEE Communications Surveys Tutorials 8(1), 24–37 (2006)

3. Rowstron, A., Kermarrec, A.-M., Castro, M., Druschel, P.: Scribe: The design of a large-scale event notification infrastructure. In: Crowcroft, J., Hofmann, M. (eds.) NGC 2001. LNCS, vol. 2233, pp. 30–43. Springer, Heidelberg (2001)
4. Pietzuch, P.R., Bacon, J.M.: Hermes: a distributed event-based middleware architecture. In: Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, pp. 611–618 (2002)
5. Farooq, U., Parsons, E.W., Majumdar, S.: Performance of publish/subscribe middleware in mobile wireless networks. SIGSOFT Softw. Eng. Notes 29, 278–289 (2004)
6. Toor, Y., Muhlethaler, P., Laouiti, A.: Vehicle ad hoc networks: applications and related technical issues. IEEE Communications Surveys Tutorials 10(3), 74–88 (2008)
7. Besse, F., Garcia, F., Pirovano, A., Radzic, J.: Wireless ad hoc networks access for aeronautical communications. In: AIAA 28th International Communications Satellite Systems Conference (2010)
8. Willke, T.L., Tientrakool, P., Maxemchuk, N.F.: A survey of inter-vehicle communication protocols and their applications. IEEE Communications Surveys Tutorials 11(2), 3–20 (2009)
9. Besse, F., Pirovano, A., Garcia, F., Radzik, J.: Interference estimation in an aeronautical ad hoc network. In: 2011 IEEE/AIAA 30th Digital Avionics Systems Conference (DASC), pp. 4C6-1–4C6-11 (October 2011)
10. Blum, J., Eskandarian, A., Hoffman, L.J.: Performance characteristics of inter-vehicle ad hoc networks. In: Proceedings of the 2003 IEEE Intelligent Transportation Systems, vol. 1, pp. 114–119 (October 2003)
11. Xi, S., Li, X.-M.: Study of the feasibility of vanet and its routing protocols. In: 4th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2008, pp. 1–4 (October 2008)
12. Li, F., Wang, Y.: Routing in vehicular ad hoc networks: A survey. IEEE Vehicular Technology Magazine 2(2), 12–22 (2007)
13. Harri, J., Filali, F., Bonnet, C.: Mobility models for vehicular ad hoc networks: a survey and taxonomy. IEEE Communications Surveys Tutorials 11(4), 19–41 (2009)
14. Hartenstein, H., Laberteaux, K.P.: A tutorial survey on vehicular ad hoc networks. IEEE Communications Magazine 46(6), 164–171 (2008)
15. Gallagher, S., Blythe, D., Decker, S.: Communication system architecture for air traffic management and weather information dissemination. In: IEEE Proceedings of the Aerospace Conference, vol. 3, pp. 3/1327 –3/1337 (2001)
16. Kitaori, J.: A performance comparison between vdl mode 2 and vhf acars by protocol simulator. In: IEEE/AIAA 28th Digital Avionics Systems Conference, DASC 2009, pp. 4.B.3-1–4.B.3-8 (October 2009)
17. Eugster, P., Garbinato, B., Holzer, A.: Pervaho: A specialized middleware for mobile context-aware applications. Electronic Commerce Research 9(4), 245–268
18. Meier, R., Cahill, V.: Steam: event-based middleware for wireless ad hoc networks. In: Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, pp. 639–644 (2002)
19. Holzer, A., Eugster, P., Garbinato, B.: Alps - adaptive location-based publish/subscribe. Comput. Netw. 56(12), 2949–2962 (2012)
20. Frey, D., Roman, G.-C.: Context-aware publish subscribe in mobile ad hoc networks. In: Murphy, A.L., Vitek, J. (eds.) COORDINATION 2007. LNCS, vol. 4467, pp. 37–55. Springer, Heidelberg (2007)
21. Eugster, P.T., Felber, P.A., Guerraoui, R., Kermarrec, A.-M.: The many faces of publish/subscribe. ACM Comput. Surv. 35, 114–131 (2003)

22. Leontiadis, I., Costa, P., Mascolo, C.: Persistent content-based information dissemination in hybrid vehicular networks. In: IEEE International Conference on Pervasive Computing and Communications, PerCom 2009, pp. 1–10 (March 2009)
23. Eugster, P., Garbinato, B., Holzer, A.: Design and implementation of the pervaho middleware for mobile context-aware applications. In: 2008 International MCETECH Conference on e-Technologies, pp. 125–135 (January 2008)
24. Cugola, G., Picco, G.P.: Reds: a reconfigurable dispatching system. In: Proceedings of the 6th International Workshop on Software Engineering and Middleware, SEM 2006, pp. 9–16. ACM, New York (2006)
25. Mottola, L., Cugola, G., Picco, G.P.: A self-repairing tree topology enabling content-based routing in mobile ad hoc networks. IEEE Transactions on Mobile Computing 7(8), 946–960 (2008)
26. Cugola, G., Margara, A., Migliavacca, M.: Context-aware publish-subscribe: Model, implementation, and evaluation. In: IEEE Symposium on Computers and Communications, ISCC 2009, pp. 875–881 (July 2009)
27. Parzyjegla, H., Graff, D., Schröter, A., Richling, J., Mühl, G.: Design and implementation of the rebeca publish/subscribe middleware. In: Sachs, K., Petrov, I., Guerrero, P. (eds.) Buchmann Festschrift. LNCS, vol. 6462, pp. 124–140. Springer, Heidelberg (2010)
28. Fiege, L., Gartner, F., Kasten, O., Zeidler, A.: Supporting mobility in content-based publish/subscribe middleware. In: Endler, M., Schmidt, D. (eds.) Middleware 2003. LNCS, vol. 2672, pp. 103–122. Springer, Heidelberg (2003)
29. Martins, J.L., Duarte, S.: Routing algorithms for content-based publish/subscribe systems. IEEE Communications Surveys Tutorials 12(1), 39–58 (2010)

A Survey on Security in Vehicular Ad Hoc Networks

Saira Gillani[1], Farrukh Shahzad[2], Amir Qayyum[1], and Rashid Mehmood[3]

[1] CoReNeT, M. A. Jinnah University, Islamabad, Pakistan
[2] nexGIN RC, FAST-National University, Islamabad, Pakistan
[3] University of Huddersfield, UK
sairagilani@yahoo.com, farrukh.shahzad@nexginrc.org,
aqayyum@ieee.org, r.mehmood@hud.ac.uk

Abstract. Vehicular Ad-hoc Networks (VANETs) are the most prominent *enabling network technology* for Intelligent Transportation Systems. VANETs provide many new exciting applications and opportunities albeit transportation safety and facilitation applications are their core drivers. Security of vehicular networks remains the most significant concern in VANETs deployment – because it is mandatory to assure public and transportation safety. In this paper, we review the various dimensions of VANETs security including security threats, challenges in providing security in vehicular networks environment, requirements and attributes of security solutions. We also provide taxonomy and critically review of the notable security solutions – available for VANETs in literature.

1 Introduction

The inherent human desire for change, progress, mobility, entertainment, safety and security are leading the way to the development of intelligent transportation systems (ITS) [37]. Vehicular Ad-hoc networks are the most prominent enabling technology for ITS. They are formed on the fly by vehicles equipped with wireless communication capability. The participating nodes in such networks (i.e. vehicles) interact and cooperate with each other by short-range direct communications, by hoping messages through vehicles (Vehicle-to-Vehicle) and road side masts (Vehicle-to-Infrastructure). Traditionally, information about traffic on a road is only gained through inductive loops, cameras, roadside sensors and surveys. VANETs provide new venues for collecting realtime information – from onboard sensors on vehicles – and its quick dissemination. The information collected through individual vehicles participating in the network can be integrated together to form a realtime picture of the road situation. Many new applications have been enabled through VANETs, though safety and transportation efficiency applications are the most important driver for them. The various ITS stakeholders such as governments, tele-communication companies and car manufacturers are working together to make "VANET based ITS" a reality. Hundreds of projects are underway in the US, Europe, Japan, China,

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 59–74, 2013.

Singapore and other countries of the world, supporting the cause with research, innovation, testing and standardization activities [1][2].

It is expected – up till 2015 – that software and electronics will cover 50% of the total cost of an automobile [9]. Today, automobiles are fully equipped with IT and software technologies. However, these systems are not mainly concentrating on security aspects. On the other hand, Vehicular ad hoc networks are mainly designed for safety applications (to ensure the safety of drivers and vehicles[1] and also to avoid or minimize the road accidents). VANET is a special type of Mobile ad hoc networks (MANET) [5] customized for automobiles with some distinctive features e.g. predictable mobility patterns, movement of nodes along predefined paths instead of random directions. There are no battery constraints in VANET therefore it is suitable for long range communications through vehicles. In addition to safety applications, various other applications like collision avoidance, traffic management, trip planning and infotainment are also developed for them [6]. Safety applications must be protected from hackers and crafty attacker because a compromised safety application could result in the loss of human life [8]. Commercial applications need security to protect the potential loss of revenue. Without security, a Vehicular Ad hoc network can be affected by many attacks like denial of service, message suppression and propagation of false message attacks etc. that may cause accidents.

Our focal point in this survey paper is to flash out and emphasize the major security issues, threats, core requirements & challenges to design a fail-safe security framework. We also outline a taxonomy – in which categories are devise on the basis of mandatory sets of security requirements – for already proposed protection solutions. Moreover, we analyze and critically review the trustworthiness, privacy and confidentiality aspects of different security solutions – available in literature – as a guideline for security researchers & products' vendors.

The rest of the paper is organized as follows. Section 2 summarizes the security threats in VANET landscape. In section 3, we enlist the background challenges for designing the security frameworks for vehicular networks. We provide mandatory requirements for VANET security in section 4. In section 5, we analyze previously proposed categories and define our own taxonomy for security solutions. Section 6 provides the survey and critical review of existing solutions. Finally, we conclude the survey in section 7 with a few possible future directions.

2 Major Security Threats in VANETs

Due to open wireless nature, there exist number of security threats & attacks which are quite non-trivial for VANETs. In this section, an overview of attacks is presented that may ensue in the background of VANETs. Obviously, all possible attacks can't be covered in this paper. Therefore, we are enumerating some generic and significant attacks that are commonly available in literature. For the sake of brevity, we portray the possible nature of attacks and their likely scenarios.

[1] In this survey , we use terms vehicles, nodes and cars interchangeably.

2.1 Denial of Service (DoS)

In DoS attacks, the main objective of the crafty attacker is to disturb the communication channel or overwhelms the vehicle's available services from the legitimate users. This attack makes the system useless, and uselessness of the system in realtime vehicular networks – even for a tiny instant of time – is not affordable for users. A few exemplary scenario of DoS are enumerated below.

1. *Flooding:* The network can be flooded by sending high volume of traffic. In this way, attacker takes up a vehicle's computing resources and seize legitimate network traffic – by overloading the communication channel. In this way, the critical information can't be conveyed to other vehicles on time. Moreover, it can cause or increase danger to the driver if she/he is depending on the application's information for making decisions. For example, on a highway, an illegitimate user or attacker can create massive pile up by just prevent the warning messages – by launching a DoS attack [11][12][13][14] – which are being generated by other legitimate drivers.
2. *Jamming Attack:* Jammers deliberately generate interfering transmissions or signals to prevent communication across the network. Since, the network coverage areas are well-defined in VANETs; therefore, an adversary can partition the network. For this purpose, he neither require much transmission power nor compromising the cryptographic mechanisms. It's because jamming is known as low-effort exploit.
3. *Broadcast Tampering/Spamming:* The hackers inject false (spam) messages into the network in broadcast tempering attack. Due to these false messages, serious problems can be initiated in traffic flows. For example, suppression of safety messages may cause an accident or manipulating the false information may cause disturbance in traffic management.
4. *Malware:* Different categories of malicious software – i.e. virus, worm and trojans etc. (populated in network) – may cause threats for service availability in VANET. Moreover, these malicious programs can cause traffic related threats ranging from congestion to large scale accidents. Inside attacker usually introduce malware in network but some worm-like malware can spread themselves in VANET without human intervention (usually injected into the network when vehicles receive updates) [15].
5. *Black Hole Attack:* Another availability problem is black hole attack. In this type of attack, an illegitimate user advertises its routing advertisements by using its own routing protocol. In these advertisements, he claims to have shortest path to the destination node. As a result, the benign traffic of network nodes fails to reach the desired destinations.

2.2 Authentication Attacks

Authentication is a mandatory requirement to provide effective security. In VANETs, safety applications especially require efficient authentication mechanism because unauthenticated message may cause threats to human lives. In absence of adequate authentication mechanisms, attackers can launch different types of attacks. Some potential attacks are enumerated here.

1. *Masquerading:* In this type of attack, attacker pretends to be a legitimate vehicle by using stolen identity. In this act, the objective of attacker may be malicious or rational [16]. By pretending another vehicle, attacker can launch different attacks like message fabrication, alteration and replay. For example, the attacker pretends to be emergency vehicle to mislead other vehicles to slow down or leave place for it [17].
2. *Impersonation Attack:* In such attacks, hacker imitates the identity of a legitimate user – by using MAC and IP spoofing. He hide his original identity – uses identity of another node – to perform any illegal activity in the network. In case of using authentication certificates by the users, impersonation attacks are almost impossible because certificates couldn't be forged.
3. *Sybil attack:* In such attack, attacker uses multiple identities to send multiple messages; different identities are used at the same time. In this way, the attacker creates an illusion that messages are being sent from different nodes. The basic objective of this attack is to mislead other vehicles. Moreover, this attack can also disrupt routing protocols, leading to distributed DoS and unfair distribution of resources. [17].
4. *Replay attack:* In this type of attack, hacker can resend or replay a (captured) valid message of a benign user. The attacker replays the transmission of earlier information at a later time. The purpose of attacker would be to confuse authorities and to take the advantage of the situation – in order to propagate false information into the network. The security mechanism of basic 802.11 protocol doesn't provide protection against this attack [18].
5. *GPS Spoofing:* Malicious user tries to deceive the legitimate users – by hiding its actual location. This is possible by giving false GPS information to users. This attack can be launched by using a GPS satellite simulator that generates stronger signal than that of genuine satellite [19].

2.3 Attacks on Privacy

The hackers illegally obtain the sensitive information about vehicle or driver. Some sample privacy attacks are listed here.

1. *Identity revealing:* In the case of identity revealing attack, the attacker gets vehicle's identity and put its privacy at risk. In most of the cases, driver of the vehicle is its owner; so in this way, the attacker can obtain personal data of vehicle's owner.
2. *Location tracking:* Attacker can locate and track a vehicle through its transmitted messages – during communication to any other vehicle or roadside unit. By tracking a vehicle, it becomes possible to build vehicles profile; in this way, the privacy of the driver is breached.

2.4 Attacks on Confidentiality

Basically, the eavesdropping (getting illegal access to user's data) is an attack on user's confidentiality. The attackers record & use the information about vehicles – without the permissions of their owners. Later on, this information is used by the market vendors & companies for data mining and pattern findings purposes.

3 Challenges for VANET Security

VANET environment is quite different from other networks due to its high speed mobility nodes and distributed nature. Therefore security threats and security requirements in VANETs are also different from other networks. This section will discuss the design challenges for security solutions in VANETs.

3.1 Mobility

Mobility challenge is difficult to handle in VANETs in general and in security frameworks in particular. In VANET, vehicles move with high velocity on predefined paths; so these moving vehicles make connections for very short duration due to high speed. Therefore, quality of communication can be affected by the high velocity vehicles and due to high mobility; handshake based mechanisms cannot be used in VANET.

3.2 Network Scalability

VANET (worldwide) is a large scale network which is covering more than 75 million vehicles all over the world [9]. The management of control of such a huge network and its security aspects – exchanging certificate etc. – is a big problem; despite the fact that there doesn't exist a global authority who governs the standard of DSRC[2]. Security protocols that required pre-stored information about participating nodes are not suitable.

3.3 Heterogeneity

Due to the availability and implementation of different network infrastructures in different parts (countries) of the world, future vehicular networks can be envisioned as a heterogeneous networks. Therefore, different manufacturers will implement different technologies according to their perspective country's privacy and security policies [39].

3.4 Secure Positioning

GPS equipment may exhibit several drawbacks e.g. precision issues when used in security solutions. Although, recently introduced devices have reduced precision problems [13] but many attacks are related to GPS such as signal jamming and spoofing etc. [29].

3.5 Privacy

In VANET, there is a close relationship between user and vehicle. Drivers want their privacy and are concerned about the disclosure of their location and behavior; as the movement pattern of a person can be determined by tracking his vehicle. Furthermore, financial transactions carried out on VANET also include the privacy concerns [38].

[2] Dedicated Short Range Communications, a bandwidth of 75MHz for VANET applications.

3.6 Volatility

As mentioned earlier, due to high velocity vehicles in VANET, connections are established for short period of time; while the connectivity of user-devices to a hot spot require long-time communication for authentication purposes; therefore, it would be difficult to secure the communication of vehicular networks.

3.7 Usability

It is also a considerable factor when deploying security solution on VANET [38]. A vehicle's driver may not be willing to deal with any issue related to electronic system; it may be an annoying task for a driver to operate an electronic system. Hence, security application should be automatically configurable.

4 Requirements of VANET Security Solutions

For widespread deployment of secure VANETs, security solution designers should meet (at least) some basic and significant requirements. In this section, we primarily focus on security issues and requirements for safety related applications [39].

4.1 Authentication

To incorporate the necessary trust in the network, authentication is a mandatory requirement. Authentication of nodes or vehicles and authentication of messaging among them – are equally important. By authenticating the nodes, it can be made sure that messages are being generated from a valid source and not from a virtual or malicious node. Otherwise, the launch of *Sybil attack* is possible – in which a single node pretends as arbitrary number of nodes. If safety messages are compromised then *replay attack* is possible and it can be avoided only by inserting an authenticated timestamp in the messages. For authentication purpose, multiple solutions have been proposed but overhead of such schemes is larger than the message content (of DSRC protocol) of VANET. Researchers have proposed cryptography based solutions to enable authentication. However, due to large scale network of VANET, even symmetric cryptographic techniques (having lower overhead than asymmetric cryptography [4]) cannot be used as a generalized solution for all types of communication.

4.2 Privacy

It is to ensure that the information is not being leaked to the unauthorized people who are not allowed to view the information [6]. For authorization, authentication is required. Whereas for authentication purposes, if some authority publicly discloses permanent identities of vehicles then it may be dangerous for vehicles' drivers – in case their travel activities are monitored. So, in such type of networks, anonymity will always be required [3]. But in liability related cases, maintaining the anonymity is not possible because some specified authorities like traffic police etc. needed to be allowed, the tracing of user identities for legal investigation.

4.3 Non-repudiation

Any illegal activity of illegitimate user can equally harm people's life. So it is probable that such a user deny or avoid to accept the ownership of sent message or the contents of a message. Non-repudiation is useful to detect such compromised nodes.

4.4 Availability

VANET is specially designed for safety related applications. For such applications, network availability must be made sure at all times (24/7) because life critical information is being communicated between sender and receiver. Adversaries will always try to reduce availability while delay of seconds can render the message meaningless.

4.5 Location Accuracy

An adversary may report false information about its location and misguide others; so it is critical to determine whether the sending vehicle is at a given location or is on logical place. Therefore, to get accurate location of sending node is very important in VANETs.

4.6 Realtime Guarantees

Since, VANET is specially designed for safety applications – i.e. collision avoidance – that are dependent on strict time guarantees. Adversaries may try to delay the messages in VANET; delay of seconds means that the message become meaningless and its result may be devastating [7][8].

5 Taxonomy of Security Solution for VANETs

Moharrum et al. [28] distinguished two main categories of cryptography based security algorithms. In first category, the authors included public key infrastructure based schemes (basic PKI based approaches, pseudonym-based approaches and group signature based approaches) while in second category they listed non-fully PKI based schemes (identity based cryptography and hybrid approaches etc.). In [30], Haseeb et al. defined the different methods of authentication protocols like node level authentication, group level authentication, unicast authentication, multicast authentication and broadcast authentication. But, they didn't mention other security aspects like privacy. In [31], Xiong et al. described the main approaches of anonymous authentication protocols for Vehicular Ad Hoc Networks. Moreover, they divided authentication protocols into following approaches i.e. RSU-based approach, Group-oriented signature-based approach, Pseudonyms-based approach and Priori-based approach. They only discussed, just one aspect i.e. authentication of VANET security. In [32], authors did not defined any taxonomy rather they just discussed VANET security in broader sense. They discussed data centric trust approaches; secure localization approaches, integration, hybrid vehicular communication and privacy. We also define a taxonomy for security solutions – in broader sense (four categories) – and try to cover all aspects of security in it. According to the our literature survey & review, there exist four main aspects of VANET security – which are major heads of our taxonomy.

1. Reputation: It means how to trust the message sender. For example, if a vehicle sends a message that *there is road block ahead.* How other vehicles can trust this message?. In literature, some models have been proposed that ensure this required trust level; they are called reputation systems. Some reputation models are: (a) entity centric trust model, (b) data centric trust model, and (c) combined trust model etc.
2. Authentication: Vehicles should response only to legitimate messages. The receiver should be able to authenticate the legitimacy of message sending entity. Different approaches have been defined for authentication in vehicular ad hoc networks. Some prominent approaches are: RSU-based approach, group-oriented approach, pseudonyms-based approach and priori-based approach.
3. Privacy: Privacy of the drivers – against unauthorized observers – should be protected in VANET. Different cryptographic schemes have been proposed in literature for privacy protection purposes. Some well-known of them are: public key-based approach, group signature-based approach, identity-based approach and symmetric-key cryptography based approach etc.
4. Confidentiality: It is to make sure that the message would be read only by authorized users. Majority of the researchers argue that the confidentiality is not required in VANET because safety messages don't contain sensitive information [10]. But numerous solutions are available in literature like [33] which handle confidentiality by using anonymous key pairs. Such solutions are under great criticism by research community because they take large storage space and undergo high level of complexity.

6 Security Solutions for VANETs - Review

After defining the taxonomy, now we concentrate towards survey and critical review of different security solutions for VANET.

6.1 Event-Based Reputation System

Nai-Wei et al. [20] proposed a dynamic event-based reputation system (ERS) to provide secure communication in VANET. For this purpose, they proposed four functions to compute confidence and trust threshold. Event reputation value indicates the intensity level of a specific traffic event. The initial value of this event reputation is always zero. When a vehicle detects an event with its – on board – sensors, the reputation value is increased by one. The number of distinct vehicles who generate messages regarding the same event defines the event confidence value. Every time when a vehicle detects an event, ERS uses a simple algorithm that adds the reputation value in the received message – from already stored reputation value. The event confidence list is appended in the message. Now, if the event reputation value and confidence value are matched with the already defined threshold of event reputation and confidence, then the traffic event will be considered a real event and the vehicle will send the traffic warning message to its neighboring vehicles. In this way, false messages are prohibited by ERS. However, this system has some issues: (1) In VANET, we assume that vehicle's id will be changed over time but authors didn't mention, how to guarantee' that vehicle's id will not change during an event, and (2) In VANET, when the speed of vehicles will get high then there will be a short time to detect a specific event. If it happens again and again, it will degrade the accuracy of the system.

6.2 Vehicle Ad Hoc Reputation System (VARS)

Dotzer et al. [21] presented the first Vehicle Ad Hoc Reputation System (VARS). It is a distributed entity centric reputation system that uses opinion piggybacking method, where every forwarding peer appends its own opinion to enable confident decisions on event message. An algorithm is defined by the authors to allow a peer to append an opinion about the message. This opinion is based on aggregated opinions appended to the message. According to their opinion generation algorithm every forwarding node generates an opinion on the worthiness of coming message. This generated opinion is appended to the message before forwarding it. The peer opinion is also dependent on some other metrics like direct trust, indirect trust, sender based reputation level and geo-/situation-oriented reputation levels. However, detail of this system is not sufficiently provided in the paper. They did not mention how sender based information will be updated. The authors claim that their scheme will use direct and indirect trust methods but they didn't mention indirect trust method in detail. Another problem of the system is that every node is giving opinion about distributed content. The opinions of earlier nodes will have greater influence as compared to nodes that provide opinion later. As soon as different nodes add their opinion, the message size will increase eventually. In this way, continuously added overhead to the packet make this scheme unsuitable for an ephemeral environment. This scheme is based on event based messages while beacon messages can be useful in decision making to determine trustworthiness.

6.3 Data-Centric Trust Based Security

Raya et al. [22] proposed a framework that provides data centric trust based security. It is a trust management scheme which is applied to traffic safety application in vehicular ad hoc networks. The main emphasis of this approach is on the evaluation of the trustworthiness of the message or data reported by other entities. They proposed data centric trust establishment approach. This approach combines multiple evidences for trust. In this framework, a collection of multiple reports is passed to a decision logic module. Specific weights associated with these reports are also passed along with these reports to module where different techniques are applied to derive the level of trust of the given data. For evaluation, they use Bayesian inference and Dempster Shafer theory

(DST) [23]. Finally, an assessment is produced by this decision logic module. This assessment defines the level of trust of the event and also analyzes, whether in real this event has taken place or not. There are some shortcomings of this technique: (1) the model will not be suitable for spars environment (VANET with a low node or vehicle density). In spars environment, limited reports are available regarding to an event. So it would not possible to establish trust again and again between entities. By using this technique in realtime, we can't combine local views. Another problem of this technique is that it relies on data sensed by sensors or received by any other entity. In case, entity doesn't detect event properly due to inefficiency of its sensors then evaluation result – that depends on this received information – may not be fully accurate.

6.4 ID-Based Authentication Framework for VANET

Hung lu et al. [24] proposed an authentication system for VANET. This proposed framework uses identity based encryption. For privacy purpose, pseudonyms are used. These pseudonyms are self generated identifiers. For the authentication of Road Side Units (RSU) and vehicles ID-Based Signature (IBS) scheme is used and ID-Based Online/Offline Signature (IBOOS) technique is used for authentication between vehicles. Authentication process is divided into three parts, vehicle to roadside (V2R) authentication, vehicle to vehicle authentication and road side to vehicle authentication. Before a vehicle comes on a road, it requests to a Regional Trusted Authority (RTA) for registration. In response of authentication request, RTA publishes certified domain parameters and it computes and stores the hash values of all registered vehicles – against their real word ids. The RSU periodically broadcast beacon messages. Vehicles will use self created pseudonyms. A vehicle uni-casts its new generated pseudonym to RSU. The RSU verifies this pseudonym by checking signature and accepts the authenticated message. Then the RSU generates offline signature for (requesting) vehicle and then uses IBS for authentication and broadcasts an allocation set message to all vehicles in the network for vehicle to vehicle communication. Each vehicle accepts this message if the signature verification is valid otherwise drops it. There are some limitations of this proposed authentication scheme. This architecture uses ECC based signatures scheme that takes less time to generate signature and is more efficient in storage as compared to RSA but ECC take more time for signature verification as compared to verification through RSA. Another problem occurs in this scheme, when a vehicle A receives an authentication message from another vehicle B. In this message, storage for the pseudonym and POI set of vehicle B is checked by the vehicle A. If it is not found then vehicle A send a query with POI set of vehicle B to nearest RSU for authentication. RSU further sends this query to other RSU or RTA for authentication; finally the reply is received from current RSU whether POI is authenticated or not. VANET is a wireless network, therefore it can take much time for this whole process. Eventually, this scheme will become unsuitable for safety critical applications – where we cannot afford a delay of even milliseconds.

6.5 Message Linkable Group Signature (MLGS)

Qianhong Wu et al. [25] proposed a new privacy preserving technique called Message Linkable Group Signature (MLGS). This technique provides anonymous authentication. In this technique it is assumed that most of the vehicles in network are honest. In this system, a threshold mechanism is used as a priori countermeasure. A message is considered a trustworthy if at least n vehicles endorse this message. This n is a threshold which is adaptive. The sender can change the threshold. If a node produces two signatures on one message then a trusted authority will identify it as an attacker. *Sybil attack* can be avoided by using this technique. If a vehicle receives a message with multiple signatures, it can be checked whether these multiple signatures are from a single vehicle or from multiple honest vehicles.

6.6 Trust Management System - RaBTM

Yu-chih et al. [26] presented a new trust management system (RaBTM) that is based on a RSU and beacon messages. To check whether coming message is trustworthy or not is handled by using direct trust methods as well as indirect trust methods. Through this scheme, message opinions quickly propagate in the network. Using the beacons based trust, position, velocity and drive direction of vehicular node can be determined. The authors use *Tanimoto coefficient measure* to determine the trustworthiness of the vehicles; they compare the sent and estimated position of the vehicles and calculate the difference to estimate the trustworthiness. In indirect method, the trustworthiness of message is determined by checking the trust value between sender and receiver. This trust value should not be more than the value of trust between receiver/sender and forwarder. To determine the trust worthiness of event messages, this scheme doesn't rely only on beacon messages & event messages but also gets assistance from the RSU. According to authors, all transmitted messages are protected, but they did not describe any cryptography or pseudo identity scheme in their paper.

6.7 Identity-Based Cryptosystem for VANET

Sun et al. [27] presented a cryptosystem for VANETs which is identity based. This scheme uses pseudonym to provide privacy and traceability. This system avoids the use of certificates and in this way reduces overheads. The pseudonym may be generated by node itself or by the fixed RSUs. This system uses identity based cryptography that allows public key to be derived from user's public identity like name or email address. This cryptosystem uses a defense scheme in which threshold signature and threshold authentication are to be considered for achieving non frame ability and privacy preservation against misbehaving nodes. This system uses proof of knowledge technique for authentication based defense scheme. This nonframeability scheme is basically designed for law enforcement authorities. It is desirable that there should be two or more authorities for identity retrieval process. This scheme has some issues like there is no simulations or experimentations so it is not clear that how id based cryptosystem will operate in vehicular ad hoc networks. Second issue is about revocation of user's public key. In pseudonym-based scheme the public key of user is his/her name or email address so when revoke a user's public key means change his/her identity which is impractical in real traffic scenarios.

6.8 Crypto-Based Security for VANET

Wang et al. [34] provides a novel secure scheme for VANETs. This scheme not only provides privacy rather also provides confidentiality and non-repudiation. For secret communication, this scheme uses symmetrical cryptography. To make more secure, they used Advanced Encryption Standard (AES); and to provide confidentiality and non-repudiation, message is signed by sender's private key. This scheme is based on certificate based public key cryptography so overhead of such scheme can create delay in transmission. So this time consuming verification process is not suitable for for time critical safety applications in VANET. Another problem of this scheme is this, it uses AES for encryption which extremely processor intensive algorithm. So efficiency can be affected even system would simply not work by using such algorithms.

6.9 Pairing-Based Decentralized Revocation

Wasef et al. [35] proposed a revocation protocol for VANETs. This novel protocol is based on pairing Efficient Decentralized Revocation (EDR). It's decentralized nature makes it possible that a group of neighboring vehicles – that would be legitimate vehicles – will revoke a nearby vehicle which is misbehaving. In this protocol a master key is divided mathematically into number of parts and these parts are distributed in whole network. This protocol is based on probabilistic random key distribution. When a node misbehaves, its neighbors vote to revoke this misbehaving node. The node whose accumulates votes exceeding the threshold that defined to revoke a node act as a revocation coordinator and sends a revocation request to its one hop neighbors. When a neighbor receives this request message verifies the signature of the coordinator by using coordinator's public key which is in this request message. This node calculates its share by using its mathematical part and then sends it to revocation coordinator. Again, revocation coordinator verifies this share; if it fails then it simply drops it. If it is passed then it waits for all other shares. After receiving all required revocation shares, it computes the total revocation message signature and after verification, the revocation coordinator broadcasts a certificate revocation message. When any neighbor node receives this message, after verification it validates the revocation. In this protocol, it is necessary to determine the average number of vehicles within a defined communication range and according to this average value, setting the value of N and in this way calculating the revocation success probability. Though this average number of vehicles can vary according to scenario like in traffic jam, number of vehicles are more in a communication range of any vehicle as compared to normal traffic flow. The authors didn't discuss about the value of N, would it be adaptive or fix? Or if it is fixed then in any specific scenario like in traffic jam the revocation success probability could be low. Moreover, the proposed protocol is based on detection of misbehaving nodes but the authors didn't explain the specific misbehavior of nodes which is being detected.

6.10 Anonymous Authentication Protocol

X.Wang et al. [36] proposed an anonymous authentication protocol – based on Certificate-based Cryptography (CBC) – that provides conditional privacy and non-repudiation. It is for V2I communication. In first step, a certificate-based signature scheme is proposed. This scheme employs bilinear pairing in signature scheme that further consists of five algorithms. In these algorithms, Transportation Regulation Center (TRC) determines the system public parameters by using cryptographic hash functions. This TRC generates public and private keys for user. It also constructs the certificate for a user and sends it through a secure channel to him. User signs the message with his private key and certificate. User's signatures can be checked by the signer's public key. Later on, in second step, this certificate-based signature scheme is used in anonymous authentication protocol. This authentication protocol consists of three phases. In first initialization phase, TRC determines & publishes the system public parameters and loads two cryptographic hash functions along with a message authentication code. In second phase, TRC generates private, public key pair along with certificate as described above. TRC also creates a secret account for *On Board Unit* (OBU) which includes an index of OBU's account, unique identifier and public key of OBU. It also includes the verification result of OBU's public key and related information (like traffic information that it has sent) as a part of this secret account. TRC maintains an Account List (AL) for all OBUs' accounts. TRC sends this AL to Road Side unit (RSU) through a secure channel. In the third phase, OBU and RSU authenticate each other. In this phase, a secure session key is built which provides perfect forward secrecy because if primary keys of OBU and RSU are compromised even then illegitimate user can't compute this session key due to known computational problem of Diffie-Hellman.

Finally, we provide the correlation of the security solutions with their provided security features according to proposed taxonomy – in addition to misc. features – in Table 1.

Table 1. Correlation (matrix) of the security solutions to taxonomy & misc. features

VANET Security Solutions	Entity-based Rep.(VARS)	Data-Centric Trust-Model2	Event-based Rep.	ID-Based Auth.	MLGS	RaBTM	ID-Based Security	D. Sig. & Pwd	Cert. Auth.	EDR
Taxonomy Connects										
Reputation	✓	✓	✓			✓				
Authentication				✓	✓		✓	✓	✓	
Privacy				✓	✓	✓	✓		✓	✓
Confidentiality							✓			
Misc. Features										
Non-Repudiation				✓			✓		✓	
Integrity		✓						✓		
Revocation										✓
Pseudonymity				✓	✓	✓		✓	✓	

7 Conclusions

In near future, it is expected that Vehicular ad hoc networks will deploy in different countries. Security of such networks is very essential because people's lives may be at stake due to it. In this paper we have explained why this problem has such particular requirements. We also describe different types of threats that are possible in vehicular ad hoc networks. We also surveyed the literature on several security issues specifically related to VANETs. These security issues make a potential stumbling block to deploy VANETs. From the analysis in survey, we came to know that – up till now – there doesn't exist a comprehensive security protocol or framework that covers all security aspects of VANET. So, there is need to develop such a framework which mitigate all these security problems; more research is required in this area.

References

1. Schweiger, B., Ehnert, P., Schlichter, J.: Simulative Evaluation of the Potential of Car2X-Communication in Terms of Efficiency. In: Strang, T., Festag, A., Vinel, A., Mehmood, R., Rico Garcia, C., Röckl, M. (eds.) Nets4Trains/Nets4Cars 2011. LNCS, vol. 6596, pp. 155–164. Springer, Heidelberg (2011)
2. Mehmood, R., Nekovee, M.: Vehicular Ad hoc and Grid Networks: Discussion, Design and Evaluation. In: Proc. of the 14th World Congress on Intelligent Transport Systems, p. 8 (2007)
3. Lee, K.C., Lee, U., Gerla, M.: Survey of Routing Protocols in Vehicular Ad Hoc Networks in Car2Car communication consortium (2010)
4. Kumar, S., Narayan, K.D.: Qualitative Based Comparison of Routing Protocols for VANET. Journal of Information Engineering and Applications 1(4) (2011)
5. Gillani, S., et al.: Vehicular Ad hoc Network, Enabling Secure and Efficient Transportation System, Technical Journal, University of Engineering and Technology, Taxila, Pakistan (2008)
6. Sakib, R., et al.: Security Issues in VANET, Research Thesis (2010)
7. Karagiannis, G., et al.: Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards and Solutions. IEEE Communications Surveys & Tutorials
8. Toor, Y., et al.: Vehicle ad hoc networks: Applications and related technical issues. IEEE Communications Surveys & Tutorials 10(3), 74–88 (2008)
9. Saad, A., Weinmann, U.: Automotive software engineering and concepts. GI Jahrestagung, Frankfurt, Germany, pp. 318–319 (2003)
10. Raya, M., Hubaux, J.-P.: Securing vehicular ad hoc networks. Journal of Computer Security, 39–68 (2007)
11. Raya, M., Papadimitratos, P., Hubaux, J.P.: Securing Vehicular Communications. IEEE Wireless Communications 13 (2006)
12. Parno, B., Perrig, A.: Challenges in Securing Vehicular Networks. In: Proc. of HotNets-IV (2005)
13. Aad, I., Hubaux, J.P., Knightly, E.W.: Impact of Denial of Service attacks on Ad Hoc Networks. IEEE/ACM Transactions on Networking 16 (2008)

14. Raya, M., Pierre Hubaux, J.: The security of VANETs. In: Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks (2005)
15. Khayam, S.A., Radha, H.: Analyzing the Spread of Active Worms over VANET. In: ACM Mobicom International Workshop on Vehicular Ad Hoc Networks (2004)
16. Raya, M., Papadimitratos, P., Hubaux, J.-P.: Securing vehicular communications. IEEE Wireless Communications Magazine 13(5), 8–15 (2006)
17. Gada, D., et al.: A Distributed Security Scheme for Ad Hoc Networks. ACM Crossroads, Special Issue on Computer Security 11(1), 1–10 (2004)
18. Samara, G., et al.: Security Analysis of Vehicular Ad Hoc Nerworks. In: International Conference on Network Applications Protocols and Services (2010)
19. Zeadally, S., Hunt, R., Chen, Y., Irwin, A., Hassan, A.: Vehicular Ad Hoc Networks: Status, Results, and Challenges (2010)
20. Nai-Wei, L., Hsiao-Chien, T.: A reputation system for traffic safety event on vehicular ad hoc networks. EURASIP Journal on Wireless Communications and Networking (2009)
21. Dotzer, F., et al.: VARS: A Vehicle Ad-Hoc Network Reputation System. In: Symposium on a World of Wireless Mobile and Multimedia Networks, pp. 454–456 (2005)
22. Raya, M., Papadimitratos, P., Gligor, V.D., Hubaux, J.P.: On data-centric trust establishment in ephemeral ad hoc networks. In: INFOCOM - Conference on Computer Communications, pp. 1238–1246. IEEE (2008)
23. Chen, T.M., Venkataramanan, V.: Dempster-shafer theory for intrusion detection in ad hoc networks. IEEE Internet Computing 9(6), 35–41 (2005)
24. Lu, H., Li, J., Guizani, M.: A novel ID-based authentication framework with adaptive privacy preservation for VANETs. In: Computing, Communications and Applications Conference (ComComAp), pp. 345–350. IEEE (2012)
25. Wu, Q., Domingo-Ferrer, J., González-Nicolás, U.: Balanced trustworthiness, safety, and privacy in vehicle-to-vehicle communications. IEEE Transactions on Vehicular Technology 59(2), 559–573 (2010)
26. Wei, Y., Chen, Y.: An Efficient Trust Management System for Balancing the Safety and Location Privacy in VANETs. In: TrustCom, pp. 393–400 (2012)
27. Sun, J., et al.: An Identity-Based Security System for User Privacy in Vehicular Ad Hoc Networks. IEEE Transactions on Parallel and Distributed Systems 21(9), 1227–1239 (2010)
28. Moharrum, M.A., Al Daraiseh, A.A.: Toward Secure Vehicular Ad-hoc Networks: A Survey. IETE Technical Review 29 (2012)
29. Karagiannis, G., et al.: Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards and Solutions. In: IEEE Communications Surveys & Tutorials, pp. 584–616 (2011)
30. Haseeb, K., Arshad, M., Yasin, S., Abbas, N.: A Survey of VANET's Authentication, Islamia College Peshawar, Pakistan (2010)
31. Xiong, H., Guan, Z., Hu, J., Chen, Z.: Anonymous Authentication Protocols for Vehicular Ad Hoc Networks: An Overview (online) (2012)
32. Papadimitratos, P., et al.: Secure vehicular communication systems: Design and architecture. IEEE Commun. Mag. 46(11), 100–109 (2008)
33. Klaus, et al.: A privacy aware and efficient security infrastructure for vehicular ad hoc networks. Computer Standards & Interfaces 30, 390–397 (2008)

34. Wang, N.W., Huang, Y.M., Chen, W.M.: A novel secure communication scheme in vehicular Ad Hoc networks. Journal of Computer Communications 31(12), 2827–2837 (2008)
35. Albert, W., Shen, X.: EDR: Efficient decentralized revocation protocol for vehicular ad hoc networks. IEEE Transactions on Vehicular Technology 58(9), 5214–5224 (2009)
36. Wang, X., Liu, T., Xiao, G.: Certificate-based anonymous authentication protocol for vehicular Ad-hoc network. IETE Technical Review 29 (2012)
37. Figueiredo, L., et al.: Towards the development of intelligent transportation systems. IEEE Intelligent Transportation Systems, 1206–1211 (2001)
38. André, W., et al.: Data security in vehicular communication networks. VANET: Vehicular Applications and Inter-Networking Technologies (2010)
39. Antonios, S., Chai, Z.: Survey of Security in Vehicular Networks. Project CPSC 534 (2007)

Wireless Vehicular Network Standard Harmonization

Tsutomu Tsuboi

Hamamatsu Agency for Innovation
3-5-1, Johoku Naka-ku, Hamamatsu, Shizuoka 432-8561, Japan
tsuboi@hai.or.jp

Abstract. This paper describes next generation vehicular network standard comparison among North America, Europe and Japan and provides harmonization proposal for Japanese DSRC (Dedicated Short Range Communication). The basic DSRC standards are in IEEE (Institute of Electrical and Electronics Engineers Inc.) and in ETSI (European Telecommunication Standard Institute). The IEEE and the ETSI have regular meetings for harmonization, which makes expanding common portion of each standard. However Japanese DSRC standard is ARIB (Association of Radio Industries and Businesses) who has well established previous vehicular communication for ETC (Electric Toll Collection) System application in advance since 1997. Therefore Japanese DSRC standard has been established independently. And it becomes problem for standard harmonization especially in Japan. Japanese ETC system has already expanded into not only Japan but also into Asia pacific countries and has established as ITU (International Telecommunication Union) for ETC. Asian countries including Japan face to improvement for their DSRC standard based on standard harmonization.

Keywords: DSRC, WAVE, IEEE802.11p, Standard harmonization.

1 Introduction

ITS (Intelligent Transport System) becomes near future safety technology for automotive application since 2007 when European Telecommunication Standard Institute has started TC-ITS (Technical Committee – Intelligent Transport System) group. There are several field trials especially European automotive committee such as Car to Car Communication Consortium (C2C-CC). There are also same activities in North America such as "IntelliDrive SM" [1] project under MOT (Ministry Of Transportation). In Asia especially in Japan, ETC system has been well established since 1997 and try to use this technology for ITS. The ETC market in Japan is 4.3 million units in 2011[2]. The ETC technology is used 5.8GHz frequency band. After 2010 when analog terrestrial services has terminated, UHF (Ultra High Frequency) band especially 700MHz band has been open from 2011. Japanese MIC (Ministry of Internal Affairs and Communications) has assigned 10MHz bandwidth in 700MHz band for ITS, not only ETC 5.8GHz band. Therefore there are two types of ITS standard[3] potential system in Japan.

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 75–83, 2013.

In Section 2, it is described each ITS standard comparison and explains standard harmonization. Section 3 describes ITS standard issues especially standard harmonization and explains harmonization challenges by new proposal. In Section 4, it is described ITS application analysis and shown validation of new proposal. In Section 5, it is summarized wireless vehicular communication standard harmonization as summary.

2 ITS Standard

2.1 DSRC/WAVE Standard Comparison

Each regional ITS standards are listed in Figure 1. There are Japanese two types ITS standards, North America IEEE standard, and European ITS standard ETSI. The IEEE standard is IEEE802.11p-2010/IEEE1609 and ETSI standard is ETSI ES202 663/EN102 731.

	Japan DSRC	Japan 700MHz	USA (DSRC)	EU (DSRC)
Application Layer	**application** **ETC**	**application** **Anti-collision**	**application** ●**Anti-collision** ●**High speed data**	**application** ●**Anti-collision** ●**High speed data**
Upper Layer Protocol	**DSRC Protocol**	**Dedicated (current)**	**Application mng.** IEEE1609	**Application mng.** **UDP/TCP, IPv6** **WSMP (non-IP)** **LLC**
Access typ.	**TDM/FDD**	**CSMA/CA**	802.11p **CSMA／CA**	
Modulation	**QPSK**	**OFDM (BPSK/QPSK/16QAM)**		
No. of channel	5MHzx7ch x2(up/down)	■CCH 10MHzx1ch	■CCH ■SCH 10MHzx7ch (20MHz option)	■CCH ■SCH 10MHzx3ch
Frequency Band	**5.8GHz**	700MHz	5.9GHz	5.9GHz

Fig. 1. WAVE Worldwide Standard comparison

There are individual regulations in regard to frequency allocation because of historical development in each country: 5.8GHz and 700MHz in Japan, 5.9GHz in Europe and North America. There is major difference among those standards, which is channel service allocation. There are two types ITS service channels, Control Channel (CCH) and Service Channel (SCH) in Europe and North America. However there is only one CCH in current Japanese standard proposal because of ETC historical development. And there are also several differences among standards such as access type and upper layer protocols. According to Figure 1, there are common parts, which are access type and upper layer protocols between North America and Europe. The reason why NorthAmerica and Europe harmonization is established, is a regular meeting between both standard activities in order to make more common technology application. On the other hand, there is also Asia Pacific reguraty meeting but harnomization becomes complicated because of the differences among those countries frequency allocation[4].

2.2 ITS Frequency Allocation

As for frequency allocation, it is defined by each country regulation. The ETSI EN 302 571[5] is defined ITS. The IEEE 802.11p is followed Federal Communication Commission (FCC)[6] licensing. The summary of ITS related frequency allocation for each North Amrica, Europe and Japan is shown in Figure 2. In Japanese 700MHz frequency band, the previous trail was used between 710MHz and 730MHz including 5MHz guard band each edge of frequency. However it has been planed to shift from 720MHz center band frequency to 760MHz because of avoidance for digital terrestrial service interferencecrosstalk for lower 700MHz freqency. The total bandwidth is 10MHz.

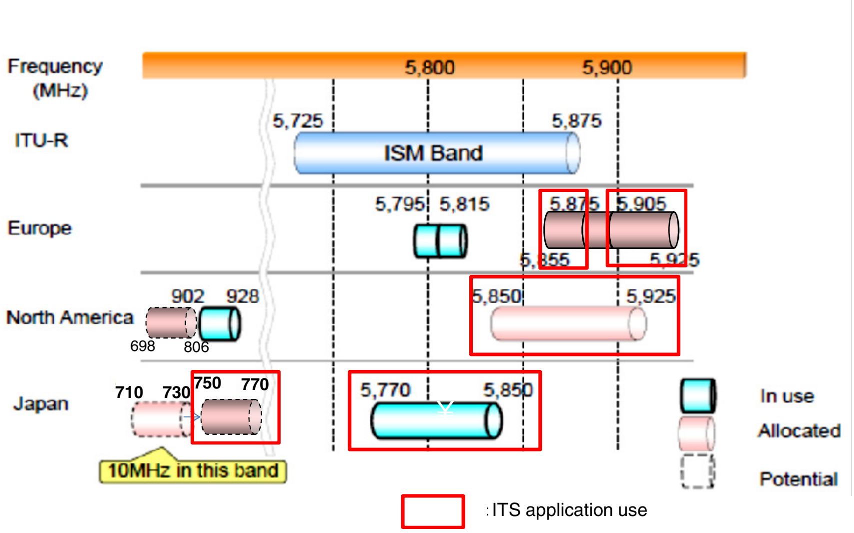


Fig. 2. World ITS frequency Allocation

3 ITS Standard Issues and Challenges

According to Figure 1, there are differences of upper layer protocol, access type and number of channel between Japanese standard and North America / Europe standards. Author has proposed counter measures for standard harmonization into Japanese ITS standard especially for 700MHz frequency band. As mentioned in Section 1, 5.8GHz frequency band is ETC application related band, therefore there is already lots of software established for ETC. Accordingly, there is condition between ITS and ETC in case of changing parameters such as access type, modulation and service channel. Therefore author focuses on 700MHz specification at this tme.

As for standard harmonization, there are several solutions for standard harmonization. One idea is to split a CCH service into two (CCH and SCH) by frequency division multiplexing (FDM) like 5MHz and or 2.5MHz band width[7]. In case of 5MHz band width, there are one CCH and one SCH with each 5MHz (Two Channel Solution) band width. In case of 2.5MHz band width, there are one CCH and three SCHs for example (Four Channle Solution). Second approach is time division multiplexing (TDM) access type such as 50 millisecond CCH and 50 millisecond SCH in 100millisecond time frame span. Therefore the half time span is used as each CCH or SCH. Once channel service is divided into CCH and SCH, it is easy to implement existing SCH service protocol such as UDP (User data Program), Network Service (IPv6), and LLC (Logical Link Control) followed by resource management (IEEE1609.1) The UDP/IPv6/LLC structure is followed IEEE/ETSI specification, which is able to support standard harmonization in regard to SCH. When it is necessary to support existing current Japanese 700MHz ITS standard condition, the sprit CCH is able to use independently without any damage of standard harmonization of SCHs.

Author proposes FDM type because it is more useful for assign multiple SCHs and more flexible in future expand ITS frequency allocation. This idea is also used for European standard. There is diffeneces between number of SCH. The IEEE standard has six SCHs and ETSI standard has two SCHs. Therefore the proposal idea is able to use for expantion of ETSI to prepare more SCHs.

The harmonization proposal idea is shown in Figure 3.

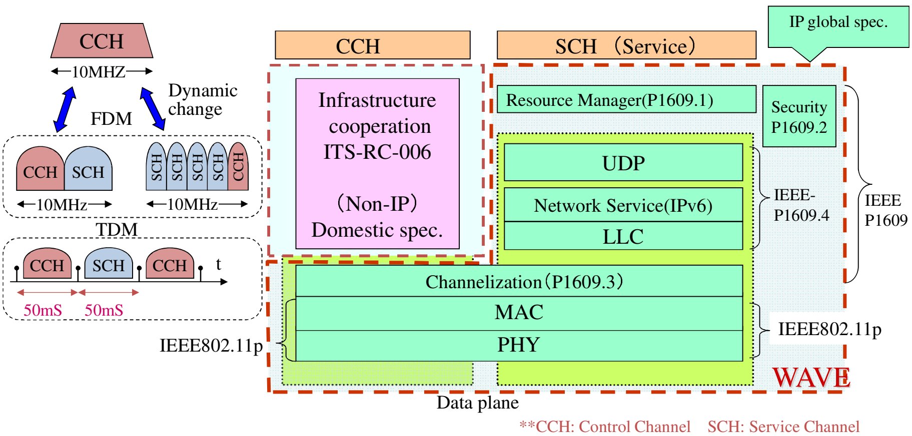


Fig. 3. WAVE New Proposal Structure

As this section summary, Figure 4 shows among those standards comparison. According to Figure 4, the new Japanese ITS proposal of 700MHz band is able to be compatible for de-facto ITS standards (IEEE and ETSI). In Figure 4, upper layer is also followed de-facto ITS standards. However it is not necessary to follow de-facto ITS standards in regard to CCH in case of keeping existing Japanese CCH as described in Figure 3.

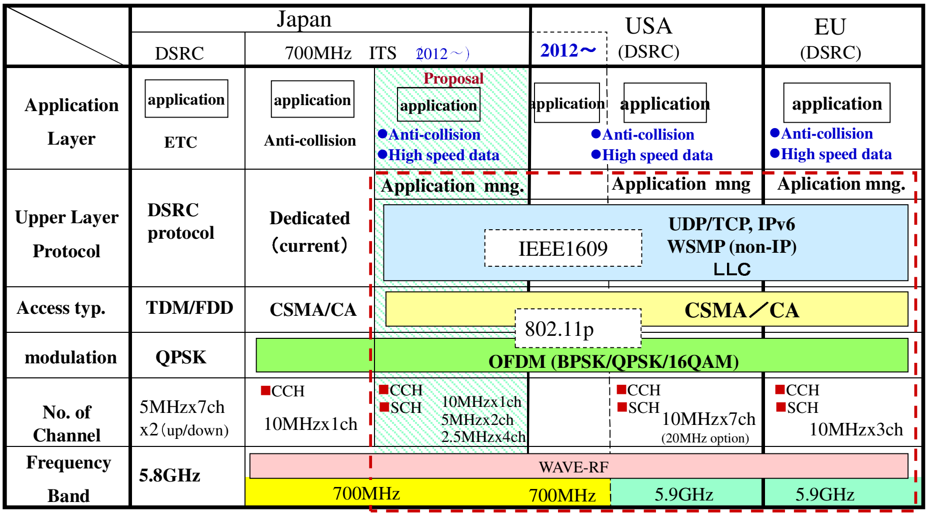


Fig. 4. WAVE standard and harmonized proposal

4 ITS Application Analysis and New Proposal Validation

4.1 Validation Conditions

The next generation ITS application is mainly for safety. There are several key information in ITS and Table 1 shows safety related information factors in DSRC/WAVE application. The detail Basic Safety Message (BSM) is defined by SAE International standard (SAE J2735)[8]. The total messaage size of safety application becomes 80 bytes. The Table 1 shows each size of items of standard. In case of more information and headers in future, the total length of data in this paper sets 100 Bytes.

Table 1. ITS Safety Information

Information	Content	Unit (Byte)
Mangement info	Data Version etc.	2
Basic Data	Vehicle ID, target vehicle ID, position of vehicles, velocity, direction	27
Vehcile status	shift position, Break condition, winker status, hazard status	2
Traffic condition	moving condition (Emergency vehicle, bus traffic)	5
Positioning	Cross section condition for moving direction etc.	22
Situation info	mutual concession, course info, slow down, warning (road, jam)	2
Free data	free data for users	20
Total		80

In order to validation of new Japanese ITS proposal, it is examined to use one of ITS access condition as follows;

- Case study place: Tokyo Japan
- Density of vehicle: 2,000 vehicles per km^2
- Maximum velocity of vehicle: 120 km/hour
- Space between vehicles: 5 meters (one regular vehicle size)
- WAVE RSU (Road Side Unit) coverage[9,10]: 315 vehicles
- Transmission cycle : 100 millisecond
- Wireless access modulation: 16QAM and QPSK
- Multi-channel type: FDM
- WAVE band width: : 5 MHz and 2.5 MHz

The condition of traffic average density in metropolitan Tokyo[11] is 500 vehciles km^2 which becomes more safety side than the case study with 2,000 vehicles km^2.

4.2 16QAM(Quadrature Amplitude Modulation)

Accpording to Section 4.1 condition, WAVE transmission capacity (D4) is calluculated as follows;

D4 = D1 x D3 / D2 -- (1)
D1: Total vehicle capacity in cell
D2: Transmission cycle for WAVE data transmission
D3: Transmission data size

The requried transmission band width D6 is calluculated as follows;
D6 = D4/D5 -- (2)
D5: Spectral Efficiency[12]

The result of the required transmission bandwidth is 1.84MHz in Figure 2. Therefore it is able to be covered by 2.5MHz freqency band. It is satisfied with Four Channel Solution.

Table 2. Validation of 16QAM WAVE

	Item	Unit	Value	Note
D1	Vehicle number		315	
D2	Tx cycle	ms	100	
D3	Tx Data	Byte	100	
D4	Transmit Capacity	Mbps	2.52	D1*D3/D2
D5	Spectral efficient	bit/Hz	1.37	16QAM[12]
D6	Transmit Bandwidth	MHz	1.84	D4/D5

4.3 QPSK(Quadrature Phase Shift Keying)

In case of QPSK, the required transmission band width is calluculated by same method of 16QAM. The result of the required transmission band width is 3.68MHz in Table 3. Therefore required channel bandwidth can be covered by 5MHz. It is satisfied with Two Channel Solution.

Table 3. Validation of QPSK WAVE

	Item	Unit	Value	Note
D1	Vehicle number		315	
D2	Tx cycle	ms	100	
D3	Tx Data	Byte	100	
D4	Transmit Capacity	Mbps	2.52	D1*D3/D2
D5	Spectral efficient	bit/Hz	0.684	QPSK[12]
D6	Transmit Bandwidth	MHz	3.68	D4/D5

4.4 Image Data Information

In case of using image picture for WAVE such as moving picture information in ITS application, it is necessary to more band width because of the data capacity. The Table 4 shows typical urban district information including image data.

The highest data rate is 11.5Mbps (Mega bit per second) as moving picture data. The picture information is used the following typical conditions;

Image data length: 8 bits
Picture resolution: 680 x 480 picels
Data rate: 30 Frames per second
Image compression: 1/8 (typical cammera specification)

Accroding to Table 5, the required channel bandwidth is 8.39MHz, which is over 5MHz limitation band width. Therefore it is not coverd by Two Channel Solution. It has to use CCH for moving picture transmission.

Table 4. Metoropolitan Typical Information example

No.	Item	data rate(kbps)	Notes
1	Picture info	11,500	10 bitsx680x480x30 frame x 1/8(compression)
2	data info	0.8	Vehicle ID, position, velocity, direction, contrl info
3	jam info	384	equal to Max speed of cellular
4	telematics	278	Music info ex. MP3

Table 5. Moving Picture Transmission Bandwidth

	Item	Unit	Value	Note
G1	Picture Tx capacity	Mbps	11.5	
G2	Frequency efficiency	bit/Hz	1.37	16QAM
G3	Transmit Bandwidth	MHz	8.39	G1/G2

5 Summary

Author proposes FDM technology for ITS harmonization standard especially for Japanese ITS specification as the new proposal. The summarized new proposal for 700MHz band ITS shows to work with other standards IEEE and ETSI under the current existing Japanese study. It achieves the followings;

1) Current Japanese ITS standard has only one CCH against multiple channels under North America and Europe ITS standard. The new proposal with FDM technology for current Japanese ITS standard is able to be harmonized with other worldwide ITS standard, which has CCH and SCH(s) with 5MHz or 2.5MHz channel bandwidth.
2) The new proposal is able to satisfied with the current Japanese standard study (RC-006).
3) The new proposal shows confidence to support typical safety information by FDM technology.
4) In case of moving picture transmission, the required channel bandwidth is 10MHz however it is also covered under new proposal which needs one channel application.
5) Japanese 5.8GHz ITS standard is original ETC application. It has to be applied for dedicated application in regard to standard harmonization.

Acknowledgements. Author provides appreciation to Dr.Noriaki Yoshikawa of Cyber Creative Institute for study of standard harmonization proposal suggestions. Author provides deep appreciation to deceased Dr.Jun Yamada of Cyber Creative Institute for many technical suggestions and guides in driving new proposal.

References

1. Ammana, A.: Overview of IntelliDrive / Vehicle Infrastructure Integration report (VII) (2009),
   http://filebox.vt.edu/users/aamanna/web%20page/VII-IntelliDrive%20Report-edited.pdf
2. Automobile Reserach Institute,
   http://www.jari.or.jp/resource/pdf/H22jigyo1/0711-4.pdf
3. ARIB RC-005, RC-006 ITS standard proposal
4. Asia Pacific Telecommunity Task Group 4,
   http://www.apt.int/awg-technology-tg
5. The Eurpean table of frequency allocations and applications in the frequency range 9 kHz to 3000 GH,
   http://www.erodocdb.dk/docs/doc98/official/pdf/ERCRep025.pdf
6. Federal Comunication Commision, Dedicated Short Range Communications (DSRC) Service,
   http://wireless.fcc.gov/services/index.htm?job=service_home&id=dedicated_src
7. Yoshikawa, N.: Information Telecommunication technology committee Netwroked car working group (NWedCarWG) 2nd meeting (2010)
8. SAE International standard SAE J2735 - Dedicated Short Range Communications (DSRC) Message Set Dictionary
9. Tsuboi, T.: WAVE system architecture design in Urban area / ITST (2011)
10. ITS Promotion Committee, ITS FORUM RC-007 (2009)
11. Tsuboi, T.: UHF Band WAVE System Architecture Design / Intelligent Vehicle (2011)
12. Ministry International Affairs and Communications (MIC) Repoer,
    http://www.tele.soumu.go.jp/resource/j/ref/portal

Multi-technology Vehicular Cooperative System Based on Software Defined Radio (SDR)

Nathalie Haziza[1], Mohamed Kassab[2], Raymond Knopp[3], Jérôme Härri[3], Florian Kaltenberger[3], Philippe Agostini[1], Marion Berbineau[2], Christophe Gransart[2], Joëlle Besnier[4], Jacques Ehrlich[4], and Hasnaa Aniss[4]

[1] Thales Communications & Security, Gennevilliers, France
{nathalie.haziza,philippe.agostini}@thalesgroup.com
[2] Univ. Lille Nord de France, IFSTTAR, LEOST, Villeneuve d'Ascq, France
{mohamed.kassab,marion.berbineau,christophe.gransart}@ifsttar.fr
[3] EURECOM, Sophia Antipolis, France
{raymond.knopp,jerome.haerri,florian.kaltenberger}@eurecom.fr
[4] IFSTTAR, LIVIC, Versailles, France
{joelle.besnier,jacques.ehrlich,hasnaa.aniss}@ifsttar.fr

Abstract. Within the scope of the European policy for Intelligent Transport Systems (ITS), the PLATA-PROTON project proposes a multi-technology cooperative Advanced Driver Assistance System (ADAS), based on the integration of Software-Defined Radio (SDR) devices in vehicles. With the choice of significant road-safety related scenarios, V2V and V2I communications have been both implemented and simulated. This paper proposes an overview of the Software-Defined Radio (SDR) platform development and the performance evaluation based on network simulation, performed within this project.

Keywords: Software-Defined Radio, vehicular communication, V2V, V2I, driver assistance application, proof of concept, network simulation, IEEE 802.11p.

1 Introduction

Nowadays, the integration of cooperative functions is the main focus of advanced driver assistance systems (ADAS). Following the European policy for Intelligent Transport System (ITS), the cooperative functions are used to enhance road safety, to optimize the traffic and to reduce the impact of transports on the environment. A key potential is the combination of local environment data (exchanged between neighbor vehicles or between vehicles and Road Side Units) and regional traffic data broadcast. In other words, the drivers can get detailed information from immediate surroundings and longer-term forecasts and alerts controlled from a central infrastructure. The communication system associated to such functions must support vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (broadcasting) communications. Several wireless technologies are able to support these communications

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 84–95, 2013.

e.g. WLAN technologies for V2V and V2I and technologies such as Digital Audio Broadcasting (DAB), Digital Multimedia Broadcasting (DMB) and Digital Video Broadcasting (DVB) for broadcasting from the infrastructure. The technologies used for cooperative systems will depend on communication standards and infrastructures defined in each country. Several working groups are active in the field of communications for ADAS. The International Organization for Standardization (ISO) TC204 Working Group 16 [1] produced specifications for the Communications Access for Land Mobile (CALM) [4], providing a reference framework for future implementations. Harmonization at the European level has been provided with C2C-CC (with the European Telecommunications Standards Institute-ETSI [2]), and with the network working group for the IETF and WAVE for IEEE [3]. Among them, the architecture proposed by CALM is particularly interesting regarding interoperability and flexibility. It is based on heterogeneous cooperative communication framework to provide continuous communications to users. The CALM concept is based on the juxtaposition of several communication modems in the vehicles [4]. The need of such heterogeneity is justified by several characteristics of vehicular communications.

As an alternative to several modems to answer to the need of multiple standards, the Software-Defined Radio (SDR) is a wireless system implemented by software routines so that various wireless radios can be supported by the same hardware based on software changes. SDR hardware allows to modify easily the communication technology supported by a terminal in accordance with its environment. The use of SDR allows vehicle manufacturers to adapt communication technologies supported by vehicles to fit to the local standards.

Despite the prospective advantage of interoperability and reconfigurability, several studies to optimize the SDR signal processing chain have shown implementation issues impacting performances that have to be solved. The first main argument concerns the processing speed, needing an optimal sharing in the baseband processor (implementation choices between DSP and FPGA). The second argument deals with the power consumption in the digital signal processing, becoming of major influence for terminal batteries and handset devices. Nevertheless, interest for SDR implementations (vehicular communications, payload systems) kept growing since the 90's, starting with military applications and spreading to civilian areas, thanks to the recent progress in chip development. Semiconductors and automotive manufacturers perceived the combination of telematics and SDR as a promising area.[5][6]

On the other hand, the adequation of the IEEE 802.11p/WAVE standard for vehicular context is now well-known: the average needed data rate stays low, and the WAVE protocol mechanims simplify V2V communications in a restricted area. Comparison with other systems has been performed (LTE in [7] giving advantage to the WAVE system). However, performance assessment for the IEEE 802.11p/WAVE standard performed in LOS and NLOS scenarios (straight roads, urban context with intersections and shadowing buildings) have shown a need to combine two systems, for local information and regional broadcast.

The PROTON-PLATA [12] project lies on a German-French collaboration (Deufrako/ANR agreement) for the development of a common platform, able to deal with various standards, emerging in each country; industrial and academic partners interested in ITS and particularly in vehicular systems were involved. The goal of this project is to show the feasibility and the interest of a cooperative ADAS making use of the SDR technology to enhance communication performances. The main idea is to propose a multi-technology communication infrastructure and to equip On-Board Units by SDR devices that support simultaneously V2V communications, V2I communications and broadcasting (infrastructure-to-vehicle). The scope of the project includes "proof of concept" part and a performance study part. The first one proposes the development of a prototype, including SDR device and driver-assistance applications. The second is a performance evaluation based on network simulation tools.

This paper is organized as follows. In section 2, we propose an overview of research studies related to cooperative ADAS. Section 3 introduces the multi-technology cooperative system designed within the PLATA-PROTON project. In section 4, we describe the prototype development and related experiments. Section 5 propose an overview of results obtained with the simulation based performance study. Concluding remarks are presented in section 6.

2 Related Work

Several researches deal with to the enhancement of ADAS to vehicular cooperative systems. These works vary from specification frameworks proposed by international standardization institutes, *e.g.* ETSI, ISO, research projects managed by academic and industrial consortiums and academic works. The research scope is very large from PHY layer to applications, protocols, security, mobility management, communication architecture, *etc.* In this section, we propose a survey of the most interesting works in this area.

Regarding international-wide, three main groups are working on this subject: CALM for the ISO, WAVE for the IEEE and C2C-CC for the ETSI [2].

At the European level, we have identified several key projects related to the enhancement of vehicular cooperative systems.

The Cooperative Vehicle Infrastructure System (CVIS) [8] (2006-2010) is a very well known European research project related to the enhancement of cooperative services based on interaction between vehicles and centralized infrastructure. The goal of the project was to develop an open application framework that enables a wide range of potential cooperative services to operate in vehicles, roadside equipments and centralized management centers. It assumes some diversity in communication technology (802.11p, WiMAX, UMTS, DVB / DAB, *etc.*) and is based on IP network mechanisms (IPv6, MIPv6, NEMO, *etc.*).

The EVITA (E-safety Vehicle Intrusion proTected Application) [9] (2008-2011) focused on the design of an on-board automotive network architecture, to ensure protection against unexpected or unauthorized manipulation, for sensitive data transmission. The CO-OPerative SystEms for Intelligent Road Safety

(COOPERS) [10] (2006-2011) concern was about the development of *Co-operative Traffic Management* applications on the road infrastructure with an interaction between vehicles and infrastructure. The approach aims to extend the concepts of in-vehicle autonomous systems and vehicle-to-vehicle communication (V2V) with tactical and strategic traffic information which can only be provided by the infrastructure operator in real time (traffic jam warning and guidance, in-car display and alert of area-specific speed limits, car breakdown/emergency services, *etc*). The SAFESPOT project [11] (2006-2010) considered the combination of the information from vehicles and from the infrastructure, to improve the driver perception of the surrounding vehicles. It focuses on the identification of cooperative solutions that will first be applied to the critical areas, such as the so called *black spots*. Sensing systems on the infrastructure side are proposed in combination with the information coming from vehicles to detect critical conditions and events.

All these projects have contributed to the definition of new applications and services for driving assistance and the specification of cooperative architecture models. The PROTON-PLATA project is a follow up of these works as it focuses on the use of multi-technology architecture and SDR devices to enhance the communication capabilities of ADAS.

3 Multi-technology Cooperative System

In this section, we present the main components of the vehicular cooperative ADAS proposed in PROTON-PLATA. We detail the drivers assistance applications and the communication architecture.

3.1 Driver Assistance Applications

We have defined four applications that cover several types of services in vehicular environment.

- **GPS position exchange:** This application allows vehicles to exchange their GPS data. Each vehicle broadcasts a message containing its own information: vehicle category (car, truck, motorbike, bike, etc.), geographical position, speed, course and geolocation time. *Geolocation time* is the time given by the geo-location system.
- **Local weather condition:** This application informs the driver about the weather condition where he/she is and according to the travel direction. A centralized server broadcasts information over a given geographic area using a wireless technology.
- **Incident hazard warning:** This application keeps drivers aware of traffic hazards based on information provided by central server, roadside units, road equipment and other vehicles. The hazards that can be supported by this application are multiple such as wrong way driver also called *Ghost driver*, vehicles at a standstill, obstacles on the road, black ice, *etc.* When a vehicle

detects a hazard on its way, it sends a message to its environment to share the information with other vehicles. Two versions of this application are defined: *V2V hazard warning* and *V2I hazard warning*. *V2V hazard warning* is fully decentralized, a vehicle that detects a hazard sends messages to neighboring vehicles via peer-to-peer communications. With *V2I hazard warning*, a vehicle that detects a hazard sends messages to a server located in the management network via a Road Side Unit (RSU). The server forwards the message to appropriate areas using broadcast communications.
– **Speed limit:** This application defines a database for speed limits in a road network. A reference database is located in a centralized server and speed limits are stored on-board the vehicles in a local database. A new edition of the reference database is published at regular intervals. Vehicle on-board systems have to update their local databases periodically. This application defines dynamic and static versions of updating mechanisms. The *Static speed limit* detects and reports any obsolescence of the embedded database by comparing its version number with the reference database. The *dynamic speed limit* is executed when dynamic changes are decided in the center. Updates are sent to vehicle by means of broadcasting transmissions into the related area.

3.2 Communication System

PROTON-PLATA project proposes a communication infrastructure that allows to fully exploit the SDR capacity of the devices in vehicles. The communication infrastructure manages the centralized part of the cooperative system, namely Network-Head (NH). This architecture connects the road traffic management center to two wireless networks. The first one is a wireless access network based on Road Side Units (RSU) along the roads. The second one is a broadcast network based on Base Stations (BS) with large coverage. Figure 1 shows an example of a deployment. The uplink communications between On Board Unit (OBU) and the Network-Head (NH) are supported by the RSUs network, while the downlink communications are supported by both the BS and the RSU.

The key innovation of the project lies in the design of the multi-technology OBU, which role is to exchange information with the NH (through the RSUs or the BS) and with other OBUs. The OBUs implement different protocol stacks on a *Software Defined Radio (SDR)* device. The SDR is a wireless system implemented by software routines so that various wireless radios can be supported by the same platform based on software changes. SDR hardware allows to modify easily the communication technology supported by a terminal in accordance with its environment and/or needs. The PLATA-PROTON specifications suppose that OBU are equipped with SDR devices that manage V2V communications with other OBUs, V2I communications with RSUs, I2V broadcast reception from BS.

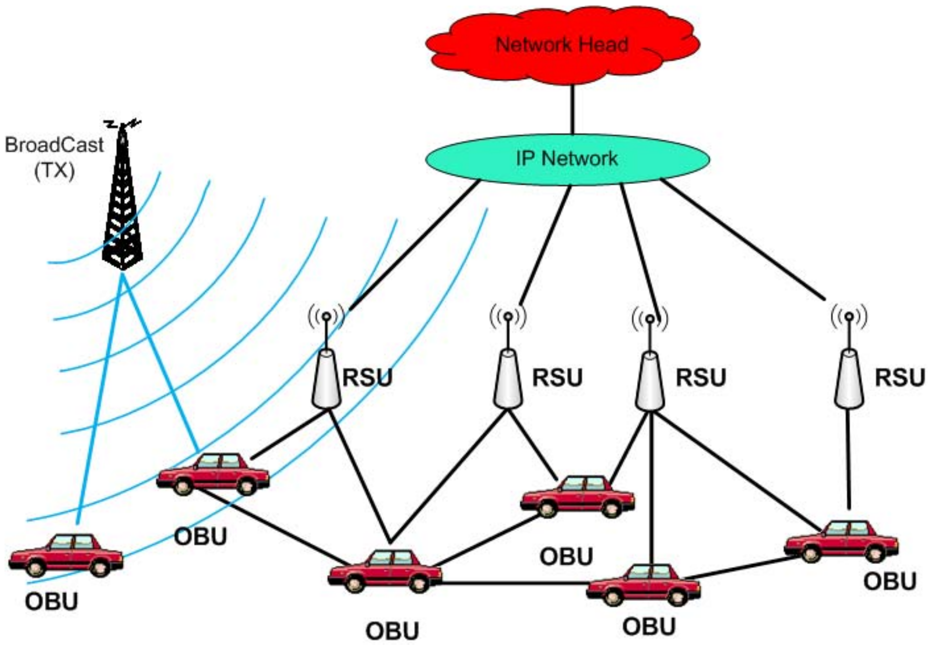


Fig. 1. Communication architecture

4 SDR Prototype

For the SDR protoype design, the choice of the open-source hardware/SDR DSRC prototype on the *OpenAirInterface* platform [14] has been identified as a real asset, as it answers to the overall aim of the cooperative PROTON-PLATA system, precisely to adapt to standards specific to each country. The standards for the PROTON-PLATA have thus been set to the promising LTE wireless standard associated to the 802.11p vehicular technology.

Figure 2 depicts the software architecture of the OpenAirInterface SDR platform. It is segmented into three parts, the *User Space*, the *Kernel Space* and the *SDR Express-MIMO board*. The user-space contains the code regarding the 3GPP LTE protocol stack and the IEEE 802.11p soft-modem from the OpenAirInterface. It also contains the ITS safety applications that sends IP packets to the PROTON-PLATA virtual interface on the linux subsystem. The kernel-space contains deep linux routines and kernel extensions modules required by PROTON-PLATA. In particular, it contains a hook from the linux subsystem either to the linux wireless subsystem modified to support the IEEE 802.11p (OCB mode) and to a IEEE 802.11p driver, or to the LTE NASMesh device. Finally, the PCI-Express MIMO SDR board is connected via a PCI-Express and an OpenAirInterface firmware to the OpenAirInterface main driver. The salient aspect is that PROTON-PLATA designed it so to be able to support both LTE and 802.11p signals on the same board, and as such shall access to the

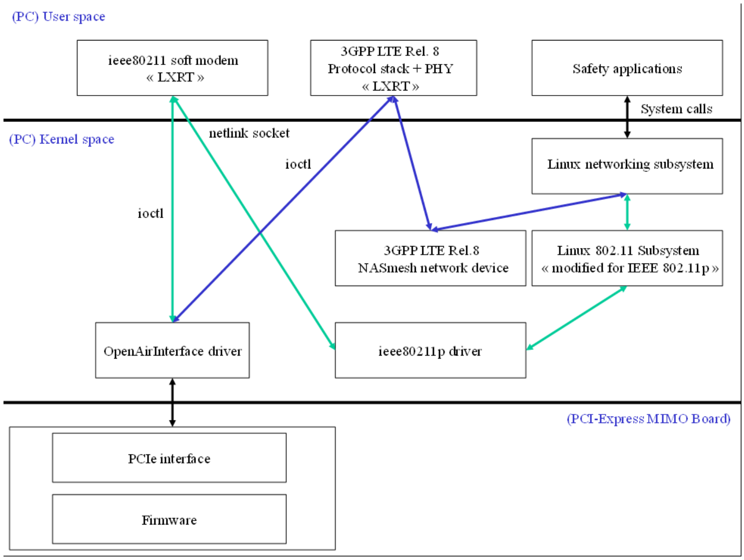


Fig. 2. OpenAirInterface software architecture

PCI-Express MIMO board through the same OpenAirInterface driver. Both protocol stacks are capable to work at the same time using different RF chains of the PCI-Express MIMO board.

4.1 IEEE 802.11p Protocol Stack

Following Figure 2, the SDR architecture of the IEEE 802.11p protocol stack is composed of three blocks. The upper block contains an extension for IEEE 802.11p of the Linux kernel 802.11 subsystem. This subsystem is composed of *nl80211*, a netlink configuration interface for user-space applications, *cfg80211* which is the Linux wireless configuration interface bridging user-space and drivers and *mac80211*, which offers a framework for driver developers writing soft-MAC wireless devices. The *mac80211* subsystem is the Linux stack for IEEE 802.11.

The second block is the *IEEE 802.11p driver*, which bridges the Linux 802.11 subsystem and the hardware. One major difference with standard architecture for 802.11 systems, is that the IEEE 802.11p driver does not link to the SDR chipset directly, but rather to a *soft-modem* via netlink sockets. We chose this architecture for flexibility in the development and configuration of the low layer functionalities of the IEEE 802.11p stack and to ease the reconfigurability of our radio. The *soft-modem* is the placeholder of all the functionalities of IEEE 802.11p physical layer. Being located out of the chipset, it is totally accessible and reconfigurable. The soft-modem is finally connected to the SDR board via an IOCTL link and a dedicated OpenAirInterface driver, which composes the last block.

4.2 3GPP LTE Protocol Stack

As for the IEEE 802.11p, the software architecture of the 3GPP LTE rel. 8/9 is composed of three blocks. The upper block is located in the Linux user-space and contains an open-source implementation of the 3GPP LTE rel-8/9 protocol stack, including PDCP, RLC, MAC and the upper PHY. It contains 3GPP LTE rel.8/9 compliant functionalities and a LTE soft-modem. Located in user-space, they are easily extensible to future LTE rel. 10 and beyond features. The second block is the LTE NASMesh device, which is located in the kernel space, due to real-time constraints for low PHY operations. The last block is the OpenAirInterface driver, which acts as interface with the Express-MIMO board.

4.3 SDR Express-MIMO Board

The prototype is based on an Agile Radio Front-end (RF) and ExpressMIMO SDR board. They provide a fully reconfiguration RF and baseband DSP, and include "developper-friendly" tools for real-time hard-modem development and validation. The Express-MIMO board is a FPGA platform on PCI-Express with up to 8 MIMO capacity and providing full SDR descriptions of three air interfaces: *LTE*, *802.11p*, and *DAB/DMB*.

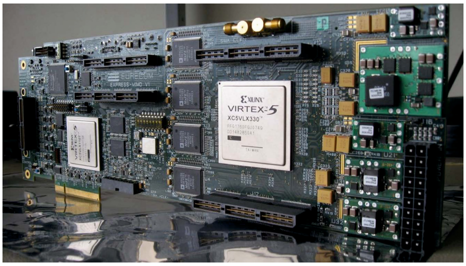


Fig. 3. Express-MIMO SDR Board

4.4 Experimentation

As a "proof-of-concept", the major goal within the scope of the project was the demonstration of communications with the defined scenarios. It opens further work on more detailed measurements and field-test characterization.

We performed in-situ tests on a dedicated IFSTTAR test-tracks dedicated to road-safety experimentation (Versailles-Satory, France). We equipped two vehicles with the previously described 3GPP LTE/IEEE 802.11p OBU prototype, including an on-board GPS (Figure 4b), and one RSU equipped only with the 3GPP LTE technology. As a test case, we evaluated a *contextual speed warning* application as well as a *ghost driver* applications as specified by the French SCOREF project [13], using V2I communication over 3GPP LTE for the former, and V2V over IEEE 802.11p for the latter. Figure 5 is a screen-shot of the contextual speed warning application, which also depicts the IFSTTAR test track.

(a) Integration of the SDR platform on a BMW X5 test vehicle

(b) Warning displayed on the vehicle driving on a wrong direction

Fig. 4. Illustration of the experimental test

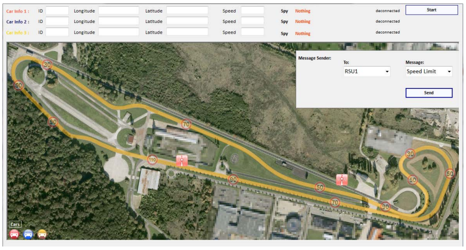


Fig. 5. Screenshot of the test-tracks in Versailles/Satory

On that test case, using the 3GPP LTE RSU, vehicles provides to the NH their instantaneous speed and in return the NH provides contextual speed limitations that are displayed on the car HMI.

5 Performance Study

In this performance study, we are interested in the communication efficiency of multi-technology cooperative system, and particularly the communication performances experienced by driver assistance applications. This cooperative system is able to balance the traffic generated by V2I communications and I2V Broadcast over two different wireless networks in order to increase the capacity of the system and provide better performances to the applications.

To model communications in the multi-technology cooperative system, we consider the IEEE 802.11 technology for V2V and V2I communications and the DVB-T technology for I2V broadcast. The performance study is based on a simulation framework, developed in the *OPNET Modeler* network simulator, that includes realistic vehicular mobility models [16] and SDR device modeling [15].

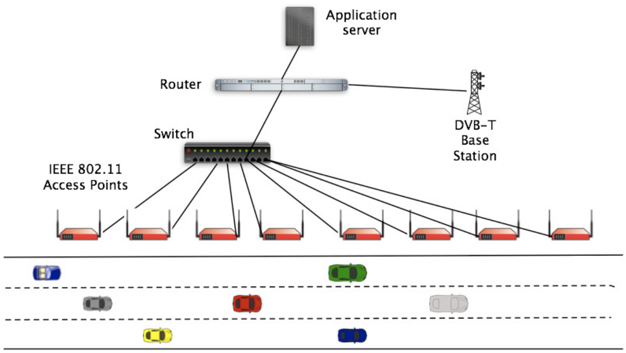


Fig. 6. Simulation scenario for the IEEE 802.11 and DVB-T based architecture

Evaluation Context. As an evaluation context, we consider a three ways straight road. Two hundred vehicles travel on this road. Vehicle mobility is based on the SuMo tool [16]. The Vehicles are equipped with SDR devices that manage IEEE 802.11 and DVB-T radios [15]. We model the network architectures conforming to the PLATA cooperative system specification 3. We deploy IEEE 802.11 access points along the roadside as RSUs and a DVB-T base station that cover all the road. These wireless networks are connected to a wired network that hosts the application server. Figure 6 shows this architecture. In this architecture, we deploy 35 access points over the roadside, one access point every 300 meters. The access points manage an 802.11 network with a throughput of 54 Mbps and a transmission power of 100 mW. This wireless network is connected to a wired network over a router. Vehicles are equipped with IEEE 802.11 hardwares with a transmission power of 20 mW. The DVB-T base station is broadcasting over the 800 MHz frequency on a 8 kHz bandwidth and a 32 Mbps throughput.

We modeled the driver assistance applications that define V2I communications (*V2I hazard warning* and *Static speed limit*) and I2V broadcasting (*Local weather condition* and *Dynamic speed limit*). In this configuration, data exchanges, defined by the *Static speed limit* and *V2I Incident Hazard Warning*, are carried through the RSUs. Broadcasts, defined by the *Local weather condition* and *Dynamic speed limit*, are sent through the broadcast base stations.

Simulation Results. The results show that the driver assistance application operates with good performances.

For *static speed limit* response time values vary between 0.11 s and 21.62 s as shown in table 2. Detailed results show the mean response time is about 1.026s with a standard deviation of 3.38. The disparity of response time values is due to the message exchange which consists of four 2-way-handshakes. The obtained values for response time are acceptable, even the highest, considering the nature of this application that aims to synchronize the speed database of the vehicles with the database of the management center. In addition, there is no packet loss for these applications as shown in table 1.

Table 1. Packet Loss

	Sent data (Pkt)	Received data (Pkt)	Packet loss rate
Local weather condition	109654	109654	0
Dynamic speed limit	14	14	0
V2I Incident hazard warning	37438.5	43959.66	0,14
Static speed limit	40536.833	40595.5	0.001

For *V2I Incident hazard warning*, results shows that values vary between 0.001 ms and 9.50s as shown in table 2. The mean response time is about 0.06s with a standard deviation of 0.93. In the case of a vehicle velocity equal to 50 km/h, a delay equal to 3.5 ms enables the information to reach a neighbor vehicle after a move of less than one meter. This value is more than sufficient to get the driver or the onboard system informed in time.

We look to the state of communications on IEEE 802.11 cells (cell load, data dropped and medium access delay) for both architecture. Results in table 3 show that the association of IEEE 802.11 with DVB-T, thanks to SDR in the vehicles, enables an increase of the available bandwidth in each cell.

Table 2. Response Time

	Mean (sec)	Min (sec)	Max (sec)	Stdev (sec)
Static Speed Limit	1.026	0.11	21.62	3.38
V2I Incident Hazard Warning	0.06	0.001	9.50	0.93

Table 3. IEEE 802.11 cell metrics

	Mean	Min	Max	Stdev
Media Access Delay (sec)	0.0002	0.0003	0.0001	0.00002
Cell load (bps)	128354.82	201554.44	0	25930.79
Dropped data (bps)	1055.91	0	26449.99	2142.32

These results open the way for the specification of additional applications that will make use of this bandwidth without disturbing the good operation of the system.

6 Conclusion

Vehicular cooperative systems rely on information obtained from immediate surroundings and longer-term forecasts and alerts controlled from a central infrastructure to offer assistance to drivers. The enhancement of these systems depends closely on wireless communication performances that enable cooperation between vehicles and between vehicles and infrastructure. The communication systems associated to cooperative systems have to support vehicle-to-vehicle

(V2V), vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (broadcasting) communications. The use of SDR represents an asset to ensure both a wide area coverage and the use of complementary applications to answer to various events on a travel. In this context, the PLATA-PROTON project proposes a multi-technology cooperative Advanced Driver Assistance System (ADAS), based on the integration of Software-Defined Radio (SDR) devices in vehicles. In this paper, we have shown the feasibility and interest of such ADAS through the development of a proof of concept that included SDR and driver assistance application prototypes and a performance study of communications performances, experienced by applications and based on network simulation tools.

Acknowledgment. The PROTON-PLATA research leading to these results has received funding as a Deufrako collaboration from the German *Federal Ministry of Economics and Technology* and the French *Agence Nationale de la Recherche*, with reference ANR-08-VTT-010-01.

References

1. ISO TC 204 working group 16 web page, `www.iso.org/iso/iso_technical_committee?commid=54706`
2. Car 2 Car communication consortium web site, `http://www.car-to-car.org`
3. IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments, IEEE Std 802.11p-2010, pp.1–51, doi: 10.1109/IEEESTD, 5514475 (July 15, 2010)
4. TC 204. ISo 21217:2010 : Intelligent transport systems "communications access for land mobiles (CALM)" architecture. Technical report, ISO/ TC 204 (2010)
5. Seely, J.: Altera Corporation: The programmable supercomputer on wheels (2005), `http://wwwdsp-fpga.com/articles/seely`
6. Alsliety, M.: General Motors: How does SDR fit the telematics model? In: SDR forum - Workshop on SW Portability (2004)
7. Vinel, A.: 3GPP LTE versus IEEE 802.11p/WAVE: Which technology is able to support cooperative vehicular safety applications? IEEE Wireless Comm. Letters 1(2) (April 2012)
8. CVIS project web site, `http://www.cvisproject.org`
9. EVITA project web site, `http://evita-project.org`
10. COOPERS project web site, `http://www.coopers-ip.eu/`
11. SAFESPOT project web site, `http://www.safespot-eu.org/`
12. PROTON-PLATA project web site, `http://www.proton-plata.fr`
13. SCOREF project web site, `http://www.scoref.fr`
14. OpenAirInterface web site, `http://www.openairinterface.org`
15. Kassab, M., Berbineau, M., Gransart, C., Wahl, M.: A software defined radio module for vehicular communications. In: OPNETWORK, Washington D.C. (2010)
16. Kaisser, F., Gransart, C., Berbineau, M.: Simulation of vanet scenarios with opnet and sumo. In: Vinel, A., Mehmood, R., Berbineau, M., Garcia, C.R., Huang, C.-M., Chilamkurti, N. (eds.) Nets4Trains 2012 and Nets4Cars 2012. LNCS, vol. 7266, pp. 103–112. Springer, Heidelberg (2012)

Performance of Inter-Vehicle Relay Network Based IR-UWB

Yamen Issa, Iyad Dayoub, and Abdelaziz Bensrhair

University of Lille Nord de France, F-59000 Lille,
IEMN, DOAE, F-59313 Valenciennes
INSA de Rouen, LITIS, BP 08 Avenue de l'Universite 76801
Saint-Etienne-du-Rouvray

Abstract. Inter-vehicle communication (IVC) have lately attracted a great interest in commercial and military fields to develop applications that can make transportation safer and more efficient. Ultra wideband (UWB) technology is a candidate as a license-free radio link for inter-vehicular transmission. Reliable data transmission between vehicles is difficult to accomplish due to the presence of noise and obstacles. The objective of this paper is to outline the effectiveness of cooperative transmission in vehicular network using impulse radio ultra wideband (IR-UWB) transmission. We consider a relay network applying decode-and-forward (DF) protocol and evaluate the bit error rate (BER) performance of direct sequence binary phase shift keying (DS-BPSK). Multiple access capability is set up using direct sequence spreading codes (DSSS) and we analyze the effect of multiple access interference (MAI). Our results demonstrate the diversity gains and BER performance improvement in two scenarios; perfect and imperfect source-relay channels.

Keywords: DS-BPSK, inter-vehicle communication (IVC), relay networks, UWB, virtual MIMO.

1 Introduction

Communication between vehicles are intended to improve efficiency and safety of road traffic. A dynamic exchange of sensor gathered data between nearby vehicles will improve the vehicle capabilities to calculate risk; sense threats and hazards; issue driver advisories or warnings; enhance airbag functionality; or activate braking system to avoid and mitigate crashes. It is expected that the user interest in this kind of applications to become a big market driver in a near future. To come with these applications and services, vehicles have to build up ad hoc wireless communication networks that can provide reliable connectivity, typically at data rates between 1 and 10 Mb/s [1]. Inter-vehicles wireless channel creates challenges for satisfying reliable data transmissions; this is due to the unreliable nature of wireless communication and the mobility of transmitter, receiver and scatterers around.

Among all the under development wireless technologies, (UWB) is one of the most competing for short-range future wireless communication. It uses very

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 96–106, 2013.

short pulses of duration of fraction of nanosecond to convey information, thus the spectrum of the emitted signals may spread over several GHz. UWB technology is a license-free frequency radio that can provide high data rate at low cost with low power consumption. It has owned a prior interest in the research during the last decade, in particular since the allocation of uwb spectrum by the federal communication commission (FCC)'s in February 2002 [2]. The FCC has defined UWB signal as signals having a fractional bandwidth of at least 20% or occupying at least 500 MHz in the allocated spectrum. The fractional bandwidth B_{Δ} is defined as

$$B_{\Delta} = \frac{2(f_H - f_L)}{f_H + f_L} \tag{1}$$

where f_H and f_L are the upper and lower frequencies of the -10 dB emission point, respectively.

The UWB transmission covers a large range of frequencies and underlay existing narrow-band communications. To avoid interference with existing narrow-band wireless systems operating in the bandwidth-limit regime, UWB works in the power-limit regime. For this reason, the maximum emission power for UWB transmission in the $3.1 - 10.6$ frequency range is limited to $-41.3\,dBm/MHz$ as specified in [3]. For any narrow-band channel, the UWB signal power is very small and so close to the noise floor; thus, the interference to any other existing technology like 802.11a and 3G mobile communication can be ignored [2].

Despite its promised services and functionalities, IVC based UWB have some drawbacks. Actually, the propagation characteristics of the channel between vehicles are significantly different from those of cellular channels. In addition, the low power spectral density (PSD) and the extremely short pulse duration of UWB pulses complicate the detection of the signal at the receiver. Due to the frequency-selective nature of UWB channels, the number of resolvable multipath components (MPC) at the receiver is very large, each one can involve attenuation, reflection, diffraction, and so on. These effects severely corrupt the radio transmission between vehicles.

In wireless communications, diversity techniques are the most practical and effective countermeasures to combat wireless channel effects in order to improve transmission reliability. Diversity provides receivers with redundant signal information. The well known forms of diversity are space, time and frequency. However, for cooperation diversity, space and time diversity schemes are mainly considered. In conventional wireless communications, spatial diversity technique requires implementing multiple transmit and/or receive antennas in order to provide independent versions of the transmitted signal to the receiver. The well known system models that use spatial diversity technique are: single-input multiple-output (SIMO), multiple-input single-output (MISO), and multiple-input multiple-output (MIMO) systems which combine transmit and receive diversity [4]. Recently, cooperative diversity was introduced as an alternative in multiuser networks, where users can cooperate; in such a way that they form a virtual antenna array (virtual MIMO system) [5] - [7]. Our objective in this paper is to introduce cooperative diversity to IVC using UWB transmission technology. In this paper, a relay network is composed of three or more vehicles:

source (S), one or more relay (R), and destination (D). The considered channel model is the UWB IEEE 802.15.4a for outdoor environments. We evaluate the BER performance of DS-BPSK considering MAI for two scenarios; errors occurs at the relay nodes and free-errors at the relay nodes.

User cooperation has been first introduced in narrow-band transmissions. In [5, 6], the advantages of cooperation were shown for two active users using conventional code-division multiple-access (CDMA). The most common cooperation protocols are amplify and forward (AF), decode and forward (DF), and compress and forward (CF). In the first, the relay amplifies the received signal of its partner, while in the second, it decodes, re-encodes, and retransmits an estimate of the partner message. For the CF protocol, the relay forwards a quantized and compressed version of the received signal to the destination. Employing DF protocol, the authors in [7] investigated the error probability performance of cooperative diversity for asynchronous CDMA system in frequency selective slow fading environment. Vehicular Ad hoc relay wireless networks are studied in [8], it is shown that the motion of vehicles on a highway can contribute to successful message delivery. Cooperative diversity was extended to the context of UWB systems in [10] - [13]. Employing DF and AF protocols, the authors in [10, 11] present the construction of space-time (ST) codes suitable for IR-UWB. In [13], the authors analyzed the bit error probability of time hopping UWB relay networks under IEEE 802.15.4a channel models.

In our work, we consider vehicular relay network with IEEE 802.15.4a outdoor environment standard as the channel model between vehicles. We show that, the contribution of a relay in the vehicular source-destination transmission significantly enhances the BER performance as well as the reliability of detection. In the presence of MAI, the overall system performance is significantly degraded and is dominated by the source-relay S-R link, especially, when the source-destination S-D link exhibits poor channel conditions.

The rest of the paper is organized as follow. In section 2, we describes the proposed DS-BPSK UWB system for the vehicular relay network including channel model, signal model, and receiver structure. The results are discussed in section 3 and finally conclusions are given in section 4.

2 System Description

2.1 Inter-Vehicle Communication

In near future, vehicles will be equipped with wireless communication equipment that allow them to exchange messages with each other and also with a roadside infrastructure. A number of applications and services well be available icluding safety application, driver assistance, on board internet surfing, multimedia application and so on [14]. In vehicular communication, nearby vehicles will be able to communicate with each other without the need to backbone infrastructure or pre-installed base stations network. There are several proposals for data transmission between nodes. IEEE 802.11p is an international standard intended for both vehicle-to-vihicle (VTV) and vehicle-to-infrastructure (VTI) in the licensed band of 5.9 GHz (5.85-5.925 GHz). It defines enhancements to the popular WiFi standard (IEEE 802.11) required to support intelligent wireless transportation applications [15].

Based on the unlicensed IR-UWB system, we propose another alternative for wireless access in vehicular environments (WAVE). We use DSSS as a multiple access technique for the network nodes; in that several nodes can share the same frequency band. We consider multiple vehicle transmitters signaling through the wireless access medium using IR-UWB signals, where each transmitter employs DS-BPSK modulation. The considered transmission scheme is illustrated in Fig. 1. The relay R assists the S-D transmission using DF protocol in order to improve the detection reliability at the destination. In our scenario, two time slots with equal signaling intervals are used to accomplish the transmission. In the first time slot, the source broadcasts its signal to the destination and the relay through S-D, S-R links. In the second time slot, the relay R decodes the source signal, re-encodes and retransmits an estimate of the source signal to the destination through R-D link, whereas the source is silent.

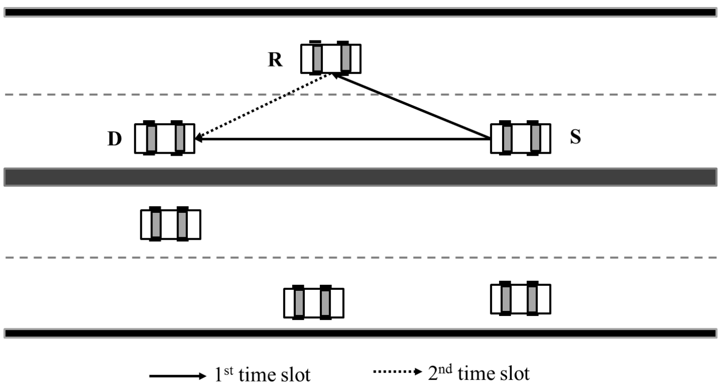


Fig. 1. Relay inter-vehicle transmission scheme

2.2 Transmitter Structure

Since a typical UWB system is carrierless, the choice of the pulse shape should be convenient for physical implementation and for mathematical modeling. Thus the second derivation of the Gaussian pulse, called Gaussian doublet, is typically considered in the literature [13,16]. It is assumed to include the differential effects of the transmitter and receiver antennas and is given by

$$p(t) = -\frac{2A}{T_p^2}(1 - \frac{2t^2}{T_p^2})exp(-\frac{t}{T_p})^2 \tag{2}$$

where A and T_p determine the amplitude and the width of the pulse, respectively. In a DS-BPSK, users employ DSSS to ensure multiple access capability, the equiprobable binary bit stream $b_m^s \in \{1, -1\}_{m=-\infty}^{\infty}$ of the source is BPSK modulated by the pulse $p(t)$. That is, $b_m^s = 1$ is mapped to $p(t)$ and $b_m^s = -1$ is mapped to $-p(t)$ with $P\{b_m^s = +1\} = P\{b_m^s = -1\} = 1/2$. The transmitted signal generated by the source is then given by

$$s_s(t) = \sqrt{\frac{E_s}{N_c}} \sum_{m=-\infty}^{\infty} \sum_{n=0}^{N_c-1} b_m^s c_n^s p(t - t_s - mT_b - nT_c) \quad (3)$$

where E_s is the bit energy of the source signal, t_s is the relative transmission time difference. Without loss of generality, we suppose $t_s = 0$. N_c is the number of chip per bit, also called the processing gain. $T_b = N_c T_c$, where T_c is the chip duration and T_b is the bit duration. $\sum_{n=0}^{N_c-1} c_n^s$ is the source spreading sequence with $c_n^s \in \{1, -1\}$. To avoid pulses overlapping and reduce the effect of intersymbol and intrasymbol interference; the chip interval is chosen to be comparable to the delay spread of the typical IEEE 802.15.4a channel ($T_c > T_p + max(T_{m1}, T_{m2}, T_{m3})$) [13, 18], where T_{mi} is the maximum excess delay spread of link i.

2.3 Channel Model

For IR-UWB transmission, the channel does not follow narrow-band scenarios. Based on a modified Saleh-Valenzuela model [9], the IEEE 802.15.4a channel modeling subgroup has provided channel models for UWB systems where the multipath components arrive in a random number of clusters; within each a random number of rays is considered [3]. In our system we use IEEE 802.15.4a channel model for outdoor environment to model the channel between two vehicles. This model was extracted based on measurements that cover a range from 5-17m in the frequency band 3-6 GHz [3]. The discrete channel model is expressed as

$$h(t) = \tilde{\beta} \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} \alpha_{k,l} exp(j\varphi_{k,l}) \delta(t - T_l - \tau_{k,l}) \quad (4)$$

where $\alpha_{k,l}$ is the k path gain in the l cluster, $\tilde{\beta}$ models the pathloss and shadowing. Here the effects of transmit and receive antennas are ignored. The phase $\varphi_{k,l}$ is considered as a uniformly distributed random variable in the range [0, 2π]. In what follows, we use $\alpha_{k,l}$ instead of $\tilde{\beta}\alpha_{k,l} exp(j\varphi_{k,l})$ for notational simplicity. L is the number of clusters assumed to be Poisson-distributed, K is the number of paths per cluster and its value depends on the required dynamic range of the model, T_l and $\tau_{k,l}$ represent the delay of the lth cluster and the delay of the kth multipath component (MPC) relative to the lth cluster arrival time, respectively. The details of the joint probabilistic model of these parameters are described in [3]. Through this study, we assume that the minimum path resolution time is equal to the sample width T_{samp}. Hence, multipath components arrive at some integer multiple of the minimum path resolution time.

2.4 Signal Model

Under multiple access scenario, the simultaneous transmission of multiple signals is a typical source of interference in communication. When using the DSSS multiple access technique, the data stream is combined with a high-speed digital code where each data bit is mapped into the user's unique pseudo noise (PN) code. This code is known only to the transmitter and the corresponding receivers. In a frequency-selective multipath fading UWB channel, MAI and self multipath interference are caused by multipath dispersion. The general received signal model is given by

$$r_{xy}(t) = \sqrt{\frac{E_x}{N_c}} \sum_{l=0}^{L_{xy}-1} \sum_{k=0}^{K_{xy}-1} \sum_{m=-\infty}^{\infty} \sum_{n=0}^{N_c-1} \alpha_{k,l}^{xy} b_m^x c_n^x p(t - mT_b - nT_c - T_l^{xy} - \tau_{k,l}^{xy}) + \sum_{i=1}^{N_I} s_i(t) \otimes h_{iy}(t) + n_{xy}(t). \quad (5)$$

where $\otimes$ denotes convolution, x and y determine the channel link, i.e. during the first time slot, the received signal at the destination through the S-D link is obtained by replacing x, y by s, d, respectively; and the received signal at the relay through the S-R link is obtained by replacing x, y by s, r, respectively. In the second time slot, the received signal at destination through the R-D link is obtained by replacing x, y by r, d, respectively. N_I is the number of simultaneous asynchronous interfering users in each time slot with transmitted signals $s_i(t)$; $h_{iy}(t)$ are the complex channel coefficients of the i-y links. The first, second and third terms are respectively, the received signal, the MAI and the complex additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2 = N_0/2$.

2.5 Receiver Structure

For the well demodulation of the received UWB signal, the template signal and the target path of the received signal must be perfectly aligned at each of the Rake correlators. The aim is to determine the relative delay of the received signal tap with respect to the template signal. Channel estimation, in UWB system, is a critical issue due to the precise channel characteristics. There are several used algorithms to estimate the channel taps with different complexity degree. One of the most used algorithms is the data-aided (DA) approach that is based on using a number of known pilot symbols called (training sequence) in the beginning of the data packet, while the rest of the packet is interpreted based on the acquired channel characteristics. Two suboptimal estimation algorithms with reduced complexity are examined in some papers. The first is denoted by Successive Channel (SC) estimation and the second is the sliding window (SW) algorithm [17]. The sliding window algorithm cross-correlates an identified pilot with the received pilot sequence and calculates gains and delays for all channel taps at the same time. The SC algorithm enhances the performance of SW algorithm by employing the iterative notion. The algorithm begins by using the sliding window to uncover the coefficients of the strongest tap, then the delayed version of the transmitted signal corresponding to the estimated tap is subtracted from the received sequence and the algorithm is repeated by a number of iterations equal to the number of taps to be estimated. In this work, we consider that the receiver has a prior knowledge of the channel state information (CSI) and we use Rake receiver to exploit the multipath diversity. The Rake receiver consists of a bank of correlators, or matched filters. Each filter is matched to a delayed version of the template waveform to match the delay of a particular path. However, due to the large number of resolvable paths and for complexity constraints, we use partial Rake receiver (P-Rake) which processes only a subset of the total number of resolvable paths [13, 18]. The P-Rake receiver used in our simulations captures the first $K^{'}$ rays from the first cluster of the targeted UWB channel link. The general template waveforms used in the correlators of the P-rake receivers are given by

$$v_{xy}^{k^{'}}(t) = \alpha_{k^{'},0}^{xy^{*}} \sum_{n=0}^{N_c-1} c_n^x p(t - nT_c - T_0^{xy} - \tau_{k^{'},0}^{xy}) \tag{6}$$

where $0 \leq k^{'} \leq K^{'} - 1$, x and y determine the channel link as explained previously. It is assumed that the channel impulse response remains stable over several consecutive data bits durations. This assumption is justified by considering that the symbol duration is of the order of tens of nanoseconds, and the coherent time of inter-vehicles UWB wireless channel is on the order of milliseconds [19]. Without loss of generality, we limit our interest to the detection of the information bit b_0^s in the observation interval $[0, T_b]$ and drop the symbol index 0. Hence, the decision variables at the output of the Rake receivers at destination in the first and the second time slots are given respectively as

$$\hat{Z}_{sd} = \sum_{k=0}^{K^{'}} \int_0^{T_b} r_{sd}(t) v_{sd}^k(t) dt \tag{7}$$

$$\hat{Z}_{rd} = \sum_{k=0}^{K^{'}} \int_0^{T_b} r_{rd}(t) v_{rd}^k(t) dt. \tag{8}$$

At the destination, the receiver combines the two decision variables with a factor λ which controls the near-far effect. This effect consists of the pathloss difference and depends on the transmitted power and propagation distances. This type of detector has been refereed to as the λ-MRC [6]. The final decision variable is then given by

$$\hat{Z} = \hat{Z}_{sd} + \lambda \hat{Z}_{rd}. \tag{9}$$

The receiver then makes hard decision following

$$Z = \begin{cases} 0 & if\, \Re\left[\hat{Z}\right] > 0 \\ 1 & if\, \Re\left[\hat{Z}\right] \leq 0 \end{cases}. \tag{10}$$

Note that the receiver extracts the energy from a limited number of paths and that decision errors may occur due to the self multipath interference, multiple access interference, and noise which render the receiver non-optimal.

3 Simulation Results

In this section, we present simulation results for the BER performance of the DS-BPSK inter-vehicle relay network based IR-UWB system. Monte Carlo simulations have been carried out with and without MAI for perfect and imperfect S-R channels. We use Gaussian doublet UWB impulse with a duration of $0.7ns$. The line of sight (LOS) UWB IEEE 802.15.4a outdoor channel model is used to model the channel between each pair of vehicles. The UWB channels have been chosen randomly from a set of 500 channel realizations and the BER performance has been averaged over at least 200 channels. The channel bandwidth is set to 3 GHz. The pathloss is considered as function of distance and its frequency dependency is ignored.

A five-finger Rake receiver is employed with a combination factor $\lambda = 1$. Without loss of generality, we assume that all users' signals travel the same distance to arrive the relay and/or the destination receivers; which yields equal received power from all nodes. The emitted power is at most equal to the FCC maximum emission power limit. In the simulations, each frame consists of 10 bits; where each bit is spread into a set of 7-chip PN code. Each chip is presented by a single pulse. The chip duration T_c equals $20ns$ and thus the bit rate equals $7.14Mbits/s$. Higher data rates can be obtained by making the chip duration shorter. This may leads to inter-chip interference when the chip intervals becomes smaller than the delay spread of the UWB channel impulse response. In all results, we show the BER performance versus the SNR under UWB IEEE802.15.4a outdoor LOS channel model.

Fig. 2 shows the BER performance of the system without interference considering perfect S-R channels. We note that in the absence of MAI, a significant BER performance improvement is obtained by the use of one relay node. This is due to the diversity gain provided by the relay node. When using two vehicular nodes to relay the source message, an additional performance improvement is obtained for high SNR (above 4 dB). In fact, the two relay nodes use the same timeslot to relay the source message in asynchronous transmission.

Indeed, when cooperating with two relay, the full diversity gain can be obtained by assigning different time slot for each of them or by using MAI suppression algorithm.

Fig. 3 illustrates the BER performance of the system with the existance of interfering signal considering perfect S-R channels. In comparison to the former case of Fig. 2, we note that the overall system performance is significantly degraded when MAI is present even when using relay cooperation (e.g. at SNR $= 4$ dB, the BER probability equal to 2.738×10^{-2} and 2.946×10^{-3} for the direct transmisson without relay and transmission when cooperating with one relay in fig. 2, respectively, wherease the BER probability equal to 6.853×10^{-2} and 1.558×10^{-2} for the direct transmisson without relay and transmission when cooperating with one relay in fig. 3, respectively). This is due to the fact that the received signals at the destination from the source and the relay nodes suffer from MAI interference. But, the cooperation with two relays in this case brings more performance improvement. Actually, if a pulse is being corrupted by

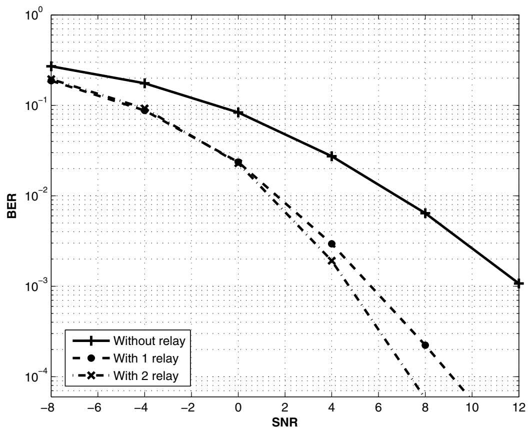


Fig. 2. BER of DS-BPSK UWB system for inter-vehicle relay network and a LOS channel. Perfect S-R channel, without MAI.

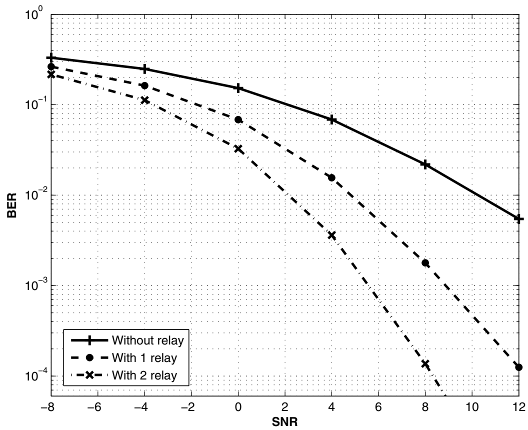


Fig. 3. BER of DS-BPSK UWB system for inter-vehicle relay network and a LOS channel. Perfect S-R channel, with MAI.

interference, it is highly possible that only one interfering pulse collides with the pulse position of the desired user.

Considering errors in the S-R channel, the BER performance of the system is compared in Fig. 4 in with and without MAI. We note that in the absence of MAI, the BER performance improvement obtained by the use of one relay node is not as much as in the perfect S-R channel case. Indeed, the errors occur at the relay node well propagate to the destination. As can be seen from these results, the system can achieve the full diversity gain when considering perfect

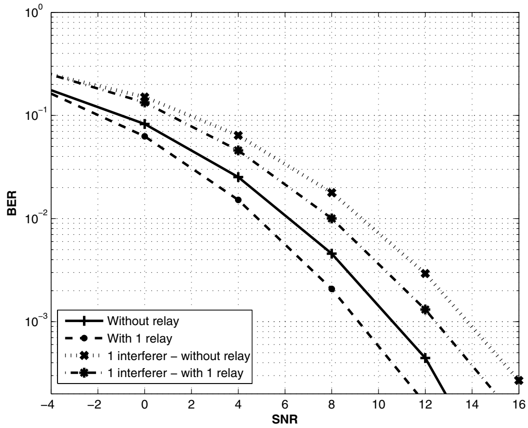


Fig. 4. BER of DS-BPSK UWB system for inter-vehicle relay network and a LOS channel. Imperfect S-R channel.

S-R channels. However, the full diversity gain cannot be obtained when considering imperfect S-R channels due to the errors propagation.

4 Conclusions

We examined the performance of DS-BPSK IR-UWB system for inter-vehicle relay network. Simulation results have demonstrated the benefits of relay cooperation in different cases. We have showen that a performance improvement can be obtained due to the diversity gain provided by the relay nodes. However the overall system performance can be degraded in the presence of MAI. We have analyzed the BER performance of the proposed system in both cases, perfect and imperfect S-R channels. Our results show that the full diversity gain can be achieved when no errors occur at the relay nodes. However, the diversity gain of the system is degraded when considering imperfect S-R channels which leads to errors propagation.

References

1. Molisch, A., Tufvesson, F., Karedal, J., Mecklenbrauker, C.: A survey on vehicle-to-vehicle propagation channels. IEEE Wireless Communications 16(6), 12–22 (2009)
2. Federal Communications Commission (FCC): Revision of Part 15 of the communication rules regarding ultra-wideband transmission systems, First report and order. ET Docket 98-153, FCC 02-48, Washington (2002)
3. Molisch, A.F., Balakrishnan, K., Cassioli, D., Chong, C.C., Emami, S., Fort, A., Karedal, J., Kunisch, J., Schantz, H., Schuster, U., et al.: IEEE 802.15. 4a channel model-final report. IEEE 802 15(4), 0662 (2004)
4. Gibson, J.D.: Mobile communications handbook, vol. 45. CRC Press (2012)

5. Sendonaris, A., Erkip, E., Aazhang, B.: User cooperation diversity - Part I System description. IEEE Trans. Communications 51(11), 1927–1938 (2003)
6. Sendonaris, A., Erkip, E., Aazhang, B.: User cooperation diversity - Part II Aspects and performance analysis. IEEE Trans. Communications 51(11), 1939–1948 (2003)
7. Eid, A., Hamouda, W., Dayoub, I.: Cooperative Diversity in asynchronous CDMA over Frequency-Selective Fading Channel. IET Proceedings-Communications 5(5), 683–692 (2011)
8. Chen, Z.D., Kung, H.T., Vlah, D.: Ad hoc relay wireless networks over moving vehicles on highways. In: Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing, pp. 247–250 (2001)
9. Saleh, A., Valenzuela, R.A.: A statistical model for indoor multipath propagation. IEEE J. Selected Areas Comm. 5, 128–137, 247–250 (1987, 2001)
10. Abou-Rjeily, C., Daniele, N., Belfiore, J.-C.: On the decode-and-forward cooperative diversity with coherent and non-coherent UWB systems. In: Proc. IEEE Int. Conf. UWB, pp. 247–250, 435–440 (2001, 2006)
11. Abou-Rjeily, C., Daniele, N., Belfiore, J.C.: On the amplify-and-forward cooperative diversity with time-hopping ultra-wideband communications. IEEE Trans. Communications 56(4), 630–641 (2008)
12. Maichalernnukul, K., Zheng, F., Kaiser, T.: UWB MIMO Cooperative Relay Systems: BER Analysis and Relay Regions. In: Proc. 44th Annual Conference on Information Sciences and Systems, CISS, Princeton University, NJ, USA (2010)
13. Yazdi, Z.Z., Kenari, M.N., Aazhang, B.: Bit Error probability Analysis of UWB Communications with a Relay Node. IEEE Trans. Communications 9(2) (2010)
14. Festag, A., Hessler, A., Baldessari, R., Le, L., Zhang, W., Westhoff, D.: Vehicle-to-Vehicle and Road-Side sensor communication for enhanced road safety. In: 15th World Congress on Intelligent Transport Systems (2008)
15. IEEE P802.11p/D9.0: Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment: Wireless Access in Vehicular Environments (WAVE), Draft 9.0 (September 2009)
16. Li, Q., Rusch, L.A.: Multiuser Detection for DS-CDMA UWB in the Home Environment. IEEE J. Sel. Areas Commun. 20(9), 1701–1711 (2002)
17. D'Amico, A.A., Mengali, U., Morelli, M.: Multipath channel estimation for the uplink of a DS-CDMA system. In: IEEE International Conference on Communications (ICC 2002), vol. 1, pp. 16–20 (2002)
18. Choi, J.D., Stark, W.E.: Performance of Ultra-Wideband Communications With Suboptimal Receivers in Multipath Channels. IEEE J. Sel. Areas Commun. 20(9), 1758–1766 (2002)
19. Okonkwo, U.A.K.C., Hashim, S.Z.M., Ngah, R., Nanyan, N.F., Rahman, T.A.: Time-scale domain characterization of nonstationary wideband vehicle-to-vehicle propagation channel. In: 2010 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE), pp. 1–6 (2010)

The Effects of Increasing Antenna Arrays and Spatial Correlation on Loading Algorithm for Closed-Loop MIMO Vehicle-to-Infrastructure Communications

Imade Fahd Eddine Fatani[1,3], Mohamed Gharbi[1], François-Xavier Coudoux[1], Marion Berbineau[2], Patrick Corlay[1], and Marc Gazalet[1]

[1] Université Lille Nord de France, F-59000, UMR 8520, IEMN, Department OAE, UVHC, F-59313, Valenciennes
{firstname.name}@univ-valenciennes.fr
[2] Université Lille Nord de France, F-59000, IFFSTAR, LEOST, F-59666, Villeneuve d'Ascq
marion.berbineau@iffstar.fr
[3] Université Hassan Premier, ENSA, TCI, BP 77, 25000, Khouribga, Morocco
imade.fatani@uh1.ac.ma

Abstract. In this paper, an efficient bit-loading algorithm in the context of SVD-aided MIMO transmission for vehicle-to-infrastructure (V2I) communications is presented. This algorithm, which exploits the intrinsic properties of the eigen-channels resulting from SVD, tries to minimize the error probability of the system, assuming that the error rates per subchannel are equal, without considering the well-known Gap approximation. Furthermore, we examine the effects of increasing the number of antennas and spatial correlation when using a bit loading policy. Several simulation results for different transmission scenarios are presented and discussed. We show that when using a high number of antennas, bit and power allocation scheme is reduced to an adaptive power allocation scheme only which is due to the respect for the hypothesis on error rate, and for a reduced number of antennas, the effect of correlation becomes more pronounced while the allocation of ressources becomes sensitive to the variation of the number of antennas.

Keywords: Vehicle-to-Infrastructure (V2I) communications, Closed-Loop MIMO System, Bit and Power loading, Spatial correlation, Diversity.

1 Introduction

The use of multiple antennas at both the transmit and receive ends for modern radio access systems have drawn much attention during the last decade [10,27,26]. This is mainly due to its high spatial degrees of freedom through which enhanced capacity or reliability or a desired combination of the two aspects is made possible [28]. The so-called Multiple Input Multiple Output (MIMO)

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 107–120, 2013.

technology is nowadays an essential part in next generation wireless telecommunication standards, such as IEEE802.11n, WiMAX, LTE and DVB, hence, it constitutes one of the most promising technologies in vehicular communications [21]. For instance, recent DVB-NGH (Next Generation Handheld) experiments have shown that the use of MIMO technology leads to improved robustness of services delivered to moving devices in cars or trains [14]. However, MIMO system performances depends essentially on the channel characteristics and show serious degradations in the presence of spatial correlation as in the case of tunnel environments [3,5].

Performance improvement is allowed thanks to precoding schemes which rely on the channel's knowledge at the transmitter [12]. The problem of precoding designs for multiple antenna systems has been widely studied in the literature; most of the proposed schemes are based on the Singular Value Decomposition (SVD) of the channel reponse and propose to distribute the available power on eigenmodes under a variety of criteria [24,22,6]. Recently, a limited feedback precoding scheme that achieves orthogonality between transmitted symbols was proposed simplifying the design of ML detection with no need to implement SVD. Such precoding scheme named Orthogonalized Spatial Multiplexing (OSM) [20], relies on a phase rotation operation at the transmitter. To enhance the OSM performance, the authors in [18,19] proposed an adaptive power allocation scheme for the OSM approach that will maximize the minimal Euclidean distance d_{min} between the received symbols.

MIMO performances can be further significantly enhanced by employing adaptive bits and power loading algorithms also called adaptive modulation (AM) algorithms, i.e., optimization of the signal constellation size and power allocation over the different eigenmodes according to their correponding signal-to-noise ratio (SNR) values [29,11,8]. Several loading schemes for distributing power and bits have been proposed in the literature [4,2,15,23]; they are based on the well known waterfilling approach, which enables reliable high bit rate transmission.

In this paper, we focus on optimal bit and power loading design for closed-loop MIMO transmission. We study and analyze the effect of number of antennas as well as the impact of spatial correlation on the bit and power distribution. We examine in particular the effect of the number of antennas at the receiver side by considering the uplink case, as it should easier to install a large number of antennas at the infrastructure side than at the vehicle. Hence, in order to fully benefit from MIMO technology advantages, the equipment of the terminals with a large number of antennas have received much attention: for example, the new LTE communication standard allows for up to 8 antennas at the base station. The obtained results remain general and have been used in our case for vehicular-to-infrastructure communications dedicated to CCTV video surveillance. In our research work, indeed, the applicative scenario we address consists in transmitting video compressed streams over a wireless link from a vehicle like a subway or a bus (equipped with on board camera) to a distant control centre. For further details, please see [8,9].

The remainder of this paper is organized as follows: Section II describes the MIMO system model in detail and then the considered loading algorithm is described in Section III. Simulation results are presented and analyzed for several transmission scenarios in Section IV. Finally, Section V concludes the paper.

2 SVD Aided MIMO Communication System

This section describes the signal model for SVD aided MIMO systems. Assume a flat fading MIMO channel consisting of N_t transmit antennas and N_r receive antennas. The received signal vector can be expressed in linear form as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{1\times N_t}$ is the transmitted signal vector, $\mathbf{n}$ is a complex Gaussian noise vector with covariance matrix $\sigma_n^2 \mathbf{I}_{N_t}$ with $\mathbf{I}_{N_t}$ is the identity matrix of size $N_t \times N_t$ and $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the random channel matrix.

If the transmitter has accurate knowledge of the channel response, either by means of a feedback channel from the receiver to the transmitter or by applying reciprocity between the uplink and the downlink propagation channels, several independent channels can be highlighted [13]. To do that, SVD (Singular Value Decomposition) of the MIMO channel is used, i.e., $\mathbf{H} = \mathbf{U}\varSigma\mathbf{V}^H$ where $\mathbf{U} \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{V} \in \mathbb{C}^{N_r \times N_r}$ are complex unitary matrices, and $\varSigma \in \mathbb{C}^{N_r \times N_t}$ is a diagonal matrix whose diagonal elements are the singular values of $\mathbf{H}$. Hence, the multiplexed channel can be decoupled into a series of r parallel channels, with r denoting the rank of the channel matrix. The performances of each independent SISO channel depends on the singular value amplitude, i.e. $\left\{\sqrt{\lambda_i}\right\}_{1\le i \le r}$. Therefore, the input-output relationship becomes:

$$\tilde{\mathbf{y}} = \varSigma\,\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{2}$$

Here, $\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y},\ \tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x},\ \tilde{\mathbf{n}} = \mathbf{U}^H\mathbf{n}$.

The general block diagram for the SVD-based MIMO system is depicted in Fig. 1. Thus the overall system's performance can be enhanced by adapting the transmitted signals ,i.e., the input bits as well as the available transmit power are optimally distributed between the eigenmodes according to the distribution policy defined by the proposed adaptive modulation (AM) algorithm. At the receiver side, after applying a post-compensation filter in order to accomplish a SVD decomposition of the channel matrix, the received signal is demodulated.

The optimal allocation strategy such that the average mutual information across the channel is maximized is the traditional water filling technique [27]. Assuming the noise level power to be the same on each eigenmode, the water filling solution allocates the highest power to the channel with the highest gain such that :

$$p_i = \left(L - \frac{1}{\rho\lambda_i}\right)^+,\ i = 1,\ldots,r \tag{3}$$

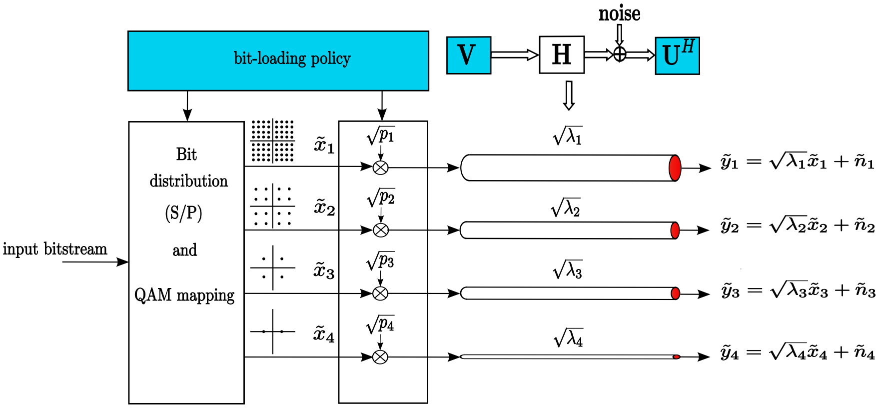


Fig. 1. Bit-loading algorithm for orthogonalized MIMO radio channel

Here $\rho = \frac{E_S}{\sigma_n^2}$ is the Signal-to-Noise Ratio (**SNR**) and L is a constant chosen so as to satisfy the power constraint $\sum_{i=1}^{r} p_i = 1$ where $(x)^+$ denotes $\max\{0, x\}$.

3 Optimal Discrete Loading Algorithm for MIMO System

In this section, an efficient discrete loading algorithm is investigated and explained in detail. This strategy is an enhanced version of the algorithm proposed in [2], which allows optimal bits allocation as well as power allocation for a given set of independent sub-channels [7]. We consider practical two-dimensional QAM-coded scheme. Therefore the number of bits b_i (in bits/symbol) that can be transmitted over the ith subchannel as a function of the amount of allocated power p_i is given by the following expression :

$$b_i = \log_2 (1 + p_i g_i), \tag{4}$$

where g_i is the subchannel gain-to-noise ratio which is given by $g_i = \frac{\lambda_i}{\sigma_n^2 \Gamma_i}$, Γ_i is the SNR gap [4]. The SNR gap is computable according to the *gap-approximation* hypothesis [15,4,2,23] which can be expressed as follow :

$$\Gamma = \frac{1}{3} \cdot \left[Q^{-1} \left(\frac{\mathrm{SER}_i}{4} \right) \right]^2, \tag{5}$$

where SER_i is the target Symbol-Error-Rate (SER) of the ith subchannel, for convinience the error rates per subchannel are assumed to be equal, Q^{-1} is the inverse of the error function of the gaussian statistics defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt, \tag{6}$$

However, the gap approximation is not valid for a reduced number of bits. Hence, in order the bit and power allocation to be optimal, we will not consider in the rest of this work the conventional SNR gap formula (eq. 5), but we use instead the following exact expression:

$$\Gamma_{i,b_i} = K\left(b_i\right)^{-1} \times \left[Q^{-1}\left(\frac{\mathrm{SER}_i}{A\left(b_i\right)}\right)\right]^2 . \tag{7}$$

The two parameters $K\left(b_i\right)$ and $A\left(b_i\right)$ are defined as follows:

$$A\left(b_i\right) = \begin{cases} 4\left(1-\frac{1}{2^{0.5b_i}}\right) & \text{, for rectangular constellation } (b_i \text{ even}), \\ 4\left(1-\frac{1}{2^{0.5(b_i+1)}}\right), & \text{for cross constellation } (b_i \text{ odd}), \\ 2 & , \text{ for } b_i = 3, \\ 1 & , \text{ for } b_i = 1. \end{cases}$$

and

$$K\left(b_i\right) = \begin{cases} \frac{7}{3}, & b_i = 3, \\ 2, & b_i = 1, \\ 3, & \text{otherwise.} \end{cases}$$

Fig. 2 shows the evolution of the gap according to the SER_i for different constellation sizes. From this figure, it can be denoted that for values of $b \geq 4$, the gap becomes insensitive to variations of constellation size. In contrast, for 2-QAM, 4-QAM and 8-QAM constellations, the value of the gap strongly depends on the modulation order. Using the exact value ensures that the bits and power allocation is optimal.

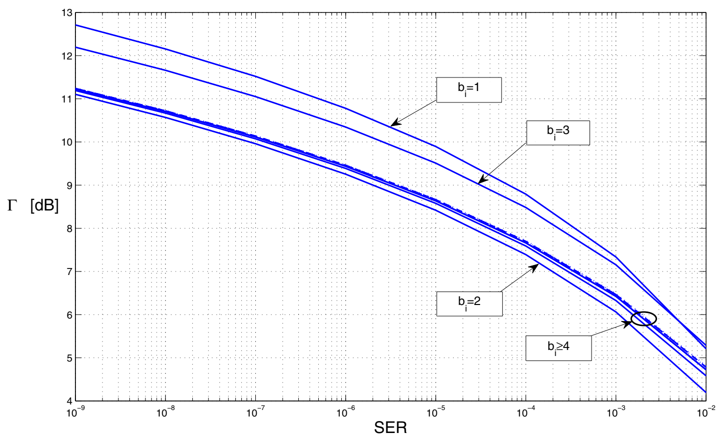


Fig. 2. Evolution of the gap Γ according to the SER value : curves merge when $b_i \geq 4$

Finally, the loading problem consists in distributing the rate and power in order to minimize the error probability, this constrained optimization problem can be described as

$$\begin{aligned} & \underset{b_i, p_i}{\text{minimize}}\ \text{SER} \\ \text{subject to} & \\ & \textstyle\sum_{i=1}^{r} b_i = R_T \text{ and } \sum_{i=1}^{r} p_i = \text{P}_T \end{aligned} \tag{8}$$

where R_T is the total supported data rate and P_T represents the available transmit power.

4 Simulation Results

In this section, we analyze the effect of increasing number of antennas as well as spatial correlation on the bit and power loading strategy. Consider a flat fading $(N_t \times N_r)$ MIMO system and $\mathbf{H}$ the channel complex matrix. When considering the uncorrelated full rank Rayleigh fading channel, the mean gain of all eigenvalues of $\mathbf{H}$ is bounded below and above according to the following relationship [1,16]:

$$\left(\sqrt{N_t} - \sqrt{N_r}\right)^2 < E\left[\lambda_i\right] < \left(\sqrt{N_t} + \sqrt{N_r}\right)^2 \tag{9}$$

From these inequalities, we see that increasing the number of transmit (resp. receive) antennas for a fixed number of receive (resp. transmit) antennas has as consequence the increase of the subchannels gains; all the eigenvalues will tend to approach $\max\{N_t, N_r\}$. In the special case, where $N_r \to \infty$ we have $\lambda_i \to \infty$. In this way, with reference to the relation 3, the adaptive allocation of bits and power will converge to an equal allocation.

As first step, we investigate the effect of increasing the number of receive antennas in the case of uncorrelated Rayleigh channel. Then we consider the case of spatially correlated channels which is known to be more realistic and representative of V2V and V2I propagation conditions. Although the algorithm assigns discrete integer bits for each of eigenchannels, the results presented below are real numbers obtained in averaging over several simulation.

4.1 Uncorrelated Rayleigh Channel

Figures 3 and 4 present the evolution of ressource allocation, i.e, the available power and bits, according to the number of receive antennas in the case of uncorrelated Rayleigh channel. It can be seen that for the channel with lower N_r (i.e., lower eigenvalues) the algorithm optimizes both data rate en power allocations. For the higher order MIMO systems (i.e., high number of antennas), the bits and power allocation is reduced to power allocation, while the number of bits per sub-channel remains the same (from 20 antennas, in the case of a system with two transmit antennas, and from 30 for a system with four receive antennas). On the other hand, for a lower number of antennas, the loading policy becomes more sensitive to the variation of the number of antennas; especially, the distribution of power varies significantly. This is because the SER constraint is ensured by selecting the appropriate power distribution factor.

As it can be seen, the two curves of the power allocation factors for ($N_t = 2$,N_r) systems present an intersection mainly due to the SER constraint; for lower N_r the second eigenvalue λ_2 is so poor that the optimization algorithm assigns most of the rate and power to the first eigenchannel. As well as N_r increases (i.e. λ_2 increases and tends to have the same value than λ_1), the second subchannel becomes able to transport more information so that b_2 and p_2 increase so, given that $\lambda_1 \geq \lambda_2$ is always verifyed, for an appropriate value of N_r the algorithm assigns more power to the second subchannel compared to the first one to ensure the same target error rate. We can drawn the same conclusions in 4 for the system with $N_r = 4$. It is interesting to note that in general the knowledge of the channel state at the transmitter will give significant difference compared with the channel unknown, if $N_t > N_r$. This is true, since high order of transmitter array gain will be exploited in this case. In the Figures 3 and 4, it can be observed that also for $N_r > N_t$, the difference between knowing and not knowing the channel at the transmitter is significant to ensure the target error rate, this is due to the adaptive power allocation however the modulation scheme (data rate) per eigenmode is the same.

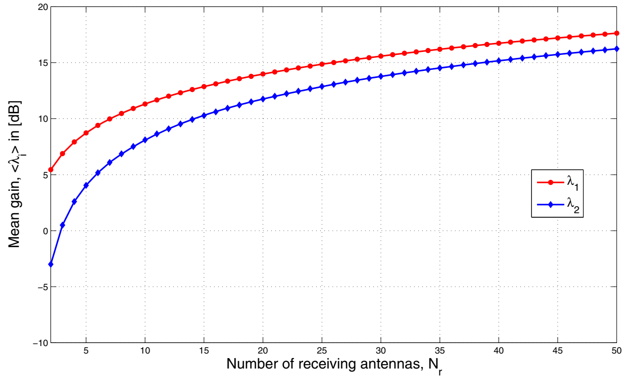
4.2 Spatially Correlated Channel

As previously mentioned, the bits and power allocation is also known to be sensitive to variation in the degree of spatial correlation of the channel. In order to take into account spatial correlation characteristic to realistic confined propagation environments, we consider in this work the well-known Kronecker radio channel model with different values of the correlation coefficient ρ [25]. This amounts to modeling $\mathbf{H}$ as

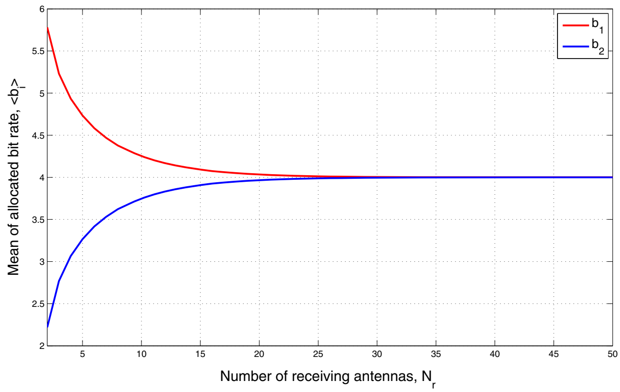
$$\mathbf{H} = \mathbf{\Sigma}_r^{1/2}\mathbf{H}_w\mathbf{\Sigma}_t^{1/2} \tag{10}$$

where $\mathbf{H}_w \in \mathbb{C}^{N_r \times N_t}$ is the uncorrelated MIMO channel matrix. The matrices $\mathbf{\Sigma}_t \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{\Sigma}_r \in \mathbb{C}^{N_r \times N_r}$ are the correlation matrices respectively at transmitter and receiver side. The Kronecker model is popular in the literature despite some existing limitations and has been shown experimentally to model satisfactorily the statistics of many physical channels (see, for example, [17]). Consequently, we chose to use in a first approach the theoretical Kronecher model in our simulations; these preliminary results will be futher confronted against the results obtained using real channel measurements.

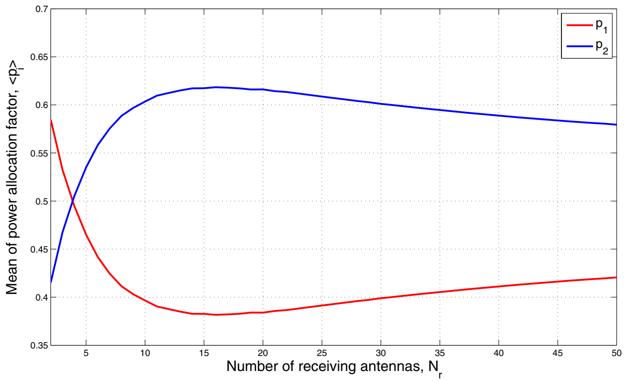
Figure 5 illustrates the evolution of the ergodic channel capacity versus the correlation coefficient ρ for various values of N_t and N_r. The capacity curves highlight a serious decrease of the capacity in the presence of spatial correlation, specially when $\rho \geq 0.6$.



(a) Mean power gain of the two eigen-channels for MIMO (N_t = 2,N_r) systems

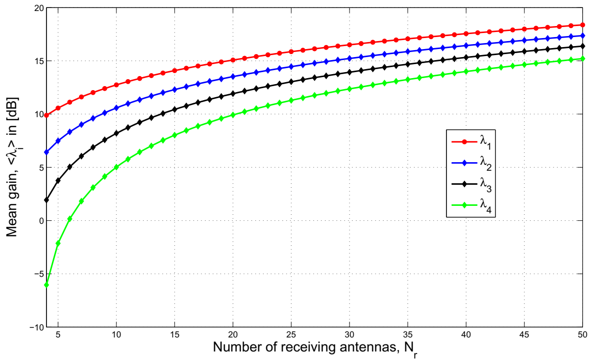


(b) Mean allocated bit rate for MIMO ($N_t = 2$,N_r) systems

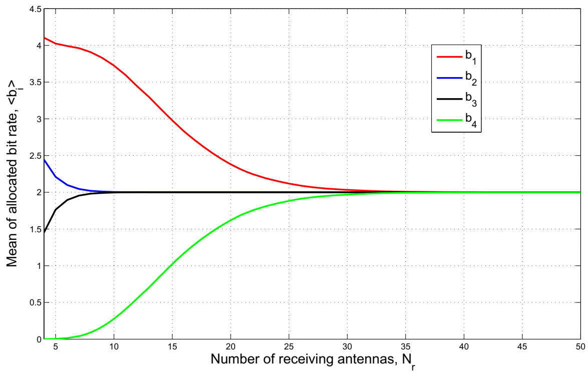


(c) Mean power allocation factor over the two eigen-channels for MIMO ($N_t = 2$,N_r) systems

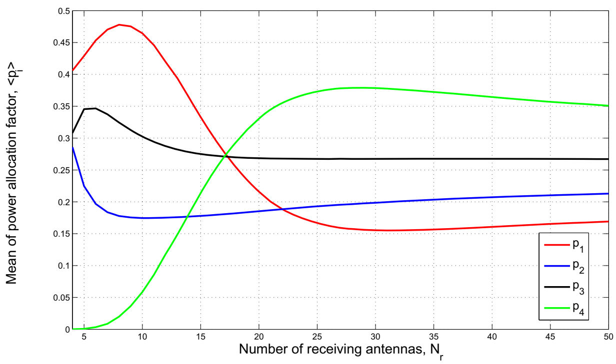
Fig. 3. Effect of increasing N_r on bit and power allocation for (N_t = 2,N_r) with $R_T = 8\,\text{bps/Hz}$ and $\frac{E_b}{N_0} = 10\,\text{dB}$



(a) Mean power gain of the forth eigen-channels for MIMO ($N_t = 4$,N_r) systems



(b) Mean allocated bit rate for MIMO ($N_t = 4$,N_r) systems



(c) Mean power allocation factor for MIMO ($N_t = 4$,N_r) systems

Fig. 4. Effect of increasing N_r on bit and power allocation for ($N_t = 4$,N_r) with $R_T = 8$ bps/Hz and $\frac{E_b}{N_0} = 10$ dB

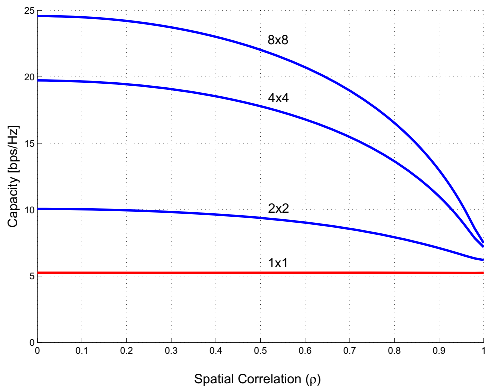


Fig. 5. MIMO system capacity vs. spatial correlation

As explained previously, the algorithm proposed here tries to minimize the error probability without increasing the overall throughput of the system, i.e., it consists to find the efficient way of distributing the rate and power among the eigenmodes so as to minimize the error probability. Thus, when the subchannels gain increase due to increasing the number of antennas, the target error rate decrease because the algorithm performances are strongly depend on the sub-channels gain. Table 1 shows an example of the obtained error rate for various number of receiving antennas for $\rho = 0.5$ and $\frac{E_b}{N_0} = 10\,\text{dB}$.

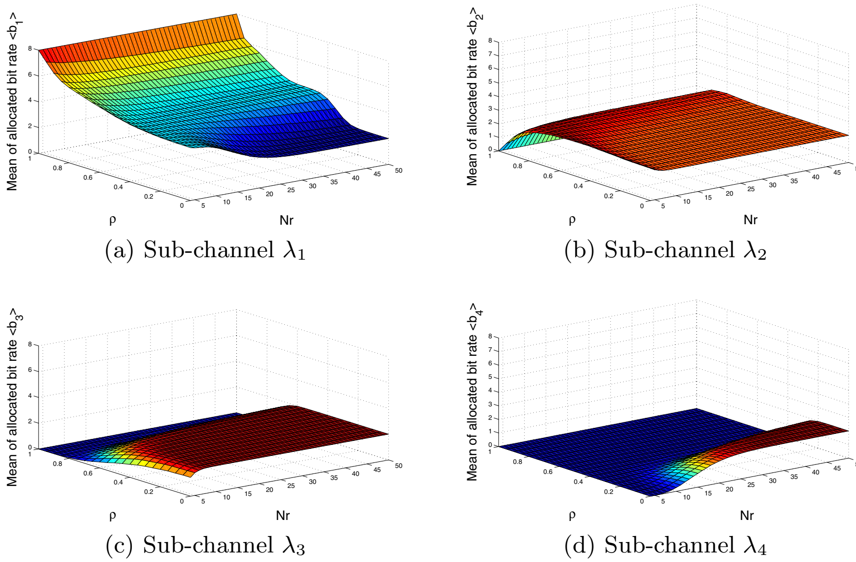
Table 1. SER versus number of receive antennas for $\rho = 0.5$ ($N_t = 2$, $R_T = 8\,\text{bps/Hz}$)

N_r	2	15	20	30	50
SER	$1.0000e-015$	$9.8000e-016$	$7.8182e-016$	$5.1495e-017$	$1.0000e-045$

We present in figures 6 and 7 the evolution of ressource allocation as a function of both number of antennas and spatial correlation. The transmission system includes four transmit antennas with a total data rate of 8 bps/Hz with $\frac{E_b}{N_0} = 10\,\text{dB}$. Our objective here is to investigate the effects of number of antennas and channel correlation.

From these three-dimensional presentations, we find that the allocation of available power is much more sensitive to the variation of ρ and N_r. Indeed, the degree of spatial correlation and the number of receive antennas result in the increased sub-channels gains. Furthermore, the condition of an identical SER imposed at first has a major impact on the power distribution.

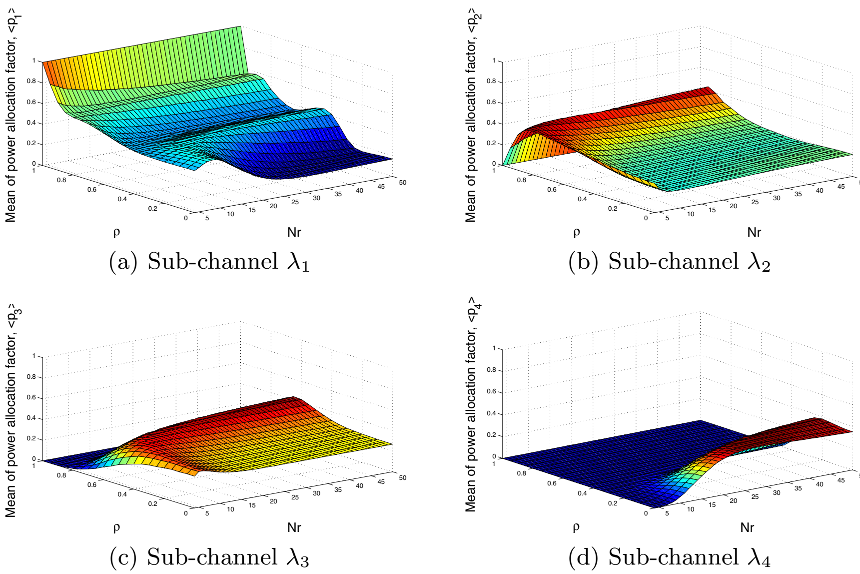
As N_r increase, substantial decreases in both mean of allocated bit rate and mean of power allocation factor are observed for the first eigenmode (λ_1). In fact, for low-order size arrays the maximum information is transmitted over the first eigenmode. This mean that when the quality of the other of eigenmodes



(a) Sub-channel λ_1 (b) Sub-channel λ_2

(c) Sub-channel λ_3 (d) Sub-channel λ_4

Fig. 6. Evolution of bit allocation as a function of both number of receive antennas and spatial correlation ($N_t = 4$, $R_T = 8\,\text{bps/Hz}$)



(a) Sub-channel λ_1 (b) Sub-channel λ_2

(c) Sub-channel λ_3 (d) Sub-channel λ_4

Fig. 7. Evolution of power allocation as a function of both number of receive antennas and spatial correlation ρ ($N_t = 4$, $R_T = 8\,\text{bps/Hz}$)

increases, we will have an information transfering form λ_1 to the other eigenchannels so that the variation of the allocation ressources over this eigenmode becomes more pronounced. On the other hand, the allocated ressources over the weak eigenchannel (λ_4) increase only when N_r increase, this because for high-order size arrays, all eigenmodes have a significant channel gain and for reduced number of antennas λ_4 is so poor that is neglected in the transmission even if $\rho = 0$.

The bits allocation is generally more sensitive to the spatial correlation than to the number of antennas variation in particular fo sub-channels λ_2 and λ_3 in the treated configuration.

5 Conclusion

In this paper, we focus on the loading scheme for SVD-aided MIMO transmissions, aiming to increase system performances where the bit rate and the available power are distributed in order to minimize the error probability assuming that the error rates per eigenchannel are equal. We investigate the channel correlation and number of antennas effects when using a bit loading policy. Numerical results show that if the number of receive antennas increases, the bits and power allocation is reduced to power allocation and that the bits allocation is more sensitive to the spatial correlation than to the number of antennas. These results may help to develop innovative MIMO systems with large number of antennas, thanks to rapid progress in RF design and miniaturization.

Further work will concern several aspects. The case of more realistic channel conditions obtained experimentally thanks to ray tracing will be considered in order to overcome the shortcomings of the Kronecker model. Moreover, the analysis of system behavior using imperfect channel state estimation should be also investigated. In this investigation, we will consider the impact of errors on the return link on the overall performances of the MIMO transmission system.

Acknowledgment. This work has been performed in the framework of the regional CISIT (Campus International Sécurité et Intermodalité des Transports) program, task 3.2 : intelligent infrastructure and Vehicle-MIMO-OFDM for V2V and V2I communications. The authors would like to thank the Nord Pas de Calais Region for financial support of the project.

References

1. Andersen, J.: Array gain and capacity for known random channels with multiple element arrays at both ends. IEEE Journal on Selected Areas in Communications 18(11), 2172–2178 (2000)
2. Campello, J.: Optimal discrete bit loading for multicarrier modulation systems. In: Proc. Int. Symp. Inform. Theory, p. 193 (August 1998)
3. Chizhik, D., Foschini, G.J., Valenzuela, R.: Capacity of multi element transmit and received antennas: Correlation and keyholes. Electron. Lett. 36(13), 1099–1100 (2000)

4. Chow, P., Cioffi, J., Bingham, J.: A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels. IEEE Trans. Commun. 43(2), 773–775 (1995)
5. Chuah, C.N., Tse, D.N.C., Kahn, J.K., Valenzuela, R.A.: Capacity scaling in MIMO wireless systems under correlated fading. IEEE Transactions on Information Theory 48(3), 637–650 (2002)
6. Collin, L., Berder, O., Rostaing, P., Burel, G.: Optimal minimum distance-based precoder for MIMO spatial multiplexing systems. IEEE Trans. Signal Process. 52(3), 617–627 (2004)
7. Fatani, I.F.E.: Contribution to the study of joint source-channel optimization for video transmission over mimo wireless channel- application to video surveillance in public transportation systems. PhD dissertation, Valenciennes University, France (October 2010)
8. Fatani, I.F.E., Cocheril, Y., Coudoux, F.X., Zwingelstein-Colin, M., Gharbi, M., Nsiala-Nzéza, C., Berbineau, M.: An improved adaptive modulation scheme for closed-loop mimo transmission in tunnel environments. In: 11th IEEE International Conference on Inteligent Transportation System & Telecommunications, pp. 162–167 (August 2011)
9. Fatani, I.F.E., Cocheril, Y., Nsiala-Nzéza, C., Vrigneau, B., Berbineau, M., Coudoux, F.X.: Train-to-wayside video communications in tunnels using multiple description coding and roi combined with orthogonal spatial multiplexing schemes for multiple input multiple output techniques. Transportation Research, Part C: Emerging Technologies 25(2), 168–180 (2012)
10. Foschini, G.: Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. Bell Labs Technical Journal 1(2), 41–59 (1996)
11. Getu, B.N., Andersen, J.: Mimo systems in random uncorrelated, correlated and deterministic radio channels. Wireless Personal Communications: An International Journal 30(1), 27–61 (2004)
12. Getu, B.N., Andersen, J.: On the performance of link adaptation techniques in mimo systems. Wireless Personal Communications 42(4), 543–561 (2007)
13. Goldsmith, A.: Wireless Communications. Cambridge University Press, Standford University (2005)
14. Heuberger, A.: Beyond dvb-x2. In: Int. Symp. on BMSB, pp. 27–29. IEEE, Yonsei (2012)
15. Hughes-Hartogs, D.: Ensemble modem structure for imperfect transmission media. U.S. Patents no 4,679,227, July 1987, 4,731,816, March 1988, and 4,833,706 (1989)
16. James, A.T.: Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Stat. 35(2), 475–501 (1964)
17. Kermoal, J., Schumacher, L., Pedersen, K., Mogensen, P., Frederiksen, F.: A stochastic MIMO radio channel model with experimental validation. IEEE J. Sel. Areas Commun. 20(6), 1211–1226 (2002)
18. Kim, Y., Lee, H., Park, S., Lee, I.: Optimal precoding for orthogonalized spatial multiplexing in closed-loop mimo systems. IEEE J. Sel. Areas Commun. 26(8), 1556–1566 (2008)
19. Kim, Y., Lee, H., Park, S., Lee, I.: Optimal precoding for orthogonalized spatial multiplexing in mimo wireless systems. In: Proc. IEEE International Conference on Communications, ICC 2008, pp. 4739–4743 (May 2008)
20. Kim, Y., Park, S., Lee, I.: Orthogonalized spatial multiplexing for closed-loop mimo systems. IEEE Trans. Commun. 55(5), 1044–1052 (2007)

21. Moniak, G., Berbineau, M., Pardonche, J.: Robust and high data rate transmissions for security between a bus and a control center. In: VTC 2004-Fall Proceedings, vol. 2, pp. 1377–1381. IEEE, Los Angeles (2004)
22. Palomar, D.P., Cioffi, J.M., Lagunas, M.A.: Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization. IEEE Trans. Signal Process. 51(9), 2381–2401 (2003)
23. Papandreou, N., Antonakopoulos, T.: A new computationally efficient discrete bit-loading algorithm for DMT applications. IEEE Trans. Commun. 53(5), 785–789 (2005)
24. Sampath, H., Stoica, P., Paulraj, A.: Generalized linear precoder and decoder design for mimo channels using the weighted mmse criterion. IEEE Trans. Commun. 49(12), 2198–2206 (2001)
25. Shiu, D.S., Foschini, G.J., Gans, M.J., Kahn, J.M.: Fading correlation and its effect on the capacity of multielement antenna systems. IEEE Trans. Commun. 48(3), 502–513 (2000)
26. Tarokh, V., Seshardi, N., Calderbank, A.: Space-time codes for high data rate wireless communications: Performance criterion and code construction. IEEE Transactions on Information Theory 44(2), 744–765 (1998)
27. Telatar, I.: Capacity of multi-antenna gaussian channels. European Transactions on Telecommunications 10(6), 585–595 (1999)
28. Tse, D., Viswanath, P., Zheng, L.: Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. IEEE Transactions on Information Theory 49(5), 1073–1096 (2003)
29. Zhou, Z., Vucetic, B., Dohler, M., Li, Y.: Mimo systems with adaptive modulation. IEEE Trans. Veh. Technol. 54(5), 1828–1842 (2005)

Increased Communication Reliability for Delay-Sensitive Platooning Applications on Top of IEEE 802.11p

Magnus Jonsson, Kristina Kunert, and Annette Böhm

CERES – Centre of Research on Embedded Systems, Halmstad University, Halmstad, Sweden
{magnus.jonsson,kristina.kunert,annette.bohm}@hh.se

Abstract. Cooperative driving in platooning applications has received much attention lately due to its potential to lower fuel consumption and improve safety and efficiency on our roads. However, the recently adopted standard for vehicular communication, IEEE 802.11p, fails to support the level of reliability and real-time properties required by highly safety-critical applications. In this paper, we propose a communication and real-time analysis framework over a dedicated frequency channel for platoon applications and show that our retransmission scheme is able to decrease the message error rate of control data exchange within a platoon of moderate size by several orders of magnitude while still guaranteeing that all delay bounds are met. Even for long platoons with up to seventeen members the message error rate is significantly reduced by retransmitting erroneous packets without jeopardizing the timely delivery of regular data traffic.

Keywords: Platooning, real-time communication, vehicular communication, retransmission scheme, real-time scheduling analysis.

1 Introduction

Intelligent Transport Systems (ITS) in general and cooperative driving in particular have received much attention from both researchers and media recently. Enabled by sensing and communication capabilities, as well as recent international standardization efforts, the introduction of cooperative driving shows great potential for increased safety and efficiency on our roads. The basic building blocks of a variety of ITS application are the exchange of status information [1] within a concerned group of vehicles, making it possible for the driver or the system itself to e.g. adjust to sudden changes in the current traffic situation. These changes usually need to be performed within a strict deadline and with high demands on reliability.

An application area where real-time and reliability demands are particularly obvious is platooning. It has been shown that a considerable reduction in fuel consumption can be achieved for trucks driving in platoons with reduced vehicle-to-vehicle spacing [2]. Additionally, the controlled and thereby more predictable behaviour of trucks on a highway increases the overall road safety not only for platoon members but for all surrounding road users. To enable a distance between platoon vehicles of 10 meters or

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 121–135, 2013.

less control data has to be exchanged within the platoon at a regular basis and with a high and predictable success probability.

The recently approved IEEE 802.11p standard for inter-vehicle communication [3] defines two message types, periodic status updates and event-triggered warning messages, that are allowed to share the dedicated ITS control channel in the 5.9 GHz ITS frequency band. To access that shared communication resource the standard employs CSMA/CA, a random access Medium Access Control (MAC) method, which by definition does not support the reliable and timely exchange of safety-critical and highly deadline-dependent control data needed for platoon control applications. Furthermore, there is no support for retransmissions to tackle the potentially high message error rate found in highly mobile wireless communication networks. Measurements of the packet reception ratio for different scenarios using IEEE 802.11p are reported in [4].

In this paper we therefore propose two enhancements used on top of the unaltered IEEE 802.11p standard in form of real-time functionality containing a deterministic, polling-based MAC approach as well as a transport layer retransmission scheme over a specific service channel dedicated to inter-platoon communication only. Both concepts are explained in detail in Chapter 2 and 3 in the context of a platooning scenario. Our approach improves the packet error rate (PER) in the platooning use case by several orders of magnitude while still maintaining a reasonable platoon size and is based on earlier work on infrastructure supported communication [5]. The inadequacy of IEEE 802.11p to obtain the level of reliability required in platoon control applications was recognized by [6] where infrared is suggested as a complementary communication technology for increased reliability. This adds additional hardware costs that our IEEE 802.11p-based approach manages to avoid. A polling-based MAC scheme similar to ours was presented in [7]. The lack of a real-time schedulability test, however, fails to provide timing guarantees. The authors of [8] show how to obtain strict priorities in 802.11p, but collisions among equal-priority packets can still appear. STDMA has been suggested as a promising alternative to the 802.11p MAC method [9], but even though the medium access delay is bounded, there still remains a possibility of collision. Studies on how to increase quality of service through retransmissions have been presented before. The combination of retransmissions with deadline-dependent traffic, however, is studied less thoroughly. [10] presents a retransmission scheme where packets only are retransmitted in case their deadline has not already passed, but the lack of any queuing analysis might imply the fact that the worst case delay of packets cannot be upper-bounded. [11] provides guarantees for timely treatment of real-time traffic, but lacks timing details and uses an analysis method shown to result in lower network utilization as the analysis method used in our approach [12]. For more general information and surveys on vehicular communication, see, e.g., [13-15].

2 Protocol Details

Imagine a platoon of trucks driving at reduced inter-vehicle distance at relatively high speed. To maintain that distance, every platoon member has to be well informed of the behaviour of its surrounding platoon neighbours, especially in case sudden changes in speed call for immediate actions performed by the vehicles' control systems. It can be assumed that one particular vehicle within the platoon assumes a central role in collecting the necessary status information, processes it and spreads the necessary control data to the platoon members. This role often fall the first vehicle in the platoon, the platoon leader. Furthermore, our proposed MAC method is based on a polling scheme administered by one node in the system in a master-slave fashion. As the length of a platoon easily can exceed the transmission radius of the platoon, we assign that role to a vehicle in the middle of the platoon, ensuring the best platoon coverage.

As the ITS control channel only allows the aforementioned IEEE 802.11p message to be transmitted, the polling messages our MAC approach relies on are not accepted on that channel. Due to the frequent and highly deadline-dependent data traffic needed to make a platoon function properly and safely it is reasonable to expect that a second transceiver will be available in platoon members, dedicated to platoon safety-related data exchange. Besides the ITS control channel there are several service channel available that can be used for that purpose. By using a service channel, we are not restricted to a specific message type and, at the same time, we are able to use the entire bandwidth of that service channel for platooning purposes while still listening to and participating in the data exchange with other road traffic users over the ITS control channel.

2.1 Medium Access Control

In the MAC protocol defined for e.g. IEEE 802.11a/b/g used for Wireless Local Area Networks (WLAN) the bandwidth is divided into Contention-Based Phases (CBPs) and Collision-Free Phases (CFP). The former is based on CSMA/CA random access where nodes compete for channel access and packet collisions are possible and, in the case of heavy channel utilization, even likely. The latter uses a polling-based scheme granting a specific node undivided access to the channel for a given period of time. As a slimmed version of the original IEEE 802.11 protocol suit, 802.11p does not offer collision-free channel access, despite the fact that many of its applications are safety-critical and put very high requirements on the timeliness and reliability. As proposed in previous work [16] [17], we reintroduce the possibility to use the CFP and propose a polling scheme administered by the master vehicle by placing a real-time layer on top of the IEEE 802.11p MAC layer. Admission control comprises a real-time schedulability test that insures the timely treatment of safety-critical data.

Time is divided into superframes (see Fig. 1) consisting of a CFP and a CBP, including a beacon frame, where channel access during the CBP is simply handled by IEEE 802.11p MAC in its random access fashion. Depending on the types of data traffic present in the network, the CBP can either be exploited by non-real-time packets in case those are present or reduced to zero (except for the beacon frame), freeing all bandwidth for safety-critical real-time traffic handled during the CFP. The beacon is used for synchronization purpose, but can also be used for, e.g., special short-message services [18] coordinated by the master. The transport layer hands down packets in Earliest Deadline First (EDF) [19] order and only at a point in time when the polling mechanism can ensure that the medium is available, thereby ensuring a deterministic treatment of the packet.

The master utilizes a polling scheme to grant the other vehicles access to the medium during the CFP. Knowing real-time traffic demands in the network, the polling scheme schedules traffic according to EDF. In case data traffic is to be sent from a vehicle to the master, the master will poll the vehicle and allow it exclusive access to the medium. In case the data traffic flow is from the master to any other vehicle, the master will allocate time for its own transmission, including an ACK answer from the receiving vehicle. The polling protocol enables deterministic medium access during the CFP for real-time communication necessary for the platooning application.

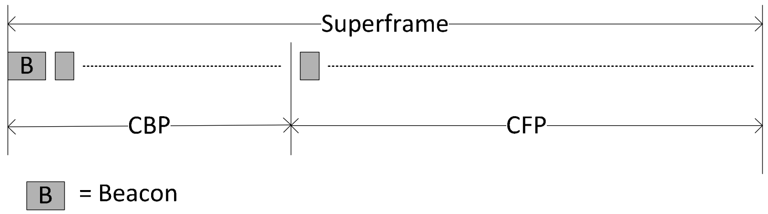


Fig. 1. Superframe structure

2.2 Real-Time Transport Layer

In order to make it possible to decide upon retransmissions on a message level, the transport layer added to the protocol stack implements EDF scheduling of both ordinary transmissions and retransmissions, and a real-time ARQ (Automatic Repeat Request) scheme. The ARQ scheme allows us to retransmit erroneous packets before their deadline expires and without negative interference with other real-time data traffic. This scheme has been proven successful by the authors in other types of networks (industrial networks [20] [21] and infrastructure-based vehicular networks [5]) and is here applied to the specific requirements of a platooning scenario.

Real-time data traffic in the network is modelled as periodic real-time channels (RTCs) called τ_i (with $i = 1, 2,\ldots$) and defined as $\tau_i = \{S_i, R_i, P_i, L_i, D_i\}$, where S_i is the sending node, R_i the receiving node, P_i the period, L_i the packet length, and D_i the relative transport layer deadline of the traffic flow. In order to have time for retransmissions before the end of the deadline, D_i is divided into two components: the deadline for the ordinary transmission ($D_{ord,i}$), and the deadline for retransmission(s) ($D_{retr,i}$). Retransmissions are modelled as retransmission real-time channels (ReRTCs) and defined by $\tau_{re,i} = \{ P_{re,i}, L_{re,i}, D_{re,i}\}$. The period is a predetermined system parameter indicating the minimum interval possible between two retransmissions on the ReRTC and by that upper-bounding the bandwidth dedicated to retransmissions. The length specified is the maximum packet length in order to accommodate any retransmission packet, and the deadline is the maximum time allowed for the current retransmission.

Retransmission of a data packet from the master is triggered if no ACK is received directly following the data packet, i.e., before time out. In case the master does not receive a packet it polled for, it will simply schedule a retransmission of its polling packet (no link layer acknowledgements or retransmissions from the underlying standard are used). This mechanism leads to that merely the master has to make the decisions about retransmissions. Regular vehicular nodes only need packet queues for ordinary transmission and a possibility to store packets in case they need to be retransmitted. However, all packets can be discarded after their deadline. ACKs are sent directly upon the reception of a correct data packet (after the predefined highest-priority arbitration interframe spacing (AIFS). Erroneous packets are detected at the MAC layer and not handed to the transport layer at all.

The master vehicle will schedule all ordinary transmissions and retransmissions in the form of EDF queues of polling and data packets. No packet can be sent before knowing how long it would take to transmit (T_{wait}), as no transmission can be initiated if T_{wait} is longer than the time remaining in the current CFP. These times are different for polling ($T_{wait,poll,i}$) and data packets ($T_{wait\_data,i}$) and are calculated as follows:

$$T_{wait\_poll,i} = T_{procM} + T_{AIFS} + T_{poll} + T_{prop} + T_{procS} + T_{AIFS} + T_{data,i} + T_{prop} + T_{procM\_CRC} + T_{margin} \quad (1)$$

$$T_{wait\_data,i} = T_{procM} + T_{AIFS} + T_{data,i} + T_{prop} + T_{procS\_CRC} + T_{AIFS} + T_{ACK} + T_{prop} + T_{procM} + T_{margin} \quad (2)$$

where

- T_{procM} = processing time at the master before the start of the packet transmission
- T_{AIFS} = Highest-priority arbitration interframe spacing (always needed)
- $T_{poll} = L_{poll} / r$ = transmission time of a polling packet, where L_{poll} denotes the length of a polling packet and r the bit rate
- T_{prop} = maximum propagation delay
- T_{procS} = processing time at the slave after reception of packet
- $T_{data,i} = L_i / r$ = transmission delay of a data packet of RTC i
- T_{procM_CRC} = processing time at the master for error checking in data packet
- T_{margin} = time margin compensating for, e.g., other protocol delays
- T_{procS_CRC} = processing time at the slave for error checking in data packet
- $T_{ACK} = L_{ACK} / r$ = transmission time of ACK, L_{ACK} being the ACK packet length

Having sent a data or polling packet, the master waits the calculated waiting time before moving on to the next packet in the EDF queue. However, at the end of the waiting time, the master's transport layer will first check for a data or ACK packet handed up from the MAC layer. In case the received data packet was erroneous or no ACK arrived, the possibility of retransmissions has to be investigated. A packet can only be retransmitted a predefined number of times ($N_{attempt}$). Moreover, a ReRTC has to be available, i.e., it must not have been used during the length of one retransmission period P_{re}. Only if both checks are positive, a retransmission can be scheduled in the EDF queue, otherwise the packet has to be discarded.

3 Real-Time Analysis

In the following, a timing and feasibility analysis is described, adapting the analysis framework introduced in [22] [23] to work with this platooning single-hop star topology network with a central master node as defined above.

3.1 Timing Parameters

We are assuming superframes consisting of one CBP of length T_{CBP} and one CFP of length T_{CFP}. Data traffic with real-time demands will only be sent during the CFP. Considering a worst-case situation, not the whole CFP can be utilized for transmissions, as the last interval left during this phase might be too short to schedule the next packet from the queue. The length of this interval is called blocking time and denoted T_b. In our worst case analysis, T_b corresponds to the time it would take to transmit the longest packet pertaining to any of the traffic flows in the network, i.e.,

$$T_b = \max\left\{\max_i\left\{T_{wait\_poll,i}\right\}, \max_i\left\{T_{wait\_data,i}\right\}\right\}. \tag{3}$$

In consequence, the length of the reduced CFP, T_{rCFP}, is calculated as:

$$T_{rCFP} = T_{CFP} - T_b \, . \tag{4}$$

As real-time traffic can only be sent during the length of the reduced CFP, we model the bit rate for the real-time traffic as an experienced bit rate r_e, i.e., we scale down the actual bit rate r by the ratio of the reduced CFP and superframe length, i.e.,

$$r_e = \frac{T_{rCFP}}{T_{SF}} \cdot r \, . \tag{5}$$

This leads to a scaling down also of the waiting times defined in Eq. 1 and 2:

$$T_{e,wait\_poll,i} = T_{wait\_poll,i} \cdot \frac{T_{rCFP}}{T_{SF}} \, , \tag{6}$$

$$T_{e,wait\_data,i} = T_{wait\_data,i} \cdot \frac{T_{rCFP}}{T_{SF}} \, . \tag{7}$$

3.2 Timing Analysis

For the analysis we are assuming that the number of RTCs in the network is Q, while there are M ReRTCs available. In conjunction with admission control (if not done offline during design stage), the master uses the real-time schedulability analysis described in the next chapter to only admit data traffic for which required delays can be guaranteed. One of the parameters needed in this analysis is the maximum queuing delay that any packet can experience, and which can be calculated knowing the transport layer deadline D_i and the worst case delay the packet can experience. As D_i is divided into $D_{ord,i}$ and $D_{retr,i}$, the worst case delay consists of two components, $d_{ord,i}$ and $d_{retr,i}$, respectively. For the highest priority packet this will occur if the packet arrives to the EDF queue just as the transmission of a lower priority packet starts and if after the completion of this transmission the interval left in the CFP is just too short for the delayed packet to be sent. The worst case delay experienced by that packet would in this case comprise two waiting times plus a complete CBP and consequently the maximum queuing delay can be calculated as:

$$d_{ordM2S} = D_{ord,i} - T_{CBP} - T_{wait\_data,i} - \max\left\{T_{wait\_data,i}, T_{wait\_poll,i}\right\}, \quad (8)$$

$$d_{ordS2M} = D_{ord,i} - T_{CBP} - T_{wait\_poll,i} - \max\left\{T_{wait\_data,i}, T_{wait\_poll,i}\right\}. \quad (9)$$

When calculating the maximum queuing delay for packets using retransmission channels, d_{retr}, we assume the retransmission deadline $D_{retr,i}$ to be the same for all values of i, i.e., we can set $D_{retr,i} = D_{retr}$. Additionally, D_{retr} has to be divided between all of the allowed retransmission attempts:

$$D_{re} = \frac{D_{retr}}{N_{attempt}}. \quad (10)$$

The maximum queuing delay for retransmission packets can then be calculated as:

$$d_{retrM2S} = D_{re} - T_{CBP} - T_{wait\_data,i} - \max\left\{T_{wait\_data,i}, T_{wait\_poll,i}\right\}, \quad (11)$$

$$d_{retrS2M} = D_{re} - T_{CBP} - T_{wait\_poll,i} - \max\left\{T_{wait\_data,i}, T_{wait\_poll,i}\right\}. \quad (12)$$

3.3 Real-Time Scheduling Analysis

In the admission control, the master uses a real-time schedulability analysis originally developed for EDF scheduling of periodic tasks on a uniprocessor [19] in order to check the feasibility of the real-time traffic allocation as proven possible in [24]. The check is implemented in two stages, a utilization check and a workload check, and is only done when a new RTC is to be allocated. In case both checks are positive, the deadline for the checked traffic flow can be guaranteed.

The utilization check calculates the bandwidth utilization U of all RTCs and ReRTCs during the CFP and ensures it to be no bigger than 1. The calculation is done as follows:

$$U = \sum_{i=1}^{\alpha}\left(\frac{T_{e,wait\_poll,i}}{P_i}\right) + \sum_{i=1}^{\beta}\left(\frac{T_{e,wait\_data,i}}{P_i}\right) + \sum_{i=1}^{\gamma}\left(\frac{T_{e,wait\_retr\_poll,i}}{P_{re,i}}\right) + \sum_{i=1}^{\delta}\left(\frac{T_{e,wait\_retr\_data,i}}{P_{re,i}}\right). \quad (13)$$

Greek lettering indicates the number of traffic flows in the different summations. While α is the numbers of RTC from the master to the slaves, β denotes RTCs in the opposite direction, and $\alpha + \beta = Q$. γ and δ are the number of ReRTCs from master to slave and slave to master, respectively, where $\gamma + \delta = M$. $T_{e,wait\_retr\_poll}$ and $T_{e,wait\_retr\_data}$ are defined in the same manner as $T_{e,wait\_poll}$ and $T_{e,wait\_data}$, and only introduced here to help understanding.

The second check calculates the workload that is put onto the network by all accepted RTCs. It is done by the means of a workload function, $h(t)$, in which all transmission times of all packets of all traffic flows (including the retransmission flows) with a deadline before time t are summed up. Time starts at the start of a hyperperiod, i.e., when all traffic flows' periods start simultaneously. Hyperperiods continue until the periods start at the same time again. Using the workload check during one hyperperiod has been shown to constitute the worst case scenario [25-27]. The $h(t)$ function is calculated as follows, where summands one and two calculate the workload imposed by RTCs and summands three and four that of the ReRTCs:

$$\begin{aligned} h(t) = & \sum_{\substack{i\in[1,\alpha],\\ d_{ordS2M,i}\le t}} \left(1+\left\lfloor\frac{t-d_{ordS2M,i}}{P_i}\right\rfloor\right)\cdot T_{e,wait\_poll,i} + \\ & \sum_{\substack{i\in[1,\beta],\\ d_{ordM2S,i}\le t}} \left(1+\left\lfloor\frac{t-d_{ordM2S,i}}{P_i}\right\rfloor\right)\cdot T_{e,wait\_data,i} + \\ & \sum_{\substack{i\in[1,\gamma],\\ d_{retrS2M,i}\le t}} \left(1+\left\lfloor\frac{t-d_{retrS2M,i}}{P_{re,i}}\right\rfloor\right)\cdot T_{e,wait\_retr\_poll,i} + \\ & \sum_{\substack{i\in[1,\delta],\\ d_{retrM2S,i}\le t}} \left(1+\left\lfloor\frac{t-d_{retrM2S,i}}{P_{re,i}}\right\rfloor\right)\cdot T_{e,wait\_retr\_data,i} \end{aligned}. \quad (14)$$

For the workload check to hold, the following demand has to hold:

$$h(t) \le t \quad \forall t. \quad (15)$$

According to [28], this computationally heavy computation (due to the complexity introduced by the notion of continuous time) can be simplified to only include the following points in time at which the workload function has to be evaluated:

$$\begin{aligned} t \in & \bigcup_{i=1}^{\alpha} \left\{ w \cdot P_i + d_{ordS2M,i} : w = 0,1,2... \right\} + \\ & \bigcup_{i=1}^{\beta} \left\{ x \cdot P_i + d_{ordM2S,i} : x = 0,1,2... \right\} + \\ & \bigcup_{i=1}^{\gamma} \left\{ y \cdot P_{re,i} + d_{retrS2M,i} : y = 0,1,2... \right\} + \\ & \bigcup_{i=1}^{\delta} \left\{ z \cdot P_{re,i} + d_{retrM2S,i} : z = 0,1,2... \right\} \end{aligned} \tag{16}$$

and

$$t \in \left[1; P_{busy} \right]. \tag{17}$$

The busyperiod P_{busy} is the length of time between the start of a hyperperiod until the the link becomes idle the first time. Here P_{busy} denotes the first busyperiod, i.e., the busyperiod that starts at the beginning of the first hyperperiod.

4 Performance Evaluation

We evaluated the performance of our framework by simulating a series of platooning test cases and present the results from two representative cases in this chapter. The simulator was implemented in MatLab. The bit rate is set to be 6 Mbps and the frame durations are calculated and set according to IEEE 802.11p. A polling packet has aduration of 154 μs, corresponding to a packet length of 80 bits plus MAC and physical layer headers and highest priority AIFS. ACKs are assumed to have the same duration as polling packets, while beacon frames have a duration of 370 μs, comprising 200 bytes including MAC and physical layer headers and highest priority AIFS. The corresponding values for data packets are 642 μs and 400 byte. To account for a deterioration of the channel quality due to fading and shadowing (by vehicles situated in-between sender and receiver), the simulated packet error rate is calculated as:

$$PER = N_{hop} \cdot 0.05 \tag{18}$$

where N_{hop} = 1 if the sending and receiving vehicle are direct neighbors, N_{hop} = 2 if there is one vehicle in-between etc.

In this performance study, we compare our retransmission scheme to a case without retransmissions. For each simulation setup, the average performance is based on

250 simulations runs of a duration of 1000 hyperperiods. An initial real-time feasibility check determines whether the timing requirements of a given task set can be met. Each vehicle is assumed to have one RTC for transmission to the master (middle vehicle), while the master has one RTC to the platoon leader and a RTC has a period and deadline of 50 ms. Due to the broadcast nature of the transmissions, every vehicle can overhear and make use of all data traffic sent within its reception range. Only the single destination RTCs described in Chapter 3, however, are simulated and considered for evaluation.

In the first simulated case, Case 1, we have a superframe length of 25 ms and a CBP length of 5 ms (including beacon). There are $M = 4$ retransmission RT channels, each having a period of $P_{re,i} = 25$ ms. The maximum number of retransmission attempts is set to $N_{attempt} = 2$. The deadline for the retransmission RT channels is always calculated automatically to the minimum possible value by using the workload function and the real-time analysis backwards. With this configuration, the real-time analysis showed that a maximum of seventeen vehicles could be supported without deadline misses. Fig. 2 shows the experienced message error rate (MER) without ARQ, after maximum one or, if required, two retransmission attempt. As the number of vehicles (and RTCs) increases, the retransmission RTCs become a bottleneck and retransmission attempts must be rejected. Moreover, the PER increases with distance. Disregarding that fact in a simulation with a distance-independent PER (not shown here) we could still see a higher average MER in longer platoons. As seen in Fig. 2, a MER below 10^{-3} can be achieved with only a few vehicles in the platoon, while the MER reaches 6.3×10^{-2} for 17 vehicles, still a considerable reduction compared to the case without retransmissions. Fig. 3 shows the MER for the master only, reaching 0.13 for 17 vehicles and a reduction of one to two orders of magnitude for moderate-sized platoons. Fig. 4 shows that the channel busy time, i.e. the percentage of time when the medium is occupied, in most cases stays below 25% and that penalty introduced by retransmissions is moderate. The average delay (not shown) for all platoon lengths remains below 2.2 ms and thereby far below the stated guarantee.

In the second simulated case, Case 2, we have a superframe length of 50 ms and no CBP. There are $M = 9$ retransmission RTC, each with a period of $P_{re,i} = 20$ ms. The maximum number of retransmission attempts is set to $N_{attempt} = 3$. According to the real-time analysis, this configuration allows for a maximum of 15 vehicles in the platoon. Fig. 5 compares the MER without retransmissions and to cases of up to three retransmission attempts where a reduction by roughly two and three orders of magnitude can be achieved for longer and shorter platoons, respectively.

The MER for the master sending to the leading vehicle is shown in Fig. 6. For 15 vehicles the reduction in MER is more than one order of magnitude. For shorter platoons an improvement of several orders of magnitude can be achieved. The average delay (not shown) is below 1.3 ms for all platoon sizes, while the channel busy time (not shown) reaches 30 % for a platoon length of 15 vehicles and 25% for 13 vehicles, respectively.

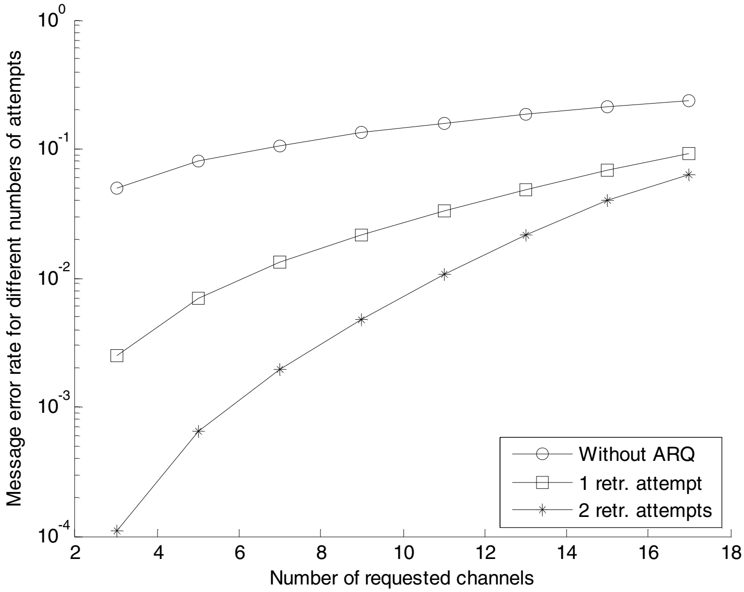


Fig. 2. Message error rate for Case 1

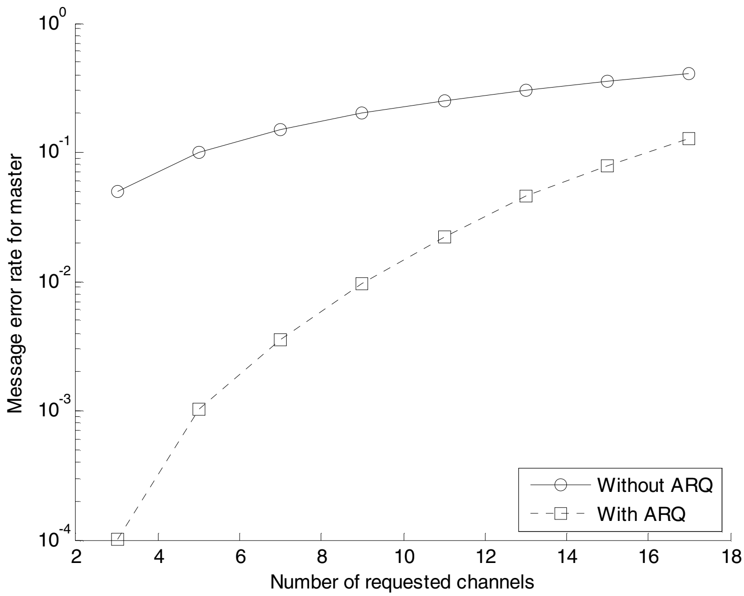


Fig. 3. Message error rate for master to leading vehicle in Case 1

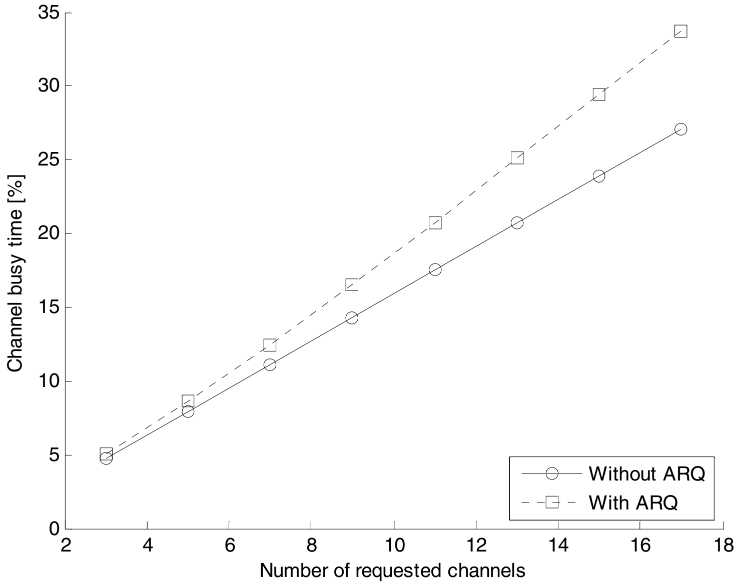


Fig. 4. Channel busy time for Case 1

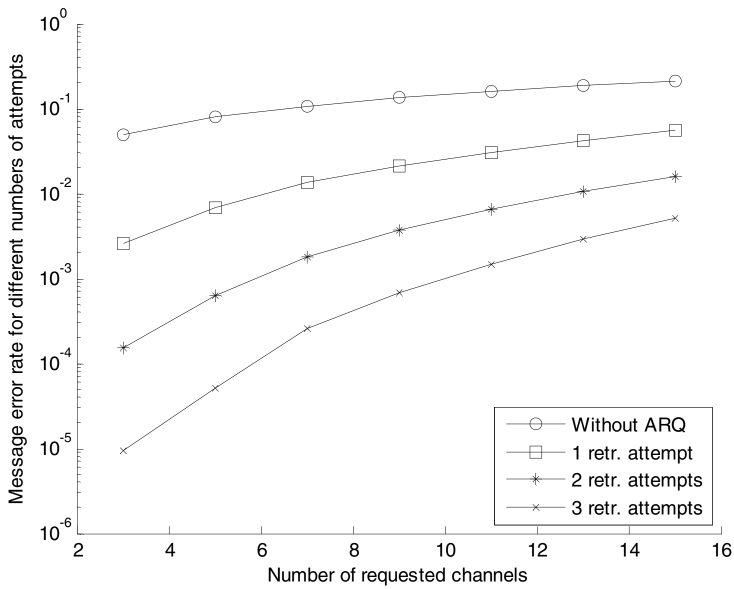


Fig. 5. Message error rate for Case 2

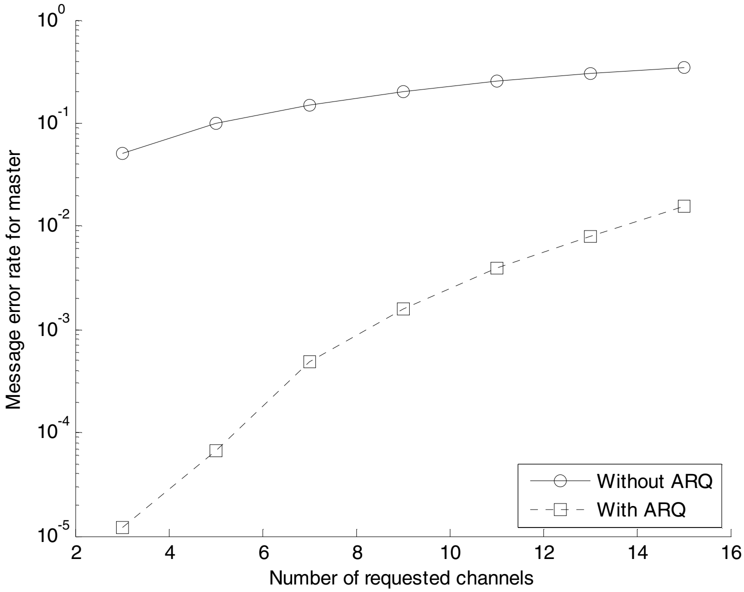


Fig. 6. Message error rate for master to leading vehicle in Case 2

5 Conclusion

In this paper we presented an enhancement to the IEEE 802.11p standard for the reliable and timely exchange of safety-critical control data in platooning applications. The framework combines guaranteed delay bounds with the possibility of retransmitting erroneous packets. Although we showed that our retransmission scheme significantly reduces the experienced message error rate, our simulation studies even shed light on the overall effect of the number of platoon members on the reliability that can be achieved. For moderate sized platoons, a reduction in message error rate by several orders of magnitude could be achieved.

References

1. Vinel, A., Staehle, D., Turlikov, A.: Study of Beaconing for Car-to-Car Communication in Vehicular Ad-Hoc Networks. In: Proc. IEEE International Conference on Communications, Dresden, Germany. IEEE (June 2009)
2. Alam, A., Gattami, A., Johansson, K.H.: An Experimental Study on the Fuel Reduction Potential of Heavy Duty Vehicle Platooning. In: International Conference on Intelligent Transportation Systems, Madeira, Portugal. IEEE (September 2010)
3. Standards Committee, IEEE Standard 802.11p-2010, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments, pp. 1–51 (July 2010)

4. Böhm, A., Lidström, K., Jonsson, M., Larsson, T.: Evaluating CALM M5-based Vehicle-to-Vehicle Communication in Various Road Settings Through Field Trials. In: The 4th IEEE LCN Workshop On User MObility and VEhicular Networks (ON-MOVE), Denver, CO, USA, pp. 613–620. IEEE (October 2010)
5. Jonsson, M., Kunert, K., Böhm, A.: Increasing the Propability of Timely and Correct Message Delivery in Road Side Unit Based Vehicular Communication. In: Intelligent Transportation Systems Conference (ITSC 2012), Anchorage, AK, USA. IEEE (September 2012)
6. Fernandes, P., Nunes, U.: Platooning with DSRC-Based IVC-Enabled Autonomous Vehicles: Adding Infrared Communication for IVC Reliability Improvement. In: Intelligent Vehicle Symposium, Alcalà de Henares, Spain. IEEE (June 2012)
7. Mak, T.K., Laberteaux, K.P., Sengupta, R.: A Multi-Channel VANET Providing Concurrent Safety and Commercial Services. In: Proc. 2nd ACM International Workshop on Vehicular Ad Hoc Networks (VANET 2005), pp. 1–9. ACM, New York (2005)
8. Barradi, M., Hafid, A.S., Gallardo, J.R.: Establishing Strict Priorities in IEEE 802.11p WAVE vehicular networks. In: GLOBECOM 2010, Miami, FL, USA, pp. 1–6. IEEE (December 2010)
9. Bilstrup, K., Uhlemann, E., Ström, E.G., Bilstrup, U.: On the Ability of the 802.11p MAC Method and STDMA to Support Real-Time Vehicle-to-Vehicle Cmmunications. EURASIP J. on Wireless Comm. and Networking, Article number 902414, 1–14 (2009)
10. Butt, M.M.: Provision of Guaranteed QoS with Hybrid Automatic Repeat Request in Interleave Division Multiple Access Systems. In: IEEE Singapore Int. Conference on Communication Systems (ICCS 2006), Singapore, pp. 1–5. IEEE (October 2006)
11. Giancola, G., Falco, S., Di Benedetto, M.G.: A Novel Approach to Error Protection in Medium Access Control Design. In: 4th Int. IEEE Workshop on Mobile and Wireless Communication Networks (MWCN 2002), Stockholm, Sweden, pp. 337–341. IEEE (September 2002)
12. Fan, X., Jonsson, M., Jonsson, J.: Guaranteed Real-Time Communication in Packet-Switched Networks with FCFS Queuing. Comp. Netw. 53(3), 400–417 (2009)
13. Kosch, T., Kulp, I., Bechler, M., Strassberger, M., Weyl, B., Lasowski, R.: Communication Architecture for Cooperative Systems in Europe. IEEE Communications Magazine 47(5), 116–125 (2009)
14. Pereira, P.P., Casaca, A., Rodrigues, J.J.P.C., Soares, V.N.G.J., Triay, J., Cervelló-Pastor, C.: From Delay-Tolerant Networks to Vehicular Delay-Tolerant Networks. IEEE Communications Surveys & Tutorials 14(4), 1166–1182 (2012)
15. Jarupan, B., Ekici, E.: A survey of cross-layer design for VANETs. Ad Hoc Networks 9, 966–983 (2011)
16. Böhm, A., Jonsson, M.: Real-Time Communication Support for Cooperative, Infrastructure-Based Traffic Safety Applications. Int. J. of Vehicular Technology, Article ID 541903, 17 (2011)
17. Böhm, A., Jonsson, M.: Supporting Real-Time Data Traffic In Safety-Critical Vehicle-To-Infrastructure Communication. In: The 2nd IEEE LCN Workshop On User MObility and VEhicular Networks (ON-MOVE), Montreal, Canada, pp. 614–621. IEEE (October 14, 2008)
18. Jonsson, M., Bergenhem, C., Olsson, J.: Fiber-Ribbon Ring Network With Services For Parallel Processing And Distributed Real-Time Systems. In: Proc. ISCA 12th International Conference on Parallel and Distributed Computing Systems (PDCS 1999), Fort Lauderdale, FL, USA, August 18-20, pp. 94–101. ISCA (1999)

19. Liu, C.L., Layland, J.W.: Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment. J. of the ACM 20(1), 46–61 (1973)
20. Kunert, K., Uhlemann, E., Jonsson, M.: Predictable Real-Time Communications with Improved Reliability for IEEE 802.15.4 based Industrial Networks. In: Proc. 8th IEEE International Workshop on Factory Communication Systems (WFCS 2010), Nancy, France, May 18-21, pp. 13–22. IEEE (2010)
21. Kunert, K., Jonsson, M., Uhlemann, E.: Exploiting Time and Frequency Diversity in IEEE 802.15.4 Industrial Networks for Enhanced Reliability and Throughput. In: Proc. 15th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2010), Bilbao, Spain, September 13-16, 9 pages. IEEE (2010)
22. Jonsson, M., Kunert, K.: Towards Reliable Wireless Industrial Communication with Real-Time Guarantees. IEEE Tr. on Ind. Informatics 5(4), 429–442 (2009)
23. Jonsson, M., Kunert, K.: Reliable Hard Real-Time Communication In Industrial And Embedded Systems. In: Proc. Third IEEE International Symposium on Industrial Embedded Systems (SIES 2008), Montpellier, France, June 11-13, pp. 184–191. IEEE (2008)
24. Hoang, H., Jonsson, M.: Switched Real-Time Ethernet in Industrial Applications - Deadline Partitioning. In: 9th Asia-Pacific Conference on Communications (APCC 2003), Penang, Malaysia, vol. 1, pp. 76–81. IEEE (September 2003)
25. Spuri, M.: Analysis of Deadline Scheduled Real-Time Systems. Technical Report No. 2772, INRIA, France (1996)
26. Baruah, S.K., Mok, A.K., Rosier, L.E.: Preemptively Scheduling Hard-Real-Time Sporadic Tasks on One Processor. In: IEEE Real-Time Systems Symposium (RTSS 1990), Lake Buena Vista, FL, USA, pp. 182–190. IEEE (December 1990)
27. Baruah, S.K., Rosier, L.E., Howell, R.R.: Algorithms and Complexity Concerning the Preemptive Scheduling of Periodic, Real-Time Tasks on One Processor. Real-Time Systems 2(4), 301–324 (1990)
28. Stankovic, J.A., Spuri, M., Ramamritham, K., Buttazzo, G.C.: Deadline Scheduling for Real-Time Systems – EDF and Related Algorithms. Kluwer Academic Publishers, Boston (1998)

Development of Car2X Communication and Localization PHY and MAC Protocol Following Iterative Spiral Model Using Simulation and Emulation

Axel Sikora[1], Dirk Lill[2], Manuel Schappacher[2], Simon Gutjahr[2], and Eugen Gerber[2]

[1] Offenburg University of Applied Sciences,
Faculty of Electrical & Information Technology Engineering, Offenburg, Germany
axel.sikora@hs-offenburg.de
[2] Steinbeis Innovation Center Embedded Design and Networking, Heitersheim, Germany
{dirk.lill,manuel.schappacher}@stzedn.de,
{simon.gutjahr,eugen.gerber}@stzedn.de

Abstract. The communication between objects, i.e. between cars (car-2-car, C2C), between cars and infrastructure (car-2-infrastructure, C2I) and between cars and vulnerable road users (car-2-VRU, C2VRU) is a major stepping stone towards traffic applications to enable efficient and safe traffic flow. However, these applications pose very high requirements to the communication protocols, which go beyond the capabilities of an available standardized solution.

This contribution shows how iterative design processes can help to fulfill these requirements, while re-using a maximum of elements from one level to the next and thus avoiding unrealistic overhead. In especially, the added value of simulation and emulation in this iterative process is elaborated.

Keywords: Car-2-Car communication, Car-2-X communication, simulation, emulation, iterative design process.

1 Introduction

The communication between objects, i.e. between cars (car-2-car, C2C), between cars and infrastructure (car-2-infrastructure, C2I) and between cars and vulnerable road users (car-2-VRU, C2VRU) is a major stepping stone towards traffic applications to enable efficient and safe traffic flow. However, these applications pose very high requirements to the communication protocols, which go beyond the capabilities of an available standardized solution. Consequently, a specific protocol must be designed, which take into account all these requirements. Due to the complex requirements, the design of the communication protocols must be performed extremely careful. Systematic approach, thorough testing and extensive verification are major preconditions for the successful development of communication units.

This contribution shows how iterative design processes can help to fulfill these requirements, while re-using a maximum of elements from one level to the next and thus avoiding unrealistic overhead. In especially, the added value of simulation and emulation in this iterative process is elaborated.

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 136–145, 2013.

In the remainder of this contribution, the activities of the development for the layer-2-protocol of a C2X-communication subsystem are described, where testing and verification are tightly integrated into a multi-level iterative protocol design flow. In the remainder of this contribution, the general flow for protocol design is described (ch. 2), before the concrete protocols that are already available for C2X-communication are presented (ch. 3). This chapter includes a short description of the authors' project background from the Ko-TAG project. In ch. 4, the hardware and software setup of the emulator is presented, before first measurement results are shown and discussed in ch. 4.3.

2 State of the Art of Protocol Design

2.1 Process Models

A plethora of process models has been described for the development of software, hardware, and systems [1], where typically linear and iterative process models are distinguished. Iterative models differ from the linear models in such a way that the design and the verification are performed iteratively in a sequence of different abstraction levels. Iterative models are regularly used for the development of very complex microelectronic systems, i.e. in defense, aerospace or automotive industry, but find much less acceptance for the development of smaller systems. However, the extensive use of simulation and emulation can help to increase the number of iterations and thus to decrease the distance between two design cycles.

The spiral model for the design of new communication protocols is shown in Fig. 1 and contains the following steps:

- Level 1: The requirements are described in a customer requirements specification. These requirements include parameter like number of nodes, activity behavior of the nodes, channel conditions, real-time and energy requirements, necessity of multi-hop operation, etc.
- Level 2: Existing algorithms are researched and analyzed, and a proof of concept is elaborated. This might be done with abstract behavioral models, like sequence flow and state diagrams, preferably using Unified Modeling Language (UML). If necessary and suitable, top-level (abstract) simulations or calculations can help to proof the concept.
- Level 3: For communication protocols, event driven network simulators might help to support the transition from abstract description to concrete firmware implementation. This level is shortly described in subchapter 2.2.
- Level 4: Next step can be an emulator platform, where hardware can be used in a closed and reproducible test environment. The general approach is described in subchapter 2.3. Our concrete development is presented in ch. 4.
- Level 5: The tests conclude in real field trials, where as many of the test scenarios should be executed. However, full control of all parameters – and thus reproducibility - cannot be guaranteed. It is important to understand that the tools of the earlier steps can help to accelerate the implementation of the field tests. This is described in subchapter 2.4.

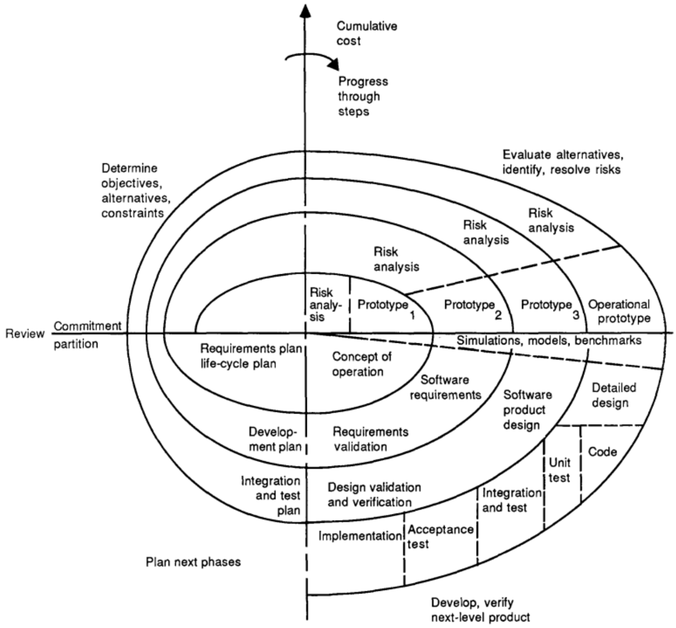


Fig. 1. Spiral model of the software process [2]

2.2 Simulation

Running simulations is experimenting with models. Models describe the real behavior on an abstract level, i.e. leaving out all irrelevant implementation details

Event driven simulators are the platform of choice for an efficient execution of scenarios using models of communication nodes. The most popular event driven network simulators are the open-source simulators NS-3 [12] and the OMNet++ Network Simulation Framework [13], as well as the commercial OPNET simulator [14]. Simulation is an important element in many protocol design processes. Therefore, no further references are given to show examples of simulation results.

It is important to understand that with the help of the simulator also more than one abstraction level can be described. In most of the simulators, the node behavior can be programmed on a very (!) abstract level. But – if reasonable hardware abstraction layer (HAL) is provided – it is also possible to have full firmware level description. Thus, the firmware development of the nodes can be nearly completely implemented on the simulator level – without any piece of hardware.

2.3 Emulation

As soon as hardware is available, a transition can be made from simulator to emulator level, where physical networking nodes are used, but where RF environment is still modeled using RF-waveguides. In such an environment, it is well possible to

- control the path loss between any of the nodes with the help of RF attenuators, which possibly can be programmable (or at least interchangeable)
- control the connectivity between any of the nodes with the help of RF switches, which are programmable.

With this, arbitrary topologies can be selected.

We call this step emulation, although many others make use of this terminology for still virtual equipment, like [3] or [4]. The terminology "testbed" is also misleading, as others use this denomination already for field test level, like for example [5].

We have made the experience that – although a certain effort at first time – the use of such an emulation platform decreases the overall development and testing effort. This especially holds true if further elements from the emulation phase can be re-used in the field trial phase.

It should be mentioned that the authors made earlier use of such an emulated platform [11].

The authors very much appreciate that this approach has already found acceptance in the community and many examples can be found in literature, e.g. [6] [7] [8]. On the European level, there are various further testbeds with different objectives and characteristics, i.e. [9] [10].

2.4 Field Tests

If and when the earlier phases of design and verification have been successful, field tests are the very last element of the overall development. It is important to mention that – if carefully designed – control and monitoring equipment from the emulator can be reused.

3 C2X-Protocols

3.1 Requirements of C2X-protocols

C2X-applications pose very high requirements to the communication protocols and are thus object to

- They are highly dynamic, with potentially only short periods of interaction of partners.
- They must be very scalable and should work with only very few sparsely distributed nodes, but should be functional also with very many nodes in dense environments.
- Thus, they potentially might by very complex due to the large number of potential nodes.
- A major element of the complexity comes from manifold existence of hidden-stations (near-far-problem).
- They are potential subjects to security attacks [15].

3.2 Project Background

Within the research initiative Ko-FAS from the German Ministry of Economics and Technologies (BMWi) [19], cooperative pedestrian protection systems (CPPS) shall be developed that follow the communication model of secondary surveillance radar from air traffic control and shall help to overcome the weaknesses of legacy advanced driver assistance systems (ADAS). The sub-project Ko-Tag concentrates on the development of a pedestrian tag that can communicate with a vehicle and can be localized by the vehicle. Within the Ko-Tag project, it is the task of the authors to provide a stable, secure and scalable network architecture between vehicles and VRUs that integrates into the upcoming C2C- and C2I-networks.

3.3 Enhancements to C2X-Protocols

For the communication between (car-2-car, C2C) and between cars and infrastructure (car-2-infrastructure, C2I), the "IEEE 1609 Family of Standards for Wireless Access in Vehicular Environment (WAVE)" [16] defines an architecture and a complementary, standardized set of services and has found wide acceptance.

However, as already motivated in [17], the pure 802.11p protocol is not suitable for the communication between cars and vulnerable road users (car-2-VRU, C2VRU), as it does not sufficiently support fast real-time characteristics. In addition, it does not integrate localization functionality.

Therefore, in the first iterations of the spiral design cycle, a protocol extension to 802.11p has been developed and simulated [18]. As of now, first hardware is available and thus it is well possible to enter into the next design flow loop on an emulator platform.

4 Emulator Development

4.1 Objectives

Although there are already various emulator testbeds available, there is the necessity to design an extended platform:

- The Ko-TAG protocol is sought to be an extension to C2C-communication, and thus makes use of the 5.9 GHz-band. This leads to the necessity to design all RF elements and paths with a capability up to 6 GHz.
- The Ko-TAG protocol works in a time-synchronized manner, where accuracy requirements are in the µs- range. Therefore, it is required to have separate high-precision clock lines, i.e. for the On-Board Units (OBU).
- The Ko-TAG protocol is not restricted to pure communication, but also includes distance measurements. Therefore, the connection lines must be calibrated with regard to Time of Flight (TOF) measurements.

4.2 Setup

Fig. 3 shows the logical setup of the Ko-TAG emulator platform.

- The Ko-TAG RF nodes are mounted into shielded boxes (cf. Fig. 3).
- The Ko-TAG RF nodes (Ko-TAG VRU and Ko-TAG OBU) are interconnected via RF-waveguides, splitters, and attenuation elements.
- The signals at the OBU can be observed via a Signal Analyzer (Agilent EXA N9010A).
- Additional noise and signals can be fed into the RF channel from the Vector Signal Generator (Agilent MXG N5182A). The Vector Signal Generator can work continuously, but can also be triggered, so that the interference can be inserted in relation to the overall time synchronization. I.e. single packets or certain parts of single can be subject of interference.
- The complete setup is globally controlled and observed by Matlab from a PC.
- The local control of the RF switches is performed by an ARM Cortex-M3 based microcontroller unit (STM32F107VCT6).

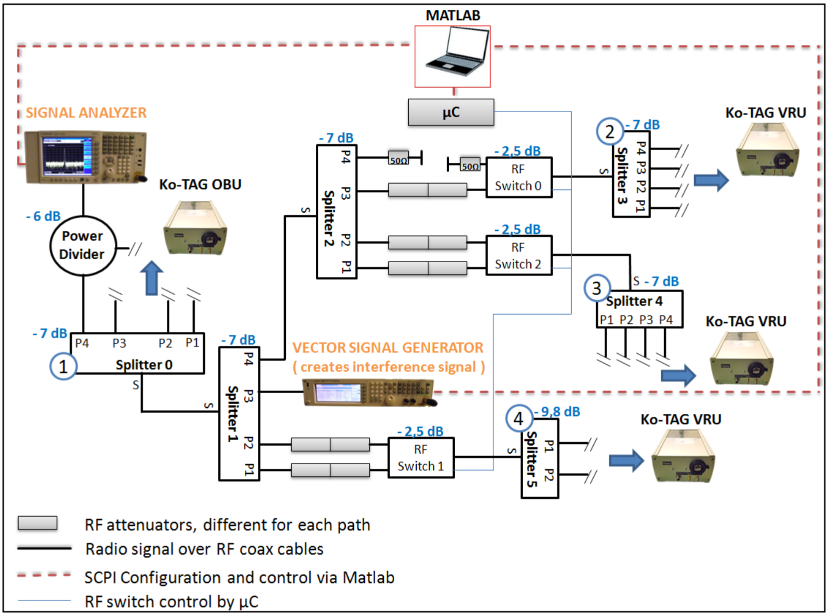


Fig. 2. Logical topology of Ko-TAG emulator

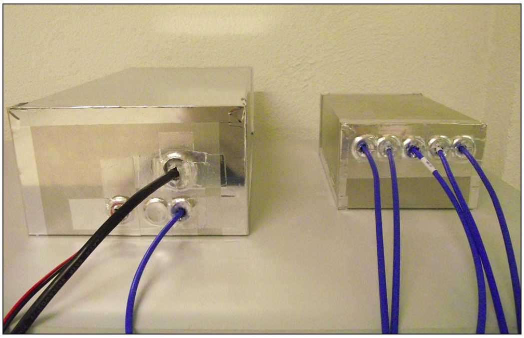


Fig. 3. RF-nodes being mounted in shielded boxes

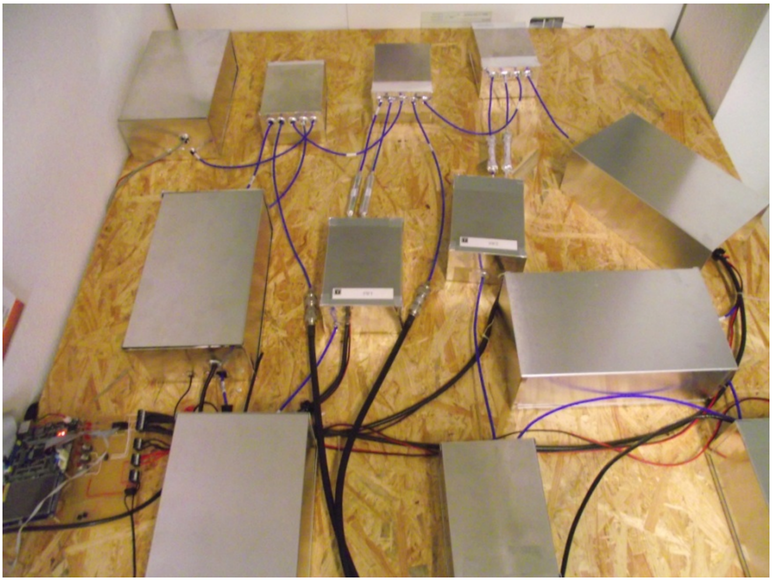


Fig. 4. Physical topology of Ko-TAG emulator (current functional prototype status)

4.3 Calibration

Based on the experience from earlier emulator setups, all elements are carefully mounted and additionally shielded (cf. Figs. 3 and 4). The links are fixed with a torque wrench. All elements and connections are characterized after they have been built-in with regard to attenuation, reflection, and phase shift (delay).

4.4 Results

Typical measurement results are shown in Fig. 5 and 6. Fig. 5 shows the frequency characteristics of the different channels, as they are described in [18]. The beacon frame in the control channel and the data packets in the TOF channel are depicted. Both channels are 10 MHz wide.

Fig. 6 shows a screenshot from the Matlab based GUI at the Control PC with the quality of the transmission, by reading out the receiver internal registers, and the anticipated distance values. In addition, Fig. 7 shows the packet success rate and the signal quality in dependence of the signal to noise ratio (SNR).

5 Outlook

An emulator platform has been presented that allows very good controllability and observability of 5.9 GHz-based Car2X-communication, including precise synchronization and range measurements. Based on this platform, further evaluation of the Ko-TAG system will be performed.

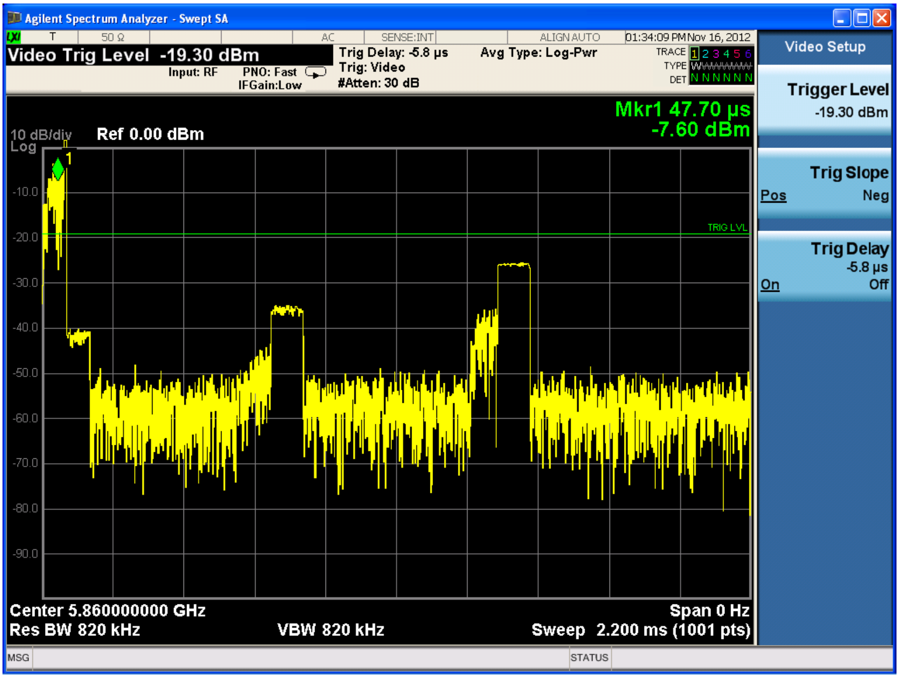


Fig. 5. Measurement results from Ko-TAG emulator: Frequency Analysis at Signal Analyzer

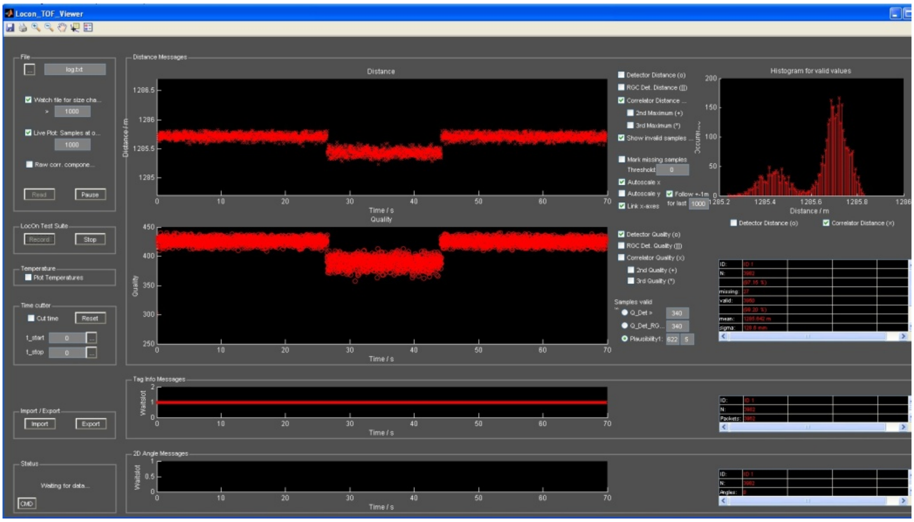


Fig. 6. Measurement results from Ko-TAG emulator: Matlab based Analysis Window at the PC for Statistical Analysis

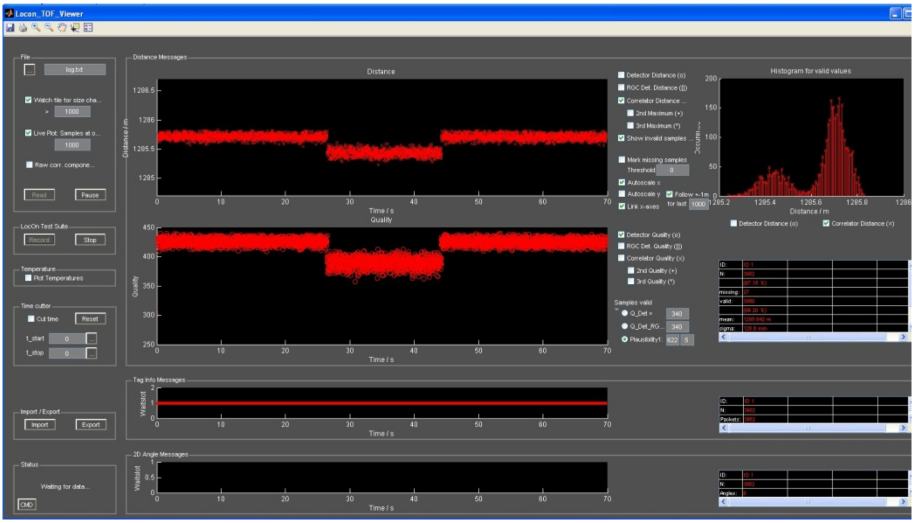


Fig. 7. Measurement results from Ko-TAG emulator: Packet Success Rate and Signal Quality in dependence of Signal-to-Noise-Ratio

References

1. Sommerville, I.: Software Engineering. Pearson Education
2. Boehm, B.W.: A Spiral Model of Software Development and Enhancement. IEEE Computer 21(5), S.61–S.72 (1988)

3. Staub, T., Gantenbein, R., Braun, T.: VirtualMesh: an emulation framework for wireless mesh networks in OMNeT++. In: 2nd Int'l. Conference on Simulation Tools and Techniques (Simutools 2009) (2009)
4. Evans, N.S., Grothoff, C.: Beyond simulation: large-scale distributed emulation of P2P protocols. In: 4th Conference on Cyber Security Experimentation and Test (CSET 2011) (2011)
5. Iwanicki, K., Gaba, A., van Steen, M.: KonTest: A Wireless Sensor Network Testbed at Vrije Universiteit Amsterdam. Technical Report IR-CS-045, Vrije Universiteit Amsterdam (August 2008), http://www.few.vu.nl/~iwanicki/ (visited December 2012)
6. Nourbakhsh, E., Dix, J., Johnson, P., Burchfield, R., Venkatesan, S., Mittal, N., Prakash, R.: ASSERT: A Wireless Networking Testbed. In: Magedanz, T., Gavras, A., Thanh, N.H., Chase, J.S. (eds.) TridentCom 2010. LNICST, vol. 46, pp. 209–218. Springer, Heidelberg (2011)
7. Su, Y., Gross, T.R.: Validation of a miniaturized wireless network testbed. In: 3rd ACM Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization (WINTECH 2008) (2008)
8. De, P., Raniwala, A., Sharma, S., Chiueh, T.: MiNT: a miniaturized network testbed for mobile wireless research. In: 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005) (2005)
9. http://www.ict-openlab.eu/technologies/testbeds.html
10. Chatzigiannakis, I., Koninis, C., Mylonas, G., Fischer, S., Pfisterer, D.: WISEBED: an Open Large-Scale Wireless Sensor Network Testbed. In: Proc. of the 1st Int'l. Conference on Sensor Networks Applications, Experimentation and Logistics, 2009. Springer LNICST, Social Informatics and Telecommunications Engineering, vol. 29, pp. 68–87 (2010)
11. Möllendorf, L., Schauenberg, D., Braun, N., Rahusen, D., Sikora, A.: A Distributed Embedded Web2.0 Based Automated Testbed for Wireless Mesh Networks. In: 1st IEEE Int'l. Workshop on Autonomic Wireless Networking (AWN 2009) (2009)
12. http://www.nsnam.org/
13. http://www.omnetpp.org/
14. http://www.opnet.com/
15. Sikora, A.: Privacy and Trust Management in Safety-Related Car2X Communication. In: Thampi, S.M. (ed.) Managing Trust in Cyberspace. Chapman and Hall/CRC (2013)
16. U.S. Department of Transportation, IEEE 1609 - Family of Standards for Wireless Access in Vehicular Environments (WAVE), Intelligent Transportation Systems Standards Fact Sheet (September 25, 2009), http://www.standards.its.dot.gov/fact_sheet.asp?f=80
17. Lill, D., Schappacher, M., Gutjahr, A., Sikora, A.: Development of a Wireless Communication and Localization System for VRU eSafety. In: 2nd Int'l. Workshop on Communication Technologies for Vehicles (Nets4Cars 2010) at 7th IEEE Int'l. Symposium on Communication Systems, Networks and DSP (CSNDSP 2010), Newcastle, UK (2010)
18. Lill, D., Schappacher, M., Gutjahr, A., Sikora, A.: Protocol and System Design of a Cooperative Pedestrian Safety System. In: 11 th Int'l. Conference on ITS Communication (ITST 2011), St. Petersburg, Russia (2011)
19. http://www.kofas.de

Characterization of a Laser Scanner Sensor for the Use as a Reference System in Vehicular Relative Positioning

Fabian de Ponte Müller[1], Luis Martín Navajas[2], and Thomas Strang[1]

[1] Institute of Communications and Navigation,
German Aerospace Center, Munich, Germany
[2] University of Málaga, Málaga, Spain

Abstract. Advanced Driver Assistance Systems (ADAS) play an important role in increasing the safety on today's roads. Forward collision warning systems, lane change assistants or cooperative adaptive cruise control are examples of safety relevant applications that rely on accurate relative positioning between vehicles. Current solutions found in commercial automobiles estimate the position of surrounding vehicles by measuring the distance with RADAR, cameras or IR-sensors. It is envisioned that the advent of inter-car communication will provide onboard relative positioning systems with further information about other vehicles in the surrounding area. While performing research in this field, the need of a proper reference system for testing new approaches originates. In the ideal case, such a reference system would yield the exact and continuous 3D baseline between two vehicles at any time in any circumstance. In this paper we will characterize the use of a laser scanner as a reference system for relative vehicle positioning.

Keywords: Laser Scanner, LD-MRS, Relative Positioning, RTK, Vehicles.

1 Introduction

The knowledge about other vehicle's position is a fundamental prerequisite for numerous Advanced Driver Assistance Systems (ADAS) in the Intelligent Transportation Systems (ITS) domain. Specially many safety relevant applications require robust relative positioning of surrounding vehicles rather than an absolute position of vehicles on the globe. Forward Collision Warning (FCW) or Cooperative Adaptive Cruise Control (CACC) are examples of such ITS safety applications. Relative positioning in traffic environment is typically addressed using a ranging sensor as for instance a RADAR. Laser scanners, due to their high cost, are currently not considered as a mass market solution in the automotive industry. Camera systems, on the other hand, are a promising solution that is still under research. However, single sensor solutions might not always meet the requirements imposed by future safety applications. For this reason numerous research groups currently work on sensor fusion approaches where the information

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 146–158, 2013.

from different sources is combined. The future availability of a wireless communication link between vehicles enables to extend the vehicle's perception range and develop cooperative approaches to estimate the target vehicle's position by combining sensors from different vehicles.

When evaluating the designed relative positioning algorithms in a real world environment, developers use a variety of systems to determine the "true" range between the vehicles. The system under test is then compared with the reference using this value. However, the employed reference systems are often not suited for the stated purpose. In some cases the chosen system lacks the required accuracy or a position accuracy analysis is not properly undertaken. In other cases the reference system cannot be universally employed, as it is only locally available and thus reduces the evaluation to a certain scenario. Also situations have been recorded where due to an unappropriate selection of the reference system test runs with an insufficient statistic were evaluated.

This paper presents a collection of different reference systems found in the literature stating their advantages and disadvantages by analyzing various parameters like accuracy, range, availability, price, etc. A solution for a relative positioning reference system by employing a laser scanner is presented in detail and compared to the previous systems. The laser scanner reveals certain important advantages over the mentioned systems and these have been verified experimentally in a set of tests.

The paper is structured as follows: The following section gives an overview of possible reference systems for vehicle relative positioning by analyzing approaches published by different research groups in this field. The third section presents the laser scanner in detail as a further reference system for vehicle relative positioning analyzing its advantages and disadvantages over the aforementioned systems. The laser scanner has been characterized for its accuracy, reliability and scanning range in a set of measurements. The experimental setup and its results are presented in section four. The paper ends with a conclusion.

2 Current Approaches

There are several systems that can be used as a reference system to evaluate new positioning approaches. When evaluating absolute or relative localization systems the key parameter characterizing the reference system is the accuracy in the position. A reference system might also give a measurement for the speed vector of the tracked object, absolute or relative. Further performance parameters are the coverage range, the sight line, the dynamic performance, the measuring rate, synchronization ability and the latency.

Although this paper will discuss a reference systems for relative localization of vehicles, we will first list different approaches for measuring the ground truth in a global or local coordinate frame. By duplicating any of the following systems and differencing its absolute positions the relative baseline between two vehicles can be retrieved. However, the relative position or speed accuracy, will fall off in quality by "adding" the position or speed error.

Global Navigation Satellite Systems (GNSS), like the US Global Positioning System (GPS) or the European Galileo system, can be used as a rough reference system when evaluating rough positioning techniques. In multipath free environments and with good sky visibility GNSS position accuracy usually lies below $10m$ [1]. Satellite based augmentation systems like the European EGNOS system can improve this accuracy by having geostationary satellites broadcasting correction signals. Many research groups favor to use Differential GPS (DGPS) with a local base station broadcasting pseudorange corrections. The position accuracy in the horizontal plane is in the order of $1m$ [1]. For instance [2] uses a DGPS approach to evaluate a radio ranging technique based on the received signal strength. In [3] differential GPS is used as a ground truth when comparing two movement models for vehicles and for determining the true lateral location of lanes by performing repeated runs. [4] compares a vehicle trajectory estimation algorithm to WAAS-GPS, the US equivalent to EGNOS.

Real Time Kinematic (RTK) systems can offer a far better solution for the absolute position than using stand-alone or differential GPS. In Real Time Kinematic (RTK) systems the carrier phase to the satellites in view is tracked and, along with the measurements of a reference base station (real or virtual, online or post-processed), the carrier phase ambiguity is solved. In case of a correct carrier resolution towards at least four satellites, this method is able to yield position accuracies in the order of 1% of the GPS wavelength, i.e. $2mm$ [1].

Schubert et. al. present in [5] a series of movement models which were evaluated with experimental data. A DGPS receiver with RTK capability was used for calculating the reference trajectory of the vehicle. The performance of an RTK receiver highly depends on its capability of resolving the unknown integer number of carrier cycles from the satellite to the receiver. Only in case of a fix ambiguity resolution the stated sub-centimeter accuracy can be achieved. Schubert's group took this into account during evaluation and discarded test runs with non-fix solutions. However, in multipath environments it is possible that the RTK device might be tracking the phase of a reflected signal and therefore give a wrong position solution. Also Alam et. al. [6] use this technology to acquire a reference for the relative position of two vehicles. During their measurement run they encountered the carrier-fix issue and could finally only utilize a 12 minute period out of a 45 minute journey for evaluation. To assess an Ultra-wideband system, in [7] GPS and GLONASS carrier phase measurements are post-processed along with IMU data using Waypoint's Inertial Explorer and GrafNav software packages. The resulting absolute positions of two vehicles are used to compute the bearing and range between them whenever a fixed ambiguity solution is found. In [8] and [9] two further examples for carrier phase based relative positioning by differencing absolute positions can be found.

GNSS based systems have the advantage of being globally available and offering a position solution in global coordinates. When moving from stand-alone GNSS, to SBAS, to DGPS and to RTK the accuracy of the system increases, so does the cost of the system. The time to fix increases while the lock robustness degrades. DGPS and RTK might need a reference station and a permanent communication link to the target vehicle. The systems have a good performance under clear-sky conditions but suffer severely in obstacle rich scenarios and therefore are unsuited for urban measurements.

Infrastructure based techniques might also offer advantages in determining the absolute position of a vehicle. Huang et al. [10] used an automated vehicle with on-board magnetometers following a track of magnetic markers. The paper states a lateral accuracy of 3cm without mentioning the accuracy, if even available, along the magnetic track. The clear disadvantage of techniques based on fix infrastructure is their limited deployability to different scenarios.

A further solution for tracking the absolute position of a vehicle is to use an optical measurement equipment. Tachymeters, commonly used in surveying, calculate the polar coordinates of a target prism by comparing phase shift measures from a reflected laser beam. A servomotor control system makes it possible to automatically track the target. If the geographic position of the tachymeter is known, the absolute position of the target prism can be computed. The usage of the tachymeter as a reference system is limited by its range and requires constant line of sight between the device and target vehicle. This limits its application to wide obstacle-free areas. As a kinematic measuring device it is limited in speed to around 20m/s and in angular velocity to around $45°/s$. Its high update rate of over 10Hz and its sub-centimeter accuracy are clear advantages of this system. Schönber at. al. use this technique to evaluate their GPS/INS solution for an autonomous vehicle in [11].

When it comes to determine the relative baseline between two land vehicles one alternative would be to duplicate any of the above listed absolute position sensors. However, the errors of both systems will add in quadrature (assuming independent errors). One simple solution is to force a static baseline and compare the system under test against the constant known range. In [12] Travis et al. use a towed trailer to force a constant distance between two GNSS antennas for evaluating a differential RTK approach for estimating the relative position between vehicles. The immediate drawbacks of this system are the limits in baseline length and driving speed. To avoid the problem of error growth, a solution that measures directly the baseline between the vehicles is preferred. A laser scanner sensor, the chosen device to be used as a reference system by our research group, will be analyzed in the following sections.

3 Laser Scanner

Our experiments aim to proof the usage of the SICK LD-MRS laser scanner as a reference sensor for relative positioning. For this purpose the maximum range and the range accuracy of the scanner are addressed. In the scope of this text the range vector is defined in a coordinate system centered at the foremost center point in the detector vehicle with the x-axis pointing in the driving direction and the y-axis to the left of the vehicle (see Fig. 1). The laser scanner's measurement frame is co-located and aligned with this frame (see Fig. 2). The range vector points to the rearmost center point of the target vehicle.

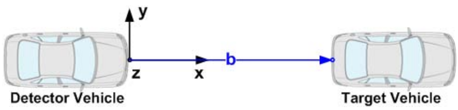


Fig. 1.

The LD-MRS is a four layer laser scanner which uses a rotating mirror and several pulsed laser beams to calculate the distance towards reflecting objects based on time of flight (TOF) technology. The angular resolution of the laser scanner is dependent on the scanning frequency, which is configurable to 12.5Hz, 25Hz and 50Hz. At the default 12.5Hz frequency, an angular resolution of 0.5° is achieved on each layer. The scanning aperture ranges from 50° to −60°.

The product sheet states a maximum range of 50m with 10% remission and 160m for 100% remission surfaces (the light reflecting from a perfectly diffuse reflecting white surface corresponds to the definition of 100%). The data sheet further states a measurement resolution of 40mm and range error (1σ) of 100mm. These values are going to be verified in the following experiments.

Fig. 2. SICK LD-MRS laser scanner mounted on the SOL-Car

The laser scanner outputs both scan data as well as object data. The scan data is the raw measurement including the distance of each measurement point on each layer along with the echo pulse width in cm. One abstraction layer above, the object data contains the result of the detection and tracking of single objects out of the clouds of scan points. Objects are either localized by a reference point, their contour line or their bounding box. Fig. 4 shows an example of a static scan on a parking lot. The corresponding environment is shown in Fig. 3.

By using the object data from the sensor for relative positioning of vehicles the work of detecting and tracking objects is already done in the sensor, thus shortening the development time and decreasing the processing requirements on the application unit. In Fig. 4 it can be observed how clouds of scan points are clustered into single detected objects. The front line of the reflection points

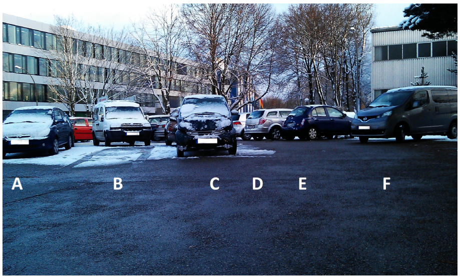


Fig. 3. View from the laser scanner

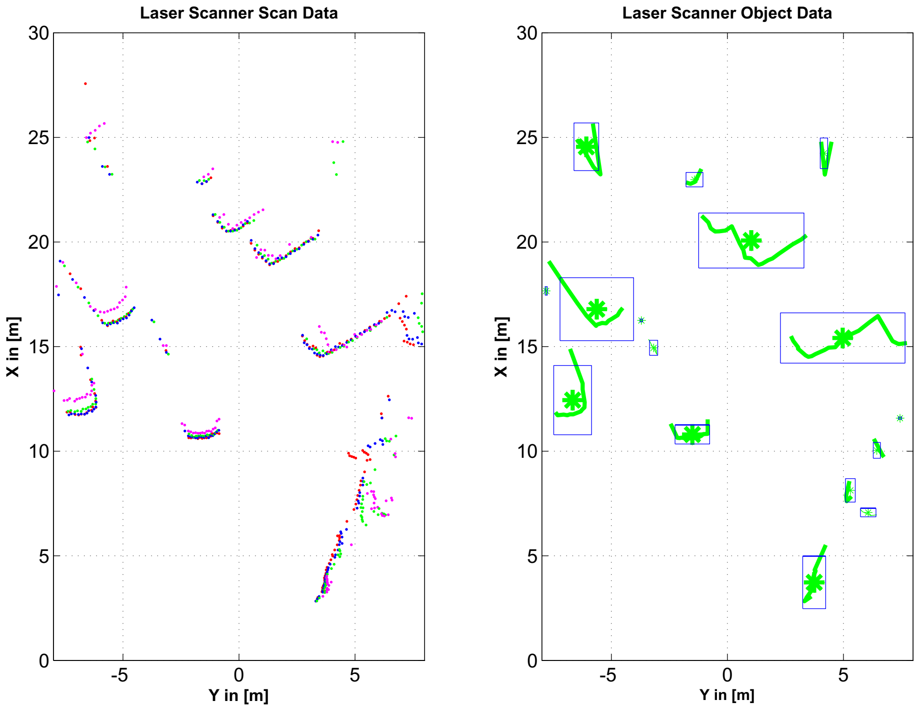


Fig. 4. Output of the SICK LD-MRS scanner: raw scan data and object data

generate the contour line (green line) and its center of mass is taken as the reference point for the object position (green asterisk). However, the usage of this data comes along with certain drawbacks that should be taken into account.

It can be noticed that nearby target objects might not be correctly resolved and merged into the same detected object (e.g. vehicles D and E). However, this behavior was rarely noticed in real driving environments. Fig. 4 also reveals a circumstance occurring when a target vehicle is in an oblique position to the sensor (e.g. vehicle F). Such geometry between two vehicles can be observed in real traffic in bends, roundabouts or intersections. The displacement of the center of mass of the contour line will introduce a systematic error in the object's position estimate, which will depend on the dimensions and the relative position of the target vehicle and its relative heading towards the detector vehicle.

Also in scenarios with an in-front placed target object a series of considerations have to be made. Depending on the shape of the rear end of the vehicle the measurement point might be displaced longitudinally by up to several decimeters. At short distances several scanning planes will hit the rear side of the vehicle. The resulting contour line will more likely coincide with its actual distance (e.g. vehicle B). However, when increasing the distance, only one of the scanning planes might hit the target vehicle and thus producing an unknown offset (see Fig. 5).

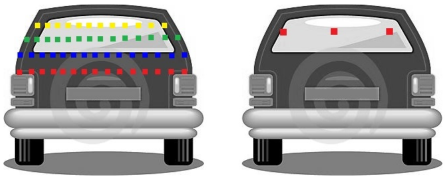


Fig. 5. Sketch showing the scanning planes hitting on the rear side of a vehicle. The left shows a nearby vehicle hit by four planes. the right picture shows a distant vehicle hit by only one plane.

Further on, with increasing distance the scanning points get laterally further separated leading to fewer points measuring the position of the target vehicle. Whilst at 20m distance and measuring with two planes the points are separated 8cm from each other, at 100m the points are as far as 87cm by measuring with only one plane. The number of laser points actually measuring the distance has a direct impact on the lateral position resolution.

Table 1 lists the theoretical expected error in lateral direction due to the limited angular resolution. The number of scanning planes and points per distance

Table 1. Theoretical maximum lateral error in dependence of the distance

Distance	10m	20m	30m	50m	80m	100m	120m
Number of Planes	4	4	3	1	1	1	1
Hor. Angular Resolution	0.25°	0.25°	0.25°	0.5°	0.5°	0.5°	0.5°
Hor. Point Distance	0.04m	0.08m	0.13m	0.43m	0.69m	0.87m	1.04m
Number of Points on Target	40	20	13	4	3	2	1
Maximum lateral Error at 0°	0.02m	0.03m	0.05m	0.19m	0.2m	0.4m	0.37m
Maximum lateral Error at 30°	0.03m	0.08m	0.15m	0.27m	0.40m	0.31m	0.50m

have been retrieved from a test run with a 1800mm wide and 1500mm high target vehicle. These values, along with the longitudinal position accuracy stated in the data sheet ($0.1m$) can be taken as the best performance that can be expected from the sensor.

The next section presents a series of measurement runs performed to determine the accuracy of the laser scanner sensor in terms of its bias and standard deviation.

4 Evaluation

A series of measurement runs have been performed to understand and characterize the behavior of the sensor in real world situations. The static measurements were performed with a vehicle equipped with a SICK LD-MRS four-layer laser scanner under test mounted on its front bump. The laser scanner is connected via Ethernet to the on-board automotive computer, which receives the detected objects at a rate of 12.5Hz. For each object its position in the form of x/y coordinates in the vehicle's body frame along with the estimated line representing its contour are output.

A reflector has been used as the target object. Its rectangular shape, with 1400mm width and 1500mm height, and the material, aluminum, resemble that of a vehicle. The plain surface of the reflector enables to measure the distance unambiguously, as the laser beam reflects directly on its surface, unlike a vehicle where the reflection point cannot be determined definitely.

The environment in which the measurement run was performed consisted of a straight single lane road inside the premises of DLR. The left side of the road is free of obstacles, while on the right side a metallic fence runs along the road. The fence, as well as other obstacles located behind the target vehicle that fell into the detection range of the laser scanner, produce objects that had to be filtered out manually. Several measurement procedures have been used to assess the laser scanner performance in such scenario.

Tape Measure. A tape measure was used to determine the distance between the reflector and the vehicle. The tape was laid perpendicularly to the reflector keeping the center of the laser scanner always on top of it. As the maximum length of the tape was 50m, 22 measuring point were taken between 0 and 40m. At each measuring point the vehicle was stopped between 5 to 15 seconds, which corresponds to 60 to 180 laser scanner object detections. For each point the distance between reflector and vehicle has been measured "by hand" using the tape measure. From the large list of detected objects the one corresponding to the reflector is taken, its distance is calculated using the law of cosines and compared to the distance measured by hand.

For each point measured with the tape measure the laser error bias and its standard deviation are calculated. Fig. 6 shows these values for the 22 measured points. The values lie between 1 and 6cm, what is as well inside the range of the stated laser scanner's resolution and in the order of the accuracy of the measurement technique.

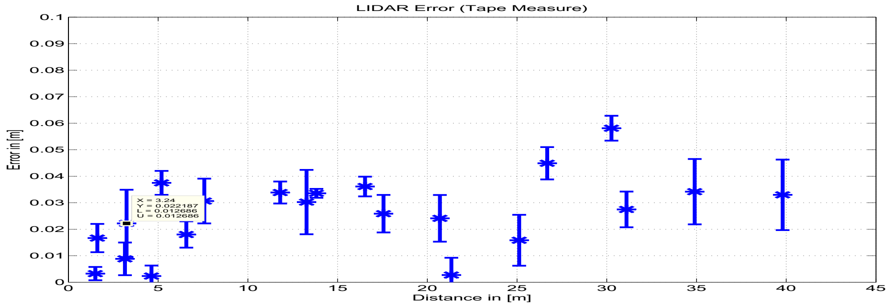


Fig. 6. Bias and standard deviation for different distances compared to a tape measure

GNSS Carrier Phase Measurement. A second test series using RTK carrier phase solution was chosen. On both the reflector and on the front of the detector vehicle a GNSS antenna was placed. The distance between the vehicle and the reflector was changed from 0 to 50m. At each measuring point the vehicle was stopped for several minutes to ensure that a carrier phase solution would be achieved during post processing. Also the movements between two different measurements were done smoothly and slowly to avoid losing carrier phase fix. For the absolute geo-position of the reflector in earth-centered-earth-fixed (ECEF) coordinates over the measurement period of 53 minutes a standard deviation of $\sigma_x = 0.0082m$, $\sigma_y = 0.0072m$ and $\sigma_z = 0.0120m$ was measured, suggesting a lock on the carrier phase. The average over the measurement period yields the position of the reflector. The vehicle's position is converted to the ECEF frame, the baseline is calculated by subtracting both coordinates and the norm of the vector yields the distance from the vehicle to the reflector. After subtracting the antenna offset to the laser scanner sensor (0.3m), both the distance between the correct detected object and the relative distance calculated by subtracting the RTK positions is compared to each other. 14 measurement points were taken from 1.5m to 45m distance. Fig. 7 shows again the resulting bias and standard deviation in each measurement. The figure shows laser scanner errors between 0cm and 8cm when compared to RTK. This error is in the order of the previous measurements performed with the measuring tape, thus validating both approaches and giving an idea of the expected laser scanner error in this range.

Tachymeter. A TCRP 1201 tachymeter from Leica has been used to measure the position of the target vehicle with respect to the laser scanner. Internally the tachymeter measures the range, azimuth and elevation angles towards a reflector point. To precisely orient the tachymeter a set of points have been taken on the detector vehicle. A prism was located on the roof on the rear part of the target vehicle. In post-processing the retrieved coordinates of the prism were translated and rotated to transpose them from the global coordinate frame to the sensor's coordinate frame of the laser scanner. The offset between the tachymeter and the most rear point of the vehicle (20cm) were corrected.

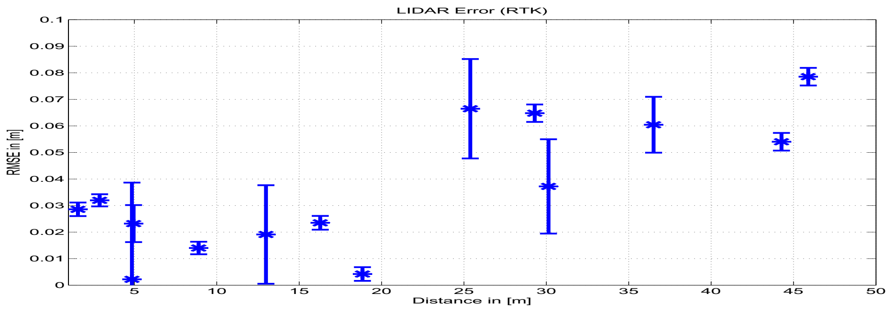


Fig. 7. Bias and standard deviation for different distances compared to the differenced RTK positions

The stated accuracy of the tachymeter (2mm in range and 0.27mdeg in angle) is achieved under static conditions, whereas in tracking mode, measurements in range and angle might not be synchronized and the resulting coordinates and their associated timestamps would be inconsistent [13]. For these reasons, only static phases where the target vehicle was standing still were taken for evaluation. Fig. 8 shows the along-track and cross-track components for the laser scanner (blue) and tachymeter (green) in dependence of time. The distance between the vehicles was increased stepwise from 6m to about 120m and decreased again down to 7m. The cross-track distance varied in a range of 1.5m.

The measurement run consisted of 35 points in the range of 0m to 120m. Fig 9 displays the resulting along-track bias and standard deviation for each of the measurements.

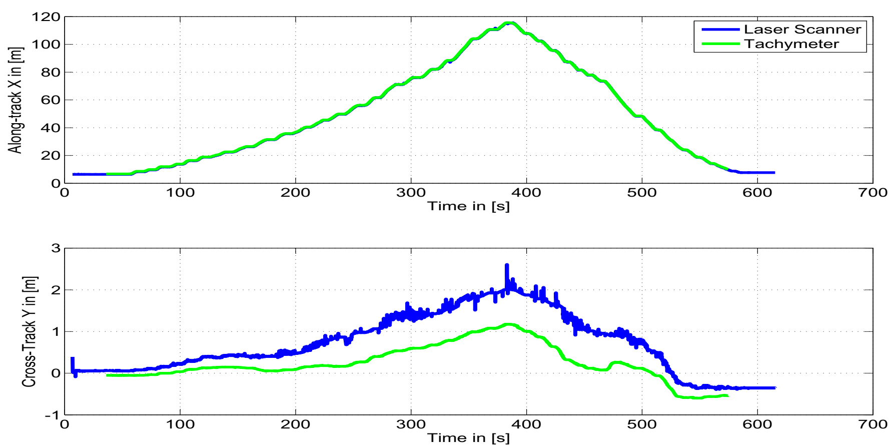


Fig. 8. Along-track and cross-track coordinate for laser scanner (blue) and tachymeter (green)

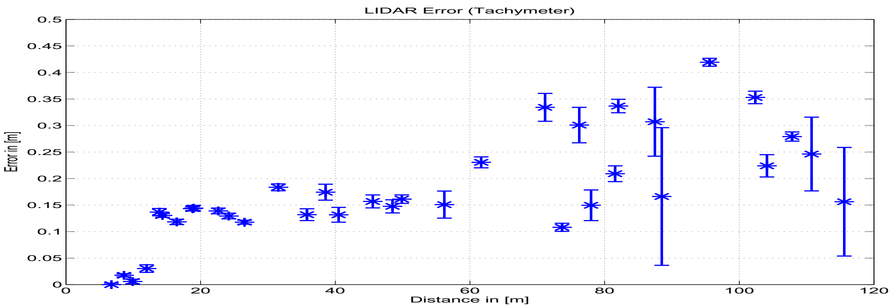


Fig. 9. 35 measuring points comparing the laser scanner along track coordinate to the correspondent point measured with the Leica Tachymeter

A minor offset of less than two centimeters can be observed at short distances. This offset is increased to about 15cm and is maintained from 15m to 60m. From this distance on, and coinciding with the measurement from only one laser plane, the offset varies from about 10cm to 40cm along with an increased standard deviation. The fact that this measurement run was performed using a vehicle as a target object instead of the reflector can explain the increased error in comparison to the previous experiments using the tape measure and and RTK. Further on, it can be observed that the cross-track error is marginal for short distances and increases with distance. At a distance above 100m this error is about 1m, matching the theoretical analysis performed in the previous sections.

5 General Conclusion

The authors conclude that the laser scanner is a convenient and relatively inexpensive solution to be used as a reference system for vehicle relative positioning. This paper has listed the particularities of the sensor, its drawbacks and special characteristics that have to be taken into account for its usage in static scenarios. The ranging accuracy of the laser scanner sensor has been characterized in static scenarios by using different techniques including a tape measure, a GNSS carrier phase based positioning engine and a tachymeter. It has been shown that the distance accuracy of the laser scanner is below 10cm in the range of 0 to 50m. The shape of the rear of the target vehicle introduces an unknown range ambiguity due to the variation in the measuring point. Further on, it has been shown that the limited angular resolution of the laser scanner leads to a lateral position error when increasing the distance.

To conclude we present the following recommendations when the object data from the laser scanner is to be used as a reference system for relative vehicle position:

- With distance the longitudinal accuracy of the laser scanner degrades. Up to 60m an accuracy of about 15cm could be verified.
- With distance the lateral accuracy degrades due to the limited angular resolution.
- In bends the object's center point is displaced from the vehicle's true center point. Therefore, we recommend using the device for tracking the in-front driving vehicle in straight maneuvers.
- Using a vehicle with a planar rear end is recommended to avoid errors caused to measurements at different heights.
- The use of wider vehicles is also recommended, in order to guarantee a certain lateral accuracy at higher distances.

Acknowledgements. The authors would like to thank EDMO-Flugbetrieb GmbH for the opportunity to run the experiments under almost ideal GNSS coverage conditions and for their safe coordination of accessing the runway alternating with normal air traffic.

References

1. Kaplan, E.: Understanding GPS - Principles and applications, 2nd edn. Artech House (December 2005)
2. Kukshya, V., Krishnan, H., Kellum, C.: Design of a system solution for relative positioning of vehicles using vehicle-to-vehicle radio communications during GPS outages. In: 2005 IEEE 62nd Vehicular Technology Conference, VTC 2005-Fall, vol. 2, pp. 1313–1317 (January 2006)
3. Rezaei, S., Sengupta, R.: Kalman filter based integration of DGPS and vehicle sensors for localization. In: 2005 IEEE International Conference Mechatronics and Automation, July 29-August 1, vol. 1, pp. 455–460 (2005)
4. Kim, Z., Cao, M.: Evaluation of feature-based vehicle trajectory extraction algorithms. In: 2010 13th International IEEE Conference on Intelligent Transportation Systems, ITSC, pp. 99–104 (September 2010)
5. Schubert, R., Richter, E., Wanielik, G.: Comparison and evaluation of advanced motion models for vehicle tracking. In: 2008 11th International Conference on Information Fusion, June 30-July 3, pp. 1–6 (2008)
6. Alam, N., Tabatabaei Balaei, A., Dempster, A.G.: Relative positioning enhancement in VANETs: A tight integration approach. IEEE Transactions on Intelligent Transportation Systems PP(99), 1–9 (2012)
7. Petovello, M., O'Keefe, K., Chan, B.: Demonstration of inter-vehicle UWB ranging to augment DGPS for improved relative positioning. In: Proceedings of IEEE/ION GNSS (2010)
8. Basnayake, C., Kellum, C., Sinko, J., Strus, J.: GPS-based relative positioning test platform for automotive active safety systems. In: Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation, ION GNSS 2006 (2006)
9. Luo, N., Lachapelle, G.: Relative positioning of multiple moving platforms using gps. IEEE Transactions on Aerospace and Electronic Systems 39, 936–948 (2003)

10. Huang, J., Tan, H.-S.: A low-order DGPS-based vehicle positioning system under urban environment. IEEE/ASME Transactions on Mechatronics 11, 567–575 (2006)
11. Schönberg, T., Ojala, M., Suomela, J., Torpo, A., Halme, A.: Positioning an autonomous off-road vehicle by using fused dgps and inertial navigation. In: 2nd IFAC Conference on Intelligent Autonomous Vehicles, pp. 226–231 (1995)
12. Travis, W., Bevly, D.: Trajectory duplication using relative position information for automated ground vehicle convoys. In: 2008 IEEE/ION Position, Location and Navigation Symposium, pp. 1022–1032 (2008)
13. Stempfhuber, W., Wunderlich, T.: Auf dem Weg zur Sensorsynchronisation von GPS und TPS für kinematische Messaufgaben. In: Allgemeine Vermessungs Nachrichten, AVN, vol. 111, pp. 175–184 (May 2004)

Estimating the Scheduling Discipline of an Ethernet Switch Using Constant Bit-Rate Probes

Kasper Revsbech, Tatiana K. Madsen, and Henrik Schiøler

Dept. of Electronic Systems, Aalborg University, Denmark
{kar,tatiana,henrik}@es.aau.dk

Abstract. In this paper we present a method to estimate the scheduling discipline applied on the output port on an Ethernet switch. We use well known formulas to calculate the expected output rate of a scheduler. Assuming that it is one of the following scheduling algorithms FIFO, Fixed Priority or Round Robin, we measure the actual output bandwidth and find the most likely scheduling algorithm. We present both a simulation study and a practical implementation where the proposed principle is implemented on a FPGA based high-pression Ethernet testbed.

Keywords: Safety-critical networks, Ethernet Modeling, In-car Networks, Bandwidth Measurement, Performance Evaluation.

1 Introduction

Until recently, safety critical systems and hard real-time systems have mostly relied on dedicated bus systems/networks to facilitate the communication between the nodes in the network. As an example the aviation industry often uses e.g. SpaceWire or Avionics Full-Duplex Switched Ethernet (AFDX). The automotive industry often uses a number of bus systems conjunction, e.g. CAN, FlexRay, LIN etc. Furthermore, TTEthernet, Media Oriented Systems Transport (MOST) Bus and Profinet are frequently applied in such systems. However, these technologies each have their advantages and disadvantages. In general the field-bus systems such as CAN and FlexRay are not applicable for high data rates. AFDX and TTEthernet provide high bandwidth, however the per-unit cost is still high. As a consequence the automotive industry among others are investigating the feasibility of using commodity Ethernet technology/hardware and the TCP/IP protocol suite, for in-car networks [1]. This is motivated by the possibility of constructing a network primarily based on one single technology, with a low unit cost and high bandwidth.

Introducing new technology and new protocols leads to a requirement of updating design and validation methods. Especially for safety-critical networks the validation of end-to-end delay guarantees is essential. A number of well established methods to validate end-to-end delay and backlog in safety critical systems

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 159–168, 2013.

already exists. As an example, Real-Time Calculus (RTC) is an analytical approach inspired by Deterministic Network Calculus (DNC), specially intended for design and validation of safety-critical systems [2]. In [3] we give an example of how RTC can be used to model an Ethernet/IP based in-car network. Furthermore, discrete-time event simulators such as Network Simulator (NS) 2, NS-3 and OMNET++ are often also used in the design process, as an example [4] uses OMNET++ to evaluate shaping methods in an Ethernet based in-car network. However, common for all modeling approaches is the need of precise model parameters such as service rates, frame forwarding delays and scheduling disciplines.

As the automotive industry seek to use off-the-shelves equipment such as commodity switches, they have to accept the fact that not all needed parameters will be shared by the vendor. In this work we present a method for deriving the scheduling discipline of the output queue in an Ethernet switch by use of two constant bit-rate flows and bandwidth measurements.

Estimating the performance/service for networks in the context of DNC is not a new concept. As an example Liebeherr et al. [5] introduce a foundational approach to estimate available bandwidth for min-plus linar systems, where well known probing methods are related to the min-plus system theory. Moreover, in [6] Liebeherr et al. present a concept called Δ-scheduler where a single closed form formula can be derived for the three scheduling disciplines: First In First Out (FIFO), Fixed Priority (FP), and Earliest Deadline First (EDF). However, the available literature on service estimation assumes that the cross traffic remains the same as when measured, which is perfect when estimating available bandwidth. However, in the design process of a network, it is necessary the be able to calculate delay bounds for various cross-load traffic scenarios. In this paper we propose a method for estimating the scheduling discipline from a set of known schedulers using constant bit-rate probes. This allows the designer to use the formulas already defined for a specific scheduler e.g. by use of the ones defined by Boudec in [7]. Furthermore, this allows the designer to model various cross loads.

The solution presented uses formulas derived for constant-bit rate flows, to pre-calculate the expected bandwidth for a given scheduler provided the input flows. By creating a scenario where we configure the bandwidth of the input flows and are capable to measure the bandwidth of the same flows leaving a scheduler, it is possible to derive the scheduling that took place. We show a simulation study in which we study the effect of using frames and not constant bit-rate flows. Furthermore, we describe the problems that can arrise in an environment with perfect timing. It is also shown how this method can be used in practice, by implementing it on a high-precision FPGA based Ethernet testbed.

In the remainder of this paper we first present the method to calculate the expected bandwidth followed by a simulation study of the proposed method. In Sec. 3, we show a practical implementation of the method followed by Sec. 4 which presents the conclusions and discussions.

2 Scheduling Estimation with Constant Bit Rate Probes

As shown in Fig. 1 a switch can be divided in three major components: *Input ports*, *Backplane*, and *Output ports*. As described in Sec. 1 a key component in constructing a switch model is the model of the output queue/scheduler. Here we show how the theoretical output rates for fluid flows can be used to estimate the scheduling discipline of an output queue. We assume that the scheduling discipline is one of the following: *First In First Out (FIFO)*, *Fixed Priority (FP)* or *Round Robin (RR)* as these are the most common used in commodity switches.

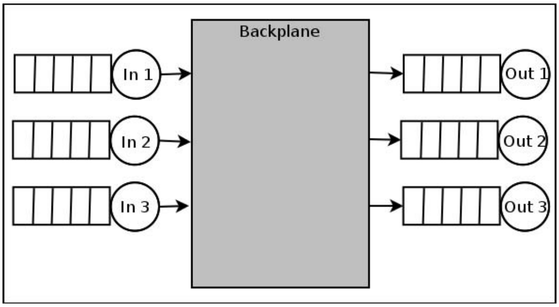


Fig. 1. A simple layout of a switch with input ports, a backplane and output ports

Assuming a fluid flow $\mathcal{F}_i$ with a constant arrival-rate r_i, a fluid cross flow $\mathcal{F}_j$ with a constant arrival-rate r_j and a processor with a constant processing rate C, it is straight forward to derive the equations to find the theoretical output rate r_i^* for the various scheduling disciplines:
FIFO:

$$r^*_{i,\text{FIFO}} = \begin{cases} r_i * \frac{C}{r_i+r_j} & r_i + r_j \geq C \\ r_i & r_i + r_j < C \end{cases} \tag{1}$$

FP - $\mathcal{F}_i$ has lowest priority:

$$r^*_{i,\text{FP-LP}} = C - r_j \tag{2}$$

FP - $\mathcal{F}_i$ has highest priority:

$$r^*_{i,\text{FP-HP}} = \begin{cases} r_i & r_i \leq C \\ C & r_i > C \end{cases} \tag{3}$$

RR

$$r^*_{i,\mathrm{RR}} = max\left\{C - r_j; \frac{C}{2}\right\} \tag{4}$$

Using Eq. (1) - (4) it is possible to pre-calculate the theoretical output-rate for the three scheduler types, given two input flows $\{\mathcal{F}_i, \mathcal{F}_j\}$. Hence by choosing r_i and r_j and knowing/estimating C, an output-rate of each scheduling discipline can be calculated. Thereby by comparing the measured output bandwidth $\tilde{r_i}^*$ with the theoretical r_i^* per scheduler, it can be estimated which scheduler type is most likely. Note that we have four possible outcomes of r_i^* for the three scheduling cases, as the outcome of FP scheduling depends on the priority.
Note that for Eq. (2) and (4), where the output rate of r_i does depend on the input rate, the output rate can not be higher than the input rate.

As the processing rate C of the processor is a parameter to the model, it is necessary to know it before the scheduling estimation can be performed. The capacity can either be estimated empirically by probing with a constant bit-rate, or assumed to be the same as the link speed. The probing can be constructed in a way where a constant bit-rate flow $\mathcal{F}_p$ is sent through a switch with rate r_p, then r_p is incremented until $r_p > r_p^*$ [8]. However, in practice modern hardware switches are cable of serving flows at line speed, hence it is a fair assumption that the capacity of the output queue/scheduler is the link capacity. Naturally, it should be tested by performing the afore mentioned rate scan, which if the assumption is true, would show that $r_p = r_p^* = L_c$, where L_c is the link capacity. The input rate r_p also is upper limited by the link capacity L_c.

Delimiting the estimation to the three scheduler classes (FIFO, FP, and RR), four possible theoretical output rates, for a given input flow $\mathcal{F}_i$, can be observed: $r^*_{i,FIFO}$, $r^*_{i,FP-LP}$, $r^*_{i,FP-HP}$, and $r^*_{i,RR}$. As seen from Eq. (1) - (4) it is sufficient to use two input flows, to infer the scheduling principle. To ensure maximum separation between the possible output rates it has been chosen to let the measurement flow $\mathcal{F}_i$ arrive with a constant bit rate $r_i = C$ and find the cross flow rate r_2 that maximises the "distance" between the theoretical output rate of $\mathcal{F}_i$. Hence by solving Eq. (5)

$$r_j = \underset{0 \le r_j \le C}{\arg\max}\left\{\left(\frac{1}{|d_1(r_j)|} + \frac{1}{|d_2(r_j)|} + \frac{1}{|d_3(r_j)|}\right)^{-1}\right\} \tag{5}$$

where, d_1, d_2 and d_3 are the distances between the output rates as a function of (r_j):

$$d_1(r_j) = r^*_{i,\text{FP-HP}}(r_j) - r^*_{i,\text{FIFO}}(r_j) \tag{6}$$

$$d_2(r_j) = r^*_{i,\text{FIFO}}(r_j) - r^*_{i,\text{RR}}(r_j) \tag{7}$$

$$d_3(r_j) = r^*_{i,\text{RR}}(r_j) - r^*_{i,\text{FP-LP}}(r_j) \tag{8}$$

we find the optimal crossflow rate (r_2).

2.1 Simulation

As described above Eq. (1) - (4) assume constant bit-rate, fluid flows. However, as Ethernet is frame based the flow will not be fluid. Hence, a simulation study using the described method has been conducted, with different frame-sizes to check the effect of framed flows vs. fluid flows. The proposed method has been applied in an OMNET++ [9] simulation, using the INET [10] framework. A scenario with three nodes and one switch has been created as depicted in Fig. 2, where *node 1* and *node 2* each emits one flow dedicated for *node 3*. Hence, relating to Fig. 1 *node 1* is connected to *In 1* on the switch, and *node 2* to *In 2*, and both flows are send to *Out 1* where they are scheduled according to a scheduling algorithm, which is the one we are estimating. The switch has been configured to employ the various scheduling algorithms in the output scheduler connected with *node 3*. The network is configured as a 100 Mbit/sec full duplex Ethernet network, assuming a lossless channel. Flow $\mathcal{F}_1$ from *node 1* to *node 3* is the flows measured, and has a constant rate of (r_1) 100 Mbit/sec. It has been chosen to conduct the same simulation scenario with two different frame sizes to see if this packetization has an impact using the formulas assuming constant bit-rate. In Table 1 we present the measured bandwidth for the different scheduling policies where both $\mathcal{F}_1$ and $\mathcal{F}_2$ have been configured with a frame size of either 100 B or 1400 B. In both cases the cross flow $\mathcal{F}_2$ has been configured to have a rate r_2 of 67 Mbit/sec, as this is the optimal rate for this scenario found by Eq. (5). As seen from Table 1 the measured and expected bandwidth for both frame sizes remains the same, hence we conclude that there is no significant effect on the measured bandwidth compared to the expected bandwidth due to framing.

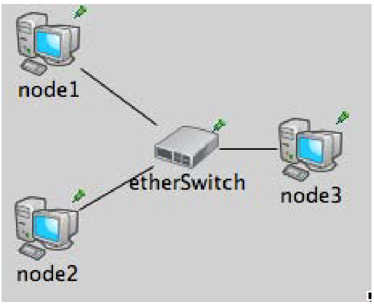


Fig. 2. OMNET ++ network setup

In the simulation study we observed an important detail which one has to be aware of when using the proposed method in a real-life setting. The results presented in Table 1 are generated when the output queue in the switch connected to *node 3*, is configured to be so large that no frames are dropped. If we allow tail dropping in the queue, it should in principal be "fair" and the probability of drop for a certain flow should be proportional to the input rate. However, as

Table 1. Expected and measured bandwidth from the simulation. Where r_1^* is the expected/theoretical bandwidth of $\mathcal{F}_1$ and $\tilde{r}_1^*$ the measured bandwidth.

Scheduling Discipline	r_1^*	$\tilde{r}_1^*$ [100 B]	$\tilde{r}_1^*$ [1400 B]
FIFO	59.9	59.88	59.88
FP (HP)	100	100	100
FP (LP)	33	33	33
RR	50	50	50

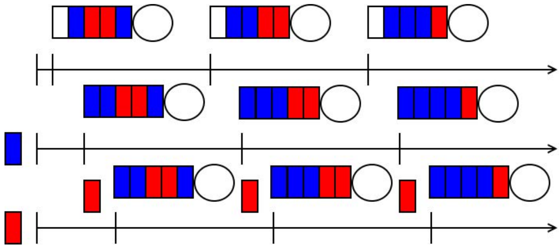


Fig. 3. Example of a FIFO scheduler with a drop tail queue, with synchronized arrival flows

we have configured the flows as constant bit-rate flows with deterministic Inter Frame Gap (IFG) the tail dropping will depend highly on the original offset between the frames. In Fig. 3 an example is shown with a constant offset between the two flows. In this case the IFG is the same, hence the "Blue" will get the empty slot first all the time. As a further example Table 2 presents the measured bandwidth for the same simulation scenario as described above just with a finite queue-size of 100 frames, and for the frame size of 100 B. As seen from the results the measured bandwidth for the FIFO scenario is the only one affected. This is due to the fact that in FIFO the queue is shared for all flows, where a (conceptual) queue is used per flow for the other disciplines. For the simulation the solution to prevent this behavior simply is to increase the buffer to ensure that tail-dropping is not happening. However, in real-life this is not an option, thus one must design the experiment with this in mind. In Sec. 3 we present our real-life setup and how we handle this effect.

Table 2. Expected and measured bandwidth from the simulation, as in Table 1 but with finite queue length

Scheduling Discipline	r_1^*	$\tilde{r}_1^*$ [100 B]
FIFO	59.9	67.0
SP (HP)	100	100
SP (LP)	33	33
RR	50	50

Having presented the equations for calculation of the estimated output rate as well as simulation results for the presented method, is seems convincing that the proposed method is valid in theory for a controlled simulation environment.

3 Real-Life Implementation

In this section we describe how the estimation method in Sec. 2.1 can be implemented in a real life equipment.

A manageable switch where we can control the scheduling algorithm on the output port, has been chosen for the experiment. In this case it is a *Cisco Small Business SG 300-10* switch, which has 10 Ethernet ports each operating at up to 1 Gb/sec. The generation and measurement of the measurement flow $\mathcal{F}_1$ and the cross flow $\mathcal{F}_2$ are done with a special purpose FPGA based high precision testbed presented in [11]. The testbed is used as it offers high precision both generating and receiving frames, which enables us to send with a precise bandwidth and to measure the received bandwidth with a high accuracy. As presented in [11] the max deviation of the IFG between two frames send, is measured to 90 ns. Furthermore, the testbed is capable to measure the frame arrival time with a precision of ± 55 ns.

The test follows the same principle as explained in Sec. 2.1 where two flows ($\mathcal{F}_1$ and $\mathcal{F}_2$) traverse the switch, and share the same output buffer/scheduler. Relating to Fig. 2, $\mathcal{F}_1$ is entering the switch at *In 1* and $\mathcal{F}_2$ at *In 2*. Both flows leave the switch at *Out 2* which is the scheduler under test. As the Switch operates at 1 Gb/sec, we choose $r_1 = 1$ Gb/sec and using Eq. (5) we find that r_2 should be 670 Mb/sec. Both flows have been configured with a frame size of 1400 B on the wire (Including the Ethernet preamble). However, as indicated in Sec. 2.1 the frame size should not have any practical effect. The switch is preconfigured to a known scheduling algorithm, and the frames are marked, before sending, using Differentiated Services Code Point (DSCP). This allows the switch to separate the flows, and act according to the preconfigured scheduling algorithm.

As described in Sec. 2.1 the simulation results are obtained in a setting where the queue on the output buffer in the switch is configured to be so large that no frames are dropped due to tail dropping. As explained this is done to prevent the effects of an "unfair" drop due to the strict periodic frame generation in the simulator. Normally this problem/effect will not be significant in real-life settings as the frame generators impose a high jitter. However as we use a FPGA based frame generator there are effectively no jitter and the offset between frames of $\mathcal{F}_1$ and $\mathcal{F}_2$ will remain the same throughout an experiment. This is not only true for the Napatech [12] adapter used in our testbed but also if e.g. the open source NetFPGA [13] adapter is used together with their packet generator [14]. Naturally it is not possible in a real-life switch to configure the buffer such that it would be sufficiently large to hold all frames. It would be possible to let the IFG between frames follow a random distribution e.g. Poisson to introduce jitter. However, that would require that Eq. (1) - (4) are proven for stochastic flows instead of constant bit-rate flows. Instead we propose to use packet trains for the measurement flow $\mathcal{F}_1$ where the length of the train is small enough not to fill the queue. To achieve a sufficiently large number of samples we propose to run each test a number of times.

To find a suitable frame train size we also use the testbed and a setup as described in Sec. 2.1 where the two flows $\mathcal{F}_1$ and $\mathcal{F}_2$ traverse the same switch and the same output port of the switch. Flow $\mathcal{F}_2$ is configured to send 1400000 frames at 1 Gb/sec with a frame size of 1400 $Bytes$, which means that $\mathcal{F}_2$ is active for 15 sec. While $\mathcal{F}_2$ is active, $\mathcal{F}_1$ is started also with a frame size of 1400 and a rate of 1 Gb/sec, however, with various train sizes. In this way we can "count" the number of received frames and if the sum is less than the aggregate of the number of frames sent for $\mathcal{F}_1$ and $\mathcal{F}_2$, the packet train was so large that the queue was filled and frames were dropped. In Table 3 we show the results with packet trains of $\{10, 20, 30, 40, 50, 60\}$ frames, each test repeated 5 times. As seen from the table the buffer starts to drop frames when the packet train contains 50 frames or more, thus we have chosen to use packet trains with 40 frames to ensure no losses.

Table 3. Received number of frames with packet trains of varios size

Packet train size [Frames]	**Expected frames**	**Received frames**
10	1400010	1400010
20	1400020	1400020
30	1400030	1400010
40	1400040	1400040
50	1400050	1400045
60	1400060	1400045

Table 4. Measured and expected bandwidth measured by the testbed

Scheduling Discipline	r_1^* $[Mb/sec]$	$\tilde{r}_1^*$ $[Mb/sec]$
FIFO	599	592
SP (LP)	330	327
SP (HP)	1000	956
RR	500	498

In Table 4 we show the measured and estimated bandwidth for the experiment described above, where the measured bandwidth is the mean of 100 repetitions. Furthermore, in Fig 4 the measured and the expected bandwidth are plotted and a 95% confidence interval is shown for the measured bandwidth. As seen from Table 4 and Fig. 4 the estimated/expected bandwidth is close to the measured bandwidth for all scheduling classes, which indicates that the proposed method can be used, also in a real-life setting. It can also be seen that for the case where the scheduler is prioritized and $\mathcal{F}_1$ has the highest priority the estimated

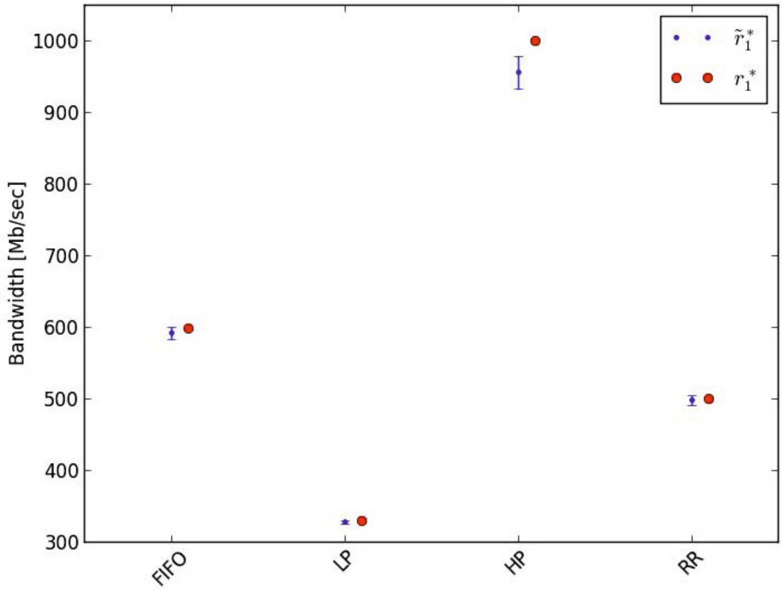


Fig. 4. Measured and expected bandwidth plotted with a 95% confidence interval

bandwidth is deviating more than for the other disciplines. We believe this is mainly due to Head-of-line Blocking (HOL) from the lower priority flow.

4 Conclusion and Discussion

In this paper we have presented a method to derive the scheduling algorithm of an Ethernet switch, which is one of the key parameters needed to model it. We have presented a simulation study where it is seen that the frame size has no practical effect, and that using synchronized periodic packet generation can lead to an "unequal" drop effect in the tail drop queue. Furthermore we have presented the results where the proposed method has been tested with a *Cisco Small Business SG 300-10* switch using an FPGA based testbed to generate the packet flows and to measure the bandwidth. The results indicate that the proposed method is also applicable for a real-life switch. In this work we presented the method for FIFO, FP and RR scheduling algorithms; future work will include other scheduling algorithms. Furthermore, a natural extension is to design methods to derive other important model parameters, e.g. a method to estimate the frame processing time for a switch.

In this work we have used an high precision FPGA based Ethernet testbed, which enables us to conduct experiments with high precision, both when generating frames, and measuring the bandwidth. An interesting future study is to quantify the effect of applying such a method in commodity off-the shelves hardware and software, using higher level packet generators and commodity NIC's to capture the traffic and hence measure the bandwidth. Naturally variation of the measurements will increase due to higher "noise" in the measurement, but methods could be implemented to account for this.

Acknowledgement. The authors of the paper would like to thank professor Markus Fidler for his valuable inputs and ideas for this work.

References

1. Steffen, R., Bogenberger, R., Hillebrand, J., Hintermaier, W., Winckler, A., Rahmani, M.: Design and realization of an ip-based in-car network architecture. In: ISVCS (2008)
2. Wandeler, E., Thiele, L., Verhoef, M., Lieverse, P.: System architecture evaluation using modular performance analysis: a case study. International Journal on Software Tools for Technology Transfer 8, 649–667 (2006)
3. Revsbech, K., Schiøler, H., Madsen, T.K., Nielsen, J.J.: Worst-Case Traversal Time Modelling of Ethernet Based In-Car Networks Using Real Time Calculus. In: Balandin, S., Koucheryavy, Y., Hu, H. (eds.) NEW2AN/ruSMART 2011. LNCS, vol. 6869, pp. 219–230. Springer, Heidelberg (2011)
4. Rahmani, M., Tappayuthpijarn, K., Krebs, B., Steinbach, E., Bogenberger, R.: Traffic Shaping for Resource-Efficient In-Vehicle Communication. IEEE Transactions on Industrial Informatics 5, 414–428 (2009)
5. Liebeherr, J., Fidler, M., Valaee, S.: A System-Theoretic Approach to Bandwidth Estimation. IEEE/ACM Transactions on Networking 18(4), 1040–1053 (2010)
6. Liebeherr, J., Ghiassi-Farrokhfal, Y., Burchard, A.: Does Link Scheduling Matter on Long Paths? In: 2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 199–208 (2010)
7. Le Boudec, J.-Y., Thiran, P.: Network calculus. In: Thiran, P., Le Boudec, J.-Y. (eds.) Network Calculus. LNCS, vol. 2050, pp. 3–81. Springer, Heidelberg (2001)
8. Jain, M., Dovrolis, C.: Pathload: A measurement tool for end-to-end available bandwidth. In: Proceedings of Passive and Active Measurements (PAM) Workshop (2002)
9. Omnet ++, `http://www.omnetpp.org/`
10. Inet, `http://inet.omnetpp.org/`
11. Revsbech, K., Madsen, T.K., Schioler, H.: High Precision Testbed to Evaluate Ethernet Performance for In-Car Networks. In: ITST (2012)
12. Napatech homepage, `http://www.napatech.com/`
13. The netfpga project homepage, `http://netfpga.org`
14. Covington, G., Gibb, G., Lockwood, J., Mckeown, N.: A Packet Generator on the NetFPGA Platform. In: 17th IEEE Symposium on Field Programmable Custom Computing Machines, FCCM 2009, pp. 235–238 (2009)

Bridging Physical and Digital Traffic System Simulations with the Gulliver Test-Bed*

Christian Berger[1], Erik Dahlgren[2], Johan Grunden[2], Daniel Gunnarsson[2], Nadia Holtryd[2], Anmar Khazal[2], Mohamed Mustafa[2], Marina Papatriantafilou[2], Elad M. Schiller[2], Christoph Steup[3], Viktor Swantesson[2], and Philippas Tsigas[2]

[1] Göteborg University, Göteborg, Sweden
christian.berger@gu.se
[2] Chalmers University of Technology, Göteborg, Sweden
{mohmus,ptrianta,elad,tsigas}@chalmers.se,
{erikda,it4grjo,dangunn,holtryd,khazal,vikswa}@student.chalmers.se
[3] University Magdeburg, Magdeburg, Germany
steup@ivs.cs.uni-magdeburg.de

Abstract. We propose a cyber-physical platform that combines road traffic simulation, network simulation, and physically simulated vehicles to facilitate extensive testing on various levels of vehicular systems. Our design integrates physical and digital vehicle simulation into a common development and testing environment. This paper describes the platform design and presents prototypical implementations that use Simulator of Urban Mobility (SUMO), TinyOS Simulator (TOSSIM), a 3D sensor simulation environment, and a test-bed of miniature vehicles called Gulliver. As a prototypical implementation, we demonstrate the development of cooperative applications, and by that we achieve: (a) a cyber-physical system that provides a common environment for physically and digitally simulated vehicles, (b) a platform to interface communication between physically and digitally simulated vehicles, and (c) the ability to tailor testing scenarios in which some system components are simulated digitally and some physically.

The suggested design provides flexibility, cost efficiency, and scalable testing opportunities for future vehicular systems. Furthermore, the proposed system is able to support novel steps towards intelligent transportation systems for smart cities.

1 Introduction

Modern vehicular systems require extensive testing to ensure the safety and reliability of active safety and driver assistance systems. New ideas and first concepts of new systems that rely heavily on data from the surroundings are not

* The authors appear in alphabetical order. This work was partially supported by the EC, through project FP7-STREP-288195, KARYON (Kernel-based ARchitecture for safetY-critical cONtrol).

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 169–184, 2013.

only time-consuming but also resource-intense with respect to testing equipment and proving grounds. Thus, designers of vehicular systems often use simulation tools for validating and testing systems behavior before carrying out the needed and often costly test-runs in proving grounds. Currently, universities and public research and engineering institutes are faced with the challenge to get access to testing ground facilities, and are most often out of the development and prototyping cycle. Therefore, we consider an inexpensive test-bed of miniature vehicles named Gulliver [18] for validating new ideas in vehicular systems. [1] In this paper, we present a cyber-physical platform that bridges between physical and digital simulations, such that a small set of physically simulated vehicles can coexist with a larger set of digitally simulated vehicles in a common testing environment. This enables simpler and faster development and testing of proof-of-concept prototypes, such as driver assistance mechanisms and traffic control, in addition to experiments of a larger scale, where traffic phenomena can be observed and investigated on physically simulated vehicles. We expect the proposed platform to influence the development and testing of full-scale vehicles in testing grounds used to validate purely digitally simulated results.

The physical simulation part uses the Gulliver platform, which is a test-bed of miniature vehicles that facilitates the validation of new ideas in vehicular systems. The miniature vehicles have on-board sensors and radio and are capable of autonomous driving. The vehicles can be remotely accessed to upload their routes and their movement through a route is traced and visualized. The network simulations are performed using TinyOS Simulator (TOSSIM), a discrete event simulation tool used in wireless sensor network simulations [15]. We use a microscopic traffic simulation tool, namely Simulator of Urban Mobility (SUMO) [11], for digital road traffic simulations. For digitally simulating raw sensor data, we use a 3D simulation environment described in [4]. Having the ability to bridge these physical and digital simulators provides a platform that can give researchers powerful tools for testing and evaluating new ideas. Fig. 1 depicts the proposed platform.

1.1 Related Work

Cyber-physical systems target varying areas from water distribution modeling [16], to medical applications [12], and data center performance and energy management [6], to name a few. Several works have studied digital vehicular systems by considering digital road traffic and network simulators [1, 2, 8, 21]. While these works can benefit from digital simulations before their physical deployment, the lack of physical validation in these works limits applications development to digital models of the physical world. Considering only a physical test-bed, as in [25], to demonstrate new ideas is often very complicated due to a number of constraints that one has to account for at the beginning of the development phase. We propose a way to simplify the development and testing process.

[1] The Gulliver test-bed web-site is accessible via `http://www.gulliver-testbed.net/`

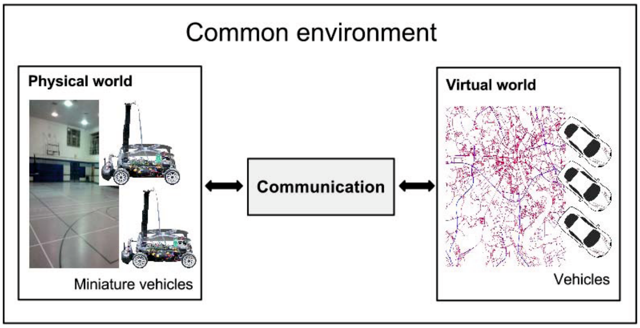


Fig. 1. Illustration of the Gulliver test-bed as a cyber-physical platform, where physical and virtual worlds communicate in a shared environment

The authors of [10] consider bridging digital and physical simulations, where full-scale vehicles are used on public roads. The work in [10] is limited to collecting and replaying empirical data traces for radio channel analysis. Furthermore, we note the costs of test ground facilities and their lack of availability for a broad range of research institutes. We are not the first to study combined digital and physical simulations, e.g., [9, 19, 20], we are the first to consider a platform in the context of experiments that consider traffic behavior, in addition to the simulation of the vehicular systems and networks.

1.2 Our Contribution

We propose a cyber-physical platform that brings together road traffic simulation, network simulation, and physically simulated miniature vehicles to facilitate rapid application development and extensive testing. The platform is based on the Gulliver test-bed in a way that integrates digitally and physically simulated miniature vehicles into a common environment, see Fig. 2.

In the proposed platform, the physical and the virtual worlds must be able to interact. The proposed platform includes the ability to generate maps so that physically and digitally simulated vehicles perceive the same surroundings. Moreover, a digital representation of each physically simulated vehicle is presented in the virtual world. This facilitates interaction among digitally and physically simulated vehicles. This is done by a communication middleware that facilities a bridge between simulators and the test-bed.

The result presented in this paper shows how to use the Gulliver test-bed for faster development and simpler testing of proof-of-concept prototypes and allow experiments that consider more vehicles than the test-bed's physically available vehicles. We review a number of case studies that exemplify the development process of different applications, such as autonomous driving, adaptive cruise control, crash avoidance, virtual traffic light and more. In each case study, we highlight our development steps and provide details to how we simplify the development environment (see Section 3).

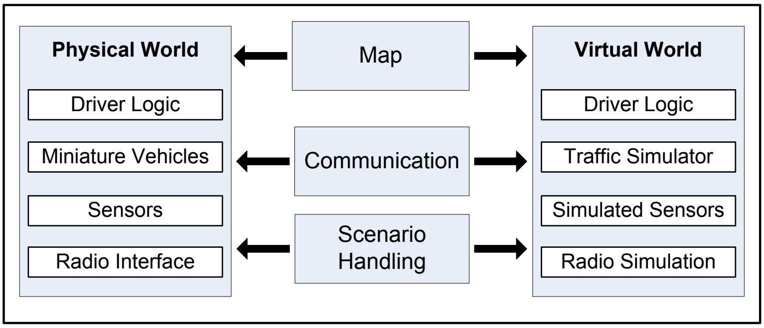


Fig. 2. Detailed logical overview of the proposed cyber-physical platform

2 The Gulliver Test-Bed

The tested includes a system of N vehicles, where each vehicle, $v_i \in V$, is represented by a digital vehicle, $v_i \in V_{ds}$, and possibly also by a physically simulated one, $v_i \in V_{ps}$. We note that each physically simulated vehicle, $v_i \in V_{ps}$, has also digital representation, $v_i \in V_{ds}$, in the virtual world.

2.1 Mapping and Consistency

We use a two-dimensional map for positioning. The map consists of roads that are defined as vectors (*segments*) interconnected via nodes to create a road network. We use the microscopic traffic tool, Simulator of Urban Mobility (SUMO) [11], which uses XML files to define scenarios and road maps. This feature is used to let each physically simulated vehicle, $v_p \in V_{ps}$, navigate through the same map like the simulated vehicles, $v_d \in V_{ds}$.

The first step is to define the map in a *Map Client*, which is provided by the Gulliver test-bed. The map client provides a Graphical User Interface (GUI) that allows construction of scenarios and road networks. Once the map is defined, the GUI automatically generates XML files, which are used directly by SUMO. Note that the GUI uploads the same map to every physically simulated vehicle. During simulation, SUMO has a representation of every physically simulated vehicle, $v_p \in V_P$, in the platform. This facilitates several benefits apart from a common map, such as an on-going overview of the simulation in SUMO-GUI and the possibility to create detailed logs.

2.2 Radio-Based Ranging for Indoor Localization

The miniature vehicles determine their position by filtering internal data, such as the distance traveled and the steering angle, vehicle's current heading provided by the magnetometer, and external data in the form of the distance measured to a predefined anchor using Kalman filter. The external data is provided by a high precision positioning system that uses PulsON 410 RCM as radio-based ranging for indoor localization. [2] It consists of three stationary positioning modules, which are placed in the test area. The distance between the modules is measured and fed to the map client, which defines the grid. The use of three modules provides the ability to define a 2-dimensional grid. A positioning module is then attached to every physically simulated vehicle. The modules communicate at 4.3 GHz and perform two-way time-of-flight ranging technique for obtaining their ranging measurements. Using these measurements the vehicles perform triangulation and receive their position. This approach provides precision in few centimeters, however it is subject to communication delays. Therefore, we have devised a dedicated medium access algorithm [23], and used internal input parameters from vehicle actuators to provide an estimation of vehicle position, e.g., dead reckoning [7]. We combine the benefits from both approaches, see [23] for details.

2.3 Miniature Vehicles for Physical Simulation

We use an open-source platform of miniature vehicles that consists of 1:8 and 1:10 scale miniature vehicles [18]. All vehicles are equipped with IR-distance and ultrasonic sensors for collision avoidance and some exhibit also monocular vision systems. High precision positioning system [23] provides vehicles with their absolute location on the map. The vehicles are fitted with several computer systems, namely an engine controller, a main board that connects and commands peripheral subsystems, MicaZ mote [22] that provides radio communications and either a mini-ITX based PC with an attached WiFi card or an ARM-based PandaBoard ES board. The mini-ITX based PC as well as the PandaBoard ES provide several convenient abilities, such as video streaming, remote upload of maps and binaries to the vehicle, and the possibility for a human driver to remotely control the vehicle through the use of a joystick. This data is sent over WiFi to a server program running on the vehicle.

Currently, the test-bed includes eight miniature vehicles that are available for physical simulation, but so far this number is limited mainly by the fact that the system is at its prototypical stage. The number of vehicles that is available for digital simulation is bounded mainly by memory and processing constraints of the simulation server.

2.4 Vehicles for Digitally Simulation

Digitally simulated vehicles enhance the test-bed scale by considering a greater number of vehicles than the physical ones. They also provide an ability to simplify the software development and testing processes, as the case studies show (Section 3). Some of the digital simulation is generated by the Simulator of Urban Mobility (SUMO) [11]. SUMO allows the use of a variety of variables for defining the driving quality. The system connects SUMO to a sensor networks

[2] `http://www.timedomain.com/p400.php`

simulator, TinyOS Simulator (TOSSIM) [15], in order to facilitate radio environment simulation. In TOSSIM, a digitally simulated node is created for every vehicle in SUMO. The digitally simulated nodes hold properties such as position, speed, and map related information (*segment*). These properties are provided by SUMO via the TraCI interface [26], and are used in generating radio link gains among nodes in TOSSIM. By having vehicles being represented as nodes in TOSSIM, the vehicles' radio communication can be simulated. The access to variables, such as position and speed, together with radio communication ease the implementation of several driver assistance mechanisms within TOSSIM that affect traffic behavior in SUMO.

The digitally simulated vehicles can also be equipped with functionalities similar to the one in the physically simulated ones. The examples of functionalities that we consider are based on radio communication and implementations of driver assistance mechanisms, such as adaptive cruise control and collision avoidance. We note that the exact set of digitally simulated vehicles that have these abilities can be defined in a way that resembles testing of new technologies on the open road, where prototype vehicles coexist with legacy ones. Furthermore, some digitally simulated vehicles have the ability to provide raw sensor input data for camera, ultra sonic, infrared, single layer laser scanner sensor, to name a few.

2.5 Radio Communication

Inter-vehicle communication allows vehicles to exchange useful data among themselves. Radio communication is enabled via motes that are TinyOS portable, such MicaZ [22], which use direct sequence spread spectrum for transmitting data at 2.4 GHz. Our experiments sometime use Chameleon-MAC [13, 14], which employs the scheduled approach when accessing the communication medium, rather than the random access approach, such as Carrier Sense Multiple Access (CSMA). This choice is explained by CSMA's unbounded delay in high contention periods, which leads to high packet drop as shown in [5, 24]. Chameleon-MAC serves also as the bases for vehicles to synchronize via TDMA frame alignment, see the Grasshopper algorithm [17].

The common environment created by the proposed design promotes communication across the digital and the physical simulation domains. We implement a repeater to provide this functionality. The repeater captures messages from the radio environment and injects them into simulated vehicles within sender's range. Moreover, when a simulated vehicle is transmitting to a physically simulated vehicle, the message is injected into the repeater, which then broadcasts it over the radio.

2.6 Middleware for Communication Handling

To connect the different subsystems of the proposed design a middleware for *Communication Handler* (CM) was implemented, see Fig. 3. The communication handler maintains a link to every subsystem during a simulation run. It works

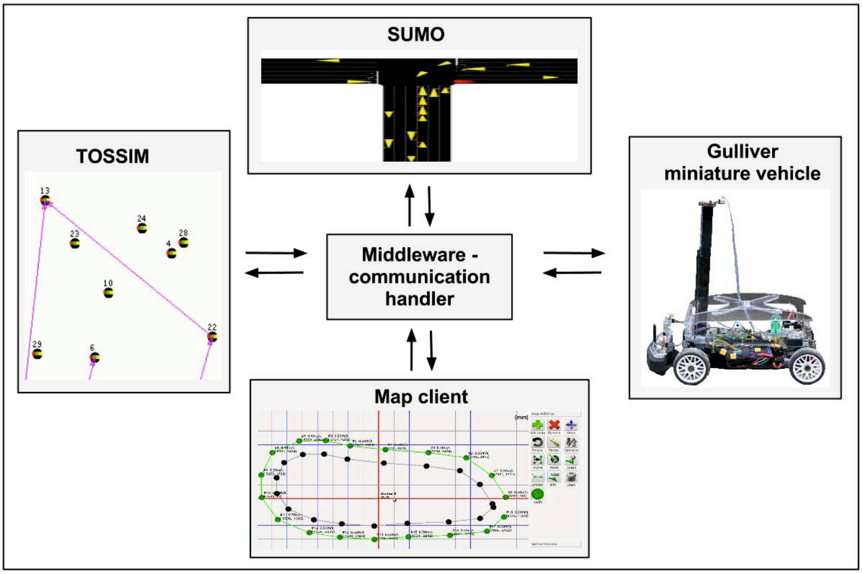


Fig. 3. The proposed platform design with arrows demonstrating communication within the system

like a packet forwarding router, where packets are sent based on a set of rules, e.g., different packets are to be sent to different destinations. The connections to different subsystems are set up in the following manner:

– CM's link with SUMO uses the traffic control interface TraCI [26], which allows control over SUMO's digitally simulated vehicles during simulation.
– CM's link to TOSSIM is based on the *serial-forwarder* (SF) extension, which enables bidirectional communication to nodes within the TOSSIM environment. In practice, the SF is a TCP server that delivers packets to nodes in TOSSIM based on their node ID. Communication in the other direction works in a similar way; broadcasted packets are sent to the SF and later received by CM.
– CM's link to the physically simulated vehicles uses a physical MicaZ mote connected, via USB, to the PC that is running the simulation. The MicaZ node works as a two-way repeater, sending packets received from the physical radio to CM, and packets from the CM to the physical radio, see the Broadcasting Unit box in Fig. 4.

We note that, at this platform prototyping stage, the communication loads and the speed of the physically simulated were kept low so that the digital simulator could keep up with the physical simulator.

3 Case Studies

We show how to simplify the design, development, and evaluation process by reviewing a number of applications that we developed for the Gulliver test-bed. The applications are developed using an iterative development process. In each iteration, a prototype is developed, tested in digital simulations and validated in the physical test-bed. Validation results are then used to improve the design functionalities in the next iteration starting with digital simulations. This cyclic prototype improvement is facilitated by having a bridge between digital and physical simulations.[3]

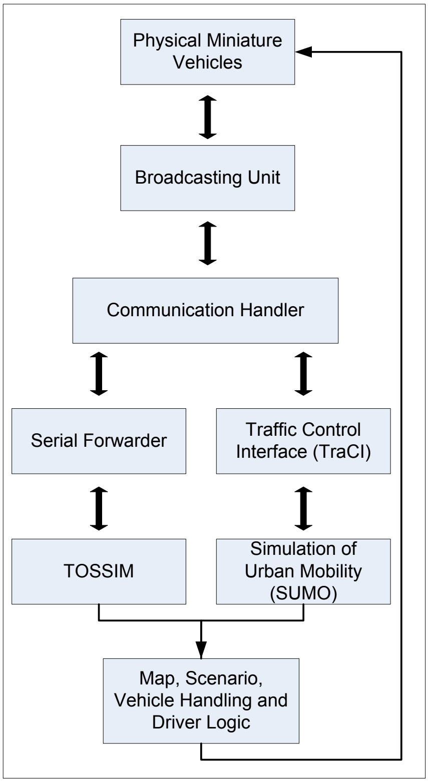


Fig. 4. Block diagram of communication between different systems components within the platform

3.1 Autonomous Driving

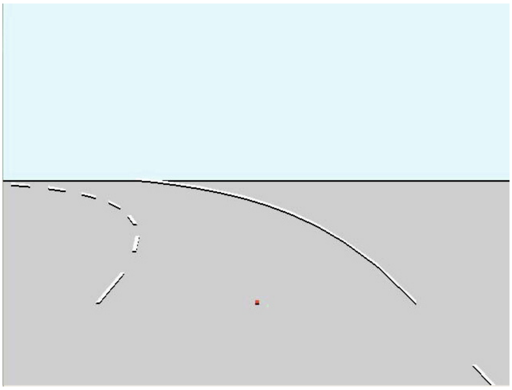
We consider vehicles that run applications for vision and sensor-based lane following. During operation, vehicles determine their deviation from their respective lane markings using feature detection and extraction algorithms. The algorithms are developed and evaluated in the virtualized environment, see Fig. 5, before they are integrated and validated in the physical test-bed.

3.2 Adaptive Cruise Control

This application adjusts the vehicle's cruising speed by monitoring and maintaining a safety distance to the vehicle in front. We study a (cooperative) adaptive cruise control (ACC) application that uses direct and indirect sensing information. In the case of direct sensing, a pair of ultrasonic distance sensors mounted at the front of each vehicle provides an estimated distance to the vehicle ahead. Indirect sensing uses a network protocol that frequently reports the locations of nearby vehicles. Each sensing channel is iteratively and independently tested via digital simulations, and then crossed over to the physical test-bed for validation.

[3] See demonstration videos via `http://www.chalmers.se/hosted/gulliver-en/documents`

(a)



(b)

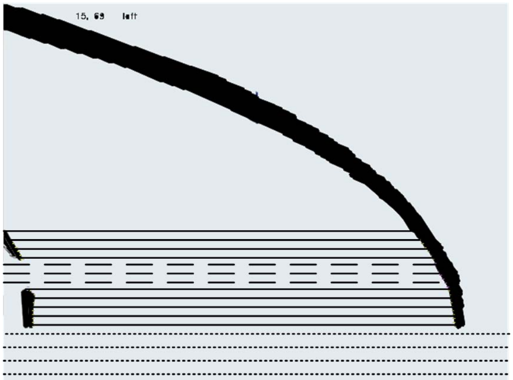
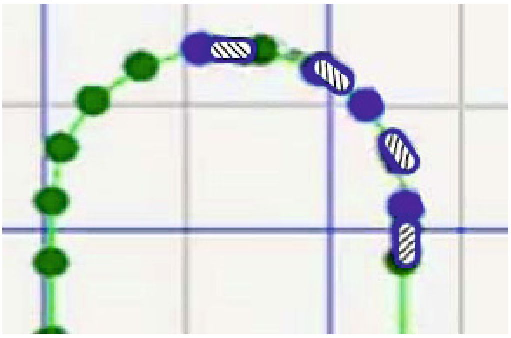


Fig. 5. Vision based approach is used to detect lane markings. (a) The input data is retrieved directly from the 3D simulation environment [4] by the virtual camera. (b) Based on virtual camera data input, lane detection algorithms are developed and tested. The horizontal lines indicate whether both left and right lane markings are detected (solid), one lane marking is detected (dashed), or none is detected (dotted).

The use of direct sensing allows the control loop to retrieve measurements within nearly constant and short delays, but with limited sensing range. The indirect sensing approach provides the control loop with longer and wider range of all nearby vehicles. However, the use of a network protocol implicates longer delays and communication interferences. We combine the benefits of direct and indirect sensing by joining the two approaches under one control loop, and use the test-bed for validation, see Fig. 6.

This case study presents the ability to independently test application components. This allows parallel validations before their integration. A sequence diagram illustrating the development and testing process of ACC is shown in

(a)



(b)

Fig. 6. (a) Map part of the application is showing physically simulated vehicles that travel on a curve while maintaining a safety distance to vehicle ahead using the developed adaptive cruise control application. (b) Rear view of leading vehicle showing trailing vehicles maintaining a safety distance.

Fig. 7. The development and testing of the application was simplified by having the possibility to bridge between digital and physical simulations.

3.3 Crash Avoidance

Testing active safety applications in non-destructive settings is crucial for the development of vehicular safety systems. We consider such experimentations using mixed approaches for digital and physical simulations. We develop a test scenario where digital and physical vehicles travel in an arena that has digital obstacles on the road. The vehicles are equipped with digitally simulated infrared distance sensors that provide the vehicles with the detection data.

In this setting, both digitally and physically simulated vehicles are allowed to drive and apply maneuvering techniques around (digital) obstacles. This setting allows crash avoidance applications that execute on the physical vehicles to be validated without the risk of crashing. Furthermore, the iterative development of

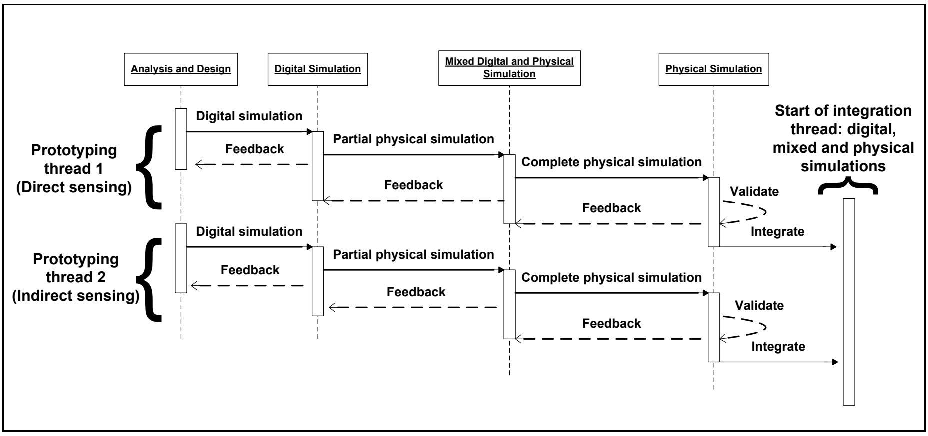


Fig. 7. Sequence diagram of ACC. Both direct and indirect sensing are developed and crossed over from digital to physical simulations in parallel by the protoyping threads. The validation feedback loops can be used for making sure that the design assumptions are covered when testing the application via digital, mixed and physical simulations. The integration thread combines both protyping threads and evaluates the resulting application.

crash avoidance applications via digital simulations, and the ability to selectively move components to the physical test-bed according to test-case needs provide flexibility and safety in validating the application.

3.4 Virtual Traffic Light

This case study considers an application for coordinated intersection crossing. The coordination is made via a distributed scheduler that the vehicles communicate over the radio. We design the scheduler such that at any time, only one direction can have a green light. This is required to prevent vehicles from crashing. Before the application can be tested, we need to validate the behavior of the network protocol. Therefore, our development process includes two parts: one where only the network protocol is validated, and the other where the system is validated. This implies that we need to have a mixed phase where we combine both physical and digital simulations under one test case.

The Gulliver test-bed allowed the application to be developed in stages. We started with a fully digital phase, where the application is only tested via digital simulations. Before validating the application in the physical test-bed, we had to validate the network protocol the delivers to the vehicles information about the traffic light state. Therefore, the next step was to use a mixed digital and physical phase that validates the network protocol of the virtual traffic light via physical simulations. In this test, the vehicle's position was given by the digital representation of the vehicles in the simulator. We then installed the entire application on the physical vehicles, but disabled the vehicle's own mobility. This phase allowed us to validate the vehicle's behavior before the full physical simulation of the application (with autonomous vehicles), see Fig. 8.

This case study demonstrates the ability and usefulness of approaches that mix digital and physical simulation components. This becomes beneficial when there is a need to validate component, such as the virtual traffic light protocol, that are suppose to prevent vehicles from crashing. By considering a limited set of physical components, we were able to exclude such risks before the application was fully used in the test-bed.

3.5 Emerging Traffic Patterns and Phenomena

A number of real-world traffic phenomena, such as shock-waves and stop-and-go traffic jams, can appear when the road contention is very high. The

(a)

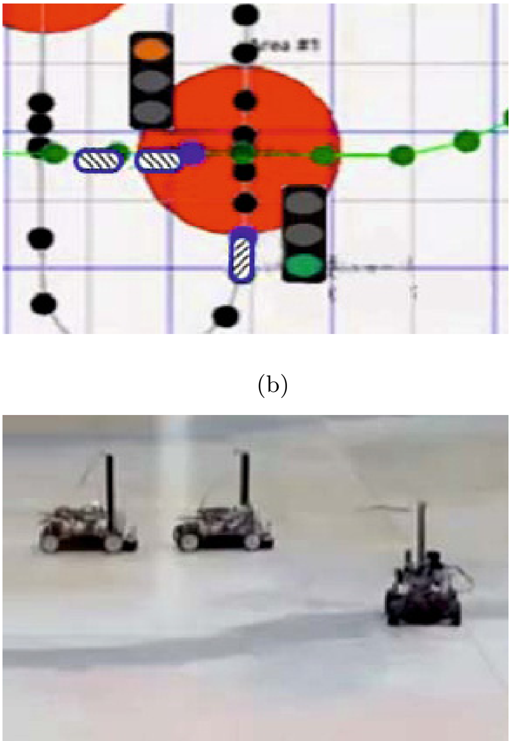
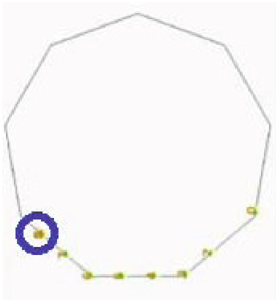


Fig. 8. The virtual traffic light application running on the test-bed. (a) The map client monitors and displays the current position of vehicles as they approach the intersection. (b) The miniature vehicle at the intersection as seen on test-bed.

(a)



(b)

Fig. 9. (a) Digital vehicles traveling counter-clockwise on a single lane road, where the last (circled) vehicle is physically simulated. (b) The physically simulated vehicle is imposed to move according to the stop-and-go phenomenon in test-bed.

investigation of such phenomena is imperative when examining applications, such as the virtual traffic light and cooperative cruise control, that have safety-critical and contention concerns. For example, a shock-wave can surprise the driver and impose an emergency brake. Cooperative cruise control mechanisms can assist with meditating the shock-wave propagation.

Thus far, such phenomena were studied via microscopic (digital) analytical simulation tools such as SUMO [3]. One may aim to produce such phenomena with miniature vehicles. Such experiments could be costly and complicated because they will involve hundreds of miniature vehicles. Gulliver's approach overcomes this difficulty by taking a mixed approach for physical and digital simulation, where some of the simulated vehicles are physically present in the test-bed, while a large number of vehicles are digitally simulated.

We consider both digital and physical simulation of vehicles. A physically simulated vehicle can be either autonomous or controlled via the digital simulator. The experiment demonstrates this ability using a circular single lane road, where a physically simulated vehicle is controlled by its digital image. The physically simulated vehicle follows a number of digitally simulated vehicles that travel counter-clockwise, see Fig. 9.

As the contention on the road created by the digitally simulated vehicles increases, the physically simulated vehicle is imposed to move according to the produced stop-and-go phenomenon. We note that a number of autonomous physically simulated vehicles can follow the controlled one and move according to the stop-and-go phenomenon.

This case study shows that the mixed physical and digital approach can cause physically simulated vehicles to move according to the digitally generated traffic phenomenon. Given this ability, we can then use physically simulated vehicles to investigate the behavior of safety critical applications with traffic phenomena by considering simple and inexpensive testing environment.

4 Conclusions

Vehicular systems require extensive testing in the digital domain before they are validated with physical experiments. This paper proposed an inexpensive cyber-physical platform that bridges digital and physical simulations in order to simplify the development and testing environment of vehicular systems. The proposed platform facilitates iterative testing, where lessons learned from each iteration are used to improve the application starting from digital simulations.

By reviewing case studies of applications that were developed using the proposed platform, we were able to demonstrate our mixed digital and physical simulation approach. According to the application needs, the platform allows some components to be digital and others physical. For example, applications of high speed object avoidance can be tested in a mixed simulation, where physically simulated vehicles maneuvers around objects that are digitally simulated. This mixed approach is extended to demonstrate that real world traffic phenomena, such as shock-waves and stop-and-go traffic jams, can be generated using mixed digital and physically simulated vehicles that navigate through the same map. This facilitates the investigation of safety critical applications that are related to collision avoidance in the context of different traffic phenomena. One may also add intelligent balloon vehicles to the test-bed to validate safety critical applications. Outcomes learned from using miniature vehicles can then be used with full-scale vehicles on testing grounds.

The designed platform opens new possibilities for validating new ideas for traffic situations in complex systems. Gulliver long-term goals include decreasing the gap between full-size and miniature scale vehicles in order to provide an even more flexible and low-cost test-bed.

Acknowledgment. We would like to thank Benjamin Vedder, who designed the vehicle's computer system setup and developed many of the key components.

References

1. Adachi, M., Morita, Y., Fujimura, K., Takatori, Y., Hasegawa, T.: On an autonomous cruising traffic flow simulator including inter- vehicle and road-to-vehicle communication networks. In: The IEEE 5th International Conference on Intelligent Transportation Systems, pp. 645–650 (2002)
2. Al Jaafari, M., Al Shamisi, M., Al Darmki, M., Al Kaabi, M., Lakas, A., Boulmalf, M.: Vie: A simulator for road traffic and inter-vehicular communication. In: Int. Conf. Innovations in Information Technology, pp. 548–552 (2008)
3. Behrisch, M., Bieker, L., Erdmann, J., Krajzewicz, D.: Sumo - simulation of urban mobility: An overview. In: The Third International Conference on Advances in System Simulation, SIMUL 2011, Barcelona, Spain (2011)
4. Berger, C.: Automating Acceptance Tests for Sensor- and Actuator-based Systems on the Example of Autonomous Vehicles. Shaker Verlag, Aachener Informatik-Berichte, Software Engineering Band 6, Aachen, Germany (2010)
5. Bilstrup, K., Uhlemann, E., Ström, E.G., Bilstrup, U.: Evaluation of the ieee 802.11p mac method for vehicle-to-vehicle communication. In: IEEE 68th Vehicular Technology Conference: VTC2008-Fall, pp. 1–5. IEEE (2008)
6. Chen, H., Xiong, P., Schwan, K., Gavrilovska, A., Xu, C.-Z.: A cyber-physical integrated system for application performance and energy management in data centers. In: IGCC, pp. 1–10. IEEE Computer Society (2012)
7. Cho, B.-G., Koo, J.-K., Yoon, B.-J., Kim, B.-W.: The research of dead reckoning stabilization algorithm using different kinds of sensors. In: International Conference on Control, Automation and Systems, pp. 1089–1092 (2010)
8. Hoehmann, L., Kummert, A.: Mobility support for wireless sensor networks simulations for road intersection safety applications. In: Midwest Symposium on Circuits and Systems, pp. 260–263 (2009)
9. Kaiser, J., Schulze, M., Zug, S., Cardeira, C., Carreira, F.: Sentient objects for designing and controlling service robots. Proccedings of IFAC 8, 6–11 (2008)
10. Khalfallah, S., Ducourthial, B.: Bridging the gap between simulation and experimentation in vehicular networks. In: VTC Fall, pp. 1–5 (2010)
11. Krajzewicz, D., Hertkorn, G., Rössel, C., Wagner, P.: Sumo (simulation of urban mobility) - an open-source traffic simulation. In: 4th Middle East Symposium on Simulation and Modelling, pp. 183–187 (2002)
12. Lee, I., Sokolsky, O.: Medical cyber physical systems. In: Proceedings of the 47th Design Automation Conference, DAC 2010, pp. 743–748. ACM, New York (2010)
13. Leone, P., Papatriantafilou, M., Schiller, E.M., Zhu, G.: Chameleon-mac: Adaptive and self-* algorithms for media access control in mobile ad hoc networks. In: Dolev, S., Cobb, J., Fischer, M., Yung, M. (eds.) Stabilization, Safety, and Security of Distributed Systems (SSS 2010). LNCS, vol. 6366, pp. 468–488. Springer, Heidelberg (2010)
14. Leone, P., Schiller, E.M.: Self-stabilizing tdma algorithms for dynamic wireless adhoc networks. In: Bar-Noy, A., Halldórsson, M.M. (eds.) ALGOSENSORS 2012. LNCS, vol. 7718, pp. 105–107. Springer, Heidelberg (2013)
15. Levis, P., Lee, N., Welsh, M., Culler, D.E.: TOSSIM: accurate and scalable simulation of entire tinyos applications. In: ACM SenSys, pp. 126–137 (2003)
16. Lin, J., Sedigh, S., Miller, A.: Towards integrated simulation of cyber-physical systems: A case study on intelligent water distribution. In: Proceedings of the 8th IEEE Int. Conf. Dependable, Autonomic and Secure Computing, DASC 2009, pp. 690–695. IEEE Computer Society, Washington, DC (2009)

17. Mustafa, M., Papatriantafilou, M., Schiller, E.M., Tohidi, A., Tsigas, P.: Autonomous TDMA alignment for vanets. In: VTC Fall, pp. 1–5. IEEE (2012)
18. Pahlavan, M., Papatriantafilou, M., Schiller, E.M.: Gulliver: a test-bed for developing, demonstrating and prototyping vehicular systems. In: MOBIWAC, pp. 1–8 (2011)
19. Schulze, M., Zug, S.: A middleware based framework for multi-robot application development. Relation 10(1.115), 9498 (2010)
20. Schulze, M., Zug, S., Campos, F., Carreira, F.: Exploiting the famouso middleware in multi-robot application development with matlab/simulink. In: Middleware (Companion), pp. 74–77 (2008)
21. Schumacher, H., Schack, M., Kürner, T.: Coupling of simulators for the investigation of car-to-x communication aspects. In: APSCC, pp. 58–63 (2009)
22. Crossbow Technology Inc. MicaZ specs (2009), http://bit.ly/roPGqJ
23. Vedder, B.: Gulliver: Design and implementation of a miniature vehicular system. Master's thesis, CSE, Chalmers Univ. of Tech. (2012)
24. Vinel, A.: 3gpp lte versus ieee 802.11p/wave: Which technology is able to support cooperative vehicular safety applications? IEEE Wireless Communications Letters 1(2), 125–128 (2012)
25. Wan, J., Suo, H., Yan, H., Liu, J.: A general test platform for cyber-physical systems: Unmanned vehicle with wireless sensor network navigation. Procedia Engineering 24, 123–127 (2011); International Conference on Advances in Engineering (2011)
26. Wegener, A., Piórkowski, M., Raya, M., Hellbrück, H., Fischer, S., Hubaux, J.-P.: Traci: an interface for coupling road traffic and network simulators. In: Proceedings of the 11th Communications and Networking Simulation Symposium, CNS 2008, pp. 155–163. ACM, New York (2008)

Open-VSeSeMe: A Middleware for Efficient Vehicular Sensor Processing

Zubair Nabi[1], Atif Alvi[2], Gary Allen[3], David Greaves[4], and Rashid Mehmood[3]

[1] Information Technology University, Pakistan
zubair.nabi@itu.edu.pk
[2] University of Engineering and Technology, Pakistan
atif.alvi@uet.edu.pk
[3] University of Huddersfield, UK
{g.allen,r.mehmood}@hud.ac.uk
[4] University of Cambridge, UK
david.greaves@cl.cam.ac.uk

Abstract. The recent increase in the number of sensors within cars has resulted in various fragmented software stacks and development frameworks. In this ecosystem, applications have to make sense of raw sensor data themselves. As a remedial solution, we present Open-VSeSeMe, a middleware atop TinyOS that converts raw sensor streams into data units with semantic meaning. These data units can be shared between applications leading to efficient use of resources. In addition, we argue that the use of a common software stack leads to hardware standardization and opens up the platform to third-party developers, making a Car App Store possible. Furthermore, the entire architecture is event-driven which frees the applications from the clutches of constant polling. Finally, using a number of illustrative examples we show the utility and usefulness of Open-VSeSeMe.

Keywords: In-car sensors, sensor stream-processing, car safety systems, automotive standardization, vehicular telemetry, contextual data.

1 Introduction

Over the course of the last few decades vehicles have evolved from having zero electronic components and sensors to having hundreds of them [1]. This proliferation has been engendered by a number of factors including automotive systems applications, safety, regulation, entertainment, infotainment, and telemetry. In turn, this has also led to a number of vendor-specific on-board networks to connect these sensors. Unfortunately, this has pushed network stacks towards fragmentation, lack of reusability, and maintainability, hampering evolution [2]. To remedy this, vendors and researchers have been pushing a number of standardization initiatives with AUTOSAR [3] at the forefront. Realizing the potential of intra- and inter-car communication, traditional tech companies and network vendors have also jumped into the fray [4]. Although standardization solves issues

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 185–196, 2013.

Table 1. Typical sensors in modern cars

Sensor	Usage	Type
Engine coolant temperature	Measures the temperature of the coolant inside combustion engines	Critical
Parking (proximity)	Detects unseen obstacles	User-centric
Oxygen	Measures the air-fuel ratio	Critical
Tyre-pressure monitoring	Monitors tyre air pressure	Critical
Knock	Detects and prevents engine knocking	Critical
Crankshaft position	Monitors rotational speed and position of the crankshaft	Critical
Vehicle speed	Measures vehicle speed from the transaxle	User-centric
Power steering pressure switch	Detects power steering load	Critical

of interoperability and evolution, the use of wired networks [5,6] still imposes space, weight, and inflexibility overheads. Wireless networks resolve some of these problems but introduce their own set of shortcomings including privacy [7].

Similar to the networks that connect on-board sensors, the software stack and the hardware in sensors are closed and vendor-specific, leading to poor interoperability, and software and hardware reuse. A natural solution to these problems is to develop and deploy application-agnostic hardware and a common software stack [8,9]. On one hand this opens up the platform to third-party vendors and on the other it forces the application developer to follow well-establish code specifications. Marked advantages include increase in code reliability, improvement in programmer output, and masking of hardware heterogeneity.

As stated earlier, sensor data is consumed by a number of diverse applications. These can be generalized into two categories: (1) Critical automotive and (2) User-centric [10]. Critical automotive applications rely on sensor data to ensure error-free, smooth, and reliable functioning of the vehicle. Examples include anti-lock braking system (ABS), automated parking, and adaptive cruise control. User-centric applications on the other hand convert sensor data to a human-readable form and/or enhance the driving experience. Examples include speedometer, automotive night vision, and lane departure warning system. Table 1 lists a small subset of sensors present in modern cars.

One common thread that runs through both categories of applications is that they read raw data from proprietary sensors. This has 2 main implications: (1) The data received by the applications is in raw form and must be converted to some higher level semantics by the application themselves, and (2) There is no sharing of sensor data due to vendor-specific heterogeneity. It is important to highlight that the former problem is similar to sensor data mapping for contextual applications in smart phones [11]. To remedy this we present Open-VSeSeMe, Open-*V*ehicular *Se*mantic *Se*nsor *M*iddlewar*e*, an abstraction middleware built on top of TinyOS [12] that receives raw sensor data and converts it to *Semantic Data Units* (SDUs) for consumption by applications. The use of TinyOS naturally allows the same sensor data to be shared between multiple applications and in essence enables sensor and platform homogeneity [9]. It also opens up the platform to third-party applications and makes a *Car App Store* possible, on the same lines as smart phone app stores. In addition, Open-VSeSeMe implements an event-driven architecture (another advantage of using TinyOS underneath) which blends well with car applications: Actions are only performed when a subscribed sensor event fires. Furthermore, as Open-VSeSeMe can be likened to a stream processing system–which converts a constant stream of data into a higher-level series of events [13]–the platform is amenable to stream processing optimizations. Finally, a rule-based system can use SDUs to implement a car-wide policy infrastructure [11,14].

The rest of the paper is organized as follows. The problem is laid out in detail in §2. §3 describes the design and implementation of Open-VSeSeMe. A number of applications are presented in §4. We summarize related work in §5 and conclude in §6.

2 The Problem

A modern car contains a multitude of applications that perform some form of automotive functionality. A large number of applications also rely on information from in-car components and the external environment, enabled by sensors. For instance, the engine coolant temperature sensor relays its data to a central ECU. This ECU then looks them up in a table and takes the required action. Depending on the look up, it adjusts engine actuators, turns on the cooling fan, and displays information on a gauge on the console. Due to the conflation between policy specification and implementation, and the synchronous nature of this model, any fault in one functionality (say, displaying information on the gauge) might have global side effects. With its large number of sensors and applications, a modern car can be treated as a pervasive environment or more generally a distributed system. Therefore asynchronous broadcast primitives are preferable for such an environment [15] as they can decouple the fate of individual functionalities. On the other hand, this makes it hard to reason about the sequence of events and whether any inadvertent interlocks have been introduced. It is important to highlight that the multicast event bus model of CAN was designed specifically to avoid the synchronous problem at the outset but application design so far has failed to take advantage of this facility. The conflation between policy specification and implementation can be generalized to a problem that afflicts most vehicular multi-tasking applications: There is no compartmentalization between different tasks that rely on the same sensor(s).

Another problem is the lack of sensor sharing between applications. It is pertinent to mention that the multicast event bus nature of CAN directly facilitates sensor data sharing but in the status quo, each application makes use of proprietary sensors and the lack of trust and standardization make it impossible for another application to benefit from those sensors. As a result, modern cars can potentially have redundant sensors performing the same task but exclusively for a particular application. For instance, both the curb feeler and the parking system rely on proximity sensors but each has its own set of redundant sensors. Similarly, the vehicle stability control and the tyre pressure monitoring system are dependent on wheel speed from their own set of sensors with no opportunity of exploiting each other's sensor data for either redundancy, fault-tolerance, or data quality improvement. In addition, this proliferation in redundant physical sensors also increases the energy footprint of applications, takes up physical space, and increases the weight of the car.

Above all, each application takes raw sensor data and gives it application-specific semantic meaning. This effectively rules out any data sharing: Application A might interpret sensor value 10 as "high" while for Application B that value might be "low". One way to tackle this problem is to give each interested application raw data from shared sensors. In principle, this would effectively pass the data interpretation buck to the applications enabling them to suit their needs. But practically this has a high overhead as the same raw sensor data stream would be replicated across the network bus, leading to congestion. This would in turn also increase the energy footprint of the bus.

Finally, as each sensor and application stack is closed-source and proprietary, only factory fitted applications can be deployed within cars [3]. As a result, it is non-trivial to extend an existing service or to add a completely new application to a car after it has passed the factory pipeline. Hence, the status quo makes the formation of a vehicular app store impossible.

To concretize the discussion, any solution to these problems must be in the form of an application-independent middleware that receives raw sensor data and gives it semantic meaning, freeing up the applications to just perform their core functionality. Furthermore, this semantic meaning should potentially be the outcome of multiple raw sensors, for instance location and speed. Such an architecture should also enable the sharing of this semantic data allowing the application execution platform to be standardized and modularized. These requirements constitute the design goals of our solution, Open-VSeSeMe, which we discuss in detail in the next section.

3 Open-VSeSeMe

This section first introduces the design of Open-VSeSeMe and then describes its implementation.

3.1 Design

Figure 1 shows the architecture of Open-VSeSeMe at a high-level. Conceptually, Open-VSeSeMe interposes itself between sensors and applications. For each SDU type, it receives raw data from one or more sensors and converts it to a standard SDU which can be consumed by interested applications. Mathematically stated:

$$\textit{Open-VSeSeMe}(x, y, z, ..) \rightarrow SDU_{x,y,z,..} \quad (1)$$

Figure 2 shows an illustrative example where Open-VSeSeMe receives raw data from the GPS in terms of latitude and longitude coordinates and from the speed

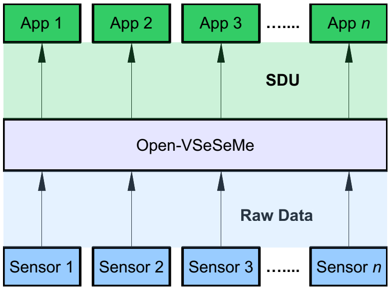


Fig. 1. Abstract design of Open-VSeSeMe

sensor in terms of numerical values in speed units. For instance, latitude la_x and longitude lo_x and speed 50 mph might be translated to an SDU value of "cruising on the highway" while latitude la_y and longitude lo_y and speed 0 mph might be translated to "parked in the garage". It is noteworthy that there might be some applications that require raw sensor data or that sensor data might not be amenable to any SDU conversion. In such circumstances, applications can subscribe to raw data via Open-VSeSeMe. For instance, the speedometer application requires raw and precise speed readings which it can obtain through our middleware. Open-VSeSeMe itself does not perform any non-trivial computations, it only reads raw sensor data and assigns a semantic tag to individual values or an entire range. Therefore, due to its minimal computation and processing footprint, the middleware will never be a performance bottleneck.

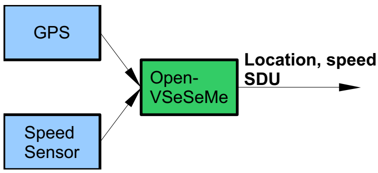


Fig. 2. Open-VSeSeMe converting raw data from the GPS and speed sensor to a location and speed SDU

A natural consequence of this two-level sensor data manipulation is that event firing can be made efficient. Open-VSeSeMe components are invoked every time a new data value is available at the sensor while application events are less frequent. For instance, in case of a temperature sensor, Open-VSeSeMe would be invoked (an event will be fired) whenever there is a change in the temperature reading. Assuming the temperature semantic translator within Open-VSeSeMe considers everything below 50° Celsius as "low" and everything above it as "high", an application event will only be fired if there is a transition between

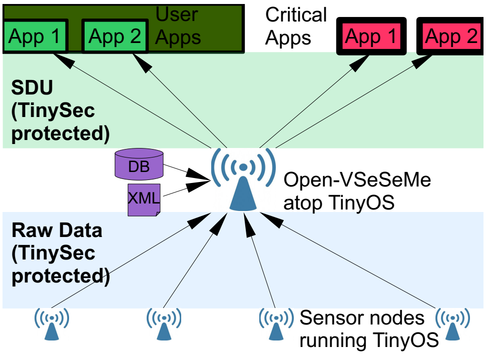


Fig. 3. Architectural overview of the Open-VSeSeMe implementation

these two states. The reduced frequency of application events will yield a substantial reduction in computation cycles across multiple applications.

3.2 Implementation

Open-VSeSeMe is built on top of TinyOS, an event-driven, energy-efficient operating system for wireless sensor networks with a minimal memory footprint (around 400 bytes). An application is compiled into a TinyOS instance (no distinction between kernel and user-space) and has exclusive access to the entire system (no multithreading or multiprocessing). Therefore, unlike conventional operating systems, TinyOS only provides software abstractions and components to read sensor values and communicate them over the network. The operating system is written in nesC [16], an event-driven and concurrent extension of C designed for networked embedded systems. The compiler uses various optimizations (such as function inlining) to ensure reliability and efficient use of resources. TinyOS applications revolve around components, most of which are system provided. Components–which provide system services–are stitched together using configurations. As a result, the same components can be reused across applications. Naturally, only referenced components are compiled into the application resulting in a compact binary image.

We build atop the infrastructure described in [9]. This reference architecture makes use of TinyOS-enabled motes as sensors within a car. Sensors communicate with each other and other artifacts via a radio interface[1]. We augment this setup by defining a component for each SDU. This per SDU component subscribes to raw sensor events, potentially from multiple sensors (both local and from remote ones through the radio interface), and converts them to SDUs. All

[1] While this prototype implementation makes use of wireless links, SDU as a concept can be employed across wired alternatives, including CAN [5], FlexRay [6], FlexCAN [17], and dedicated writing.

SDU components exist within a common TinyOS binary/namespace. Figure 3 depicts this setup. Sensors implemented as wireless sensor nodes run TinyOS and monitor their respective physical quantities. Each node generates events for every new data value. Open-VSeSeMe, also implemented as a TinyOS instance, receives raw values over the radio interface from each sensor node. It then matches these values to a user-provided database or XML file and retrieves the corresponding SDU. An event is only fired off if there is a transition in state. To this end, for each SDU Open-VSeSeMe caches the previous reading (a single value per SDU) and any change in this value is counted as a state transition. Listing 1 shows the pseudo-code of Open-VSeSeMe operation.

The system maintains a distinction between user-centric (non-critical) and internal automotive (critical) applications. The former are all staged within a common embedded OS instance and share the same hardware. Critical applications in contrast are deployed atop their own dedicated ECU and software stack. Both types of applications subscribe to required SDU events with Open-VSeSeMe and are notified on an SDU state transition. Applications are then free to use these SDUs in any way they deem necessary to perform their functionality.

Data: Sensor event
read current;
if *current is not previous* **then**
    look up value in DB;
    broadcast SDU to all applications;
    previous = current;
end

Algorithm 1: Pseudo code of Open-VSeSeMe execution

As stated earlier, in-car wireless systems are vulnerable to attacks [18]. Wireless tyre pressure monitoring systems take minimal security measures to ensure privacy or integrity. So much so that even remote attacks can be launched, resulting in serious safety risks for the vehicle. To avoid such pitfalls altogether, we make use of TinySec [19], a TinyOS add-on with security extensions. It seamlessly and transparently adds encryption and/or authentication to TinyOS applications with a minimal overhead.

4 Applications

This section presents a number of illustrative applications for Open-VSeSeMe enabled sensors. All of these applications can be downloaded from a Car App Store and deployed on existing infrastructure within the car.

4.1 Telemetry

Vehicles have recently been explored as conduits for telemetry. One such effort is CarTel [20], a distributed infrastructure in which each car is treated as a node with sensing capability. It also comes coupled with a stream processing system, ICEDB, and a delay-tolerant network stack, CafNet. A central portal issues ICEDB queries which in turn cause the mobile nodes to return results via CafNet. Under Open-VSeSeMe CafNet can be hosted as a user-centric application allowing the vehicle to act as a CarTel node. In addition, ICEDB specific SDUs can be defined that directly return semantically meaningful results based on a particular query.

4.2 Network Optimizations

Wireless protocols that incorporate contextual information within their adaption algorithms have been proven to substantially improve performance [21]. These algorithms rely on sensor data to define context based on the location, speed, and bearing of mobile devices. This information is then used to, say, adapt the bit rate of the device if the user is stationary. Open-VSeSeMe already converts sensor data to SDUs which can directly be used to establish context and benefit from all of these algorithms.

4.3 Location-Aware Speed Monitor

A natural usage of Open-VSeSeMe is to derive contextual SDUs [11]. Using both GPS coordinates and network proximity, Open-VSeSeMe can give human-centric meaning to location values, such as "Highway X". A location-aware speed monitor can use this SDU to query online sources and get the speed limit for that particular highway. It can then match that value to the current speed of the car and inform the driver if she is exceeding the speed limit.

4.4 Policy Definition

A number of rule based languages have been used to define policy in an entire array of domains including automotive [14]. For instance, SWRL, a first-order predicate logic language enables policy definition in the form of an antecedent with conjunctions and a consequent. For instance, the following rule from [9] automatically takes care of the seasonal calibration of the tyre pressure monitoring system ECU:

$$\mathsf{TPMSECU(?s)} \wedge \mathsf{temp(?t)} \wedge \mathsf{swrlb:lessThan(?t,5)} \rightarrow \mathsf{hasCode(?s,Winter)} \tag{2}$$

We can see that the rule requires access to raw temperature data from the sensor. On the other hand, Open-VSeSeMe can generate an SDU with two states: "SummerTemp" and "WinterTemp" and allow the policy engine to reason over it. The previous rule can thus be re-written as:

$$\mathsf{TPMSECU(?s)} \wedge \mathsf{hasSeason(?s,WinterTemp)} \rightarrow \mathsf{hasCode(?s,Winter)} \tag{3}$$

Similarly, consider the following rule from [14] that automatically turns on the air conditioning when the temperature goes over 30° Celsius:

$$\mathsf{Car}(?\mathsf{c}) \ \wedge \ \mathsf{hasTemperature}(?\mathsf{c}, 30) \rightarrow \ \mathsf{isACOn}(?\mathsf{c}, \mathsf{true}) \tag{4}$$

In the presence of a temperature SDU under Open-VSeSeMe which describes values in terms of "Cold", "Warm", and "Hot", the previous rule can be rewritten as:

$$\mathsf{Car}(?\mathsf{c}) \ \wedge \ \mathsf{hasTemperature}(?\mathsf{c}, \mathsf{Hot}) \rightarrow \ \mathsf{isACOn}(?\mathsf{c}, \mathsf{true}) \tag{5}$$

4.5 Stream Processing System

As noted before, Open-VSeSeMe is very similar to stream processing systems that convert a constant stream of sensor data to higher-level semantic events. As a result, it opens up the Open-VSeSeMe platform to various stream processing optimizations, such as redundancy elimination, load balancing, and state sharing. Most notably, the use of fault tolerance techniques from these systems to deal with both software and hardware failure.

4.6 Maintenance and Diagnostics

Different SDUs can be combined to make higher level diagnostics possible. For instance, "Oil Life Monitoring" systems in certain cars monitor low level characteristics such as engine statistics, driving patterns, temperature, duration of trips, etc. to notify the driver when it is time to change the motor oil. Under Open-VSeSeMe this decision can be taken by logging the relevant SDUs of all of these components.

4.7 Privacy

Information leakage by smart phone apps is a prevalent problem [22,23] leading to serious privacy concerns. We have argued that Open-VSeSeMe enables a car app store to exist but at the same time, it is noteworthy that this will also open up the platform to malicious apps. The gravity of the situation can be gauged from the fact that signals from intra-car wireless networks escape the car chassis and enable malicious users to remotely track the car [18]. The general unit of sensor data in Open-VSeSeMe is an SDU and we believe that a large number of applications will only operate on SDUs as opposed to raw data. Therefore, the onboard system can easily generate fake SDUs to feed to untrusted applications on the same lines as certain solutions for smart phones [22].

5 Related Work

In this section we summarize relevant related work. The design and raison d'etre of Open-VSeSeMe has been inspired by CondOS [11] which generates *contextual* *data units* (CDUs) for mobile devices. Similar to Open-VSeSeMe it also advocates the use of these CDUs for user-centric applications and the optimization of OS services. Our work takes CondOS two steps further by (a) Focusing on all types of sensors to generate SDUs, and (b) Implementing a concrete solution on top of TinyOS. Our work is also complementary to Nabi et al. [9], who advocate the use of TinyOS to standardize the software and hardware stack of in-car sensors. In contrast to their work, which maintains the overhead of raw sensor data, Open-VSeSeMe generates flexible SDUs for application consumption. Similarly, our work is orthogonal to all vehicular standardization efforts [3,8] as we showed that the use of sensors as a platform automatically leads to standardization.

Alvi et al. [14] describe an architecture that makes use of first order rules to define vehicular policy. We showed that SDUs enabled by Open-VSeSeMe can simplify the definition of such rules and make them more flexible and scalable. Furthermore, TinyDB [24] is a stream processing system for acquiring data from TinyOS nodes. It uses a SQL-like interface to extract data and perform in-network processing. TinyDB can easily be extended to query over Open-VSeSeMe SDUs. Moreoever, CarTel [20] is a complete system for deploying and extracting data from telemetry nodes in cars. We have already argued that Open-VSeSeMe is a natural companion to CarTel by exposing in-car sensor data to it. Finally, the concept of converting physical quantities to set values with units and using an event-driven architecture is similar to service data objects in the CANopen standardization [25]. Open-VSeSeMe builds upon this by allowing SDUs to clump different service data objects and assign them application-specific semantic meaning.

6 Conclusion and Future Work

We presented Open-VSeSeMe, a middleware to convert raw sensor data to sensor data units. It uses TinyOS as a substrate to enable sensor standardization, data sharing, and event-driven design. We also argued that such an architecture—in which the middleware gleans meaning from data streams—makes applications more reliable and efficient by imposing modularity. We also presented a number of applications to demonstrate the efficacy of our approach.

Our future work consists of defining a database of SDUs to match the requirements of disparate applications and diverse sensors. In addition, these SDUs should also be amenable to TinyDB and ICEDB queries. Furthermore, we also plan on increasing the suite of applications that can benefit from Open-VSeSeMe. Finally, in the current design Open-VSeSeMe is a central point of failure. Thus, an exploration of soft-state design and replication is also on our list.

References

1. D'Orazio, L., Visintainer, F., Darin, M.: Sensor networks on the car: State of the art and future challenges. In: Design, Automation Test in Europe Conference Exhibition (DATE) (March 2011)

2. Heinecke, H.: Automotive System Design - Challenges and Potential. In: Proceedings of the Conference on Design, Automation and Test in Europe, DATE 2005, vol. 1. IEEE Computer Society, Washington, DC (2005)
3. Heinecke, H., Schnelle, K.P., Fennel, H., Bortolazzi, J., Lundh, L., Leflour, J., Maté, J.L., Nishikawa, K., Scharnhorst, T.: AUTomotive Open System ARchitecture - An Industry-Wide Initiative to Manage the Complexity of Emerging Automotive E/E Architectures. In: Convergence International Congress & Exposition on Transportation Electronics (2004)
4. McMillan, R.: Cisco Wants to Drive Your Car ... Onto the Internet, http://www.wired.com/wiredenterprise/2012/09/cisco-car-router/
5. Tindell, K., Hansson, H., Wellings, A.: Analysing real-time communications: controller area network (CAN). In: Proceedings of the Real-Time Systems Symposium (December 1994)
6. Berwanger, J., Ebner, C., Schedl, A., Belschner, R., Fluhrer, S., Lohrmann, P., Fuchs, E., Millinger, D., Sprachmann, M., Bogenberger, F., Hay, G., Krÿger, A., Rausch, M., Budde, W.O., Fuhrmann, P., Mores, R.: FlexRay - the communication system for advanced automotive control systems. In: SAE World Congress (2001)
7. Dötzer, F.: Privacy issues in vehicular ad hoc networks. In: Danezis, G., Martin, D. (eds.) PET 2005. LNCS, vol. 3856, pp. 197–209. Springer, Heidelberg (2006)
8. EMMA Project: Embedded Middleware in Mobility Applications, http://www.emmaproject.eu
9. Nabi, Z., Alvi, A., Mehmood, R.: Towards standardization of in-car sensors. In: Strang, T., Festag, A., Vinel, A., Mehmood, R., Rico Garcia, C., Röckl, M. (eds.) Nets4Trains/Nets4Cars 2011. LNCS, vol. 6596, pp. 216–223. Springer, Heidelberg (2011)
10. Fröberg, J., Sandström, K., Norström, C., Hansson, H., Axelsson, J., Villing, B.: Correlating business needs and network architectures in automotive applications - a comparative case study. In: Proceedings of the 5th IFAC International Conference on Fieldbus Systems and their Applications (FET). IFAC (July 2003)
11. Chu, D., Kansal, A., Liu, J., Zhao, F.: Mobile apps: it's time to move up to CondOS. In: Proceedings of the 13th USENIX Conference on Hot Topics in Operating Systems, HotOS 2013. USENIX Association, Berkeley (2011)
12. Hill, J., Szewczyk, R., Woo, A., Hollar, S., Culler, D., Pister, K.: System architecture directions for networked sensors. SIGPLAN Not. 35(11) (November 2000)
13. Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Stonebraker, M., Tatbul, N., Zdonik, S.: Monitoring streams: a new class of data management applications. In: Proceedings of the 28th International Conference on Very Large Data Bases, VLDB 2002. VLDB Endowment (2002)
14. Alvi, A., Nabi, Z., Greaves, D.J., Mehmood, R.: Intra-vehicular verification and control: a two-pronged approach. International Journal of Vehicle Information and Communication Systems 2(3) (2011)
15. Saif, U., Greaves, D.: Communication primitives for ubiquitous systems or RPC considered harmful. In: International Conference on Distributed Computing Systems Workshop (April 2001)
16. Gay, D., Levis, P., von Behren, R., Welsh, M., Brewer, E., Culler, D.: The nesC language: A holistic approach to networked embedded systems. In: Proceedings of the ACM SIGPLAN 2003 Conference on Programming Language Design and Implementation, PLDI 2003. ACM, New York (2003)
17. Pimentel, J.R., Fonseca, J.A.: FlexCAN: A Flexible Architecture for Highly Dependable Embedded Applications. In: 3rd Int. Workshop on Real-Time Networks (July 2004)

18. Rouf, I., Miller, R., Mustafa, H., Taylor, T., Oh, S., Xu, W., Gruteser, M., Trappe, W., Seskar, I.: Security and privacy vulnerabilities of in-car wireless networks: a tire pressure monitoring system case study. In: Proceedings of the 19th USENIX Conference on Security, USENIX Security 2010. USENIX Association, Berkeley (2010)
19. Karlof, C., Sastry, N., Wagner, D.: TinySec: a link layer security architecture for wireless sensor networks. In: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, SenSys 2004. ACM, New York (2004)
20. Hull, B., Bychkovsky, V., Zhang, Y., Chen, K., Goraczko, M., Miu, A., Shih, E., Balakrishnan, H., Madden, S.: CarTel: a distributed mobile sensor computing system. In: Proceedings of the 4th International Conference on Embedded Networked Sensor Systems, SenSys 2006. ACM, New York (2006)
21. Ravindranath, L., Newport, C., Balakrishnan, H., Madden, S.: Improving wireless network performance using sensor hints. In: Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation, NSDI 2011. USENIX Association, Berkeley (2011)
22. Hornyack, P., Han, S., Jung, J., Schechter, S., Wetherall, D.: These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In: Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS 2011. ACM, New York (2011)
23. Beresford, A.R., Rice, A., Skehin, N., Sohan, R.: Mockdroid: trading privacy for application functionality on smartphones. In: Proceedings of the 12th Workshop on Mobile Computing Systems and Applications, HotMobile 2011. ACM, New York (2011)
24. Madden, S.R., Franklin, M.J., Hellerstein, J.M., Hong, W.: TinyDB: an acquisitional query processing system for sensor networks. ACM Trans. Database Syst. 30(1) (March 2005)
25. Device Profile Measuring Devices and ClosedLoop Controllers. CiA Draft Standard 404

LTE Based Communication System for Urban Guided-Transport: A QoS Performance Study

Arwa Khayat[1,2], Mohamed Kassab[1], Marion Berbineau[1], Mohamed Amine Abid[2], and Abdelfettah Belghith[2]

[1] Univ Lille Nord de France, F-59000, Lille, IFSTTAR, LEOST, F-59650, Villeneuve d'Ascq
{arwa.khayat,mohamed.kassab,marion.berbineau}@ifsttar.fr
[2] University of Manouba, HANA Research Group, Tunisia
{mohamed-amine.abid,abdelfettah.belghith}@hanalab.org

Abstract. The control and management of new urban guided-transport systems such as tramways and subways are based on several IT services with heterogeneous data communication requirements. These applications are known as CBTC (Communication Based Train Control systems) and CCTV (Closed Circuit TeleVision). In order to support existing as well as emerging applications over a unique communication infrastructure, it is mandatory to be able to ensure efficient QoS management that meets the requirements of CBTC and CCTV services especially the critical ones. This paper presents an evaluation of QoS provisioning in LTE (Long Term Evolution) based communication system for urban guided-transport. The evaluation is made through simulations with the well known event driven simulator OPNET. We propose a mapping of the typical applications CBTC and CCTV to standardized 3GPP QCIs (QoS Class Identifiers) that serve as a basis for the LTE class-based QoS. After discussion of the obtained results, we provide recommendations to enhance the communication performances experienced by these applications.

Keywords: LTE, QoS provisioning, EPS bearer, urban guided-transport, CBTC, CCTV.

1 Introduction

To increase quality, reliability, safety and security of public transport systems, while increasing accessibility and productivity, new urban guided-transport systems rely on the deployment of several communication applications and services. The latter fall into three main categories: critical applications (e.g. control-command), the non-critical applications for train operation support (e.g. video surveillance, maintenance), and non-critical applications for infotainment (e.g. passenger information, internet on board). These applications define data exchanges between management infrastructures, rolling stock on-board units, devices deployed along the rails, *etc.* The emergence of these applications has to face with some limitations of conventional communication technologies, especially

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 197–210, 2013.

in terms of providing ETE (End-To-End) QoS (Quality of Services). Current commercial solutions for subways use WLAN (Wireless Local Area Network) networks based on IEEE 802.11 like modems with some proprietary development in order to support requirements of the CBTC and CCTV applications. Furthermore, it is usual today that a public transport operator deploys several different wireless networks that are mostly non interoperable, which leads to high operation and maintenance costs and a non-optimized radio spectrum use.

The emergence of 4G telecommunication systems and particularly the future wide deployment of the 3GPP LTE based networks in the cities represent a real opportunity for public transport. The 3GPP LTE is the future mobile communication technology proposed by the 3GPP. It targets a packet optimized-system that provides higher data rate and lower delay with improved coverage and spectrum efficiency. In parallel, the technology proposes a network architecture that aims at supporting heterogeneous traffic and manages transparently the QoS.

Taking into account all these promising performances, a communication system based on LTE is introduced as possible communication infrastructure able to support the variability in the QoS requirements across urban guided transport applications. However, it remains a challenge to accommodate all these applications on a single wireless access network where bandwidth is at a premium. QoS mechanism is always a much lauded basic requirement.

In this paper, we consider a LTE deployment to manage urban guided-transport. We propose a QoS configuration that maps transport applications, based on their communication requirements, to LTE QoS classes. We consider the CBTC as an example of critical applications and the CCTV as an example of non-critical applications. This configuration is evaluated using the event-driven network simulator: OPNET Modeler (Optimized Network Engineering Tool) [1]. Our goal is to investigate the performance regarding ETE delay and packet loss, when delivering mixed traffic over a single wireless LTE access. We propose typical LTE network deployment for performance evaluation. Based on the obtained results, we provide some recommendations for using LTE in the context of urban guided-transport.

The remainder of the paper is organized as follows: Section II presents some of the research literature relevant to QoS issues in LTE. Section III overviews the LTE system. It briefly describes the network architecture and the basic concepts for QoS provisioning in LTE networks. Section IV presents the proposed QoS mapping. The simulation results are detailled and analyzed in section V. Conclusions are given in section VI.

2 Related Works

Several studies, presented in the literature, have addressed various QoS aspects in LTE. But far fewer works consider mixed services scenarios. In this section, we highlight studies that evaluate LTE network capabilities to meet heterogeneous applications requirements. In mobile network, the wireless link is usually the bottleneck of the whole network. Therefore, most of these studies focus on QoS mechanisms at the radio interface, especially the performance assessment of QoS scheduling techniques.

The effects of QoS scheduling strategies on service performance in a mixed traffic consisting of VoIP (Voice over IP), streaming video, and SIP (Session Initiation Protocol) were studied in [2]. *M. Wernersson et al.* considered a video/audio call over IMS (IP Multimedia Subsystem). They showed that strict prioritization for SIP packets over other packets such as voice and data enhances the overall performances. This work showed also that prioritization of such signaling messages does not significantly affect other services performance. [3] and [4] analyze the packet scheduling of mixed traffic in LTE downlink. The authors demonstrate that prioritization of delay sensitive traffics such as VoIP traffic, is required, particularly, when this traffic coexists with delay-tolerant traffics like web surfing or TCP download. Similarly to the LTE downlink, *M. Salah et al.* conduct in [5] a performance evaluation of LTE, with regard to the QoS requirements for different types of uplink traffic. Several uplink schedulers were evaluated in terms of per-user throughput, delay, packet loss and fairness. The authors of [7] propose a combined admission control and packet scheduling framework for QoS provisioning in LTE uplink. The proposed framework is shown to be able to guarantee the respective QoS requirements of different user classes in a mixed traffic scenario. In order to improve the QoE (Quality of Experience) of the end user, *V.H. Muntean et al.* evaluate the performance of three scheduling algorithms for an e-learning application in [6]. This application generates a traffic mix and is deployed over LTE network.

The QoS perceived by the end user is an ETE issue and is affected by every part of the network. In this paper, we rely on the LTE class-based QoS to go beyond the air interface and look at QoS provisioning from an ETE network perspective. The considered traffic mixes in literature are: delay-sensitive and elastic applications, video, audio and data traffic, signaling and user data. To the best of our knowledge and to date, there is no published work related to QoS management for mixed critical and non-critical applications in general, and particularly in the context of urban-guided transport. Because of the particularity of this context, transport applications experience more stringent QoS requirements compared to common network applications such as VoIP and video streaming.

3 Overview on LTE Technology

LTE is an emerging wireless technology for mobile communications standardized in 3GPP. It represents the last progress of the UMTS (Universal Mobile Telecommunications System) standard. LTE proposes a flat all-IP architecture (Fig.1). This architecture encompasses the evolution of radio and non-radio aspects of UMTS, represented respectively by the E-UTRAN (Evolved UMTS Radio Access Network) and the SAE (System Architecture Evolution), which covers the EPC (Evolved Packet Core) network.

The E-UTRAN consists of one type of node, the eNodeB (evolved NodeB), which ensures the radio resource management and the connectivity of UEs (User Equipments) to the core network via the S1 interface. The EPC consists of MME (Mobility Management Entity), the SGW (Serving GateWay), and PDN (Packet Data Network) Gateway. The MME is responsible for mobility and session management. The SGW routes and forwards user data packets between E-UTRAN and PDN Gateway. Besides, it serves as a mobility anchor for the user plane during inter-eNodeB handovers and inter-3GPP technology handovers. The PDN Gateway provides gateway function to the operator IP services e.g. IMS or Internet. PCRF (Policy and Charging Rule Function) controls QoS policy and charging for users and services. HSS (Home Subscriber Server) contains users'subscription data such as the subscribed QoS profile.

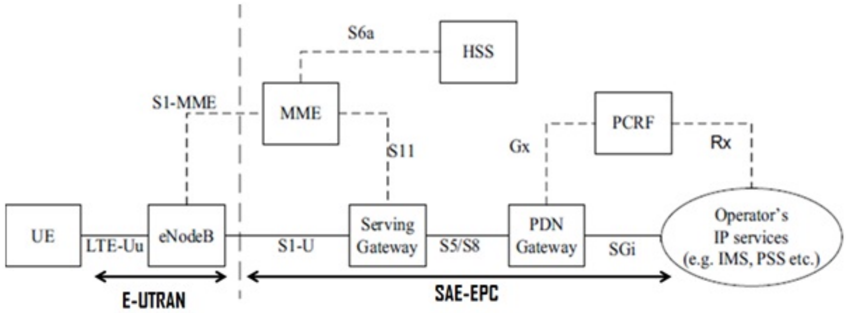


Fig. 1. Architecture of the 3GPP LTE System [10]

In order to make such an architecture able to support the variability in the QoS requirements across services and users, LTE specifies class-based QoS provisioning. Most of the QoS related functions are configurable. Hence, an operators can simply align the level of QoS support in the network with his own strategy. LTE supports ETE QoS through the concept of EPS (Evolved Packet System) bearer, which is used to identify data packets belonging to a logical IP transmission path that receives a common QoS treatment between an UE and the PDN Gateway. All functionalities required to handle the bearer signaling as well as the user plane transport are supported by a GTP (GPRS Tunneling Protocol) between the PDN Gateway and the SGW. There are two kinds of EPS bearers: GBR (Guaranteed Bit Rate) bearers and Non-GBR bearers. A GBR bearer has a specified data rates (UL (UpLink)/ DL (DownLink) GBR values) for which network resources are allocated generally by an admission control function in eNodeB. Non-GBR bearer does not guarantee any particular bit rate. During the attachment procedure to the network, an IP address and at least one EPS bearer (the default bearer) are assigned to the UE by the PDN Gateway. The default bearer remains active during the PDN connection lifetime in order to ensure to the UE an always-on IP connectivity. The default bearer serves as default tunnel to transport traffic when no QoS differentiation is applied. Since it is permanently established, the default bearer is a Non-GBR bearer. An EPC bearer has an associated QCI. A QCI is a pointer to a pre-configured set of node specific parameters (e.g. scheduling weights, admission thresholds, packet discard timer, etc.) that determines the packet forwarding behavior to be provided to a service data flow delivered over this bearer. To ensure interoperability between operators and network equipment vendors, nine QCIs have been standardized by the 3GPP (Tab.1).

Each QCI is characterized by priority level, packet delay budget and acceptable packet loss rate. The packet delay budget defines an upper bound for the time that a packet may be delayed between a UE and the PDN Gateway. The packet loss rate defines an upper bound for a rate of non-congestion related packet losses. In addition, a priority level is assigned for each QCI (the priority level 1 is the highest). The priority level and the packet delay budget determine how the MAC (Medium Access Control) scheduler, located in the eNodeB, handles packets sent over the bearers (e.g. a packet with higher priority is expected to be scheduled before a packet with lower priority).

Table 1. Standardized QCI characteristics [8]

QCI	Resource	Priority	Packet delay (ms)	Packet loss	Services
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational voice (live streaming)
3	GBR	3	50	10^{-3}	Real-time gaming
4	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
5	Non-GBR	1	100	10^{-3}	IMS signaling
6	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
7	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive streaming
8	Non-GBR	8	300	10^{-6}	TCP-based (e.g. WWW, e-mail), FTP, P2P, etc.,
9	Non-GBR	9	300	10^{-6}	

The QCI characteristics (Tab.1) are not signaled on any interface. They should be understood as guidelines for the pre-configuration of node specific parameters for each QCI, particularly the configuration of the air interface. The scope of the standardized QCI characteristics is between the UE and the PDN Gateway. The EPC is a wired network which can be easily configured and controlled by the operator. The LTE specifications focus on the performance of the radio link. The configuration of the core network is generally an operator's issue. Therefore, QCI characteristics are based on some assumptions related to EPC. For a given packet delay budget, the considered core network delay is 20 ms (average delay depending on the distance between the PDN gateway and the eNodeB). Concerning the packet loss rate, the rate of non congestion related packet losses that may occur between the radio base station and a PDN Gateway are considered negligible. Even if the most of values specified for a standardized QCI apply particularly to the radio interface, the QCI parameter determines the ETE QoS. In the core as well as the backhaul network, the QCI parameter is used to map traffic to EPS bearer and eventually to DiffServ (Differentiated Services) classes; the QCI is mapped to a DSCP (Diffserv Code Point) value in the IP header over the GTP tunnel, so that the intermediate routers between the PDN Gateway and the eNodeB can make the correct prioritization decisions. Another parameter that characterizes an EPS bearer is the ARP (Allocation Retention Priority). It is used during call admission control to decide whether a bearer establishment or modification request should be accepted or rejected in case of radio congestion. The ARP parameter contains information about the priority level, the preemption capability and the preemption vulnerability of a resource request. The priority level (from 1 to 15, where priority level 1 is the highest) defines the relative importance of a bearer request. The preemption capability and the preemption vulnerability are flags that define respectively, the ability of the bearer request to get resources that are assigned to another bearer with a lower priority level and the possibility that the bearer loses the resources assigned to it in order to admit another bearer with a higher priority level.

QoS in LTE is generally network-initiated, where a service is offered to a subscriber by the operator. QCI, ARP and UL/DL GBR are subscription parameters stored per-service in the HSS. These parameters are used by the PCRF to provide PCC (Policy and Charging Control) rules to the PDN Gateway [9]. The PCC rules define packet filters that allow the PDN Gateway and the UE to distinguish packet flows and then match them to the appropriate EPS bearer. Once the PCC rules are executed, the associated EPS bearers are established. Service data flows, having the same QoS requirements (i.e. the same QoS parameters, QCI and ARP), are delivered over one EPS bearer. If this bearer is GBR, the PCRF adjusts the UL/DL GBR values to the sum of GBR values associated to each service data flow active on this bearer. Bearer establishment procedure across the network nodes can be more complicated. It all depends on the QoS control architecture used by the operator.

4 QoS Configuration

In this paragraph, a communication system based on LTE for urban guided-transport is proposed. A key issue for this proposal is an accurate mapping of service and application requirements to standardized QCIs. In this section, we propose a QoS configuration for CBTC and CCTV applications consisting of QoS parameters per transport application. We use the values specified in Tab.1 as a basis for our mapping.

4.1 QoS Requirements of CBTC and CCTV Applications

The CBTC application relies on a bidirectional communication between on-board units and a wayside server, called ZC (Zone Controller). The size of the exchanged messages is around 200 bytes and the data rate does not surpass 10 kbps. The rolling stock communicates with the ground server to obtain reliable information to adapt its operational status (e.g. speed) to the line status. This application aims at controlling the distance between operating trains, in order to allow optimal use of the railway infrastructure while maintaining the safety requirements. CBTC is highly sensitive to transmission delay and packet loss. The limit of the acceptable ETE delay and the packet loss ratio are around 50 ms and 0.1%, respectively. Such time-critical application requires also a high network availability (99.99%). Thus, dropping an ongoing application or blocking a new application request is unacceptable.

The video surveillance application CCTV defines the use of a camera to transmit real-time videos of the inside of the vehicles to a set of monitors in a control center. The traffic generated by this application is a constant streaming from one camera to the server in the ground. The size of datagram is set to 1000 bytes and the data rate is 2 Mbps. This application is less critical then CBTC, the ETE delay threshold is around 100 ms and the packet loss ratio is equal to 1%.

4.2 Proposed Mapping of Applications Data Flows to Standardized QCI

In the sequel, we identify the requirements of CBTC application as corresponding to QCI 3 characteristics. This QCI specifies resource type corresponds to GBR bearer. It specifies a very strict delay bounds (50 ms) and a low packet loss 10^{-3}. However, the specified priority level is 3, which is considered critical if we assume that other applications can be mapped to QCIs with higher priority level. Another possibility is the use of the QCI 5. This QCI has the highest priority and a very low packet loss 10^{-6}. With regards to delay, the specified value is 100 ms. Remember that this value is an upper bound. The resource type that corresponds to QCI 5 is Non-GBR bearer. The delay can be much lower if the network is not congested. In mobile network, the QCI 5 is used for IMS signaling, Thus a real-time application such as CBTC can be mapped to such QCI. In theory, both QCI 3 and 5 can be used. The two solutions will be evaluated latter using simulation.

The video surveillance application (CCTV) imposes high and constant data rate. We propose to map the data flow of this application to QCI 2, which involves a GBR bearer and meets CCTV requirements regarding delay and packet loss. The corresponding priority level is 4, which is lower than the QCIs proposed for the critical application. The QCI 1 seems also to meet the CCTV requirements. The specified delay and packet loss are 100 ms and 10^{-2}, respectively. However, we do not recommend the QCI 1 for CCTV as we consider that the priority level is too high. A classical map of this QCI is to VoIP applications. It would be better to reserve this QCI, in the urban transport context, to voice applications such as inter-phony and voice announcement.

5 Results and Discussion

In this section, we propose a simulation based performance study to validate the QCI mapping. We use the even driven simulator OPNET. We consider two scenarios: congested and non-congested network. In this study, we focus on the performance of the uplink, as urban guided transport applications (CBTC and

CCTV) generate more traffic over the uplink than the downlink. We present below only QoS metrics calculated for the uplink (ETE delay and packet loss).

5.1 Scenario1: Low Network Load

We consider one eNodeB connected with the EPC using point-to-point links operated at 44.736 Mbps. This eNodeB serves fixed user equipment, named "Tram". An Ethernet network, including application servers, is connected to the EPC. The LTE network carries data from the "Tram" to relevant application servers and vice versa (Fig. 2).

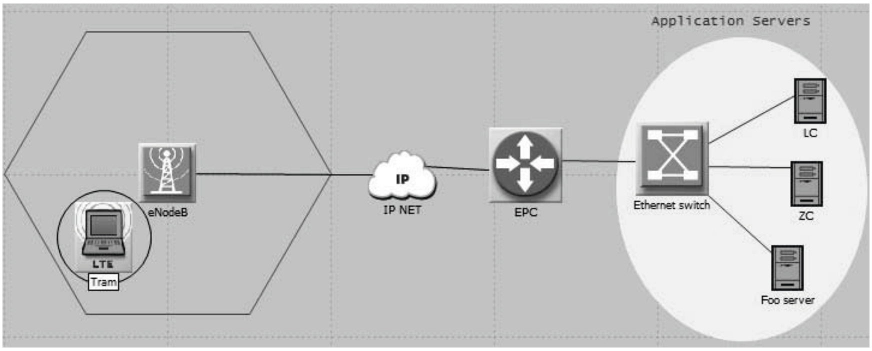


Fig. 2. Network topology

The LTE deployment is configured at 2 GHz the frequency band with a 20 MHz bandwidth used in Time Division Duplexing (TDD) mode (UL/DL 3:2). The proposed configuration offers 50.1377 Mbps on the uplink and 30.0298 Mbps on the downlink. These values are estimated by the control admission function of OPNET's LTE model. The considered applications are given in Tab.2. The user equipment "Tram" executes one CBTC instance, four CCTV video flows, and a set of additional non-critical applications. The latter are on-board file updating with 1 Mbps and three VoIP applications (inter phony, voice announcement and discreet listening) each one generates 64 kbps of traffic. The simulation runs for 600 s, all applications start at 100 s. The network utilization in the uplink and downlink are respectively, 18% and 4%.

We consider the following QoS profiles:

- **Default QoS:** all applications use the default bearer [QCI=9], best effort network.
- **QoS config1:** Non-GBR bearer [QCI=5] used only by the CBTC application, the other applications use the default bearer [QCI=9].
- **QoS config2:** GBR bearer [QCI=3, UL/DL GBR=10 kbps] used only by CBTC application, the other applications use the default bearer [QCI=9].

Table 2. The considered applications

Applications	Direction	Data rate (kbps)
CBTC	UL	8
	DL	4.8
CCTV (1 flow)	UL	2000
Inter phony	UL	64
	DL	64
Voice announcement	DL	64
Discreet listening	UL	64
File update	DL	1000

Fig. 3 and Fig. 4 show the uplink CDF (Cumulative Distribution Function) of the ETE delay for CBTC and CCTV applications. As seen in Fig. 3, mapping the CBTC application to QCI 5 offers the lowest delay (packet ETE delay does not exceed 12 ms). When comparing the graphs, the use of GBR bearer to deliver CBTC traffic gives higher delay than the best effort configuration (Default-QoS) and the QoS-config1. The difference is caused by the delay of the GTP tunnel, between the eNodeB and the EPC. While the bit rate is limited to 10 kbps through the resource allocation done in the case of QoS-config2, this bit rate can achieve much higher values with the use of Non-GBR bearer. In fact, the Non-GBR bearer has an opportunistic aspect and allows the traffic to exploit the 44.739 Mbps offered by the point-to-point links.

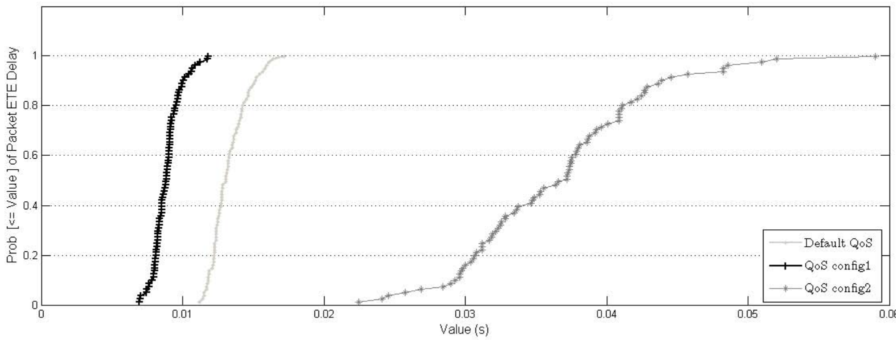


Fig. 3. CDF of ETE Delay for CBTC Application $Tram \rightarrow ZC$

For QoS-config2, about 4% of packets experienced ETE delay that exceeds 50 ms. This percentage can be much higher considering mobility and constraints imposed by the radio propagation environment. Therefore, we can assert that a Non-GBR bearer is the most appropriate to support the critical application. As mentioned earlier, the QCI 5 offers the lowest ETE delay for CBTC application. In the case of QoS-config1, we have two Non-GBR bearers but the bearer that corresponds to QCI 5 has the highest priority. Fig. 4 shows that the

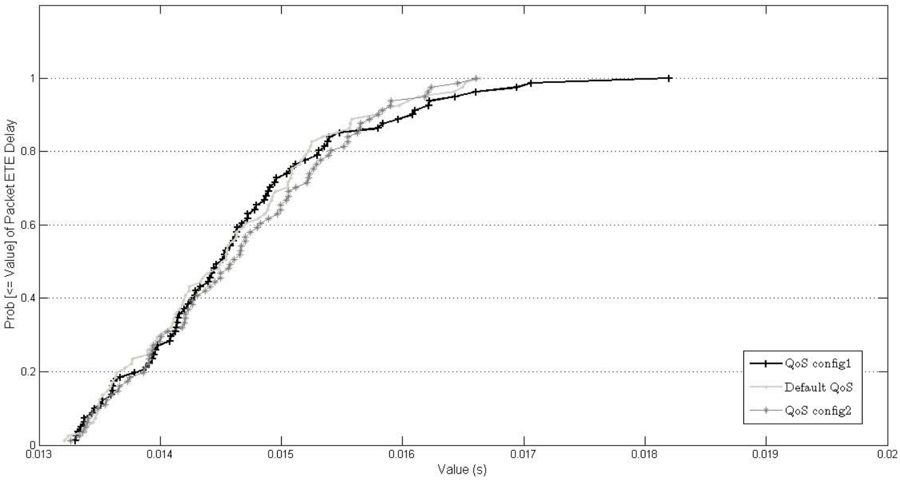


Fig. 4. CDF of ETE Delay for CCTV Application $Tram \rightarrow LC$

prioritization of the critical CBTC application does not harm CCTV traffic significantly, because of the low data rate of CBTC.

5.2 Scenario2: High Network Load

In this scenario, we consider a network topology similar to the previous one. The eNodeB serves eight UEs (tramways) distributed randomly in the coverage area of the cell. Each one executes one CBTC and four CCTV video flows (we consider CBTC and CCTV applications as described in Tab.2). The total bit rate required by CCTV application is 8 Mbps (four cameras per tramway, with 2 Mbps per camera). The simulation runs for 1000 seconds and the traffic is generated as follow:

– The applications of the first tramway start at 100 s.
– Then, every 100 s the applications of another tramway start and remain active until the simulation is ended.

Every 100s, new CBTC and CCTV applications are added to the simulation scenario. At 800 ms, the applications of the eight tramways are activated. The total network load generated by video surveillance application is 64Mbps (4*2 Mbps per tramway, 4*2*8Mbps for 8 tramways). The network uplink utilization achieves 100% at 600 s, when the applications of only six tramways are generating traffic. We consider four QoS profiles, as shown in Tab.3. These configurations are applied on each tramway.

Fig. 5 shows that mapping CCTV application to QCI 2 with GBR bearer equals to 8 Mbps blocks this application in 3 tramways (QoS-config2). The activation of CCTV application (four video flows) are accompanied with the activation of a GBR bearer. CCTV applications defined on tramway 6, 7 and 8 are

Table 3. QoS profiles

	CBTC	CCTV
QoS config1	Non-GBR bearer [QCI=5]	Default bearer [QCI=9]
QoS config2	Non-GBR bearer [QCI=5]	GBR bearer [QCI=2, UL GBR =8Mbps]
QoS config3	Non-GBR bearer [QCI=5]	GBR bearer [QCI=2, UL GBR=2Mbps]
QoS config4	GBR bearer [QCI=3, UL/DL GBR=10kbps]	GBR bearer [QCI=2, UL GBR=2Mbps]

blocked. After the activation of CCTV flows of tramway 5, no additional GBR bearers are established. Therefore, the total number of CCTV flows remains equal to 20 (4*5 flows). At 600 s, the network is saturated.There are not enough radio resources and then the control admission function rejects the new resource requests. Such behavior is not tolerated in actual wireless systems deployment. The situation could be worst, regarding the radio access network capacity in terms of data rate when transmissions are affected by interferences. Transferring data over a noisy radio channel can increase significantly the number of HARQ (Hybrid Automatic Repeat-Request) and RLC (Radio Link Control) retransmissions and reduce the modulation order when AMC (Adaptive Modulation and Coding) is activated.

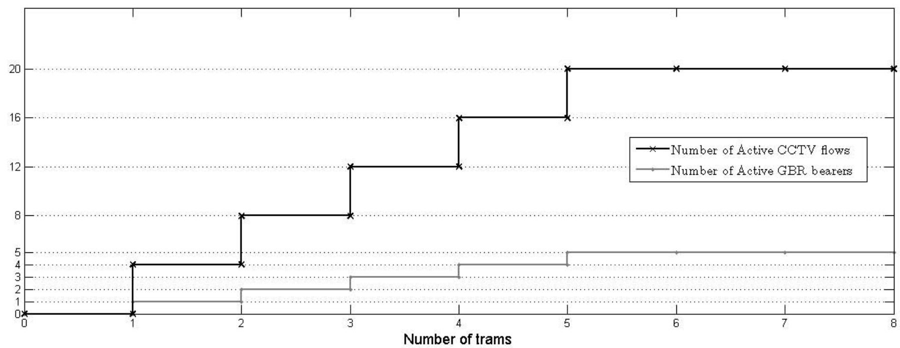


Fig. 5. The total number of CCTV flows and GBR bearers activated during simulation in the case of QoS config2

Fig. 6 gives a comparison of traffic received for CBTC application (tramway 8) with three QoS profiles. Applications defined on this tramway start at 800 s, when applications on the other tramways were already activated. In QoS-config1 and QoS-config3, we map the CBTC application to QCI 5, which specifies the high level priority. However, curves on Fig. 6 show that the use of QoS-config3 increases the packet loss of 34.1% compared with QoS-config1. In QoS-config3, we assign a GBR bearer with 2 Mbps to the CCTV application. The interest of such configuration is to generate a network load as high as the network load generated with QoS-config1, but this time we use resource allocation for CCTV application.

The results prove that using QCI 5 does not protect the critical application when the network is congested and the application coexists with CCTV traffic delivered over a GBR bearer. In QoS-config4, we mapped the CBTC application to QCI 3, corresponding to GBR resource type. Fig. 6 shows that this configuration enhances the performances compared with QoS-config3. We achieve performance level close to QoS-config1. We should notice here that these high packet losses are caused, not only by the congestion in the radio interface, but also congestion in the core network. In this scenario, the data rate of the point-to-point links between the eNodeB and the EPC is 44.739 Mbps, which is not sufficient to support a network load that exceeds 64Mbps. Based on the obtained results, we notice that the QCI 5 does not guarantee a strict priority for the critical application. In this typical LTE deployment, Proportional fair scheduling is used to serve the GBR bearers. This algorithm guarantees a minimum transport of the bit rate specified in the EPS bearer contract with delays below the values that are specified in Tab.1. Then, the remaining radio resources are distributed among Non-GBR bearers respecting their QCI's priority level.

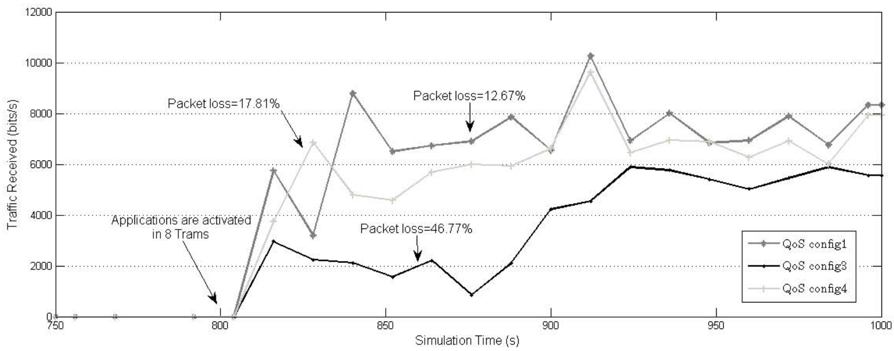


Fig. 6. Traffic received for CBTC Application $Tram \rightarrow ZC$ of tramway 8

The study of different QoS profiles when varying the network load allows us to draw the following conclusions:

- Service differentiation and prioritization of critical applications are important even with low load, when these applications share the network resources with non-critical applications. Mapping the CBTC application to a premium QCI such as QCI 5, in the first scenario, offer a minimum ETE delay.
- Using a Non-GBR bearer to deliver the critical traffic is recommended when radio resources are available. In the scenario with low load, QoS configurations, defining Non-GBR bearer (QoS-config1 and Defaut-QoS) offer the best performances in terms of ETE delay. The interest of resource allocation for

critical applications was shown in the second scenario (QoS-config4). A GBR bearer can protect the critical traffic from packet loss when radio resources are allocated to non-critical applications.
- Prioritization of critical applications has no significant impact on the performances of non-critical applications. This observation was validated for all QoS configurations presented in this work, in which the highest priority is assigned to CBTC application.
- Resource allocation is necessary to meet the stringent requirement of non-critical applications in terms of data rate. In congested network, the radio resources become insufficient, thus resource requests are rejected and the service is totally blocked. This situation is illustrated through the second scenario and the configuration QoS-config2. To avoid a high number of tramways without CCTV application we propose a controlled video coding degradation in connection with the radio resources saturation. Another solution could be the change of the admission control per flow rather that per UE. Then, when the radio resources' saturation is reached, we reduce the number of cameras per UE.

Finally, to enhance the communication performances experienced by urban guided-transport applications, we recommend the following:

- To enhance scheduling algorithms to ensure strict priority for QCI mapped to critical applications.
- To define data rate adaptation for non-critical applications like services depending on network load conditions based on cross layer functions that ensure interaction between resource management (admission control and scheduling) and application level (e.g. rate adaptive codecs).

6 Conclusion

In this paper, we consider the deployment of a LTE architecture to support CBTC and CCTV applications for a tramway. We have proposed a mapping of the two typical transport applications to standardized QCIs in LTE network as a unified communication technology for tramways. We have conducted a simulation based study to evaluate the proposed mapping. The simulation rely on the use of OPNET. We have proven that service differentiation and prioritization of critical applications is necessary even at non-congested network. The simulation results have shown that opportunistic aspect of Non-GBR QCI can be advantageous for the critical application to achieve minimum ETE delay, while GBR QCI is useful to maintain acceptable performance when network resources becomes limited. In this work, we have shown also that the prioritization of CBTC application does not harm significantly the non-critical traffic. Based on the obtained results, we have provided some general recommendations to enhance the QoS provision in such LTE deployment.

In our future work, we will enhance LTE scheduling algorithms for the use of LTE in urban-guided transport system context.

References

1. Opnet Modeler Inc., v.17.5, http://www.opnet.com
2. Wernersson, M., Wanstedt, S., Synnergren, P.: Effects of QoS scheduling strategies on performance of mixed services over LTE. In: Proc. of the 18th Intl. Symposium on Per-sonal, Indoor and Mobile Radio Communications (PIMRC 2007) (September 2007)
3. Siomina, I., Wanstedt, S.: The Impact of QoS Support on the End User Satisfaction in LTE Networks with Mixed Traffic. In: IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France (September 2008)
4. Puttonen, J., Kolehmainen, N., et al.: Mixed Traffic Packet Scheduling in UTRAN Long Term Evolution Downlink. In: IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France (September 2008)
5. Salah, M., Ali, N.A., Taha, A., Hassanein, H.: Evaluating Uplink Schedulers in LTE in Mixed Traffic Environments. In: IEEE International Conference on Communications Society ICC, pp. 5–9 (2011)
6. Muntean, V.H., Otesteanu, M., Muntean, G.-M.: QoS Parameters Mapping for E-learning Traffic Mix in LTE Networks. In: 2010 International Joint Conference on Computational Cybernetics and Technical Information (ICCC-CONTI), pp. 299–304 (2010)
7. Anas, M., Rosa, C., Calabrese, F.D., Pedersen, K.I., Mogensen, P.E.: Combined Admission Control and Scheduling for QoS Differentiation in LTE Uplink. In: IEEE 68th Vehicular Technology Conference, VTC 2008-Fall (September 2008)
8. 3GPP TS 23.203 V8.3.1, Technical Specification, Policy and charging control architecture (Release 8)
9. Olsson, M., Rommer, S., Sultana, S., Frid, L., Mulligan, C.: SAE and the Evolved Packet Core: Driving the Mobile Broadband Revolution. Academic Press (2009)
10. Sesia, S., Toufik, I., Baker, M.: LTE, The UMTS Long Term Evolution: From Theory to Practice. John Wiley & Sons (2011)

Performance of LTE in High Speed Railway Scenarios

Impact on Transfer Delay and Integrity of ETCS Messages

Aleksander Sniady and Jose Soler

Networks Technology and Service Platforms
DTU Fotonik, Technical University of Denmark
Kgs. Lyngby, Denmark
{alesn,joss}@fotonik.dtu.dk

Abstract. GSM-Railways (GSM-R) is an obsolete mobile technology with a number of shortcomings in terms of capacity and capability. These shortcomings become a major issue for railways as GSM-R may limit the number of running trains in some areas and it cannot support advanced data services. Hence, alternative technologies, such as LTE, have to be considered as a future railway communication technology.

This paper presents an analysis of transfer delay and data integrity of European Train Control System (ETCS) messages transmitted over LTE network. The analysis is made using OPNET models of a high speed railway line and LTE systems.

Keywords: GSM-R, ETCS, LTE, railway signaling.

1 Introduction

Communication networks are inevitable elements of modern railways. This is because communication networks are fundamental for vital railway services such as train command-control systems and railway emergency call. These services greatly improve railway safety and efficiency (e.g. by ensuring that trains always obey signals and by providing more detailed information to train drivers, what allows reaching higher speeds and maximize track occupancy) [1].

GSM-R is one of the most important communication networks for railways due to its growing popularity across Europe and other places around the world, where it substitutes legacy national railway communication technologies [1][2]. Despite that, GSM-R has some major shortcomings in terms of capacity and capability, which are directly inherited from the commercial GSM. Hence, the question arises if railways could replace GSM-R with a more modern mobile technology, such as LTE.

The next two sections present an overview of the shortcomings of GSM-R and benefits of LTE that could be advantageous in a railway environment. In the following part of the paper, an LTE network is analyzed on an example of one of the main Danish railway lines (Snoghøj-Odense). Using OPNET Modeler, a set of simulation scenarios is investigated, where ETCS railway signaling on the line is delivered over

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 211–222, 2013.

an LTE network. Simulation results are compared with ETCS requirements concerning data transfer delay and data integrity [3].

2 Shortcomings of GSM-R

The first major issue of GSM-R is that it offers only circuit-switched transmission. This mode of transmission is less efficient than packet-switched. Especially when considering that one of the main applications of GSM-R is delivery of bursty, low-rate ETCS data messages. The lack of packet-switched transmission leads to very low utilization of the GSM-R network [1].

Another major problem with GSM-R is its insufficient capacity, i.e. a small number of channels available for user transmission. This is a consequence of the combined effect of the circuit switched transmission paradigm and the reduced band of radio spectrum assigned. In areas with high train concentration such as central train stations there are problems with providing sufficient number of channels to serve all the trains that are to operate there simultaneously [4]. Railway companies try to overcome this problem by implementing special operational rules, such as those proposed by Banedanmark – the manager of the Danish national rail network [3], e.g.:

- train drivers are required to turn off GSM-R radio while they are at a longer stop.
- train traffic supervisors need to constantly control the number of trains in a given GSM-R cell and ensure that there are free channels available for incoming trains.

Such solutions are impractical, prone to error and they cannot solve the capacity problem entirely. This means that the capacity of the GSM-R network becomes a bottleneck limiting the number of trains to be operated in a given area. Desirably, the only limitation should be related to the capacity of the railway infrastructure.

Finally, the last shortcoming of GSM-R is its very limited support for data communication. The maximum transmission rate per connection is limited to just 9.6 Kbit/s [1], what is sufficient only for applications with very low demands. Apart from that, message delay is in the range of 400 ms [1], what is too high to allow any interactive real-time application. Lastly, long connection setup time, which is in the range of 7 seconds [1], heavily impacts applications that require rapid setup, e.g. applications for emergency situations.

These shortcomings (the lack of packet-switched transmission, limited capacity and limited support for data communication) show how outdated GSM-R technology is from a telecommunication point of view. This is especially apparent if GSM-R is compared with commercial mobile networks, which already underwent few major evolutions from the original GSM technology. The most significant standards released in Europe, after GSM were GPRS, UMTS and LTE. These new standards brought various improvements in the capabilities of the mobile networks [5]. Railways could benefit from these modern technologies. It is especially important considering that some commercial mobile operators already consider shutting down their GSM networks, so GSM is a technology becoming quickly obsolete. In the beginning of 1990's it was decided to build the railway communication technology on the basis of an already available, well-established mobile communication standard [1]. That is one of the main reasons why GSM was chosen for railways. Hence, now when GSM is being slowly abandoned by commercial operators, one of the main reasons for adopting GSM-R by railways becomes invalid.

The technology that should be considered as the most likely alternative to GSM-R is LTE. This is because it brings number of benefits over previous mobile communication technologies – both GSM and the later UMTS. These benefits concern such improvements as a more efficient core network, reduced packet delay and a high throughput radio access [5]. Details of these improvements are described in the following section. Apart from the technical benefits, LTE is the most modern mobile communication standard that is being deployed commercially around the world. That is advantageous in terms of economy, because of the reduced obsolescence span.

3 Benefits of LTE

There are over 20 years of development separating GSM and LTE technologies, what makes them very different in many aspects. However, here, the focus is put on a few main advantages of LTE that could be highly beneficial from the railway perspective.

The key element differentiating LTE from GSM-R is that an LTE network is based on packet-switched transmission. It is the first mobile technology adopting the all-IP approach abandoning circuit-switched transmission. Packet switched transmission is more flexible in managing available network resources. Thanks to this, it increases network utilization and reduces waste of limited network resources. Despite the lack of circuit-switched mode, LTE includes Quality-of-Service mechanisms that provide packet differentiation. This could be applied to protect railway safety-critical applications such as ETCS [5].

Another advantage of LTE network is the reduced packet delay, which is one of the crucial requirements for providing ETCS messages. This is achieved by simplifying the network architecture. The LTE network has less logical and physical elements and they are all based on a common technology (IP).

Finally, LTE offers much higher throughput over its radio access thanks to a more advanced radio interface. It consists of a number of improvements that increase spectral efficiency of LTE in comparison to older technologies (GSM and UMTS):

- Advanced multiplexing – Orthogonal Frequency Division Multiplexing (OFDM).
- More advanced modulation – up to 64 Quadrature Amplitude Modulation (QAM).
- Sophisticated transceiver – Multiple Input Multiple Output (MIMO) technology.

The additional throughput can be consumed in various ways: to serve more users, to provide more applications or to provide bandwidth-demanding applications, which cannot be provided over the low-rate GSM-R radio interface.

Summing up, LTE brings important improvements over GSM-R. But the question that needs to be answered is whether it can also fulfill all the railways requirements in terms of performance and reliability, in order to become a viable alternative to GSM-R.

4 LTE as a Railway Communication Technology

The general question of whether LTE can be a railway communication technology has been divided into three smaller questions:

4.1 Can LTE Support Safety-Critical Railway Applications, Such as ETCS?

GSM-R was designed to provide two fundamental services: transmission of The European Train Control System (ETCS) messages and voice communication for railways. ETCS is a digital wireless railway signaling system that replaces legacy national signaling systems used around Europe. Its main goal is to provide safe and efficient command-control system that ensures international interoperability. ETCS becomes widely adopted across Europe thanks to its technical benefits, but also due to the European Union directives, which oblige railways to adopt it.

Thus, any railway communication technology that could be considered an alternative to GSM-R needs to support ETCS, i.e. it needs to fulfill all the transmission requirements for reliable and timely delivery of ETCS messages. These requirements concern parameters such as [6]:

- Received signal power.
- End-to-end delay.
- Data rates.
- Probability of connection loss.
- Maximum break during handover.
- Bit error rate.
- Connection establishment delay.
- Connection establishment failure probability.

The requirements listed above concern circuit-switched transmission of ETCS messages. Thus, for LTE based ETCS transmission, there is a need to redefine these requirements to packet-switched transmission. This redefinition has not been finalized by the International Union of Railways (UIC). However, the Danish Signaling Program defined tentative requirements for packet-switched transmission of ETCS messages which are published in [3]. These requirements concern parameters such as:

- Data transfer delays.
- Data integrity: probabilities of packet loss, duplication, out-of-sequence delivery and corruption.
- Network attach procedure delay.
- Packet Data Protocol (PDP) context activation delay.

4.2 Can LTE Support All the Advanced Voice Functionality Provided by GSM-R?

Voice communication is still a very important service for railways. Railway communication technology needs not only to provide point-to-point calls as commercial mobile telephony does, but also to provide railway-specific features such as group calls, broadcast calls, call prioritization and advanced addressing based on location or function (e.g. calling a train dispatcher responsible for a given area). The main problem here is the lack of a single widely accepted technical solution for providing voice communication over LTE [7].

4.3 Can LTE Bring Real Improvements for Railways in Terms of Capacity and Supported Applications While Still Fulfilling the Requirements of ETCS?

It should be verified whether the benefits of LTE listed in the previous section bring an actual improvement for railways and whether QoS mechanisms are efficient enough to use LTE for combining safety and non-safety applications.

5 OPNET Simulations

Since we have already gathered and analyzed sufficient data, we focus this paper on answering the first of the previously presented questions: whether LTE can fulfill packet-switched requirements on the delivery of ETCS messages [3]. We will focus on the other questions in following papers.

There are two factors that could limit performance of an LTE network in a railway environment: train speed (User Equipment speed) and LTE network load. In the following part of the paper the first of the two limiting factors is analyzed, i.e. relation between train speed and performance of LTE network.

The proposed simulation scenarios model a high-speed railway line, where it is required to provide continuous, reliable connectivity between train On-Board Unit (OBU) and Radio Block Controller (RBC) - ETCS server supervising train movement. In such scenario, performance of LTE transmission may be limited by the train speed. Thus, the purpose of these simulations is to verify whether LTE can fulfill the ETCS requirements at high train speeds.

The modeled scenarios are based on a railway line between Snoghøj and Odense (Denmark). An overview of the line is shown in Figure 1. It is one of the most important railway lines in Denmark for both national and international traffic. Currently, the line is operated at speeds up to 180 km/h [8]. In the future, the line may be upgraded to allow higher travel speeds up to 200 km/h or more [9]. However, speeds up to 500 km/h are considered in the simulations as this is the top speed that needs to be supported according to requirements for GSM-R [6].

Fig. 1. Snoghøj-Odense railway line with the location of intermediate train stations Map source: [10]

Snoghøj-Odense is a double-track line with length of 54.5 km. In order to provide LTE coverage over the whole line it is required to set up multiple eNodeBs. This means that trains need to perform multiple handovers between LTE cells while travelling over the line. In the OPNET simulation 11 eNodeBs are placed. Their location follows the proposed location of 11 real GSM-R BTSs, which are planned to be deployed along the Snoghøj-Odense line by Banedanmark [11]. Each eNodeB has three LTE sector antennas operating in a 5 MHz bandwidth. LTE supports 1.4MHz, 3MHz, 5MHz, 10MHz, 15MHz and 20MHz bandwidths [5]. The 5 MHz bandwidth has been chosen as it is the closest to the size of GSM-R bandwidths in Europe, where GSM-R operates in 4 or 7 MHz bandwidths [1].

The eNodeBs are connected to Evolved Packet Core (EPC), which, in the OPNET model, simulates the LTE network backbone and includes functionality of a number of separate nodes from the LTE standards: Mobility Management Entity (MME), Home Subscriber Server (HSS), Serving Gateway (GW), and Packet Data Network Gateway (PDN GW). EPC provides connectivity with the Radio Block Controller (RBC). This network topology is shown in Figure 2.

There are up to 15 trains passing through the line in the peak hour [8]. Assuming that all of them are equipped with ETCS On Board Units (OBUs) then each train is a mobile terminal, which uses the LTE network to communicate with the RBC.

Every OBU sends an ETCS message to the RBC at time intervals following exponential distribution with a mean value of 30 seconds. Also the RBC sends an ETCS message to each of the OBUs at time intervals following an exponential distribution with a mean value of 30 seconds. These time values are based on the assumption that a movement authority and a position update messages are transferred every single minute. In our model, ETCS messages have constant size of 128 bytes according to the size specified in the ETCS requirements [3].

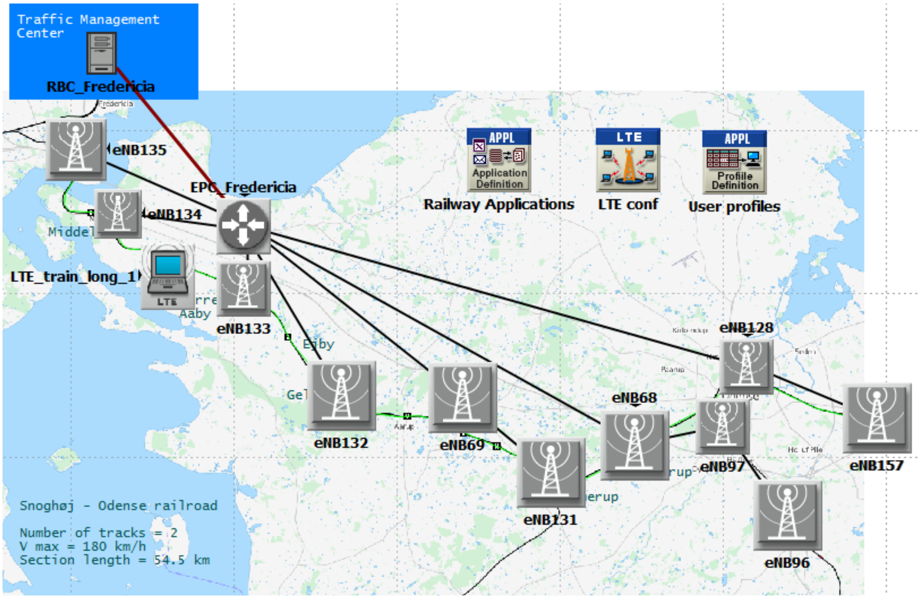


Fig. 2. Topology of the LTE network along the Snoghøj-Odense railway line

Two scenarios are considered:

Scenario 1. There is a single train travelling back and forth on the line. The purpose of this scenario is to measure performance of delivery of ETCS messages in terms of delay and data integrity in an unloaded network.

Scenario 2. Additional trains are introduced simulating train traffic in the peak hour. There are 15 trains travelling over the line per hour (total for both directions). The purpose of this scenario is to investigate how 14 additional trains affect transmission of ETCS messages by the initial single train.

Each of the two scenarios contains 13 subcases with train speeds between 25 km/h and 500 km/h. Each case was executed 20 times with varying seed numbers. The length of each simulation run was 4 hours.

6 Simulation Results

Two sets of simulation results are analyzed for each of the scenarios: one concerning the transfer delay of ETCS messages, while another concerning integrity of the delivered ETCS messages.

6.1 Transfer Delay Analysis Results

According to the requirements presented in [3] the mean transfer delay of 128 byte ETCS message is required to be lower than 0.5 s. Moreover, 95% of ETCS messages have to be delivered within 1.5 s.

Simulation results concerning ETCS message delays are presented in Table 1. All of the recorded values fulfill these transfer delay requirements:

- Mean transfer delay is in the range of 0.05 – 0.07 s
- 100 % of messages is delivered within 1.5 s

Table 1. Transfer delay of ETCS messages delivered over LTE network

Train Speed [km/h]	Single train scenario		15 train scenario	
	Mean transfer delay [s]	*Messages delivered within 1.5 s [%]*	*Mean transfer delay [s]*	*Messages delivered within 1.5 s [%]*
25	0.050	100 %	0.050	100 %
50	0.050	100 %	0.052	100 %
75	0.051	100 %	0.051	100 %
100	0.051	100 %	0.050	100 %
125	0.051	100 %	0.053	100 %
150	0.052	100 %	0.051	100 %
175	0.052	100 %	0.055	100 %
200	0.053	100 %	0.055	100 %
250	0.053	100 %	0.056	100 %
300	0.055	100 %	0.058	100 %
350	0.056	100 %	0.054	100 %
400	0.058	100 %	0.057	100 %
500	0.063	100 %	0.063	100 %

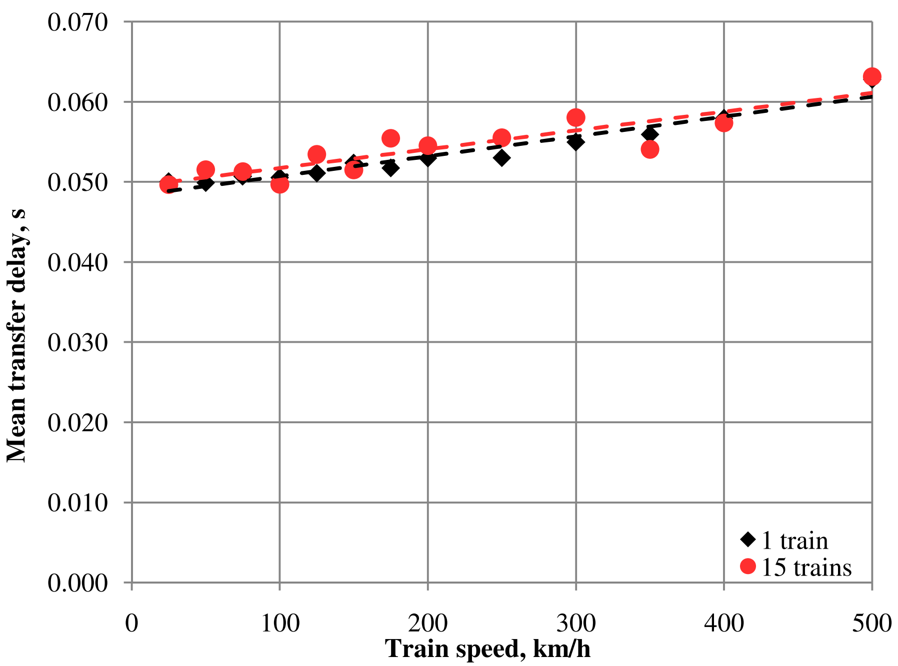


Fig. 3. Mean transfer delay of ETCS messages in relation to train speed

Furthermore, as it can be seen in Figure 3, the mean delay of ETCS messages slightly increases with the train speed. However, as the delay increase is not significant, the ETCS requirements are fulfilled even at high train speeds.

Lastly, results obtained in both scenarios (single train and 15 trains) are comparable. This means that the modeled LTE network provides sufficient resources to support ETCS signaling over the Snoghøj-Odense line, with its maximum frequency of 15 trains per hour.

6.2 Data Integrity Analysis Results

Another set of simulation results concerns data integrity. There are four requirements on the ETCS message delivery in this regard [3]:

- Probability of data loss < 10^{-4}
- Probability of data duplication < 10^{-5}
- Probability of data being out-of-sequence < 10^{-5}
- Probability of data corruption < 10^{-6}

It has to be noted that an LTE network provides retransmission mechanisms over the radio link (at the MAC layer and the RLC layer) [5]. Moreover, the data is protected by retransmission mechanisms at the end-to-end transport protocol (the TCP layer). Thus, the successful delivery of messages (lack of data loss, duplication, out-of-sequence delivery or corruption) can be measured both at the radio link and at the end-to-end connection as shown in Table 2.

Table 2. Data integrity of ETCS messages delivered over LTE network

Train Speed [km/h]	Ratio: delivered error-free packets / sent packets:			
	Single train scenario		15 train scenario	
	LTE radio link (RLC layer) [%]	*End-to-end connection (TCP layer) [%]*	*LTE radio link (RLC layer) [%]*	*End-to-end connection (TCP layer) [%]*
25	99.9995 %	99.974 %	99.9993 %	99.958 %
50	99.9990 %	99.985 %	99.9986 %	99.896 %
75	99.9994 %	99.953 %	99.9984 %	99.958 %
100	99.9992 %	99.969 %	99.9993 %	99.958 %
125	99.9993 %	99.948 %	99.9982 %	99.844 %
150	99.9989 %	99.908 %	99.9989 %	99.898 %
175	99.9990 %	99.907 %	99.9986 %	99.773 %
200	99.9990 %	99.886 %	99.9980 %	99.792 %
250	99.9991 %	99.881 %	99.9983 %	99.779 %
300	99.9988 %	99.828 %	99.9977%	99.702 %
350	99.9986 %	99.783 %	99.9982 %	99.801 %
400	99.9980 %	99.679 %	99.9976 %	99.734 %
500	99.9971 %	99.567 %	99.9952 %	99.564 %

As it can be seen in the “LTE radio link” column in Table 2, over 99.99% of packets are delivered successfully over the LTE radio link and these do not require retransmission. The remaining packets are retransmitted by the RLC layer which works in the acknowledged mode, what means that the Automatic Retransmission Request (ARQ) feature is enabled [5]. Looking at the results concerning the end-to-end connection, also shown in Table 2, it can be seen that over 99.5% of packets are delivered successfully without the need for retransmission. The remaining 0.5% of packets are retransmitted by the mechanisms of the TCP layer. It should be noted that the retransmission mechanisms at the RLC and the TCP layers are not correlated. Thus, an erroneous packet can be retransmitted by the TCP layer while its retransmission could have been already requested by the RLC layer.

Thanks to the combined multilayer error detection and correction mechanisms, 100% of the ETCS messages are delivered correctly to the ETCS application layer, what fulfills the data integrity requirements.

Figures 4 and 5 show the percentage of successful (error-free) packet transmissions in relation to train speed over the radio link and over the end-to-end connection, respectively. As it can be seen in the figures, with increasing train speeds the success rate of the transmission is decreasing. However, in all the cases, regardless of the train speed, the network was able to timely recover. This was shown in the transfer delay analysis where no packets were delayed more than 1.5 s.

Finally, as in the case of transfer delay, the results obtained in both scenarios (single train and 15 trains) are comparable. Thus, the network had no problem to serve ETCS traffic generated by the trains on the line.

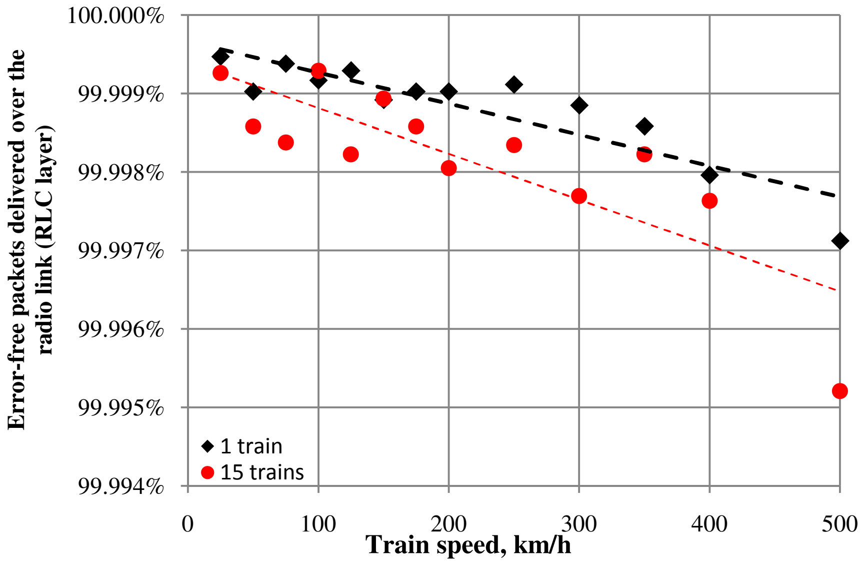


Fig. 4. Ratio of the delivered error-free packets to the sent packets over the radio link (at RLC layer) in relation to train speed

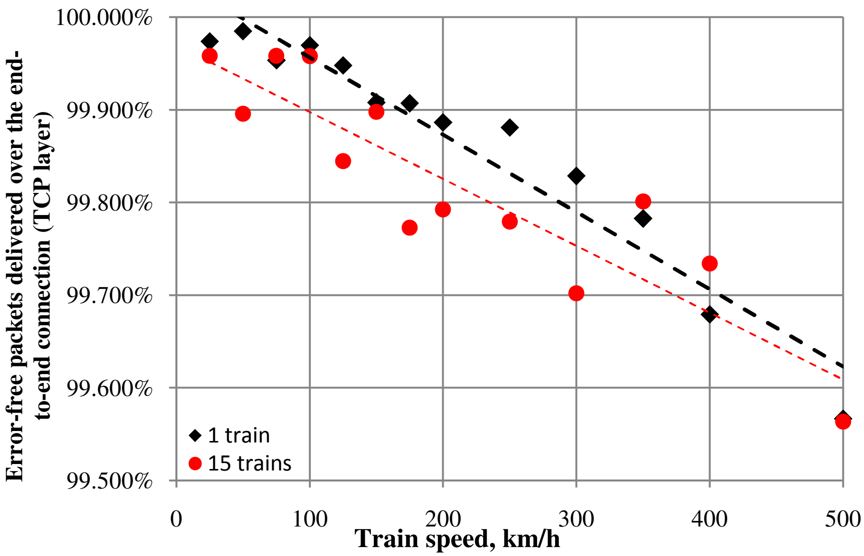


Fig. 5. Ratio of the delivered error-free packets to the sent packets over the end-to-end connection (at TCP layer) in relation to train speed

7 Conclusions

This paper presented an OPNET simulation model of a railway line where GSM-R network, which is used for providing ETCS signaling between train OBU and RBC, was substituted with the more modern LTE network.

Simulation results show that the modeled LTE network has no problems in providing connectivity between train OBU and RBC. This fulfills ETCS transmission requirements in terms of delay and data integrity. Increasing the train speed decreases the quality of the OBU-RBC communication. Nevertheless, the ETCS requirements are still fulfilled at any investigated speed in the range from 25 km/h to 500 km/h.

What is more, the recorded transfer delays, which are one order of magnitude lower than the limits set by the ETCS requirements, suggest that the LTE network has resources to serve many more users or to provide additional applications for the existing users.

Acknowledgments. This paper is supported by the Danish Council for Strategic Research through the RobustRailS project, strengthening the railway research in Denmark: http://www.dtu.dk/subsites/robustrails/English.aspx

References

1. Winter, P., et al.: Compendium on ERTMS. Eurail Press (2009)
2. UIC, ERTMS Atlas 2012, 10th UIC ERTMS World Conference in Stockholm (April 2012)

3. Fisher, D.G.: Requirements on the GSM-R Network for ETCS Support, Banedanmark (2008)
4. Radio Spectrum Committee, Working Document: Spectrum Requirements for Railway Applications, European Commission (2010)
5. Sauter, M.: From GSM to LTE: an introduction to mobile networks and mobile broadband. A John Wiley and Sons, Ltd. (2011)
6. ERTMS, SUBSET-093: GSM-R Interfaces; Class 1 requirements (2005)
7. Gessner, C., Geralch, O.: Voice and SMS in LTE; White paper. Rohde & Schwarz (2011)
8. Banedanmark, Signalling System Model, Banedanmark (2008)
9. Trafikministeriet (Denmark), Modernisering af jernbanens hovednet (1997)
10. OpenStreetMap, `http://www.openstreetmap.org`
11. Banedanmark, COWI, VVM-screening af radiokommunikationsmaster, Banedanmark (2010)

Generating Test Scenarios Based on Real-World Traces for ERTMS Telecommunication Subsystem Evaluation

Patrick Sondi[1,*], Marion Berbineau[1], Mohamed Kassab[1], and Georges Mariano[2]

[1,2] Univ Lille Nord de France, F-59000 Lille,
[1] IFSTTAR-LEOST, F-59650 Villeneuve d'Ascq, France
{patrick.sondi,marion.berbineau,mohamed.kassab}@ifsttar.fr
[2] IFSTTAR-ESTAS, F-59650 Villeneuve d'Ascq, France
georges.mariano@ifsttar.fr

Abstract. The European Rail Traffic Management System (ERTMS) allows increasing the utilization of the track by using a wireless mobile technology, the GSM-R, for dynamic signaling and control of the trains. However, the GSM-R will become the bottleneck regarding the performance of the ERTMS due to the increase of both the trains' traffic, and the resources needed by all the applications that contribute to safety in rail transportation and to satisfaction of the users. In this context, it is necessary to develop the tools that will allow evaluating the impact on the GSM-R of the scenarios that could occur, but that are costly or difficult to test on real tracks. In this paper, we propose an approach that consists in using real-world traces of trains moving on ERTMS lines to build specific scenarios for evaluating the ERTMS telecommunication subsystem. In this way, we are able to guarantee that these evaluations are very realistic due to the fact that the train moves following the functional specification of ERTMS. Moreover, our approach is particularly interesting for the future of ERTMS since the same scenarios can be used to evaluate other prospective wireless mobile technologies that could replace the GSM-R, such as LTE.

Keywords: ERTMS, GSM-R, Rail, Transport, Real Traces.

1 Introduction

The European Rail Traffic Management System (ERTMS) [1] has been introduced by the European Union in order to harmonize and ensure the interoperability of the railroads of the different countries. The ERTMS specifications define two major components. The first one is the functional subsystem identified as the European Train Control System (ETCS), which ensures train signaling and control. The ETCS consists in several applications embedded on the train or located in the control center, and which exchange information in order to ensure safety of all trains movements on the different tracks. The second component is the telecommunication subsystem which ensures the communication between the train and the control center. This subsystem ensures

* Corresponding author.

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 223–231, 2013.

the transmission of critical information for the safety of the train movement, and therefore of the passengers and the entire infrastructure. For this reason, the ERTMS specifications impose stringent constraints on the telecommunication technology that could be used in this context. Using a wireless mobile telecommunication technology presents the advantages of covering bigger area with less equipment, and controlling the trains continuously, not only at specific points. In this way, the ERTMS has improved dynamic signaling and control of the trains, while allowing an increase of the traffic that can be ensured at constant capacity of the tracks. The GSM-R (Global System for Mobile communication – Railway) was chosen as an ERTMS standard since it was the most popular mobile telecommunication technology used by the European operators and, according to the European Railway Agency evaluations [2], it met the QoS requirements imposed by ERTMS specifications.

The bandwidth allocated to the GSM-R is limited (4 MHz downlink – 4 MHz uplink, in France), thus with the increasing traffic it could become difficult to maintain all the trains simultaneously and continuously connected, especially in the areas with high track density. Some of the shortcomings of the GSM-R technology have been already emphasized by other researchers [3]. The industrials have already introduced in the equipment new mechanisms that allow ETCS applications to work with a train which is not continuously connected, in order to reuse the resources. This situation illustrates that this scenario could have been anticipated if the appropriate tests had been performed, and therefore that the ETCS applications should have not been designed to work with a train continuously connected. However, it is also obvious that performing tests on real tracks in order to study such a scenario would have been costly and difficult to realize. In this context, using simulation tools to evaluate the impact of various parameters on ERTMS telecommunication subsystem in various scenarios, would help anticipating on many situations that may have not been experienced at the moment. The same reasoning could be applied to study, with appropriate tools, and try to overcome the electromagnetic compatibility [4] and the localization [5] challenges that are involved in the ERTMS deployment.

In this paper, we propose to use the real-world traces of the trains moving on ERTMS lines in order to generate specific scenarios for evaluating the behavior of the ERMTS telecommunication subsystem. The remainder of this paper is organized as follows: in section 2, we present a brief overview of the main contributions related to ERTMS telecommunication subsystem evaluation by simulation. Section 3 presents our global approach for generating test scenarios based on real-world traces. We then describe how such scenarios can be implemented in the OPNET simulator in order to evaluate a targeted telecommunication technology. Section 5 concludes the paper.

2 Related Work

Several work proposed interesting studies based on simulation evaluation of the GSM-R [6], and other telecommunication technologies such as GPRS [6], Wimax [7] and LTE [8]. In these work, the telecommunication infrastructure is modeled using the components provided by the OPNET simulator modeler, and customized in order to adapt to railway constraints. The ETCS applications are partially modeled and interesting results are obtained for various speed related to the train movement. However, though these experiments allow gathering useful information on the behavior of the system in specific situations, the movement of the train and the messages exchanged do not reflect exactly an ERTMS scenario. Indeed, since the functional subsystem which is responsible of signaling and train control is not modeled, the mechanisms that should allow, stop or modify the train movement are not performed. Therefore, it is not possible to reproduce realistic traffic conditions; especially in areas with high track density where several trains are probably stopped, while some are moving slowly and others are moving fast on the close high speed lines. The same reasoning can be applied to the messages exchanged between the train and the control center.

In a recent work [9], the authors proposed to connect the OPNET simulator with a simulator dedicated to ERTMS functional subsystem in a co-simulation virtual laboratory. In this way, both simulators could exchange information and simulate the related component while taking into account the behavior of the other. There exist actually several ERTMS simulation tools [10] that allow simulating the functional behavior of the ERTMS, and which could be coupled with a telecommunication simulator. However, such approach requires designing and proving the accuracy of the synchronization process, while ensuring the coherency and validity of the results. Such a co-simulation platform will be useful, especially when one needs to evaluate both ERTMS subsystems.

However, if we reduce the problem only to the evaluation of the ERTMS telecommunication subsystem, a simpler solution can be investigated to obtain realistic train movement and realistic sequence of messages exchanged with the control center.

3 Generating Test Scenarios Based on Real-World Traces

The approach proposed in this paper consists in using the traces generated by onboard and roadside equipment in order to reproduce a complete and realistic scenario, with the same train movement and the same sequence of messages exchanged in the time, in the simulator that models the telecommunication subsystem such as OPNET.

Time	n	KP	Spee	Source of dat	RB	I	ETCS Message	Bytes	ETCS Application Messa
16:24:25	633			TRAIN	1	155	Initiation of a communication session	18	9B028000026500000040FF0
16:24:26	538			RBC	1	32	Configuration Determination	19	2002C00002651FFFFFE43F
16:24:27	133			TRAIN	1	159	Session established	21	9F034000028A80000040C03
16:24:27	276	34918,6	10	TRAIN	1	129	Validated Train Data	47	8109C000028AC000004000
16:24:28	037			RBC	1	8	Acknowledgement of Train Data	22	08038000028AFFFFFFE000
16:24:28	120			TRAIN	1	146	Acknowledgement	22	9203800002A30000004000
16:24:43	037			RBC	1	24	General Message	18	180280000401FFFFFFFA2
16:24:43	120			TRAIN	1	146	Acknowledgement	22	92038000041A0000004000
16:24:58	037			RBC	1	24	General Message	18	180280000578FFFFFFFF90
16:24:58	120			TRAIN	1	146	Acknowledgement	22	920380000591000000400005
16:25:13	037			RBC	1	24	General Message	18	1802800006EFFFFFFFFFF6
16:25:13	121			TRAIN	1	146	Acknowledgement	22	92038000070800000040000
16:25:28	037			RBC	1	24	General Message	18	180280000866FFFFFFFF69
16:25:28	223			TRAIN	1	146	Acknowledgement	22	92038000087F0000004000
16:25:43	037			RBC	1	24	General Message	18	1802800009DDFFFFFFFF6A
16:25:43	225			TRAIN	1	146	Acknowledgement	22	9203800009F60000004000
		…		…			…		
16:57:43	951	44336,3	120	TRAIN	1	136	Position Report	32	88060000C586C000004000
16:57:44	683			RBC	2	3	Movement Authority	100	03170000C5876D480B81E0
16:57:44	819			TRAIN	2	146	Acknowledgement	22	92038000C59F8000004000C
16:57:45	176	44369,6	120	TRAIN	2	136	Position Report	32	88060000C5A88000004000E
16:57:45	372	44369,6	120	TRAIN	1	136	Position Report	32	88060000C5A94000004000E
16:57:48	801	44503,0	120	TRAIN	2	136	Position Report	32	88060000C6030000004000E
16:57:48	982	44503,0	120	TRAIN	1	136	Position Report	32	88060000C603C000004000
16:57:49	535			RBC	1	24	General Message	32	18060000C603ED480B8550
16:57:49	658			TRAIN	1	146	Acknowledgement	22	92038000C6180000004000C
16:57:49	857			TRAIN	1	156	Termination of a communication session	18	9C028000C61900000040C4
16:57:50	537			RBC	1	39	ACK of termination of a communication session	18	27028000C61CCD480B9F60

Fig. 1. ERTMS level 2 real-world traces of a train crossing two RBCs

3.1 A Description of Real-World Traces of a Train on ERTMS Lines

The traces illustrated by a sample in figure 1 were provided by an equipement manufacturer in the context of a research project on ERTMS components evaluation. These traces report mainly the sequence of the messages exchanged between the train and the control center via the RBCs (Radio Block Centers) encountered during its movement. The data included in the report are described in table 1.

Table 1. Structure of the traces and description of the data fields

Data Field	Description
Time	the time at which the message is transmitted by the sender
Ms	a precision on the time (in milliseconds)
KP (Kilometric Point)	the relative coordinates of one specific point on the track
Speed	the maximum speed for train movement control
Source of data	the sender of the related message (the train or the RBC)
RBC	the identity of the RBC sending or receiving the message
M#	the unique identifier of the message being sent
ETCS Message	the type of the message in ETCS specifications
Bytes	the size of the message in bytes
ETCS Application data	the message in HEX data (Euroradio MAC included)

3.2 Exploiting Real-World Traces Data for Modeling Simulation Objects

The information contained in the real-world traces data described in table 1 allows a faithful modeling of the following simulation objects related to the scenario:

- the tracks and roadside infrastructure : based on the line and the KPs, it is possible to retrieve the geographic location of the tracks implied in the train movement. The handover area where the train exchanges with both RBCs allows fixing the location and the coverage of the telecommunication infrastructure components (Fig. 2);
- the trajectory : using the successive KPs and the corresponding time, it is possible to obtain, respectively, the distance on the segments and the corresponding speed. When a segment is too long due to missing intermediary KPs, the information provided by the field “Speed” allows refining the evolution of the train on the segment. Indeed, intermediary nodes positions may be introduced by linear interpolation or exponential smoothing in such a way that the speed on each segment stay lower than the value in the field “speed” during the corresponding time;
- the sequence of the messages exchanged : each message is associated with a transmission time under a precision of milliseconds. At each time a message, which size is given in the data, is sent from the source to the destination, both provided by the traces data. The position of the train on the track at the moment the message is sent can be added to the trajectory as an intermediary KP. This is particularly useful on long segments. In addition, these points can serve in order to improve the refinement of the speed over the trajectory.

Based on the traces data, the simulation objects can be implemented inside a network simulator which contains a model of the telecommunication subsystem technology. The elements of the infrastructure will be placed as described in figure 2. During the simulation, the train will be moving exactly as set in the trajectory object and it will send, or a message will be sent to it, exactly at the time provided in the traces data. In this way, it is possible to replay very realistic ERTMS scenarios from a functional point of view of several trains moving simultaneously on the different tracks in the same area, and study the impact of their communications with the control center on the telecommunication subsystem. Moreover, it is possible to change the architecture of the network, or replace one telecommunication technology by another, and observe the behavior of the entire system.

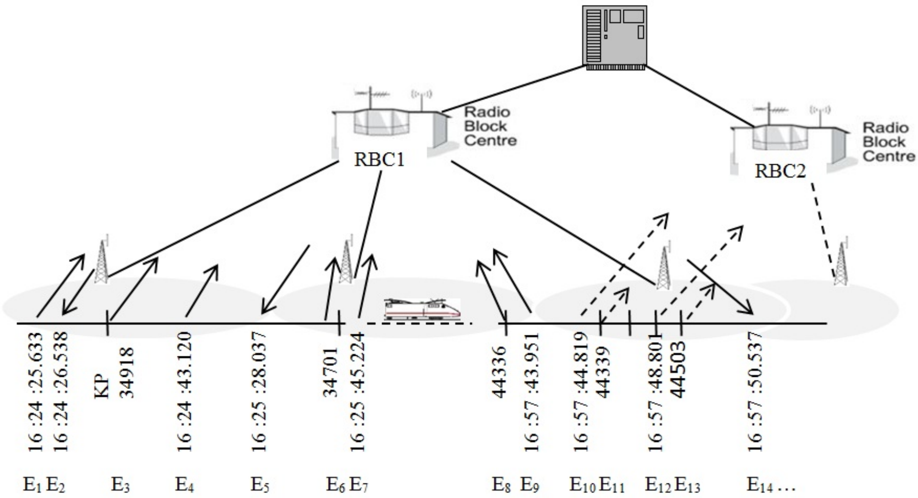


Fig. 2. The infrastructure and the events extracted from the real-world traces data

3.3 Formal Description and Transformation of Traces Data for Simulation

In order to allow automatic generation of simulation objects based on real-world traces data, it is necessary to formalize both the description and the transformation of these data, and also to propose the mechanisms that check the coherency of the result.

The Frame of Reference

The traces data are clearly based on the messages. Therefore, we can consider each message as an event described by some properties. In this way, the data form a space of events, and we can define a coordinate system *(X,t,e)* where *X* represents the coordinate of the point at which an event occur, *t* the time and *e* the description of the event. The third coordinate is necessary to support the existence of several events at the same point and the same time, due to loss of precision for example.

Transformation Rules

Each line of the data is an event E_i. At least a time t_i and an event description e_i (the message and related information) are known for this event. However, for some events, the point coordinates X_i and the speed v_i are missing. Let E_j and E_{j+k+1} two events such that X_j, X_{j+k+1}, v_j and v_{j+k+1} are known, while all the information X_i and v_i are unknown for all the k events E_{j+1}, E_{j+k}. Based on the time values t_{j+i}, the speeds v_{j+i} $(1 \leq i \leq k)$ are computed approximately using the following equation:

$$v_{j+i} = v_j + (t_{j+i} - t_j)\frac{(v_{j+k+1} - v_j)}{(t_{j+k+1} - t_j)} \tag{1}$$

The related coordinates X_{j+i} $(1 \leq i \leq k)$ can be also computed approximately using the following recursive equation:

$$X_{j+i} = X_{j+i-1} + \frac{1}{2}(v_{j+i-1} + v_{j+i})(t_{j+i} - t_{j+i-1}) \tag{2}$$

Coherency-Checking and Correction Mechanisms

The equations (1) and (2) allow rapid transformation, but they may introduce some errors. The following rules are applied to check and correct inconsistencies:

- the speed limit must be respected: *If* $v_{j+i} \geq Max\{v_j, v_{j+k+1}\}$ *then* $v_{j+i} = Min\{v_j, v_{j+k+1}\}$
- the space structure must be preserved: *if* $X_{j+i} \geq X_{j+k+1}$ then: $X_{j+i} = 0.5(X_{j+i-1} + X_{j+i+1})$

4 Use Case: Generating a Scenario Based on Traces in OPNET

The mechanisms proposed in section 3 have been applied in order to obtain a scenario based on the aforementioned real-world traces data in a telecommunication technologies simulator, namely OPNET. As a proof of concept, the simulation objects are modeled step by step following the same process.

4.1 Modeling the Infrastructure

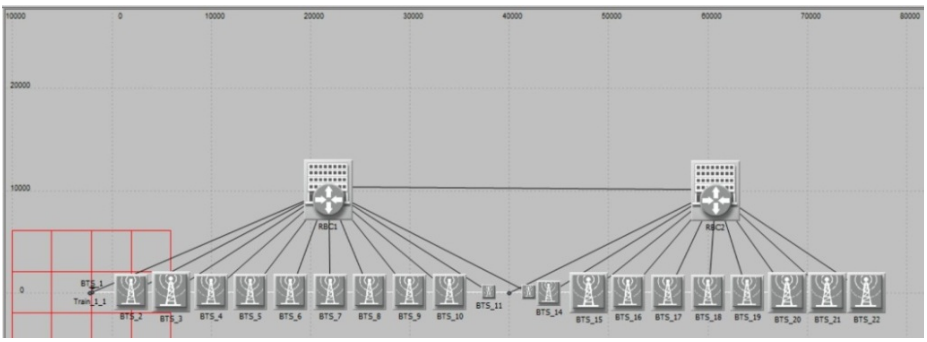


Fig. 3. The infrastructure obtained based on real world traces data

The complete traces data cover a railroad line over 78935 meters. The frame of reference *(X,t,e)* is modified in such way that -2049 ≤ X ≤ 76886. We observe that the coverages of both RBC1 and RBC2 have a common area including the line from the coordinates X_j=38503 m to X_k=44836 m. The BTS are placed every 4 km (Fig. 3).

4.2 The Trajectory

The trajectory is one of the most important objects due to the fact that, without having the functional subsystem implemented in the telecommunication simulator, it is the only element ensuring that the train follows the same movement as if it was on a real ERTMS line, and controlled by ETCS applications.

Modifying the Frame Reference Time

In order to harmonize the time of the scenario with the time of simulation, it is necessary to modify the time origin in the frame reference in such a way that the starting date of the scenario becomes 00:00:00.000ms in the simulator (see Fig. 4 under the "Accum Time" column). At the end of this column one can see the scenario duration.

Trajectory name: real_trace_ertms

	X Pos (m)	Y Pos (m)	Distance (m)	Altitude (m)	Traverse Time	Ground Speed	Ascent Rate (m/sec)	Wait Time	Accum Time
1	-2,049.500000	0.000000	n/a	0	n/a	n/a	n/a	0	00.00s
2	-2,041.800000	0.000000	7.7007	0	03.00s	9.2409	0	0	03.00s
3	-2,001.700000	0.000000	40.1000	0	15.00s	9.6240	0	0	18.00s
4	-1,960.400000	0.000000	41.2999	0	15.00s	9.9120	0	0	33.00s
5	-1,919.600000	0.000000	40.8000	0	15.00s	9.7920	0	0	48.00s
6	-1,878.900000	0.000000	40.7001	0	15.00s	9.7680	0	0	1m03.00s
7	-1,832.800000	0.000000	46.1001	0	15.00s	11.0640	0	0	1m18.00s
8	-1,796.600000	0.000000	36.2000	0	15.00s	8.6880	0	0	1m33.00s
45	-119.800000	0.000000	95.9000	0	04.00s	86.3100	0	0	10m17.00s
46	-95.900000	0.000000	23.8999	0	01.00s	86.0397	0	0	10m18.00s
47	0.000000	0.000000	95.9000	0	04.00s	86.3100	0	0	10m22.00s
48	256.000000	0.000000	256.0000	0	09.00s	102.4000	0	0	10m31.00s
49	339.500000	0.000000	83.5000	0	03.00s	100.2000	0	0	10m34.00s
610	76,686.100000	0.000000	66.6001	0	02.40s	99.9002	0	0	49m39.30s
611	76,719.500000	0.000000	33.4001	0	01.20s	100.2002	0	0	49m40.50s
612	76,752.800000	0.000000	33.2997	0	01.20s	99.8992	0	0	49m41.70s
613	76,752.900000	0.000000	0.0000	0	04.00s	0.0000	0	0	49m45.70s
614	76,886.200000	0.000000	133.3000	0	04.80s	99.9750	0	0	49m50.50s

Fig. 4. The trajectory implemented in OPNET based on the traces data

In figure 1, we saw that at the beginning the KPs were decreasing, and then they were increasing. We considered that the KPs in the first part are before the zero of the X axis and those in the second part are after this point. Moreover, instead of using

KPs, we use the actual distance from the point to zero as its X coordinate (see Fig. 4 under the column "X Pos (m)"). As a result, we have $-2049 \leq X \leq 76886$ (in meters).

Coherency-Checking and Correction Mechanisms

Before applying the correction mechanisms, one can observe that some values under the "Ground Speed" column (Fig. 4) are above the speed limit value in the traces data. This emphasizes the importance of the correction step in the proposed procedure.

4.3 Generating the Messages

Once the trajectory has been created, the sequence of messages must be created inside the communicating elements: the train, the RBC1 and the RBC2. The real-world data contained up to 1064 lines, this means the same amount of messages. In the trajectory description, we retained only 614 lines (Fig. 4) due to loss of precision in the milliseconds. However, in the message creation time, it is possible to obtain this precision in milliseconds. In this way, all the 1064 messages will be generated exactly at the same time than in real-world traces data, but according to the trajectory, if we could see the animation of the scenario, some of them would appear simultaneously.

5 Conclusion

We described how the real-world traces obtained from a train moving on ERTMS lines can be analyzed and used in order to generate a scenario that can be used in various simulation based evaluations of ERTMS components. The originality of this approach is the possibility of performing an evaluation of the ERTMS telecommunication subsystem using a network simulator such as OPNET, and in which the train will be moving exactly as it was guided by the ERTMS functional subsystem. In this way, various evaluations can be performed on the telecommunication subsystem without implementing the complete ETCS applications or without coupling the network simulator with a specific tool simulating the functional subsystem.

However, we have shown that dealing with real-world traces data is not always easy, and it is necessary to perform appropriate data transformation, coherency-checking and correction before using them in a simulator. The efficiency of these mechanisms is particularly important in a context where the scenario must be generated automatically from the real-world traces data. In our future work, we will improve the models of the GSM-R infrastructure and perform preliminary tests using the scenarios obtained using the traces.

Acknowledgments. The present research work is supported by:

- the VEGAS project (http://vegas.ifsttar.fr) funded by ANR (ANR-12-VPTT-0004) and labeled by i-Trans;
- the CISIT, the FEDER, and the Region Nord Pas de Calais.

The authors gratefully acknowledge the support of these Institutions.

References

1. UNISIG: System Requirement Specification, ERTMS specifications, subset 026 v2.3.0 (2010)
2. UNISIG: GSM-R Interfaces — Class 1 requirements, subset 093 v2.3.0 (2005)
3. Sniady, A., Soler, J.: An overview of GSM-R technology and its shortcomings. In: 12th International Conference on ITS Telecommunications (ITST), pp. 626–629 (2012)
4. Midya, S., Thottappillil, R.: An overview of electromagnetic compatibility challenges in European Rail Traffic Management System. Transportation Research Part C: Emerging Technologies 16(5), 515–534 (2008)
5. Beugin, J., Marais, J.: Simulation-based evaluation of dependability and safety properties of satellite technologies for railway localization. Transportation Research Part C: Emerging Technologies 22, 42–57 (2011)
6. Ruesche, S.F., Steuer, J., Jobmann, K.: Increase of Efficiency in Wireless Train Control Systems (ETCS level 2) by the Use of Actual Packet-Oriented Transmission Concept. In: IEEE/ASME JRC 2008, Delaware, USA (2008)
7. Aguado, M., Onandi, O., Jacob, E., Pinedo, C., Saiz, P., Higuero, M.: Wimax Role on CBTC Systems. In: ASME/IEEE JRCICE 2007, Pueblo, Colorado, USA (2007)
8. Aguado, M., Liedo Samper, I., Berbineau, M., Jacob, E.: 4G Communication Technologies for Train to Ground Communication Services: LTE versus WIMAX, a simulation study. In: 9th World Congress on Railway Research, Lille (2011)
9. Sondi, P., Kassab, M., Berbineau, M., Aguado, M., Lemaire, E.: Toward a Common Platform for Simulation-Based Evaluation of Both Functional and Telecommunication Subsystems of the ERTMS. In: JRC 2012, Philadelphia, USA (2012)
10. Mera, J.M., Gomez-Rey, I., Campos, A.: ERTMS/ETCS test simulation bench. Urban Transport and the Environment in the 21st Century, XIII, UK (2007)

Blind Digital Modulation Detector for MIMO Systems over High-Speed Railway Channels

Sofiane Kharbech[1], Iyad Dayoub[1], Eric Simon[2], and Marie Zwingelstein-Colin[1]

[1] IEMN/DOAE University of Valenciennes and Hainaut-Cambresis
59313 Valenciennes Cedex 9, France
[2] IEMN/TELICE University of Lille 1
59655 Villeneuve d'Ascq Cedex, France
{sofiane.kharbech,iyad.dayoub,
marie.zwingelstein-colin}@univ-valenciennes.fr,
eric.simon@univ-lille1.fr

Abstract. Nowadays, several wireless communication networks are widely deployed along railway lines, and future ones are already candidates for Long-Term Evolution (LTE) wireless data communications standard deployment. In fact, an efficient interoperability between these systems is crucial to increase safety, reduce maintenance costs and offer new services to passengers. Cognitive Radios (CRs) have been selected as promised alternative to answer all previous requirements. The main issue in CR is radio environment awareness, which is enhanced by modulation and waveform identification. In this paper we aim to shed light on the problem of blind digital modulation identification for Multiple-Input Multiple-Output (MIMO) technology used in high-speed railway communication systems which are associated to fast-fading channels. The intention is to distinguish among different M-ary Phase-Shift Keying (M-PSK), Amplitude-Shift Keying (M-ASK) and Quadrature Amplitude Modulation (M-QAM) modulation schemes without signal knowledge and Channel State Information (CSI). The detection process employs a Blind Source Separation (BSS) technique to sightlessly perceive the modulation type. The proposed detector proves a high M-PSK modulation identification ratio under a high velocity for LTE standard frequency band and bandwidth.

Keywords: Modulation identification, MIMO systems, Higher Order Statistics (HOS), High velocity, CR, Railway communications.

1 Introduction

MIMO is a commonly integrated technology into modern and promising wireless communication systems. The CR concept was recently introduced, it can be defined by radio awareness capabilities. Improving communications reliability and interoperability are believed to be of significant interest for high speed applications like railway. Waveform detection is one the features of CR interests.

Many modulation identification algorithms have been developed for Single-Input Single-Output (SISO) systems [1] [2] [3] [4] and non-faded MIMO systems

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 232–241, 2013.

[5]. In [5], it has been proposed an identification process through Artificial Neural Networks (ANN) for pattern recognition, where the pattern was the high order statistics features estimated from a blindly separated source signals. In this paper, we extend this work to verify the identification efficiency of the proposed process for M-ary PSK modulation schemes over fast-fading channels. The new proposal is able to take into account the variation of both the Signal-to-Noise Ratio (SNR) and the velocity. For this purpose, the ANN classifiers are updated according to the new transmission conditions.

The paper is organized as follows: section 2 defines the high-speed railway channel model and the assumed assumptions. Section 3 describes the detection algorithms while section 4 reports the performances of the detector for our system model. Finally, conclusion and perspectives of the research work are presented in section 5.

2 Channel Model

In this work, we consider a linear time-variant MIMO channel with N_t transmit antennas and N_r receive antennas. In general, to achieve the separation of source signals, some hypothesis are necessary:

H1. Source signals are independent and identically distributed (i.i.d.).
H2. The number of receive antennas is greater than or equal to the number of transmit antennas ($N_r \geq N_t$).
H3. The noise is additive and source signal independent.

In the context of this work, an additional assumption about channel fading is considered:

H4. The channel is frequency-flat fading and time-invariant during a MIMO symbol transmission.[1]

Under the above assuptions, the baseband k^{th} received MIMO symbol is done by:

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{n}(k) \tag{1}$$

where:

- k is the reception instant of a symbol vector, it also refers to a MIMO symbol.
- $\mathbf{x}(k) = [x_1(k), \cdots, x_{N_t}(k)]^T$ is the (N_t x 1) vector representing the transmitted source signals.
- $\mathbf{y}(k) = [y_1(k), \cdots, y_{N_r}(k)]^T$ is the (N_r x 1) received signal vector.
- $\mathbf{n}(k) = [n_1(k), \cdots, n_{N_r}(k)]^T$ is the (N_r x 1) vector corresponds to the additive zero-mean white circulary complex Gaussian noise with variance σ_n^2.
- $\mathbf{H}(k)$ is the (N_r x N_t) spatially-uncorrelated complex matrix of the MIMO channel at the instant k.

[1] Indeed, in MIMO-OFDMA (Orthogonal Frequency-Division Multiple Access) the frequency-flat fading assumption is valid per subcarrier only if time and frequency synchronization were achieved.

The MIMO channel matrix elements $h_{ij}(k)$ $(i \in \{1, \cdots, N_r\},\ j \in \{1, \cdots, N_t\})$ are time-varying for each MIMO symbol k. Each matrix element is Rayleigh distributed, and the normalised autocorrelation function of the fading channel at delay τ is expressed as [6]:

$$r(\tau) = J_0(2\pi f_D \tau) \tag{2}$$

where:

- J_0 is the zeroth-order Bessel function of the first kind.
- f_D is the maximum Doppler shift frequency given by $f_D = \dfrac{v}{\lambda}$, where v is the receiver velocity from the transmitter (or vice versa) and λ is the carrier wavelength.

In the frequency domain, $r(\tau)$ corresponds to a Jakes Doppler spectrum given by:

$$S(f) = \frac{1}{\pi f_D \sqrt{1 - \left(\frac{f}{f_D}\right)^2}} \tag{3}$$

This channel model is commonly used in the literature and has a few properties which make it especially appealing; first, it tends to be a good model for many fading multiple antenna channels, secondly, in many cases its mathematical simplicity enables analytical evaluation of systems where it is used to model the channel.

3 Identification Process

In this section we present the adopted process to detect the modulation type. As discribed in [5] and shown in Fig.1, the identification process has mainly four steps. We proceed – as first step – to blindly separate source signals, then to estimate some statistical features for each receive antenna. Based on the extracted features, we try to recognize the modulation type for every antenna through an ANN. As a final step, we fuse all the predicted decisions to obtain the final one.

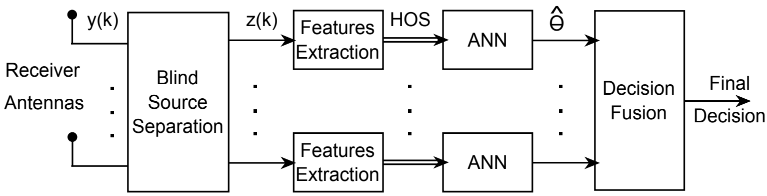


Fig. 1. Synopsis of identification process

3.1 Blind Source Separation

There are several BSS algorithms in the literature [7] [8]. Here, the Simplified Constant Modulus Algorithm (SCMA) is used to recover the transmitted symbols without prior CSI [7]. SCMA algorithm aims to determinate an N_r x N_t matrix $\mathbf{W}$ such as:

$$\begin{aligned} \mathbf{z}(k) = \hat{\mathbf{x}}(k) \stackrel{def}{=} SCMA(\mathbf{y}) &= \mathbf{W}^T\mathbf{y}(k) \\ &= \mathbf{W}^T\mathbf{H}(k)\mathbf{x}(k) + \tilde{\mathbf{n}}(k) \end{aligned} \tag{4}$$

where $\hat{\mathbf{x}}(k)$ is the estimate of $\mathbf{x}(k)$, $\tilde{\mathbf{n}}(k)$ is the filtred noise and $\mathbf{W}$ is the separator.

Note that the source separation can be achieved by optimizing a contrast function $\mathcal{J}$ [9], which introduces an arbitrary phase and permutation on the estimated symbols. SCMA simplifies the constant modulus critertion by employing a single dimension (real or imaginary part of the signal) [7]. To optimize $\mathcal{J}$, SCMA attempts to optimize the following criterion:

$$\begin{aligned} &\min_{\mathbf{W}} \quad \mathcal{J}_{SCMA}(\mathbf{W}) = \sum_{n=1}^{N_t} \mathbb{E}[(\Re(\mathbf{z}_n(k))^2 - R)^2] \\ &\text{Subject to : } \mathbf{W}^H\mathbf{W} = \mathbf{I}_{N_t} \end{aligned} \tag{5}$$

where $R = \dfrac{\mathbb{E}[\Re(\mathbf{x}(k))^4]}{\mathbb{E}[\Re(\mathbf{x}(k))^2]}$ is the dispersion constant.

BSS algorithms use-only may converge to a bound matrix $\mathbf{W}$, which means that the BSS algorithm may recover the same source several times. To overcome this drawback, we must force the BSS algorithm to obtain an uncorrelated sources. In this paper, we use Gram-Schmidt orthonormalization algorithm [10] as a post-processing to satisfy the constraint in equation (5). Since the orthonormalization of $\mathbf{W}$ assumes that the transmission channel is unitary, a whitening pre-processing is necessary and the separator will be applied on the whitened received signal $\underline{\mathbf{y}}(k) = \mathbf{B}\mathbf{y}(k)$, where $\mathbf{B}$ is the whitening matrix.

We implemented the SCMA using the Stochastic Gradient (SG) algorithm. The update equation is obtained by calculating the gradient of $\mathcal{J}_{SCMA}$:

$$\tilde{\mathbf{W}}_n(k) = \tilde{\mathbf{W}}_n(k-1) - \mu\mathbf{e}_n(k)\underline{\mathbf{y}}_n^*(k), \quad n \in \{1, \cdots, N_r\} \tag{6}$$

where μ is the SG step size and e_n is the error signal defined by $\mathbf{e}_n(k) = \Re(\mathbf{z}_n(k))(\Re(\mathbf{z}_n(k))^2 - R)$. Batch [10] is the used pre-whitening algorithm, which reduces the number of receiver sensors from N_r to N_t. Thus, since pre-whitening, N_r will be equivalent to N_t. The number of transmitter antennas and MIMO coding are supposed known at the receiver.

Given that the separation matrix resulting from the update-equation is not necessarily orthonormal, then the orthonormalization step is required at each iteration of the SG algorithm.

3.2 Features Extraction

For efficient detection, judicious choice of identification features is essential. HOS are among well-known and widely used ways for signal identification in SISO systems [1], [2]. HOS are also used for non-faded MIMO channel [5]. Higher Order Moments (HOM) of order a is expressed as [11]:

$$M_{ab}(\mathbf{x}) = \mathbb{E}[x^{a-b}(x^*)^b] \tag{7}$$

and Higher Order Cumulants (HOC) of order a of the non-zero mean signal x is done by:

$$C_{ab}(\mathbf{x}) = Cum[\underbrace{x, \cdots, x}_{a-b\ terms}, \underbrace{x^*, \cdots, x^*}_{b\ terms}] \tag{8}$$

where

$$Cum[x_1, \cdots, x_n] = \sum_{\psi} (-1)^{\alpha-1}(\alpha-1)! \prod_{\phi\in\psi} \mathbb{E}[\prod_{m\in\phi} x_m] \tag{9}$$

ψ runs through the list of all partitions of $\{1, \cdots, n\}$, ϕ runs through the list of all blocks of the partition ψ, and α is the number of elements in the partition ψ. The following table reports some theoretical values of sixth order HOS.

HOS are expected according to the estimated sources on each receiver sensor and they will be the ANN inputs for the next identification step.

Table 1. Some theoretical statistical moments and cumulants values for M-ary PSK modulation schemes [3]

	B-PSK	Q-PSK	8-PSK
M60	1	0	0
M61	1	-1	0
M63	1	1	1
C60	16	0	0
C61	16	-4	0
C62	16	0	0
C63	16	4	4

3.3 Artificial Neural Networks

The modulation detection problem can be considered as a pattern recognition problem. ANNs are among the best candidates for pattern recognition problems which many researchers have focused on to develop high performance modulation classifiers [3] [4] [5] [12]. In this paper, we make use of a multilayer feed-forward ANN. As mentioned in the previous subsection, one ANN is identically applied on each receiver sensor and provides one specific decision on the modulation scheme. It means that N_t cooperative decisions will be generated and then fused to generate the final single decision.

ANN training is a very important process which species the network efficiency level; a well-trained network is an efficient one. Training is the process that iteratively determines the gradient for weights and biases update function. Our purpose is to have a well generalized network, which has a good prediction for all samples set. For this, some processing of data set are highly recommended before launching ANN training and use. First, the extracted set of features are normalized to have zero mean and unit variance. Then, the best subset of the features set is selected through the Principal Component Analysis (PCA) technique. The used training algorithm that updates network's weight and bias values is the Resilient backpropagation algorithm (Rprop). This algorithm performs well on all the pattern recognition problems and is the fastest on them [13]. In addition, the used training style is batch. In batch training, the weights and biases are only updated after all the inputs are presented.

Notes that some other parameters can be adjusted when training the network in order to improve its generalization and to optimize its performance function. Further, using additional training data for better environment characterization is more likely to produce a network that generalizes well to new data.

In this paper, the ANN is trained to take into account the variation of both SNR and velocity.

3.4 Decision Fusion

After identifying the modulation type on each receiver sensor, the decisions on the modulation types obtained at previous step by each ANN have to be fused in order to form the final decision. This cooperative identification is called decision fusion. For that, Averaged Bayes Fusion Rule [14] is used in this paper. The idea is to estimate the probabilities P of having each modulation type from the modulation pool for each sensor, then the most probable modulation scheme is considered as final decision. Firstly, the distance d between the ANN output and each modulation scheme at each sensor is measured by:

$$d_{i,m} = ||\hat{\Theta}_i(m) - \Theta_i(m)||, \quad i \in \{1, \cdots, N_r\},\ m \in \{1, \cdots, N_M\} \tag{10}$$

where $\Theta = \{\Theta_1, \cdots, \Theta_{N_M}\}$ is the modulation pool. The equivalent probability is then determined by:

$$P_{i,m} = \frac{\frac{1}{d_{i,m}}}{\sum_{m=1}^{N_M} \frac{1}{d_{i,m}}} \tag{11}$$

Final detected modulation type Θ_f is given as:

$$\Theta_f = \max_{m=1,\cdots,N_M} \frac{1}{N_r} \sum_{i=1}^{N_r} P_{i,m} \tag{12}$$

4 Results and Performance

The proposed detector has been evaluated for various SNR values, multiple MIMO system configurations and wide velocity range for LTE standard frequency band and bandwidth. All results are based on 1000 Monte Carlo (MC) trials for each modulation scheme, SNR and normalized maximum Doppler shift frequency f_D/f_s where f_s is the sampling frequency. Each MC trial is based on 2048 MIMO symbols. The simulated modulation pools are $\Theta_1 =$ {B-PSK, Q-PSK, 8-PSK} and $\Theta_2 =$ {B-PSK, Q-PSK, 8-PSK, 4-ASK, 8-ASK, 16-QAM, 64-QAM}.

Note that the maximum simulated velocity is 540 kmph (which results in $f_D/f_s = 2.10^{-4}$) when we consider a carrier frequency of 2 Ghz and a bandwidth of 5 Mhz.

Figure 2 shows the detector efficiency where it's able to consider the variation of the SNR and the velocity at the same time, i.e. that is one ANN can be used to aware both the SNR and velocity variation. The probability of identification decreases when channel fading is faster and noise is more significant.

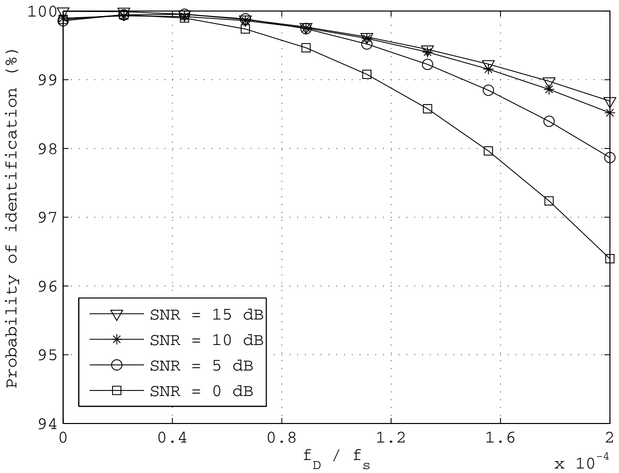


Fig. 2. Probability of identification from Θ_1 versus normalized maximum Doppler shift frequency in the case of 0-15 dB SNR range – MIMO antenna configuration is 1x4

Figure 3 shows that Q-PSK is the most ambiguous modulation scheme to identify in the same SNR and fading range; patterns of this modulation scheme are less discernible than the other ones. Indeed, up to 270 kmph ($f_D/f_s = 1.10^{-4}$) of velocity we can keep at least 99% of detection efficiency for all Θ_1 modulation types by 1x4 MIMO antenna configuration.

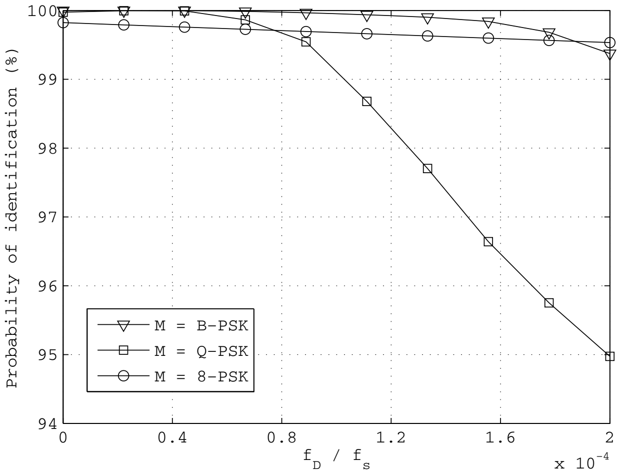


Fig. 3. Probability of identification from Θ_1 for each modulation scheme versus normalized maximum Doppler shift frequency in the case of 0-15 dB SNR range – MIMO antenna configuration is 1x4

The effect of MIMO antenna configuration on the detector efficiency is illustrated by Fig.4:

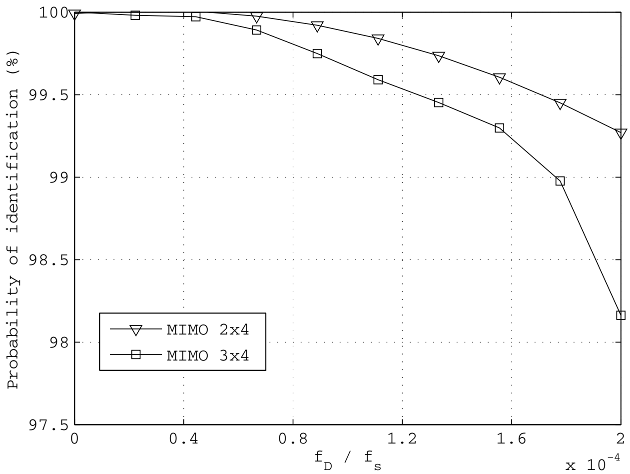


Fig. 4. Probability of identification from Θ_1 versus normalized maximum Doppler shift frequency for different MIMO antenna configurations in the case of SNR = 17 dB

In fact, a good source separation means a good recovered constellation which subsequently provides a discernible features for detection. Indeed, although the MIMO channel is non-faded, the detection efficiency decreases due to an ineffective BSS when the difference $\Delta = N_r - N_t$ increases.

As shown in Fig.5, when we try to detect from a larger modulation pool like $\Theta_2 = \{$B-PSK, Q-PSK, 8-PSK, 4-ASK, 8-ASK, 16-QAM, 64-QAM$\}$, detector performance degrades significantly especially at high velocity and even under a good transmission conditions.

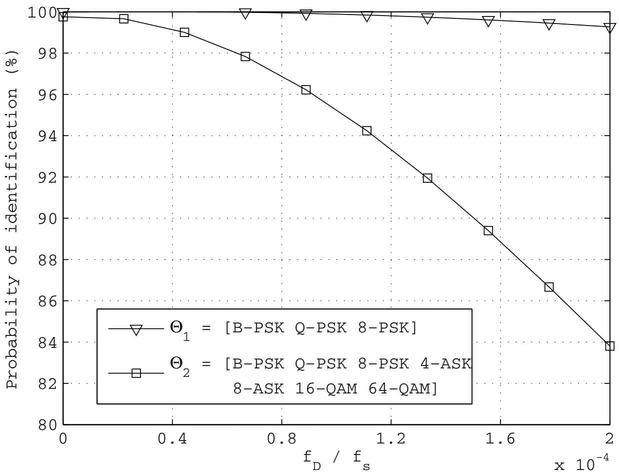


Fig. 5. Probability of identification from Θ_1 and Θ_2 versus normalized maximum Doppler shift frequency – MIMO antenna configuration is 1x4 and SNR = 25 dB

5 Conclusion and Perspectives

In this paper, we have focused on the enhancement of the modulation awareness of a CR mobile device for high-speed railways environment by identifying several features of the received signal, and recognizing the used modulation scheme. The proposed detector is based on HOS as features extraction subsystem and an ANN trained with Rprop learning algorithm as classifier subsystem. We proposed an efficient M-PSK modulation identification under different MIMO system configurations over noisy fast-fading channels. In addition, the proposed detector is able to be aware of channel variations (e.g. SNR and velocity).

For a typical SNR in high-speed railways environment and a velocity up to 450 kmph ($f_D/f_s = 1,6.10^{-4}$), the probability of false detection from a M-PSK pool is less than 1%, which is quite efficient. This performance is unlike that obtained from the modulation pool that also include M-ASK and M-QAM modulation schemes, even at a high SNR. Thus, our future work will focus on the development of a new detector able to identify from a larger modulation pool than the M-PSK one.

Acknowledgments. This work is funded by "CORRIDOR" French-ANR project. The authors would like to thank K. Hassan from IEMN-UVHC for fruitful discussions.

References

1. Wu, H.-C., Saquib, M., Yun, Z.: Novel automatic modulation classification using cumulant features for communications via multipath channels. IEEE Trans. Wireless Commun. 7(8), 3098–3105 (2008)
2. Swami, A., Sadler, B.: Hierarchical digital modulation classification using cumulants. IEEE Trans. Commun. 48(3), 416–429 (2000)
3. Dobre, O.A., Abdi, A., Bar-Ness, Y., Su, W.: Survey of automatic modulation classification techniques: classical approaches and new trends. IET Commun. 1, 137–156 (2007)
4. Wong, M.L.D., Nandi, A.K.: Automatic digital modulation recognition using artificial neural network and genetic algorithm. Signal Process. 84(2), 351–365 (2004)
5. Hassan, K., Dayoub, I., Hamouda, W., Nzéza, C., Berbineau, M.: Blind Digital Modulation Identification for Spatially Correlated MIMO Systems. IEEE Trans. on Wireless Commun. 91(2), 683–693 (2012)
6. Clarke, R.H.: A Statistical Theory of Mobile-Radio Reception. Bell System Technical Journal 47(6), 957–1000 (1968)
7. Ikhlef, A., Guennec, D.L.: A simplified constant modulus algorithm for blind recovery of MIMO QAM and PSK signals: a criterion with convergence analysis. EURASIP J. Wireless Commun. and Networking, Article ID 90401 (2007)
8. Papadias, C.: Globally convergent blind source separation based on a multiuser Kurtosis maximization criterion. IEEE Trans. Signal Process. 48(12), 3508–3519 (2000)
9. Krob, M., Benidir, M.: Fonction de contraste pour l'identification aveugle d'un modèle linéaire quadratique. In: Fourteenth Symposium GRETSI-JUAN-LES-PINS (September 1993)
10. Ikhlef, A.: Séparation aveugle de sources dans les systèmes de communication MIMO, Ph.D thesis, University of Rennes I – France (September 2008)
11. McCullagh, P.: Tensor Methods in Statistics. Chapman Hall (1987)
12. Hassan, K., Dayoub, I., Hamouda, W., Berbineau, M.: Automatic modulation recognition using wavelet transform and neural networks in wireless systems. EURASIP J. on Advances in Signal Process., Article ID 532898, 13 pages (June 2010)
13. Beale, M.H., Hagan, M.T., Demuth, H.B.: Neural Network Toolbox User' s Guide R2012b. MATLAB (2012)
14. Hassan, K.: Contributions to Cognitive Radio Awareness for High Mobility Applications, Ph.D thesis, University of Valenciennes and Hainaut-Cambresis – France (December 2012)

Enhancing the CATS Framework by Providing Asynchronous Deployment for Mobile Application

Mikael Desertot[1], Christophe Gransart[2], and Sylvain Lecomte[1,2]

[1] UVHC, LAMIH CNRS, UMR 8201, F-59313 Valenciennes
University Lille North of France, F-59000 Lille, France
[2] IFSTTAR LEOST
20 rue Elisée Reclus BP 70317 F-59666 Villeneuve d'Ascq cedex, France
{mikael.desertot,slecomte}@univ-valenciennes.fr,
christophe.gransart@ifsttar.fr

Abstract. The work presented in this paper focuses on the use of the CATS framework (a framework dedicated to the adaptation and the deployment of context aware services in the transportation context) on applications designed to assist the user in transportation activities (e.g., driving assistant, visiting a city, finding a parking place etc.). We present a solution to deploy and use these services when users do not have a connection to WAN by using asynchronous solutions based on smart cities infrastructures. A prototype has been developed and evaluated at the end of this paper.

1 Introduction

Nowadays there are many applications that users can benefit from on their handheld devices. Such capabilities are reaching our vehicles, offering drivers help for driving safely and more efficiently, thanks to the numerous services provided by applications on their devices. In this article, we focus on the transportation domain. Contrary to other application fields like finance or commerce, mobile technologies did not have a big impact on transportation until lately. The only successful application of these technologies that we can mention in the transportation domain concerns navigation devices. Nevertheless, the situation has now started to change. New applications like Waze (http://world.waze.com) have appeared on the market. Such applications rely on a community of drivers who share information about traffic congestions or accidents. For example, Waze centralizes data through the telephony network (3G), while other approaches require direct interaction between mobile nodes.

To simplify application assembly and reactivity according to transportation constraints (lack of communication infrastructure, high mobility...), several researches propose frameworks that hosts multiple applications at once, offering at the same time management functions for context-awareness and the deployment of service when the user or the system need it.

M. Berbineau et al. (Eds.): Nets4Cars/Nets4Trains 2013, LNCS 7865, pp. 242–252, 2013.

When considering the possible applications that can be used on mobile devices, we can see that many aspects have to be considered. For instance, some applications must use the telephony network to transfer data and can not work otherwise. Some of them, Locations Based Services, also need a GPS module, which is now embedded in most smartphones. Other applications are meant to be used in an ad-hoc mode, providing services to a small community of users who are at the same place at the same time. For the highly mobile users, the ad- hoc applications can bring many advantages, such as the ability to have a fast access to relevant information (e.g., traffic accidents, intervention vehicles, traffic jams). Another important aspect is the independence to the infrastructure, which suggests the interest of having services for "local" ad-hoc networks. The research for developing reliable ad-hoc networks for highly mobile users is ongoing. We mention [1] who propose a Context Management System using the NEMO (NEtworkMObility) protocol, which addresses the issue of a mobile network also capable of attaching to the Internet.

Among the possible mobile applications for handheld devices, our work focuses on those designed for the transportation domain. By transportation we mean the movement of people from one place to another. To get to her/his destination, a person can either walk, drive or take public transportation modes (or any combination of these). Common features are mostly related to the trip - the environment and the neighboring devices change as the person advances. Most of the transportation applications need positioning information at all time (preferably GPS coordinates). Communication is also an important element, allowing applications to contact either an infrastructure or surrounding devices on the ad-hoc network. Moreover, the type of environment affects the behavior of the applications (e.g., use in indoor/outdoor environments, in urban areas or on highways). The differences between applications (or between the different behaviors of the same application) are related to the considered mean of transportation. A pedestrian travels at low speed and will be able to look often at his/her device and make some adjustments if necessary. Passengers can handle their devices, but move at a higher speed, which can sometimes lead to connection failures. Drivers face the same connection problem but also need to be notified with traffic information and important events, without being disturbed since they are driving.

In our previous work we have introduced CATS (Context Aware Transportation Services), our framework for context-aware applications for the transportation domain. The goal of this framework is to provide an execution environment for service-based applications as well as management functionalities for the deployment and the adaptation to context changes. Some preliminary evaluations of CATS have been presented in [2], showing that the framework is light enough for mobile devices such as smartphones. With CATS, using the ad-hoc network we can detect services that offer a functionality specific to the area we are in, and we can benefit from them by installing them on our device. In [3] we evaluated the time necessary for service download in different situations and find results coherent with our need: services can be downloaded fast enough from one-hop neighbors.

In the same time, we are used to be connected to the network every time we need it. However, using data services can be impossible in some ways : when we travel in another country, or when we are passengers in a train (where using data network is difficult, due to the number of passengers arriving at one time in a cell). This paper proposes to enrich the CATS framework by allowing to invoke services in an asynchronous way.

The first section shortly describes the CATS Framework. The second section describes the need of asynchronous deployment in the particular nature of the transportation context we are considering and the challenges it poses, a scenario of use and the state of the art. The third section exposes a solution available in distributed computing to take into account asynchronous invocation and deployment. Next, we present an evolution of our platform, detailing our approach for the integration of asynchronous actions and the architecture. Before concluding, we present a use case, experimentation, and evaluation for our framework.

2 The CATS Framework

In this section we describe shortly the CATS Framework, which has been presented in our previous work [2], and finally the need to extend this framework to allow asynchronous invocation.

2.1 The Framework

To provide adaptation to the changes that occur during the execution of an application, there are two things necessary: the services must specify which behavior is specific to which context, and the context situations must be identified. This is where our framework comes in, by providing framework-specific services to help management of the applications by monitoring the context. We consider two kinds of adaptation. First, services can determine relationships between some parameters and values of certain context elements. The second way of adapting an application to context is to change an entire service rather than having a service with much code, trying to cover several situations. Context information is used: during execution, when one or more context elements change enough to justify a change in the behavior of an application; at the download of a new service or application, to ensure compatibility with the device.

Applications are composed of multiple functionalities, which can be divided in independent modules. First, we can identify the functionalities that are common to most applications, like for example positioning. It is always more interesting to have a single piece of code handling the localization, rather than having a lot of applications implementing similar code. Second, each part of an application providing a certain functionality can be implemented in more than one way. So, for each computation we can choose the most appropriate way to get the result.

Our framework (Figure 1) provides an unified environment with non-functional services that assure the continuous and context-adapted execution of the applications. On one hand, the CATS Framework allows the execution of three specific

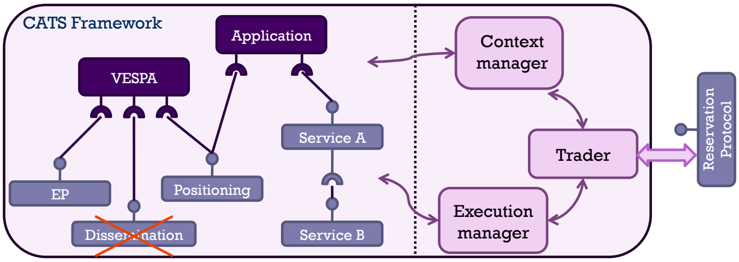


Fig. 1. The CATS Framework architecture

services (Context Manager, Execution Manager and Trader) which function continuously as long as the framework is started. On the other hand, the framework is the environment where the applications are executed.

Each specific service has a role identified as follow:

- The trader. The application framework must be able to acquire new services when these are necessary. The trader sends a message to the ad hoc network to demand the needed service. Neighboring devices able to provide the service answer to the requesting device, which then chooses from whom to download. The service trader handles both outgoing and incoming requests. If infrastructure access is available, it will be preferred, as a centralized registry is more likely to have a large number of services.
- The execution manager. The execution of the services in the framework must be monitored in order to detect failure. When a service stops working, the execution manager must react and take the necessary actions to correct the situation. The simple solution is to search for an equivalent service that is already on the device and start it. If there is no equivalent present locally, the execution manager must call the trader to launch a search for the missing service on the neighboring devices. Another type of problem occurs when services do not function correctly because of an external cause. A simple example is the loss of GPS signal; the service is still running, but unable to provide correct information. The execution manager stops the service so that another one can be bound.
- The context manager. Based on the classification we have for the context, the CATS framework service called context manager takes "snapshots" of the state of the environment and represents them in an XML file. An XML Schema is used to validate the XML representation. The context manager can evaluate the state of the context on demand, but also on a continuous basis, when certain elements need to be monitored.

We use Java and OSGi (a Java dynamic service-oriented framework) for the development, deployment and installation of services (and applications, which are built out of services). For the three parts of our framework (Context Manager, Trader and Execution Manager) we have built one or more OSGi bundles. The user applications are also built out of services exposed by the components and they are executed in the same environment as the management services. To help managing service binding and automate as much as possible dynamic and adaptive bindings, we use the iPOJO [4] service-component model (hosted by Apache) that is deployed on top of OSGi.

2.2 Needs of Asynchronous Invocation

Our CATS framework is dedicated to transportation services, because we wish to provide solutions that take into consideration the specific constraints of this domain. Our solution is based on a framework for applications with similar characteristics, designed to accompany users throughout their evolutions in different environments. To respect the constraints of the transportation domain, our solution must react sufficiently fast to the changes that occur. Moreover, services must be light enough to be easily downloaded and installed "on the fly". A local service is not subject to network connection loss.

However, in specific environments, services are still too heavy to be downloaded “on the fly” (services exchange between cars on a highway, or communication between train and car). More over, it could be impossible to use 3G connection (when the user is using roaming). To overtake this drawback we propose in this paper to enrich the CATS framework to allow asynchronous invocations to services.

More over, in a near future, smart cities will propose, through a lot of equipment, connections to services by using an ad-hoc link. Indeed, cities are increasingly integrating communication capabilities, with higher availability an quality, as well as social infrastructure. It is against this background that a modern way of interacting with the city environment has emerged, highlighting the importance of Information and Communication Technologies (ICTs). In this context, smart cities will rely on wireless networks to manage the environment, and people in the city. By multiplying communication node disseminated in the environment, users will benefit from more regular access to the network. From this fact, a user will be capable of connecting frequently and take advantage of this functionality. These connections could act as a mailbox to get the results of asynchronous invocations [5,6,7].

3 Using DDS in the CATS Framework

We presented above the CATS architecture, showing its main components: the context and the execution manager as well as the trader. This last has been first designed to only deal with synchronous requests through a classical or ad-hoc network towards an internet component farm or a user's neighbor. But because of the reasons listed in the previous sections the trader has evolved to a novel architecture to tackle asynchronous issues.

3.1 The DDS Framework

The DDS (Data Distribution Service) is a specification of a publish/subscribe middleware taking into account QoS properties[8]. The first specification has been defined by the OMG during 2004. DDS can be used independently from CORBA. This new service is suitable for a new class of applications that require real-time Quality of service. It provides an efficient publish/subscribe system. Information consumers subscribe to information of interest. Information producers publish information. DDS matches and routes relevant information to interested subscribers. DDS furnishes QoS suitable for real-time systems: deadline, levels of reliability, latency, re-source usage, time-based filter. Different programming styles are allowed based on listener or wait-based data access.

3.2 A New Architecture for Asynchronous Deployment

As in CATS framework, every pieces of software are modular and are build together thanks to a service architecture, introducing another messaging protocol only requires the addition of a component wrapping the communication with the DDS server. Thus, the trader can switch from one protocol to another according to the context elements. For now, context information like destination, network availability or battery level are used to decide the proper communication mode. In order to bring context reactivity inside the trader, CATS capabilities have permit to smoothly add the required notifications. Indeed, by just adding an XML description of the context elements to monitor, CATS injects inside the component the appropriate code. The following scenario illustrates CATS behavior when facing asynchronous communications needs.

3.3 Scenario for Context-Dependent Asynchronous Requests

On a standard use, our mobile application adapts itself to the environment requesting components or data in a synchronous way. While a context notification about a travel of the user is raised, requests related to its future location are switched to an asynchronous mode (DDS mode) by the trader. In this way, we first minimize the cost of the communication with our neighbors, unlikely to be able to have what we want. Secondly we can make sure that the reception of the response (a component service or some data) will be done at the right time and/or the right place. For instance the most obvious situation is when the trader is notified that the user is closed from its destination, this can trigger a check of the messages posted for him (a service could have been personalized in a specific way for this user). According to battery level, data or software pieces can be downloaded but may be kept aside if the level is low. Finally if the user go through an area with free and high rate network communication capabilities, subscription to this event by the trader could initiate an anticipated download. CATS framework takes in charge those notifications as soon as thresholds and context elements are described with XML meta-data. Afterwards the trader will manage components or data retrieving, and delegate their instantiation/installation to CATS' execution manager.

4 Prototype and Evaluation

This section presents experiments and results carried out in the context of this work. It relies, on the client side, on an application built thanks to the CATS framework presented above and implemented in previous works. The goal is to illustrate the feasibility of our proposition by using industrial standard frameworks in a prototypal architecture close to a real environment.

4.1 Prototype Description

Our environments is composed of three layers : The mobile client side (i.e. a smartphone) which will send an asynchronous request for obtaining a software service/component or any kind of data. It is capable to dynamically install the software grain it will received. The server side, responsible for interfacing both clients and external component repositories with the DDS environment. Component repositories, offering the capability to potentially download any kind of known component. We use OpenSplice, the PrismTech's DDS implementation [9] as a keystone. It offers the reliable asynchronous framework we require to achieve our goal. But if it could have been smoothly included inside any smartphone application, the architecture would have been lighter as business component managing asynchronism capabilities would have been embedded inside smartphones. But it is not currently available for any smartphone OS like Android or iOS, mainly because its architecture is not designed for now for such constraint environments. It is the reason why we introduce some servers interfacing our mobile client with the DDS message bus. Those servers have in charge collecting requests from clients and delivering messages when a client is connected and is looking forward to unstack its messages.

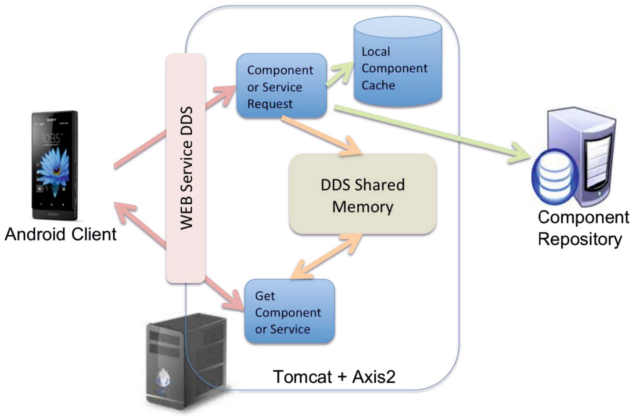


Fig. 2. Overview of the architecture

Connection and communication between clients and servers are performed using a Web Services architecture. Clients build and send a SOAP compliant message (those devices always offers on-the-shelf API to easily generate such messages) to the web services hosted by servers. In our prototype, two web services interfaces are available : one offering the possibility for a client to request a component, service or data and another one allowing to consult message box.

When a query is received by the server through a Web Service, four eventualities can occur :

– The component/service/information is available locally (managing a cache minimizes the access time to a component). The response is packaged and posted to the client via DDS.
– Nothing is available locally. Query is sent to remote component or data repositories. Response is cached and posted to the client via DDS.
– The component has to be adapted. Depending of the kind of customization, such component can be cached or not. Afterward the message is posted to DDS.
– If the request is not understandable or if there isn't any component available a post referencing the issue is made. The interest is twofold, ensuring a response to the client and allowing him to eventually react to the problem.

To optimize the bandwidth, the client can specify a component name to prioritize a particular message containing the information he needs as fast as possible. Here the CATS framework on client side managing context elements is very helpful to determine what is the preferred behavior. A client can then dynamically install the downloaded component, or process the received data. If an error is raised, the request can be re-sent, identicilly or modified if needed. The behavior is in the smartphone application hands at this point.

In the architecture presented in figure 2, only one server is shown but obviously we imagine a constellation of such servers, transparently exchanging information and messages in a reliable way thanks to DDS.

4.2 Experimentation

Lets first describe devices used at server side. To implement this architecture, we used a Mac Book Pro with a dual-core processor and 8 Gb of memory to host the server. Server is run using Tomcat[1] coupled with AXIS 2. The community edition of OpenSplice offers the DDS instance. Its runs on a Linux ubuntu operating system on the Mac Book Pro. On client side we use different kind of mobile devices, running Android in different versions (2.3, 3.1 and 4.0.4). Multiplying the number of devices shows that performances depends for a certain part on the device used but also show that in most of the cases, any device is capable of offering enough power to make the experiment possible for a user.

From this architecture, we evaluate different measures to illustrate request and message un-stacking namely A : The time spent by client to send a request.

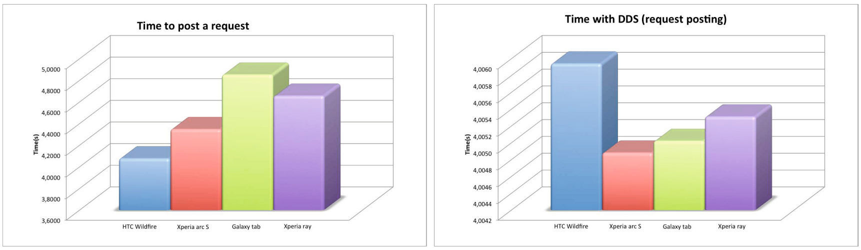
[1] http://tomcat.apache.org/

It includes connecting the client to the web service and receiving an acquittal from the server. B : The time, starting from the client request, for the server to put down the message to the DDS queue if the response is cached locally. C : Same as B but response downloaded from a remote component repository. D : The time spent by a client to connect, authenticate and retrieve a component posted for him

4.3 Results

Figures 3 and 4 show the results obtained with our prototype. They consist in an average based on a hundred values, removing the higher and the lower one.

Figure 3 (A) shows the time a client spend connecting the web service and obtaining an acquittal. This time varies from 4 to 5 seconds depending on the device used. The difference is due to the power of the devices, their ease to manipulate XML and the version of Android used. Here the best result is obtained by the HTC. Independently from the characteristics it is mostly because the HTC overlay is light and the Android version used is also less powerful than newer ones (less services running).

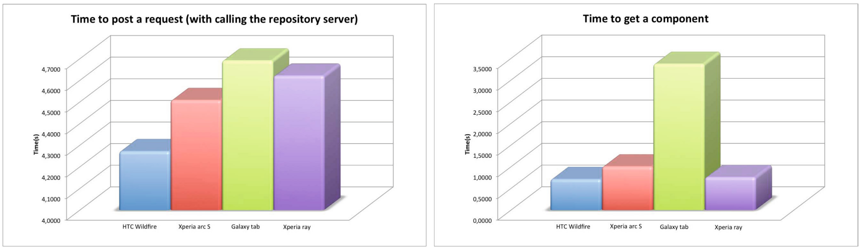


(A) (B)

Fig. 3. A : Sending a request and B : Package a local message via DDS

Figure 3 (B) gives the time spent from a request sent from a client to the moment the response message is effectively posted to DDS. Each of the devices spent the same time to achieve this experiment, showing that the difference obtained in (A) with different devices is mostly due to processing the acquittal. Results obtain in (B) are increased in (C) if the needed component has to be downloaded from an external server. There, the time depends on the device used but in each case, time spent to process a request is still acceptable, knowing that the client is supposed to potentially look for a response later.

Finally, in figure 4 (D), the time spent by a client to authenticate and download a component posted for him is described. The difference between devices is mostly due to the Android environment. 3.1 version manages data in a specific way for os 3.XX. Whereas for Android version 2, the older version includes less services which make XML processing and data downloading longer. This time include the installation time of the component, i.e. the installation and



(C) (D)

Fig. 4. C : Package a distant message via DDS and D : Get a message posted

activation of the component on the Android device until it is available to other components.

5 Conclusion

In this paper we showed the interest of bringing a context-aware framework for transportation services together with asynchronous communication capabilities. Indeed, network connection loss, cost or speed could become an important limitation for mobile user, especially as travel distances increase and public transports are widely used.

Mobile applications are built from different services and the assembly depends on the context. The main issues in this domain are both connecting services and obtaining data. If services could be accessed remotely (like web services), they can also be downloaded locally on the mobile device. And it goes even with data. By relying on asynchronous communication we are able to post request for both data or service downloading. Afterwards, the receiving of components or information could be performed at the appropriate time.

We illustrate the architecture we are proposing with a prototype and results showing feasibility as well as acceptable performances for asynchronous context-aware interaction. Different experiments has been made over a concrete platform to show that using asynchronous request with DDS is not an important additional cost compared to synchronous interaction.

However, the architecture can be still improved. Here we are using a number of context elements identified as the most representative for the mobile domain. These context elements could still be increased in order to refine dynamic adaptation opportunities and be able to switch from asynchronous to synchronous (and vice versa) at the most suitable time. Moreover, due to the multiplicity of communications capabilities, it is still necessary to handle the case where asynchronous request is send but the need is satisfied before the reception of the answer. This could occur when a neighbor or a local access point has been reachable and had the data or services requested at disposal.

References

1. Gaultier, B., Ben Rayana, R., Bonnin, J.-M.: Context management systems applied to mobility. In: Berbineau, M., Itami, M., Wen, G. (eds.) 9th International Conference on Intelligent Transport Systems Telecommunications, ITST 2009, pp. 330–335. IEEE Computer Society, Piscataway (2009)
2. Popovici, D., Desertot, M., Lecomte, S., Peon, N.: Context-aware transportation services (cats) framework for mobile environments. International Journal of Next-Generation Computing 2 (2011)
3. Popovici, D., Desertot, M., Lecomte, S., Delot, T.: When the context changes, so does my transportation application: Vespa. Procedia Computer Science 5, 401–408 (2011); The 2nd International Conference on Ambient Systems, Networks and Technologies (ANT 2011) / The 8th International Conference on Mobile Web Information Systems (MobiWIS 2011), doi:10.1016/j.procs.2011.07.052
4. Escoffier, C., Hall, R.S., Lalanda, P.: Ipojo an extensible service-oriented component framework. In: IEEE International Conference on Service Computing (SCC 2007), Salt Lake City, USA, pp. 474–481 (2007)
5. Bodhani, A.: Smart transport. Iet Journals and Magazines 7(6), 70–73 (2012)
6. Xiong, Z., Sheng, H., Rong, W., Cooper, D.: Intelligent transportation systems for smart cities: a progress review. Science China Information Sciences 55, 2908–2914 (2012), http://dx.doi.org/10.1007/s11432-012-4725-1, doi:10.1007/s11432-012-4725-1
7. Anthopoulos, L., Fitsilis, P.: Considering future internet on the basis of smart urban cities a client-city architecture for viable smart cities. In: The Fourth International Conference on Evolving Internet, INTERNET 2012 (2012)
8. OMG, Data distribution service v1.0, Tech. rep., OMG, document formal/04-12-02 (2004)
9. Prismtech, Opensplice dds v5 reference guide, Tech. rep., Prismtech (2011), http://www.prismtech.com/opensplice

Author Index