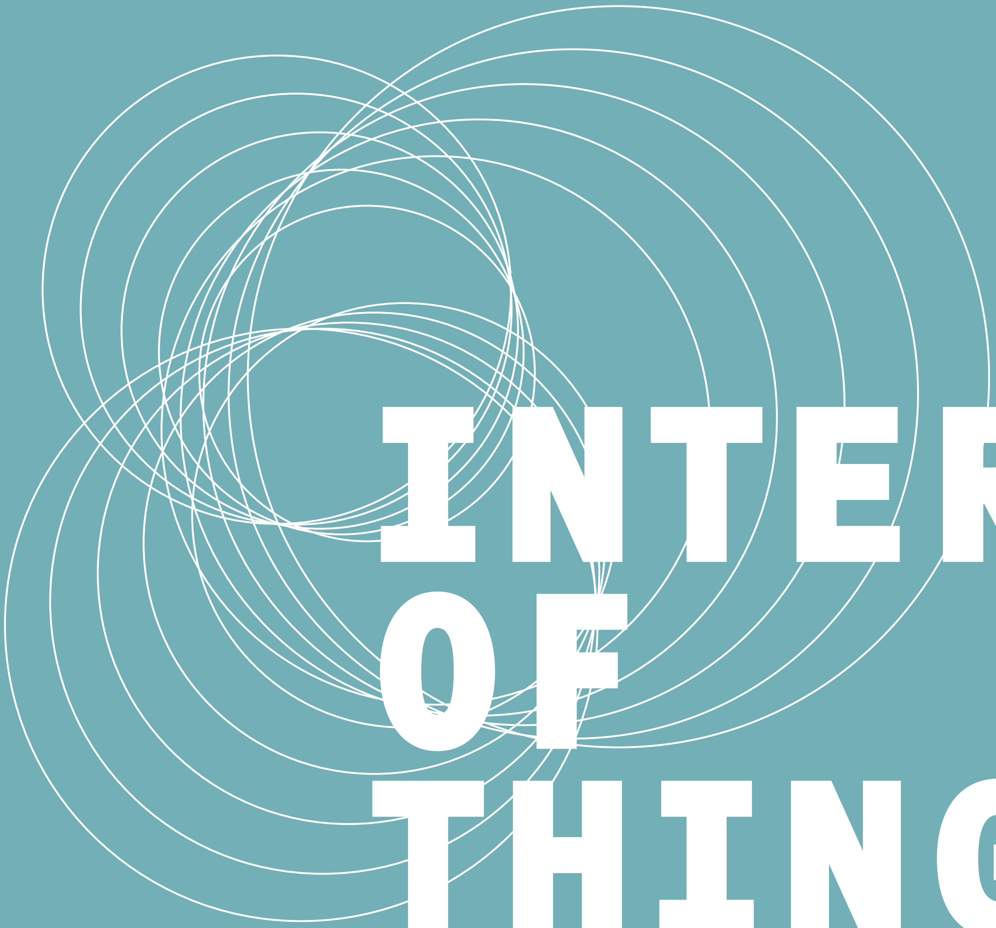
Florian Michahelles (Ed.)

Adjunct Proceedings

First International Conference on The Internet of Things



INTERNET OF THINGS 08

ORGANIZED BY >

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

University of St.Gallen

Massachusetts Institute of Technology

Table of Contents
First International Conference on The Internet of Things
IOT 2008 WORKSHOPS

PROSIT 2008
Producing Standards for the Internet of Things

IOV 2008

1st International Workshop on Interoperable Vehicles

Social-IoT 2008

Designing the Internet of Things for Workplace Realities: Social and Cultural Aspects in Design and Organisation

IOT 2008 Demos

Message from the Workshops and Demo Chair

It is my great pleasure to present the first Internet of Things Conference (IoT 2008) Workshops and Demos. The Workshops associated with IoT 2008 feature a specific range of topics beyond the main conference tracks to better cover the research areas of standards, interoperable vehicles, and the social and cultural aspects in the exciting field of an emerging internet reaching out into the real world. Accordingly, we are proud to play host of the following three workshops:

- Workshop A: Producing Standards for the Internet of Things (PROSIT)
- Workshop B: 1st International Workshop on Interoperable Vehicles (IOV 2008)
- Workshop C: Designing the Internet of Things for Workplace Realities (Social-IoT)

While the main conference features keynotes from industrial and academic visionaries, technical presentations of cutting-edge research, reports on the user-experience from seasoned practitioners, the purpose of the workshops is to provide a less formal and more open environment for the free exchange of views where novel and stimulating ideas and experiences among researchers both from industry and academia are shared.

Our major criterion for selecting this three workshops (out of 15 workshop proposal submissions!) was how they could provide and increase means for discussions among the participants, e.g., by panels, break out or specific discussion sessions .

All IoT attendees are welcome to register for any workshop. Although this imposes a challenge on workshop organizers with respect to the integration of an unknown number of participants, we expect to create an interactive atmosphere for lively discussions and exchange of knowledge and experience.

In addition we proudly host ten demonstrations - being shown during the conference breaks - that provide a first insight into what applications of a future internet of things may look like. The topics of the demos are rather diverse ranging from human to internet of things interactions to RFID data visualization and moving sensor networks placed on robot cars. We are looking forward to these exciting demonstrations inspiring our minds and illustrating late-breaking research concepts during the conference breaks.

I would like to thank all the workshop organizers, the involved program committee members, the general and program committee chairs, reviewers and all the authors for their contribution to an exciting and high quality program! I am indebted to Marc Langheinrich, program co-chair, for creating the cover artwork and for helping to put the program together. I also want to thank Steve Hinske, organizational chair, for mastering all the challenges concerning space allocation, registration of the participations, and the special needs of the demo exhibition!

Welcome to IOT 2008, enjoy the lively discussions, and get inspired!

Florian Michahelles, ETH Zurich
Workshop and Demo Chair

Workshop ‘Producing Standards for the Internet of Things’ (PROSIT 2008)

Message from the Workshop Co-Organisers

Many of the virtually innumerable (intelligent?) nodes that will eventually form the Internet of Things will be found inside cars, at home, and in the shopping mall. Typical applications will include, for example, intelligent homes, car safety, item tracking, and probably a whole lot of others we can hardly imagine today. Many of these applications will become part of our daily lives, and are prone to collect information that would be considered as private by many. For the average user, it will be next to impossible to establish who has access to these information, and for which purposes.

This unprecedented penetration of virtually everyone’s life suggests the need for a close scrutiny of the various processes to be associated with the development of such a technology and its subsequent wide deployment. International standardisation is among the most important of these processes. It is linked to both the technological development and the policy and legal frameworks within which the technology is to be developed and deployed.

To the best of our knowledge, this is the first ever workshop to address these non-technical topics that (will) surround the Internet of Things. Unfortunately, being (at least among) the first also has its drawbacks. In our case, the number of submissions was rather low (though luckily of good quality). We concluded that we are probably ahead of our time, and that the window of opportunity isn’t fully open yet. But we are certain that the focus will (have to) move away from the technical nuts and bolts to these non-technical aspects the Prosit-Workshop will discuss.

The Workshop will start with a ‘traditional’ paper session, to set the scene. Session number two will be an open discussion, guided by a number of questions. We hope that this discussion will help identify the problems and issues the Internet of Things will face once it has actually materialised. And while we do believe that the IoT will eventually become reality (in whichever exact shape and colour), we also have to note that – like many other technologies – the IoT so far has been just around the corner for quite a while now ……. Perhaps the Prosit-Workshop will provide some explanations here as well.

Kai Jakobs, RWTH Aachen University, Germany
Robin Williams, The University of Edinburgh, Scotland

PROSIT 2008 Workshop Organization

Workshop Co-organizers
Kai Jakobs, RWTH Aachen University, Germany
Robin Williams, The University of Edinburgh, Scotland

Technical Program Committee
Knut Blind, FhG ISI & TU Berlin, Germany
Yves Chauvel, ETSI, France
Tineke Egyedi, TU Delft, The Netherlands
Vladislav Fomin, Vytautas Magnus University, Lithuania
Ole Hanseth, University of Oslo, Norway
Eric Iversen, NIFU STEP, Norway
Ken Krechmer, ICSR, USA
Kalle Lyytinen, Case Western Reserve University, USA
Roy Rada, University of Maryland, USA
Kai Reimers, RWTH Aachen, Germany
Mostafa H. Sherif , AT&T, USA
Henk de Vries, ERASMUS University, The Netherlands
Willem Wakker, ACE Cons., The Netherlands
Marc van Wegberg, University of Maastricht, The Netherlands

Near Field Communication: State of Standardization

Gerald Madlmayr, Jürgen Ecker, Josef Langer
University of Applied Sciences of Upper Austria, Hagenberg, NFC Research Lab
{gmadlmay, ecker, jlanger}@fh-hagenberg.at

Josef Scharinger
Johannes Kepler University, Department of Computational Perception
josef.scharinger@jku.at

Abstract

Near Field Communication (NFC) is an RF-based proximity coupling technology allowing transaction within a range of up to 10 cm. Whereas the physical and the data link layer of NFC are already standardized by ISO, ECMA and ETSI, it's the NFC-Forum's task to get a common agreement on data protocols, security measures and layers above. The NFC-Forum is a non-profit industrial consortia cooperating with other standardization and special interest groups like the GSMA, the JavaCard Form, GlobalPlatform as well as industry partners from transport and payment businesses and many more. With NFC, a key technology is on its way into the consumer's most personal device. Interoperability is one of the most important goal to be achieved prior to roll out devices and services, in order to satisfied the consumer's expectations. In order to realized a vision where goods and device can interact requires the industry players to adopt and stick to a standard. However, relying on a standard for exchanging bits is not the overall solution. Interoperability must be granted also on all layers. Additionally, already during standardization security must be considered closely as well. This paper discuses the actors' goals and strategies during the process of NFC's standardization, which parties investigate on with part of the technology and fields where standardization would be required but has not yet begun. The reader gets a sound understanding on NFC technology and its holistic impact on applications and services in the Internet of Things.

1 Introduction

RFID (Radio Frequency Identification) is used by many daily applications. For the consumer unnoticed and simple to use, they offer a popular alternative to conventional communication channels. Starting with simple access control possibilities up to complex data memories, quite different applications can be realized. A further development represents NFC [13], a technology for the fast and uncomplicated exchange of small amounts of data. It opens new perspectives regarding the application development on mobile phones and other devices. Meanwhile NFC found introduction to mobile phones, one of the most common means of communication.

Whereas the dream of extensive item level tagging is still not fully implemented, ABI Research [1] is convinced that by 2010 350 million (equal to 20 % of all mobile phones) of mobile phones can be used as an RFID smartcard, an RFID reader/writer or a device which is able to establish an RF based peer-to-peer connection with other NFC devices. While logistics and health care make use of long-range RFID technology, NFC relies on the smartcard standard ISO 14443 [7] allowing wireless transactions only over a distance of up to 10 centimeters. This is part of the *Touch and Go* philosophy giving the user a new dimension of usability. Hence, NFC enabled handsets allow the end consumer to interactively participate in the *Internet of Things* in a way like never before. Consumers can use their handsets to retrive further information by touching tags integrated in prodcuts or shelves. The other way around, the handset itself also can be used as a transponder and therefore provides additionaly functionallity in terms of applications and identification. This vision requires interoperablity on different layers and a common agreement of industry players integrating technology and applications.

2 Near Field Communication

An NFC device offers different operating modes which are the following [12]:

Reader/Writer Mode: Operating in this mode, the NFC device can read and alter data stored in NFC compliant passive (without battery) transponders. Such tags can be found on *SmartPoster* e. g., allowing the user to retrieve additional information by reading the tag with the NFC device. Depending on the data stored on the tag, the NFC device takes an appropriate action without any user interaction. If a URI was found on the tag, the handset would open a web browser for example.

Card Emulation: An NFC device can also act as smart card (ISO 14443) after being switched into card emulation mode. In this case an external reader cannot distinguish between a smart card and an NFC device. This mode is useful for contactless payment and ticketing applications for example. Actually, an NFC enable handset is capable of storing different contactless smartcard applications in one device.

Peer-to-Peer: The NFC peer-to-peer mode (ISO 18092) allows two NFC enabled devices to establish a bidirectional connection to exchange contacts, Bluetooth pairing information or any other kind of data [8]. Cumbersome pairing processes are a thing of the past thanks to NFC technology. To establish a connection a client (NFC peer-to-peer initiator) is searching for a host (NFC peer-to-peer target) to setup a connection. Then the NDEF (NFC Data Exchange Format) is used to transmit the data.

NFC technology integrated in a mobile device typically consists of two integrated circuits. The NFC controller is required for the analog digital conversion of the signals transferred over the proximity connection. An HCI (host controller interface) allows the host controller to set the operating modes of the NFC controller and process data sent and received. The second IC, a secure smartcard chip also referred to as the secure element, is used for the tag emulation mode. The secure element is connected to the NFC controller for proximity transactions (external mode e. g. for payment at point of sale) through the Single-Wire Protocol (SWP). The host-controller as well is able to exchange data with the secure element (internal mode e. g. for top up of money into the secure element over the air). Different implementations of secure elements are extensively discussed in [2]. The components of an NFC device are illustrated in Figure 1.

3 Participants in Standardization and their Achievements

3.1 ISO, ECMA and ETSI

The RF-layer and the NFCIP (Near Field Communication - Interface and Protocol) of NFC are already standardized (ISO 18092, 21481, 22536 and 23917; ECMA 340, 352, 356 and 365; ETSI TS 102 190). NFC is operating at 13.56 MHz and transferring data up to 424 Kbits/second. ECMA 340/352 and ISO 18092/21481 are more or less equivalent and describe the Near Field Communication Interface and Protocol (NFCIP-1 and -2) and ECMA 356/362 and ISO 22536/23917 with test methods for interfaces and protocols. Theses official standards do not considers data or applications on top and have been adopted between 2004 and 2006.

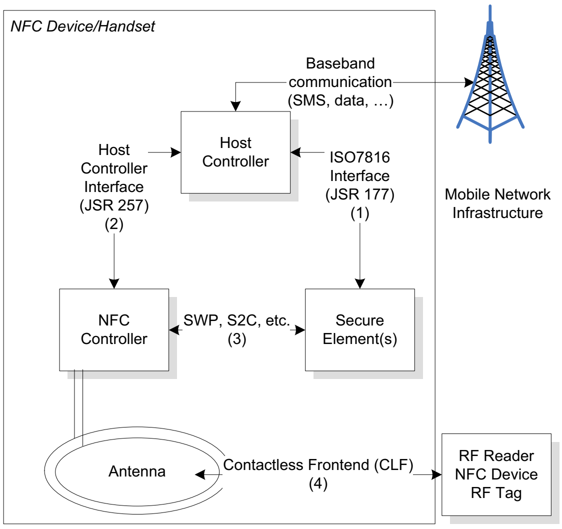

NFC Device/Handset
Baseband communication (SMS, data, ...)
Host Controller
Host Controller Interface (JSR 257) (2)
ISO7816 Interface (JSR 177) (1)
Mobile Network Infrastructure
NFC Controller
SWP, S2C, etc. (3)
Secure Element(s)
Antenna
Contactless Frontend (CLF) (4)
RF Reader NFC Device RF Tag

Figure 1. Architecture of NFC integrated in a mobile device.

3.2 The NFC-Forum

The NFC-Forum[1] was founded in April 2004 by Nokia, NXP Semiconductors, and Sony. The goal of the NFC-Forum is to deal with all related issues above the NFCIP. The most important topics are the already specified data format NDEF, security, and interoperability. Hence, common standards for payment (like EMV[2]) are not considered. More generally speaking the NFC-Forum covers all functional details of the NFC controller (see Figure 2). Aspects with regard to architecture are outside the scope of the NFC-Forum. However, the NFC-Forum is cooperating with the GSMA in order to have a common understanding of how to integrate the functionality into a mobile phone.

NXP and Sony do already have a long experience in the contactless industry. NXP holds about 80 % of the contactless transponder market with their product *Mifare* [15] whereas Sony dominates the Asian market with *Felica* [17]. Mifare, based on ISO 14443-A, and Felica, based on ISO

[1] `http://www.nfc-forum.org`

[2] Europay, Visa, MasterCard

14443-B, are both contactless memeory cards but do have a proprietary implementation of ISO 14443-4 (protocol layer) and therefore are not compatible. For example London's public transport systems *Oyster* uses Mifare-Cards whereas the Japaneese *Suica* Card used in Tokyo's public transport system is based on Felcia. Now the idea was to create a common standard for contactless technology instead of two different implementations: Near Field Communication. This would users allow to use their handsets for applications and services like never before. Soon MasterCard and Visa, covering the payment industry joined the NFC-Forum in order to standardize the technology to be suitable for applications on top. Other members of the steering committee like Microsoft, Samsung or NEC are planning to integrate the technology into their products as well.

Until today (Q1/2008) the NFC-Forum has agreed upon three specifications:

- NFC Tags: A set of tags which must be readable by any NFC device.
- NDEF: The NFC Data Exchange Format between NFC Devices and NFC Tags.
- RTD: Four Record Type Definitions which are information junks to be stored on NFC Tags and transmitted using the NDEF format.

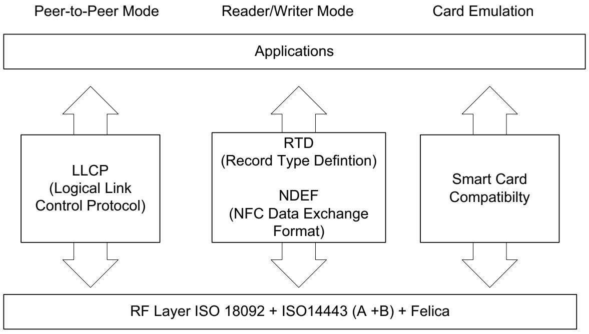

Peer-to-Peer Mode
Reader/Writer Mode
Card Emulation
Applications
LLCP (Logical Link Control Protocol)
RTD (Record Type Defintion)
NDEF (NFC Data Exchange Format)
Smart Card Compatibilty
RF Layer ISO 18092 + ISO14443 (A +B) + Felica

Figure 2. NFC-Forum's scope during standardization of Near Field Communication.

3.3 GSMA

The Global System for Mobile Communications Association (GMSA[3]) is a global trade association representing over 700 GSM mobile phone operators, more than 200 manufacturers and suppliers support the Association's initiatives as associate members. The primary goals of the GSMA are to ensure that mobile phones and wireless services work globally and are easily accessible, enhancing their value to individual customers and national economies. Thus the GSMA is investigating NFC technology as well.

The GSMA established a special NFC working group in the 2nd half of 2006 and already published two white papers dealing with NFC services [5] and NFC technical guidelines [6]. After a careful process of evaluation the GSMA announced in November 2007 that the Single-Wire-Protocol (SWP) will be used as the missing link for communication between the NFC controller and the SIM card (Figure 1, (3)). Noll et al. raised this issue already in [14]. This does not exclude the use of other communication interfaces for additional secure elements.

The GSMA is also driving the SIM as the favored secure element, thus the importance of a mobile network operator will be leveraged. For Q1/2008 the semiconductor manufacturers are expecting the Host Controller Interface (HCI), which is the missing link between the NFC controller and the host-controller, to be defined (Figure 1, (2)). Then all necessary interfaces required for a smooth integration of NFC into a handset are adopted.

3.4 Java Community

The Java Community Process (JCP[4]) holds the responsibility for the development of Java technology. As an open, inclusive organization of active members and non-member public input, it primarily guides the development and approval of Java technical specifications. The JCP already adopted two important API specifications for NFC devices:

Contactless Communications API (JSR257) This JSR was released in 2006 and describes the necessary interfaces in order to allow contactless transactions with a J2ME application running on the handset. Thus this API makes use of the reader/writer mode as well as the NFC peer-to-peer mode. The JSR257 already implements the NDEF format and the basic RTDs published so far by the NFC-Forum [10].

Secure and Trust Service API (SATSA, JSR177) The JCP released the SATSA in 2004. The intended goal of this API was to provide cryptographic functionality of a smartcard chip to J2ME applications. Also, the use of a secure storage for DRM certificates and digital signatures was a use case during the definition. With the introduction of NFC and the use of a smartcard chip for tag emulation, this API received a boost in importance. Recently in 2007 a maintenance release was published [9].

[3] http://www.gsmworld.com

[4] http://jcp.org/

Both APIs allow an application developer to use the functionality of an NFC device. By providing a J2ME API for using NFC technology, the barrier to application developers is lowered in comparison to using proprietary Symbian OS or Windows mobile APIs.

3.5 GlobalPlatform

GlobalPlatform[5] (GP) represents the interests of the smartcard industry, card issuers, and vendors. GlobalPlatform has defined a command set for handling data, applications, and files on a smartcard. The command set relies on the definitions given in ISO 7816-4, inter-industry command for interchange, and can be used for contacted as well as contactless smartcards. In terms of NFC, GlobalPlatform commands can be sent though the JSR177 (ISO 7816) or the RF interface (ISO 14443 over SWP). The smartcard chip in the handset is likely to be GP compliant but can also feature propriety products like NXP's Mifare or Sony's Felica. GlobalPlatform is eager to make sure that the integration of the smartcard chip still fits the need of the consortia.

3.6 Financial Industry

The major driver regarding applications behind NFC is the financial industry. Although applications are not the goal of standardization, all use cases for applications are well considered. As the financial industry like banks and credit card companies are dealing heavily with smartcards, they are also interested in NFC: they expect falling cost for cards being issued, abuse of cards as well as raising revenues and therefore are keen on seeing this technology to be integrated into every handset. Interest groups of the financial industry like the Mobey Forum or the Mobile Payment Forum are members of the NFC-Forum. The standardization of a technology with financial institutions is difficult as Lim lines out in [11]. But the standardization problem itself is not merely a technical but a business problem.

3.7 Others

Besides parties directly involved in the standardization of NFC and the integration of NFC in mobile devices, there are also other groups effecting and being effected by NFC. With regard to services and application the Open Mobile Alliance (OMA[6]) and the JavaCard Forum[7] play a key role regarding the implementation of the SmartCard WebServer (SCWS). The SCWS will play an important role in terms of NFC, as it would allow both, wide area and proximity connections in one component. Whereas the OMA specifies only the interfaces to the smartcard itself, the JavaCard Forum defines the integration of the whole server into the JavaCard OS. The SCWS is specified for JavaCard OS 3.0 (cards expected by 2010), but there are already implementations working with JavaCard OS 2.2.

4 Upcoming challenges

All interested groups and industrial players expect additional revenue and market share from introducing NFC technology. Each player wants to have the biggest share of the pie and thus uses his power to have as much influence of the process standardization as possible. Thus the process of standardization moves away from a technical discussion and ends more or less in a negotiation process. Economical and political debates are the order of the day [16].

4.1 OTA Management

According to Collins' analysis in [3] the OTA (over the air) management of NFC applications plays a key roll in an NFC ecosystem. OTA management allows an instance to remotely load data into the secure element. Hence, the definition of the air-interface as well as the processes and protocols for uploading and managing applications needs to be handled. The air-interface definition is in the scope of competences of the GSMA and the OMA. GlobalPlatform and GSMA would need to agree on a definition for remotely accessing the secure element. OTA management will add additional complexity to the standardization, as a new stakeholder acting as a platform manager for managing the secure elements will be involved. As there are no consortia dealing with this issue yet, there are already companies implementing proprietary solutions for OTA management. In case this issue will not be turned to immediately, a standard that might be adopted too late would not be used, as proprietary solutions might already dominate the market. This has been the case in the mid 1990s during the standardization of contactless smartcards (ISO 14443) and ended in three different implementations (Type A, Type B and Felica). Ironically, this is among the reasons for creating NFC.

Whereas the reader/write mode and the peer-to-peer mode of NFC do not yet have an extensive business model, replacing smartcards by the secure element in the phone does. The holder of this secure element can lease space to application providers. Hence, the discussions around the secure element and its implementation are of serious interest, especially to mobile network operators and the financial industry. Besides reducing the cost for issuing a smartcard, the possibility of sending data to applications proposes a benefit to customers as well as application providers (see figure 3): updates within the application can be performed

[5] http://www.globalplatform.org

[6] http://www.openmobilealliance.org/

[7] http://www.javacardforum.org/

just in time and data such as tickets can be delivered to the mobile phone over the air. Hence, the provider of the secure element, also referred to as the platform provider, as well as the platform manager is going to play major role in an NFC ecosystem. In order to have an open system a standard for the OTA management is required to leverage the market entry barrier for service providers and the penetration of applications.

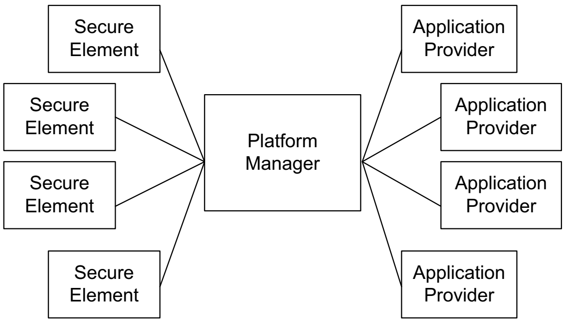


Figure 3. Platform Manager acting as aggregator between secure elements and application providers.

4.2 Multiple Secure Elements

An NFC device is typically made up of two parts: the NFC part responsible for RF communication, and a secure element storing sensitive data. Traditionally, they are treated as separate devices represented by different UIDs. But this approach poses several problems. Despite of possessing multiple IDs, an NFC device has only one antenna. This resource must be shared. From a technical point of view it is possible to mix both data streams thus allowing simultaneous access. However, certain existing RFID infrastructures – which NFC should be compatible with – do not allow more than one device in range (e.g. Mastercard's PayPass). This is because of security concerns. Thus, an NFC device should only be represented by one ID during the process of anticollision.

Additionally, the internal handling and routing of data streams needs to be considered. The specification in ETSI for a Multi Host Controller (MHC) for multiple secure elements in one device is still ongoing [4]. This standard specifies the protocol layer for attaching multiple secure elements to the NFC controller (see Figure 4). Unfortunately, it does not consider yet the management of the secure elements themselves. Thsu, another interface, such as a Secure Element Controller (SEC) for routing the data streams between the contactless frontend (CLF) and the appropriate secure element is required. Considering this issue would be part of the NFC-Forum's as well as the GlobalPlatform's task.

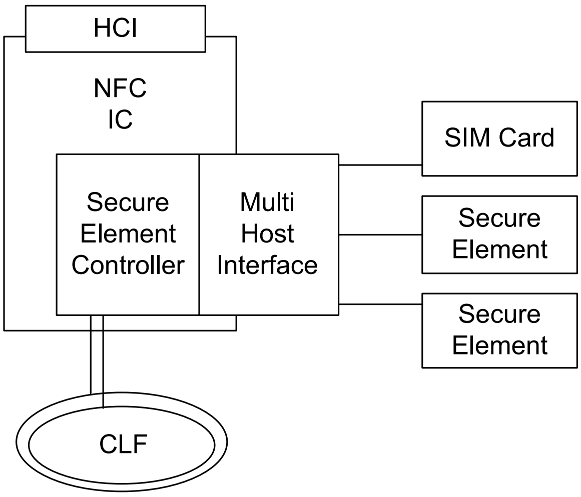


Figure 4. Integration of a central management instance for all secure elements

4.3 Tag Management

As NFC devices can read external tags, the issuing process as well as the content of such tags requires protection. A user reading information of a SmartPoster, a tagged shelve or product can easily be missled, by destroying the original tag in the poster and replacing it with a tag with malicious content. Hence, in order to put trust in the content stored in the tag, it would require a signature. To verify this signature, the handset requires the correct certificate. The management of the certificates as well as the setup of the whole public key infrastructure needs to be considered. Of course, already established certificate authorities are able to issue such certificates, but an NFC tag has a very limited capacity and thus needs a specially designed signature format. The integration of such a features requires the work of the NFC-Forum in cooperation with industry players already acting as certificate authorities. This is especially necessary in terms of tagging products or shelves in order to provide additional information n a trusted way. However, also the combination of logistic information and consumer/product information in one tag/chip should be considered as well, in order not to have two different tags on one product. At the moment the standards for logistics and NFC are not compatible on several layers. This is a potential barrier for the *Internet of Things* for the end consumer.

5 Conclusion

In order to provide an interoperable service for consumers, it requires more than having the bottom layer standardized. A device that features ISO 18092 may not be able to act as a contactless creditcard or a reading device for a smartposter. Therefore, also all layers above require a clear definition of data exchange format, interfaces, and APIs. The present problem during the standardization of a technology is, that some stakeholder may not disclose their intention in the beginning. Thus a top-down standardization with a holistic overview is not possible. With regard to NFC, the major challenges in standardization are the synchronization with other consortia as well as the different views of the members of the NFC-Forum.

The standardization of NFC goes hand-in-hand in with the applications and services. Therefore, many different institutions with different core-competences are involved. The setup up of working groups and the definition of the interfaces between the groups as well as the parts to work on, is a time consuming process. The major problem is that there is no central institution coordinating the standardization approaches. It may happen that not the best technical solution is chosen, but the one of the player with the most power.

As the standardization of NFC technology shows, many different parties are involved in such a process. This kind of distributed standardization causes a significant overhead for dividing up tasks and the definition of interfaces. Additionally, the participation of groups with different interests make the process of standardization tough, as it moves back and forward without a common agreement. But on the other hand many different views and opinions are considered during the standardization.

Contactless technology for the *Internet of Things* requires all parties to agree one common definition and implementation. Having different implementation of one technology blocks interoperability, confuses users and raises the market entry barrier for companies. The idea of having a technology in mobile phone allowing everybody to participate in the *Internet of Things* will only work if tags, devices, readers and application work seamlessly together and if there is an open market for these components.

References

[1] ABI Research. Near Field Communications (NFC) - Leveraging Contactless for Mobile Payments, Content and Access. Research Report, 01 2007. Report Code: RR-NFC.

[2] C. Bishwajit and R. Juha. *Mobile Device Security Element*. Mobey Forum, Satamaradankatu 3 B, 3rd floor 00020 Nordea, Helsinki/Finland, 02 2005.

[3] J. Collins. ABI Research Insight: No OTA, No NFC. http://www.abiresearch.com/, 10 2007.

[4] ETSI. Smart Cards: UICC-CLF interface; host Controller Interface. www.etsi.org, 12 2007. Draft (Release 7).

[5] GSMA London Office, 1st Floor, Mid City Place, 71 High Holborn, London WC1V 6EA, United Kingdom. *mobile NFC Services*, 1.0 edition, 02 2007. 1st Revision.

[6] GSMA London Office, 1st Floor, Mid City Place, 71 High Holborn, London WC1V 6EA, United Kingdom. *mobile NFC technical guidelines*, 1.0 edition, 04 2007. 1st Revision.

[7] International Organization for Standardization. Proximity cards. ISO/IEC 14443, 2003.

[8] International Organization for Standardization. Near Field Communication - Interface and Protocol (NFCIP-1). ISO/IEC 18092, 2004.

[9] Java Community Process (SM) Program. Java Security and Trust Services API (SATSA). http://java.sun.com/products/satsa/, 09 2004. JSR177 Final Release.

[10] Java Community Process (SM) Program. Java Contactless Communications API. http://jcp.org/en/jsr/detail?id=257, 09 2006. JSR257 Final Release.

[11] A. S. Lim. Pre-standardisation of mobile payments: Negotiations within consortia. *ICMB*, 04:392–399, 2005.

[12] G. Madlmayr, O. Dillinger, J. Langer, C. Schaffer, C. Kantner, and J. Scharinger. The Benefit of using SIM application toolkit in the context of Near Field Communication applications for mobile applications. In *Proceedings of the 6th International Conference on Information and Communication for mobile Business (ICMB2007)*, volume 06, page 7. IEEE Computer Society, 07 2007.

[13] F. Michahelles, F. Thiesse, A. Schmidt, and J. R. Williams. Pervasive RFID and Near Field Communication Technology. *IEEE Pervasive Computing*, 6(3):94–96, c3, 2007.

[14] J. Noll, J. C. L. Calvet, and K. Myksvoll. Admittance Services through Mobile Phone Short Messages. *ICWMC*, 1:77, 2006.

[15] NXP Semiconductors. Mifare - contactless authentification. `www.nxp.com/products/identification/mifare/`, 2008. Online; accessed 02/24/2008.

[16] R. M. Richard Hawkins and J. Shea, editors. *Standards, Innovation and Competitiveness: The Politics and Economics of Standards in Natural and Technical Environments*. Edward Elgar Publishing Company, 1995.

[17] Sony Global. Felica. `www.sony.net/Products/felica/`, 2007. Online; accessed 02/24/2008.

The Internet of Things: On Standardisation in the Domain of Intralogistics

Lars Nagel
Fraunhofer Institute for Material Flow and Logistics, Dortmund
lars.nagel@iml.fraunhofer.de

Moritz Roidl
Chair for Materials Handling and Warehousing, Technische Universität Dortmund
moritz.roidl@tu-dortmund.de

Guido Follert
Chair for Materials Handling and Warehousing, Technische Universität Dortmund
guido.follert@tu-dortmund.de

Abstract

The Internet of Things (IoT) in the domain of intralogistics represents a broad decentralisation effort in the areas of data management and control flow and provides a promising concept for the control of material flow systems. Several research projects are carried out at the Fraunhofer Institute for Material Flow and Logistics in Dortmund (IML) as well as at the Technische Universität Dortmund. These projects deal with the standardisation of decentralised storage and material flow systems. A well-founded know-how in the field of multi-agent based control of material flow systems and experiments in AutoID systems sets a solid knowledge base for standardisation efforts. Insights from research projects with industrial partners in connection with the standardisation of the IoT are presented.

1. Introduction

Intralogistics is a cutting-edge term that comprises all technical systems, services and related business involved in the in-house materials handling of industrial enterprises, wholesalers, retailers and government institutions. The processes of the intralogistics domain are vital for managing the flows of goods along the entire supply chain (supply chain management) as they provide the reliable and predictable flow of physical goods in the nodes of a supply network.

The domain of intralogistics spreads in the area of conflict between systems with simple manual operations and few technical solutions adding little value up to highly automated, complex systems containing several organisational areas and functionally specialised solutions. An extract of applications of intralogistic systems are bulk-good terminals of forwarders, picking systems for the mail order business, assembly lines in the automotive industry, baggage handling systems in airports and high dynamic distribution centres for wholesale and retail. Intralogistics fills a key position between engineering and economy for companies acting in global supply chains (cp. [1], [2], [3]).

A distinguishing mark between intralogistics and external logistics is the insularity of the respective systems. Intralogistic systems can be seen as closed systems with clearly defined interfaces to the outside world. May these systems be as complex as described before, they are anyway completely controllable. The difficult task trying to control a material flow system including the occurring obstacles will be described in advance. The mere possibility to theoretically shape an intralogistic system autonomously distinguishes this domain from the external logistics and opens a perfect test-bed for the development of material flow control in the range from proprietary solutions over solutions with de facto standards up to those using de jure standards (cp. [4], [5]).

The Internet of Things (IoT) in the domain of intralogistics represents a broad decentralisation effort in the areas of data management and control flow. In the past years, the development of the Radio Frequency Identification (RFID) technology has opened up the possibility to eliminate the central data warehouses in logistic systems: the data about a physical object are held in an RFID tag and travel with it; the material flow is united with the information flow (cp. [6]).

The current research concentrates on the decentralisation of decision-making, resulting in the elimination of the central material flow controller. The IoT approach in the intralogistics domain is based on the agent concept from the field of Artificial Intelligence. In one of its simplest definitions an agent is an entity which observes its environment through sensors and reacts by means of actuators (cp. [7]). This definition equals that of a material flow automaton. There it is the automaton which observes a logistic process through sensors and modifies it by means of actuators (cp. [8]). Such an automaton can be viewed as an agent who resides in a logistics system. The difference between artificial intelligence and material flow automation is in the approach. A classical agent has to act in a rough or even hostile environment; it has to learn how to adapt and to survive. Things are different in the domain of intralogistics. In the automation of logistics systems, and as mentioned before in a higher dimension in intralogistics, the environment is heavily standardised to facilitate the supervision and action of the automatons. It has been shown, that controlling a material flow system by the means of agents is feasible. During the processs of realising a material flow system that is controlled with the concept of the Internet of Things, a large number of necessities to standardise came across.

The process of research and development leads from fundamental research over applied research to experimental development and ends up with a marketable product. In the first stage, the participants in research have to agree upon a consistent terminology. Such a semantic standard empowers several different parties with different aims in the same projects. With proceeding results in the research, the stage of applied research is reached. This stage is characterised by the need of integrating the solutions of the different partners into a functional system. At this point, interoperability becomes important and technical or de facto standards in a shape of the definition of interfaces, whether they are physical or data based, arise. The goal of further experimental development is the marketable product and finally the diffusion into the market. In this stage, compatibility and quality gain weight and de jure standards concerning these aspects must be developed. These efforts point on reliability, trust, flexibility and economies of scale and are usually accompanied by industrial organisations and federations.

During current research at the Fraunhofer IML and the TU Dortmund, the two aspects interoperability and compatibility were topical.

Standardisation in the domain of intralogistics works on two sectors. Firstly, compatibility defines the interaction of components on the facility level and raises topics like modularity of technical components, for example of conveyors. Standardisation in this sector allows the operator to flexibly design his facility and to easily change systems' layouts (cp. [4]). From the operator's point of view compatibility is crucial, because certain groups of components must be exchangeable, so that in case of maintenance, for example, a third-party product may be chosen.

As compatibility focuses on events during the lifecycle of a facility, interoperability points to the simultaneous cooperation between different components or systems. The interaction of all components and systems involved generates the capacity of the facility. Information systems must be able to interact with actuators and sensors; control nodes must identify codes on the tags as well as address certain actuators. The interaction between single points in the network has to be standardised to streamline the logistic processes and to bring them to life at all. Hence, interoperability allows the operator to enhance the capacity and scale of benefits. Two or more subsystems work together in order to efficiently and usefully exchange information or physical capacities so that a more significant service will be provided for the user. Interoperability is the ability to make two heterogeneous and independent – and often completely different – systems work together. Standardisation is the vehicle to enable this. As a consequence, the process of standardisation in the focus of IoT has to deal with materials handling equipment on the one hand and with matters of information and communication on the other hand.

2. Existing Standardisation Efforts

Industrial federations in Germany, like the VDMA (Verband Deutscher Maschinen- und Anlagenbau/Association of German mechanical engineering and plant construction), which focuses on mechanical and plant engineering and pushes research and development in Logistics, or professional institutions, like the VDI (Verein Deutscher Ingenieure/Association of German Engineers), which embraces the concerns of all kinds of engineers, bring forward efforts in modularisation of materials handling equipment. As large associations and representatives of the German industry, VDMA and VDI influence a big part of the German GDP. Their attributes of neutrality and non-profitability help to come to cross-industry agreements on solutions and standards. Technical guidelines and symposia as well as cross-functional working groups are examples of these efforts.

One concept with a strong interrelation to IoT is called SAIL (German acronym for "system architecture for intralogistics"). It was developed and promoted by a group of industrial and software engineers out of the VDMA (cp. [9]). Basically, it is a recommendation for the architecture of an intralogistics system design. It is a result of standardisation efforts to gain more efficient interfaces between the individual crafts within a comprehensive intralogistics system. The approach of SAIL is a systematic arrangement of core functions in an intralogistics system, defining standard control functions and interfaces between them. SAIL abolishes the old hierarchical allocation of functions and replaces it by independent functions, which only refer to each other by predefined and standardised interfaces.

The standardisation of materials handling equipment has not yet become an industry-wide purpose. Furthermore, the components, like motors, drive elements and sensors, are standardised, while on the other hand the companies internally administer design sets for such equipment. Thus they facilitate the process of setting up new layouts and reduce the manufacturing costs of the equipment.

In the future, more flexibility will arise from new methods of material flow control, like decentralisation and IoT, and furthermore it will be emphasised by innovative strategies which go along with them. As a consequence, an increasing demand for independent but standardised materials handling equipment can be forecasted. Those tendencies are already discussed in the automotive industry and indicated as “alterable material flow systems” (cp. [10]).

3. Past and Current Research

Intralogistics systems are nowadays challenged by the increasing global integration of supply chains, which can only be managed sufficiently by means of complex IT-systems. New materials handling and information technologies in the logistics sector have to accomplish a complete coverage of objects and information in real time. RFID is the technology which fulfils these requirements on a very basic level and implements by this the long since desired alliance of physical and information flow.

At the Fraunhofer Institute for Material Flow and Logistics, Dortmund (IML), the *openID-center* was founded in order to cope with the challenges of this task and to create an open source platform for logistics software and AutoID systems. In the experimental field of the *openID-center* a community of logistics and IT-companies has room to cooperate with academics and scientists, designating their work to the creation of innovations and standards. In this environment a comprehensive supply chain of returnable transport units is set up. All echelons of a supply chain, i. e. the unitising, goods income and order-picking, are covered by this demonstration process. Complementary to the development of technical devices the benchmarking of solutions based on performance and economical indicators is accomplished in the *openID-center,* which offers a suitable environment for experiments dedicated to standardisation procedures (cp. http://www.openid-center.de).

In addition to the efforts made in the *openID-center* in the past, Fraunhofer IML could gain insight into the communication with material flow systems running more than 100 autonomous storage vehicles. It could develop the Multishuttle system in cooperation with Dematic, prove its feasibility and develop it into a product. Within the Multishuttle system, first experiences have been made with the standardisation of vehicle designs as well with multi-agent based communication (cp. [11]).

While the Multishuttle, as a dedicated solution, was successful in a market niche, a more general decentralisation effort is being made with the help of simulation. The integration of the concept of multi-agent networks within an existing discrete event simulation model allows for the direct comparison between classical centralised material flow control and a multi-agent based control. The use of automated analysis and code generation enables the reuse of large-scale, industrial models which represent a more realistic testing ground than common test-bed environments. A created environment could demonstrate the simulation of a large-scale baggage handling system, in which the routing is controlled by a multi-agent system using an adapted version of Dynamic Source Routing. The agents and the communication infrastructure are integrated into the model to facilitate the analysis of the interaction between agents. The results proved that a decentralised routing algorithm is able to control a large real-world system as effectively as a centralised controller. Another lesson learned was that the decentralisation effort opened up a field which traditionally belongs to the proprietary part of logistics systems: the planning of routes. A subsequent standardisation of routing protocols will enable global routing in heterogeneous environments and, as a result, optimize the transport over complete logistics chains (cp. [12]).

4. Standardisation Approach

Stakeholders in logistics projects generally can assume different roles. Usually they assume two or more roles at the same time or share roles. Common archetypes of stakeholders in the process of planning logistic systems are:

- Material handling equipment supplier
- Control hardware manufacturer
- System planner
- System integrator
- User of the logistic facility

In the following, just the core tasks of the stakeholders, which may become crucial in the process of standardisation, are worked out.

Material handling equipment suppliers provide mechanical components and actuators which fulfil the physical movement of goods in logistic systems. They design components compatible along their product lines to reduce production complexity. The components usually provide interoperability due to manifest existing logistic standards like the euro-pallet. In addition, most manufacturers broaden their product spectrum by selling control hardware, since margins for mechanical hardware decrease. In this context innovations are nearly always achieved by specialised proprietary solutions which combine mechanics and control hardware.

The next stakeholder in the field of building up material handling systems are control hardware manufacturers, who provide sensors, computational hardware, the communication infrastructure and software development toolkits. Their business normally is not logistics centred but evenly spread across all industry sectors. Therefore, they enter the process of engineering a material flow system with highly standardised components in terms of type series.

System planners analyse processes, plan layouts and carry out the arrangement of mechanical parts and control hardware. They assure the mechanical performance of the system and substantiate the requirements for the control software. System planners are mostly the connecting factor amongst all stakeholders planning a material flow system and are responsible for the overall capacity of the system.

The system integrator is responsible for the implementing of the working system. The duties are dependant on what the other stakeholders provide but very often include the actual programming of the control software.

The system user expects an operative system and pays for the facility. Sometimes the user additionally takes on the roles for planning and integrating the system to keep the inherent implementation and the process knowledge inside the company. Capacity, flexibility and robustness are his core interests.

As a result of this comparison can be stated, that manifold interests interfere during the process of implementing the philosophy of the IoT as a concept of material flow control into the market.

Furthermore not one single successful strategy for a company in the market is currently observable. For example, a company like Siemens sells complete solutions to system users as well as its control hardware (the SPS) separately to system integrators. Another aspect is the existence of various examples of successfully working proprietary systems. To put it in a nutshell, from the single stakeholder's point of view there is not necessarily a natural need for standardisation. But it is common knowledge in the intralogistics domain that a large part of the failed projects showed their biggest problems with interface problems and standardisation issues. The need of a reasonable standardisation process can be read from the experiences of failed projects.

Standards in general point on harmonising interfaces and in the consequence on generating products. This leads to a smaller market segmentation and finally to stronger competition between the suppliers. In the first stage, this creates lower prices and produces reluctance concerning further collaboration of the suppliers for fear of rivalry. However, this fear is superficial for a larger market leads to higher sales of the respective products. Therefore participating in the process of standardisation always holds a benefit for the stakeholders (cp. [5], [13]).

Our approach to standardisation in the IoT before the aforementioned background can be described as a bottom-up process. It starts with the creation of a research project with industrial partners. Those partners usually represent all major players in the respective German market. Essentially, the research project is a collective effort to build a prototype which proves the validity of new technologies. During the building process interfaces are created in order to make the prototype executable. The design of these interfaces is based on mutual agreements between the technical representatives of all partners. The result can be described as an technical or de facto standard since it is agreed upon by all partners on a technical level.

Beyond our research there exists a multitude of approaches trying to prove the feasibility of IoT technologies in prototypical environments. Surely, several parallel de facto standards by different teams have already been created. Hence, it is the challenge to create one official, de jure standard based on the common findings of these teams. One core obstacle that has to be overcome is the concerns of the participating partners, especially of those from the industry. Technical solutions in a proprietary shape form unique selling points for industrial partners as described before. Abandoning these technical strongholds is often a task of the academics and institutions during the standardisation process by persuading the industrial stakeholders that a vision will become a rewarding business case. A balance must be agreed upon where unique selling points and individual business models concerning certain technical features can be retained and simultaneously the best possible design of the material flow system achieved by an agreement on a certain amount of standards can be created. In the discussions on how to develop standards in the IoT, it is a challenge for all participants to push the borderline as far as possible, so that the common optimum – the overall benefit and capacity of a material flow system that is controlled by the means of the IoT – will be reached on a win-win basis.

Nevertheless, crossing borders this way occasionally provokes resistance. How it can be transformed into an advantage was shown by a current research project of the Bundesministerium für Bildung und Forschung (BMBF – the German Ministry for Education and Research). The title of the project is "Internet der Dinge" ("Internet of Things" in German) and it focusses on the creation of a prototypical IoT-enabled intralogistics environment as described above. In the technical sessions a discussion circled around the specification of communication protocols. The attendant technical representatives tried to agree upon a communication interface on a telegram basis with specific data structures. This approach provoked the resistance of a couple of companies which did not want to commit to a specific technical solution. As a result, a compromise was made based on the specification of messages and their content on a more semantic level, accompanied by a guideline and a reference implementation. This implementation then comprised the original telegram specifications. The advantage of the compromise was the continued contentment of all participants and the avoidance of a stalemate in the project. As shown by this anecdote, standardisation can be realised by solutions and compromises on a technical level. The better structured an idea solves an evolved conflict, the more participants are willing to agree upon this solution and help to establish this newly specified standard. It will be the mission of the IoT standardisation community to produce a spectrum of these kinds of elaborate solutions that will set the basis for a lasting and widely spread standard.

5. Summary

The standardisation process in IoT-related technologies in the domain of intralogistics is far from being completed. Nonetheless, there are various research projects running right now which focus on the problem. A significant number of industrial partners participate in most of these projects. From interviews with these partners, we could identify two motivations for participation: the fear of missing an opportunity and the possibility to control the evolution of potential standards. On the one hand, a standardisation process opens up new business opportunities. On the other hand, it can mean a direct loss of unique selling points for a company.

6. Conclusion

The pursuit of standardisation in research projects with industrial partners is a delicate affair. The best efforts by all participants can become thwarted if standardisation approaches unique selling points of participating companies, be they either real or imagined. We think that standardisation will not only generate a larger market in general but will also form the basis for innovative logistics concepts which are currently not feasible. The applied scientist has to transfer that knowledge to the industrials partners. Only if the prospect of development opportunities outweighs the fear to lose traditional technical strongholds the standardisation process can become a success. Applying the IoT-technologies means a broad decentralisation effort which touches the complete information infrastructure in intralogistic systems. The decentralisation of control algorithms will generate a heap of additional communication interfaces. Networks of distributed entities will need standardised communication languages to stay controllable. If this can be achieved, it will very likely end the dominance of the central server as a classical proprietary black box. In return for that loss, from a business perspective, the current research will have to come up with new algorithms, and in the end, allow for new business models, which are only possible in a completely decentralised environment.

[1] Tompkins, J., *Facilities Planning*, 3rd Edition. Wiley, New York, 2002.

[2] ten Hompel, M., Schmidt, T., Nagel, L., *Materialflusssysteme – Förder- und Lagertechnik*, 3rd Edition, Springer, Berlin, 2007.

[3] Arnold, D. (Ed.), *Intralogistik – Potentiale, Perspektiven, Prognosen*, Springer, Berlin, 2007.

[4] Farrell, J., Saloner, G., "Standardisation, Compatibility, and Innovation", *The RAND Journal of Economics*, Vol. 16, No. 1, Spring, 1985, pp. 70-83.

[5] Stango, V., "The Economics of Standards Wars", *Review of Network Economics*, March/2004, Volume: 3 , Issue: 1 , Pages: 1-19.

[6] Bullinger, H.-J., ten Hompel, M. (Eds.), *Internet der Dinge*, Springer, Berlin, 2007.

[7] Russel, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Englewood Cliffs, 1995.

[8] Ten Hompel, M., Schmidt, T., *Warehouse Management: Automation and Organisation*

of Warehouse and Order Picking Systems, Springer, Berlin, 2006.

[9] Balbach, U., “Standards für erfolgreiche Logistikprojekte“, *Logistik für Unternehmen*, Issue 3/07, Springer, Berlin, 2007.

[10] Wilke, M., *Wandelbare automatisierte Materialflusssysteme für dynamische Produktionsstrukturen*, Dissertation, Technische Universität München, Herbert Utz Verlag, München, 2006.

[11] Trautmann, A., Daniluk, D., “Hardwarenahe Emulation des Multishuttle für ein Multiagenten-basiertes Steuerungssystem durch Automod und HLA“*, Produktion und Logistik 2006, ASIM Tagungsband zur 12. Fachtagung*, SCS Publishing House, Erlangen, 2006.

[12] Roidl, M., Follert, G. “Simulation von multiagentenbasierten Materialflusssteuerungen“, *Informatik 2007 – Informatik trifft Logistik, Beiträge der 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, *Volume 1*, 2007.

[13] Blind, K., *Theory of Standards: Theory, Evidence, Policy*. E. Elgar Publishing Ltd., Cheltenham, 2004.

Supporting Standardisation of an Open Governance for the EPCglobal Network:
the French Initiative to Put a Theory into Practice

Nicolas PAUVRE
GS1 France
nicolas.pauvre@gs1fr.org

Abstract

As we move forward towards the ambient intelligence environment where most devices are connected to seamless, ubiquitous networks, the inter-enterprise interoperability is an essential condition. While that sounds fairly clinical, the exciting opportunities come from the fact that the process will be developing a complex network which will have the characteristics of 'The Internet of things'. The icing on the cake comes from the opportunity to develop an open governance model for this style and size of network.

That is the subject of this paper focusing on the large scale EPCglobal network in which only one ONS (Object Naming Service) root is in existence today.

In order to create a richer offering able to address the increasing complexity of the intelligent networks, GS1 France is developing an "ONS in Europe" that would be managed on a shared basis, linked to a number of local and possible industries specific ONS.

1. Concept and objectives

A world where global supply chains are the norm requires that RFID tags and associated sensors can operate, can be seen and can be interrogated anywhere in the world.

As we move on from localised RFID applications towards the 'Internet of Things'' in a networked world, we can identify three levels of standards development in ambient intelligence supply chains:

1) Development of the operating characteristics, intelligence and sophistication of the physical objects (through readers, tags, sensors etc.)
2) Clear definition of user-based models and data definition exchange standards
3) A federated network infrastructure, based on an openly defined and implemented network architecture which provides the communication and information transport for the ambient supply chains. This network needs to be able to operate both locally and to be linked globally (ie the ability to be de-centralised and to appear to be centralised).

Related to the last level (ie. level 3) and through the ONS root project led by GS1 France, we suggest to practice on the basis of a concrete case during this workshop.

This workshop will allow us to demonstrate that an increase in scale of the EPCglobal network implies to support a new open governance model, in the practical perspective offered by the proof-of-concept root ONS plateform.

2. Progress beyond the state-of-the-art

To understand the need of the EPCglobal network evolution we have to remember that, at the beginning, it was based predominantly on the needs of food manufacturers and retailers. Therefore, the architecture of the current EPCglobal network is heavily focused on the needs of these business scenarios.

In the course of time, organisations are beginning to adopt RFID further up and down the supply chain and also beyond small scale or sporadic deployments, involving a growing number of industries in various sectors such as healthcare, aerospace, automotive, defence etc.

So the next phase of the EPCglobal network development will have to allow flexible integration of product information provided by a large number of organisations horizontally across the supply chain, and also vertically across various other industries.

This move from small localised activities to large cross-company and cross-country networks will require both more complete and more comprehensive data sets. This implies efficient data synchronisation, guaranteed data availability and improved data security. There is, as a result, a need for data alignment and standards evolution, including one for a so-called Object Naming Services (ONS), which defines the interface for lookup services by providing quasi-permanent or relatively static links between the identity of a company responsible for an object (often the manufacturer) and the authoritative information services which that company provides.

The extension and improvement of current solutions to other domains originate some questions about effective management of increasing amount and variety of data that will be exchanged between partners. In this way, the participants of the workshop will have to react on further developments of scenarios to characterise the network architecture and infrastructure (definition and implementation), having regard to imagine the future value chain.

To support various organisations, this increase in scale for the network also demands the development of an open governance model. Subsequently, this open governance model can be extended to incorporate various ONS systems from other parts of the world, both on technical and business aspects that would be administrated under a common set of rules. Drawing on the GS1 France project to initiate an ONS root in a European context, we would take advantage of the workshop to define a set of rules for the governance, including for instance standards for naming issues and the use of security tools such as certificate authority, privacy management, etc.

Furthermore, the aim of this project is to give the European Community a leadership role in developing ambient intelligence in the supply chain and thereby enhancing competitiveness through leadership in implementing broadly-based, open business enterprise networks. This development represents the important move onwards from RFID supply chain applications the participants will have to figure out what it implies.

3. Methodoly and process

Through a governance Committee made up of end user companies, the project management and the standard development process are driven by user requirements.

While following a usual path by developing requirements for the network, running a pilot then analysing the results before moving to production, this project focuses on very innovative and challenging topics that will be discussed and exploited to bring the benefit not only to the project's partners or the participants to the workshop but the whole European community.

The project is also developing in connection with other IST projects and standards groups (e.g. BRIDGE, CERP cluster, GS1, ISO/GS1 coordination activity) so that project results can be discussed to influence standards development.

In addition GS1 has a significant solution provider membership, especially in the auto-identification and data exchange arena, with which we work closely to ensure that general developments and standards are feasible ('the art of the possible').

The Global Interoperability Forum for Standards (GRIFS)

Patrick Guillemin
ETSI
Strategy & New Initiatives
Patrick.Guillemin@etsi.org

GRIFS **http://www.grifs-project.eu** is a support action project initiated and funded by the European Union with the aim to improve collaboration and thereby to maximise the global consistency of RFID standards.

The two year project GRIFS started on 1st January 2008. The GRIFS project will initiate a forum that will continue to work constructively after the end of the project through a Memorandum of Understanding between key global standard organizations active in RFID.

GS1 in Europe is leading the GRIFS project in cooperation with the European Telecommunications Standards Institute (ETSI) and the European Committee for Standardization (CEN), and with the help of 12 GS1 national organizations from Europe, Africa, Asia and America.

In the FP7 IST work programme, the European Commission has recognised the importance of inter-enterprise interoperability as we move forward towards the ambient intelligence environment where most devices are connected to seamless, ubiquitous networks.

As we move on from localised RFID applications towards the 'Internet of Things' we can identify three levels of technology-related standards in a world of ambient intelligence supply chains:

- Standards for the operating characteristics of physical objects (readers, tags, sensors)
- Infrastructure standards to define the communications, addressing and structures
- Data exchange standards

There are two other important areas of standardization that must be considered as well. These are:

- Standards related to the technical aspects of allocation by regulators of appropriate spectrum for RFID use;
- Standards related to privacy and security issues affecting RFID use, regulatory and otherwise.

The GRIFS support action project will:

- Work to characterise the variety of standards activities taking place globally;
- Create a number of liaison activities to disseminate information about the importance of global standards and to align RFID standards development globally;
- Put in place the 'Global RFID Interoperability forum for Standards' (GRIFS), comprising global stakeholders, to ensure continuing close collaboration between standards activities.

This should create synergy, catalyse co-operation and avoid duplication of developments - thereby minimising unnecessary business expense caused by incompatible localised standards development and maximising the use of a scarce standards development resource. As enterprise networks and intelligent supply chains grow in number, size and reach the requirement for coherent global standards becomes even more of an absolutely necessary requisite.

How broad is the scope of this project?

This Support Action will focus on the use of RFID in supply chain and related activities. These activities primarily encompass the tracking and tracing of objects and items – physical goods – as they move through supply chains in many different businesses, both in the public and private sector. This also includes the tracking of assets, such as returnable assets (pallets, kegs etc.) involved in logistics, tracking assets to ensure their pedigree (anti-counterfeiting activities) and to maintain, service and support objects throughout their life cycle (such as TVs or railway engines).

1st International Workshop on Interoperable Vehicles (IOV2008)

Message from the Workshop Co-Organizers

Welcome to the First International Workshop on Interoperable Vehicles 2008. The workshop aims to establish a forum to bring together research professionals in both academia and industry to address the technical and economic systems and application issues related to Interoperable Vehicles. With new wireless communication capabilities available like WLAN, WiMAX or UMTS, many services become feasible. This will encourage the already existing telematics strategy from the automotive manufacturers.
Contributions to the IOV2008 workshop have been received from five different countries both from appreciable universities and worldwide operating companies. The submitted papers have been reviewed by the workshop's program committee by a double blind review process, where nine have been accepted for a presentation. The acceptance criteria where defined by the originality, quality and the match to the goals of the workshop.

The workshop program is split into three blocks:

- Channel Allocation and Networking
- Development Techniques and Applications
- Security Techniques

This workshop intends to give a state of the art overview about the existing and ongoing ideas and thoughts towards the interoperability between vehicles and their environment. It aims to originate interesting discussions and know-how transfer between the workshop's participants.

Markus Strassberger, BMW Group Research and Technology
Robert Lasowski, Cirquent GmbH

IOV 2008 Workshop Organization

Workshop Co-organizers
Markus Strassberger, BMW Group Research and Technology
Robert Lasowski, Cirquent GmbH

Technical Program Committee (partial list):
Hannes Hartenstein, University Karlsruhe
Claudia Linnhoff-Popien, University Munich
Irene Karanasiou, National Technical University of Athens
Tim Leinmüller, DENSO Automotive
Andeas Festag, NEC Europe Ltd
Helmut Dohmann, University of Applied Sciences Fulda
Claudia Eckert, Darmstadt University of Technology

Extension of the Rice channel model for deterministic consideration of obstacles in urban traffic scenarios

Steffen Hiebel
IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany
hiebel@ihp-microelectronics.com

Abstract

Vehicle communication in urban areas is influenced by the topology of the road system (topology layer) as well as by current positions of the mobile subscriber nodes (node layer). By overlapping of both layers arise the current forwarding conditions of the urban Vehicular Ad Hoc Network (VANET). Most channel models consider buildings and other obstacles only statistically, a deterministic description and simulation of an urban road system is impossible. Therefore in this paper we will present an extension of the Rice channel propagation model for deterministic consideration of obstacles. We will show that the extended model is an acceptable alternative to the use of a more complex radiation-optical channel model.

1. Introduction

Radio signals suffer from a number of effects while propagating between the sender's and the receiver's antenna. These effects can be classified into reflection, diffraction, absorption and scattering. The received signal is a composition of a *line-of-sight* (LOS) component and all reflected, diffracted, and scattered signals (*non-line-of-sight* components, NLOS), which is summarized under the term *multipath fading*.

This composition is modeled by *radio propagation models*, also referred to as *channel models*. The receiver's signal strength at a certain distance is predicted by a *large-scale propagation model* (for example the propagation models *freespace*, *two-ray ground* and *log-distance shadowing*) and a *small-scale propagation model* (for example the propagation models *log-normal shadowing* and the *Rayleigh or Ricean fading*). Large-scale models predict the average signal strength at a particular location. Small-scale models describe signal fluctuations that arise due to the movement of a fraction of the wavelength (for example between 5 and 6 centimeters for IEEE 802.11a/p). [01]

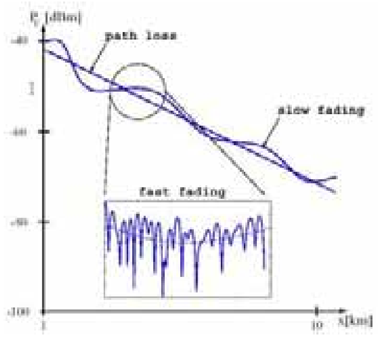


Figure 1. Path loss, slow and fast fading [05]

Figure 1 shows the relation between *path loss* (in decibel [dB]) and *slow fading* (for example because of shadowing of the signal caused by growing transmitter-receiver-distance or mountains) as well as *fast fading* (for example because of *multipath scattering* and *Doppler shift*). The linear *path loss* correlates with the decrease of the signal power in an ideal channel (also referred to as *freespace loss*).

Propagation models can also be categorized as *deterministic* and *probabilistic*. A deterministic propagation model always predicts the same signal strength for a certain distance (for example the *freespace model*), whereas a probabilistic model has a random component leading to a variability of the signal at a certain distance (for example the *shadowing model* and *Rayleigh* respectively *Ricean fading model*).

Most channel models for simulations of wireless communication consider buildings and other obstacles only statistically, a deterministic description and simulation of an urban road system is impossible. Vehicle communication in urban areas is influenced by the topology of the road system (*road system layer*) as well as by the current positions of the mobile subscriber nodes (*node layer*). The current *forwarding conditions* of the urban *Vehicular Ad Hoc Network*

(VANET) arise by overlapping of both layers. That is the reason why we need a channel model with deterministic obstacle consideration for simulations of urban vehicle networks [06].

2. The Rayleigh fading model

Rayleigh fading is a reasonable model when there are many objects in the environment that scatter the radio signal before it arrives at the receiver. The *central limit theorem* holds that, if there is sufficiently more scatter, the channel impulse response will be modelled as a *Gaussian process* irrespective of the distribution of the individual components. [01]

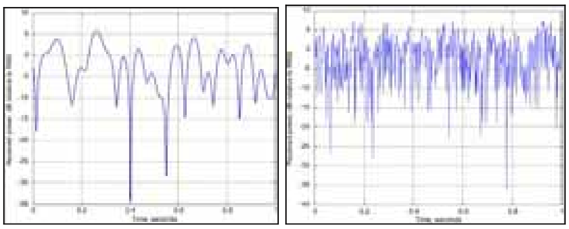


Figure 2. One second of Rayleigh fading with a maximum Doppler shift of 10 Hz (left side) and 100 Hz (right side) [03]

Figure 2 shows the power variation over one second of a constant signal after passing through a single-path *Rayleigh fading channel* with a maximum Doppler shift of 10 Hz (left side) and 100 Hz (right side). These *Doppler shifts* correspond to velocities of about 6 km/h and 60km/h respectively at 1800 MHz, one of the operating frequencies for GSM mobile phones. In *Rayleigh channels* one can observe signal fluctuations of 30 to 40 dB. [03]

The superposition of the many signals at the receiver dues to a fast variation of the receiving amplitude is also referred to as *temporal fading*. The effect of the channel to the transmitted signal *u(t)* can be described as multiple disturbance *c(t)*. The receiving signal *v(t)* therefore can be calculated as the following term [04], [05]:

$$v(t) = u(t) \cdot c(t)$$

The statistical characteristics of the random process *c(t)* have to be known for the mathematical description or modelling of the *Rayleigh channel*. The process *c(t)* is modeled as a zero-mean complex *Gaussian noise process* and has to be defined for a concrete simulation scenario [04]. Because of the only statistical description the process *c(t)* is independent from the geographical (geo-based) definition of the obstacle arrangement (for example buildings) in a specific simulation environment.

The requirement that there may be many scatters present means that *Rayleigh fading* can be a useful model in heavily built-up city centres where there is no line of sight between the transmitter and receiver and many buildings and other objects attenuate, reflect, refract and diffract the signal.

Note that *Rayleigh fading* is a small-scale effect. The model uses the *freespace propagation* or the *two-ray ground propagation* for long-distance (long-scale) prediction and therefore has the same property. [05]

3. The Ricean fading envelope model

The *Ricean fading envelope* extends the *Rayleigh fading model* and models also time-correlated variations of the received signal strength in small-scale. The model uses also the *freespace propagation* or the *two-ray ground propagation* for long-scale prediction. It considers a line-of-sight component superimposed by a number of multipath components (NLOS). Since sender, receiver, and reflecting objects move, the superposition of the multipath signals, as well as the resulting signal, varies over time. This results in intervals of good and poor signals leading to bursts of errors. [01]

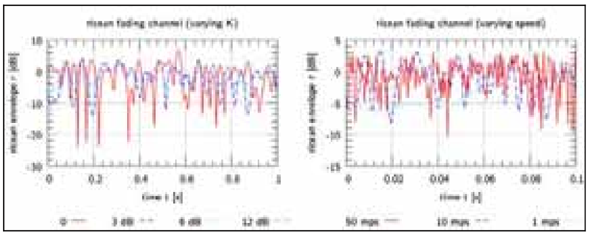


Figure 3. Signal strength over time in a Ricean channel for different values of parameter K (on left side) and different speeds (on right side) [01]

Figure 3 shows the received signal strength over time in a *Ricean channel* for different values of parameter *K* (on left side) and different speeds (on right side). The frequency of variation depends on the object's speed, whereas the amplitude depends on the ratio between the LOS signal and the variance of the multipath. This ratio is also referred to as the *Ricean parameter K*, generally in dB [01], [05]:

$$K = \frac{P_{LOS}}{E[P_{NLOS}]}$$

In the term above P_{LOS} means the signal strength of the *line-of-sight* component, $E[P_{NLOS}]$ means the sum of signal strength of all *non-line-of-sight* components. If there is no *line-of-sight* signal, the *Ricean parameter K* decreases to 0 ($-\infty$ dB) and the *Ricean small-scale fading* degenerates to the *Rayleigh small-scale fading*, which is therefore included in the *Ricean model*. [01]

The *Ricean fading channel* is of high interest to *Vehicular Ad Hoc Network* (VANET) simulations since high speed objects cause rapid environment changes. But the model simulates buildings and other obstacles only statistically, by choice of a suitable value for the *Rice factor K*, which can be obtained by measurement for a specific simulation environment. So especially for the simulation of urban vehicle networks the *Rice channel model* is not adequate (see also [06]).

4. Extension of the Rice model for deterministic consideration of obstacles

4.1. Idea

To take into consideration buildings and other obstacles, for example in an urban road system scenario, in the introduced *Ricean channel model* deterministically, this model has to be extended by a dynamical, geo-based modulation of the *Rice factor K*. A new calculation of *K* has to be done when a mobile node has changed its position.

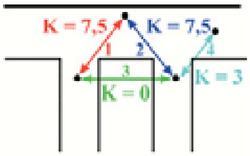


Figure 4. Dynamical, geo-based modulation of the Rice faktor K

The dynamical, geo-based modulation of the *Rice factor K* is a compromise, which makes if possible to simulate scenarios with deterministically defined obstacle arrangements without usage of a more complex radiation-optical channel model. Without such dynamical variation of *K* all the four communication links from figure 4 would underlie the same statistical *Rice channel conditions*.

4.2. Estimation of the Rice factor K for an urban traffic scenario by measurements

If the average K-value by an existing unobstructed *line-of-sight* component for a representativ urban area is known, the *Ricean propagation channel* for this environment can be simulated. If there is no LOS signal the *K*-value goes to zero. In a related scientific thesis [07] measurements for the estimation of *K* were carried out in the urban area of Karlsruhe in Germany.

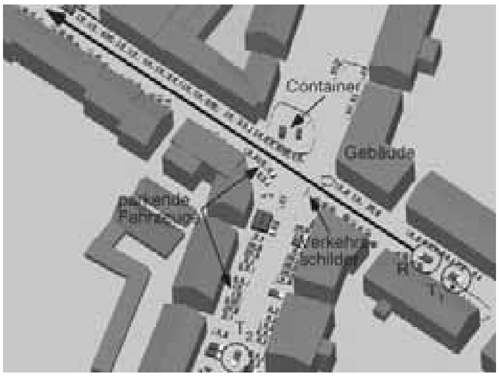


Figure 5. Three-dimensional model of the measuring scenario for the ascertainment of the Rice factor K [07]

Figure 5 shows the three-dimensional model of the measuring scenario, used within the mentioned related scientific thesis [07]. Channel measurements for two different positions of the transmitting vehicle denoted as *T1* and *T2* were carried out. The receiving vehicle *R* moved in the direction of the arrow along the marked line in both measurements. The measuring distance was about 140 meters long. As you can see in figure 5 there was a *line-of-sight* link between transmitter and receiver at position *T1*. After moving the transmitter vehicle to position *T2* the propagation conditions changed during the measurement period and there were time intervals with as well as without a LOS-component. During both measurements the speed of the receiver vehicle *R* was about 10 kilometers per hour; this corresponds to 50 seconds measuring duration. [07]

The *Rice factor K1 = 7,4 dB* (transmitter's position *T1*), measured within [07], describes the statistical *Rice channel conditions* for an urban scenario in a heavily built-up city centre, with a dominating *line-of-sight* signal component and several *non-line-of-sight* components between transmitter and receiver. The *K*-value for position *T2* (*K2 = 6,0 dB*), also measured within [07], is not important for our further work, because we only need one *K*-value (here *K1*) for our *Rice channel model extension*. This *K* has to describe an urban scenario with a dominating LOS signal component.

4.3. Extension of the Ricean channel model

As a simplification, we suggest that only the signal strength of the LOS component changes during gradual shadowing of the *line-of-sight* link between transmitter and receiver; but the summary of the signal strengths of all NLOS-components stay constant. The LOS signal strength P_{LOS} and the value of *Rice factor* K behave to each other directly proportionally. The following terms describe the mathematical circumstances:

$$\frac{P_{LOS}}{E[P_{NLOS}]} = K = 7{,}4dB \quad \text{with } P_{LOS} \sim \text{K},$$

from which results:

$$K_H = 7{,}4 \cdot \frac{P^H_{LOS}}{P_{LOS}} \text{ [dB] with } P_{LOS} = twoRay(P_t, d)$$

The function *TwoRay*(P_t, *d*) describes the received signal strength P_{LOS} for an undisturbed LOS-component without obstacles, after the *two-ray ground propagation model*, for the transmitter-receiver-distance *d*, depending on the strength of transmitted signal P_t. To calculate the obstacle-specific *K*-value K_H the signal strength of the attenuated *line-of-sight* link with obstacles P^H_{LOS} has to be determined. Figure 6 shows the used algorithm graphically.

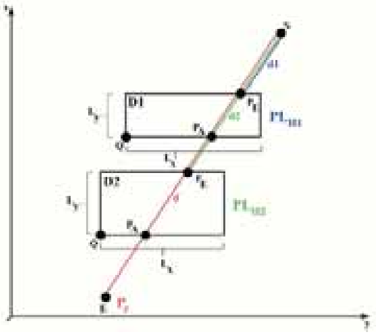


Figure 6. Calc. of the receiver's signal strength of the attenuated LOS signal with obstacles

The path loss PL_{H1} results from the first obstacle (with length L_x and width L_y and the specific obstacle attenuation *D1* in percent per decimeter of obstacle way length d_{H1}) and depends on the resultant obstacle attenuation D_result_1 and can be calculated as:

$$PL_{H1} = twoRay(P_t, d_1) \cdot D\_result_1 \,,$$

$$\text{with } D\_result_1 = \left(\frac{d_{H1}}{0{,}1m}\right)^N \cdot \frac{D_1}{100\%}$$

The length of the signal path by an obstacle d_H corresponds on the distance between the points P_E and P_A. As a function of d_H you can not assume a linear increase of the resulting obstacle attenuation *D_result*. With the support of the freely selectable exponent *N* the increase of *D_result* can be described lineary (*N = 1*) as well as non-linear (*N ≠ 1*), depending on d_H. Then the path loss by the second obstacle can be calculated as:

$$PL_{H2} = \left[twoRay(P_t \cdot d_2) - PL_{H1}\right] \cdot D\_result_2 \,,$$

$$\text{with } D\_result_2 = \left(\frac{d_{H2}}{0{,}1m}\right)^N \cdot \frac{D_2}{100\%}$$

The obstacle attenuated received signal strength of the LOS component at the receiver can be calculated now as:

$$P^H_{LOS} = twoRay(P_t, d) - PL_{H1} - PL_{H2}$$

To calculate the current (obstacle attenuated) value of the *Ricean factor* K_H, the strength of the LOS received signal P_{LOS} between transmitter *S* and receiver *E* without obstacle attenuation has to be ascertained:

$$K_H = 7{,}4 \cdot \frac{P^H_{LOS}}{P_{LOS}} \quad \text{with} \quad P_{LOS} = twoRay(P_t, d)$$

If there are more than two obstacles between transmitter *S* and the receiver *E* the number of the considered partial distances increases. The calculation can be carried out in an adequate implemented while-loop then. As soon as the strength of the obstacle attenuated LOS-signal P^H_{LOS} reaches the value zero the while-loop can be canceled and K_H can be taken as minus infinity (in dB).

4.4. Validation of the Ricean channel obstacle extension

The extended *Rice channel model* with the deterministic obstacle consideration was implemented for the Networksimulator-2 (release version *ns 2.28*). The validation of the model was carried out with several useful simulation scenarios. We will introduce only one of them now. Figure 7 shows a test scenario with seven rectangular obstacles (*H1* to *H7*) with the same specific attenuation *D* and one node pair communicating with each other. Only an unidirectional link from node *n0* to node *n1* was simulated. Node *n1* moves with a speed of 0.5 meters per second parallel to the x-axis, while node *n0* remains still.

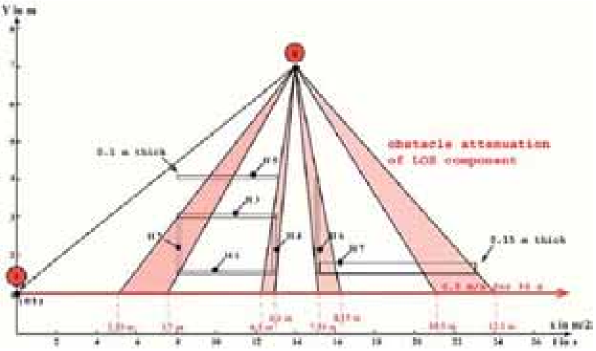


Figure 7. Simulation scenario with rectangular obstacles

Node *n0* sends data packets to node *n1* with a constant data rate of 27 Mbits per second. The specific attenuation *D* of all obstacles in the shown simulation scenario was varied during several simulations, the exponent *N* has the constant value of 1.3. The red triangles in the figure 7 mark way sections on which the LOS-component of the received signal is attenuated particularly strong by the defined obstacles (white rectangles).

Figure 8 and figure 9 show the time-, respectively the distance-dependent trends of the value of the *Ricean factor K* (figure 8) as well as the total received signal strength *Pr_tot* (figure 9) by a specific obstacle attenuation *D* of all simulated obstacles of 10 percent.

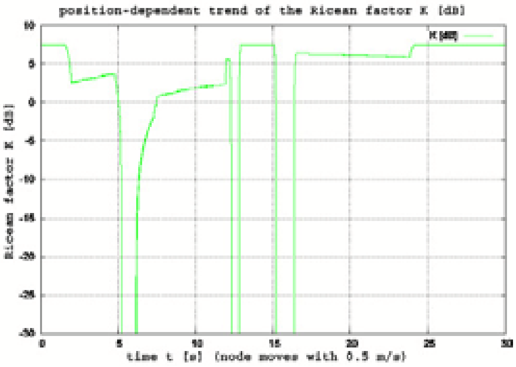


Figure 8. Ricean factor K_H by a specific obstacle attenuation D of 10 percent

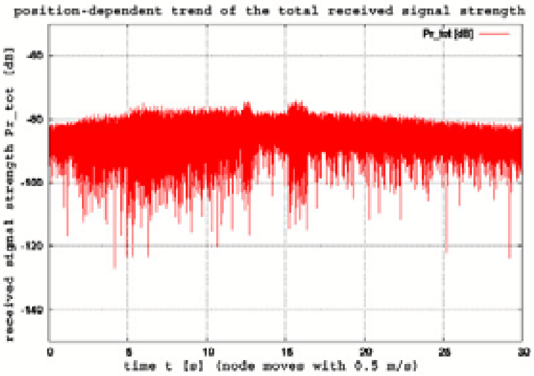


Figure 9. Received signal strength (LOS + NLOS) Pr_tot by a specific obstacle attenuation D of 10 percent

In the same way the strength of the obstacle attenuated LOS signal P_{LOS}^{H} will vary in comparison to two-ray (respectively freespace) attenuated LOS-component P_{LOS}, the actual value of *Rice factor K* changes (figure 8). This dynamical change of *K* affects crucially the actual total power of the receiver's signal *Pr_tot* (figure 9). Within the areas with attenuated LOS-component (remember the red triangles in figure 7) the signal strength amplitude variations, especially downwards, are significantly stronger (small-scall effects).

How an increasing respectively a decreasing transmitter-receiver-distance affects the actual large-scale signal attenuation is shown in figure 9 too. The *two-ray ground propagation model* respectively the *freespace propagation model* (if the actual transmitter-receiver-distance is smaller than the *cross over distance*) were used as large-scale propagation models in all simulations.

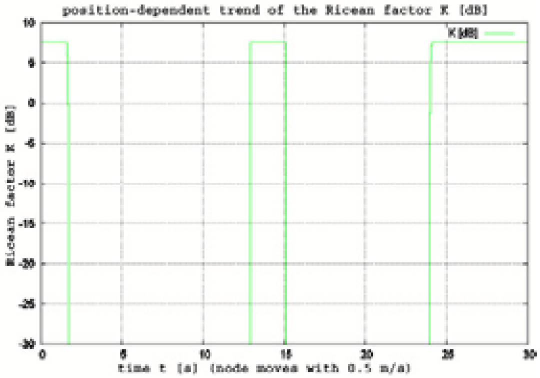


Figure 10. Ricean factor K_H by a specific obstacle attenuation D of 100 percent

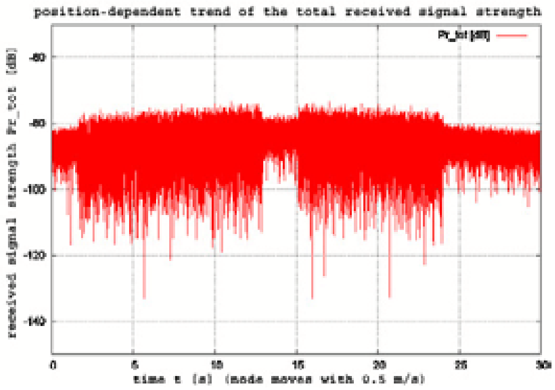


Figure 11. Received signal strength (LOS + NLOS) Pr_tot by a specific obstacle attenuation D of 100 percent

Figure 10 and figure 11 show the time-, respectively the distance-dependent trends of the value of the *Ricean factor K* (figure 10) as well as the total received signal strength *Pr_tot* (figure 11) by a specific obstacle attenuation *D* of all simulated obstacles of 100 percent. Now a LOS signal component the receiver can only receive within those areas where an unobstructed *line-of-sight* path exists. The value of the *Ricean factor K* varies consequently discretely between minus infinity and 7,4 dB (see figure 10). On the basis of the different intensities of the signal strength amplitude variations for the total received signal strength *Pr_tot* on the figure 11 you can differentiate well areas with an existing LOS link from the areas were the LOS-component is attenuated completely.

5. Summary and Outlook

The dynamical, geo-based modulation of the *Rice factor K* is a compromise, which makes it possible to simulate scenarios with deterministically defined obstacle arrangement without usage of a more complex radiation-optical channel model.

A further optimisation of the extended *Rice model* is possible when the obstacles in the simulation scenarios are also more realistic considered by the used *large-scale propagation model* (here, two-ray respectively freespace attenuation).

6. References

[01] Enrico Minack, “Evaluation of the influence of channel conditions on Car2X Communication”, Diploma Thesis, Chemnitz University of Technologie, 2005

[02] LAN/MAN Standards Committee of the IEEE Computer Society, \Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Wireless Access in Vehicular Environments (WAVE), Draft IEEE P802.11p/D1.0, February 2006

[03] Rayleigh fading, from Wikipedia, the free encyclopedia, available: http://en.wikipedia.org/wiki/Rayleigh_fading, last access: 11.2007

[04] Dr.-Ing. habil. H. Nuszkowski, Script of lecture “Mobile Information Systems“, Vodafone Chair for Mobile Information Systems, Technical University Dresden, 2005

[05] Theodore S. Rappaport, Wireless Communications: Principles and Practice, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2002

[06] Steffen Hiebel, „ Concept and evaluation of an efficient geo-based forwarding mechanism for vehicle communication within an urban network”, IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany, 2007

[07] Jürgen Maurer, “Strahlenoptisches Kanalmodell für die Fahrzeug-Fahrzeug-Funkkommunikation”, Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Fridericiana Karlsruhe, 05.2005

[08] LAN/MAN Standards Committee of the IEEE Computer Society, \Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 5 GHz Band, IEEE Std 802.11a-1999, 1999

[09] LAN/MAN Standards Committee of the IEEE Computer Society, \Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Wireless Access in Vehicular Environments (WAVE), Draft IEEE P802.11p/D1.0, February 2006

[10] Committee SCC32 of the IEEE Intelligent Transportation Systems Council, IEEE P1609.4™, Standard for Wireless Access in Vehicular Environments (WAVE) - Multi-Channel Operation, 2005

[11] CMU-extension of the network simulator NS-2 for the consideration of Ricean- and Rayleigh-Fading, available: http://www.ece.cmu.edu/wireless/downloads.html, last access: 09.2006

Analysis of Approaches for Channel Allocation in Car-to-Car Communication

Long Le[1], Wenhui Zhang[1], Andreas Festag[2], and Roberto Baldessari[1]
[1]NEC Laboratories Europe
[2]NEC Deutschland GmbH

Abstract

Car-to-car and car-to-infrastructure communication (C2X communication) has received considerable attention worldwide. The main goals of C2X communication are reduction of road accidents and fatalities and improvement of traffic efficiency. In Europe, the European Telecommunications Standards Institute (ETSI) is currently leading the spectrum allocation dedicated to road safety. It is expected that 30 MHz spectrum in the range from 5.875 to 5.905 GHz will be allocated for critical road safety and traffic efficiency applications. This paper provides a detailed analysis of channel allocation for the 30-MHz spectrum dedicated to safety-related C2X communication. Our contributions are:

(1) We provide a comprehensive overview of existing approaches on the usage of the 30-MHz frequency band dedicated for safety-related C2X communication.

(2) We analyze advantages and disadvantages of these approaches based on an extensive set of evaluation criteria.

(3) We provide a recommendation for the channel allocation of the 30-MHz frequency band dedicated for safety-related C2X communication in Europe.

1 Introduction and Motivation

Recently, C2X communication has received considerable attention in both academia and industry because it has the potential to improve road safety and to reduce road accidents and fatalities. For this purpose, the mature, inexpensive, and widely available IEEE 802.11 technology appears very attractive. In C2X communication, cars are equipped with IEEE 802.11-based wireless network interfaces and can spontaneously form an ad hoc network among themselves. Cars can use the ad hoc network to communicate with each other in order to support safety applications such as cooperative collision warning. This allows drivers to receive emergency warnings from the C2X communication system and reduce speed before they can actually see an accident or the brake light of the cars in front. Further, road side units (RSUs) equipped with sensors can also communicate with cars via the ad hoc network to provide warnings about road conditions or speed limit.

C2X communication is considered as an important part of future *Intelligent Transportation Systems* (ITS). An overview of the C2X communication system is depicted in Fig. 1. Beside enabling safety applications, the C2X communication system provides non-safety applications such as infotainment applications and Internet access. These will have lower priority than safety applications.

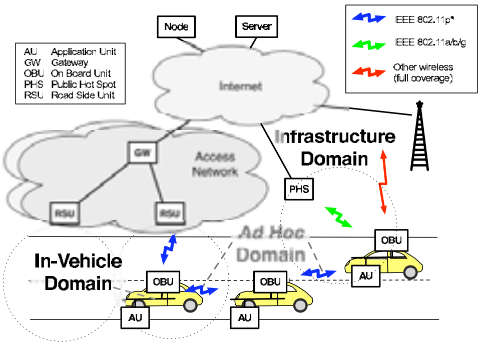


Figure 1. C2X communication scenario.

Since C2X communication's major goal is to support critical road safety applications, it is desirable that C2X communication experience as little interference from other wireless applications on the wireless medium as possible. For this reason, there are ongoing discussions that a spectrum allocation will be used as protected bandwidth for C2X communication (and ITS in general).

In Europe, it is expected that the frequency bands 5.855-5.875 and 5.875-5.925 GHz will be used for ITS non-safety and safety applications. Further, the frequency band 5.875-5.925 GHz will be divided into two parts 5.875-5.905 GHz and 5.905-5.925 GHz in an initial and a later deployment phase [5]. An overview of the expected spectrum allocation for ITS applications in Europe is illustrated in Fig. 2.

A number of proposals have been made for the usage of the 30-MHz frequency band dedicated for road safety and traffic efficiency. This paper provides an analysis of the channel allocation for the 30-MHz frequency band dedicated to safety-related C2X communication. Our contributions are:

(1) We provide a comprehensive overview of existing approaches for the usage of the 30-MHz frequency band dedicated for safety-related C2X communication.

(2) Based on a set of evaluation criteria for multi channel operation [7], we perform a detailed analysis of the existing proposals for channel allocation.

(3) We provide a recommendation for the channel usage of the 30-MHz frequency band dedicated for safety-related C2X communication. Our recommendation is intended as input for further discussions in different standardization bodies such as ETSI, ISO, and Car 2 Car Communication Consortium (C2C-CC).

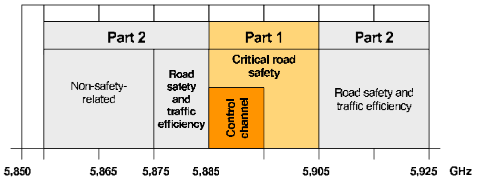


Figure 2. Anticipated spectrum allocation for ITS applications in Europe.

The rest of this paper is organized as follows. Section 2 reviews the background and requirements for C2X communication. Section 3 discusses related work. Section 4 presents our analysis. Section 5 concludes the paper.

2 Background and Requirements

In this section, we review the basic protocol operations and the requirements for C2X communication. This section covers the background in C2X communication before we present our analysis in section 4.

2.1 Background for C2X Communication

In vehicular ad hoc networks (VANETs) vehicles support safety applications by broadcasting and processing of two types of messages: periodic and event-driven safety messages [2]. These safety messages typically need to be delivered within a geographical area with certain reliability and delay limit. The periodic messages, also called beacons, carry vehicles' status information such as positions and speeds. Beacons can be generated at the application layer or at the network layer, and are used by neighbouring vehicles to become aware of their surrounding and to avoid potential dangers. Event-driven safety messages are generated when an abnormal condition or an imminent danger is detected, and disseminated within a certain area with high priority. Critical event-driven messages usually have strong reliability and delay requirements.

It is well known that vehicular communication environments are characterized by highly mobile vehicles, extremely frequent topology changes and a great variation in the number of vehicles in a certain region. To meet the specific requirements of V2X communications in such environments, geographical routing is applied. Geographical routing assumes that vehicles acquire information about their own positions (i.e. geodetic coordinates) via GPS or other positioning systems. If a vehicle intends to send data to a known target geographic location, it chooses another vehicle as message relay, which is located in the direction towards the target position. The same procedure is executed by every vehicle on a multi-hop path until the destination is reached. Results from extensive network simulations and measurements have indicated that geographical routing has good performance in realistic environments.

2.2 Requirements

Recently, a set of criteria for evaluating multi channel operation in C2C-CC has been proposed [7]. We review these criteria in this section and use them to evaluate different approaches for channel allocation in Section 4.

Usability: This criterion represents the main requirements for safety-related C2X communication: low latency and high reliability for critical safety messages. In Section 4, we will focus on latency since network and/or applications should be responsible for reliability.

Robustness: This criterion evaluates the wireless link's robustness in two aspects: (1) it has to be robust in terms of bit errors (e.g. the bit error rate should be as low as possible) and (2) it has to be robust in terms of interference.

Cost: This criterion considers the material costs for mass production and deployment. Obviously, an inexpensive solution is preferred in order reduce the market barrier.

Efficiency: This criterion evaluates the effectiveness of channel allocation in terms of bandwidth usage. Given the scarcity of available bandwidth allocated for C2X communication, this precious resource must be used effectively.

Scalability: This criterion evaluates the impact of channel allocation on the flexibility of the overall C2X communication system in different scenarios such as highways, cities, and rural areas.

Development effort: This criterion considers development costs apart from material costs. A solution for channel allocation that allows a simple design and implementation of the C2X communication system is clearly preferred.

3 Related Work

3.1 Spectrum Allocation and Measurement Reports in North America

In North America, a 75-MHz frequency band in the 5.9 GHz range is reserved for ITS. This spectrum may be allocated to seven 10-MHz bandwidth channels. One of the 10-MHz channels will be designated as a control channel (CCH) that is used to transmit critical safety applications and beacons. Other channels are used for other purposes such as traffic efficiency and infotainment and are called service channels (SCHs). It is also possible to obtain a 20-MHz SCH by combining two 10-MHz SCHs. Since wireless devices are usually incapable of simultaneously monitoring and exchanging data on different channels, the standard IEEE 1609 for Wireless Access in Vehicular Environments (WAVE) [3] suggests that wireless devices operate in the control channel during a periodic common interval.

Recent measurements evaluated the robustness of 802.11p's channel width (5, 10 and 20 MHz) against BER when operated in different scenarios: suburban, highway, and rural environments [10]. Measurements were conducted without channel interference. Analysis and measurements showed that 802.11p's guard interval is not long enough in a 20-MHz channel while errors increase from lack of channel stationarity over the packet duration in a 5-MHz channel. This study concluded that a 10-MHz channel is the best choice in terms of robustness against BER.

Regarding channel interference, a recent measurement study [9] reported that interference between adjacent channels leads to substantial packet error rates while interference between non-adjacent channels is much less of an issue, although still measurable in some environments. These results indicate that interference is a serious issue for deployment models employing adjacent channels.

3.2 Multi Channel Operation in NoW

In project Network on Wheels (NoW) [8], the task force *Multi Channel Operation* has investigated the channel allocation problem on how to use 2x10-MHz channels [1]. We observe that multi channel operation covers problems related to both the physical and network layer. Several possible channel usage scenarios are analyzed, namely, WAVE compliant mode, symmetric channels, priority and traffic channel, and combined channel mode.

WAVE compliant mode: this follows the WAVE standard [3] in the U.S. where channel switching is performed between a single CCH and multiple SCHs.

Symmetric channel layout: this considers the two channels as identical, and each transmitter decides which channel it uses or it may use the channels arbitrarily.

Priority and traffic channel: this uses one channel exclusively for high priority safety messages and the other channel for all non-priority safety messages.

Combined channel mode: this combines two 10 MHz channels to a single 20 MHz channel.

It is argued that the support of high priority low latency messages is vital for critical safety applications. Thus, the Priority and Traffic channel usage scheme is chosen for multi channel operation.

3.3 C2C-CC Channel Allocation Proposal

Within C2C-CC, there have been intensive discussions on the usage of 30-MHz frequency band. There are a number of proposals. Generally saying, three basic approaches have been discussed: (1) Single 30-MHz channel, (2) one 20- and one 10-MHz channel, and (3) three 10-MHz channels. However, the single 30-MHz channel approach requires new hardware that supports 30-MHz bandwidth. It is envisaged that considerable amount of development efforts are needed. Thus, it is not a preferred solution.

Table 1. Spectrum mask (in dBc)

MHz	4.5	5.0	5.5	10	15
Class A	0	-10	-20	-28	-40
Class B	0	-16	-20	-28	-40
Class C	0	-26	-32	-40	-50

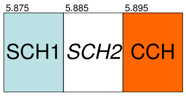


Figure 3. Proposed channel allocation for 30-MHz frequency band in C2C-CC

Recent discussions in C2C-CC tend to propose the following channel usage (Fig. 3): a service channel (SCH1) in the frequency band 5.875 - 5.885 GHz will be used for low-priority safety messages and traffic efficiency applications, another service channel (SCH2) in the frequency band 5.885 - 5.895 GHz will be used for transmission in small distances and with low transmit power to minimize the interference to SCH1 and CCH, and a control channel in the frequency band 5.895 - 5.905 GHz will be used for high-priority safety messages and beacons.

3.4 Spectrum Mask and Link Budget

For better understanding of channel interference, we will review the relevant specification of IEEE 802.11p and methods for calculating link budget in vehicular environments.

ETSI has specified the emission limits of different classes of IEEE 802.11p radio equipment. The maximum radiated power for Class A, B and C is 10, 20 and 33 dBm, respectively [5]. The transmit spectrum masks for these classes follow IEEE 802.11p [6] and are listed in Table 1. The minimum receiver sensitivity in a 10-MHz channel is also listed, e.g. -85dBm for data rate of 3 Mbits/s and $-72dBm$ for data rate of 6 Mbits/s.

ETSI also specifies method for link budget calculation. Without considering the gain of antenna, the link budget is calculated in dBm as

$$P_e = P_s + L_0 + L_l, \tag{1}$$

where Pe is the received power in dBm, P_s is the transmit power in dBm, L_0 is the path loss in dB up to the breakpoint distance d_0, L_l is given by

$$L_l = -10log(d/d_0)^n, \qquad (2)$$

where d is the distance between a transmitter and a receiver, n is the path loss factor, which is typically 2.7 for vehicular environments [5]. In the 5.9-GHz band, the breakpoint distance is given as $d_0 = 15m$ and accordingly $L_0 = -71dBm.$

4 Analysis of Different Approaches

4.1 Overview of Existing Approaches

Considering the basic channel usage scenarios from C2C-CC, we can eliminate some scenarios based on qualitative observations. First, since the 30-MHz channel scenario requires new hardware and would incur significant development cost, it must be ruled out. Second, the measurement study for channel interference [9] indicates that the simultaneous usage of two adjacent channels causes significant packet loss and hence is unacceptable for ITS safety applications. Although measurements have been conducted only for 10-MHz channels, we expect that simultaneous usage of 20- and 10 MHz channels will have similar effects and should also be avoided. Further, since measurement results [10] show that 20-MHz channels are more susceptible to BER than 10-MHz channels, usage of 20-MHz channels will not be further considered in our analysis.

While the usage of adjacent channels is possible, certain mechanisms such as WAVE channel switching have to be in place to prevent simultaneous transmissions on these channels. Although interference also occurs between non-adjacent channels, it is much lower than in the case of adjacent channels. We believe that this smaller interference acceptable as packet losses caused by interference between non-adjacent channels can be recovered by reliability mechanisms at the network layer and/or applications.

From recent measurement results [9, 10] presented in Section 3 and our reasoning above, we analyze and compare two channel usage schemes in the rest of our paper: SCH1 + SCH2 + CCH (Scheme A) and CCH + 2 * SCH with WAVE channel switching (Scheme B). Scheme A needs two transceivers while scheme B requires one transceiver performing WAVE channel switching.

4.2 Usability

4.2.1 Latency

Scheme A: Both SCH1 and CCH experience the media access latency according to 802.11 (denoted as T_m). SCH2 has lower priority and will also be subject to adjacent channel interference from both SCH1 and CCH. A node using SCH2 may sense channel as busy if other nodes are transmitting on SCH1 or CCH. The channel access latency of SCH2 depends on activities on SCH1 or CCH, and its minimum value is T_m. The latency for CCH and SCH1 is

$$T_A = T_m. \qquad (3)$$

Scheme B: CCH and each SCH have the switching cycle of 100ms and are only active every $T_c = 50ms$. A message may arrive when CCH is either active or inactive. We assume that the arrival time of a message is uniformly distributed within a switching cycle. Thus, the probability that a message finds CCH to be active or inactive will be the same, and the average time between the arrival of a message and the end of an active or inactive period is $0.5T_c$.

When CCH is active, the message will not experience channel switching latency. Here we assume that the channel load is low and the message can be transmitted in a switching cycle. Otherwise, the message will experience a switching delay (denoted as T_s) and the media access latency T_m depending on the channel load. When CCH is inactive, the message will experience an additional average waiting time of $0.5T_c$. Since the probability of these two cases are equal, the average waiting time before CCH becomes active is $0.25T_c$. Therefore, we get the total average channel access latency for CCH is

$$T_{BC} = 0.25T_c + T_s + T_m. \qquad (4)$$

For latency on SCH, we note that each SCH will be active after its service announcement on CCH. We assume that a SCH is activated immediately after a CCH's active period. Applying the same method in calculating Eq. (4), we get the total average channel access latency for SCH

$$T_{BS} = T_c + T_s + T_m. \qquad (5)$$

For latency, scheme A is preferred over scheme B.

4.2.2 Prioritization of Different Message Types

Since both scheme A and B can support prioritization of different message types by using CCH and SCHs for different priorities, there is no preference between the schemes.

4.3 Robustness

4.3.1 Channel interference

Scheme A: Here we present a theoretical analysis of channel interference similar to [1]. We assume that a message will be correctly received if the received power from the transmitter is over 10dB higher than that of the interfering node. Based on the link budget calculation model presented in (1), if the transmit power of the transmitter on one channel is 20dBm, the received power will be -85dBm at the distance of $276m$. This means, the transmitter has a communication range of $276m$, in which the data rate of 3Mbits/s is supported. Depending on the distance between the transmitter and the interference node, there may be a jammed area within the communication range of the transmitter, where the reception from the transmitter will fail due to the interference node as shown in Figure 4.

Figure 4(a) shows the jammed area caused by the interference node in an adjacent channel with centre frequency offset 10MHz. The light-grey areas are the communication range of the transmitter without interference, the grey areas indicate the jammed area of class A/B equipment (with spectrum mask -28dBc) and the black areas indicate the jammed area of class C equipment (with spectrum mask -40dBc). We may observe that the jammed area is relatively large for class A/B equipment, especially when the distance between the transmitter and the interference node is large. Because class A/B equipment from non-adjacent channels with channel spacing 20MHz also has the same spectrum mask of -40dBc, its jammed areas are the same as the jammed areas in black.

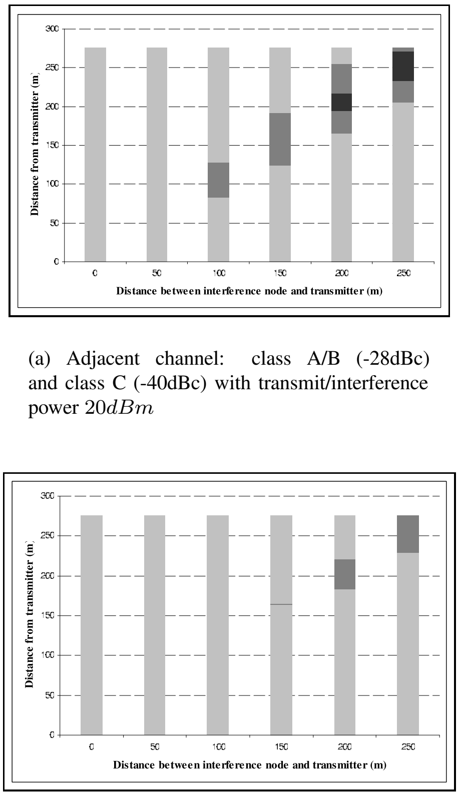
Figure 4(b) shows the jammed areas caused by the interference node with transmit power 10dBm in an adjacent channel. The light-grey areas are the communication range of the transmitter without interference, the grey areas indicate the jammed area of class A/B equipment (with spectrum mask -28dBc). Further analysis shows that in case the transmit power is 5dBm in an adjacent channel, no jammed area will be observed.

Comparing the theoretical analysis with field measurements [9], we have several observations. We find that the measurement of the non-adjacent channel interference is slightly different from theoretical analysis, but all indicating that non-adjacent channel interference is rather small if the frequency offset is over 20 MHz. However, theoretical analysis shows that there is no jammed area between adjacent channels if the distance between the transmitter and the receiver is below $50m$ (Figure 4(a)). But field measurements show that packet loss probability is considerably high if the distance between the interference node and the receiver is below $10m$. One possibility to cope adjacent channel interference is to reduce interference power. Theoretical analysis shows that in case the transmit power in adjacent channels is 15dB less, no jammed area will be observed, but no similar measurement data are available.

For scheme A, we conclude that: (1) There is very limited interference between the CCH and SCH1 if an interference node is over 2m away from a receiver. (2) Theoretical analysis shows that if the transmit power in SCH2 is 15dB lower than that of the CCH and SCH1, the CCH and SCH1 will be free from interference. It is still necessary to measure adjacent channel interference with relatively small interference power in order to use SCH2 with other channels simultaneously. (3) SCH2 will suffer severe interference from the CCH and SCH1.

Scheme B: If the CCH is located between two SCHs, there will be no adjacent channel interference since WAVE does not allow simultaneous transmissions in the CCH and a SCH. However, there exists non-adjacent channel interference similar to the case shown in Figure 4(a).

Here scheme B slightly outperforms scheme A.



(a) Adjacent channel: class A/B (-28dBc) and class C (-40dBc) with transmit/interference power $20dBm$

(b) Adjacent channel: class A/B (-28dBc) with transmit/interference power 20dBm/10dBm

Figure 4. Adjacent channel interference

4.3.2 Reliability

In Scheme B, all nodes need to synchronise with a reference time system such as UTC. Unsynchronized nodes or synchronization inaccuracy could render WAVE channel switching inoperable. Since scheme A has no requirement on synchronization, it is preferred over scheme B.

4.4 Hardware and Development Effort

Both scheme A and B will be based on standard IEEE 802.11p hardware and are equal regarding availability. However, scheme A needs two wireless network interfaces where scheme B requires only one. On the other hand, scheme A can use off-the-shelf IEEE 802.11 drivers whereas scheme B requires an implementation of WAVE synchronized channel switching. Here it is likely that scheme B is slightly preferred over scheme A. However, depending on deployment scenarios, the hardware cost can outweigh the development effort or vice versa.

4.5 Efficiency

Bandwidth usage efficiency. In order to compare the bandwidth usage efficiency we define the following metrics.

$$\text{E} \triangleq \sum \text{bandwidth} \times \text{percentage of active time} \quad (6)$$

For Scheme A, both the CCH and SCH1 can be active the same time. In this case, SCH2 will suffer from interference. Thus, the efficiency for Scheme A is

$$E_A = 2 * 10MHz * 100\% = 10MHz * 200\% \quad (7)$$

For Scheme B, every channel can only be active half of the same. Due to channel switching, not all the active time can be used for transmission. Assume the channel switching time is $x\%$, the efficiency for Scheme B is

$$E_B = 3 * 10MHz * (50\% - x\%) < 10MHz * 150\% \quad (8)$$

Here scheme A is preferred over scheme B.

4.6 Scalability

4.6.1 Node Density

With high node density, the amount of data traffic has to be controlled. This requires proper congestion control mechanisms, such as reducing packet size, packet generation rate and transmit power. Since available congestion control mechanisms appear to be applicable for both schemes, there is no preference for neither scheme A nor scheme B.

4.6.2 Additional Frequency Band

Addition frequency band, especially the additional 20-MHz bandwidth between 5.905GHz and 5.925GHz will have different implications on the allocation schemes. With scheme A, the lower part of the additional bandwidth may be used as a SCH with low transmit power similar to SCH2, and the higher part may be used as a SCH with relative higher power similar to SCH1. This will add the channel usage efficiency by $10MHz * 100\%$. With scheme B, another two SCHs with 10-MHz bandwidth will be available. However, the two new channels, along with the SCH between 5.895GHz and 5.905GHz, cannot be used simultaneously due to high adjacent channel interference. One possible approach to reduce adjacent channel interference between these three channels is to reduce the active time of the channels, e.g. to limit the activity on the SCH between 5.905GHz and 5.915GHz. This will only add channel usage efficiency by $10MHz * 50\%$, which is much less than that of scheme A. Here scheme A is preferred over scheme B.

5 Summary and Conclusion

We present a comprehensive overview of channel allocations for the spectrum allocated to C2X communication in Europe. We use an extensive set of evaluation criteria for channel allocation and present an analysis of some proposals. In particular, we consider and compare two channel usage schemes: SCH1 + SCH2 + CCH (Scheme A) and CCH + 2 * SCH with WAVE channel switching (Scheme B). Scheme A requires two transceivers and uses low transmit power on SCH2. Scheme B needs one transceiver performing WAVE channel switching. Overall, the advantages of scheme A outweigh those of scheme B. We recommend scheme A for C2X communication in Europe.

An issue that is not considered in this paper is the interference caused by transmissions in the 30-MHz frequency band dedicated to safety-related C2X communication to other frequency bands (below 5855 MHz and above 5925 MHz). This issue was investigated in a related study [4]. When results of this study is taken into consideration, one possible solution is to swap CCH and SCH1 (as shown in Figure 3). The main results of our analysis still hold.

References

[1] A. Brakemeier. Network on Wheels Project Report. Task Force: Multi Channel Operation. Technical report, July 2005.

[2] Car-to-Car Communication Consortium. C2C-CC Manifesto. Version 1.0, July 2007. Available at http://www.car-to-car.org/fileadmin/dokumente/pdf/C2C-CC_manifesto_v1.0%_2007-05-17.pdf.

[3] Committee SCC32 of the IEEE Intelligent Transportation Systems Council. Draft Standard for Wireless Access in Vehicular Environments – WAVE Resource Manager. IEEE P1609, November 2005.

[4] Electronic Communication Committee (ECC) within the European Conference of Postal and Telecommunications Administrations (CEPT). Compatibility studies in the band 5855-5925 MHz between intelligent transport systems (ITS) and other systems, February 2007.

[5] ETSI Technical Committee Electromagnetic compatibility and Radio spectrum Matters (ERM). Technical Characteristics for Pan-European Harmonized Communication Equipment Operating in the 5 GHz Frequency Range and Intended for Critical Road-Safety Applications. Technical Report ETSI TR 102 492-1/2, ETSI, 2005.

[6] IEEE. Draft Standard for Information Technology - Telecommunications and information exchange between systems - LAN/MAN Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 7: Wireless Access in Vehicular Environments (WAVE), July 2007.

[7] A. Lübke. Evaluation Criteria for Multi Channel Operation. Technical report, Car 2 Car Communication Consortium, November 2007.

[8] Network on Wheels. http://www.network-on-wheels.de.

[9] V. Rai, F. Bai, J. Kenney, and K. Laberteaux. Cross-Channel Interference Test Results: A report from the VSC-A project. Technical report, IEEE 802.11 technical report, July 2007.

[10] D. Stancil, L. Cheng, B. Henty, and F. Bai. Performance of 802.11p Waveforms over the Vehicle-to-Vehicle Channel at 5.9 GHz . Technical report, IEEE 802.11, September 2007.

Dual Receiver Communication System for DSRC

Tsutomu Tsuboi[1], Jun Yamada[2], Naoki Yamauchi[2], Masato Hayashi[1]

[1] New Business Development Center, Renesas Technology Corp.,
[2]Automotive Semiconductor Business Unit, Renesas Technology Corp.,
2-6-2, Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

Abstract

This is the first study and field test analysis for DSRC (Dedicated Short Range Communication) vehicle wireless communication system design with WAVE (Wireless Access in Vehicle Environment) technology under IEEE802.11p and IEEE1609 standard for evaluating basic functions of WAVE. In the field test, packet reach ability has been measured for both WAVE announce frame and IP (Internet Protocol) packet data and makes sure about WAVE announce frame communication between On-Board-Unit (OBU) and Road-Side-Unit (RSU). After basic WAVE functional test, it has been considered about safety application with new DSRC technology and it is also evaluated in the field testing. Then dual receiver concept has been examined in order to support "Dual receiver concept" required by Car to Car Communication Consortium (C2C-CC) and been evaluated advantage of dual receiver concept.

At first, WAVE system platform has been developed and evaluated under current existing IEEE802.11p specification which supports Control Chanel (CCH) and Service Channel (SCH) with alternative 50msec interval. Then it has been measured basic WAVE system functions result of WAVE announcement frame receiving and traffic sign receiving after communication link establishment with GPS (Global Positioning System) synchronization as one of example application.

Then it has been compared between single receiver and dual receiver architecture and figured out advantage of dual receiver concept in terms of safety application by new DSRC (Dedicated Short distance Radio System) system.

Then lastly dual receiver concept of WAVE system is proposed to C2C-CC and ETSI (European Telecommunications Standards Institute) standardization which has been started since 2007 December.

I . INTRODUCTION

For the next ITS (Intelligent Transport System) communications, there are two types of communication necessary. In terms of real time application such as collision avoidance, traffic warning and traffic jam information, WAVE becomes more important between car to car and also car to infrastructure communication. And on the other hand, broadband wireless communication system is also important for informing traffic road work and map update etc. The broadband wireless access is not necessary to be real time such as DSRC because all information like traffic road work and map update is not so time critical information. There are broadband wireless access systems in current network such as mobile telephony (3G, GSM: Global System for Mobile communications, CDMA: Code Division Multiple Access etc.) and WiMAX in future. (Refer to Figure 1. Vehicle wireless communication)

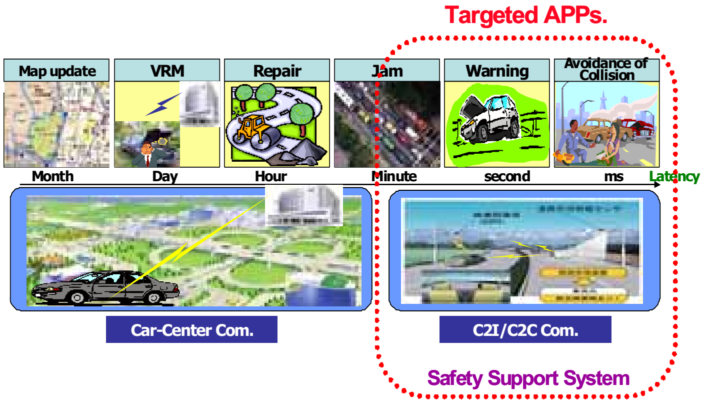


Figure 1 Vehicle wireless communication

It is focused on DSRC technology in this paper and how it works. And even in DSRC system, there are also several systems available in the world such as WiFi (current Wireless Local Area Network IEEE802.11b/g), Japanese DSRC which has been used ETC (Electric Toll Collection) system with 5.8GHz band and or 700MHz new frequency allocation for next TELEMATIC system. which is under discussion. But IEEE802.11p and IEEE1609 (so called WAVE) are most promising technology in DSRC in world-wide. In terms of real time, WAVE has very short link establishment time for synchronization among vehicles (OBU) and infrastructure base stations (RSU). It is described about this link establishment timing in the field evaluation later section.

The WAVE specification consists of IEEE802.11p and IEEE1609. The IEEE802.11p is now draft 3.0 stage and basic key system specification is based on draft 1.0 which defines MAC (Media Access Control layer) and PHY (Physical layer) for wireless communication. The IEEE1609 is upper MAC layer and Network pulse Session layer. The main IEEE1609 specification defines networking communication terminal allocation and dedicated safety critical message information under WSMP (Wave Short Message Protocol). In WAVE communication, there are two basic channels defined, which are CCH and SCH. According to IEEE802.11p, CCH and SCH are switching each 50msec time interval and uses only single channel at once. The CCH is mainly used for Emergency and Data control information and the SCH is open for users to information any type of data like UDP (User Datagram Protocol) in internet application. The unique technology concept in this paper is "dual receiver" concept for WAVE system and explains the importance of this concept compared with single receiver concept based on safety application as real time application. Currently it has been discussed under car manufacturers' consortium such as C2C-CC. In order to showing and evaluating such a real time application, WAVE system has been actually developed in our laboratory made real field system evaluation.

This paper provides important information for safety application system with WAVE system.

Ⅱ. WAVE specification

The WAVE system specification consist of IEEE802.11p and IEEE1609 as mentioned before. In Figure 2, it shows about IEEE802.11p and IEEE1609 and relation about those specifications. IEEE802.11p is used 5.9GHz frequency band (5.85~5.925 GHz) .For CCH and SCH, each channel has 10MHz bandwidth and total bandwidth is 20MHz but it is also under consideration more bandwidth for safety application because total IEEE802.11p has 75MHz between 5.85GHz and 5.925 GHz assignment. In IEEE1609, there are two basic systems. One is WSMP is defined dedicated WAVE system and it is non-IP data (Internet Protocol). And the other is UDP (User Dataram Protocol) and it is IP data, which means that it is able to communicate with Internet Protocol network and send any type of data packet to IP network.

Here is interesting comparison between IEEE802.11p and current existing WiFi specification such as IEEE802.11a/b/g$^{1)}$. As for frequency assignment, IEEE802.11a is 5.8GHz (5.725~5.835GHz) which is closest frequency with IEEE802.11p band assignment and IEEE802.11b and IEEE802.11g are ISM band (2.4GHz). The all IEEE802.11a/b/g are same link establishment procedures such as first scanning of free channel for communication, then parameter setting for the channel. And after authentication between terminal and base station, then association has been established between them. This procedure normally takes more than 3 sec and sometimes worst case takes 10 sec for example. On the other hand, IEEE802.11p uses GPS (Global Positioning System) signal as synchronization for all OBUs and RSUs. Therefore WAVE is capable to have short link establishment compared with WiFi. The GPS synchronization in WAVE system takes normally few ten msec.

The two unique features like IP network communication and short time synchronization among terminals and base stations are good fit for vehicle wireless communication system, especially DSRC system. Therefore WAV (IEEE802.11p and IEEE1609) is one of best choice for safety application based on this features at this present.

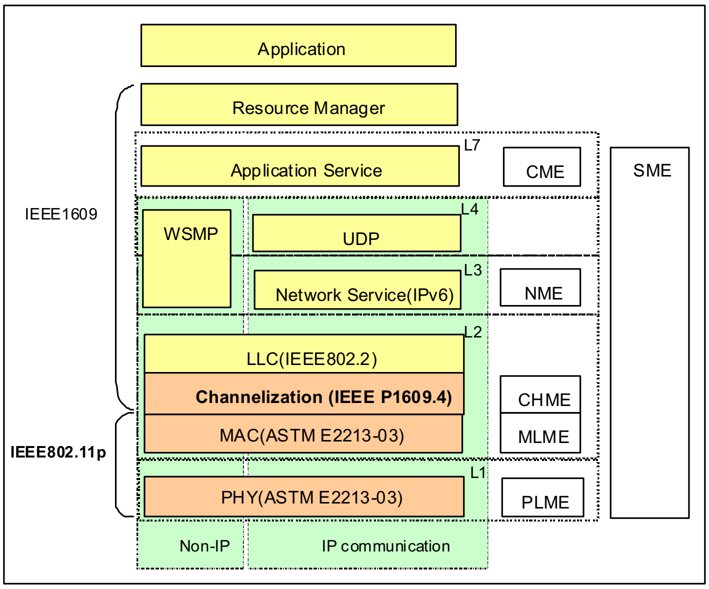


Figure.2 WAVE specification

Ⅲ. Field System Evaluation

The next step is system evaluation. In order to system evaluation, the WAVE prototype system has been developed. The first generation of WAVE system is shown in Figure 3 and brief specification is shown in table 1.

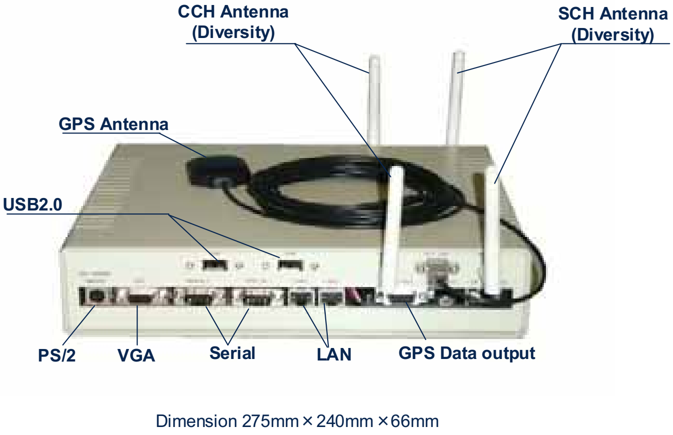


Figure. 3 WAVE system prototypes

Table 1 WAVE system specification

Item		Content
Communication method	Support standard	supporting 1609.3 D22, 1609.4-2006 , IEEE802.11p D1.0 (Part is limited and exists.)
	Frequency	5.860GHz～5.920GHz／10MHz width, 5.875GHz／20MHz width, 5.905GHz／20MHz width
	Feature	An independent wireless module is installed for CCH/SCH. No time lag by CCH/SCH switch. The service channel supports QoS. (IEEE802.11e equivalence). Time synchronization that uses GPS.
OS		Linux2.6.12
Hard Ware	Wireless LAN	Equipped with two wireless modules
	Antenna composition	Two for CCH. Two for SCH. Each space diversity
	CPU	Intel Celeron M Processor 1.3GHz
	RAM	DDR-SDRAM 256Mbyte
	Ethernet	Two ports
	GPS	The GPS reception module is built into
	Power supply	AC90-264V 47-63Hz (AC adaptor attachment) DC+12 V
	Dimension	Main body:275×240×66mm
	Weight	Main body:2.7kg(AC adaptor: 0.3kg)
Interface	PS/2	AT/8042 conforming (Mouse + Key Board)
	VGA	Analog CRT RGB Interface
	Serial interface	RS-232C Conforming ×2
	LAN1	10Base-T/100Base-TX/1000Base-T
	LAN2	10Base-T/100Base-TX
	USB Port	USB2.0 Conforming × 2
	GPS data output port	RS-232C Conforming

After developing WAVE system prototype system, it is prepared field evaluation. AT first two RSUs are set separately which are totally isolated each other in wireless communication range because GPS synchronization measurement has to be measured each different wireless zone. (Refer to Figure 4.) In this evaluation, it is measured packet reach time for each RSUs and OBU. The velocity of vehicle is up to 80km/h.

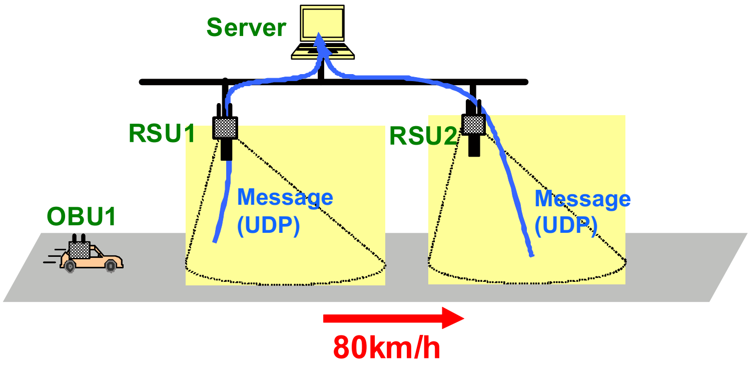


Figure 4 Field Evaluation Environments

The measurement results are shown in Figure 5. As shown in figure 4. there are two RUSs (RSU1 and RSU2). It is measured packet reach time from OBU to each RSUs once after entering each wireless communication zones. The packet reach time is measured when OBU gets into zone where wireless communication signal received area. For wireless communication zone, the packet reach time is measured after the first RSSI (Receive Signal Strength Indication) is detected in wireless communication zone as starting point of communication between RSU and OBU.

After two RSUs measurements with OBU, packet each ability time is less than 1 sec. This shows much faster than current WiFi system link establishment.

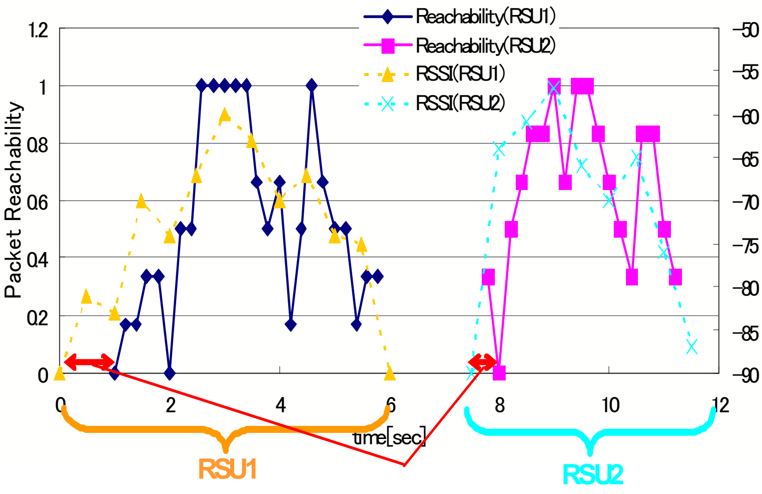


Figure.5 Field evaluation environment

Here is comparison between WiFi and WAVE system in field test. The WAVE has capable to high speed link establishment and short time packet data receiving between RSU and OBU and WAVE message such as emergency message can be delivered by using CCH because of CCH has always the certain time slot On the other hand, if there is collision in WiFi system, packet data must be wait for transmission during heavy data traffic condition. So if the message is emergency information, the delivery timing could be delayed by this congestion. (Refer to Figure 6).

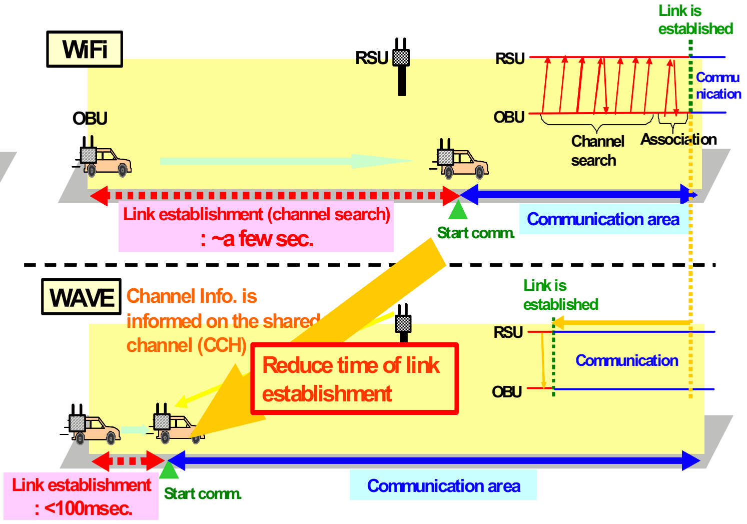


Figure 6 WAVE link establishments

Therefore WAVE has advantage short link establishment between OBU and RSUs which is critical for moving vehicle communication system. In case of safety application to use CCH for informing emergency information such as warning troubled vehicle in advance to the coming vehicle.[2] The existing WiFi system is not useful because of heavy traffic communication usage. So at this moment, WAVE (IEEE802.11p and IEEE1609) is suitable wireless communication system for safety application. (Refer to Figure 7.).

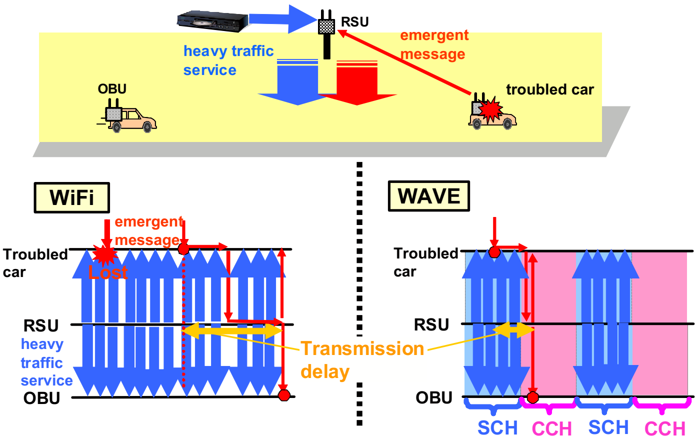


Figure 7. Emergency message transmission

IV. Dual Receiver Concept

According to IEEE802.11p D1.0 specification, CCH and SCH can be used only one at the same time because of channel allocation. The CCH is only active first 50msec time interval and then the SCH can use after continuous 50msec time interval under current IEEE802.11p draft 1 specification. (Refer to Figure 8.).

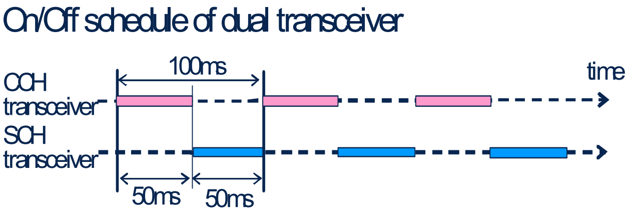


Figure 8 CCH and SCH assignment of IEEE802.11p

In this specification, CCH can only use each 50msec time interval, therefore if CCH is used to inform emergency information. It can be informed 50msec after ending contiguous SCH time frame, if CCH execution is valid during first CCH time frame. If CCH execution is missed to send during the first CCH time frame, it should be wait for the second CCH time frame, which means 150msec waiting time for CCH execution becomes valid. This causes because each CCH has to be wait 50msec after SCH time frame is finished. Therefore the worst case of CCH execution has to wait 150msec. But if dual receiver concept is used, each CCH time frame is potential to access safety information sending, therefore 50msec is waiting time to inform safety information to OBU, which means three times more than fast notification with single receiver concept defined the current IEEE802.11p. (Refer to Figure 9.).[3]

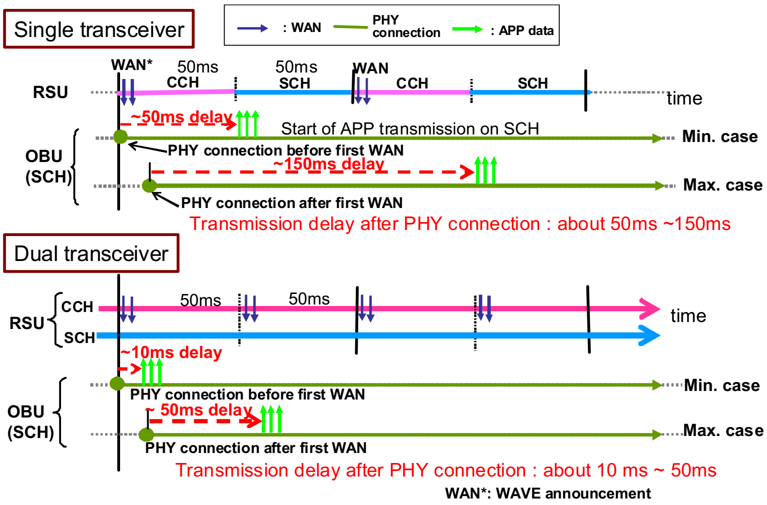

Single transceiver
: WAN
PHY connection
: APP data
WAN*
50ms
CCH
50ms
SCH
WAN
CCH
SCH
RSU
time
~50ms delay
Start of APP transmission on SCH
OBU (SCH)
PHY connection before first WAN
Min. case
~150ms delay
Max. case
PHY connection after first WAN
Transmission delay after PHY connection : about 50ms ~150ms
Dual transceiver
CCH
50ms
50ms
RSU
SCH
time
~10ms delay
OBU (SCH)
PHY connection before first WAN
Min. case
50ms delay
Max. case
PHY connection after first WAN
Transmission delay after PHY connection : about 10 ms ~ 50ms
WAN*: WAVE announcement

Figure 9 Dual Receiver Concepts for 802.11p

Summary

For the next DSRC application which is used for safety application, IEEE802.11p and IEEE1609 is most appropriate system specification because of good enough short time link establishment between OUB and RSUs which shows comparison WiFi and IEEE802.11p specification with real field evaluation with WAVE prototype system which has been developed for this evaluation. According to real field evaluation, the dual receiver concept achieves more than three times faster than using single receiver for informing emergency information to vehicles and also it has more flexibility to inform safety information with using SCH.

However it should have more detail continuous system evaluation with WAVE prototype and keep to feed back to next DSRC standardization.

In this paper, it has been evaluated with two RSUs and OBU. However there is several more detail system evaluation should be examined.

References

[1] IEEE 802.11 LAN/MAN Standards Committee home page (http://www.802.org/11)

[2] 7th ITS-Telecommunications 2007 A.Shimizu " Enhanced Functions of 802.11 Protocol for Adaptation to Communications between High Speed Vehicle and Infrastructures".

[3] 7th ITS-Telecommunications 2007 M.Hayashi "Development of Vehicle Communication (WAVE) System for Safety Applications".

Performance Analysis of Store-Carry-Forward procedure in dense Vehicular Ad-Hoc Networks

Rostam Shirani
Isfahan University of Technology
rostam_sh@ec.iut.ac.ir

Abstract

The idea of Store-Carry-Forward has been introduced for using in sparse Vehicular Ad-hoc Networks (VANETs) in order to answer the problems of VANET as a highly mobile, partitioned ad-hoc network; however, there has not been a deep analytical discussion on applications of Store-Carry-Forward in dense VANETs. In this paper, a mathematical model for urban mobility is proposed. Based on our proposed model, we analyze performance of Store-Carry-Forward procedure in dense VANETs. Simulation validates our analytical results

1. Introduction

COMFORT applications [1] are a category of applications in VANET which are used to prepare some facilities for in-vehicle passengers. Weather information, finding locations of gas stations or restaurants, music sharing and Internet access are all in the category of comfort applications. In comfort applications, usually we need to have Vehicle-to-Vehicle (V2V), or Vehicle to Roadside (V2R) communication. As an example of V2V, assume that a passenger in a vehicle needs to know the location of a gas station or restaurant. In this situation, he can ask other vehicles to answer him. Also, In-vehicle Internet is the most important use of V2R communication.

In this article, our focus is on comfort applications which are delay tolerant [2]. We assume that all vehicles use store-carry-forward (SCF) procedure [3] for sending and receiving data. When a vehicle is passing through a street, the information of that street is gathered and saved on-board in the vehicle. This information could be received by vehicle's sensors or through communication by roadside base stations. When other vehicles requested for that information, the carrier vehicle will forward the data. For the rest of the paper, we assume that each vehicle has an unlimited buffer for storing information.

SCF can be helpful in a sparse VANET scenario [4] (when the vehicles' density is low, for example at night or in a rural society). In this situation, VANET acts like a highly mobile partitioned ad hoc network. And SCF can be used to be able to connect partially disconnected parts. It is clear that this use of mobility is only suitable for delay tolerant applications like comfort applications of VANET.

In this paper, we analyze advantages of SCF in dense VANETs. A VANET is called dense when the density of vehicles is much enough and we encounter with big partitions of nodes in the network [5]. Unlike sparse VANETs, SCF will not be useful as a connector of partitioned networks; because in dense VANETs, lots of cars are found in a partition. However, we will introduce another application for SCF procedure which is applicable in dense VANETs. In order to clarify the issue, notice that number of hops is a restricting parameter in every mobile ad hoc network. As a fact, increasing mobility makes traditional end-to-end multi-hop links more challenging. This is why using traditional MANET routing protocols are usually undesirable for VANET scenarios [6].

In this paper, we want to analyze information availability in a realistic VANET scenario which uses SCF procedure. We use a Markov chain with enough accuracy to model streets and urban mobility. Then based on our model, we calculate probability of information availability that is needed for reaching to the essential information in a scenario. The less the number of hops in a high speed VANET scenario is, the more reliable links for data dissemination will be. A direct result of having reliable links is more correct data delivery, reliability and therefore better use of capacity.

The remainder of this paper is organized as follows. In Section II, we discuss related works on SCF procedure for MANETs and VANETs. Section III explains the notations and preliminaries for next discussions. Our Markov chain model is described in Section IV. Section V discusses about probability of information availability in the presence of SCF procedure. Simulation tool, scenario, and results are explained in section VI. The paper is concluded in Section VII.

2. Related works

The SCF procedure was proposed for partitioned ad hoc networks in [3]. After that, other researchers modified the idea of SCF in their works [7, 8].

After applying SCF for MANETs, researchers tried to use a kind of SCF procedure for routing in VANET scenario. VADD [9] (Vehicle-Assisted Data Delivery) is one of the most important of proposed methods. In VADD protocol, the idea is based on using predictable mobility of vehicles and SCF to be able to choose the best path for data forwarding.

SODAD [10] (Segment-Oriented Data Abstraction and Dissemination) is another proposed method that is based on SCF. In this method, it is assumed that the streets are divided into different segments and vehicles aggregate the data of different sections as they are passing through that section. After that, each vehicle store aggregated data in its on-board knowledge base. And finally, when other vehicles send request for data; the information, if existed, will be transmitted to the cars which had requested for.

Both of the aforementioned methods, which are the most powerful works on using SCF in VANET, used the assumption of information availability intuitively to be able to introduce a forwarding mechanism in VANET scenario. But in this article, we do not want to introduce a forwarding protocol. Instead, we calculate the reduction in number of hops when SCF is applied.

3. Preliminaries

Consider the scenario that is shown in Figure 1. Squares or junctions are recognized by numbers and streets are distinguished by their edge squares. For example *street 1to5* means that side of the street which vehicles are allowed to move from *square 1* towards *square 5.* The same when we say *square 5to1*, we want to clarify the direction from *square 5* to *square 1.* And as it was shown in Figure 1, *street 1-5* means both sides of the street from *1to5* and *5to1.*

Let us define some parameters that we will use in the rest of the paper. First of all, wherever we say vehicle, it means a vehicle with capability of SCF. Also, when we say *hop,* its definition is a little different from that of a traditional MANET. For clearer understanding, imagine there are three vehicles called *A, B, and C* in a VANET scenario. As an example, *vehicle A* is going to communicate with *C* that is not in its communication range. Therefore, using SCF, *A* carries the information and delivers them to vehicle *B* which is probable to meet *vehicle C* later. Then this is *B*'s duty to store and then forward the information toward *C* at the right moment. In the above scenario, by hop we mean number of mutual communications between two nodes that is needed for a piece of data to reach from an origin (*Vehicle A*) to a destination (*Vehicle C*). Thus for aforementioned example, 2-hop communication is needed. Therefore if we imagine that vehicle X has decided to have one hop communication with Y, there will be two possibilities in this situation: 1) Y is already available in X's transmission range, we name it *forwarding hop.* 2) X is forced to carry information to reach to Y's transmission range that is called *catch-up hop*.

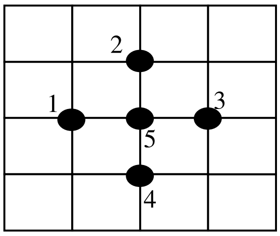


Fig. 1. Considered scenario

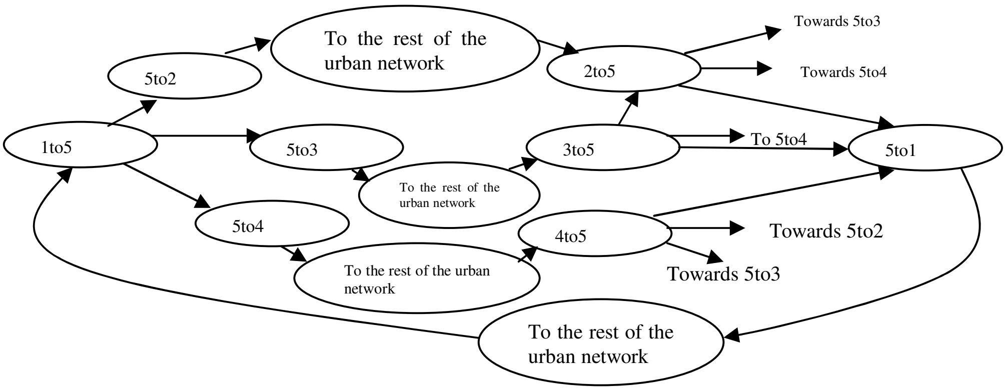
Generally having fewer hops (both forwarding hops and catch-up hops) increases QoS. Because having fewer links over an instable wireless channel, decreases number of packet losses, collisions, channel errors and so on. Also, it is obvious that forwarding hops have lower delay in comparison with catch-up hops. This is why in a forwarding hop; messages are propagated by electromagnetic waves. But in catch-up hops, vehicle is forced to carry the message. Thus propagation speed is limited by the vehicle's velocity. However, when we encountered with a partitioned vehicular ad hoc network, using catch-up hops could be helpful.

Another pre-assumption is that we assume there is a first hop neighbor for vehicle A. If there is not any first hop neighbor, so it is out of our discussion because no communication can be existed. However, even if there is not any other car in our communication range now, it will be possible to have a neighbor after a short period of time (in the order of seconds) [6, 10]. Thus it is feasible to assume existence of first hop neighbors especially when we work on dense VANETs.

Now imagine *vehicle A* in *street 2-5* needs some information about *street 1-5,* it can ask a vehicle which has been in *street 1-5* before. The more the vehicles that travel from *street 1to5* to *street 5to2*, the more information is available with fewer hops.

To be mathematically accurate, it is needed to define some other values. $P(5to2 \mid 1to5)$ is the probability that a random vehicle goes to 5to2, given it has been in street 1to5. In other word, $P(5to2 \mid 1to5)$ is that proportion of vehicles in 1to5 which will go to 5to1 in square 5. $P(1to5 \mid Car\ in\ 5to2)$ is the probability that a randomly selected vehicle had been in *1to5* before, given this is in street 5to2 now.

$P^{n^{th}-hop}(1to5 \mid Car\ \ in\ \ 5to2)$ is the probability that vehicle A can reach to the essential information (about 1-5) with at most *n* hop, given vehicle A is in *5to2* now.



Towards 5to3
To the rest of the urban network
5to2
2to5
Towards 5to4
1to5
5to3
3to5
To 5to4
5to1
To the rest of the urban network
5to4
4to5
Towards 5to2
To the rest of the urban network
Towards 5to3
To the rest of the urban network

Fig. 2. Markov chain of cross road

We define other sets of values for our next discussions. $N_{i,j}(t)$ is number of vehicles which are traveling in street *itoj* at time *t.* Also $N_{1,5}^{5,2}(t)$ is the total number of vehicles which had been in *street 1to5* before, but at time *t* they are in *street 5to2.*

$q_{i,j}(t)$ is vehicle flow in streets *itoj.* The flow *q* means the number of vehicles that pass an observer per unit time. Flow is expressed as vehicle per hour. Finally, we should mention that traffic streams are not static quantities, thus all of the previous variables are stochastic processes. So it is better to analyze the average of above parameters.

4. Markov Chain model

In general, many attempts have been done on modeling urban mobility [11, 12]. For our purpose, we use a Markov chain model to analyze mobility at an intersection. To do that, system's states should contain memory of the system. It means knowledge about present state of the system should be enough to calculate probable future states. After accurate definition of states, transition probabilities can be used to show inter-state movements. We map each unidirectional street to one state. As an example in Figure 1, *street 1to5* and *street 5to1* will be two different states. Therefore based on this definition, intersections and cross roads will act as transitions among different states.

Markov chain of the cross road in Figure 1 has been shown in Figure 2. Look at the state *1to5* and notice outcome arcs. There are five possibilities for a vehicle which is travelling in street *1to5*. It can enter to one of the streets *5to2, 5to3 or 5to4*, or it can stay in *1to5* for a while or turn to *5to1*. So, next state of Markov chain could be one of the 5to1, *5to2, 5to3, 5to4 or 1to5.*

In order to simplify our Markov chain, we only considered mobile nodes and also we neglects U-turns; it means we neglected 1to5 and 5to1 states. Therefore the possible states after 1to5 will be 5to2, or 5to3, or 5to4. Another point about Markov chain of Figure 2 is that this diagram is easily expandable to the rest of the urban network. Due to lack of space, transition probabilities are not shown. For now, we express an intuitive deduction as a corollary.

Corollary. *The probability that a first hop neighbor of vehicle A (For instance imagine vehicle A is in street 5to2 now) had been passing a desired street (e. g. 1to5) before is equal to the probability that a randomly selected vehicle in the current street had been passed through that desired street (1to5) (* $P(1to5 \mid Car\, in\, 5to2)$ *and* $P^{1st\_hop}(1to5 \mid Car\ in\ 5to2)$ *are equal).*

Proof:

For the sake of simplicity, we consider a cross road without traffic light (a priority based cross road). After that we can easily elaborate our deduction for a traffic light cross road.

Neighbors of the *vehicle A* are completely random for the *vehicle A*. In other words, *vehicle A* does not have this ability to choose its neighbors and different vehicles will get in and out of *A*'s transmission range randomly. The only knowledge about neighbors is their presence in the current street. Thus we can conclude that the probability that a random car had been in a given street is equal to the probability that a first hop neighbor had been in that street. This is due to the randomness that was the result of cross road without traffic light.

For a traffic lighted cross road, we resort to the random velocity of vehicles. Traffic light stops vehicles of some directions and allow vehicles of other directions to go, it seems that this violates our desired randomness. In spite of this fact, it is worth mentioning that an optimum traffic light is the one which gives more time to the direction with higher traffic density. If we consider an optimum traffic light and the streets with sufficient length, our desired randomness will be achieved because vehicles have different speed that seems random for others. These random speeds cause vehicles of different streets mix with each other again after an initial time.

Now, imagine a vehicle in *5to2* needs to the information of *5to2*. It is desirable for that vehicle to have a first hop neighbor which had been in *1to5* before.

$$P^{1st\_hop}(1to5 \mid Carin5to2) = P(1to5 \mid Carin5to2)$$

$$= \frac{N_{1to5}^{5to2}(t)}{N_{1to5}^{5to2}(t) + N_{3to5}^{5to2}(t) + N_{4to5}^{5to2}(t)} \tag{1}$$

$$\cong \frac{P(5to2 \mid 1to5).q_{1to5}(t)}{P(5to2 \mid 1to5).q_{1to5}(t) + P(5to2 \mid 3to5).q_{3to5}(t) + P(5to2 \mid 4to5).q_{4to5}(t)}$$

In (1), we consider that the number of vehicles which are in *5to2* now and had been is *1to5* before are proportional to the number of vehicles in *1to5*.

With the same terminology, we can define other variables like *P(1to5|Car in 5to2)*. In order to simplify the notification we put these variables in a matrix.

a_{ij}= P("i"to"5"| car in "5"to"j")
a_{ii}=0

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ a_{21} & 0 & a_{23} & a_{24} \\ a_{31} & a_{32} & 0 & a_{34} \\ a_{41} & a_{42} & a_{43} & 0 \end{bmatrix} \tag{2}$$

If we consider a_{ij} as the entry of i^{th} row and j^{th} column, we will have:

$$\sum_{i=1}^{4} a_{ij} = 1 \; ..... \; j = 1, 2, 3, 4 \tag{3}$$

Also, the scenario imposes some other conditions which are:

$$\sum_{i=1, i \neq j}^{4} P(5toi \mid jto5) = 1 \qquad j = 1, 2, 3, 4 \tag{4}$$

In addition, number of vehicles in each street is restricted by the street's dimensions. Therefore we will have the following conditions for the cross road of Figure 1.

$$0 \leq N_{i,j}(t) \leq N_{i,j}^{\max} \; .... \; i = 5 \; and \; j = 1, 2, 3, 4 \tag{5}$$
$$or \;\; i = 1, 2, 3, 4 \; and \; j = 5$$

For the rest of the paper, when we say we want to maximize a matrix or a vector, this means that we want to maximize each of entries simultaneously.

5. Probability of information availability

First of all, let's express the previous problem in a mathematically acceptable shape.

Problem Definition. *The explained problem could be considered as the following optimization problem.*

$$\max imize \; A$$
$$subject \;\; to \sum_{i=1}^{4} a_{ij} = 1 \qquad j = 1, 2, 3, 4$$
$$\sum_{i=1, i \neq j}^{4} P(5toi \mid jto5) = 1 \;\; j = 1, 2, 3, 4 \tag{6}$$
$$0 \leq P(itoj) \mid ktol) \leq 1 \; i, j, k, l$$
$$0 \leq N_{i,j}(t) \leq N_{ij}^{\max} \quad i = 5 and \; j = 1, 2, 3, 4$$
$$or \; i = 1, 2, 3, 4 \; and \; j = 5$$

The first equality constraint $\sum_{i=1}^{4} a_{ij} = 1 \; ..... \; j = 1, 2, 3, 4,$ restricts each nonzero entries of the matrix to increase more than 1/3. Because if an entry increases to (1/3)+ε, other entries of that row should decrease their values –ε. But another problem sometimes prevents us from having exact 1/3 values, this restriction is caused by inequality constraints. This problem happens at some situation where the physical characteristics of urban environments cause the maximum number of vehicles in a street to be much more than the others.

Intuitively, a vehicle in 5to2 can access to its required information, when it can access to the information of 1to5, 3to5 and 4to5 with the highest probability respectively. In other words, if the constraints of our optimization problem lets, the best state of row 2 which is (1/3,0,1/3, 1/3) will be achieved. Therefore we can say that our optimization problem moves towards (1/3,0,1/3,1/3) if the following conditions could be achievable.

$$P(5to2 \mid 1to5).N_{1to5}(t) = P(5to2 \mid 3to5).N_{3to5}(t) = P(5to2 \mid 4to5).N_{4to5}(t)$$
$$P(5to1 \mid 2to5).N_{2to5}(t) = P(5to1 \mid 3to5).N_{3to5}(t) = P(5to1 \mid 4to5).N_{4to5}(t)$$
$$P(5to3 \mid 2to5).N_{2to5}(t) = P(5to3 \mid 1to5).N_{1to5}(t) = P(5to3 \mid 4to5).N_{4to5}(t)$$
$$P(5to4 \mid 1to5).N_{1to5}(t) = P(5to4 \mid 2to5).N_{2to5}(t) = P(5to4 \mid 3to5).N_{3to5}(t)$$

If some of these conditions is violated, the optimum point will get away from (1/3,0,1/3,1/3).

$$a_{i,j} = \begin{cases} 1/3 & i \neq j \\ 0 & i = j \end{cases} \quad (7)$$

The ideal $a_{i,j}$ of matrix A is shown in (7); however, these values are not achievable in a realistic urban scenario as exact as it is shown in A. So what is the importance of these values? For answering this question, consider two sets of probabilities as possible values of the second row of matrix A: (1/3,0,1/3,1/3) and (1/2,0,1/3,1/6). In the second set, *P(1to5|Car in 5to2)* is increased and *P(4to5|Car in 5to2)* is decreased for a vehicle in *5to2*. The point is that when one of these probabilities rises up, it means that more vehicles had been in that street. Intuitively, a vehicle needs the information of more crowded streets with higher probabilities (the need to the information of a street is proportional to the number of vehicles in that street, because more crowded parts are usually commercial parts of the city and people need to know their data). The exact relationship between number of vehicles and the need to information requires deeper statistical analysis which is out of the scope of this paper. We consider the scenario in Figure 2, assume vehicles *0,1,…,N* are in *street 5to2* of Figure 1, we will calculate information availability for vehicle 0 as a variable of number of hops. In reality, it is possible to have more than one 1^{st}, 2^{nd}, ..., n^{th} hop neighbors. Thus, we consider the worst case in which the n^{th} vehicle is just achievable after *n* hops. The probability of accessing information with at most N hops is:

$$\begin{aligned} P_{availability} &= P^{1st\_hop}(1to5 \mid car\,in\,5to2) + \\ &\{(1 - P^{1st\_hop}(1to5 \mid car\,in\,5to2). \\ &P^{2nd\_hop}(1to5 \mid car\,in\,5to2)\} + \\ &\{(1 - P^{1st\_hop}(1to5 \mid car\,in\,5to2)). \\ &(1 - P^{2nd\_hop}(1to5 \mid car\,in\,5to2)). \\ &P^{3rd\_hop}(1to5 \mid car\,in\,5to2)\} + ... \end{aligned} \quad (8)$$

In (8), we supposed that the information is not available at vehicle 0. Also, if we assume that these *N* hops are in the current street which is *5-2*, by using corollary we will have:

$$\begin{aligned} &P^{1st\_hop}(1to5 \mid car\ in\ 5to2) = P^{2nd\_hop}(1to5 \mid car\ in\ 5to2) = \\ &... = P(1to5 \mid car\ in\ 5to2) \end{aligned} \quad (9)$$

With applying matrix A with entries of (7)

$$\begin{aligned} P_{availability} &= \frac{1}{3} + \frac{1}{3}(\frac{2}{3}) + \frac{1}{3}(\frac{2}{3})^2 + \frac{1}{3}(\frac{2}{3})^3 + \cdots + \frac{1}{3}(\frac{2}{3})^{N-1} \\ &= 1 - (\frac{2}{3})^N \end{aligned} \quad (10)$$

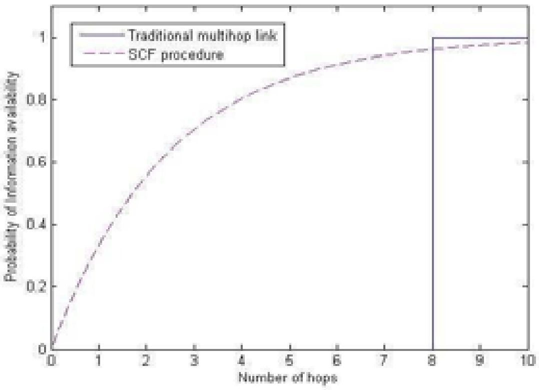


Fig. 3. SCF versus traditional routing

Imagine a vehicle in 5-2 needs the information of 1-5, without using SCF; it is needed to establish a multi-hop link from 5-2 to1-5. When SCF procedure is applied, number of links is reduced. Assume the distance from the vehicle to the place is *d*. The transmission range of the vehicle is also named d_{tr}. In an ideal traditional routing, at least $\lceil d / d_{tr} \rceil$ nodes are required to establish a connection, we considered it 8. When we use SCF, accessing information is probable with lower number of hops. We can see the improvement of SCF in comparison with the traditional source to destination link in Figure 3. With an ideal traditional routing, it is possible to reach information with 8 hops. In SCF, it is probable to reach information with lower number of hops.

6. Simulation

We used SUMO (Simulation of Urban Mobility) to produce real traffic patterns. SUMO is an open source, highly portable, microscopic road traffic simulation package designed to handle large road networks. More information about SUMO can be found in [13].

We consider the city structure of Figure 1. In this scenario, all of roads are two-way one-lane streets with 5km length. Maximum speed of vehicles was considered 60 km/h. Of course, vehicles' velocities depend on urban traffic. It means a vehicle cannot move with its maximum speed all the time, because its movement is restricted by other vehicles' velocities. In the following city structure, we define 10000 vehicles with random sources and destinations. This number of vehicles guarantees having a dense VANET.

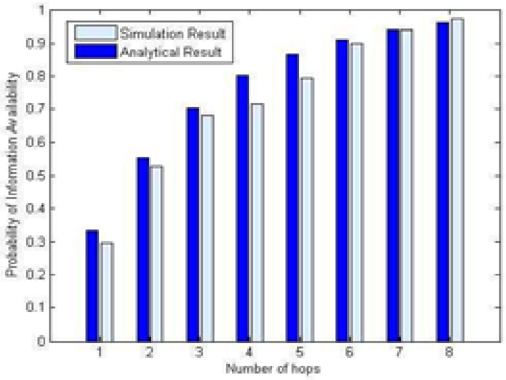


Fig. 4. Simulation results versus analytical results

Except these 10000 cars, we defined a car which travels from square 5 to square 2. We named this car VIP. Then, an exact coordination of street 5to2 was considered as sampling point. In our simulation, sampling point is located 2000 m away from square 5 towards square 2, as soon as our defined car reached to the sampling point the information will be saved in an XML file. This file contains the information about the exact location of different cars in the street in that moment. Also, traveling paths of other vehicles are gathered in another file. With comparing information of these two files, we clarified that where have the first neighbors of VIP been before. We did the same analysis for second, third, and … neighbors. Then, we ran the simulator again for 50 times, and at each time we recorded the same information. Finally, we calculated probability with the measure of occurrence states. We counted how many times neighbors have been in the desired street, *1-5* in our simulation. This number divided by total number of vehicles was named P_i and it is shown in (11).

$$P_i = \frac{\#\,of\ i^{th}\ hop\,vehicles\,which\,has\,been\,in\,1-5}{\#\,of\ i^{th}\ hop\,vehicles} \qquad (11)$$

After measuring $P_i s$, P was calculated and depicted as simulation result in (12), and it was compared with $\mathrm{P}_{\mathrm{availability}}$ of (10).

$$P = \prod_{i=1}^{n-1}(1-P_i).P_n \qquad (12)$$

7. Conclusion

In this paper, we introduced the idea of SCF to be used in dense VANETs. We explained qualitatively, and discussed analytically. Then, we validated our analytic results by simulation. For future works, we want to mix SCF procedure with traditional ad-hoc routing protocols. The idea is to use SCF to join partitioned parts, and for dense mode, we can use SCF and traditional routing when the VANET turns to a static network in rush hours or deadlocks.

References

[1] S. Yousefi, M. S. Mousavi, M. Fathy, “Vehicular Ad Hoc Networks (VANETs): Challenges and Perspectives,” Proc. 6th International Conference on ITS Telecommunications, 2006.

[2] Delay Tolerant Networking Research Group, http://dtnrg.org, 2007.

[3] J. Davis, A. Fagg, B. Levine, “Wearable computers as packet transport mechanisms in highly-partitioned ad-hoc networks,” International Symp. Wearable Computing, October 2001.

[4] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, O. Tonguz, “Routing in Sparse Vehicular Ad Hoc Wireless Networks,” IEEE Journal Selected Areas in Communications, Volume 25, Issue 8, Oct. 2007 Page(s):1538 - 1556

[5] Hao Wu, “Analysis and Design of Vehicular Networks,” PhD thesis, Georgia Institute of Technology, December 2005.

[6] J. J. Blum, A. Eskandarian, L. J. Hoffman, “Challenges of Intervehicle Ad Hoc Networks,” IEEE Trans. Intelligent Transportation System.

[7] Q. Li, D. Rus, “Sending Messages to Mobile Users in Disconnected Ad-hoc Wireless Networks,” MobiCom, 2000.

[8] J. Wu, S. Yang, F. Dai, “Logarithmic Store-Carry-Forward Routing in Mobile Ad Hoc Networks,” IEEE Trans. Parallel and Distributed Systems, vol. 18, no. 6, June 2007.

[9] J. Zhao, G. Cao, “VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks,” Proc. IEEE INFOCOM, 2006.

[10] L. Wischhof, A. Ebner, H. Rohling, “Information Dissemination in Self-Organizing Intervehicle Networks,” IEEE Trans. ITS, VOL. 6, NO. 1, MARCH 2005.

[11] L. A. Pipes, “An Operational Analysis of Traffic Dynamics,” J. Appl. Phys., vol. 24, no. 3, pp. 274-281, Mar. 1953.

[12] K. Nagel, M. Schreckenberg, “A Cellular Automaton Model for Freeway Traffic,” J. Phys., I France, vol. 2, no. 12, pp. 2221-2229, Dec. 1992.

[13] http://sumo.sourceforge.net/docs/documentation.shtm

A new Simulation Method for the Rapid Development of Car-to-X Communication Applications

Alois Mauthofer, Markus Glaab
carhs.communication GmbH
alois.mauthofer@carhs.de

Abstract

The development of future co-operative safety and comfort systems in vehicles as well as efficient traffic management need tools in an early stage of the technology introduction which support the engineers in the development process. Such a tool is viilab which assists automotive manufacturers and suppliers in developing Car-to-X communication applications in the field of driver assistance, vehicle and traffic safety, comfort or infotainment systems. Within this toolkit a simulation method called "Extended Reality", a new kind of "in-the-loop" simulation, is able to create virtual scenarios which enrich the real car driving experience with virtual components like other cars, pedestrians, Road Side Units or traffic signals.

1. Introduction

Future integrated safety and comfort systems in vehicles as well as efficient traffic management require networking of vehicles with one another and with traffic infrastructure. Car-to-X (Car2X) communication serves to optimise the traffic flow and noticeably increase traffic safety through the targeted transmission of information. If, for example, a car runs into a critical situation such as traffic congestion, fog, glazed frost or an accident, it will analyse the situation and pass on the relevant information by using wireless communication technologies to all concerned traffic participants in the immediate perimeter of the hazard area as well as to the infrastructure. Other traffic participants can thus be warned in due time and can react accordingly. Car2X communication is also the basic technology for a broad range of future comfort and infotainment applications.

Future scenarios in traffic can be versatile and very complex. But in an early stage of such a new technology only a small number of units is available for development and testing. The standards for open, non-proprietary interfaces for communication, applications and services are going to be defined in the United States [1][2] and in Europe [3][4]. Nevertheless, to define standards and to establish such a new technology a lot of research and development is needed, many future traffic scenarios have to be taken into consideration including a large number of communicating units in vehicles and infrastructure.

2. Development Tool

The development of future intelligent co-operative systems requires a new generation of development tools which assist the automotive manufacturers and suppliers in developing Car2X communication applications. The software viilab (vehicle infrastructure integration laboratory) [5] is especially designed for Car2X communication application developments. It incorporates two parallel concepts: it supports both the rapid development / rapid prototyping of communication designs and the low-level and thus highly efficient development of algorithms and hardware drivers. For the first concept – rapid prototyping – a scripting language is used which has been specifically adapted to the needs and requirements of vehicle communication, e.g. for the logic combination of events in the communicating vehicle environment with vehicle electronics, actuating elements and navigation. The viilab software can be installed as On Board Unit (OBU), Road Side Unit (RSU) or Monitoring Device (MON). Typical vehicle related interfaces (navigation, positioning, Display/HMI or bus systems) as well as infrastructure related interfaces (camera, traffic light, remote admin, traffic control) can be included. A typical application of viilab can be seen in Figure 1.

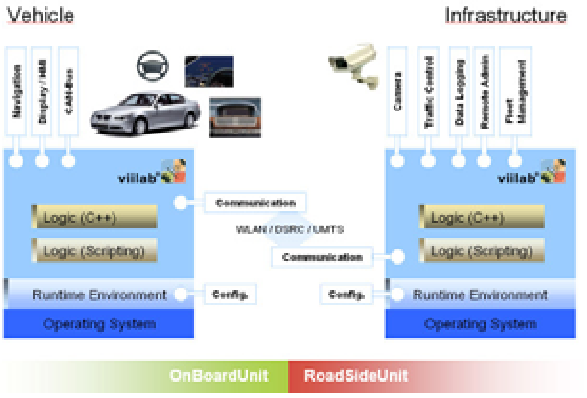


Figure 1.
viilab – Car2X communication development platform

3. A new simulation method

For the development of complex traffic scenarios with many communication units simulation tools are indispensable. Common simulation tools are used e.g. for simulating the computer network (e.g. ns-2 [6]) or the traffic flow (e.g. VISSIM [7]). But such tools are not designed for the combination of computer network and traffic with specific individual Car2X communication scenarios.

An engineer who is developing an algorithm for future Car2X communication applications in a car needs real time simulation of complete traffic scenarios. Car2X communication standards have to be taken into account as well, which are going to be defined for USA [1][2] and Europe [3][4]. A standard message set defines the messages which will be exchanges via wireless communication between vehicles and infrastructure. The purpose of this standard in USA [8], SAE J2735 - Dedicated Short Range Communications (DSRC) Message Set Dictionary, is to support interoperability among DSRC applications through the use of standardized message sets, data frames and data elements.

Therefore there are two tasks for the engineer. The first task is the development of the algorithm of the specific Car2X application which is based on the defined standard protocols and messages. The second task is the development or integration of co-operative units which can communicate with his application in the way that is described in the standard scenario. Co-operative external units can be for example OBUs in other cars or RSUs in the infrastructure.

viilab is designed in such a way that algorithms for either OBU or RSU can be developed independently of real mode or simulation mode. Many units can be combined in any user-defined way from a full simulation mode on one computer to a full real mode including many cars and infrastructure units. Each unit can run with an own specific algorithm. A basic scenario library with predefined algorithms according to the standard can be used as development kit for complex scenarios.

By means of an intersection application (see figure 2) the different modes will be explained in detail. In this figure are shown three cars, one RSU, two traffic lights, one velocity limit and one pedestrian. The cars and the RSU are connected via wireless communication. The control unit of the traffic light has an interface to the RSU. The position and type of traffic sign is stored in the RSU and the dynamic position of the pedestrian is detected by a camera connected to the RSU. Thus the traffic light status, the position of the traffic sign and of the pedestrian can be sent to the cars.

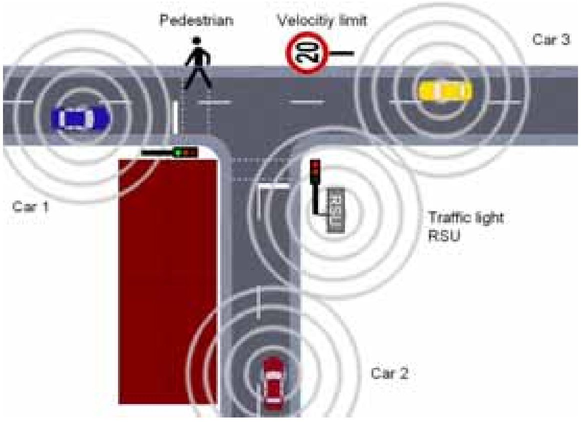


Figure 2.
viilab intersection scenario

The first step in the development process is the full simulation mode. In this mode all viilab processes are running on one computer in the laboratory. In the example three cars with OBUs and one RSU are simulated including virtual movement and the virtual communication. The monitoring unit is “receiving” the messages and positions of all units and assists the engineer in designing and watching the scenario from a top view. In this mode basic algorithms can be developed and tested without any risk. Besides the computer and the viilab software no specific hardware is needed.

A typical full simulation scenario on the developers screen is shown in figure 3. In this scenario five viilab processes are running simultaneously. One RSU process generates traffic sign, traffic light and a virtual pedestrian. The RSU messages are “sent” to the three OBU processes which represent virtual cars driving on the virtual road. The fifth process is the monitoring process. In figure 3 the monitor on the upper right illustrates the scenario. The three other displays show the in-car display / HMI with basic information and warning messages of each virtual car using a standard development GUI. The monitor and the user interfaces can be customized with the viilab GUI development environment.

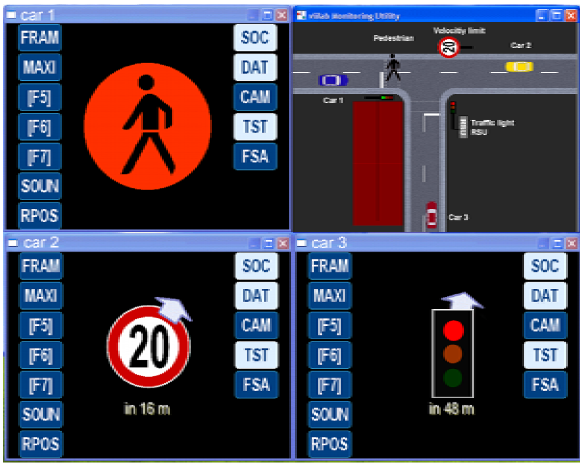


Figure 3.
viilab full simulation mode

In the simulation mode the communication ranges and the warning ranges of cars are virtual (figure 4).

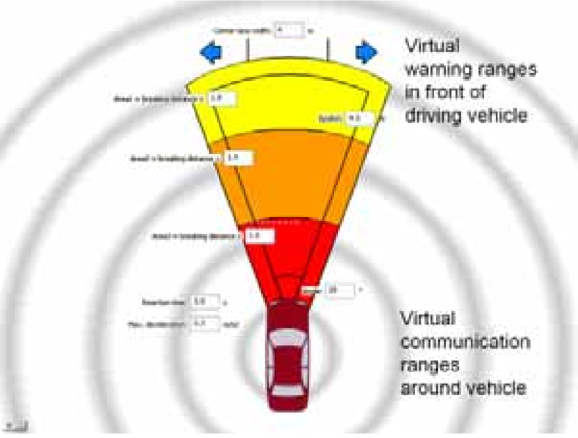


Figure 4.
viilab simulation - virtual warning and communication range

The cars are driving virtually e.g. by replaying previously recorded GPS data ("virtual GPS"). A specific module in viilab called "virtual air" is checking, if a car unit is in the communication range of any other unit. Only in this case the messages are exchanged. The communication range can be defined individually in the configuration file for each process.

In a similar manner the hazard warning is activated if there is for example another car, a pedestrian or a red light in front of a car and the car is "driving" too fast. In figure 5 the three cones in front of the vehicle show the dynamic warning ranges depending on the application and the direction and velocity of the car. If, for example, a pedestrian gets into the yellow range in front of the vehicle the first information action can be initiated like a short acoustic beep. In the orange range an action can be triggered via the CAN bus like a short braking jerk and in the red zone the full Brake Assistant is activated. The warning zones and the related actions can be configured individually for each process.

The next stage from simulation towards reality is the combination of a real application in a car enriched by virtual devices like other cars, pedestrians, Road Side Units or traffic signals for development and test. The external simulation devices behave like a real environment, like an "Extended Reality".

In the extended reality simulation mode the OBU process is installed on a computer in a real car, the driver assistance systems, display and safety systems are connected via interfaces (e.g. CAN bus). The virtual environment processes of other car OBUs and the RSU can run on this computer as well. Thus real applications can be demonstrated in an early stage. E.g. a warning message can be shown on a display in a car, the interface to the CAN bus can be tested and the ergonomics of the HMI can be evaluated.

Step by step virtual components can be replaced by real components. In the next simulation mode the environment can be simulated by a second computer, e.g. a computer in the test car or an external stationary device close to the road (see figure 5).

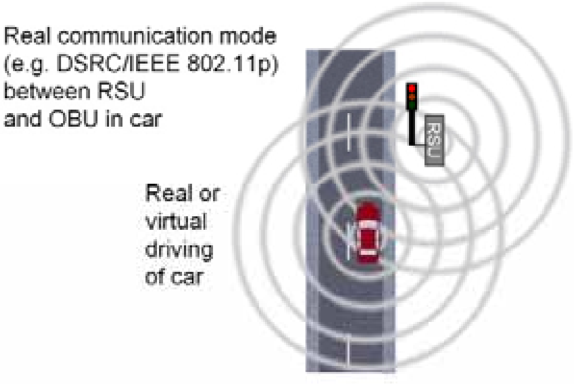


Figure 5.
viilab partial simulation - real application in car with RSU

In this case both units are connected via wireless communication according to the Car2X communication standard (IEEE 802.11p).

In the final step, after a thorough testing of the algorithms in full simulation and extended reality mode, the algorithms can be used without changes in the real scenario as can be seen in figure 2.

Thus, using the viilab simulation environment the development of a Car2X communication application can be performed without a disruption in the development process in a short time and at low risk for driver and hardware.

The versatile simulation modes are enabled by the modular and scalable viilab design. The schematic architecture of viilab for an On Board Unit in a partial simulation can be seen in figure 6. The part which is configured for the simulation mode is marked in the figure.

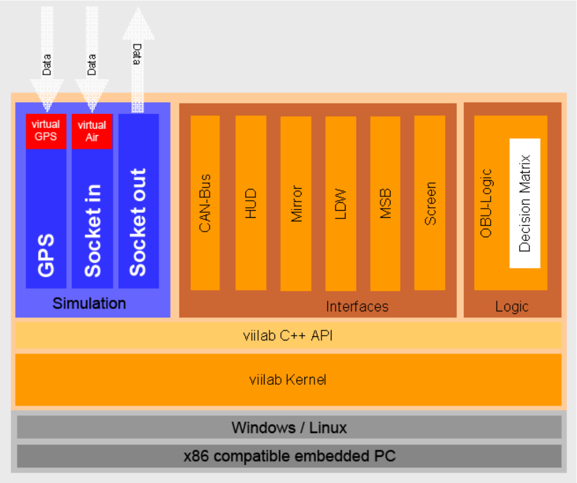
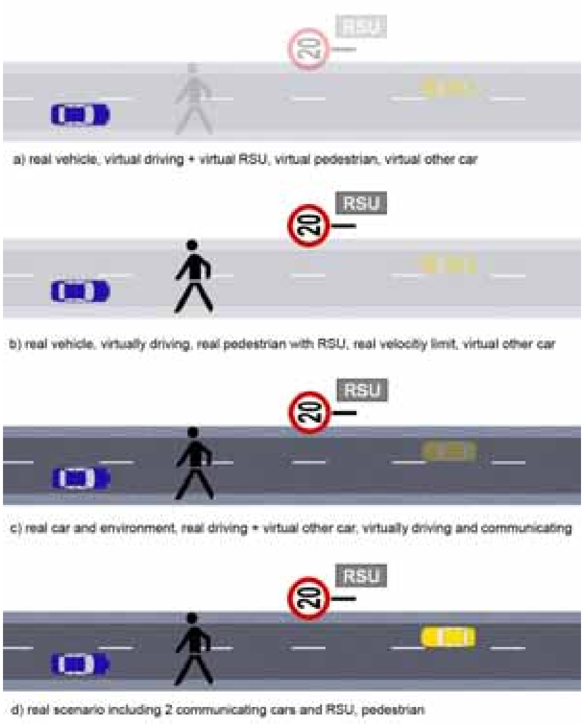


Figure 6.
viilab architectur
OBU simulation mode with “virtual Air” and “virtual GPS”

A typical development process is shown in figure 7. The first step, the full simulation of all processes on one computer (see figure 3) is not included in this overview. In case a) a real vehicle, which is virtually driving on a virtual road, has real interfaces to on-board systems and is enriched by virtual components. A RSU process generates the virtual traffic sign and a virtual moving pedestrian. The RSU messages are sent to the OBU processes which represent virtual cars driving on the virtual road while using programmed or previously recorded GPS data. In case b) a real RSU is added. The position information of the real traffic sign and of the real pedestrian is sent to the car by wireless communication. In case c) the only difference to case b) is the real driving of the car on the road. Real GPS data are used. In case d) the entire scenario is real including a second car which is communicating with the first car and the RSU.



a) real vehicle, virtual driving + virtual RSU, virtual pedestrian, virtual other car

b) real vehicle, virtually driving, real pedestrian with RSU, real velocity limit, virtual other car

c) real car and environment, real driving + virtual other car, virtually driving and communicating

d) real scenario including 2 communicating cars and RSU, pedestrian

Figure 7.
viilab development process - from simulation to reality

The extended reality simulation as can be seen in figure 8 b), c), d) is a breakthrough for developers of Car2X On Board Units and similar control units who don’t want to create the whole environmental scenario by themselves. Since some standardized ready-to-use scenarios as well as additional tools (see chapter 2) are available the virtual or real environment can be included with minimal effort into the own development process. Based on communication and message standards the wireless communication between the own unit(s) and the viilab unit(s) is guaranteed.

4. Summary

The viilab “Extended Reality” simulation is a powerful tool to rapidly develop and test Car2X communication applications and services for traffic scenarios at low risk for driver and hardware. In a very short time it is possible to develop, control and debug the applications and algorithms in detail. In particular a smooth transfer without a disruption in the development process is possible from complete virtual simulation on one computer to a complete real testing on many distributed units.

5. References

[1] National ITS Architecture,
http://www.iteris.com/itsarch/index.htm

[2] ITS Standards Program Website,
http://www.standards.its.dot.gov

[3] Car 2 Car Communication Consortium, Manifesto,
http://www.car-2-car.org

[4] CVIS, Cooperative Vehicle-Infrastructure Systems,
http://www.cvisproject.org/

[5] viilab architecture, http://www.viilab.de,
carhs.communication GmbH,
Siemensstr. 12, D-63755 Alzenau

[6] ns2 Network Simulator,
http://nsnam.isi.edu/nsnam/index.php/Main_Page

[7] VISSIM Traffic Simulation,
http://www.english.ptv.de/cgi-bin/traffic/traf_vissim.pl

[8] Dedicated Short Range Communication (DSRC) Message Set Dictionary, SAE J2735, SAE International, U.S. Department of Transportation

The Automatic Green Light Project – Vehicular Traffic Optimization via Velocity Advice

Martin Goralczyk, Jens Pontow, Florian Häusler, Ilja Radusch
Technische Universität Berlin,
Sekr. DCAITI, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{martin.goralczyk,jens.pontow,florian.haeusler,ilja.radusch}@dcaiti.com

Abstract

Car rides through inner cities are characterized by frequent changes of driving speed resulting in high fuel consumption and CO_2 emission. This paper presents two approaches, which result in car rides through congested urban areas with a reduced amount of acceleration by calculating adaptive speed advices. Hence, we describe a system partly running on a GPS-enabled smartphone in the car. Thereby, the smartphone collects GPS data and gives speed advices. We show that the driver who adopts the velocity advices will drive with much more constant speed while reaching his or her destination location in the same time. The first approach tries to identify traffic lights and advices optimal speed to hit no red traffic light. The second approach tries to forecast traffic flow conditions and advices the best speed for the actual situation. Both approaches are able to give useful speed advices to the driver.

1. Introduction

Due to the worldwide increasing traffic in combination with the limited capacity of road networks, traveling by car gets more and more frustrating in congested urban areas. Most of the car stops during a route through the city are imposed by junctions and their rules and protocols. When traffic lights are used to control the traffic flow, a simple rule is that a junction can only be passed by one direction at a time. Traffic lights determine alternating time slots where cars are allowed to pass or where cars are not allowed to enter the junction. In an ideal setting without any traffic lights every participant would enter the junction instantly without colliding with other cars. However, such as a scenario requires additional information and a complex and reliable communication infrastructure. This charming vision of highly frequented junctions without traffic lights, just based on communication, suffers from severe safety issues and lacks of an implementation in the near future as envisioned by Dresner and Stone [1]. Hence, in this paper we concentrate on the current scenario with traffic lights, wherein Letia [2] has done research, and try to optimize it as far as possible.

From the ecological and economic point of view every type of acceleration, indifferent whether positive or negative, has to be avoided. For this reason, succeeding traffic lights on frequently used streets are connected to each other or chronologically synchronized. So the cars on such roads have some kind of automatic green light if they are driving at the speed-limit or a specific velocity. Even though driving on such a "green wave" is very simple, a route through a city is a combination of many of those and at connection points the probability is high to get out of flow of the “green wave”. Furthermore, obstacles like a slowly driving bus or garbage truck can force the driver to reduce his speed or even stop so he loses his connection to the next green traffic light. In addition, some traffic lights are not connected to each other, maybe because they are not on the same street, but are using the same time intervals. In other words, there exists a perfect speed to get from one traffic light to the other hitting the green light, but in the majority of cases it is not the usual speed-limit. These facts show that there is some hidden potential in our present traffic system, but a trivial way to use this potential is missing.

One way to reveal this hidden potential is to optimize the velocity of vehicles. Our previous goal is to reduce CO_2 emission while having no disadvantage regarding the arrival time.

Our vision is to drive through a city in a similar way we drive through the countryside. This means that our aim is to drive with a more or less constant speed and with just a few stops.

In this paper we present a concept called Automatic Green Light (AGL). Part of this concept is a software which runs on a GPS-enabled PDA or smartphone inside the car. While driving, this software collects information about the traffic lights or current traffic flow conditions and gives advice how fast to drive to catch the next green light according to the collected data. To reduce the need for possibly inaccurate map data the application runs without a map and gathers its information just from the tracked behavior of all the participants and saves the generalized data into a database, where it can be served to all clients.

In the following chapters, we will present related work to reduce CO_2 emissions as well as our approaches, their results, discuss them and give a brief overview about future work we are intending to do on this issue.

2. Related Work

A couple of different potential approaches exist to reduce the fuel consumption and thereby CO_2 emissions of cars. In this section, we will present five selected measures.

2.1. New Technologies

One possible way of CO_2 reduction is to encourage new technologies by adequate laws and taxes. In 1996, DeCicco et al. [6] forecasted that fuel consumption could be reduced by 49% by applying new technologies. Unfortunately, new technologies need plenty of time until they take effect because the whole vehicular fleet needs to be substituted by vehicles featuring these newest technologies of fuel conservation.

2.2. New Infrastructure

Another room for improvement is to build new roads or extend the existing road network. However, extending the road network is obviously an inappropriate solution considering the general goal of environmental protection. In addition, in congested urban areas, there is no further space available to expand the road network usually. Hence, this option is inapplicable in most scenarios.

In our opinion, most of the same problems apply to developing public transportation.

2.3. Temporal and Spatial Reallocation of Traffic

Another chance to avoid unnecessary high fuel consumption of individual motor car traffic is to arrange the traffic in a way that avoids congestion. For example, flexible work time may lead to better temporal distribution of traffic volume. The same applies to spatial reallocation of traffic volume where road users with the same destination use different roads to get to their destination. Yamashita et al. [7] have developed a cooperative car navigation system that harmonizes planned car routes to avoid congestion.

2.4. Traffic Management

Improvements in the management of junctions as envisioned by Dresner et al. [1] are able to minimize the amount of necessary car stops and thereby minimize CO_2 emissions. In this scenario, traffic lights are substituted by a system that synchronizes car movements at junctions in a more flexible way than traffic lights are able to. In our opinion, this approach is inapplicable in the near future because every car needs to be equipped with such a sophisticated onboard unit that manages the speed and direction of all participating cars. Hence, as long as there are cars without such a device, this approach will not completely work.

2.5. Driver Behavior Improvement

It is a well-known fact that different driving characteristics essentially influence fuel consumption. Research of Waters et al. [8] shows that different driver behavior may lead to a difference in fuel consumption of up to 50%. In this context, Voort et al. [9] have done research on a driver support tool that gives advice to the driver on how to minimize fuel consumption. The advice of their system relates to shifting gears at optimal moments and choosing the best rate of acceleration. These behavioral advices result in fuel savings of about 23% compared to ‘normal driving’ without any advice. We think that the most promising short-term approaches of CO_2 reduction of vehicular traffic in urban areas are changes in the driving characteristics. Especially the avoidance of unnecessary car stops seems to be promising.

The AGL-approach described in the following chapters is such an approach, because it aims to influence or improve individual driving behavior.

3. Work

The AGL project could be divided into three separate blocks: (1) Traffic Data Recording, (2) Data Analysis and (3) Velocity Advice, see Figure 1. The modular implementation of our approach reflects these building blocks.

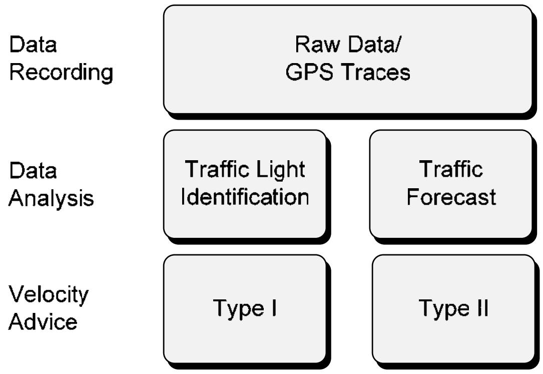


Figure 1. AGL architecture overview

We decided to implement two different approaches, both leading to a velocity advice.

3.1. Data Recording

In order to implement a first prototype and proof of concept, we recorded a certain driving route through the city of Berlin for several times. We have driven the car during the rush hour on crowded streets in a normal way. That means we drove as fast as possible respecting the traffic flow and legal speed-limits.

During the complete trip of about 10 kilometers, we recorded our location periodically with help of a common GPS receiver. A central database stores the complete recorded (historic) data. Until now, the number of recorded trips is less than ten.

The AGL project aims at developing a system in which every participating user profits from the data of other users. Moreover, the different algorithms should work on the same most accurate data, so all the data from every GPS event from the mobile devices is stored to the public part of the database. Thus, every user is able to add new raw data.

3.2. Data Analysis

The recorded raw data (as described in Section 3.1 Data Recording) is the input to the Data Analysis building block. It generates traffic models, which serve as basis to recommend velocities regarding certain criterions. For evaluation, we implemented two data analysis approaches completely separated.

The first one focuses on minimizing stops because of red traffic lights. In the second one, we try to identify situations, where acceleration is of no use and could be avoided without causing any disadvantages regarding the arrival time.

Both approaches include typical strengths and weaknesses. On the one hand, we are able to pre-compute traffic models on a powerful computer with complex algorithms when we employ the Traffic Light Identification-approach. On the other hand, we expect the velocity forecasting approach to have a greater potential regarding overall velocity optimization. At the same time, we depend on employing rather simple algorithms to fulfill real-time requirements when driving.

Hence, both approaches generate so-called tracks from the provided raw data, which serve as an input to the velocity recommender building block. The special character of a track is that the optimal velocity is constant. This data preparation helps to reduce the complexity of computation on the PDA.

Tracks consist of

- a starting and ending point (which could be two traffic lights),
- the length on the road between those two,
- a heading value that corresponds with the heading a few seconds after the car started from the starting point, and
- the average speed which has to be driven to get to the ending point within a certain time.

3.2.1. Traffic Light Identification

The automatic green light problem, based on the precondition that we have no maps but just historic car positions and velocities that are recorded from GPS receivers, can be divided into two major problems:

- the identification of traffic lights and
- the calculation of the optimal speed to hit the next green light.

Our first algorithm is based on the rather trivial assumption that we expect a red light every time the speed of a car is zero. We improved the approach by excluding start- and endpoint of each trace assuming the majority of such cases determine parking processes. Furthermore, we re-started recording the driving direction after a seven second delay to ensure that the car has left the expected junction area.

3.2.2. Velocity Forecast

The goal of our second approach is to identify situations, where acceleration is of no use and could be avoided without any disadvantages. One of such situations could be explained with help of Figure 3. The continuous line illustrates a typical movement of a car driving through an urban site. The dotted line's slope in Figure 3 determines the optimal velocity regarding the section between t_a and t_b concerning the "minimizing acceleration"-optimization criterion.

Whereas the post-calculation of the optimal velocity is rather easy to implement, the pre-calculation or estimation of the optimal velocity of an ongoing track is rather complicated.

The basic idea of this approach is to forecast forthcoming velocities the individual driver is going to drive for a certain time horizon under normal circumstances. This relevant forecasting time horizon should be chosen based on the optimization criterion. Based on the forecasted time series of velocities, it is possible to calculate and recommend a certain speed concerning any optimization.

Velocity and density are common parameters in general traffic forecasting. Different to former traffic forecasting works, we require less initial traffic data to perform an individual forecast. As previously described, we do not employ live traffic monitoring data. That means, we do not monitor traffic flows at different locations at a certain time but the traffic flow at a certain location at a certain time. That corresponds to the individual velocity time series recorded by the relevant car. Thus we could not employ common forecasting methods, such as ARIMA modeling employed in [3], the Kalman Filter in [4], or non-parametric regression models in [5].

Based on the limited pool of traffic data, we decided to employ a non-parametric regression method belonging to non-linear time series analysis techniques. Concerned time series are the previously recorded trips called historic time series as well as the trip currently monitored by a device inside the car called current time series.

Based on the monitored data during a ride (location x, time t, driving direction d, and velocity v), we define time series as a chain of states s. A state is a combination of scalar measurements:

$$s_t = \begin{pmatrix} x_t & d_t & v_t \end{pmatrix} \qquad \textbf{(1)}$$

The basic forecasting idea is to find one or more parts of time series within the pool of historic time series, which are similar to the time series monitored during the current trip. Historic time series are extracted, which include the most relevant (most similar) parts concerning the current trip. In case the most similar parts within the historic time series have successive states, we assume that there is coherence between these successive states and the future states of the currently monitored time series.

In order to find similar time series' parts, we employ the method of phase space embedding realized by delay reconstruction in k dimensions. The vectors within this phase space are given by:

$$\vec{s}_n = \left(s_n, s_{n-1}, s_{n-2}, \ldots, s_{n-k+1}, s_{n-k}\right) \qquad \textbf{(2)}$$

According to the recorded data, the time difference (time units) between two following states is exactly one second. Thus, we take the k most recent states of the current time series and search for the most similar phase space vectors within the pool of historic time series. Similarity computing corresponds to the calculation of the Euclidian distance between two k-dimensional points.

Successive states of the r most similar parts of time series serve as a calculation basis for the estimated oncoming time series (regarding the current time series). The number of forecasted states is limited by the previously determined forecasting time horizon h. To provide adaptability, we re-estimate the forecasted time series periodically every p time units (seconds). A forecasted time series with p = 15 is illustrated in Figure 2.

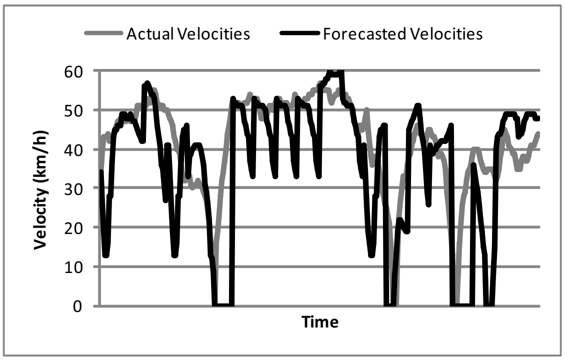


Figure 2. Actual velocities and forecasted velocities

The iterative forecasted series becomes better with smaller p but less helpful for a meaningful velocity recommendation.

Based on the results of forecasting, we cut the forecasted time series into parts. It is delimited by states, where we expect the driver to start accelerating after he or she will have stopped, slowed down or driven with constant speed. The section between these two points is a track. Figure 3 illustrates two of these delimiting states in t_a and t_b, which correspond to the start and end point of a track.

3.3. Velocity Advice

Based on the results of data analysis, we calculate the constant velocity, which is needed for a track previously computed in the data analysis building block. Concerning our forecasting approach, we calculate the constant speed v regarding the oncoming track, which is needed to reach the location of the last state s within the relevant track in the same time.

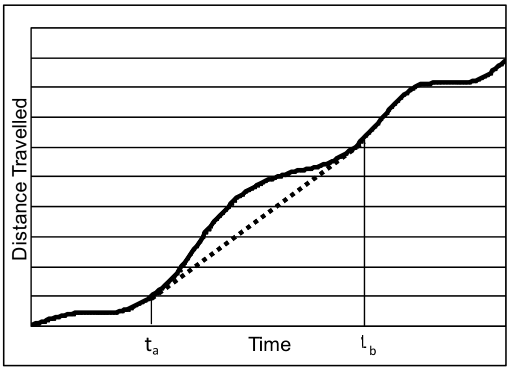


Figure 3. Typical trip through an urban side and optimal velocity

That means that we generate velocity recommendation during the trip in real time.

4. Results

Figures 4 and 5 visualize the velocity of a trip through the city of Berlin and the according velocity recommendations as an output of our two approaches.

The results based on the traffic light identification approach (as shown in Figure 4) computes the positions of traffic lights and advices the average speed between two identified traffic lights.

The Velocity Forecast-based approach is illustrated in Figure 5. It recommends optimized velocities based on a pool of historic trip data in real time. Even though only few data is available, we achieved promising results. Whereas the driver had to stop (or drive very slow) for five times, we recommended velocities in real time, which results in a very balanced driving with velocities between 31 and 47 kilometers per hour.

The difference between the two algorithms is shown by their recommended velocity. While the trivial algorithm tries to give constant recommendations for the complete distance between two detected traffic lights, the Forecast-based velocity recommendations are less constant, which is reasoned by non-satisfying forecasts.

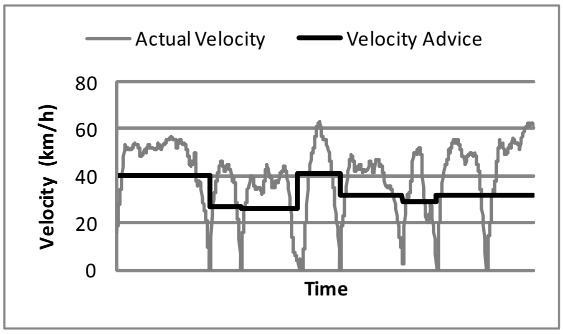


Figure 4. Velocity advice type I

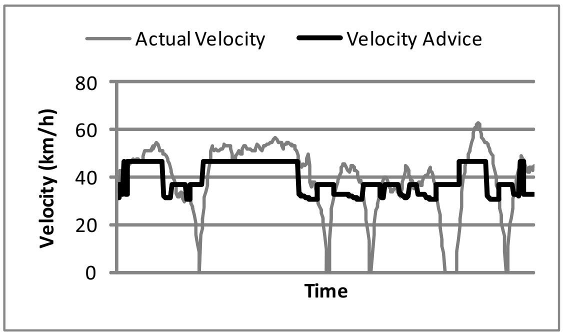


Figure 5. Velocity advice type II

5. Discussion

First, the results demonstrate the feasibility of our approaches to get useful speed advices that base upon historic data. The generated speed recommendations will lead to drive in a more relaxing, economic and energy-efficient way. For now, the most severe problem is the small data set. Currently, simulated data is no option due to our need for realistic variance in driving behavior. We need many data recordings of the same routes that ideally differ from each other in the driver's behavior like the acceleration and speed, because this is the data our algorithms rely on. At the moment we will have collected this required amount of real data, we will present more detailed results. We expect great advances of our results as soon as we have a bigger pool of historic data, which will help us refining parameters in both approaches.

One of our central future tasks is the AGL evaluation. First of all, we will test our application live to discover potential reciprocal effects with normal traffic (participants) as well as to confirm our current results. Additionally, we will run the applications against common quality metrics in order to compare and enhance our approaches.

6. Future Work

The current implementation of the AGL application provides a lot of room for improvement and further development.

Due to data inaccuracies, we discuss employing better GPS sensors or an inertial GPS receiver for future trace recordings.

As previously denoted, both of our approaches show different shortcomings and strengths. We will combine both methods to a new hybrid approach to exploit the differences.

Both approaches provide chances for improvements for themselves. The traffic light identification approach assumes that adjacent traffic lights have the same cycle times. Indeed, most traffic lights have variable cycles depending on different influences. We have already detected inconsistencies in velocity recommendations (flickering data), which is an evidence for such a situation. We expect further improvements by computing the length of the traffic lights' cycle times. Additional parameters will be considered. That includes the absolute time, the knowledge of the cycle's length regarding the daytime, and an absolute timestamp from the history when this particular traffic light turned green. To determine variable cycle times with a certain confidence, the expansion of our data basis is essential.

One current severe problem regarding the trivial AGL algorithm is false detection of traffic lights. That is because cars stop at various locations waiting for the same traffic light to turn to green. But the current approach will detect multiple traffic lights at all these positions. Thus, we planned to extend the algorithm to a weighted algorithm. The weight corresponds to a statistic confidence whether there is a traffic light or not. We defined a set of rules, which manipulate this weight. Every traffic light, found by the trivial algorithm, will be examined regarding the predefined rules and gets a weight as a result of that.

Another very fundamental and important approach to improve the speed advices is based on Car-to-Car or Car-to-Infrastructure communication.

Thereby, forecasting improvements could be achieved through considerations of traffic densities and traffic states at locations different than the one monitored by the relevant car.

In our current solution all participants use a central server to deliver their collected data or to get new data sets. That requires a scalable infrastructure. Moreover, all users carry a lot of information and the portable devices have to try to match their position to the prepared historic data and the effort increases with the amount of available data. If we would implement an adequate peer-to-peer technique based on wireless Car-to-Car Communication, the devices would share their knowledge on the fly when another participant gets in their range, which would also lead to a preference of local data and smaller datasets.

The list of potential benefits could be continued. In the long run, we will focus our research on the AGL project and on approaches considering Car-to-Car or Car-to-Infrastructure capabilities.

7. References

[1] Dresner, K., P. Stone, "Multiagent Traffic Management: A Reservation-Based Intersection Control Mechanism", *Proceedings of the Third international Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2* (New York, New York, July 19 - 23, 2004). International Conference on Autonomous Agents. IEEE Computer Society, Washington, DC, 530-537.

[2] T. S. Letia, "Real-Time Approaches of Urban Vehicle Traffic Control", *Automation, Quality and Testing, Robotics, 2006 IEEE International Conference on* , vol.1, no., pp.346-350, May 2006

[3] Ahmed, S.A., Cook, A.R., "Applications of time series analysis techniques to freeway incident detection", *Transportation Research Board Record 841*, Transportation Research Board, 1982, pp. 19—21.

[4] Whittaker, J., Garside, S., Lindveld, K., 1997. "Tracking and predicting a network traffic process," *International Journal of Forecasting*, Elsevier, vol. 13(1), pp. 51—61

[5] Camilleri, M., "Forecasting Using Non-Linear Techniques in Time Series Analysis: An Overview of Techniques and Main Issues", *Proceedings of Computer Science Annual Research Workshop*, 2004, pp. 19—29

[6] DeCicco, J., Ross, M., "Recent Advances in Automotive Technology and the Cost-Effectiveness of Fuel Economy Improvement", Transportation Research Part D: Transport and Environment Vol. 1 No. 2, 1996, pp. 79-96

[7] Yamashita, T., Izumi, K., Kurumatani, K., and Nakashima, H. 2005. Smooth traffic flow with a cooperative car navigation system. In *Proceedings of the Fourth international Joint Conference on Autonomous Agents and Multiagent Systems* (The Netherlands, July 25 - 29, 2005). AAMAS '05. ACM, New York, NY, 478-485.

[8] Waters, M.H.L., and Laker, I.B., 1980, "Researchon fuel conservation for cars", TRRL-Report 921,1980

[9] Voort, M., Dougherty, M. S., Maarseveen, M., "A Prototype Fuel-Efficiency Support Tool", Transportation Research Part C: Emerging Technologies, Volume 9, Issue 4, August 2001, Pages 279-296

A Panorama on Vehicular Networks Security

Christian Tchepnda[1], Hassnaa Moustafa[1], Houda Labiod[2], Gilles Bourdon[1]
[1]Orange Labs - France Telecom Group
38-40 rue du Général Leclerc, 92794 Issy-les-Moulineaux Cedex 9 – France
{christian.tchepnda, hassnaa.moustafa, gilles.bourdon}@orange-ftgroup.com
[2]Institut Telecom – Telecom ParisTech
46 rue Barrault, 75634 Paris Cedex 13 – France
houda.labiod@telecom-paristech.fr

Abstract— **This paper provides a panorama on the security in vehicular networks' environments. The special characteristics of these environments are presented and a general classification for the different types of attacks illustrated by some relevant attacks examples is introduced. Consequently, some key security requirements and security challenges are derived, considering ITS (Intelligent Transportation System) or Safety services as well as non-ITS or non-Safety services. Finally, some existing contributions in this subject are presented, and their deployment feasibility is discussed.**

Index Terms— **Security, Vehicular networks, Safety services, non-Safety services.**

I. Introduction

Vehicular communication is an emerging class of mobile communication enabling mobile users in their vehicles to communicate to the road and to each other. Currently, Inter-Vehicle Communication Systems (IVCS) are widely discussed, attracting considerable attention from the research community as well as the automotive industry. These networks have special behavior and characteristics, distinguishing them from other types of ad-hoc networks. The nodes' (vehicles') mobility in these networks is high and may reach up to 200Km/h, the network topology is dynamic but constrained by roads' topology, these networks may scale to a very large number of nodes (vehicles) according to the traffic condition and finally these networks probably have a potentially heterogeneous administration. In this context, we consider that vehicular communication, opposing the wireless mobile communication, does not suffer from resource limitations (energy, CPU, memory, etc.) as vehicles are not tiny nodes and are capable of providing large resources.

Although, vehicular networks are considered as one of the promising concrete applications of ad hoc networks, their special behavior and characteristics create some communication challenges (for network operators and service providers), which can greatly impact the future deployment of these networks. An important research and development aspect in vehicular communication concerns the development of security mechanisms that allow trust among the communicating parties (whether vehicles or infrastructure elements) and guarantee only authorized users' access to network resources and services offered by the provider as well as secure data transfer. In fact, security requirements differ according to the type of applications, where different security levels are needed. We notice that vehicular communication security is a young research domain, showing few contributions and lacking concrete security solutions.

This paper gives a panorama on vehicular communication security. We discuss the different types of attacks in this environment and present from a deployment perspective the security requirements that should be satisfied, taking into consideration safety applications and commercial applications expected in the future. The remainder of this paper is organized as follows: Section 2 figures out services and potential architectures in vehicular networks. Section 3 presents a classification for potential attacks giving some attacks examples. Section 4 introduces our view of the main security requirements and security challenges for vehicular communications deployment. In Section 5, we give an overview on the related work discussing the main contributions and mentioning their limitations as well as some open issues. Finally, the paper is concluded in Section 6.

II. Services and architectures

A. Services

Vehicular communications are expected to provide a wide set of useful services to drivers and passengers. We classify these services into two main classes: i) ITS (Intelligent Transportation System) or Safety services, ii) non-ITS or non-Safety services.

ITS was the main objective in the emergence of vehicular communications, where the primary works aimed at providing ITS solutions. ITS target is to minimize accidents and improve traffic conditions through providing drivers and passengers with useful information, e.g. road conditions alarms, congestions alarms, fire alarms, accident-ahead warnings, speed limit reminder, and traffic messages' exchange that is useful in avoiding collision at intersection, optimizing traffic flows, and avoiding crash situations.

On the other hand, non-ITS services aim at providing commercial, leisure and convenience services. Non-ITS services have taken recent attention in vehicular communications, being a target of some recent research contributions in this domain. Such services should guarantee data transfer between vehicles and are expected to provide passengers and drivers with Internet connections facility exploiting an available infrastructure in an "on-demand" fashion. Examples of useful non-ITS services are: electronic tolling system, multimedia services (e.g. interactive video games, VoIP, streaming, etc.), web browsing, email access, file sharing, and discovery of local services in the neighborhood (e.g. restaurants, bars, movie theatres, etc.).

We notice that the messages' dissemination in vehicular networks depends on the type of provided services, where diffusion (broadcast) seems more relevant for ITS or Safety-related services.

B. Architectures

Vehicular communications are expected to take place in urban zones, rural zones and highways through providing some network functionalities, protocols and integration strategies for services' delivery to users. Vehicular networks can be provided by network operators, service providers or through integration between operators, providers and a governmental authority. The recent advances in wireless technology as well as the new ad hoc scenarios defined within the IETF allow several possible vehicular network architectures. Three deployment alternatives, that could be mutual, include: i) a pure wireless Vehicle-to-Vehicle ad hoc network (V2V) allowing standalone vehicular communication with no infrastructure support, ii) a wired backbone with wireless last hops, iii) and a hybrid Vehicle-to-Road (V2R) architecture that does not rely on a fixed infrastructure in a constant manner, but can exploit it for improved performance and service access when it is available. In addition, the Car-to-Car Communication Consortium (C2C-CC) specified some architectural considerations for vehicular networks deployment, including: i) Road-Side Units (RSUs) existing along the road, and ii) vehicle equipment with an On Board Unit (OBU), and potentially multiple Application Units (AUs) executing a single or a set of applications while using the OBU communication capabilities. Vehicles' OBUs and RSUs can form ad hoc networks, where communication can be: i) V2V taking place directly between OBUs via multi-hop or single-hop without involving any RSU, or ii) Vehicle-to-Infrastructure (V2I), in which OBUs communicate with RSUs in order to connect to the infrastructure.

III. Attacks in vehicular networks

In this section we discuss the most relevant vehicular communications attacks through providing a classification and some concrete examples.

A. Attacks classification

An important key in securing vehicular communications is determining the types of attacks threatening the communication in such environment. Different types of attacks may exist according to the type of environment as well as the usage scenario. Based on the taxonomy provided in [6] with few enhancements, these attacks are classified as follows:

- *Internal or External*: An internal attack can be mounted by an authenticated member of the vehicular network. In other words, this member is identified by other members as a legitimate member. This type of attack is probably the most critical one. On the other hand, an external attack can be mounted by a non authenticated entity that is hence considered as an intruder by legitimate members. Unlike an internal attacker, an external attacker is limited in the diversity of attacks he can mount.

- *Intentional or Unintentional*: An intentional attack is mounted by an entity aiming voluntarily to disrupt the network operation. Conversely an unintentional attack is mostly due to potential transmission or network operation errors.

- *Active or Passive*: An active attack is mounted by an attacker who generates or modifies the network traffic. In contrast, a passive attack is mounted by an attacker who will only eavesdrop the wireless channel for later unauthorized use.

- *Independent or Coordinated*: An independent attack is caused by a unique attacker whereas a coordinated attack is caused by a group of attackers sharing the same interest.

B. Attacks examples

Since it is hardly possible to give an exhaustive list encompassing all attacks in vehicular networks, we have rather chosen to derive a general classification of attacks. In the following subsections, we only introduce some important and relevant attacks examples that have been identified.

- *Privacy attack:* This is considered as the Big Brother case where an attacker actively monitors the network traffic in order to disclose vehicles identities and trace them. Some examples of traceable identifiers are IP addresses, MAC addresses, certificates IDs, etc. Even if individual messages do not contain such identifiers, a particular string that does not in itself identify the vehicle could appear in a series of messages. If an attacker ever unambiguously observes the vehicle emitting that string, they can use the string as an identifier for the vehicle. Moreover, using radio fingerprinting (*i.e.* a physical layer attack) is another way to identify and trace vehicles. Figure 1 illustrates an identity disclosure example. Considering the attacks classification given in the previous subsection, this attack is characterized as: (*Internal or External, Intentional, Passive, and Independent*).

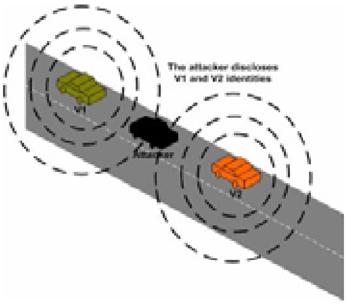


Figure 1: Identity disclosure

- *Information inconsistency:* In this case the attacker injects wrong information in the network in order to affect the behavior of other vehicles. An example scenario is a vehicle cheating with positioning information in order to alter its perceived position, speed, direction, itinerary, etc. By doing so, the attacker can divert the traffic from a given road and hence free that road for himself. Figure 2 and Figure 3 illustrate this case. In Figure 2, the attacker diffuses bogus traffic information and in Figure 3, we have some attackers cheating with positioning information. In Figure 2 the attack is characterized as: (Internal, Intentional, Active, and Independent) whereas in Figure 3 the attack is characterized as: (*Internal, Intentional, Active and Coordinated*).

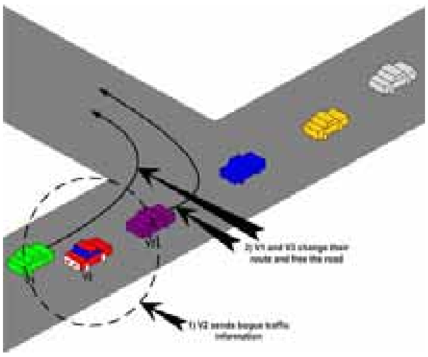

V3
V1
2) V1 and V3 change their route and free the road
1) V2 sends bogus traffic information

Figure 2: Bogus traffic information injection

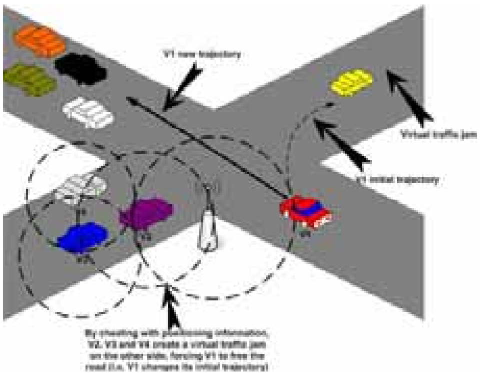

V1 new trajectory
Virtual traffic jam
V1 initial trajectory
V3
V2
V4
V1
By cheating with positioning information, V2, V3 and V4 create a virtual traffic jam on the other side, forcing V1 to free the road (i.e. V1 changes its initial trajectory)

Figure 3: Cheating with positioning information

• *Impersonation/Masquerading:* In this case, the attacker uses a false identity pretending to be another vehicle. More generally, the attacker uses a false credential in order to be granted another vehicle privileges. This specific attack example is characterized as: (*Internal or External, Intentional, Active and Independent*).

• *Denial of Service (DoS):* In this case the attacker prevents legitimate vehicles to access the network services. This can be done by jamming the wireless channel, overloading the network or having a non cooperative behavior (*e.g.* dropping packets). Figure 4 illustrates an attack example where the attacker causes a crash by preventing legitimate vehicles to exchange critical traffic information. This attack is characterized as: (*Internal or External, Intentional, Active and Independent or Coordinated*).

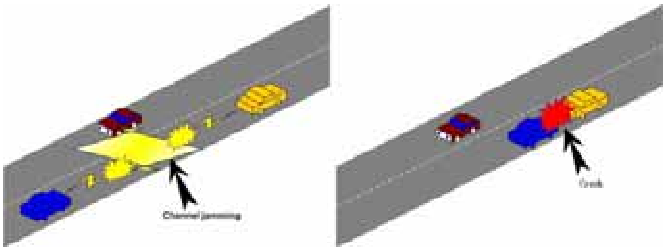


Figure 4: DoS - Channel jamming

• *Eavesdropping:* In this case, the attacker monitors the network traffic in order to extract any sensitive information. Figure 5 shows the case where an attacker eavesdrops and extracts a commercial transaction password. This attack is characterized as: (*Internal or External, Intentional, Passive and Independent*).

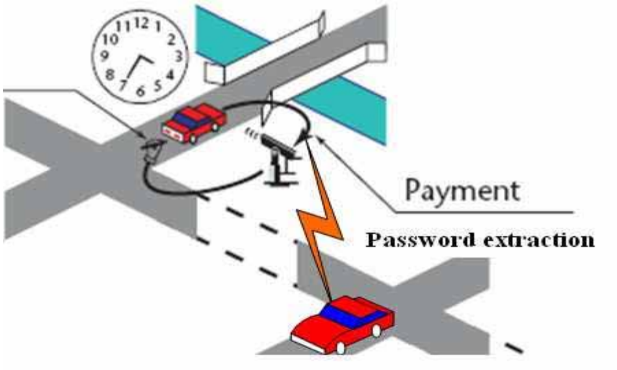


Figure 5: Eavesdropping a commercial transaction

IV. Security requirements and security challenges

In this section, we show the main security requirements and security challenges for vehicular communications considering both ITS and non-ITS services. These security requirements and security challenges should be considered during the design of the network architecture, security protocols, cryptographic algorithms as well as software and hardware architectures.

• *Confidentiality:* We believe that ITS services are directly related to people physical safety. If a vehicle that is not authenticated (*i.e.* an illegitimate vehicle) is in accident, authenticated vehicles (*i.e.* legitimate vehicles) may also be affected. Consequently, to efficiently achieve the main goal of ITS services which is mainly peoples' safety, every vehicle (legitimate and illegitimate) should be capable of receiving and processing ITS data. Thus, unlike non-ITS services which may require confidentiality due to their commercial nature, ITS services should not require protection against eavesdropping.

• *Source authentication:* In ITS services, the only restriction lies in avoiding illegitimate vehicles to generate ITS data. This is simply implemented through discarding ITS data not having a proof of the authenticity of their sources (*e.g.* a digital signature). We notice that, this security requirement appears more relevant for ITS services as we consider only broadcast communication for these services.

- *Mutual authentication, authorization and access control:* The commercial or transactional nature of non-ITS services requires them to support mutual authentication between each client (vehicle) and the service provider (or the network operator) on one hand and between each communicating vehicles on the other hand. Mutual authentication aims at preventing attacks such as Man In The Middle (MITM) attack.

A simple solution to carryout authentication in such environment is to employ a symmetric key shared by all nodes in the network. Although this mechanism is considered as a *plug and play* solution and requires less processing and communication overhead, it is limited to closed scenarios of small number of vehicles, mostly belonging to the same provider. For wide scale commercial deployment of vehicular networks, the symmetric group key authentication has two main pitfalls: firstly, an attacker only needs to compromise one node (vehicle) to break the security of the system and paralyze the entire network. Secondly, mobile nodes (vehicles) can impersonate each other and can access each other messages breaking the non-repudiation and the confidentiality security requirements. As regards the symmetric pairwise key authentication, the main problem is its inherent key establishment non-scalability as the number of keys grows linearly with the number of vehicles. Hence, public key cryptography with few performance enhancements seems to be the way to go. Furthermore, authorization and access control are important counter-attack measures in vehicular networks deployment, allowing only authorized mobile nodes to be connected and preventing adversaries to sneak into the network disrupting the normal operation or service provision.

- *Non repudiation:* Vehicles causing accidents or injecting malicious data must be reliably identified. Hence, a vehicle should not be able to deny the transmission of a message. If used carefully, digital signatures can provide the non-repudiation property. However, the main reason for using signatures is not to provide non-repudiation but to allow authentication between two entities who have not previously encountered each other, without having to make an online query to a third party. Although non repudiation security requirement is mostly critical for ITS services, it may be desirable to also have it for some sensitive non-ITS services especially those involving online payments.

- *Privacy:* As people increasingly worry about the Big Brother enabling technologies, private individuals' anonymity and non-traceability should be guaranteed. It should be noted however that anonymity and non-traceability are strictly conditional as non-repudiation must also be enforced. Although traceability is a legitimate process for some governmental authorities and networks operators, the non-traceability is an important security requirement in order to assure peoples' privacy. Thus a complex problem arises in this issue. In fact, a tough requirement in vehicular networks environments is to manage traceability in terms of allowing this process for the concerned authorities and at the same time assuring the non-traceability between mobile clients (vehicles) themselves. Nevertheless, the latter is difficult to achieve and so far no promising solutions exist to resolve this issue in the vehicular networks dynamic and open environment. It is noticed that the traceability can include: i) who is talking to who, ii) what one is sending, iii) which site one is accessing or which application one is using, and iv) where is the mobile client now (his location) and where is he going to be after a while. In fact, the privacy security requirement applies to both ITS and non-ITS services.

- *Real-time constraints:* A critical feature in ITS services is their time sensitiveness, where ITS data are mostly real-time data with about 100ms critical transmission delay [1]. ITS services are expected to carry out much more signature verification than signature generation. So, an important challenge, achieving real-time constraints, is choosing the fastest public key cryptosystem in signature verification which at the same time performs well in signature generation. Also, the selected cryptosystem should be as compact as possible. In any case, any security mechanism for ITS services should take into consideration these real-time constraints.

- *Data Consistency / Liability:* This is also an important security issue for ITS services as even authenticated vehicles could become malicious by sending bogus information in order to gain an undue advantage, and thus can cause accidents or disturb the network operation. As a mechanism example ensuring data consistency or liability, we can quote verification by correlation which consists in correlating, through a reputation-based or a recommendation-based system, data received from a given source with those received from other sources. Some other approaches consist in enabling any node to search for possible explanations for the data it has collected based on the fact that malicious nodes may be present. Explanations that are consistent with the node's model of vehicular networks are scored and the node accepts the data as dictated by the highest scoring explanations [5].

- *Integrity:* This security requirement applies to ITS and non-ITS services as it protects against altering a message in transit. In practice, authenticity and integrity go together since there's no point correctly identifying a message origin if the message content is altered.

- *Availability:* Denial of Service (DoS) attacks due to channel jamming, network overloading or non cooperative behaviors may result in network unavailability. Hence, a continuous network operation should be supported by alternative means. The non-availability risk can be mitigated by exploring the numerous security mechanisms (*e.g.* monitoring, reputation-based systems, etc.) which can be used for non-cooperative nodes detection or by exploring channel and technology switching and cognitive radio techniques which can be used against jamming attacks. The availability security requirement applies to any vehicular service.

- *High mobility support:* This is a crucial challenge in designing vehicular networks security system. We assume that the vehicles' computing platforms have the same computational capability and energy supply as wired clients such as desktop PCs. However they significantly differ in their mobility support and their resulting throughput capability. These factors result in a mismatch between security protocols execution time in vehicular networks and their execution time in wired networks. This security execution gap is an important issue that must be faced by vehicular network designers. An attempt to lower this gap is making vehicular security protocols and their inherent cryptographic algorithms, lightweight and fast (without loosing security robustness). For instance, it has been demonstrated that by implementing only a subset of security protocols' features, it is possible to reduce the overhead and the execution time. This goal can also be achieved by selecting optimal software or hardware implementations for cryptographic algorithms and adapting the encryption policies based on the content of the data that is being encrypted (*e.g.* video encryption).

An important scheme reducing the execution time and adapting high mobility is the low complexity security algorithms. For example, the current security protocols such as SSL/TLS, DTLS, WTLS, etc. generally use RSA-based public key cryptography for authentication. The security of the basic RSA algorithm is derived from the integer factorization which is NP-hard. Hence, RSA can provide high security if the modulus is a large integer (*e.g.* 1024 or 2048 bits) whose factoring is extremely complex. This means that the basic computation for decrypting data is performed using large keys, making it computationally and time expensive. We can take advantage of alternative public key cryptography standards (PKCS) that provide security robustness while requiring less execution time. Elliptic Curve Cryptosystems (ECC) and lattice-based cryptosystem NTRU are examples of such alternative public key cryptosystems that are increasingly being used in wireless security software toolkits. For bulk data encryption/decryption, a protocol such as AES is preferred since older ones like DES or 3DES appear less attractive due to security limitations or computational and time expensiveness.

An additional way helping to lower the security execution gap, is to carefully choose the transport layer over which security protocols are implemented when securing transactions over IP. For example, TLS which secures application-layer traffic over TCP/IP is discouraged for mobile use as it operates over TCP. Conversely, DTLS which secures application-layer traffic over UDP/IP is better accepted as it operates over a connectionless transport layer (*i.e.* UDP). A protocol like IPSec which secures IP traffic, should be avoided as its secure connections (known as Security Associations or SAs) are cumbersome to set up, requiring too many messages. However, when vehicles are not in motion (*e.g.* in a parking space), protocols like IPsec or TLS might become appropriate.

Although high mobility support is a global challenge for any vehicular service security, we however conclude that ITS services security requirements are a little bit different from non-ITS services security requirements. In Figure 6, we highlight the main security requirements of both services, showing the commonality and the difference between them.

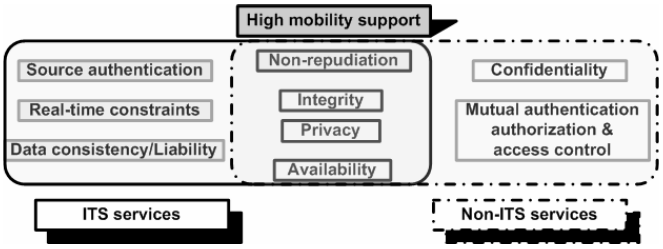


Figure 6: Main security requirements and security challenges for ITS and non-ITS services

V. ANALYSIS OF RELATED WORKS

A. Main contributions

Vehicular communications security is a young research domain, showing few research contributions and lacking real solutions development. The existing contributions mainly focus on securing ITS services without providing general security architectures considering both ITS and non-ITS services.

In [2], Zarki et al. address ITS services security in vehicular environment. They consider that there is no routing and no hand-over in their scenario as they use a broadcasting communication model that is one way with the Base Station (BS). They show that through providing efficient PKI and digital signatures mechanisms, the confidentiality and key distribution do not need to be considered. These mechanisms are required to have acceptable delay with respect to the ITS services real-time constraints. Since this contribution does not support privacy issues, and more importantly does not consider non-ITS services security requirements, this limits the possibility of it's wide scale deployment, as ITS services show far less potential compared to non-ITS services when it comes to the possibility of services' commercialization. Privacy issues for ITS services are more specifically tackled in [3]. The authors introduce the concept of entropy anonymity metric, recommending the use of PKI and some location verification tools to implement ITS security. The main advantage of this work is that privacy mechanisms which are introduced can also be considered for non-ITS services. In [4], a secure communication architecture for Inter Vehicles Communication (IVC) network is presented which comprise a PKI, a distributed IDS (Intrusion Detection System) and a virtual network infrastructure. Although this work is enough general for securing ITS and non-ITS services, it is extremely restrictive in considering a stand-alone vehicular ad-hoc network which does not show any great interest for the envisioned business model of network operators where services delivery are mainly done through access points or base stations. Golle et al. propose a general approach to evaluate the validity of ITS data in vehicular networks, focusing on artificial intelligence issues in such environment [5]. A more network-oriented work introducing a threat analysis and a security architecture for ITS services is presented in by Raya et al. in [6] and [10]. The authors provide a set of security protocols making use of PKI, digital signatures, anonymous public keys implementing privacy and secure positioning among other points. Nevertheless, the security architecture is once again strictly limited to ITS services. An approach considering a real deployable solution from a network operator perspective is presented in [7]. This solution aims at securing ITS and non-

ITS services based on EAP-Kerberos using public key certificates at entry points of highways. Actually this solution needs to support the continuous access to infrastructure services, as vehicles must always re-authenticate whenever they change their Access Points (APs). To enhance vehicular networks ubiquitous secure access from a network operator perspective, a novel architecture and security mechanisms are proposed in [9] taking advantages of: i) the ad hoc multi-hop authentication concept, ii) the smart card-based authentication allowing authentication before the V2R communication, and iii) the grid paradigm for security resources' aggregation. In the Now (Network On Wheels) project context, Matthias Gerlach et al. [11] introduce a security architecture integrating existing individual solutions for vehicle registration, data integrity, data assessment, authentication, pseudonyms, certification, revocation and so on. They distinguish in their architecture the high-level views i.e. the functional layers and the organizational structure describing how the overall security system should look like and the implementation-near views describing an implementation design for the security system in the vehicle's on-board unit and presenting the information flow among the architecture components. In [12], the authors depict a trusted authority architecture including a governmental authority and private authorities aiming to offer reliable communications on top of which services can be deployed. These authorities are in charge of providing pseudonyms, key distribution and key management for on-demand requesters namely users and services providers. Single-hop and multi-hop secure service provisioning scenarios are illustrated although not sufficiently detailed to be thoroughly analysed.

We had in the past many industrial projects on vehicular communications in Europe, Asia and USA (e.g. CarTalk, Chauffeur1, Chauffeur2, Fleenet, PATH, etc.) but unfortunately these projects didn't take security issues into consideration. However some efforts are made to address this limitation through ongoing projects or consortiums such as Car2Car Communication Consortium (C2C-CC), SEVECOM project (SEcure VEhicular COMmunications), NoW project and the DSRC-based IEEE P1609.2 project [8] aiming to specify safety applications security mechanisms based on public key cryptography.

B. Summary and outlook

From our investigation to the existing contributions, we notice that most of them focus on ITS services, neglecting the tremendous potential of non-ITS services in driving the development and deployment of vehicular networks. Although extensively tackling security of ITS services, these contributions fail in considering some peculiarities of such services like broadcast transmission and its resulting effect on security requirements. All these limitations fuel the motivation in providing a broader and a more complete view of vehicular communications security challenges.

In summary the future security mechanisms of vehicular networks should comply with the special characteristics of these networks *i.e.* high mobility, constrained bandwidth, large resources (CPU, energy, memory), dynamic topologies constrained by roads' topologies, large number of nodes, heterogeneous administration, etc. Moreover, these security mechanisms should be transparent for non-ITS applications (serving network operators and services provides business needs) allowing vehicles not to re-authenticate from scratch whenever they move. Overall, the main open issue here lies in designing multi-purpose security architectures complying with vehicular networks characteristics and meeting all the respective requirements of ITS and non-ITS services in a single framework. This type of framework can be depicted as a context-aware or a service-aware security framework for vehicular communications.

VI. Conclusion

Security is one of the significant challenges impacting mobile ad-hoc networks and vehicular networks specifically. A point that complicates this issue is that securing vehicular communication is service-related. For instance, safety-related services should be granted to every vehicle on the road while assuring the secure messages' transfer. In this paper, we illustrate vehicular networks applications and potential architectures, addressing the main security requirements and challenges in this specific environment. We noticed that the existing contributions are not general for all vehicular networks' environments and are restricted mostly to ITS services. Moreover these contributions hardly comply with the main security requirements that are important in designing security solutions for real deployments of vehicular networks. Thus, those solutions can restrict vehicular networks potential, ignoring its importance in providing ubiquitous communications. Consequently, appropriate security mechanisms are required allowing trust and secure transmission, while taking into consideration the dynamic and none fully centralized nature of vehicular networks as well as the different types of applications and their time-sensitivity.

References

[1] Xue Yang, Jie Liu, Feng Zhao and Nitin Vaidya, "A vehicle-to-vehicle communication protocol for cooperative collision warning", MobiQuitous, August 2004

[2] Magda El Zarki, Sharad Mehrotra, Gene Tsudik and Nalini Venkatasubramanian, "Security Issues in a Future Vehicular Network", EuropeanWireless, 2002

[3] Jean-Pierre Hubeaux, Srdjan, Capkun and Jun Luo, "The Security and Privacy of Smart Vehicles", IEEE Computer Society, 2004

[4] Jeremy Blum, Azim Eskandarian, "The Threat of Intelligent Collisions", IEEE Computer Society, 2004

[5] Philippe Golle, Dan Greene and Jessica Staddon, "Detecting and Correcting Malicious Data in VANETs", ACM VANET, October 2004

[6] Maxim Raya and Jean-Pierre Hubaux, "The Security of Vehicular Ad Hoc Networks", ACM SASAN, 2005

[7] Hasnaa Moustafa, Gilles Bourdon and Yvon Gourhant, "Providing Authentication and Access Control in Vehicular Network Environment", IFIP SEC, 2006

[8] IEEE P1609.2 - Standard for Wireless Access in Vehicular Environments - Security Services for Applications and Management Messages, 2006

[9] C. Tchepnda, H. Moustafa, H. Labiod and G. Bourdon, "Securing Vehicular Communications: An Architectural Solution Providing a Trust Infrastructure, Authentication, Access Control and Secure Data Transfer," IEEE AutoNet - Globecom, 2006

[10] Maxim Raya and Jean-Pierre Hubaux, "Securing Vehicular Ad Hoc Networks", Journal of Computer Security (JCS) - special issue on Security on Ad Hoc and Sensor Networks, January 2007

[11] M. Gerlach, A. Festag, T. Leinmüller, G. Goldacker and C. Harsch, "Security Architecture for Vehicular Communication", International Workshop on Intelligent Transportation (WIT), 2007

[12] Etienne Coronado and Soumaya Cherkaoui, "Secure Service Provisioning for Vehicular Networks", UBIROADS, 2007

Efficient Memory Integrity Protection for Vehicular Networks

Giovanni Di Crescenzo*

Abstract

As vehicle-based networks approach practicality and expand their service capabilities and interoperability, a number of questions on guaranteeing the security of transactions of vehicles or devices take place in it. Contrarily to other types of networks, most of the software involved in these transactions, is expected to be run on clients that remain subject to long and uncontrolled hacking activity. Thus, guaranteeing the integrity of this software and data becomes crucial for the success of these transactions. In this paper we model and investigate the problem of checking the integrity of client software via a client-server architecture, where vehicles are clients or other devices and servers are run by road-side equipment. In this model, we consider the problem of designing and analyzing efficient cryptographic techniques to speed up integrity verification and detection of integrity corruptions. We design three memory integrity protection schemes based on non-trivial applications of cryptographic hashing, that attempt to minimize an ordered sequence of metrics, including communication complexity, round complexity and client's running time. All three schemes give different and significant improvements over the standard strategy of using one cryptographic hash for each monitored file.

1 Introduction

Vehicular networks are making their entrance in everyday human life, enhancing safety of the driving experience and increasing the amount of commercial transactions engaged by vehicle drivers. In such networks, at any given time, vehicles may take part into a number of more or less complex protocols, including, for instance, the collection, processing and forwarding of traffic-related or personal data. Here, vehicles are assisted by a number of on-board sensors and interact with other vehicles as well as trusted road-side servers. For instance, in a traffic-safety transaction, vehicle software will combine data received from on-board sensors (e.g., about time, vehicle speed, position) and combine it into some meaningfully represented information to be sent to other vehicles or to the road-side servers. The latter will combine information received from al vehicles with information received by their own environment's sensors to issue, for instance, traffic warnings to vehicle clients. While servers are running these protocols, it is of crucial importance for servers to check the integrity of the clients' programs, so to avoid that a malicious entity, even given long and uncontrolled access to these programs, can modify them to generate attacks or improper behavior or even seemingly proper behavior that misleads road-side servers into false traffic conclusions. This is especially true in the case of interoperable vehicles that are subject to interactions with a number of foreign devices that can attempt to implant viruses in their software clients. Although privacy and security for vehicular networks is a rapidly emergent research area which we do not review here (but see, e.g., [2]), virus protection questions have so far been neglected.

Verifying the integrity of computer programs and detecting specific integrity violations are hard problems with non-trivial security and efficiency constraints. As it remains unclear whether hardware-based solutions (e.g., a tamper-proof coprocessor that monitors and checks integrity of the vehicle's programs [7, 6]) may be a financially viable solution in vehicular networks, it is natural to investigate software-only solutions. Specifically, such networks exhibit a client-server scenario where vehicle clients, possibly under constant attack, communicate with road-side servers that retain a copy of the vehicle's programs. Here, the natural adaptation of the security literature's integrity verification procedure consists of the client showing some 'proof of integrity' to the server (e.g., a cryptographic hash or digital signature of the vehicle's programs). This verification process is trusted to be successful under the trust assumption on the server itself (see, e.g. [3, 5] and references therein). In this paper, we investigate improvements of the basic cryptographic hashing/signing integrity verification technique for vehicular networks, in terms of various performance metrics.

*Telcordia Technologies, Piscataway, NJ, 08854, USA. E-mail: giovanni@research.telcordia.com

Our contributions and related areas. We start by presenting a cryptographic model for memory integrity protection in vehicular networks' client-server scenario. In this model we consider the problem of designing and analyzing efficient cryptographic hashing techniques to speed up memory integrity verification and detection of memory integrity corruptions. Our definitional approach is that of allowing and extending techniques such as the basic cryptographic hashing/signing technique so to meet various security and efficiency constraints. Contrarily to previous fundamental work on (single-computer) memory integrity protection also known as the area of memory checking [1], which our model builds on, we focus the costruction of schemes on certain efficiency metrics that are of greater importance for client-server models such as those arising in vehicular networks. In particular, we focus on minimizing the communication and round complexity between each vehicle client and the road-side trusted server, while paying attention to not significantly increase the client's running time. We design three memory integrity protection schemes based on non-trivial applications of cryptographic hashing that attempt to minimize these performance metrics. All three schemes significantly improve over the popular strategy of using one cryptographic hash or signature for each monitored file (see, e.g. [8]).

Due to space restrictions, proofs and some formal definitions, theorems and protocol descriptions are omitted.

2 Model and Definitions

We present a formal model for memory integrity protection for vehicular networks, including three main requirements: correctness, security, and performance.

Model and basic notations. We will consider memory integrity protection schemes addressing a *client*'s goal of guaranteeing that any unauthorized changes inserted by some *adversary* into its memory are detected within a reasonably short amount of time. We use $\mathrm{C}, \mathrm{S}, Adv$ as shorthands for vehicle client, road-side server, and adversary. To simplify our formal treatment, critical programs and data to be protected on the client are modeled with a generic *client memory* containing a collection K of n files $K_1, \ldots, K_n$; here, every file will be considered as an atomic quantity. In this model, both C and Adv are given read and write access to K, C is given only read access to K, and S can communicate with C. The adversary Adv is given read and write access to K but not to S's memory. We also assume that the adversary cannot modify the content of the information exchanged through the communication line (this assumption is only made to simplify our formal treatment as it could be removed by having both C and S use standard cryptographic tools such as message authentication codes). S's storage contains an *authenticating area* where data is stored there only to protect the integrity of C's memory.

Scheme syntax and scenario. A memory integrity protection scheme in the client-server model may use a tuple of parameters $\vec{p} = (\lambda, n, m)$ and can be formally defined as a triple $\mathrm{MP} = (\textsc{tag}, \textsc{upd}, \textsc{find})$, where TAG is a *tagging* algorithm, run by the trusted server S, UPD is an *update* algorithm, run by the client C, and FIND is a *corruption finding* protocol, run by the client C and the trusted server S. Let λ be a security parameter represented in unary notation; these three algorithms/protocols have the following syntax. On input 1^λ, an n-file storage area K, algorithm TAG returns an m-bit output *tag* (representing the authenticating tags that will be used by S to detect any future corruptions). On input 1^λ, an n-file storage area K, an authenticating string *tag* and an index $j \in \{1, \ldots, n\}$, algorithm UPD returns an output tag' of length at most m (representing the update that C sends to S as a result of an authorized update from U to K). In protocol FIND, party C takes as input $(1^\lambda, K)$, and party S takes as input $(1^\lambda, tag)$; at the end of the protocol, S returns: $\perp$ (meaning that no file in K was corrupted), or $V \subseteq \{1, \ldots, n\}$ (meaning that all files in $\{K_i | i \in V\}$ were corrupted).

We will consider a scenario that consists of of a sequence of the following three events, which possibly trigger the execution of the algorithms TAG, UPD and of protocol FIND. First, in a setup phase, when it is assumed that the adversary is not active yet, a *tag computation event* happens, where the trusted server runs the tagging algorithm on input K, λ and produces the authenticating values *tag*. Afterwards, the following two events may happen. On a request by either S or C, a *tag update event* happens, where authorized modifications are performed to K, possibly including updates to parameters n, m. Moreover, a *corruption finding event* happens periodically or on S's request, where C and S run protocol FIND to check whether any files in K have been corrupted by Adv.

Correctness requirement. A basic requirement that we expect from a memory integrity protection scheme is that, at any time, uncorrupted files are almost always recognized to be so at the end of the corruption finding protocol, even after an arbitrarily large number of tag update events.

Definition 2.1 Let MP $= (\textsc{tag}, \textsc{find}, \textsc{upd})$ be a memory integrity protection scheme with parameters $\vec{p} = (\lambda, m, n)$. The *correctness* requirement for MP is as follows: for any n-file storage area K, the following event happens with probability negligible (in λ):

1. $tag \leftarrow \textsc{tag}(1^\lambda, K)$
2. $\{tag \leftarrow \textsc{upd}(1^\lambda, j_i, K, tag)\}_{i=1}^{\ell}$
3. $out \leftarrow \textsc{find}((1^\lambda, K); (1^\lambda; tag))\ :\ out \neq \perp$.

Security requirement. We would like a memory integrity protection scheme to be able to find out corrupted files in the storage area K with sufficiently high probability, regardless of how many files were corrupted by an adversary. More specifically, our model considers an adversary *Adv* that is, for sake of generality of results, *not* time-bounded. (However, our corruption finding requirement is defined so that a formalization for the particular case of a polynomial-time bounded *Adv* can follow with minimal modifications). In its attack experiment, after S has generated its authenticating tags, *Adv* is given read access to the storage area K and can choose to modify up to all files in it. Now, using protocol FIND, S and C try to find all corrupted files. We are now ready to define the security requirement for any memory integrity protection scheme.

Definition 2.2 Let MP $= (\textsc{tag}, \textsc{upd}, \textsc{find})$ be a memory integrity protection scheme in the client-server model with parameters $\vec{p} = (\lambda, n, m)$. We say that MP is (t, ϵ)*-secure* if for any algorithm *Adv*, and any n-file storage area K, it holds that

$$\text{Prob}\left[b \leftarrow \mathsf{Exp}^{\text{MP},Adv}(\vec{p}, t, K)\ :\ b = 1 \right] \leq \epsilon,$$

where experiment $\mathsf{Exp}^{\text{MP},Adv}$ is defined as follows (here, the notation $y \leftarrow Alg(x_1, x_2, \ldots)$ denotes the process of running the possibly probabilistic algorithm Alg on input $x_1, x_2, \ldots$ and the necessary random coins, and obtaining y as output):

```
$\mathsf{Exp}^{\text{MP},Adv}(\vec{p}, t, K)$
1. $tag \leftarrow \textsc{tag}(1^\lambda, K)$
2. $(K'_{i_1}, \ldots, K'_{i_t}) \leftarrow Adv(\vec{p}, t, K)$
3. $K_i = K'_i$ for $i = i_1, \ldots, i_t$
4. $out \leftarrow \textsc{find}((1^\lambda, K), (1^\lambda, tag))$
5. if $out \neq \{i_1, \ldots, i_t\}$ then
       return: 1 else return: 0.
```

In the rest of the paper we target the construction of (t, ϵ)-secure schemes, for any arbitrary $t \in \{1, \ldots, n\}$ and ϵ equal to 0 or negligible in λ.

Performance requirements. Designing memory integrity protection schemes that find all t corrupted files is actually easy if one disregards performance metrics. For instance, consider the well-known approach, based on cryptographic hashing, defiend as follows. First, the tagging algorithm computes a cryptographic hash tag for each file in K; then, protocol FIND consists of C sending all n hash tags to S, and S returning the indices for which the received hash tag does not match the previously stored tag; finally, UPD returns a single cryptographic hash tag associated with the updated file. The obvious drawback for this construction is that the message sent from C to S is too long, as it is equal to the number of files in K times the size of a conventional hash from a collision intractable hash function. We call this scheme MP_0 and use it later as a benchmark to evaluate our improved schemes.

Indeed, quite a few performance metrics for a memory integrity protection scheme MP seem of interest in our client-server model. These include the *communication complexity* (resp., *round complexity*) defined as the maximum of the sum of the lengths (resp., the number) of the messages exchanged by C and S during any possible execution of protocol FIND; the *verification time*, defined as the maximum running time of C during an arbitrary execution of protocol FIND; the *update complexity*, defined as the maximum of the number of files in K that are updated as a result of an arbitrary execution of algorithm UPD; and the *storage complexity*, defined as the maximum length m of the output *tag* from any possible execution of algorithm TAG.

In vehicular networks the communication complexity and the round complexity seem the most crucial metrics. Thus, in the rest of the paper we focus on designing memory integrity protection schemes that minimize these two metrics first, while maintaining a good performance on verification time and update complexity, and not worrying much about storage complexity (as storage is today an extremely cheap resource and that typically integrity verification protocols only require a constant number of bits of protection per file).

3 A Scheme with Efficient Communication Complexity

In this section we consider the problem of designing memory integrity protection schemes that attempt to achieve minimal communication complexity, while maintaining a good performance on the update complexity. Our scheme is based on a variant of the widely used Merkle-tree hashing construction [4]. We achieve the following

Theorem 3.1 Let λ, n, m be positive integers, let $t \in \{1, \ldots, n\}$, and let ϵ be a function negligible in λ. Assuming the existence of a family of collision-intractable hash functions, there exists (constructively) a (t, ϵ)-secure memory integrity protection scheme $\mathrm{MP}_1 = (\textsc{tag}_1, \textsc{upd}_1, \textsc{find}_1)$ with parameters (λ, n, m) and with the following performance: $\Theta(s \cdot t \cdot \log(n/t))$ communication complexity, $\Theta(\log n)$ round complexity, $\Theta(n \cdot t(H_u))$ verification time, $\Theta(\log n)$ update complexity, and $\Theta(ns)$ storage complexity, where $s, t(H_u)$ denote the length of the output and running time, respectively, of the assumed hash function.

The communication complexity of scheme MP_1 is essentially optimal (up to a constant) in a model where every hash tag reveals one bit about the location of the t files modified by the adversary. The rest of this section is devoted to the proof of Theorem 3.1. We start with an informal description of the scheme and then give a formal description of the scheme and of the proof that the scheme satisfies the theorem.

An informal description of MP_1**.** The basic idea behind this scheme is that C and S should be able to collaboratively perform a procedure similar to a binary search of all t files modified by the adversaries in the storage area K. Then the tagging algorithm consists of not only computing one hash tag for each file in K, as done in MP_0, but also for contiguous hash tags, so to build a binary hash tree where the leaves are associated with hash tags of the files and the hash tag associated with an internal node is computed by hashing the hash tags associated with its two children. The update algorithm consists of updating the hash tag associated with the updated file at its tree leaf, and all hash tags on the path from this leaf to the root. The corruption finding protocol consists of C and S interacting so that C can search for the t files modified by the adversary with S's help, as follows. Starting from the root of the hash tree, C sends to S the hash tags associated with the two children, and S replies by saying which of the two are different from the version of these tags previously computed with the tagging algorithm. Then C continues the search with the node(s) for which S found different hash tags, if any.

Useful tools: Collision-Resistant Hash Functions and Merkle trees. Among the most basic cryptographic primitives, CR hash function families can be informally described as families of compression functions for which it is hard to compute two preimages of the same output. Starting from any collision-resistant hash function family, with hash functions H_u mapping 2ℓ-bit inputs to ℓ-bit outputs, Merkle defined in [4] the following tree-like construction to compress a polynomial number $m = 2^t$ of ℓ-bit strings $x_0, \ldots, x_{m-1}$ into a single ℓ-bit string y. The output $y = \mathrm{MT}(H_u; x_0, \ldots, x_{m-1})$ is recursively defined as

$$H_u(\mathrm{MT}(H_u; x_0, \ldots, x_{m/2-1}) \,|\, \mathrm{MT}(H_u; x_{m/2}, \ldots, x_{m-1})\,),$$

where $\mathrm{MT}(H_u; x) = x$, for any ℓ-bit string x. With respect to the tree analogy, note that the output y is associated with the root and the inputs $x_0, \ldots, x_{n-1}$ to the leaves. We extend this definition so that it also returns all intermediate hashes associated with the tree's internal nodes. Specifically, we specifically use denote the m-depth binary complete tree T and define the following 'generalized Merkle tree function'. The output $\{y_i \,|\, i \in T\} = \mathrm{GMT}(H_u; T; x_0, \ldots, x_{m-1})$ satisfies the equalities

- $y_i = \mathrm{MT}(H_u; x_0, \ldots, x_{m-1})$ if i is T's root;
- $y_i = H_u(y_{lc(i)} | y_{rc(i)})$ if i is an internal node of T, $lc(i)$ is i's left child, and $rc(i)$ is i's right child;
- $y_i = x_i$ if i is T's i-th leaf.

In the rest of the paper, we use the following notation: for any node r in tree T, we denote as $lc(r)$ (resp., $rc(r)$) the left (resp., right) child of node r.

Formal description of MP_1**.** Recall that we denote as T a logical tree having as leaves the $n = 2^t$ files from K. We first formally describe algorithms $\textsc{tag}_1$, $\textsc{upd}_1$, and then describe a recursive procedure $\textsc{rfind}$ used by $\textsc{find}_1$ and finally describe $\textsc{find}_1$.

The algorithm $\textsc{tag}_1$. On input $(1^\lambda, K)$, run the following instructions:

1. Let $K = (K_0, \ldots, K_{n-1})$ and compute $x_i = H_u(K_i)$, for $i = 0, \ldots, n-1$
2. compute $\{y_i \,|\, i \in T\} = \mathrm{GMT}(H_u; T; x_0, \ldots, x_{n-1})$
3. store $tag = (T, \{y_i \,|\, i \in T\})$

The algorithm $\textsc{upd}_1$. On input $(1^\lambda, j, K, tag)$, run the following instructions:

1. Let $tag = (T, \{y_i \,|\, i \in T\})$
2. Let $(j_0, \ldots, j_t)$ be the nodes on the path from T's leaf j to its root
3. Set $y_{j_0} = H_u(x_j)$ and $y_{j_i} = H_u(y_{lc(j_i)} | y_{rc(j_i)})$ for $i = 0, \ldots, t-1$
4. Return: $tag' = (\{y_{j_i} \,|\, i = 0, \ldots, t\})$

The subprotocol $\textsc{rfind}$. Let r denote a node in tree T and *list* denote a sequence of indices. On input $(1^\lambda, K, r)$ for C and $(1^\lambda, tag, r, list)$ for S, the following instructions are run:

1. C sends $z_{lc(r)}, z_{rc(r)}$ to S
2. S sets $b_l = 0$ if $z_{lc(r)} = y_{lc(r)}$ and $b_l = 1$ otherwise
3. S sets $b_r = 0$ if $z_{rc(r)} = y_{rc(r)}$ and $b_r = 1$ otherwise

4. S sends (b_l, b_r) to C
5. if $(b_l, b_r) = (1, 1)$ then do the following: if $lc(r), rc(r)$ are leaves then S adds $lc(r), rc(r)$ to *list* and the subprotocol returns *list*; otherwise C and S run recursively and in parallel two executions of RFIND: using as inputs $lc(r), list$ in one execution and $rc(r), list$ in the other
6. if $(b_l, b_r) = (1, 0)$ then do the following: if $lc(r)$ is a leaf then S adds $lc(r)$ to *list* and the subprotocol returns *list* otherwise C and S run recursively RFIND using $lc(r), list$ as inputs
7. if $(b_l, b_r) = (0, 1)$ then do the following: if $rc(r)$ is a leaf then S adds $rc(r)$ to *list* and the subprotocol returns *list* otherwise C and S run recursively RFIND using $lc(r), list$ as inputs
8. if $(b_l, b_r) = (0, 0)$ then the subprotocol returns: $\perp$

The protocol FIND$_1$. On input $(1^\lambda, K)$ for C and $(1^\lambda, tag)$ for S, the following instructions are run:

1. S writes *tag* as $\{y_i \mid i \in T\}$ and initializes $list = \emptyset$
2. C computes $\{z_i \mid i \in T\} = \text{GMT}(H_u; T; K_0, \ldots, K_{n-1})$
3. C and S set r as equal to T's root
4. C and S run RFIND using $r, list$ as (additional) inputs
5. let *list* be the output returned by RFIND
6. return: $out = list$

4 Efficient Communication and Round Complexity

In this section we consider the problem of designing memory integrity protection schemes that attempt to simultaneously minimize round and communication complexity. We discuss two schemes: a first scheme MP$_2$ based on repeated hashing using a halving-subset set system; and a second scheme MP$_3$ based on a recursive composition of MP$_2$. Both schemes are non-interactive (thus improving the scheme from Section 3 that has logarithmic round complexity). MP$_2$ maintain logarithmic communication complexity but detects up to a single corrupted file in K. MP$_3$ detects an arbitrary number t of corrupted files in K but has communication complexity $\Theta(\log n)^{\log t}$.

Halving-sequence set systems. Combinatorial constructions analogue to the set system that we recall here have been used in several papers and even in several different contexts in computer science.

Given a ground set $GS = \{x[0], \ldots, x[n-1]\}$, where we assume for simplicity that $n = 2^k$, for some positive integer k, we define the set system $SS = \{(S_{1,0}, S_{1,1}), \ldots, (S_{k,0}, S_{k,1})\}$ as follows:

- $S_{i,0} = \{x[j] \mid j \in [n] \wedge j \bmod (n/2^{i-1}) < n/2^i\}$
- $S_{i,1} = \{x[j] \mid j \in [n] \wedge j \bmod (n/2^{i-1}) \geq n/2^i\}$,

for $i = 1, \ldots, k$, where we denote the set $\{0, \ldots, p-1\}$ as $[p]$, for any positive integer p.

An important property of this construction is that there exists a bijection between the ground set GS an the k-tuples $(S_{i,l_1}, \ldots, S_{k,l_k})$, for $l_1, \ldots, l_k \in \{0, 1\}$, which we will refer to as the *bijection property* of halving-sequence set systems.

Our scheme MP$_2$**.** The basic idea behind this scheme is that the set of files in K is considered a ground set, and C should help S to localize the (single) corrupted file by sending hash tags for each of the subsets from SS (here interpreted as sequences). Then S compares these hash tags with those previously computed using the tagging algorithm, and the unique corrupted file will be discovered thanks to the above bijection property.

More precisely, the tagging algorithm consists of running a collision-intractable hash function using as input all subsets from SS, and thus obtaining a sequence *tag* of $2 \log n$ hash tags. The update algorithm consists of only updating the $\log n$ hash tags affected by a modification of a file in K (or, an element in GS). Then the corruption finding protocol goes as follows. First, C recomputes the output of the hash function over the current version of K and sends this output tag' to S. Now, S compares *tag* with tag' by computing the k-tuple $(S_{i,l_1}, \ldots, S_{k,l_k})$, for some $l_1, \ldots, l_k \in \{0, 1\}$, for which the hash tags differ. By the mentioned bijection property, this difference is only due to one element from GS, and thus one file from K.

We note that MP$_2$ achieves $\Theta(s \log n)$ communication complexity and $\Theta(1)$ round complexity.

An informal description of MP$_3$**.** As mentioned, this scheme extends MP$_2$, where the server was able to detect a single corrupted file, into a scheme where the server is now able to detect an arbitrary number t of corrupted files, while keeping round, communication and update complexity as small as possible. The basic idea is to use the halving-sequence set systems to select $2 \log n$ subsets from $\{K_1, \ldots, K_n\}$ on which to recurse, where the parameter t is being halved at each recursive step. The proof that this suffices to detect all t corrupted files is based on a non-trivial covering property of halving-sequence set systems. We obtain the following theorem.

Theorem 4.1 Let λ, n, m be positive integers, let $t \in \{1, \ldots, n\}$, and let ϵ be a function negligible in

Communication complexity	Round complexity	Storage complexity	Corrupted files
sn	1	sn	t
$st\log(n/t)$	$\log n$	sn	t
$s\log n$	1	$s\log n$	1
$st(\log n)^{\log t}$	1	$st(\log n)^{\log t}$	t

Figure 1: The performance of schemes MP_i is on the i-th line (constant multiplicative factors are omitted), for $i = 0, 1, 2, 3$. Here, n is the number of files to be protected, t is the number of corrupted files, s is the length of the output from the collision-intractable hash function.

λ. Assuming the existence of a family of collision-intractable hash functions, there exists (constructively) a (t, ϵ)-secure memory integrity protection scheme $\mathrm{MP}_3 = (\textsc{tag}_3, \textsc{upd}_3, \textsc{find}_3)$ with parameters (λ, n, m) and with the following performance: $\Theta(s \cdot t \cdot \log(n)^{\log t})$ communication complexity, $\Theta(1)$ round complexity, $\Theta(t^2 \cdot \log(n)^{\log t} \cdot t(H_u))$ verification time, and $\Theta(t \cdot \log(n)^{\log t})$ update complexity, $\Theta(s \cdot t \cdot \log(n)^{\log t})$ storage complexity, where $s, t(H_u)$ denote the length of the output and running time, respectively, of the assumed hash function.

5 Conclusions

We have formally defined a cryptographic model for the design and analysis of efficient cryptographic hashing schemes for memory integrity protection. Motivated by applications to vehicular networks, we have considered a client-server model, where the trusted server helps avoiding the classical problem of verification of the integrity verifier's program. In this model, we show that cryptographic hashing can be made significantly more efficient than the well known strategy of hashing every single file. The performances of our schemes $\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MP}_3$, as well as the performance of the mentioned well-known strategy, called MP_0, are summarized in Table 1.

Acknowledgement. This work was performed under the United States VII initiative [9]. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

References

[1] M. Blum, W. Evans, P. Gemmell, S. Kannan, and M. Naor, *Checking the Correctness of Memories*, in Proc. of IEEE FOCS 1991, and Algorithmica, 1994, pp. 225-244.

[2] http://bbcr.uwaterloo.ca/ rxlu/sevecombib.htm

[3] Y. Chen, R. Venkatesan, M. Cary, R. Pang, S. Sinha, and M. H. Jakubowski, *Oblivious Hashing: A Stealthy Software Integrity Verification Primitive*, in Information Hiding: 5th International Workshop 2002, LNCS 2578, Springer Verlag.

[4] R. Merkle, *A Certified Digital Signature*, in Proc. of CRYPTO 1989, LNCS 435, Springer-Verlag.

[5] A. Seshadri, M. Luk, E. Shi, A. Perrig, L. van Doorn, and P. Khosla, *Pioneer: Verifying Code Integrity and Enforcing Untampered Code Execution on Legacy Systems*, in Proc. of ACM Symposium on Operating Systems Principles, 2005.

[6] G. E. Suh, D. Clarke, B. Gassend, M. van Dijk, S. Devadas, *Efficient Memory Integrity Verification and Encryption for Secure Processors*, in Proc. of the 36th International Symposium on Microarchitecture (MICRO-36 2003), IEEE Press.

[7] B. Yee and J. D. Tygar, *Secure coprocessors in electronic commerce applications*, in Proc. of 1st Usenix Workshop on Electronic Commerce, 1995.

[8] http://www.tripwire.com

[9] http://www.its.dot.gov/vii/

An Interoperable Security Architecture for Vehicular Software Protection

Michael Scheibel
Sirrix AG, Germany
m.scheibel@sirrix.com

Christian Stüble
Sirrix AG, Germany
c.stueble@sirrix.com

Marko Wolf
escrypt GmbH, Germany
mwolf@escrypt.com

Abstract

Software is becoming the most innovative and valuable part of current and future automotive vehicles. Therewith, software updates after delivery of the vehicle have become a common issue also within the automotive area. In order to prevent misuse that could lead to disadvantages for the manufacturer or even affect the operational safety of a vehicle, vehicular software and data need to be protected. However, most currently deployed software distribution mechanisms rely on trusted networks despite their inherent weaknesses in flexibility and reliability and their enormous organizational complexity, which, moreover, provide only low resistance against attacks of skilled adversaries.

In this paper we present the design of an interoperable security architecture for software protection that is capable of binding arbitrary digital content to a certain vehicular hard- and software configuration. The underlying protocol in particular ensures that the content can never be accessed unauthorized even though it is being distributed using a fully untrusted infrastructure. We show how to implement this architecture efficiently by means of virtualization technology, an (open source) security kernel, trusted computing functionality, and an interoperable legacy operating system.

1 Motivation

Embedded software is becoming the most innovative and valuable part of current and future automotive vehicles. A today premium vehicle has up to several 100 megabytes of embedded code providing more than 2000 individual functions [5]. Future vehicles will employ more than one gigabyte of embedded code, which then control nearly all functionality of a vehicle including steering and braking [9]. Due to the ubiquitous use of flashable Electronic Control Units (ECUs)[1], software updates take place also after delivery of the vehicle. Reasons are warranty-based updates in recalls, allowing to correct defective software, but also value-adding updates, such as software-based features sold in the after-market. In addition, delivery of digital content such as routing and location-based information for navigation systems or multimedia content for in-car entertainment has become important. Intellectual property and know-how implied in software and digital content needs to be protected in order to prevent misuse such as unauthorized feature activation, unauthorized updates or modifications, illegal copies, and know-how theft. Such attacks could lead to real loss or damage such as undermined business models, false warranty claims, and damage the manufacturer's public reputation. Since software already controls very critical components such as airbags, anti-lock brakes, or the engine control, a wireless attacker or an (intentionally or unintentionally) unauthorized software update may even affect the operational safety of a vehicle. While considering malicious attacker scenarios so far has been mostly only a subject of matter for a few single components such as immobilizer or tachograph [11], information security today becomes a crucial design and implementation issue within automotive electronic development.

So far, most deployed software distribution mechanisms rely on proprietary trusted networks despite their inherent weaknesses in flexibility and reliability and their enormous organizational complexity. Hence, current vehicle ECUs provide only low resistance to attacks of skilled adversaries. With little and easily obtainable equipment, an adversary can easily read out and install any software at will. Considering this, the awareness of the need for efficient software protection mechanisms grows [1], and countermeasures start to appear in the form of digitally signed software updates [10]. However, as already known from the PC world, solely software based mechanisms cannot provide reliable protection. An adversary can for instance read cryptographic secrets from unprotected memories, thus being able to break most software based protection measures. Protection measures based on tamper-resistant components, and cryptographic techniques are still seldom in the automotive domain. Since the design of fully secure ECUs is considered to be too costly or even infeasible in practice [2], it is highly desirable to design interoperable security archi-

[1] A flashable ECU is a microcontroller capable of reprogramming its memory for application programs and data based on flash memory [10].

tectures where only few components are secure and fully trusted, while all others can be built from common off-the-shelf hardware.

1.1 The Goal

The proposed protocol allows a content provider (cf. Section 4.2) to "bind" arbitrary software to a "certain" vehicle hard- and software configuration. Hence, a content provider can reliably ensure that his bound and encrypted content can be decrypted and accessed only by a previously authorized vehicle configuration while being distributed using a fully untrusted infrastructure. The given functionality can be used to limit the usage of a software update or digital content to a certain vehicle brand, a certain vehicle type or even to a single car, thus preventing for instance unauthorized feature activation, unauthorized updates or modifications, and unauthorized copies. Moreover, since decryption of bound content is possible only on previously defined hard- and software configurations, a content provider can enforce the usage of trustworthy platform configurations that implement appropriate access control mechanisms.

Our software protection prototype is realized on top of the Turaya security kernel that is suited for embedded devices based for instance on ARM or MIPS architectures. We assume the security kernel to be running at least on the central gateway or the so-called head unit. Since current development moves from many small, individual ECUs to a few powerful multi-purpose ECUs that combine the functionality of several ECUs [9], we suggest to run the security kernel also on these larger ECUs. The security kernel basically uses two important mechanisms, namely virtualization technology (VT) and Trusted Computing (TC). Virtualization technology—well known in server environments—employs a small OS kernel (hypervisor), which allows to run multiple, full-fledged operating systems on one host processor at the same time. Thus, VT enables the efficient reuse of existing applications, while running even different OS versions isolated on a single computing platform in parallel. Moreover, VT prevents that the state of one OS or isolated application could affect the state of another, thus security vulnerabilities and bugs, e.g., of a legacy OS, cannot affect security-critical components. The Turaya security kernel actually uses a lightweight virtualization technique called para-virtualization that is realtime capable and even able to outperform even native software configurations [14]. TC technology enables hardware-based software protection based on a standardized, interoperable, tamper-resistant security chip tightly bound to the corresponding computing platform. Thus, TC serves as root of trust that enables several security mechanisms (cf. Section 3). However, the Turaya security kernel is required at the client side only, where it has to prevent circumvention of existing security policies by (malicious) software. Hence, our software protection protocol requires no changes of the existing software environment at the (trusted) content provider side.

1.2 Related Work

Various authors have identified the need for information security and software protection in vehicles [3, 5]. Ehlers presents an approach for vehicle system integrity by binding software IDs based on digital signatures or hashes to fixed hardware IDs of the ECUs, but does not address malicious actions at all [6]. Seshadri et al. [16] introduce a software-based attestation mechanism for verifying embedded software at run-time, but cannot prevent subsequent manipulations and exclude all types of hardware attacks. Adelsbach et al. introduce a requirement model and a secure installation protocol for software installation in embedded systems [1], whereas Fibikova describes several (access control) software countermeasures to face current automotive threats [8]. However, both do not address hardware requirements. First virtualization approaches [7] for automotive operating systems based on microkernels[2] mainly have safety in mind and for instance do not address how to protect cryptographic secrets.

Nevertheless, the need for hardware-based security measures in embedded systems design has already been identified [15, 21]. There exist a few hardware-based approaches to include strong security also in embedded systems [4, 13]. However, all are proprietary solutions and none of them addresses the particular requirements and constraints within the automotive context such as the high cost pressure on automotive ECUs, their low computing power, limited storage, and sporadic network connectivity.

2 Turaya Security Kernel

As shown in Figure 1, the Turaya security kernel is a small software layer that provides an abstract interface to the hardware resources, enforces strong isolation of applications and implements elementary security services built on a hardware layer that is enhanced by TC technology. Existing operating systems and applications on top of the security kernel are running in parallel to, but strongly isolated from, security-critical applications. Hence, the isolation prevents that a legacy OS or isolated application can access or even affect the state of another. The security kernel is capable of integrating software virtualization such as microkernels [12], and is prepared to take use of emerging hardware virtualization technologies.

[2] A microkernel is a minimalized operating system kernel that provides only essential services such as logical address spaces and inter-process communication (IPC). Processes on top of the microkernel run in their own address space and are therefore strongly isolated from each other.

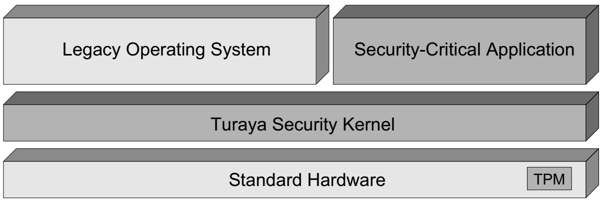


Figure 1. Turaya architecture overview.

The security kernel can be logically divided into a *hypervisor layer* and a *trusted software layer* (cf. Figure 2). The main task of the hypervisor layer is to provide an abstract interface of the underlying hardware resources like interrupts, memory and hardware devices. Moreover, this layer allows to share these resources and realizes access control enforcement on the object types known to this layer. Currently we are using a microkernel as the base of the hypervisor layer. The trusted software layer builds on the hypervisor layer and realizes security-critical services according to the designated task. In the following we briefly describe some elementary components of the trusted software layer.

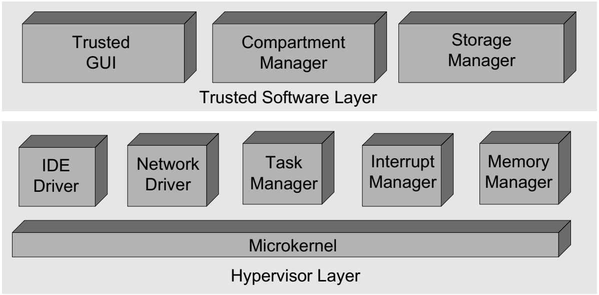


Figure 2. Turaya security kernel.

The *Trusted GUI* controls the graphic adapter and the corresponding input devices to establish a trusted path between the user and an application. The Trusted GUI labels application windows with unique application names and enforces isolation between applications on the GUI level. Thus, unauthorized applications cannot access the graphical output of other applications or fake their interface to look like the legitimate application. The *Compartment Manager* loads compartments[3] and measures the integrity of compartments. These integrity measurements can then be reported to existing local compartments and to remote compartments as well. Finally, the *Storage Manager* enables other applications to persistently store their individual states. It preserves the integrity, confidentiality, and freshness of the managed data such that only the application or the user having produced the data can later re-access it.

3 Trusted Computing

This section introduces those Trusted Computing (TC) functionalities, which are significant for the design of our security architecture for vehicular software protection.

Authenticated Boot. During an authenticated boot process, any code that will be executed is "measured" before execution, e.g., by calculating its cryptographic hash value. TC hardware is responsible for the secure storage and provision of these measurement results. Upon completion of an authenticated boot process, these measurements reflect the configuration of the currently running hardware and software environment. TC technology, however, remains passive and does explicitly *not* prevent a certain computing environment from being compromised during runtime, but integrated cryptographic mechanisms enable the platform to verifiably report their actual configuration to local and external parties.

Secured Cryptography. TC hardware implements a set of cryptographic operations to ensure that malicious software cannot compromise cryptographic secrets. Hence, key generation and decryption operations are done "on-chip", so that secret keys do not have to leave the chip. To perform a decryption operation with a specific key, several types of authorization are possible. A distinctive feature of TC hardware is the ability to not only use passwords as authorization, but also integrity measurements. That is, only a platform running previously defined software or hardware components is authorized to use a certain key. Moreover, the property that a certain key is "bound" to a platform configuration can be certified by TC hardware. This certification includes the integrity measurements that authorize a platform to employ the key. A remote party can verify the certificate and validate the embedded integrity measurement against "known good" reference configurations before encrypting data with the certified key[4].

Trusted Platform Module. The base of TC technology is the standardized Trusted Platform Module (TPM) that is considered to be a tamper-resistant hardware device similar to a smart-card and is assumed to be securely bound to the computing platform. The TPM is primarily used as a root of trust for integrity measurement and reporting and to secure all critical cryptographic operations (cf. preceding paragraphs). Current TPMs base on the open specification version 1.2 [20] published by the Trusted Computing Group (TCG) [19], an initiative led by AMD, HP, IBM, Infineon, Intel, Lenovo, Microsoft, and Sun. TPMs are available, e.g., from Atmel, Broadcom, Infineon, Sinosun, STMicroelectronics, and Winbond. A TPM basically consists of an asymmetric cryptographic engine (RSA), a cryp-

[3] An application or operating system that is logically or even physically isolated from other software components.

[4] Mostly, the data to be encrypted is a secret symmetric key used for the encryption of larger amounts of data.

tographic hash function (SHA-1), a true random number generator, a few bytes of volatile memory, some kilobytes of non-volatile memory and sensors for tampering detection. Appropriately high quantities provided, the costs for an automotive TPM are assumed to be well bellow one US dollar [19].

4 Design

This section describes the software protection protocol in detail, which relies on the Trusted Computing functionalities described in the previous section.

4.1 Security Objectives & Assumptions

This section defines the overall security objectives and considers the required assumptions.

Security Objectives. Unauthorized vehicle hardware and/or software configurations must not be able to access content (*content confidentiality*). Moreover, content cannot be altered during the transmission process without having modifications at least detected (*content integrity*).

Assumptions. The underlying hardware (e.g., CPU, devices, TPM) is non-malicious and behaves as specified. The content server, i.e., the server that holds the original content and provides the binding of the content to the vehicle certificate, is fully trusted. The vehicle has an TPM chip version 1.1b or higher integrated.

4.2 Protocol Overview

The two parties involved in our protocol are the *user* and the *content provider*. The *content provider* distributes digital content (ECU software, navigation data, media files, etc.) that will be employed by the *user*. For simplicity, we summarize OEMs, authorized maintenance service providers, infotainment data vendors, and even other vehicles in the single party that provides arbitrary digital content. The *user* refers to the person that currently uses the vehicle. As owner and user often coincide, and a distinction between the two does not affect our proposed solution, we summarize them also with a single party.

As illustrated in Figure 3, the protocol basically consists of four steps. Firstly, the TPM in the car's head unit generates an asymmetric key pair and a certificate stating that the private key is bound to the specific TPM and the car's head unit[5] configuration. Secondly, the certificate and the public key are transferred to the content provider, who validates the certificate. This validation implies verifying the signature and checking the configuration against "known good" reference values. Thirdly, on successful validation, the content provider uses the corresponding public key to establish a "trusted channel" to the car. This means that content previously encrypted using this public key can be decrypted only if the vehicle provides the "trusted" configuration as stated in the successfully validated certificate. A trusted configuration in this context means that the Turaya security kernel is installed and running on the head unit and that all components (both hardware and software) processing the decrypted content are trusted by the content provider. The decrypted and bound content is finally sent to the vehicle for decryption and usage if (and only if) the vehicle provides the stated trusted configuration. Note that steps (1) and (2) can be done already at the end of the manufacturing process by the OEM, thus reducing initial complexity. A new key pair and a corresponding certificate is thereafter only necessary, if one of the crucial and therefore measured hard- or software components has been modified due to a software update or hardware change.

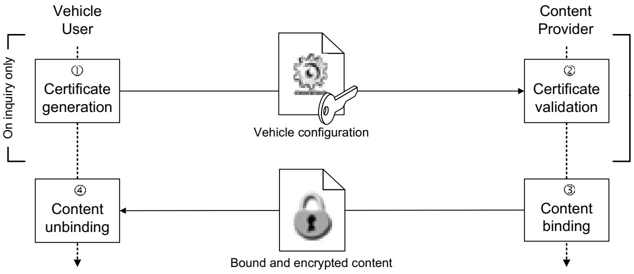


Figure 3. Certificate-based content binding.

Server Architecture. As already mentioned in Section 1.1, the content provider side requires no changes of their existing server environment (and does not need Trusted Computing technology either) when employing our proposed software protection protocol. Thus they can employ their common computing architecture to accomplish certificate validation and content binding.

Client Architecture. The software components involved in certificate creation and content unbinding on the client side are depicted in Figure 4. The two components that need to be outsourced from the untrusted legacy OS are the *Trusted Viewer*, which decrypts and renders the content, and the *Trust Manager*, which provides an abstraction of the TC technology. The Trust Manager offers an interface for key and certificate generation and for decryption (unbinding) with a provided key. Note that only the Trusted Viewer may use this interface and thus accesses TPM functionality. The isolation feature of the Turaya security kernel, e.g., as provided by a secure GUI, prevents unauthorized accesses from the legacy OS to plain content or the TPM.

[5] The head unit integrates all computing subsystems in a car.

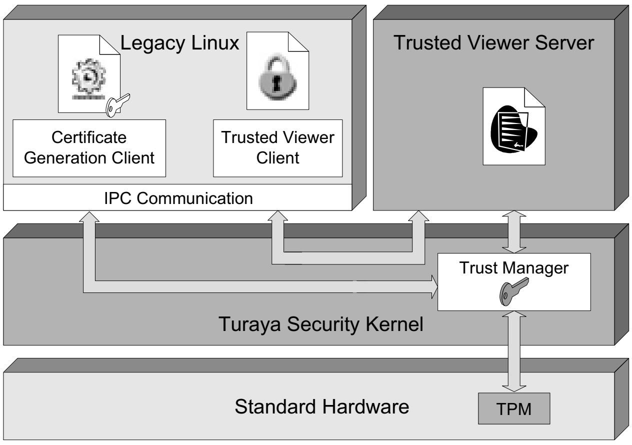


Figure 4. Client architecture.

4.3 Protocol Details

The following section gives further details on the protocol steps introduced in Section 4.2.

Certificate Generation. The legacy OS provides a client application (*Certificate Generation Client*) to invoke the Trust Manager to create an asymmetric key pair and a corresponding certificate stating that the private key is bound to the actual vehicle configuration. The actual vehicle configuration has been made available for the Trust Manager by the authenticated booting capability of the head unit (cf. Section 3) that measures all crucial computing subsystems in a car. The involved protocol steps are depicted in Figure 5. Finally, the certificate is transferred to the content provider. Since the certificate size is only about 1 kB, even small-bandwidth connections such as GSM can be therefor employed. Note again that a new key pair and thus a new certificate is only necessary if a security-critical hard- or software component has been changed or modified.

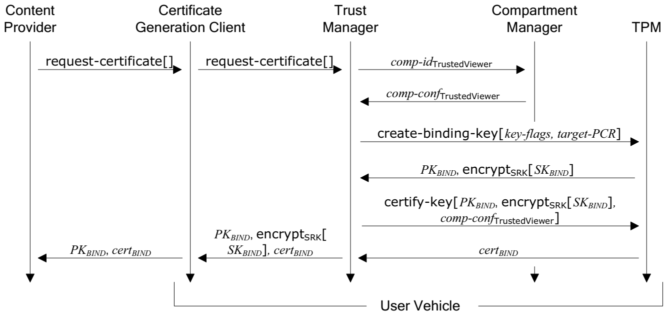

Content Provider
Certificate Generation Client
Trust Manager
Compartment Manager
TPM
request-certificate[]
request-certificate[]
$comp\text{-}id_{\text{TrustedViewer}}$
$comp\text{-}conf_{\text{TrustedViewer}}$
create-binding-key[$key\text{-}flags$, $target\text{-}PCR$]
PK_{BIND}, encrypt$_{\text{SRK}}$[SK_{BIND}]
certify-key[PK_{BIND}, encrypt$_{\text{SRK}}$[SK_{BIND}], $comp\text{-}conf_{\text{TrustedViewer}}$]
PK_{BIND}, $cert_{BIND}$
PK_{BIND}, encrypt$_{\text{SRK}}$[SK_{BIND}], $cert_{BIND}$
$cert_{BIND}$
User Vehicle

Figure 5. Certificate generation protocol.

Certificate Validation. On the arrival of the certificate, the content provider has to validate the stated values with the given TPM signature. The content provider therefore has the TPM's signature verification key available or has to establish an additional protocol to receive the TPM's signature verification key.[6] As depicted in Table 4.3, the certificate contains the public part of the RSA binding key PK_{BIND} and its cryptographic hash value H_{PK}, the value C_{Vehicle}, which represents the measured configuration of the Trusted Viewer compartment and the underlying hardware and software environment, and the TPM signature S over H_{PK} and C_{Vehicle}. On successful integrity verification, the content provider checks the given vehicle configuration C_{Vehicle} against his trusted and "known good" reference values.

Public binding key PK_{BIND}
Hash value H_{PK} = SHA-1(PK_{BIND})
Vehicle configuration C_{Vehicle} = ($comp\text{-}conf_{\text{TrustedViewer}}$, $target\text{-}PCR$)
Signature S = $sign_{TPM}$ (H_{PK}, C_{Vehicle})

Table 1. The vehicle TPM certificate.

Content Binding. On successful certificate validation, the content provider uses the given public key PK_{BIND} to encrypt and bind contents to the corresponding vehicle configuration. Virtually, the content provider encrypts and binds a secret symmetric key $symm\text{-}key$ used for the encryption of the actual content. The bound content now can be transferred to the user using a capable data link, i.e., a GSM connection or offline transfers such as by USB sticks or DVDs.

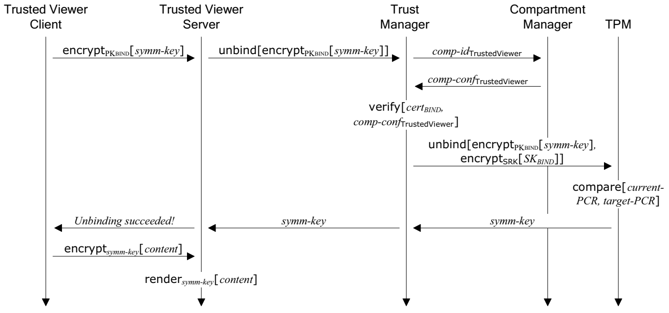

Trusted Viewer Client
Trusted Viewer Server
Trust Manager
Compartment Manager
TPM
encrypt$_{PK_{BIND}}$[$symm\text{-}key$]
unbind[encrypt$_{PK_{BIND}}$[$symm\text{-}key$]]
$comp\text{-}id_{\text{TrustedViewer}}$
$comp\text{-}conf_{\text{TrustedViewer}}$
verify[$cert_{BIND}$, $comp\text{-}conf_{\text{TrustedViewer}}$]
unbind[encrypt$_{PK_{BIND}}$[$symm\text{-}key$], encrypt$_{\text{SRK}}$[SK_{BIND}]]
compare[$current\text{-}PCR$, $target\text{-}PCR$]
Unbinding succeeded!
$symm\text{-}key$
$symm\text{-}key$
encrypt$_{symm\text{-}key}$[$content$]
render$_{symm\text{-}key}$[$content$]

Figure 6. Content unbind protocol.

Content Unbinding. The vehicle's head unit is able to decrypt and access the content, if it provides the trusted configuration as stated during certificate generation. As depicted in Figure 6, the legacy OS provides a client application (*Trusted Viewer Client*) to invoke the Trusted Viewer application that runs in parallel to, but isolated from the legacy OS. The Trusted Viewer in turn, invokes the Trust Manager to unbind the secret symmetric key. Since the unbinding function is a protected TPM hardware function, the

[6] This can be done also anonymously using an anonymous TPM attestation identity key (AIK).

unbinding will succeed only, if the actual vehicle configuration matches the one at certificate generation. On success, the Trusted Viewer uses the secret symmetric key for decryption and rendering of the actual content. Note again, since the symmetric decryption key resides only within the isolated Trusted Viewer, it cannot be compromised by a malicious legacy OS application or a legacy OS leakage.

5 Implementation

Our prototype runs on IA32 processors, but the underlying L4 microkernel [18] is known to run on embedded processors such as ARM as well. The prototype is securely booted with a modified GRUB bootloader [17]. After startup, it shows two *compartments*. The first one contains a standard Linux distribution running on a virtualized Linux kernel. The other compartment contains the Trusted Viewer, which currently supports AES-encrypted PDF files. The AES key is bound to the software configuration and has to be unbound by the Trust Manager before use. The Trust Manager builds on an open-source Trusted Software Stack. All components solely communicate through IPC calls. The GUI for initiating the certificate creation and the rendering of encrypted content is part of the untrusted compartment.

Protocol step (cf. Figure 3)	Atmel 1.1b	NSC 1.1b
(1) Certificate generation	30 – 80 s	52 – 55 s
(3) Session key encryption	< 1 s	< 1 s
(4) Session key decryption	2 – 3 s	23 – 24 s

Table 2. Performance measurement results.

We have implemented the described protocol and run it on TPMs of different vendors. The measurement results with maximum asymmetric key lengths (i.e., 2048 bits) are shown in Table 2. Note that the TPM calculations dominate the overall computation and network transfer times.

6 Summary and Outlook

In this paper we presented the design and the implementation of an interoperable security architecture for software protection that is capable to bind arbitrary digital content to a certain vehicular hard- and software configuration. We described how the underlying protocol ensures that the content can never be accessed unauthorized even though it is being distributed using a fully untrusted infrastructure. We showed how to implement this architecture efficiently by means of virtualization technology, an (open source) security kernel, standard Trusted Computing functionality, and an interoperable legacy operating system (currently Linux).

Even though Trusted Computing and virtualization technology has not yet reached the automotive area, we believe it could solve several critical security and safety issues and enable various innovative business models while requiring only minimal technical and financial resources.

References

[1] A. Adelsbach, U. Huber, and A.-R. Sadeghi. Secure software delivery and installation in embedded systems. In *Information Security Practice and Experience Conference—ISPEC*, 2005.

[2] R. Anderson and M. Kuhn. Tamper Resistance—a Cautionary Note. In *USENIX Workshop on Electronic Commerce*, 1996.

[3] K. F. Anthony Bellissimo, John Burgess. Secure software updates: Disappointments and new challenges. In *USENIX Workshop on Hot Topics in Security*, 2006.

[4] ARM Ltd. Trustzone technology overview. `www.arm.com/products/esd/trustzone_home.html`.

[5] M. Broy. Challenges in Automotive Software Engineering. In *International Conference on Software Engineering*, 2006.

[6] T. Ehlers. Systemintegrität von vernetzter Fahrzeugelektronik. In *Embedded Security in Cars (escar)*, 2003.

[7] K. Elphinstone, G. Heiser, R. Huuck, S. M. Petters, and S. Ruocco. L4cars. In *Embedded Security in Cars (escar)*, 2005.

[8] L. Fibikova. On building trusted services in automotive systems. In *Embedded Security in Cars (escar)*, 2004.

[9] H.-G. Frischkorn. Automotive Software – The Silent Revolution. In *Workshop on Future Generation Software Architectures in the Automotive Domain*, 2004.

[10] Herstellerinitiative Software. Functional specification of a flash driver version 1.3. Technical report, 2002.

[11] K. Lemke, A.-R. Sadeghi, and C. Stüble. An open approach for designing secure electronic immobilizers. In *Information Security Practice and Experience—ISPEC 2005*, 2005.

[12] J. Liedtke. Toward real microkernels. *Communications of the ACM*, 39(9):70–77, 1996.

[13] M. Milenkovic, A. Milenkovic, and E. Jovanov. Hardware support for code integrity in embedded processors. In *International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*, 2005.

[14] NICTA. L4 performance results. `ertos.nicta.com.au/research/l4/performance.pml`.

[15] A. Ravi, Srivathsand Raghunathan, P. Kocher, and S. Hattangady. Security in embedded systems: Design challenges. *Transactions on Embedded Computing Systems*, 3(3), 2004.

[16] A. Seshadri, A. Perrig, L. v. Doorn, and P. Khosla. Using software-based attestation for verifying embedded software in cars. In *Embedded Security in Cars (escar)*, 2004.

[17] C. Stueble and M. Selhorst. Trusted GRUB. `www.trust.rub.de/trusted_grub.html`.

[18] The Fiasco μ-kernel. `www.tudos.org/fiasco/`.

[19] The Trusted Computing Group (TCG). `www.trustedcomputinggroup.org`, 2003.

[20] The Trusted Computing Group (TCG). TPM Specification Version 1.2 Revision 94, 2006.

[21] G. van Battum and D. Caluccio. Physical security for automotive applications: What can we learn from other industries? In *Embedded Security in Cars (escar)*, 2005.

Workshop ‘Designing the Internet of Things for Workplace Realities: Social and Cultural Aspects in Design and Organization’ (Social- IoT 2008)

Message from the Workshop Organizers

Welcome to the first international workshop on Designing the Internet of Things for Workplace Realities: Social and Cultural Aspects in Design and Organisation.

The workshop has the goal to increase awareness of organizational issues of the Internet of Things and to provide a forum for discussion of design approaches to manage critical organizational issues. Furthermore we would like to build a bridge between the various research communities exploring organizational, social and cultural aspects of the Internet of Things and ubiquitous computing. A multitude of methods and guidelines have been developed to address organizational, human and social issues in technology design, deployment and use. However, those methods have often not yet been adopted and tested for the Internet of Things or ubiquitous computing. We discuss particular design methods that address social, cultural and organisational perspectives and experiences from modelling these processes within research and development projects. We consider the wider implications of the Internet of Things in society and culture and understand the technologies and the surrounding cultural and social logics in a co-constitutive process.

We would like to thank the members of the program committee for their fair and detailed reviews which led to a high quality program. We also thank the organisers of the Internet of Things conference 2008 for making the Social-IoT workshop possible.

Daniel Boos, ETH Zurich, Switzerland
Katharina Kinder, Lancaster University, UK
Gerd Kortuem, Lancaster University, UK

Social-IoT 2008 Workshop Organization

Workshop Co-organizers
Daniel Boos, ETH Zurich, Switzerland
Katharina Kinder, Lancaster University, UK
Gerd Kortuem, Lancaster University, UK

Technical Program Committee
Abraham Bernstein, University of Zurich, Switzerland
Monika Büscher, Lancaster University, UK
Paul Devadoss, Lancaster University, UK
Gudela Grote, ETH Zurich, Switzerland
Stephan Haller, SAP Research
Lorenz Hilty, EMPA, Switzerland
Erik Hollnagel, Ecoles de Mines de Paris, France
Saadi Lahlou, R&D, Laboratory of Design for Cognition, France
Kalle Lyytinnen, Case Western Reserve University, USA
Martina Merz, University Lucerne, Switzerland
Werner Rammert, Technical University Berlin, Germany
George Roussos, University of London, UK
Craig Shepherd, University of Leeds, UK
Nils Zurawski, University of Hamburg, Germany

I am not a machine, Sir: RFID and Customer Services

George Roussos
School of Computer Science and Information Systems,
Birkbeck College, University of London, UK
g.roussos@dcs.bbk.ac.uk

Abstract – Invited Talk

The adoption of a new information technology can be a challenge for any organisation, and the implementation of RFID---and for that matter pervasive computing in general---could not be otherwise. In this talk I will examine the specifics of the deployment of the Oyster card by Transport for London, which is of particular interest due to the very large scale of the system and the complexity of the organisation involved. I will look at the business drivers, the technology choices, the implementation plan, retraining and reallocation of staff, operational issues and last but not least the experience of commuters. I will also highlight how RFID technology allowed the extensive overhaul of the customer service system and its substitution by self-service. I will contrast the user experience created in this case with that resulting from the deployment of RFID in retail, and highlight the changing roles of staff on the ground.

Motorola’s experiences in designing the Internet of Things

Andreas Schaller, Katrin Mueller, Byung-Yong Sung[1]
Motorola,Germany
Motorola,Korea[1]
Andreas.Schaller@motorola.com
Katrin.Mueller@motorola.com
bysung@motorola.com

Abstract

The Internet of Things will enable connectivity for virtually any physical object that potentially offers a message, and will affect every aspect of life and business. This paper looks at the concepts and technologies in three application areas that Motorola is contributing to now and in the coming years.

1. Motorola’s Position in the Internet of Things

In just 20 years, the Internet has fundamentally changed the way we live, learn, do business and entertain ourselves. What makes the Internet so revolutionary is that it provides a standard way for people to connect anywhere around the world. Today’s Internet connects people to people, providing information in text, video, sound and other formats intended for use by people. The next step is to Internet-enable physical objects — connecting people with things and even things with things. The Internet of Things will enable connectivity not just between people and their computing devices, but between actual, everyday things. By enabling connectivity for virtually any physical object that can potentially offer a message, the Internet of Things will affect every aspect of life and business in ways that used to be the realm of fantasy — or even beyond fantasy. This paper looks at the concepts and technologies in three application areas that Motorola is contributing to now and in the coming years.

2. The Retail Space

2.1 Overview of Motorola’s activities

Motorola’s Enterprise Mobility group is addressing the Retail Space from three different directions: Supply Chain efficiency, associate effectiveness, and customer experience. Especially personalized shopping experience is converting browsing into buying customers by delivering tailored products and promotional information.

Motorola’s Enterprise Mobility solutions transform the customer experience by leveraging personalized information and creating new shopping experiences by connecting customers and products seamlessly while increasing sales and brand loyalty for the retailer. The goal is to provide the customer with:

- Instant access to price and availability data via personal shopping systems
- Cross-sell and up-sell opportunities through target promotions (Micro Kiosks)
- Payment systems to put your customers in control of the checkout process to utilities checkout performance.

2.2 Things to Things Technologies

To increase the customer shopping experience “person to things” (P2T) communication and therefore the P2T interaction has to be improved. This can be achieved by leveraging any kind of data capturing technologies. Most of these technologies are already available in devices used in B2B application, such as 1D, 2D barcode scanning or RFID reading. For generating new kind of customer services in retail it will be necessary to transfer these technologies in mass markets devices. To ensure a fast adoption rate it is necessary to start with low hanging fruit technologies like barcode scanning by camera, which will become a “free” feature for mobile devices morphing into high end camera phones. The drawback for this approach is that traditional laser scanning of barcodes displayed on mobile devices can not be used due to incompatibility with the phone displays properties.

Another data capturing technology is near field communication (NFC), which is build on existing smart card technology used for loyalty, access, and payment contact less cards. In addition to the checkout service improvements, this technology can provide the customer with the opportunity to interact with different objects inside of the retail store by a simple touch. Therefore it will be possible for the customer to identify themselves in front of a micro kiosk or any kind of future e-paper display, which allows the retailer to offer a selection of new services like coupons downloading to the handset or personalized advertisement based on historic or current shopping. The mayor difference between the barcode and the NFC approach is that in the barcode based scenario the personalized data are stored in the customer phone whereas in the NFC scenario a retail application managing the coupons is hosted on the phone. The pros of the second scenario are obvious as the retailer now owns a section of its customer e-wallet which can be adapted anytime to the customers needs over the air. For many other data capturing technologies the infrastructure as well as the devices has to be established. The well known example for this technology is UHF RFID tagging. Whereas in the B2B space pallet and box tagging is increasing step by step and EPC B2B services and platforms are in the focus of many companies, it is still a long way to go to see enough item level tagging to allow handset makers to enlarge the integration of UHF readers from enterprise specific to customer mass market devices. In additional the business case for the future flat rate period in the mass market driven wireless industry has not been clearly identified so far. As a result the telecom providers will not sponsor the UHF reader integration and the customer would have to pay a higher priced phone to purchase the UHF feature compared to a camera phone where the barcode feature scanning will be for free.

In addition to the data capturing technologies the wireless technologies used to transfer the information have to be mass market capable, too. In Europe these are primarily GSM, UMTS , BT, and finally WiFi based communication protocols. Most of the retailers require a penetration rate of 60+% to be used in retail. Therefore only GSM, BT, and SMS are currently available ubiquitous technologies to be built on. Especially as data flat rates are currently just appearing on the market, which offers internet connectivity opportunities for each customer inside of the retail store. Retailers can avoid creating complex databases and rely on publicly available information published by the manufacturer of the product.

2.3 New Benefits opportunities and next steps

Motorola’s retail customers have seen different benefits implementing MC 17 like self scanning approaches. In addition to the opportunity to completely dissolve queuing at cashiers for the people using self scanning devices, product information can be provided at any time during the shopping process to push purchasing. Moving from an e-commerce into an m-commerce world, wireless technology allows the retailers to stronger link themselves with their customers. Direct 1-to-1 marketing campaigns provide much higher success rates than traditional mass market promotions. As a result SMS and PUSH messages based on RFID or barcode to personal devices have a positive impact on revenue growth. The already explained future opportunity to host different business applications directly on the mobile phone will offer the opportunity to dramatically increase customer attention. Wireless push technology will be able to update the retailers’ space on the mobile phone based on customers’ behavior or other events like children’s birthdays etc. In the future loyalty programs will be able to offer coupons for children which might be linked directly to specific products e.g. books and cannot be used by the child to purchase e.g. a new video games. Finally contactless payment methods integrated into mobile devices will allow new check-out procedures in the retail floor. Especially for micro payments not requiring a printed receipt new concepts have to be developed for the different kind of retail shops e.g. apparel, supermarket, convenience store or petrol station.

To further define the next needs of the retail industry Motorola has joined the EU FP6 StoLPaN project, where a special retail track is focusing on the benefits of retailers using new wireless and in particular NFC technology. Despite the fact that some requirements of the retail industry are not quite in line with wireless providers' point of view, the targeted customer applications are common and wireless providers are eager to get into the B2C market. Upcoming visions will not only be based on over the air configurable smart cards integrated into the phone, but they also will be leveraging new mobilized wireless infrastructure integrated into different objects, e.g. point of sales terminals for loyalty payment integrated into shopping cart bars. Cheap mass market data capturing technologies are a key for success in the retail industry. Therefore it will be interesting if low cost printed and disposable RFID tags will be drifting into the retail space. Not to replace UHF RFID tags in the supply chain but to offer customer services based on low cost, large area sensing data on item level. EPC numbers and sensor data access-able by millions of users will change the way people interact with perishable food and plants as well as with moisture or pressure sensitive goods.

3. Ambient Assisted Living Space

3.1 Overview of Motorola's activities

It is becoming important to develop solutions that support public and private health and care services to manage their resources efficiently and in a cost-effective manner and at the same time to improve the quality of life by helping elderly people to live a good life in their familiar environment with the least possible dependency on care services. Motorola is collaborating in the EU funded project PERSONA on a scalable open standard technological platform to enable a broad range of AAL Services for social inclusion, for support in daily life activities, for early risk detection, for personal protection from health and environmental risks and for support in mobility and displacements within the neighborhood/town.

3.2 Things to Things Technologies

ICT offers important means to address the challenges of independent living and inclusion by, for example, extending the time during which elderly people can live independently in their preferred environment and by providing a basis for a new generation of inclusive products and services that will help integrate people who are at risk of exclusion.

Relevant technologies supporting this trend are:

- Low cost printed sensors for environmental and user condition monitoring
- Smart textiles and other un-intrusive devices for user state monitoring
- User state discovery and interpretation based on history, activity and context
- Anticipation and forecast of user intention based on user state, needs, desires, daily work flow and interaction
- User state based content delivery, service composition and security
- Seamless service access
- Multi-protocol gateways and multi-radio communication management
- Multimodal interface and interaction

Technologies to be developed shall consider the health, psychological & well-being status of the elderly, needs, preferences, fears and concerns as well as the role in a social environment. In order to make the environment and services reactive to user needs, it is necessary that further research is addressed to intelligent and adaptable user profiling, user motivation and adaptable support, user's behavior learning and prediction.

3.3 New Benefits opportunities and next steps

Based on the user needs the PERSONA project identified different scenarios and use cases and related business opportunities. For example, the service "Remote Rehabilitation" has a number of benefits for the elderly and the welfare organization contributing to the general concept of health and social integration of an elderly, as well as considering cost and quality aspects of the welfare system and therapies applied to elderly people. The PERSONA system is able to recognize any healthy risk during the rehabilitation and to inform the therapist. The PERSONA system ensures that more patients can be served at the same time using remote connections and control providing a potential for cost reduction. PERSONA helps the therapists to monitor the patients remotely to ensure the high quality standard in providing instruction and individual feedback to the patients. The personal real time feedback improves the customer relationship and the compliance to the therapy. Backbone is the PERSONA architecture enabling the seamless integration and interaction between different spaces and the access on demand to services and their personalization. PERSONA proposes a physical and a logical architecture for AAL spaces. The abstract physical architecture treats an AAL space as a dynamic ensemble of networked nodes and the logical architecture abstracts all functionality not leading to context, input, or output events (or the direct processing of such events) as service. The proposed architecture is service oriented. The explicit distinction between context events, input events, output events, and service requests and responses, the coherence of the resulting system is guaranteed based on a modeling of the basic data flow, which leads to the identification of a set of four communication buses, namely a context bus, an input bus, an output bus, and a service bus. In a next evaluation cycle the use cases illustrated in mock-ups and demonstrated to end user and welfare organizations as well as stakeholders. The feedback will be used to refine the services and interaction flow and result in an exemplary implementation of the references architecture. In a final step the implementation will be demonstrated in trials at pilot sites involved in the project.

4. The Hospital Space

4.1 Overview of Motorola's activities

Motorola T2TRC has developed a real-time asset tracking system to improve operation efficiency of modern workplaces by tracking mobile assets in real-time by applying wireless sensor networks (WSN). The real-time asset tracking system has been tried out in a real hospital environment to test out its effectiveness and how users accept the new ways of working using the system. The system reduces time to locate assets, and can assist patient in times, and ultimately it improves quality of service by medical staff and improved utilization of the high-priced medical equipments. Further, we prove the effectiveness of WSN. These asset tracking technologies can also be deployed in heavy industry, shipyards, and logistics by locating parts or packages in time.

4.2 Things to Things Technologies

The major things-to-things technologies to aid hospital operation in tracking medical equipments are designing and implementing low-power and low-cost wireless sensor networks, location algorithms, software middleware and back-end applications. Motorola's asset tracking system systems uses 802.15.4 based MAC and ZigBee-like network protocols. To support low-power operation, the system utilizes low-power MAC and network protocols, and conserves power as much as possible. The asset only updates its location when there are meaningful movements. The sensor node itself filters all the transient movements.

4.3 New Benefits opportunities and next steps

The asset tracking system brings many benefits to medical staff and hospital's administration. From post-trial survey, nurses stated that it takes less time to locate assets, and they can spend more time with patients. The survey also indicated that medical equipments are more visible now. This translates to high utilization of capital medical equipments, brings more effective operational efficiency to the hospital and ultimately improves the quality of service by medical staff. One finding worth mentioning is that nurses show slight resistance to the new technologies brought to the hospital in the post-trial survey. Motorola T2TRC is working to improving the system in the two ways. The asset tracking system does not interface/communicate with medical equipments the nodes are tracking. The value of the system can greatly improve if nodes and medical equipments are communicating. However, there are many obstacles to accomplish. One of them is to standardize sensor nodes' interfaces to medical equipment, and we are currently working with medical equipment manufacturers to collect consensus on this interface issue. The other thing we are planning on doing is to extract valuable information by mining the movement pattern and utilization of the medical devices. As the hospital administration states, the mined information can give us and hospital administration information they have never seen before, such as patients' movement patterns, operational efficiency of each medical device, and so on.

5. Summary

These are just three examples of what the Internet of Things could bring in the foreseeable future. The reality will likely become even more amazing as the Internet of Thing world evolves. New wireless solutions will upraise in the areas of ambient assisted living, retail and in the enterprise space. The Internet of Things is one component of Motorola's vision, and future wireless solutions will be able to answer not only the WHAT IS but also the HOW IS questions about each product and object.

IOT early possibilities in learning scenarios.

Gustavo Ramírez González, Mario Muñoz Organero, Carlos Delgado Kloos
Carlos III University of Madrid, Leganes Campus, Madrid, Spain.
gramirez@inv.it.uc3m.es, munozm@it.uc3m.es, cdk@it.uc3m.es

Abstract

This paper is part of the Work in progress inside the Mosaic Project [1], a project under development involving 6 universities in Spain. Some of the objectives of this project are to explore new alternatives for using mobile and pervasive technologies in education and to develop middleware technologies to support them. This paper explains a set of basic learning concepts where technologies from the Internet of Things (IOT) can be used, makes a review of some of these technologies and proposes a set of generic touching learning scenarios showing part of some implementations to be evaluated in real scenarios.

1. Problem statement

E-learning is defined in Wikipedia as "a general term used to refer to computer-enhanced learning". Although in this definition different kinds of technologies are involved the most currently relevant issues of e-learning are determined by the use of Learning Management Systems (LMS). An LMS is a set of applications and services to deliver educational content and develop classes, learning activities and assessments. The LMS concept is also related to different ways of learning and interacting as learning with tangible interaction, ubiquitous and mobile learning and informal learning. In most of the cases, the relevant underneath technology used is the Internet, but here a question emerges: What happens if an IOT is used in this process instead of the traditional Internet? To study this problem, a set of IOT enabling technologies will be listed and some of them, with some relevance experience, will be analyzed in section 2. This paper will focus on mobile devices due to their popular use among students as part of their culture [2][3]. Then in section 3 some basic concepts of e-learning will be exposed. These concepts are selected due to their high potential to be enhanced by mobile devices. Then, by joining these portions of the problem, a set of generic touching learning scenarios will be exposed and some actual and future implementations integrated in real scenarios will be shown.

2. IOT enabling technologies and some experiences

The new field of "ubiquitous computing" [4] or "ambient intelligence" [5] has brought computing capabilities to the physical context and has expanded the intelligence of objects surrounding us [6, 7]. Actually we have gone from a smart place to smart objects in which objects can interact with each other and with people. This field of "ubiquitous computing" has shown a constant evolution due to the integration of computing and communication technologies in mobile phones [8]. This increases the number of application scenarios due the fact that mobile devices tend to be personal and in most cases private [9]. It is not surprising then that the market size for GSM mobile devices is actually about two and a half million [10].

Some of the most important technologies associated are: RFID, NFC, Bluetooth, GPS, Code Bar and Sensor networks. Any work could be categorized in one or several of the following alternatives: objects with logical representations, augmented objects, Smart Shelves, Bluetooth localization applications, smart toys, objects interacting with services and semantics with objects. We are going to divide this section into three sub-sections to summarize the related work in three main categories: information and objects, Bluetooth based solutions and NFC based solutions.

2.1 Information and objects

Some projects and papers define alternatives based on specific infrastructures and the use of tags for locating objects in everyday activities. Most of the time, this means to deploy costly devices. The works in [11][12] were some of the earliest well known initiatives to explore tag information for services in tags. Many other projects have investigated linking online information and services to physical media [13] [14]. Some projects give logic representation of objects and introduce mobile interactions [15] [16]. The work in

[17] describes a prototype that studies physical hyperlink visualisations.

2.2 Bluetooth based solutions

Other alternatives for exploring the potential of the Internet of Things are based on the use of Bluetooth. Some projects explore the use of public displays for different types of advertising [18]. One of the earliest, GroupCast [19] identifies appropriate content using profiles or information channels [20].

2.3 NFC based solutions

Other interesting technologies for the Internet of Things increase the use of the mobile phone, Near Field Communications NFC enables the user to interact with things. Some research studies have concentrated on the use of cameras in mobile phones for ambient recognition [21, 22]. Another proposal uses the presence of RFID tags to present information to users through visual signs. The work in [23] presents a framework for physical-mobile interactions in everyday life using NFC and Bluetooth-based interactions between a mobile device and a public display.

3. E-Learning key points

The e-learning and learning theory is wide and deep, for the purpose of these paper a basic review of some of the key concepts will be shown, this concerns the area where IOT technologies could have their major impact.

3.1 LMS learning management systems

Basically an LMS is an information system where different actors of the learning process (teacher and students basically) can develop courses and track student progress. Some of the most popular are .LRN[24], Webct[25] and Moodle[26]. These support different tools for education as forums, educational material, assessments and others. These tools created to develop activities in the traditional Internet but don't support activities based on mobile learning, tangible interactions and informal learning.

3.2 Learning with tangible interaction

Computers are widespread in schools around the world. However, outside the classroom different approaches to interacting with digital information and representations are emerging. These can be considered under the term 'tangible interfaces', which attempts to overcome the difference between the ways we input and control information and the ways this information is represented.

Part of the work in the area can be categorized in [27]: direct manipulation objects and interfaces, controller and representation objects, container and tokens, and embodiment and metaphor.

3.3 Ubiquitous Learning and Mobile Learning

Mobile Learning or m-learning is a concept associated with the use of mobile devices to access the contents and services [28] [29] either from distance learning management systems using Internet connections or from local context-dependent learning-aware devices and services. Some of the basic areas in which most of the projects work [30] are: Location-based and contextual learning, Design of physical spaces, social-networked mobile learning or mobile educational gaming.

The concept of m-learning is sharing a new space with the ubiquitous learning [31] concept due to the evolution of mobile phones and networks creating a suitable environment not only for distance learning service access but also for learning service execution [32].

The basic scenarios where these two concepts enable experiences are [33]: students using handheld computers, PDAs, handheld voting systems in a classroom or lecture room, students using mobile devices in the classroom to enhance group collaboration, on-the-job training, learning in museums, learning outdoors or using personal technology to support informal or lifelong learning.

3.3 Informal learning

Informal learning [34] or informal education is the learning that goes on in daily life, the learning could emerge between the interaction between people, when you see someone doing any activity or task, by listening to someone who knows something. Also implicated is the contact with contextual information such as museums or street.

4. Generic Touching Learning Scenarios

According to the sections above, a set of generic scenarios are presented. At this stage, the terms "Touching Learning" or "Learning by Touching" can be defined as: the use of services and applications in a learning environment where the learning actor can interact with environment resources for learning purposes or for communication, only by touching as expression of the possibilities of IOT in education.

Due to standard implementations, future and actual cost, NFC will be proposed as a technological enabler that can be used in several ways. These scenarios are expected to be as generic as possible to be applied in any specific instances. The first one describes the basic function of "Touching for searching". The second presents a "Personal Physical Context", this is the relationship established between a person and several objects. The third one shows the action of direct interaction and control from the mobile user to the surrounding "intelligent" objects.

4.1 Touching to Search in the Physical Context

The Web is a huge repository of information that must be organized in order to be able to find useful things for the end user. Nowadays there is a well known series of search engines. Due to their friendly web based interfaces it is easy to find information. In general there are various mechanisms to do that, but the most popular is based on references and indexation. By introducing these mechanisms, the concept of searching can be defined. The main difference between the information in the Web and the information in the physical context is that the information is associated and lives attached to a particular object, describing it and providing sometimes the access mechanisms for its remote control and adaptation. Table 1 describes a comparison between searching the web and touching to search in the physical context.

Table 1. Searching the web versus touching for Searching in Physical context.

Searching the web	*Touching for Searching the Physical Context*
Searching a word related with a concept or related words.	Searching a related word within an object or a property of an object.
Searching by words that are in relation with specific information resources.	Searching by words that are in relation with the behaviour of objects.
The pages and links do not usually change their location and their associate contents.	Objects not always stay in the same location for their entire life, but usually preserve their related information.
The goal of searching is to find information.	The goal of searching is to localize, personalize and control objects and related information.

The present proposal is oriented to obtain information related to an object. The object has a tag that contains the information itself. The mobile reads and searches information directly from the object. The search engine resides in the mobile phone and requires an active action from the user. In the implemented scenario proposed in the following section (section 5), the search is made directly by touching the objects.

4.2 Touching in Personal Physical Context

Touching in personal physical Context is the relationship established between a person and several objects that personalizes the behavior of the objects according to the mobile user preferences. Mobile users with an NFC enabled mobile device can personalize the behavior of their local environment simply by touching the objects, exchanging profiles and taking the appropriate actions. A person in his daily routine interacts with diverse physical objects. This interaction in an NFC-enabled scenario is produced by the contact of mobile devices with physical objects with different purposes (see Table 2). Due to this relation, the information obtained by the physical contact with objects can be stored, analyzed and processed in the mobile device or can be shared and synchronized with other personal devices or with external computers depending on the information in the user profile.

Table 2. Purpose of touching

Purpose	*scenarios*
Get Information	Only Read specific information. I.e. smart poster, smart signals.
Search Information	Search from several objects i.e. searching for a book, searching for a specific property on a cabinet.
Edit Information	When objects have capabilities to read and write, not only passive information in tag (tags are editable).

4.3 Touching to Adapt Physical Context

This third scenario establishes an association between a user and the surrounding objects which is by itself active. The physical environment of the user adapts to the user not only in an automatic process which is the result of the exchange and processing of user and object profiles but in a proactive way which can be controlled by the user by means of a peer to peer communication process. Using NFC P2P capabilities allows us to introduce real scenarios implementing this idea. As an example we can find a classroom containing some devices such as an overhead projector, a bookshelf and some other devices which can be controlled by the professor or the students in some different ways depending on their profiles.

5. Earlier experiences

The following case studies have been implemented as a proof of concept for the support of NFC in Touching Learning Scenarios. As is mentioned at the beginning of this paper, these cases have been developed in the context of the MOSAIC Project [1]. For the implementation the NOKIA 6131 NFC mobile phone was used and some tags based on the ISO 14443-4 specification due to the relation of cost with other IOT technologies. This want to explores the technology and its possibilities, this cases are embedded in an educational context and located in real building. Other scenarios for integrating informal and formal learning are under work. This scenario will be displayed in the Telematics engineer department of Carlos III university of Madrid.

Touching Note

The "Touching Note" makes use of an NFC tag with the information of a text note, this is placed in the door of the teacher's office to give relevant information while a person is not present. Different from the paper based system, Touching Note allows for a bidirectional communication because not only can it show a note, but it can also receive comments from the student or any one looking for you. For this concrete scenario this will be displayed in Nokia 6131 NFC mobile phones for interaction as is shown in figure 1.

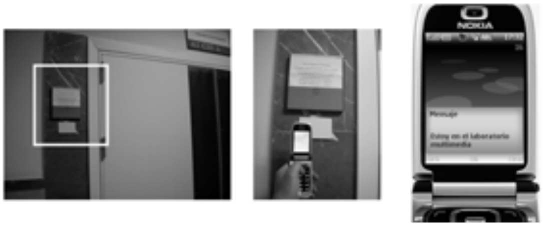


Figure 1. Scenario of "Touching Note". An office door with a tag and a screenshot of a message.

The access to the tag will be made through 2 basic profiles. The first one is the owner of the tag, who can set the information on the tag and read it later to learn about comments. The second one is the student who can read only the information and write messages flagged with the name of the student. Also in figure 1 is shown the scenario on an office door and some screenshots of the application interacting with a message.

Touching Cabinet

A place or object in a space can be tagged to give some basic information to the student only by touching. In this case the object will have a tag with textual information.

For this concrete scenario students will undertake some laboratory practice, but they will use specialized equipment. In this experiment the equipment to use are mobile phones and connection cables for different models. In this exercise a specific mobile and accessories must be found in the cabinet. The student looks for the content of the cabinet touching the tag and it displays a set of codes or names related to the content. This is a basic application where students have read only events. In figure 3 a tagged cabinet and a screenshot used are shown.

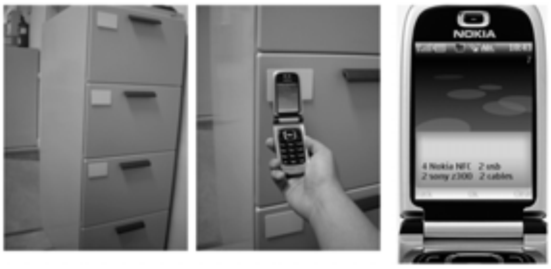
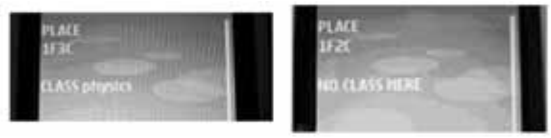


Figure 2. Scenario of "Ambient Tagging". A Tagged cabinet and a screenshot of a message.

Touching Campus

By comparing the information in the user's personal profile stored on the mobile phone and the tagged items in the learning environment, the mobile device can give recommendations or assign tasks depending on the specific context.

The application proposed for this case study is a campus recommender, which basically gives indications to new students and other people who don't know much about the campus and need to know more about the places of interest to be found or about their actual location. In this case study some places of the campus are tagged and an application running on the mobile device will provide appropriate information to the mobile user when he or she is nearby the tag containing the information. An example is presented in figure 4. In this a student with some specific subjects to take in a newly started academic year, receives information if any classes about his or her subjects are taught in that particular place. The figure shows a scenario where touching a tag in a classroom placed as in figure 2, give information if any class is there or not.



PLACE
1F3C
CLASS physics

PLACE
1F2C
NO CLASS HERE

Figure 3. Campus Recommendations by user's profile, an the working example

6. Conclusions

This paper proposes the concept of Touching Learning Scenarios using NFC as the technology to enable the interaction with physical spaces. It also proposes three generic scenarios and three cases of study. The first one tries to replace notes on a teacher's door to communicate asynchronously with students. The second one enables a searching interaction with cabinets in a laboratory scenario. The last one proposes a campus recommender for giving information to students about the classrooms.

The proposed implementations are the basic scenarios for a learning evaluation of a basic aspect of student mobility inside a physical space. According to this, the mobility in a space must be supported by pervasive options that enable transparent processes. The mobile phone is one of the best candidates for this but it can be enhanced if it is combined with other technologies such as NFC in this case.

Early analysis shows that the potentials of the IOT will derive from the use of mobile phones with enabling technologies such as NFC and Bluetooth. Their price is a predominant aspect for its mass adoption. The actors of the learning process will find the use of IOT technologies more interesting as long as the technology does not imply more effort or investment.

The early experiences presented in this paper show some of the benefits of a tagged context with personalized services. This paper has presented a deep research and new developments in this topic which are the basis for the future development of the IOT.

Currently, an evaluation is being made with a reduced number of people. Future analyses in different cultural contexts will also be appropriate.

Another key point is the suggestion of some metrics to measure the satisfaction and usefulness of all these proposals.

Currently, there is a reduced number of mobile phones supporting NFC, but in the near future and according to market tendencies, NFC could be part of the desirable features such as are nowadays represented by Bluetooth. This could help the adoption of NFC and open the door for new scenarios.

As part of our future work, more implementations are expected especially using other characteristics of NFC apart from basic tagging. Our proposal is adapted to 1kbyte and 4kbytes tags but will benefit from the current tendency in tag technologies which are bringing more and more capacity to NFC tags and at the same time, are providing more processing power to mobile devices. Additionally, the design of experiences linking the Touching Learning Environment with formal and informal learning are expected.

Acknowledgement

This work was supported by the Spanish "Programa Nacional de Tecnologías de la Sociedad de la Información" through the projects TSI2005-08225-C07-01 and -02. Gustavo Ramirez is funded by the EU Programme Alban, scholarship No. E06D101768CO and by the University of Cauca. Also thanks to Katharina E. Kinder from Lancaster University and Daniel Boos from ETH Zurich, for their contribution to the format of the paper.

7. References

[1] Mosaic Project http://mosaic.gast.it.uc3m.es/
[2] Green N. On the Move: Technology, Mobility, and the Mediation of Social Time and Space. Journal The Information Society, Published By: Taylor & Francis. Volume 18, Number 4, 1 July 2002 , pp. 281-292(12)
[3] Katz. J. Mobile Communication and the Transformation of Daily Life: The Next Phase of Research on Mobiles. Knowledge, Technology, and Policy. Vol 19 number 1. Springer Netherlands 2006
[4] Weiser, M. The Computer for the Twenty-First Century. Scientific American 265(3), pp. 94-104, September 1991.
[5] Ahola J (2001) Ambient Intelligence, ERCIM News, No 47, October 2001. Available in: http://www.ercim.org/publication/Ercim_News/enw47/intro.html
[6] K. Cheverst, N. Davies, K. Mitchell and A. Friday, The Role of Connectivity in Supporting Context-Sensitive Applications, Proc. of First International Symposium on Handheld and Ubiquitous Computing (HUC99). Lecture Notes in Computer Science No. 1707, Springer-Verlag. 1999.
[7] A. Dix, T. Rodden, N. Davies, J. Trevor, A. Friday, and K. Palfreyman. Exploiting Space and Location as a Design Framework for Interactive Mobile Systems, ACM Transactions on Computer-Human Interaction (TOCHI) 7(3), September 2000, pp. 285-321.
[8] Geser, H. Mobile phones are everywhere: Towards a sociological theory of the mobile phone. (2004). Available in http://socio.ch/mobile/t_geser1.pdf

[9] Christian Frank, Christof Roduner, Philipp Bolliger, Chie Noda, Wolfgang Kellerer. Service Architecture for Monitoring Physical Objects Using Mobile Phones. Proceedings of the 7th International Workshop on Applications and Services in Wireless Networks (ASWN 2007). Santander, Spain, May 2007

[10] GSM World web page http://www.gsmworld.com/gsmastats.shtml

[11] RoyWant, Kenneth P. Fishkin, Anuj Gujar, and Beverly L. Harrison. Bridging physical and virtual worlds with electronic tags. In CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 370–377, New York, NY, USA, 1999. ACM Press.

[12] Tim Kindberg, John Barton, Jeff Morgan, Gene Becker, Debbie Caswell, Philippe Debaty, Gita Gopal, Marcos Frid, Venky Krishnan, Howard Morris, John Schettino, Bill Serra, and Mirjana Spasojevic. People, places, things: web presence for the real world. Mobile Networks and Applications, 7(5):365–376, 2002.

[13] J. Barton, P. Goddi, and M. Spasojevic. Creating and experiencing ubimedia. HP Labs Technical Report HPL-2003-38, 2003.

[14.] M. Smith, D. Davenport, H. Hwa, and L. Mui. The annotated planet: A mobile platform for object and location annotation. In 1st Int. Workshop on Ubiquitous Systems for Supporting Social Interaction and Face-to-Face Communication in Public Spaces at UbiComp 2003, October 2003.

[15] Portolano Project homepage at http://www.cs.wasington.edu/research/portolano/

[16]Counter Intelligence Project Homepage at http://www.media.mit.edu/ci/

[17] Hovering: Visualising RFID Hyperlinks in a Mobile Phone. Pasi Välkkynen (VTT Technical Research Centre of Finland) Mobile Interaction with the Real World (MIRW 2006). September 2006.

[18] Anand Ranganathan and Roy H. Campbell, 'Advertising in a pervasive computing environment', in WMC '02: Proc. 2nd Int. Wkshp. on Mobile commerce, pp. 10–14, New York, NY, USA, (2002). ACM Press.2002

[19] Joseph F. McCarthy, Tony J. Costa, and Edy S. Liongosari, 'Unicast, outcast & groupcast: Three steps toward ubiquitous, peripheral displays', in UbiComp '01: Proc. 3rd Int. Conf. on Ubiquitous Computing, pp. 332–345, London, UK, (2001). Springer-Verlag. 2001.

[20] Salminen, T., Hosio, S, Riekki, J. Enhancing Bluetooth Connectivity with RFID. In Proceedings of PerCom, 36-41. IEEE Computer Society, 2006

[21] Scott, D., Sharp, R., Madhavapeddy, A., and Upton, E."Using visual tags to bypass Bluetooth device discovery".SIGMOBILE Mob. Comput. Commun. Rev. 9, 1 (Jan. 2005)

[22] Rohs, M., Gfeller, B. "Using Camera-Equipped Mobile Phones for Interacting with Real-World objects." In: Advances in Pervasive Computing, Austrian Computer Society (OCG), ISBN 3-85403-176-9, pp. 265-271, Vienna, Austria, April 2004.

[23] E. Rukzio, S. Wetzstein, A. Schmidt. A Framework for Mobile Interactions with the Physical World Invited paper special session "Simplification of user access to ubiquitous ICT services" at the Wireless Personal Multimedia Communication (WPMC'05) conference, Sept 18-22, Aalborg, Denmark 2005

[24] Webct http://www.webct.com/

[25] Moodle http://www.moodle.org/

[26] .LRN Web page http://www.dotlrn.org/

[27] Malley C, Stanton D. Learning with tangible technologies. Future Labs report. December 2004.

[28] A Trifonova, M Ronchetti, "A general architecture to support mobility in learning". Proc. of the IEEE ICALT Conf. 2004

[29] Sharma, S.K. and Kitchens, F.L. "Web Services Architecture for M-Learning Electronic". Journal on e-Learning, Vol. 2, No. 1, pp.203–216. 2004

[30] Naismith L, Lonsdale P, Vavoula G and Sharples M, Literature Review in Mobile Technologies and Learning. Future Labs Report 2007.

[31] Ubiquitous computing technologies for ubiquitous learning Sakamura, K.; Koshizuka, N.; Wireless and Mobile Technologies in Education, 2005. WMTE 2005. IEEE International Workshop on 28-30 Nov. 2005 Page(s):11 - 20

[32] Yu D., Zhang W., Chen X., (2006), 'New Generation of E-Learning Technologies', in IMSCCS '06, First International Multi-Symposium on Computer and Computational Sciences 2006. Volume 2, 20-24 April 2006.

[33] Attewell J, Savill-Smith. Learning with mobile devices: research and development – a book of papers. London: Learning and Skills Development Agency. At www.LSDA.org.uk/files/PDF/1440.pdf, accessed July 2004

[34] Informal Education org web site http://www.infed.org

[35] Overwien, B. (2000) Informal learning and the role of social movements, International Review of Education 46 (6) 621-640.

Application Scenarios for Cooperating Objects and their Social, Legal and Ethical Challenges

Martin Meister[1] Marcelo Pias[2] Eric Töpfer[1] George Coulouris[1]

1 Center Technology and Society
Berlin Institute of Technology
Hardenbergstrae 36a, 10623 Berlin, Germany
{meister,toepfer}@ztg.tu-berlin.de

2 Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue,
Cambridge CB3 0FD, UK
{mp431,gfc22}@cam.ac.uk

Abstract

The realisation of visions for the application of Cooperating Objects Technologies will certainly not only face technical roadblocks but also a variety of non-technical challenges when it comes to the adoption, experience, adaptation and reaction by users and the wider society. This paper aims to present a selection of application visions elicited in the context of the EU coordinated action Embedded WiSeNts, and to explore some of the non-technical challenges that might arise from these scenarios. This paper points at research methods to address these challenges and focusses on realised scenarios.

1. Introduction

In the early 1990s Mark Weiser [25, 26] outlined his vision of the computer for the 21st century at Xerox PARC as ubiquitous calm technology which weaves itself into the fabric of everyday life until becoming indistinguishable from it. Since then progress made in the domains of processing power, data storage capacities, wireless network technologies, human-machine interfaces, miniaturisation and convergence of devices is impressive.

As Greenfield [7] notes there are many definitions of ubiquitous computing and they overlap with pervasive computing, physical computing, tangible media, ambient intelligence, Internet of Things, or to use his term, 'everyware'. Greenfield argues that it has reached something like a critical mass of thought and innovation by 2005 and is thus prepared to colonise everyday life to remake the very relations that define our lives. Significantly contributing to this critical mass is the Embedded WiSeNts project funded by the European Commission [1] that integrated a network of leading European academic research labs in the areas of embedded systems, ubiquitous computing and wireless sensor networks. The WiSeNts project published a series of studies on the state of the art of Cooperating Objects [16] and devised a research roadmap that could be used for shaping future research programmes. In this context, Cooperating Objects are understood as small computing devices equipped with wireless communication capabilities that are able to cooperate and organize themselves autonomously into networks to achieve a common task [15], i.e. systems that integrate different yet complementary aspects of embedded systems, ubiquitous computing and wireless sensor networks. An important task for the Embedded WiSeNts project was to summarise application scenarios that can be well understood and characterised today. This was done by a survey of ongoing or already completed projects targeting a wide range of domains in Europe and beyond. However, the project also explored application visions and their technical scenarios that could potentially be realised once wide-ranging technology of cooperating objects becomes available. The focus was on long term visions clustered around a 10-year horizon. These visions were elicited through the organisation of two competitions, i.e. 1) WiSeNts Summer School Competition 2005 and 2) Sentient Future Competition[2] 2005 that attracted 79 entries, and the request of contributions by the projects partners and the Industry Cooperation Board.

The realisation of such visions will certainly not only face technical roadblocks but also a variety of non-technical challenges (social, legal and ethical) when it comes to the adoption, experience, adaptation and reaction by users and the wider society. This paper presents a selection of such applications and explores some of the non-technical challenges that might arise from the implementation of these or

[1] http://www.embedded-wisents.org

[2] http://www.embedded-wisents.org/competition/

similar visions. It discusses research methods to address these challenges and focusses on realised scenarios.

2. Selected Application Visions

108 applications entries were received in total from the open competitions [5]. Out of this, we provide an overview of the application areas that we consider are the most interesting and could potentially affect the direction of research in the coming years. Whenever it is appropriate, we contrast such visions with those envisaged in [11, 18, 9, 10].

2.1. Transportation

The envisaged embedded sensor systems would gather data for real-time or ”close” to real-time information services provided by governmental agencies and private organisations including insurance companies. K. Pister [11] envisions that vehicles will be fully aware of the road conditions on a user’s favourite route home, not at the level of a traffic announcer telling the users that it is slow on a given motorway, for instance, but with details of the instantaneous speed and history of every vehicle between the users and their destination, as well as the ones that are likely to get on the road, should the user choose to look at that detail. Most likely the user’s agent software will just advisel which route to take, and how many minutes it will take. Similarly, visions in [5] describe scenarios where cooperating vehicles will be aware of dangers and pro-active in making semiautonomous decisions. Proactive safety systems will be placed in every car.

The co-operation will allow sensors in each vehicle to monitor the environment condition with *in-loco* air quality measurements (e.g. nitric oxide, carbon monoxide, etc). Extensive monitoring of gas emissions - a required task to make our environment more sustainable, will support countries to meet their responsibilities under the Kyoto protocol. To make transport more economically, socially and environmentally sustainable, commitment is required from different sectors of the society including travellers (they may need to change their behaviour), manufacturers of vehicles to equip them with the latest traffic and pollution monitoring and road safety technology, and the government in the investment and management of transport infrastructures. Advanced congestion-based charging (e.g dynamic pricing), supported by a pervasive distributed traffic monitoring system, will mitigate traffic congestion in urban areas.

2.2. Environmental Monitoring

Applications within this realm are of crucial importance to the scientific community and society. Thousands of square kilometres of geographical areas may be supervised and the duration for this can be years. Application scenarios envisage that cooperating objects will monitor vegetation growth and air/water quality, oil spills and will coordinate (e.g statistical sampling and data filtering) to create a big picture of natural spaces. Because of the large-scale aspect, natural disasters such as prominent flooding and earthquakes could be anticipated through improved models of the global environment. Authorities would be alerted and actions taken quickly to respond to natural disasters. Also, the management of the population’s waste could be efficient and sustainable leading to higher life quality and less costs for the city authorities. Financial incentives may be employed to encourage the correct disposal.

Carbon emissions and absorption would be measured or estimated in order to charge/ration citizens according to their consumption. Individuals can receive carbon debits for their use of energy and carbon credits for clean energy that they generate, for example, by investing in wind farms and for carbon-absorption activities including trees and other vegetation planted or invested in. Carbon debits are then converted to a tax on the individual. The direct benefits are increased environmental and public health gains. There are the risks, however, of privacy loss and fraudulent interference with sensor systems.

2.3. Health and Fitness

Merging wireless sensor technology into health, medicine and fitness applications will make life much easier for doctors, disabled people, patients and the overall population. They will also make diagnosis and consultancy processes faster with patient monitoring entities consisting of sensors. Those sensors will provide the same information regardless of location and automatic transitions from one network in a clinic to the other installed in a patient’s home will be available. As a result, high quality healthcare services will get closer to the patients. The benefits of this are clear, although shortcomings are expected too. For example, employers can demote employees based on an analysis of biomedical data (biosensors data correlated with genetics information).

The European Commission through the ISTAG has reported on the vision of well-being in the ageing society [9, 10]. The vision is the one where a new paradigm of personalised healthcare will support EU citizens in living healthy lives, minimising time in hospital, at local doctors or in care homes. Europe’s increasingly elderly population will be able to live more independently in their home environments. The service envisaged is pro-active where more personalised and preventive health care is employed as opposed to reactive methods such as treatment for the elderly. The application scenarios foresee critical diseases diagnosed by means of tele-monitoring of individuals with specialised biosensors, with some of them implanted in the human body.

K. Pister [11] goes further to envision that there will be no unanticipated illness. In his vision, sensor implants will monitor all of the major circulator systems in the human body, and provide the monitored individual with early warning of an impending flu, or save their life by catching cancer early enough that it can be completely removed surgically.

Small sensors and actuators in our clothes through 'smart' textiles will sense our physiological signals and movements in order to understand our health condition. This should provide historical data to aid in achieving precise diagnoses. Computers should be able to interpret when we perform a physical activity, for instance, walking or jogging.

Enhanced experience in fitness exercises will be achieved with useful feedback systems (e.g. audible and haptic) from tiny computers embedded in sports clothes and equipment[3]. Entertainment systems for audio and video ubiquitous in mobile phones will be part of the overall body personal system. We will communicate with our clothes, watches and other accessories and they will coordinate with other user cooperating objects.

3. Non-technical Challenges

To identify key issues and non-technical challenges posed by the potential implementation of such applications the Embedded WiSeNts project commissioned the organisation of a workshop in which 15 experts in technology assessment (TA), sociology of technology, participatory design, system analysis, communication science, privacy protection, psychology, micro economics and the philosophy of law from across Europe met for an interdisciplinary exchange. These researchers discussed relevant social, legal and ethical issues with computer scientists and engineers from the project consortium [23]. The following section will briefly summarise some major results of this workshop which reflect key findings of TA studies conducted so far [8].

3.1. In Search of Usefulness and Usability

'Not making things but making sense of things!' is the Leitmotif of Matt Jones' research at the Future Interaction Lab in Swansea, UK. Jones and others remind us that the seductive visions of scientists and engineers are not necessarily attractive and useful for 'Joe Public'. Marketing departments and designers who currently develop applications often target a 'creamy layer' of users who skilfully embrace new technologies. However, the demands, needs, abilities and skills of these segments of the population are not necessarily the same as those of the less affluent, less techno-savvy people (not to mention the poor and illiterate population of the world) who might be envisaged as users when investors, products and applications are in search of mass markets. Therefore developers should avoid charging such new technologies with extremely positive connotations which might at the end disillusion users and scare off customers. The crucial question is what might become killer applications in the dawning age of ubiquitous computing.

3.2. Protecting Privacy and Building Trust

As already noted above most of the application visions which involve humans require the collection, storage and processing of a huge amount of personal data. Even more crucial they are likely to extract new features from such data including biometric identifiers gathered from human bodies or very intimate data displaying bodily functions (health and fitness applications), or even emotions and thoughts. Thus, not surprisingly, issues of privacy, data protection and trust are seen as crucial for the future development and design of such applications as well as their regulation [21, 20, 1, 2, 4]. Many fear the end of privacy in the 21st century [28, 22]. However, it is often pointed out that among human rights privacy is perhaps the most difficult to define [14]. Even Warren and Brandeis classic definition of privacy as the the right to be let alone raises the questions when people and users wish to exercise this right, and under which conditions they are willing to accept disturbance in exchange for economic or other benefits. It is clear that privacy in this informal and personal sense is nothing fixed but is highly contingent upon the specific context. What people are willing to reveal about their multi-layered identity depends on their trust and their knowledge about the addressee of this information. In networked ubiquitous systems even data related to simple and insignificant incidents might be collected which could remain in hardware memories on huge databases. Such data could be combined, sorted and reinvented by sophisticated algorithms and transferred at a global scale to other entities for further processing for unknown purposes. It is obvious that such complex multi-lateral systems of data processing resist attempts (not only by the common user) to know and understand their underlying purposes and rationales. Given this fact, it seems very difficult to request users to make informed decisions on revealing their personal data and to generate trust in these new systems. And even if opt-in decisions are offered the multitude and complexity of choices in smart environment systems might simply overwhelm the users.

3.3. Distributing and Locating Responsibility and Liability

An issue partly overlapping with privacy and trust is the changing nature of responsibility and liability in socio-technical environments of distributed agency. It will become more or less impossible to assign responsibility for faults or malfunctions to individual humans or technical

[3] The UK-funded Sesame project is addressing wireless sensor applications to support athletes and their coaches in day-to-day training (http://www.sesame.ucl.ac.uk)

components of the extensive networked systems of cooperating objects and related technologies. Although the increasing capability to log and track activities in such systems could help to reconstruct events in the aftermath of a possible disaster, case studies of disasters in complex systems demonstrate that they might occur even when each component is working properly if different governance regimes interfere [17, 27]. From a legal perspective a simple solution to the problem could be to attribute liability to service providers or users and thus urge them to cautiously select their applications. However, to make users responsible for their learning systems might cause new problems: the safeguard to request decisions by users regarding learning algorithms could, on the one hand, help to make their actions transparent and calculable, while limiting a key function of learning systems, i.e. context awareness. What is needed is a careful assessment of how to distribute agency in human-machine-interaction. Fair and reasonable decision on this issue will be highly contingent upon context and application and therefore require detailed research.

3.4. Bridging the Digital Divide and Being Sensitive to Digital Discrimination

Although the challenge to bridge the digital divide by ensuring equal access to ICT is widely recognised [24], another issue touching the broader question of social justice is often overlooked. Digital discrimination, the automated social sorting and prioritising of user preferences and needs, is demonstrated by the example of two users with different temperature preferences in a smart environment. The system steering the relevant conditions will be required to make a decision: it can opt either for the lower or the higher temperature, or it can opt for a compromise found in between the two user preferences with both of them eventually feeling uncomfortable. While the example might be trivial it illustrates the general problem and it is clear that scenarios with more serious implications for the users can be envisioned. Although the problem of right choice itself is not generated by the smart home but by different user preferences, the automation of environmental adaptation might hinder the dynamic negotiation between the users. Thus, visions of social relations, status and power that are intentionally or unintentionally inscribed by engineers and software programmers into design and software might be ossified over time and space [13, 6]. As these issues are closely related to issues of responsibility, again further research is needed to inform the decision making and design of cooperating object technology.

4. Addressing the Challenges: New Research Directions

The networked aspect of Cooperating Objects and related systems pose a serious challenge to traditional methods of technology assessment (TA) and technology studies. Many methodological approaches for studying and assessing the social aspects and implications of such emerging technologies are at the fingertips of interdisciplinary research: scenario building, expert interviews with marketing experts, developers and users, media analyses and ethnographic observation. But there is a problem to limit analyses, and there is a demand for a multi-method approach that combines instruments from comparative analyses, laboratory studies, user surveys and forecast. Each of these methods has its advantages and drawbacks: for instance, the selection of the interviewees in expert and user interviews is of crucial importance in particular when used as an oracle to forecast future developments of choices made about the technology to be implemented. Moreover, the right choice of methodology will depend on the features of the user groups in specific application and field of research. Not every design paradigm developed in the overall context of ubiquitous computing and ambient intelligence will meet the purpose of particular applications. Not taking into account particular technical features or social demands of applications might result on the one hand in the development of useless products and applications which simply fail to generate a market and finally prove as a waste of money. On the other hand such ignorance might turn out as user nightmares when, for instance, applications are imposed on people in a top-down approach by technophile managers dazzled by spin doctors of the supplying companies.

The workshop discussion revealed a set of multidisciplinary design issues in areas that cross boundaries between engineering, social sciences, law, economics and psychology. More research is needed drawing on these disciplines in the design of distributed object applications.

Given the broad range of possible themes, we want to highlight that many of the major issues identified so far can be addressed by conducting investigations with 'realised scenarios' with real people, in order to assess acceptability, usability and sensefulness of the visions developed empirically. Realised scenarios are real technical installations, combined with social situations of an application, on a low level of technical sophistication and social realism. They should be real enough to present experimental subjects with the scenarios to observe their behaviours. This allows ethnographic observations of man-machine-interactions, which can be combined with user and expert interviews or with scenario workshops. Empirical data gathered with such a methodology can make an important contribution to the testing and tuning of pilot applications. If it is not possible to mimic the technical functionality of an entire technical installation (e.g. for reasons of complexity), mock-up scenarios could be used. Some of the major social challenges for the design of Cooperating objects technologies may be understood through the approach of realised scenarios.

Privacy-compliant system design The important message is that early adoption of privacy issues in the design of systems should be encouraged. Otherwise, later corrections through regulatory measures can be expensive to implement, to speak in economic terms. Thus, privacy should not be dealt with as an add-on function to the system, e.g. by data filtering and minimisation measures[18]. With realised scenarios, the users' acceptance of different modes of data collection and processing can be modelled as a give-and get-game. It could be determined what type and amount of sensitive information users are willing to give to the technical system with respect to the reward they expect to receive. Instead of an overall scepticism towards these technologies, this approach could lead to a scaled and more situation-aware analysis, which will not require an overall consensus.

Reliability and Dependability A dependable system should provide, at any time, a specification of current system performance and status, often associated with levels of confidence. Dependability, then, is the property that provides the necessary support to tell the users how reliable, how available and how safe the system is. With realised scenarios, it is possible to examine the amount and the type of the necessary signals that give the users the impression of reliability and dependability of the technical system.

These signals can lead to a general visibility of the system for the users, and it is important to understand empirically whether such signals can deter or encourage certain types of user's behaviour (as examined e.g. in CCTV). This can be even a quality measure for technical systems which goes beyond the question of satisfying purely representative aesthetics (e.g. iPod).

Distributed Control: Responsibility and Liability Construction Complexity arises from the highly pervasive and distributed nature of the systems and it needs to be addressed in the design space of the cooperating objects systems. Especially the distribution of control is a crucial issue here, because the dissipation of responsibility is at stake. For example, recent developments in the aviation industry have shown that is extremely difficult to establish responsibility for accidents that occured with complex control distributed over computer systems and human decisions. To make responsibility possible, clear specifications of the functionality of systems is important, and changes and the results of processes must be properly documented. But the construction of responsibility is not a technical issue in the first place. Realised scenarios can be a way to determine how the functionality of cooperating object technologies can be designed as to appear transparent for the users and 'black box' designs can be avoided. Taking the users into account can foster transparency, activating empowerment rather than seductive convenience for specific situations. Empirical observation on this line can be oriented by the Leitmotiv of tangible, physically present technology as proposed by the concept of palpable computing [3].

Balancing technical autonomy versus human intervention To achieve the full vision of cooperating objects while addressing the legal issues calls for a balance between the autonomy of such systems and the possibilities of human intervention if necessary. Here, technical and social aspects are intermingled. One crucial point is the development of new types of user interfaces to support the interaction between humans and almost invisible autonomous objects.

Examining these issues empirically with realised scenarios will reveal different types of users with respect to different social situations and types of technical applications [12]. Moreover, classifying users by organisational roles will also have to be differentiated. Thus, the simple classifications of users applied in acceptance and usability research - e.g. the well-known typology of adopter groups [19] - will not be sufficient. For the case of the emerging technology of Cooperating Objects, it is plausible to assume from the beginning of the design process that 'the one and only human user' of Cooperating Objects technologies does not exist, but there will be different types of users and different roles of usage.

5 Acknowledgements

We would like to thank all the participants who attended the Workshop on Social Aspects of Cooperating Objects Technologies held at the Technical University of Berlin in November 2006. Further information on the workshop can be found here: http://www.embedded-wisents.org/workshop/index.html

References

[1] Bizer, Johann and Dingel, Kai and Fabian, Benjamin and Günther, Oliver and Hansen, Markus and Klafft, Michael and Möller, Jan and Spiekermann, Sarah. TAUCIS – Technikfolgenabschätzung Ubiquitäres Computing und Informationelle Selbstbestimmung: Studie im Auftrag des Bundesministeriums für Bildung und Forschung, Juli 2006.

[2] Brey, Philip. Freedom and privacy in Ambient Intelligence. *Ethics and Information Technology*, 6(7):157–166, 2005.

[3] Büscher, Monika and Christensen, Michael and Hansen, Klaus Marius and Mogensen, Preben and Shapiro, Dan. Bottom–up, top–down?: Connecting software architecture design with use. In Voß, Alex and Hartswood, Mark and Ho, Kate and Procter, Rob and Rouncefield, Mark

and Slack, Roger and Büscher, Monika, editor, *Configuring user–designer relations: Interdisciplinary perspectives*, Computer Supported Cooperative Work. Springer London, 2008.

[4] Friedewald, Michael and Lindner, Ralf. Datenschutz, Privatsphäre und Identität in intelligenten Umgebungen: Eine Szenarioanalyse. In Mattern, Friedemann, editor, *Die Informatisierung des Alltags: Leben in smarten Umgebungen*, pages 207–231. Springer, 2007.

[5] George Coulouris et al. Visions for Innovative Applications and their Social, Legal and Ethical impact. Technical Report Report available `http://www.embedded-wisents.org/studies/studies_wp3.html#Study5`, Embedded WiSeNts Project, 2006.

[6] Graham, Stephen and Wood, David. Digitizing surveillance: Categorisation, space, inequality. *Critical Social Policy*, 23(2):227–248, 2003.

[7] Greenfield, Adam. *Everyware: The dawning age of ubiquitous computing*. New Riders, Berkeley CA, 2006.

[8] Hilty, Lorenz and Behrendt, Siegfried and Binswanger, Mathias and Bruinink, Arend and Erdmann, Lorenz and Fröhlich, Jürg and Köhler, Andreas and Kuster, Niels and Som, Claudia and Würtenberger, Felix. *Das Vorsorgeprinzip in der Informationsgesellschaft: Auswirkungen des Pervasive Computing auf Gesundheit und Umwelt: Studie des Zentrums für Technologiefolgen–Abschätzung*. TA–SWISS, Bern, 2003.

[9] IST Advisory Group (ISTAG). ISTAG REPORT on Orientations for Work Programme in FP7. Technical report, European Commission, June 2006.

[10] IST Advisory Group (ISTAG). Shaping Europe's Future through ICT. Technical report, European Commission, Mar. 2006.

[11] K. Pister. Sensor Networks in 2010. Technical report, University of California Berkeley, 2001.

[12] Kindberg, T. and Sellen, A. and Geelhoed, E. Security and Trust in Mobile Interactions: A Study of Users' Perceptions and Reasoning. Consumer Applications and Systems Laboratory. Technical Report HPL-2004-113, HP Laboratories Bristol, 2004.

[13] D. Lyon, editor. *Surveillance as social sorting: Privacy, risk and digital discrimination*. Routledge, London, 2003.

[14] Lyon, David. *The electronic eye: The rise of surveillance society*. University of Minnesota Press, Minneapolis, 1994.

[15] Marrón, Pedro José and Minder, Daniel and Embedded WiSeNts Consortium. *Embedded WiSeNts research roadmap*. Logos–Verlag, Berlin, 2006.

[16] Michel Banatre and Pedro Jose Marron and Anibal Ollero and Adam Wolisz, editor. *Cooperating Embedded Systems and Wireless Sensor Networks*. Wiley, ISTE, London, Jan 2008.

[17] Perrow, Charles. *Normale Katastrophen: Die unvermeidbaren Risiken der Großtechnik*, volume 1028 of *Reihe Campus*. Campus, Frankfurt am Main, New York, 2 edition, 1992.

[18] Peter Gabriel et al. *Pervasive Computing: Trends and Impacts*. SecuMedia Verlag, Ingelheim, 2006.

[19] Rogers, E. M. *Diffusion of Innovations*. Free Press, New York (USA), fifth edition, 2003.

[20] Roßnagel, Alexander. Datenschutz im 21. Jahrhundert. *Aus Politik und Zeitgeschichte*, (5–6):9–15, 2006.

[21] Roßnagel, Alexander and Pfitzmann, Andreas and Garstka, Hansjürgen. *Modernisierung des Datenschutzrechtes: Gutachten im Auftrag des Bundesministeriums des Innern*. Bundesministerium des Innern, Berlin, 2001.

[22] Schaar, Peter. *Das Ende der Privatsphäre: Der Weg in die Überwachungsgesellschaft*. Bertelsmann, München, 2007.

[23] Töpfer, Eric and Meister, Martin. Social aspects of Cooperating Objects Technologies: Workshop report: International Workshop of the Coordinated Action Embedded WiSeNts(IST–004400), 2006.

[24] UNESCO. Information for All Programme (IFAP). Ethical implications of emerging technologies: A survey. Paris.

[25] Weiser, Mark. The computer for the 21st century. *Scientific American*, (265):94–104, 1991.

[26] Weiser, Mark and Brown, John Seeley. Designing calm technology, 1995.

[27] Weyer, Johannes. Modes of governance of hybrid systems: The Mid–Air collision at Ueberlingen and the impact of smart technology. *Science, Technology & Innovation Studies*, 2(2):127–149, 2006.

[28] Whitaker, Reginald. *The end of privacy: How total surveillance is becoming a reality*. New Press, New York, 1999.

Designing for palpability in event management and emergency response

Jesper Wolff Olsen
Department of Computer Science, University of Aarhus, Denmark
jexper@daimi.au.dk

Monika Buscher
Department of Sociology.
Lancaster University.
Lancaster, LA1 4YD UK
m.buscher@lancaster.ac.uk

Michael Christensen,
Computer Science Department
Aarhus University.
N- 8200 Aarhus, Denmark
toby@daimi.au.dk

Margit Kristensen
Computer Science Department
Aarhus University.
N- 8200 Aarhus, Denmark
margit@daimi.au.dk

Morten Kyng
Computer Science Department
Aarhus University.
N- 8200 Aarhus, Denmark
preben@daimi.au.dk

Preben Holst Mogensen
Computer Science Department
Aarhus University.
N- 8200 Aarhus, Denmark
preben@daimi.au.dk

Dan Shapiro
Department of Sociology.
Lancaster University.
Lancaster, LA1 4YD UK
d.shapiro@lancaster.ac.uk

Dominic Greenwood
Whitestein Technologies AG
Pestalozzistrasse 24,
CH-8032 Zurich, Switzerland
dgr@whitestein.com

Abstract – Invited Talk

The presentation takes a case study as a point of departure, which describes the design and test of a large scale prototype in a real life setting: The MIO (Major Incidents Overview) prototype used during the Tall Ships Races event 2007. The prototype was developed within the PalCom project, a large interdisciplinary collaboration concerned with the design of an open software architecture and conceptual framework for 'palpable' pervasive computing. The MIO prototype has been and still is being designed to support situation awareness during emergencies and in the management of large events. The background, followed by an introduction to the MIO prototype and the open architecture, and a detailed description of why and how the real-life test was carried out during the Tall Ships Races is going to be presented. Conclusively, the presentation ends with a discussion of key observations and evaluations of this experience of balancing innovative prototype technologies into real life through interdisciplinary collaboration.

NEMO project: Using technology models to explore organizational issues & Using ethnography to explore cultural logics in design, deployment and use

Katharina E. Kinder, Gerd Kortuem
Lancaster University, Lancaster, UK
k.kinder@lancaster.ac.uk , kortuem@comp.lancs.ac.uk

Abstract – Invited Talk

The NEMO project at Lancaster University investigates how the Internet of Things might affect workplaces in the road construction and maintenance industry. In collaboration between three departments (Computing, Management and Psychology) the project uses a multi-method approach to design and realize concrete technology artifacts and to assess their potential impact on people and organizations. In this talk we discuss a case study of a sensor-based system designed to reduce the number of incidents of "vibration white finger" (VWF) at construction sites. The system uses wireless sensor nodes for monitoring workers' exposure to vibrations and testing of compliance with legal health and safety regulations. Using data collected over a two year period through field trials and interviews, we clarify the impact of this technology on the relationship between management and operatives, the formulation of health and safety rules and the risk perception and risk behaviour of operatives. In particular, we contrast models of technology, in this case sensor network inspired and smart artefact inspired compliance systems, and make the case that these technology models have a strong influence on the linkage between technology and organization. Furthermore, we highlight the importance of cultural logics such as health and safety discourse, liabilities and audit for the legitimization of ubiquitous computing technologies during their introduction and deployment at industrial workplaces. This leads us to conclude that we need to look at the 'broader picture' of current social and cultural logics in contemporary society in order to understand and improve design, deployment and use of internet of things technologies.

Uncovering Cultural Issues in the Internet of Things: A Design Method

Basile Zimmermann, PhD, Maître assistant
Unit of Chinese Studies, University of Geneva
basile.zimmermann@lettres.unige.ch

Abstract

This paper presents a design method to deal with cultural issues when developing or analysing new technologies. The argument is based on the Actor-Network Theory (ANT) approach, as well as several years of research by the author on human-machine interaction in the People's Republic of China [1]. The Domain Name System (DNS) technology, as currently used in China, is discussed as a real-world example.

1. Introduction

The words 'technology' and 'culture' have both a long and complicated history. Their study, throughout their many different definitions, can be linked to several –if not all– disciplines in the humanities, natural or social sciences. Since one goal of this workshop is the interdisciplinary exchange between scholars from different fields, this paper does not discuss in detail known theoretical issues, but presents its argument mostly by means of one specific case: the Domain Name System (DNS), as it is used in today's World Wide Web, and provide two illustrations of its impact in the People's Republic of China.

2. Theoretical Background

The discussion is mainly based on three insights. First, the general statement by Melvin Kranzberg that "technology is neither good or bad, nor is it neutral" [2]. Second, the demonstration by Madeleine Akrich that technical objects embody *scripts* that tell users what their producer expects them do; "(…) like a film script, technical objects define a framework of action together with the actors and the space in which they are supposed to act" [3]. Third, the discussions by Bruno Latour on the agency of things, in which he shows how objects act in a very similar way to human beings. He compares, for example, a door-closer to a groom, and states: "(…) everytime you want to know what a nonhuman does, simply imagine what other humans or other nonhumans would have to do where this character not present" [4].

Linked together, these three points allow us to argue that technology, no matter what or where it is, has some kind of agency, and that this agency comes, one way or another, into contact with a local context.

3. 'Cultural' Aspects

In the lines that follow, a 'cultural issue' will be considered as a problem that occurs in the application of a new technology in a certain context, no matter what the technology or the context is.

This formulation may (and actually should) be looked at with suspicion by academics, who traditionally have a strong inclination for precise definitions of terms, objects of analysis, or discourse. The argument here is that the choice of the object(s) of analysis, as well as its type, to be used in the design method, is to be made *by the user* of the method, not by the method itself.

Hence, 'cultural aspects' are to be considered in their most common meaning, pointing at *ways of life* and *ways of thinking*, leading to the most general idea of *difference(s) between groups of people, groups of ideas, or groups of things*.

The design method which is at the core of this paper intends to help the user to see a particular kind of relationship between the data they have chosen –an *articulation*, as we will see–, and in no way defines what the data consist or should consist of.

3. The Design Method

The method consists of five steps. 1. *List the 'contents' of the technology*. 2. *List the differences between the listed contents and the 'context where the technology is used'*. 3. *Consider whether the technology was, is, or will be used, and if yes for how long*. 4. *Consider whether the technology can be modified by the user or not, at what cost*. 5. *Conclude* *by considering that the 'cultural issues' are the elements listed in point 2, and their importance must be evaluated with regard to the variables given in points 3 and 4.*

Here is an example of its application, in the case of the Domain Name System as it is currently used in China.

3.1. Step one: List the contents of the 'technology'

We consider here that Domain Name System (DNS) technology is what 'translates' the name of a resource to its physical Internet Protocol address (IP). For example, in today's World Wide Web, most computers have a number that helps differentiate them (in a similar way to phones). The IP of the website of the University of Zurich in Switzerland, at the moment of writing, is <http://130.60.127.170>. Since numbers are inconvenient for human beings to memorize (and for some other reasons as well), the Domain Name System has been implemented at the beginning of the 1980s, so that whenever someone types <http://www.unizh.ch/>, they are automatically redirected to <http://130.60.127.170>. Simply put, DNS makes human users and connected computers happy, as they are all able to communicate between each other in a handy way.

So what are 'the contents' of DNS? For sure, this question is a tricky one. Fortunately, since the design method states that the objects of analysis are to be chosen by the user, it doesn't really matter whether we list all of them, or only part of them. The only important point is to list at least *some of them*. If we do so, we are going to see some results. And if these are not satisfactory, we can always go back to step one and try to find some more.

Direct descriptions often give good results. Looking at <http://www.unizh.ch/>, we notice that what we have labeled with the words 'DNS contents' are Roman-alphabet letters, mixed with some sort of punctuation marks, which then disappears into a black-boxed networked process which only computer engineers have access to. The whole process' results, in the end seems to be <http://130.60.127.170>, some sort of translation where the original inputted contents have been replaced by Arabic numerals, which are communicated internally (the user, in most cases, does not see the numbers above on her browser although they are being used behind the screen).

So for step one, we can list: a) Roman-alphabet letters, b) punctuation marks, c) Arabic numbers, d) black-boxed computer process. Let us move to step two.

3.2. Differences Between 'Listed Contents of the Technology' and the 'Local Context'

Our second task is to make a second list which will consist of comparisons between the previously listed contents and the so-called 'local context'. Point d), the black-boxed process, is not within our reach, we have no choice but to leave it aside for the moment. Point c), Arabic numbers, are widely used in China, in a very similar way as in Europe. A quick comparison (e.g. by looking at printed materials in Chinese), tells us they cannot be considered as a real 'difference', at least not in a very obvious way. Same problem with b) the punctuation marks (the Chinese do have their own punctuation particularities, but it isn't something that could be easily called a huge cultural difference). However, point a), the Roman-alphabet *does* provide us with a huge difference: the Chinese use Chinese characters, they do not use Roman-alphabet letters to write.

Since we have found at least one element to write down on our list for point number 2. We can move to step 3.

3.3. The Use of Technology

Our third task is to consider whether DNS technology has been used, is being used, or will be used, and if yes for how long.

A quick look at the history of the development of the Internet in the PRC shows that China, although a few years behind, has pretty much followed the steps of its Westerners predecessors in matters of communication technologies. While DNS was first implemented in the United States at the beginning of the 1980s, China's first moves toward the Internet began at the end of the 1980s, and its wider spread in the country started at the end of the 1990s. [5]

Today, most Chinese websites and web technology are similar in many aspects to those of the West, and are mainly based on DNS (see below for a few words on IDNA). According to current statistics, China has 210 million users today (about the same as the United States of America) who surf regularly on DNS implemented web technology [6]. We can say it is widely used.

3.4. Can the 'Technology' Be Modified by the User?

In short, DNS technology cannot be modified by its average user. It relies on a complicated, a heterogenous network of machines and human beings that is not easily changed. It involves political, economical, and legal means, both at national and international levels. Although a complementary implementation is currently available (Internationalizing Domain Names in Applications or IDNA) and is slowly put into use, in China, at the moment of writing, the DNS technology is still inescapable for the average Chinese user. He or she, has no choice but to follow the rules if they want to be part of the networks it relates to.

3.5. Uncovering 'Cultural Issues'

According to the design method, *the 'cultural issues' are the elements listed in point 2, and their importance must be evaluated with regard to the variables given in points 3 and 4.* The main difference between DNS technology and the Chinese environment listed in step 2 was the one between the Roman-alphabet and the Chinese characters. Points 3 and 4 have shown us that DNS was indeed used in China, and that Chinese users could not easily adapt the DNS technology to their own needs.

This 'difference' between DNS technology and the Chinese language is not a small one. It implies that any Chinese character has to be 'translated' into Roman-alphabet, before it can be routed by the DNS to an IP address. Fortunately, the People's Republic of China does have an official Roman-alphabet transcription of the pronunciation of Chinese characters that can be used for this purpose. 中国 (« China », in Chinese) for example, is transcribed *Zhong Guo*.

For most Chinese people, who are fond of the Chinese script, the phonetic transcription has two shortcomings. First, it is ugly, and difficult to read. Most Chinese use Chinese characters to read, not Roman-alphabet letters, and many of them are not very familiar with the latin script. To a native Chinese, *Zhong Guo* is much more uncomfortable for the eyes than 中国. Second, the phonetic transcription is confusing. Chinese characters are symbols of meaning, and many share similar phonetics. The transcription *yi*, for example, corresponds to more than 300 different Chinese characters. That makes the meaning of, say, <www.yi.com>, pretty obscure for a Chinese native, who has no clue whether it relates to <www.衣.com> (« clothing »), <www.医.com> (« medical science »), <www.移.com> (« move »), or any other sound mates.

So, roughly sketched out, we see a big 'difference' between DNS technology and Chinese culture. Where can its consequences be observed in China today? As the design method states: whenever and wherever DNS technology is used and cannot be changed. Here are two illustrations of the phenomenon.

The photograph below was taken in a street next to the South gate of Peking University in August 2007.

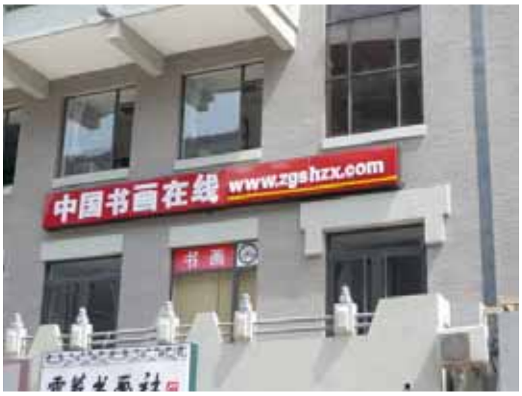
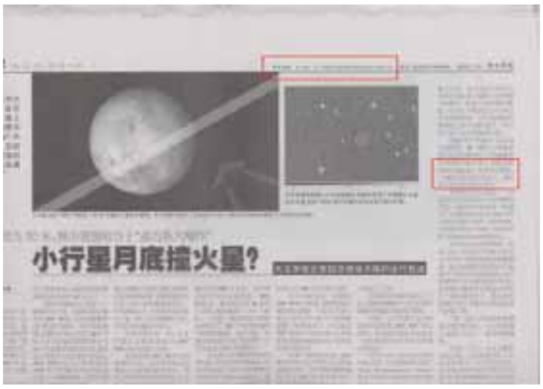


Figure 1. Street in north Beijing

The advertisement reads for 'Chinese Painting and Calligraphy Online'. The Roman transcription 'zgshzx' stands quite obviously for '中 *Zhong* – 国 *Guo* – 书 *Shu* – 画 *Hua* – 在 *Zai* – 线 *Xian'*, i.e. the first Roman-alphabet letter of each Chinese character's phonetic transcription.

The second illustration is taken from the 10 of January issue of the *Nanfang Zhoumo* 南方周末 (a famous newspaper in China). On page 26, in the *Science News* section, we find an article about whether Mars may be hit by an asteroid at the end of the month. The whole article is written in Chinese, using –of course– Chinese characters.



小行星月底撞火星?

Figure 2. Page 26 from 10 January issue of the Nanfang Zhoumo (« Southern Weekly »)

But the Chinese characters are not alone on this page. We can see Arabic numerals here and there (commonly used in modern Chinese for dates, and phone numbers), and also foreign specialists' Western names, which are indicated in Roman letters right after their transcription in Chinese characters. The enlargement below corresponds to the square on the right side of the precedent illustration.

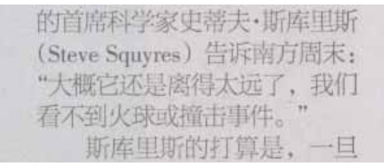

的首席科学家史蒂夫·斯库里斯
（Steve Squyres）告诉南方周末：
"大概它还是离得太远了，我们
看不到火球或撞击事件。"
斯库里斯的打算是，一旦

Figure 3. Detail (a)

Lastly, in the upper part of the page, the name of the 'responsible editor' is indicated (负责编辑 *fuze bianji*, 朱力远 ZHU Liyuan), together with a contact e-mail address (second square in the upper left part of the picture).

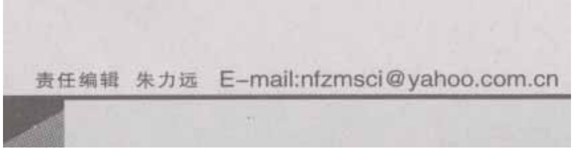

责任编辑 朱力远 E-mail:nfzmsci@yahoo.com.cn

Figure 4. Detail (b)

Here again, Roman-alphabet letters corresponding to the first one of each character's phonetic transcription are used: "nfzm" stands for 南 *Nan* - 方 *Fang*- 周 *Zhou* - 末 *Mo* (i.e. the newspaper's name, *Nanfang Zhoumo*); "sci" seems to be the English abbreviation for the *Science News* section.

Simply put, DNS technology's cultural impact can be observed on the newspaper's page, as it consists of the only 'non-Chinese character contents' of the page, other than Arabic numerals and foreign specialists' names.

4. Conclusion

Having raised two everyday-life examples showing the impact of DNS technology on Chinese culture, it is easy to get lost in thoughts on major language issues, globalization, or technological determinism, –to name but a few of the most debated ideas of the past decades.

The aim of this paper, as stated earlier, is not to suggest a explanation or an analysis of what is happening in the real or virtual worlds. Neither does it pretend to make predictions on what will happen to Chinese culture in a near or distant future.

My point here is to share the sketch of a design method which, I believe, may help both producers and users of new technologies *to see* what is really going on behind the screen(s). Since new technological advances are often quickly rendered invisible, this method may prove useful in helping us to see it anew, and imagine how things could have been, or could be different. How would DNS technology look like if it had been invented by the Chinese? How will it look like the day it will be re-invented by the Chinese, so that Chinese characters are not a problem anymore, but a very convenient way to access the Internet of things, as the ASCII technology is nowadays?

As I have tried to suggest, cultural issues in technology are, most of the time, bound by three variables: *use*, *modifiability*, and *differences between 'contents' and 'context'* [7]. The more the technology is used, the less it is modifiable, and the bigger the differences are, the bigger will be the 'cultural issue'. Inversely, in the case of an unused technology, or a modifiable technology, or an absence of difference between contents and context, there won't be any issue.

5. References and Notes

[1] Zimmermann, Basile "Tracing the Action of Technical Objects in an Ethnography: Vinyls in Beijing." *Qualitative Sociology Review,* Vol. III Issue 3 (2007): 22-45 <http://www.qualitativesociologyreview.org/ENG/archive_eng.php>

------. *De l'impact de la technologie occidentale sur la culture chinoise: les pratiques des musiciens électroniques à Pékin comme terrain d'observation de la relation entre objets techniques et création artistique*. Ph.D. dissertation, Unit of Chinese Studies, University of Geneva, Switzerland, Cyberthèses (2006). <http://www.unige.ch/cyberdocuments/theses2006/ZimmermannB/meta.html>

------. "Technology is Culture: Two Paradigms." *Leonardo Music Journal* 15, Cambridge Mass.: MIT Press (2005): 53-57.

[2] Kranzberg, Melvin. "The Information Age : Evolution or Revolution ?" *Information Technologies and Social Transformation* (1985): 35-54, p. 50.

[3] "The De-Scription of Technical Objects." In *Shaping Technology - Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker, and John Law, 205-24. Cambridge Mass.: MIT Press, 1992. p. 208

(Original French version) Akrich, Madeleine. "Comment décrire les objets techniques?" *Techniques et culture* 9 (1987): 49-63.

[4] Latour, Bruno. "Mixing Humans and NonHumans Together: The Sociology of a Door-Closer." *Social Problems* 35-3 (1988): 298-310, p. 299. For a more recent publication, see Latour, Bruno. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press, USA, 2005.

[5] Tai, Zixue. *The Internet in China: Cyberspace and Civil Society*. New York: Routledge, 2006, pp. 119-159.

[6] <http://www.cnnic.cn/index/0E/00/11/index.htm> Consulted February 2008, see the PDF report in Chinese on p. 10. At the time of writing, the English translation is not available yet.

[7] Of course, one could debate for years about the exact meaning or definition of these terms. However, as stated at the beginning of the article, this question is of no interest, since what matters is the *relation between* things, ideas, people –whatever is involved. See Latour (2005) for a broader theoretical framework [4].

6. Acknowledgments

Nicolas Zufferey and Howard S. Becker for invaluable help and friendship; Qiu Zeqi for constructive discussions and criticism at various stages of the initial research project; Deng Suo, for useful suggestions while discussing the theoretical framework; Christine Leuenberger, Mareile Flitsch and Antoine Hennion, for crucial advice and kindness at difficult times; Muriel Bowie, for the research assistance; Andrej Seàn O'Murchù for helping me improve my English; Thomas Zimmermann and one anonymous reviewer, for useful comments on this paper. The financial support of the Swiss National Science Foundation under grant PBGE 1101317 is gratefully acknowledged.

First International Conference on The Internet of Things

IOT 2008 Demos

Analyzing Product Flows with the Supply Chain Visualizer

A. Ilic[1], T. Andersen[1], F. Michahelles[1], E. Fleisch[1,2]

[1]ETH Zurich, Information Management, 8092 Zurich, Switzerland
{ailic,tandersen,fmichahelles,efleisch}@ethz.ch

[2]University of St.Gallen, Institute of Technology Management, 9000 St.Gallen, Switzerland

Abstract. Globalization and technical advances have created a new level of competition and complexity for supply chains. The risk is high that due to this complexity and limited visibility causes of inefficiencies remain undiscovered. Radio Frequency IDentification (RFID) technology has recognized potential to increase the visibility along a supply chain due to cost-efficient gathering of process event data. With the Supply Chain Visualizer, we provide a tool for analyzing this event data and show how simple rules can reveal inefficiencies. By using a map-based user interface, a supply chain manager is finally able to pinpoint the sources of inefficiencies.

Keywords: RFID, trace data analysis, rule engine, visualization

1 Introduction

Due to globalization and technical advances, it is not the single firm anymore that decides the competition, but the interplay between several organizations along the supply chain. With increased complexity, supply chains face the high risk that due to limited visibility inefficiencies remain undiscovered. While managers know about the impact of delays, inaccurate data or shrinkage, they rarely know why and where the problems in the supply chain really originate. Technologies such as Radio Frequency IDentification (RFID) provide the potential to increase a supply chain managers' visibility in a high-resolution manner [1]. RFID technology can be used to generate trace data events of business processes and thus link the physical world to information systems. As the whole supply chain is now represented digitally, the data can be used to detect and locate inefficiencies in the actual supply chain.

The goal of the Supply Chain Visualizer project is to provide an easy to use visibility tool to support supply chain managers. The tool analyzes the integrity of RFID data events to locate "hot spots" in a supply chain. This information could be relevant for supply chain managers as the hot spots are a good indicator for further investigations and hereby help to direct quality improvement efforts accordingly. We provide a bottom-up supply chain integrity analysis that is based on generic consistency rules. With the recently standardized EPCIS interface [2], generic rules for detecting inefficiencies can be built and applied to a broad context. To our best knowledge, no paper exists studying the concept of generic rules for trace data analysis to detect inefficiencies in supply chains. In the following, we describe the system design and outline the high-level architecture of the Supply Chain Visualizer. Afterwards, we propose our concept of generic trace data rules to locate problems in supply chains and show some details about our prototype. Finally, we conclude our paper and suggest future work.

2 System

We consider a supply chain where every product is equipped with an RFID tag. Every time a RFID tag is read, an event record is generated. As the Supply Chain Visualizer is focused on the inter-organizational supply chain aspects, we consider three different types of events. We use a so-called shipping/receiving supply chain model [3], and differentiate between 1) shipping, 2) receiving and 3) internal events. An example for this model can be seen on Fig 1.

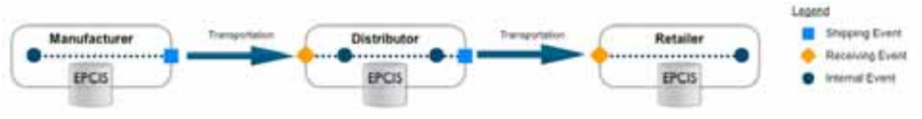


Fig 1. Illustration of the shipping/receiving data model for a generic retail supply chain

As specified in the EPCglobal architectural framework [4], the component for sharing the event data with other supply chain members are the so-called EPC Information Services (EPCIS). The Supply Chain Visualizer uses the EPCIS interfaces to collect data from various stages of a supply chain and stores them in a local data warehouse. The advantage is that several levels of analysis can be run without interfering with operational processes.

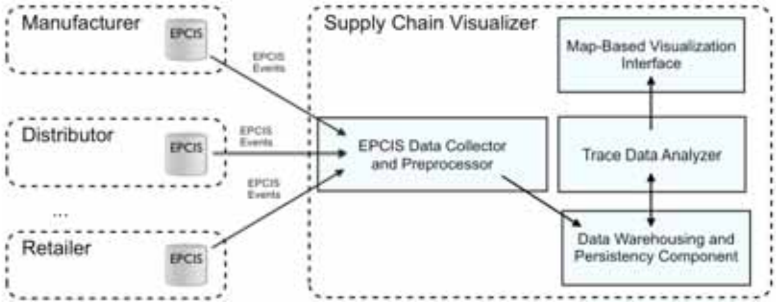


Fig 2. High-level architecture of the Supply Chain Visualizer and data flows between the components

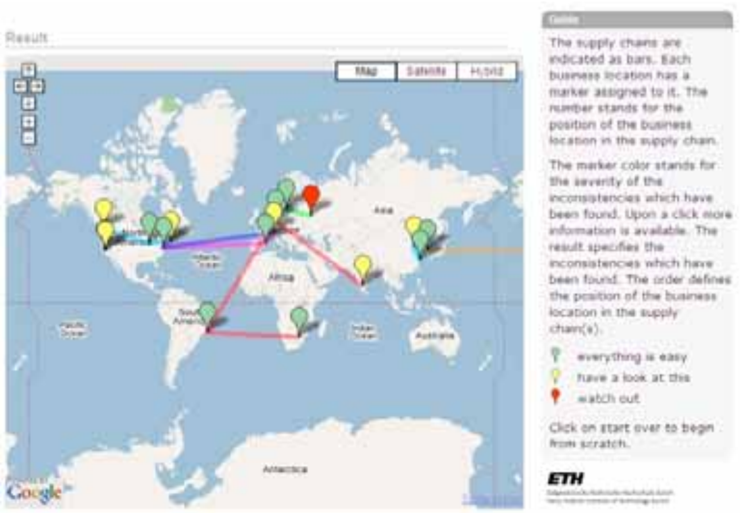
The high-level architecture of the Supply Chain Visualizer is depicted on Fig 2 and comprises four main components. The boxes represent the components of the Supply Chain Visualizer and the arrows indicate data flows. The EPCIS data collector component uses the open source Accada Framework [5] to connect to a set of preconfigured EPCIS systems. The component retrieves the event data and preprocesses them for storage in the data-warehousing component. The data-warehousing component manages persistency and access to the backend MySQL database, which contains both, the imported data sets and the results from the trace data analyzer component. On the presentation layer, a map-based visualization interface displays the results and interacts with the user.

3 Rule-based trace data analysis

From the EPCIS specification [2], we make use of the following event data fields for our analysis: *epcList*, *eventTime*, *action*, *bizStep* and *bizLocation*. In the current version of the Supply Chain Visualizer we implemented four generic rules that provide a good indication for problems in the supply chain. The *speed consistency* rule utilizes utilizes *eventTime* and *bizLocation* to calculate the transition speed from one shipping event to another receiving event. If the speed exceeds a certain threshold, an inconsistency alert is raised. This alert could for example be triggered if products are subject to counterfeiting and thus two traces of a product with the same serial number occur. The *dwell time* rule calculates the time between receiving and shipping event. If the difference exceeds a certain threshold, time-critical products (such as perishables) might not be handled accordingly by a supply chain entity. The *shipping/receiving pairconsistency* rule utilizes the *bizStep* data field and might be very useful in various industries. Every item that is shipped from one location must have an appropriate read event. Otherwise, it could be an indication that a product was actually not shipped or was stolen during transport. The fourth rule that is implemented concerns the lifecycle of a product. As the field *action* indicates the initialization (e.g. manufacturing) or inactivation (e.g. product sold) of a product's tag, the *lifecycle* rule checks that there are no events registered before an initialization event ("ADD") or after an inactivation event ("DELETE").

4 Our demo

The Supply Chain Visualizer is a web-based application. A standard computer with a web-browser and Internet connection is sufficient. No specific client software is needed. We will use data from a fictive retail supply chain (about 10000 traces) and put the visitor into the role of a supply chain manager. A screenshot of our demo can be seen on Fig 3. The visitors will be able to see that product flows are highly aggregated to single lines and thus provide a clear view on the whole supply chain. With an easy to understand three-color scheme, the visitors can now see where the problems are and track them down. Green pins hereby indicate that the integrity analysis did not find any problems at a location. Yellow means that there are some problems detected which deserve attention and red indicates that the number of problems exceed a critical threshold. By clicking on the pins, the visitor will get more details about the type of integrity violation and find out which products are involved.



The supply chains are indicated as bars. Each business location has a marker assigned to it. The number stands for the position of the business location in the supply chain.

The marker color stands for the severity of the inconsistencies which have been found. Upon a click more information is available. The result specifies the inconsistencies which have been found. The order defines the position of the business location in the supply chain(s).

🟩 everything is easy
🟨 have a look at this
🟥 watch out

Click on start over to begin from scratch.

Fig 3. Screenshot of the Supply Chain Visualizer demo

5 Conclusions

We showed how simple generic rules can increase a supply chain managers visibility of problems by analyzing RFID trace data. Our map-based interface provides an easy to use representation suitable also for larger data sets. Future research directions include the use of probabilistic rules and also the role of supply chain managers for configuring the rules to a specific problem.

Acknowledgements. This work was partly funded by the European Commission within the Sixth Framework Programme (2002-2006) IP Nr. IST-FP6-033546.

References

1. Lee, H. L., & Ozer, O. (2005). "Unlocking the value of RFID". Graduate School of Business, Stanford University, working paper.
2. EPCglobal (2007). EPCIS Specification. http://www.epcglobalinc.org/standards/epcis
3. McFarlane, D. & Sheffi, Y. (2003). "The Impact of Automatic Identification on Supply Chain Operations". International Journal of Logistics Management, 14(1).
4. Traub, K., Allgair, G., Barthel, H., Burstein, L., Garret, J., Hogan, B. et al. (2005). "The EPCglobal Architecture Framework". EPCglobal Final Version http://www.epcglobalinc.org/standards/architecture
5. Floerkemeier, C., Roduner, C., & Lampe, M. (2007). "RFID Application Development with the Accada Middleware Platform". IEEE Systems Journal, 1(2).

A Web based platform for smart spaces

Jilles van Gurp, Sasu Tarkoma, Christian Prehofer, and Cristiano di Flora

Nokia Research Center, Helsinki, Finland
http://research.nokia.com

Abstract. This demo presents our work on middleware for ubiquitous applications in — what we call – smart spaces. Our goal is to show that web technologies can be extended to interact in smart spaces. The middleware reuses various components from the open source world such as web servers; various python frameworks and other components. Key aspects of our platform such as service discovery; use of indoor positioning; use of browser based UI and other features are illustrated using a set of usecases we have developed around the theme of a shopping mall, which we regard as a useful example of a public space where it would be desirable to provision a set of services and applications that can be accessed by users in the smart space in a similar fashion as one would provision a public web site today. Our demo includes examples of both horizontal features that could be of use in other public spaces and vertical features that are more specific to the mall example.

Key words: Smart Places, web services, REST

This demo presents our work on middleware for ubiquitous applications in — what we call – smart spaces. A smart space is a multi-user, multi-device, dynamic interaction environment that enhances a physical space by virtual services [1]. These services enable the participants to interact with each other and other objects in a P2P way in the smart space. The research in the area of ubiquitous and pervasive computing has led to many interesting research demos and usage experiences. Our goal is to show that web technologies can be extended to interact in smart spaces. Building on widely spread wireless devices such as phones, PDAs and other special purpose devices, there is an enormous potential to create new smart space services and applications.

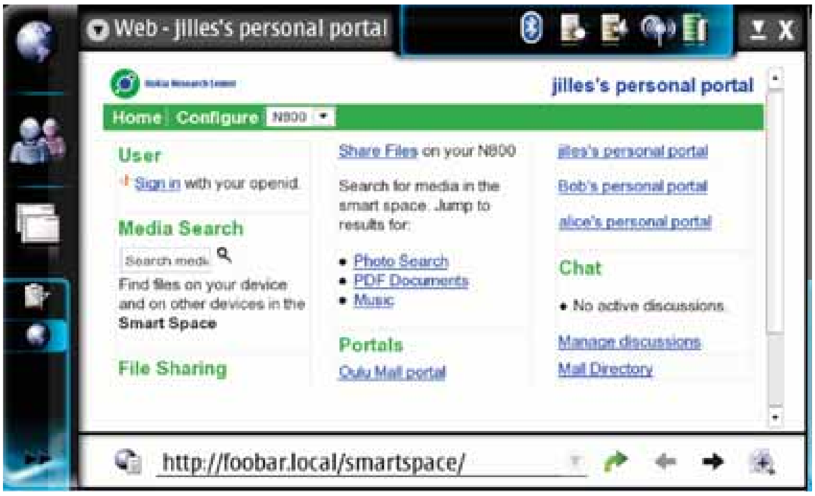
We build our work on several current trends in mobile devices, namely: mobile internet adoption is increasing; more and more phones and other user devices are becoming connected; and the internet is emerging as the glue through which these devices connect. Leveraging this trend of internet in the mobile allows us to avoid some of the pitfalls that have hindered adoption of similar technology in the past. For example, by being web based, our platform can integrate across device, vendor and platform boundaries and be accessible to users with any browser equipped device.

As partially outlined in previous publications [2, 3], our platform comprises the following major components:

- **Service Discovery** In order to allow user devices to discover the smart space, a Zeroconf MDNS based service discovery mechanism is used. Services and applications in our platform are advertised using this mechanism and can find each other when needed.
- **Web Runtime(s)** Since our platform is web based, all applications and services are hosted in a application server. Currently our demo setup consists of a mix of mobile devices with a python based web platform and Java OSGI services as well as regular servers running the same and some additional platforms. A benefit of being web based is that a lot of these components are off the shelf components such as for example blogs and feed aggregators.
- **Smart Space Services** Some important services in our platform include the service discovery that other services and applications integrate through a REST based API; smart space search and indexing; indoor positioning that allows user devices to determine where they are in the smart space; and associated services that allow associating maps; points of interest; etc with these positions. A key feature in our platform is that all these services are accessible through simple web based APIs. Where possible existing APIs are reused rather than reinvented. For example, our platform integrates blog technology using APIs that are commonly used in Blog software such as RSS, Meta weblog Post and Ping APIs, etc.
- **Web Based Security** Users, data, and services make use of common internet security mechanisms such as Openid and Oauth to protect privacy and achieve access control. This security solution will be further discussed in a forthcoming publication.
- **In device Portal** To access the smart space, users use their mobile web browser to either access a in device web portal on their own device or access a centralized portal in the mall. The portal integrates various portal applications that cover aspects of our Mall related usecases such as finding friends in the mall or finding shops near you in the mall as well as more user centric services such as sharing files on the in device portal and searching for files across all devices in the smart space.

The main services in our platform include service discovery, searching and indexing, and indoor positioning. These services are provided as a set of REST APIs, which supports easy integration with 3rd party software. The indoor positioning service allows user devices to determine where they are located in a smart space, and to correlate the location with different kinds of maps and points of interests.

The demo we present shows our platform and web portal running on Nokia N800 internet tablet devices connected to a wireless lan. In Figure 1 a screenshot is presented of the smart space portal running in the browser on an N800. The idea of this portal is that it integrates several smart space applications into one web page. The user can visit other portals in the smart space simply by clicking in the link or chat with users on other devices using our smart space chat application. Additionally, the user can use the smart space search feature to search for media such as photos and music. This search feature works by



jilles's personal portal

User

Sign in with your openid

Media Search

Search media

Find files on your device and on other devices in the Smart Space

File Sharing

Share Files on your N800

Search for media in the smart space. Jump to results for:

- Photo Search
- PDF Documents
- Music

Portals

Oulu Mall portal

jilles's personal portal

Bob's personal portal

alice's personal portal

Chat

- No active discussions

Manage discussions

Mall Directory

http://foobar.local/smartspace/

Fig. 1. Screenshot of the Smart Space portal running on a Nokia N800

discovering all registered search services in the smart space and collecting the results. The Oulu Mall portal that is listed in the portal list is a server with applications and services specific to the Oulu Mall public smart space. The idea is that when in this mall, users can interact with services such as a mall directory and a what's near feature that make use of the indoor position service available from their own device to provide an experience that is customized to the user and his/her context and location.

We demonstrate the key features of the smart space system by showing several interactions between multiple mobile devices and the mall server. These interactions pertain to finding interesting physical or logical items in the virtual neighbourhood, and communicating with people using blogs and feeds. An overview of the applications integrated into our portal and the components they build on is presented in Figure 2. A full discussion of the technical architecture is beyond the scope of this paper. However, it should be mentioned that the software integrates a large number of open source components and has been designed with portability of both software and concepts to other platforms in mind.

Building on top of many established SW components and technologies, the main novelties of the middleware include

- Indoor positioning, which includes a concept for symbolic or semantic locations.
- Distributed security built on existing technologies like openID and openAuth.
- Modeling a smart space by distribted, REST-based web-services.

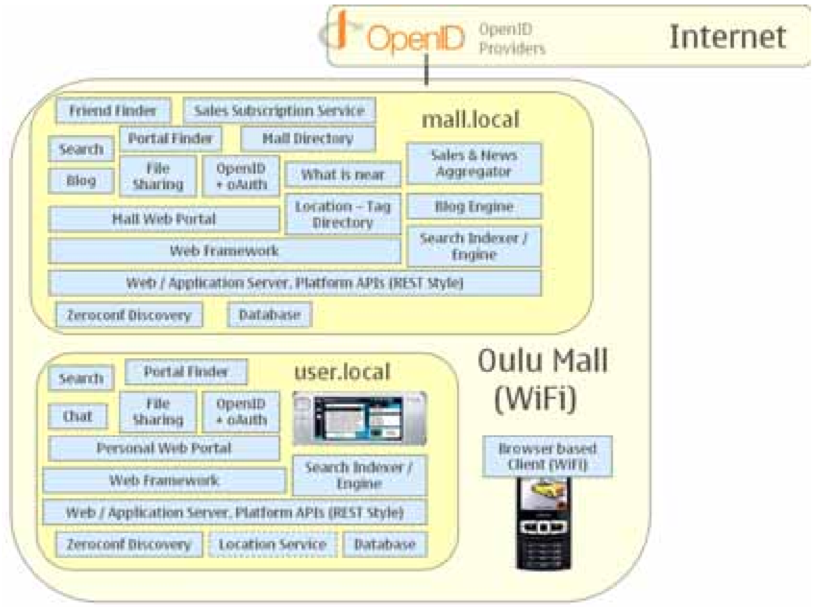


Fig. 2. Overview of usecases and components in the demo

Acknowledgements

The software has been written as part of our ongoing research in the Smart Space Lab in the Nokia Research Center. In addition to the authors, the following individuals have also been involved in developing concepts and software presented in this demo: Jaakko Kyro, Kari Ahvanainen, Pasi Liimatainen, Heikki Mattila, and Vlad Stirbu.

References

1. Wang, X., Dong, J.S., Chin, C.Y., Hettiarachchi, S.R., Zhang, D., Semantic Space: an infrastructure for smart spaces, IEEE Pervasive Computing, 3(3) pp. 32-39, 2004.
2. Jilles van Gurp, Christian Prehofer, Cristiano di Flora, Experiences with realizing Smart Space Web Service Applications, 1st Peer 2 Peer Mobile Handhelds workshop, at CCNC 2008.
3. Christian Prehofer, Jilles van Gurp, Cristiano di Flora, Towards the Web as a Platform for Ubiquitous Applications in Smart Spaces, Second Workshop on Requirements and Solutions for Pervasive Software Infrastructures (RSPSI), at UBICOMB 2007, Innsbruck, 16-19 Sebtember, 2007.

BaToo – Enabling the Rapid Prototyping of Mobile Services to Retail Products

Robert Adelmann

Institute for Pervasive Computing, ETH Zurich, Switzerland

Abstract. The mobile phone is set to change the way we shop in the future. Consumers will simply scan a product with their mobile phones and access personalized, product-related information and services while they are in the store. To facilitate the interaction with the physical products, consumer-oriented mobile applications require a convenient way to identify the product in the first place, requiring the automated recognition of products. Regarding retail products, a prominent way to do this is to use a mobile phone's built in camera to recognize the standard 1D barcode present on virtually every product world-wide. The creation of such mobile-phone based applications that provide services and information to real-world objects is currently very attractive: For prototypes, technology demos, or user studies. But even though many components of applications and prototypes are recurring, the creation of even simple applications requires a lot of time as well as know-how – both limiting the progress and development of new mobile-phone based applications and ideas. This document outlines our contribution to this problem: A rapid prototyping platform that includes the robust recognition of 1D barcodes and allows the creation of easy to use mobile-phone based applications to retail products: Within minutes, without any know-how about mobile phone programming.

1 Introduction

Today's consumer goods packaging lists a significant amount of product-related information. This includes nutritional information, ingredients, and possibly handling or recycling instructions. Some product packaging also comprise promotions with links to free song downloads or competitions. Due to the limited amount of space available on the product packaging and its static nature, the information cannot be customized for each consumer. Visually impaired people might prefer seeing allergy-related information in large print and non-natives might like to see the information in a different language. There is also a wealth of additional product-related information available that is not directly printed on the product packaging at all due to size constraints and possibly commercial considerations, e.g. reviews by consumer watch groups or price comparisons. Mobile phones have the potential to address many of these issues since they comprise display, long-range communication capabilities, processing, and user profile storage capabilities [2]. Since many of these potential applications are

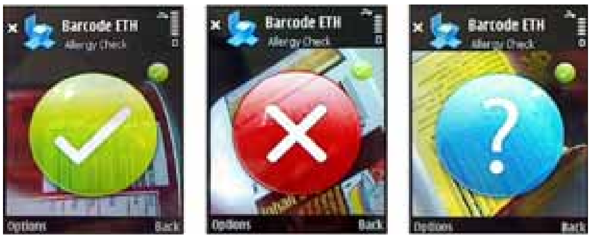


Fig. 1. Screen-shots of the Allergy-Check application

especially useful when being ”on the go”, e.g., while shopping, a simple and fast user interaction is essential, requiring the automated recognition of objects. Even though RFID technology is very promising, the widespread use of RFID tags on retail products remains unlikely for the next years. In contrast, barcodes are ubiquitous - printed on virtually all consumer items world-wide.

On one hand there is an abundance of potentially highly useful applications for both consumers as well as companies, on the other hand implementing applications ideas or prototypes is very difficult. This is due to the required know-how for the optical code recognition as well as the often necessary time consuming and difficult C++ Symbian programming on devices. In this document we are presenting both the outline of a rapid prototyping platform that eases and accelerates the development of according mobile phone applications, as well as the underlying powerful recognition of 1D barcodes on standard mobile phones, using the built-in camera.

2 Applications

Figure 1 shows screenshots of a typical application to retail products: the Allergy-Check application. It is based on the recognition of 1D barcodes. Once the user defined a profile containing all substances he or she is allergic to, holding the mobile phone in front of a product’s barcode gives the user a simple answer to the question ”Is that product fine for me?”.

3 Rapid Prototyping Platform

Like already mentioned, developing mobile phone applications that require the recognition of real-world objects is, due to several reasons, often a very time consuming and tedious process. The goal of our rapid prototyping platform is to foster the creation of novel and innovative applications by enable non-professional programmers to create mobile phone based services to real-world objects, specifically retail products. Our emphasis is hereby not on replacing, but on complementing existing general programming environment for mobile phones, such as J2ME or C++ Symbian with features that enable a very easy and rapid creation

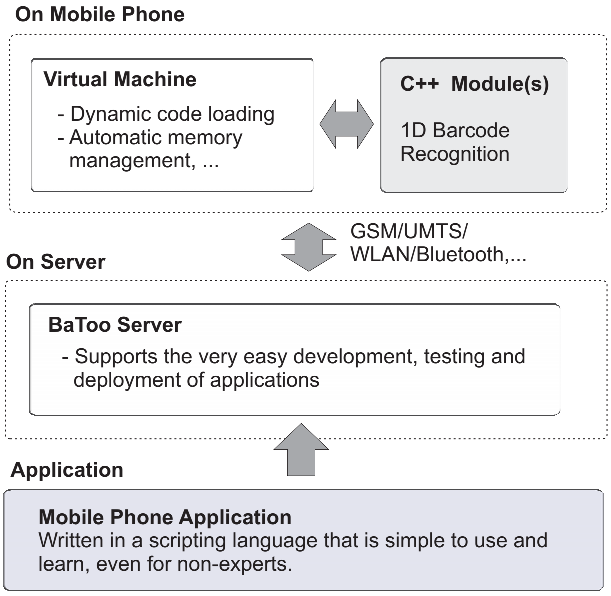

On Mobile Phone
Virtual Machine
- Dynamic code loading
- Automatic memory management, ...
C++ Module(s)
1D Barcode Recognition
GSM/UMTS/ WLAN/Bluetooth,...
On Server
BaToo Server
- Supports the very easy development, testing and deployment of applications
Application
Mobile Phone Application
Written in a scripting language that is simple to use and learn, even for non-experts.

Fig. 2. General architecture of the prototyping platform

of prototypes, without requiring any knowledge of mobile-phone programming or the involved technologies.

The general architecture of the platform is shown in Figure 2: The mobile phone contains predefined C++ Symbian components (libraries) for the major tasks that need the full processing power and capabilities available on the phone, mainly the 1D barcode recognition. Given these components, experienced C++ Symbian developers can already create functioning applications, even without special knowledge – for example regarding image recognition. But implementing even simple C++ Symbian applications still takes a considerable amount of time and won't be feasible for inexperienced developers – due to many Symbian specific concepts like ActiveObjects or Descriptors, complex memory management or the lack of documentation. J2ME is more accessible, but often lacks the required APIs and speed. In order to allow the simple and fast creation of applications, the BaToo system allows the user to write applications in a very easy to learn and use scripting language that is then executed on a virtual machine on the mobile phone, which has access to the provided C++ modules.

An additional BaToo-server-application running on a remote desktop computer handles the remaining complexity involved in developing applications: For example testing the application on the mobile phone, or creating distributable applications for today's Symbian based mobile phones, including the packaging and mandatory signing of applications. This way, developing a new service to

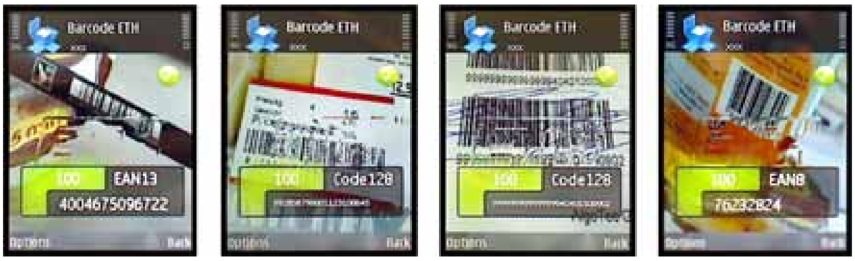


Fig. 3. Screen-shots of the barcode recognition component

retail products is limited to writing the application's code in a simple scripting language and pressing a few buttons to test and deploy the application on mobile phones.

4 Barcode Recognition

The component responsible for the recognition of barcodes has been implemented in Symbian C++ and features certain distinct advantages compared to existing commercial solutions for the recognition of 1D barcodes as well as proposed recognition algorithms like [1]: The recognition is robust, even under realistic conditions and is performed on the real-time video images of the mobile phone, requiring no key-presses by the user. This way the recognition is very fast and simple. Figure 3 shows some examples. Since no hardware specific features like DSPs (digital signal processors) were used, the software is in general working on all standard C++ Symbian devices.

5 Conclusion

We presented the outline of a toolkit that has the potential to ease as well as accelerate the creation of mobile phone based applications and prototypes that are concerned with the recognition of real-world objects. In combination with the included component for the recognition of 1D barcodes on mobile phones, this allows for a simple creation of services to retail products and the fast implementation and testing of novel application ideas.

References

1. E. Ohbuchi, H. Hanaizumi, and L.A. Hock. Barcode readers using the camera device in mobile phones. *Cyberworlds, 2004 International Conference on*, pages 260–265, 18-20 Nov. 2004.
2. Frank Siegemund, Christian Floerkemeier, and Harald Vogt. The value of handhelds in smart environments. *Personal Ubiquitous Comput.*, 9(2):69–80, 2005.

Coupling ERP systems with shop-floor web service enabled devices

Oliver Baecker[2], Dominique Guinard[2], Stamatis Karnouskos[1], Moritz Koehler[2],
Domnic Savio[1], Luciana Moreira Sá de Souza[1], Patrik Spiess[1] and Vlad Trifa[2]

[1] SAP Research, Vincenz-Priessnitz-Strasse 1, D-76131, Karlsruhe, Germany
[2] SAP Research Switzerland, Kreuzplatz 20, Zürich CH-8008, Switzerland
{oliver.baecker, dominique.guinard, stamatis.karnouskos, mo.koehler, domnic.savio,
luciana.moreira.sa.de.souza, patrik.spiess, mihai.vlad.trifa}@sap.com

Abstract. We are moving towards the Internet of Things where millions of networked such devices are interconnected, and provide their functionality as a service and seamlessly integrate in modern enterprise environments. The demo presented shows how web service-enabled devices and their services can be dynamically discovered, and integrated in business applications.

Keywords: Devices profile for web services (DPWS), ERP, SunSPOT, SAP xMII, Enterprise Applications, Web Services

1 Demonstration

Future manufacturing environments are expected to be populated with heterogeneous networked devices that can offer their functionality as a service and also consume other services, creating effectively mash-up factories. Today most of them hold proprietary communication interfaces, however with the increasing computing capabilities that will be available, their (even proprietary) functionality can be wrapped and integrated in modern service oriented environments. This will offer a tighter cooperation between shop-floor and enterprise layer, eventually drastically limiting integration gaps and error-proneness of shop-floor data due to media breaks. As we demonstrate the increased information granularity allows for more flexible and accurate enterprise services, increasing proactiveness and performance of the enterprise.

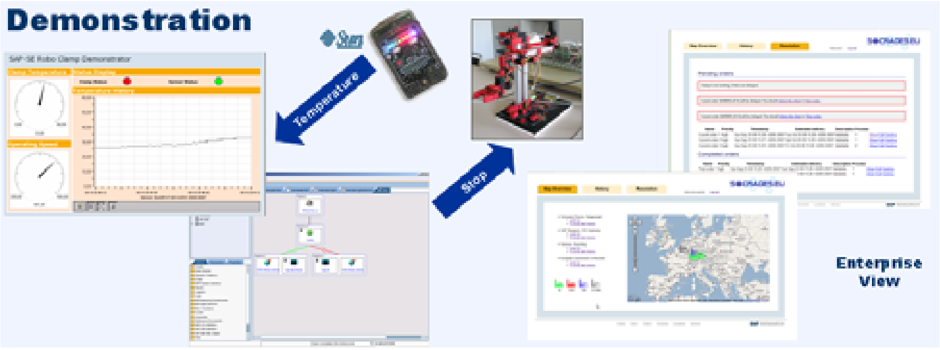


Figure 1 - Enterprise View of the shop-floor

Our prototype demonstrates the DPWS-based integration of shop-floor devices amongst each other and with enterprise systems. The combination of independent device-level DPWS services enables the composition of higher-level services and offers this functionality to top-floor applications (depicted in Figure 1). Our prototype consists of two DPWS devices on the shop-floor, which are integrated via the SAP xMII solution to SAP enterprise systems that allow controlling the devices from the top-floor. These two devices are:

- A DPWS-enabled robotic clamp which offers the services e.g. getInfo, start, stop, failure etc
- A SunSPOT wireless sensor node that is attached to the robot clamp and senses the environmental conditions of the clamp like the current temperature.
-

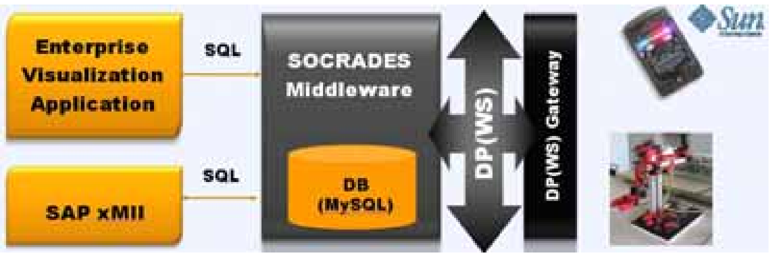


Figure 2 – Integration of Enterprise, Middleware and Device layers

As depicted in Figure 2, the described atomic services offered by the two devices are combined in a manufacturing process running on the middleware layer, which in turn interfaces with xMII that enables the modeling of business rules at run time. The middleware layer offers a mid-level business view on the services provided by shop-floor devices. Within xMII the application logic is modeled as business rules and corresponding service invocations to enterprise systems are triggered. Finally the manager can be informed via a dynamic web interface integrated with GoogleMaps about the status of all factories and their potential problems. In parallel, via the connection to the ERP system, combined data showing the effect of errors in the shop floor on the customer orders is depicted in real-time.

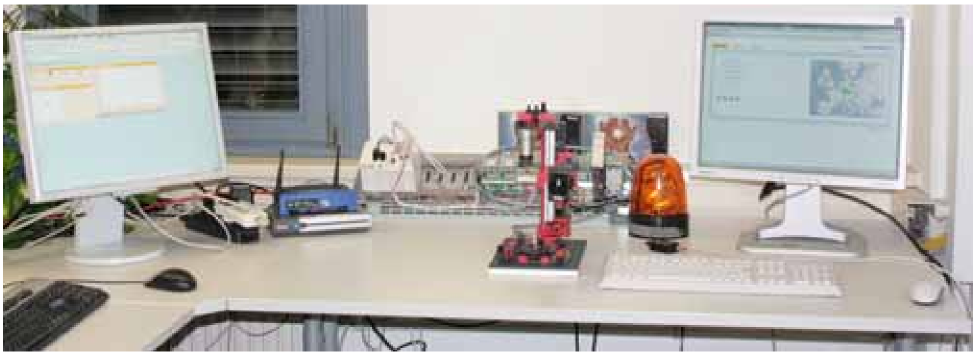


Figure 3 – The demonstrator testbed

Figure 3 depicts our demonstrator testbed. One can clearly see the two visual interfaces in the two monitors i.e. xMII on the left and dynamic web based GUI on the right. The robotic clamp, the alarm, as well as the Programmable Logic Controler (PLC) are connected via IP over Ethernet to eachother. The SunSPOT sensor is connected wirelessly via IEEE 802.15.4 with a base station attached to the USB port of the computer.

The main goal of our demonstration is to show:

- DPWS-based integration
- High level composite services
- Enterprise control via web services
- Business process monitoring
- Cross-layer alerts
- Enterprise visualization
- Automatic workflow for alert resolution
- Timely information dissemination and visibility
- Better customer relationship management

This demo has been implemented as part of the ongoing work within the European Commission IST FP6 project SOCRADES (www.socrades.eu). Further info on the architecture, motivation and rationale behind this demo can be found in [1].

References

1. Luciana Moreira Sa de Souza, Patrik Spiess, Moritz Koehler, Dominique Guinard, Stamatis Karnouskos, and Domnic Savio, "SOCRADES: A Web Service based Shop Floor Integration Infrastructure" , Internet of Things 2008 Conference, March 26-28, 2008, Zurich, Switzerland.

Demonstration of a decentralised material flow control of a large-scale baggage handling system

Moritz Roidl[1], Guido Follert[1], Lars Nagel[2]

[1] TU Dortmund, Faculty of Mechanical Engineering
Chair for Materials Handling and Warehousing,
Emil-Figge-Straße 73, 44227 Dortmund, Germany
http://www.flw.mb.uni-dortmund.de/
moritz.roidl@tu-dortmund.de
guido.follert@tu-dortmund.de

[2] Fraunhofer Institute for Material Flow and Logistics,
Joseph-von-Fraunhofer-Str. 2–4, 44227 Dortmund, Germany
http://www.iml.fraunhofer.de
lars.nagel@iml.fraunhofer.de

Abstract. In the domain of intralogistics the term *Internet of Things* subsumes a broad range of decentralisation approaches which touch the complete information infrastructure. To show the practicability of these new ideas we present an approach for integrating the concept of multiagent systems within an existing discrete event simulation model. This enables the direct comparison between classical centralised material flow control and a decentralised, multiagent-based control. The use of automated analysis and code generation enables the reuse of large-scale, industrial models which represent a more realistic testing ground than common test-bed environments. The example demonstrates the simulation of a large-scale baggage handling system in which the routing is controlled by a multiagent system using an adapted version of *Dynamic Source Routing*. The agent and communication infrastructures are integrated into the model to facilitate the analysis of the interaction between agents.

1 Introduction

Intralogistics is a cutting-edge term that comprises all technical systems, services and the relating business involved in the in-house materials handling operations of industrial enterprises, wholesalers, retailers and government institutions. The processes of the intralogistics domain are basic requirements for managing the flows of goods along the entire supply chain (supply chain management) as they provide the reliable and predictable flow of physical goods in the nodes of a supply network.

The *Internet of Things (IoT)* in the domain of intralogistics means a broad decentralisation effort in the areas of data management and flow of control. In the past years, the development of *Radio Frequency Identification (RFID)* technology has opened up the possibility of eliminating the central data warehouse in logistical systems: the data about a physical object is held in a *RFID* tag and travels with it; the flow of materials is united with the flow of information. The current research concentrates on the decentralisation of decision-making, resulting in the elimination of the central material flow controller (cp. [BtH07]).

2 Demonstration of Simulation Prototype

The agent control for the large scale baggage- handling system developed in the current case uses approx. 2,000 agents. More than 12,000 conveying elements build up the complex conveying network, which follows numerous restrictions given by the buildings and processes of the overall airport functionality. It comprises more than 150 infeeds and 100 destinations for the transport of baggage. Inside the network approx. 1,200 diverts and merges connect between the system parts. In addition several stations for luggage inspection using diverse machinery and manual encoding stations for non-identifiable pieces of luggage are distributed throughout the network. The system load of the experimental scenario covers a time span of 6 hours in which more than 60,000 pieces of luggage are conveyed in the described system.

The assigned experimentation example requires multiple implementations of agents at the relevant decisive components of the conveying network. The efficient implementation of these agents in the existing simulation model therefore takes place with automated procedures of model analysis and the generation of code. For this task the simulation model is read into a tool. This simulation model itself contains the essential information about the conveyor system in the form of a schematic topology. The reading process transforms this schematic topology into a structured model description. In a next step this model description is transformed into a graph which serves as the basis of the agent implementation. For each vertex of the graph an agent is generated. Then the new functions and the agent logic are supplemented. At the end of the process the tool transfers the model back into the format of the simulator. For very large models additional transformation processes reduce the complexity of the graph (cp. figure 1).

3 Adaptation of Dynamic Source Routing

The agents responsible for the routing of the luggage items in the conveying network use an adapted version of the *Dynamic Source Routing (DSR)* (cp. [JM96]). This algorithm is taken from the application field of mobile communication. It floods the network of the agents, which are distributed over the conveying elements, with routing requests. Starting from the direct origin of a transportation inquiry the neighbours pass this inquiry on to their successors. This way all attainable locations of the conveying network can be reached.

The hallmarks of the algorithm are the ease of implementation in the network nodes, cycle freeness and maintenance of the current network state. Compared to the original approach of the algorithm in the available form it considers the expected lead times and the current load situation of the respective relations in order to enhance the decisions between alternative routes.

Furthermore, the algorithm is influenced by the need for luggage inspection and manual encoding requests. The requests for those special operations are stored on the

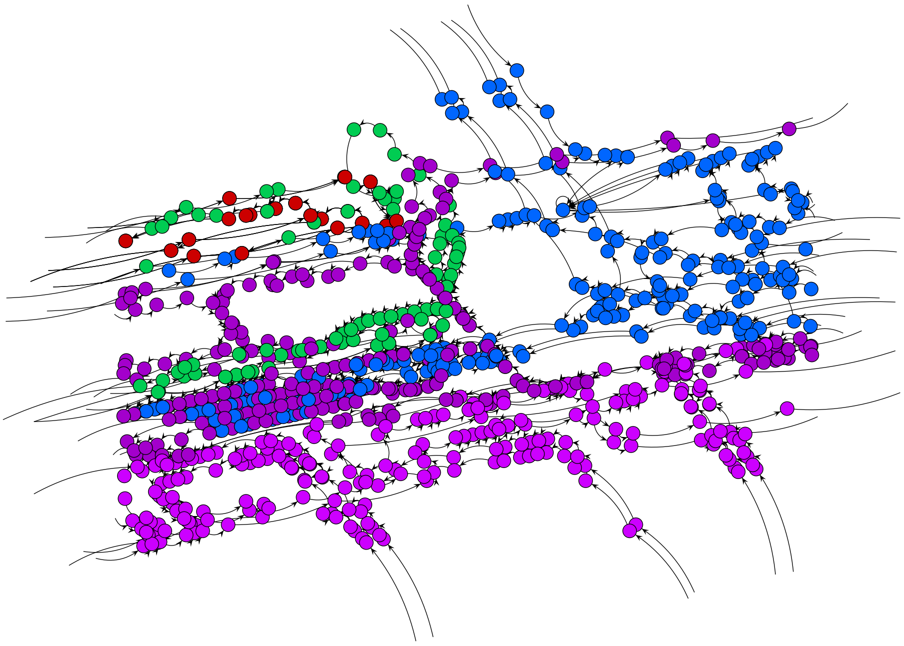


Fig. 1. Material flow graph of a small section of the demonstrated baggage handling system

identification tags of the luggage (inspection) or result from missing information on these tags. The necessary communication of the agents among themselves represents as expected an extraordinary challenge for the application of the decentralised control concept. Therefore the communication is explicitly shown in the simulation by the usage of function calls. Besides a time consumption is attributed to these function calls in order to consider the latency of the message transmission in a real communication network.

4 Analysis of Communication

Experimental analysis of the peak load of the message transmission shows that by the limitation of the routing algorithms in its flooding behaviour a significant reduction of the number of arising messages can be obtained. The flooding of the network is limited to only those subsystems that can contribute at all to the reaching of the conveying destination. In the examined scenario altogether $65 \cdot 10^6$ message transmissions take place during the 6 simulation hours of the described scenario. The model used here together with the technology for the implementation of software agents in an industrial logistics simulation possesses extensive potentials (cp. [RF07]).

5 Summary

Apart from the fundamental acceptance testing for the concept of the *IoT* and the investigations on the increasing communication intensity in large scale systems the ongoing work will focus on the increasing logistical performance of those systems and on innovative routing algorithms in general. The automated model adjustment and the comprehensive model provide a powerful and flexible investigation platform for the ongoing research. Relating to the concept of the *IoT* the conducted experiments emphasize the fact that the assigned routing agents, which cover less than 400 lines of program instructions and can be used universally at all decisive locations of the conveyor network, are able to control a large scale material flow system. Thus, the results deliver an important proof for the overall project of the *IoT*, where the methodical bases for further applications are presently compiled in different work. An outstanding conclusion from the work accomplished so far is the guarantee of the logistical function of a large scale agent-based conveying system with comparatively simple agents which is provided by the concept of the *IoT*.

References

[BtH07] BULLINGER, R. W. and M. TEN HOMPEL (editors): *Internet der Dinge*. Springer, Berlin, 2007.

[JM96] JOHNSON, D. and D. MALTZ: *Dynamic source routing in ad hoc wireless networks*. In KORTH, IMIELINSKI; (editor): *Mobile Computing*, volume 353, Kluwer Academic Publishers, Boston, 1996.

[RF07] ROIDL, M. and G. FOLLERT: *Simulation von multiagentenbasierten Materialflusssteuerungen*. In *INFORMATIK 2007 Informatik trifft Logistik, Beiträge der 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, volume 1, 2007.

Distributed Coordination in Mobile Wireless Sensor and Actuator Networks

Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu,
Hans Scholten, Paul Havinga

University of Twente, The Netherlands
s.bosch@student.utwente.nl,
{m.marinperianu, r.s.marinperianu, j.scholten, p.j.m.havinga}@utwente.nl

1 Introduction

The emergent market of wireless sensor and actuator networks (WSNs and WSANs) [2] targets a large number of applications, ranging from transport and logistics to industrial processes [3], and even planetary sensing or space exploration [4]. Even if limited in terms of computational power, memory, energy and communication bandwidth, the sensor nodes collaborate in the network, in order to accomplish complex monitoring tasks at scale.

In this work we explore the potential of WSN technology in dynamic applications requiring a *sense-and-react* control loop. More specifically, we consider the problem of distributed movement coordination of vehicles on wheels, equipped with wireless sensors and actuators. The final goal is to have a self-organizing group (or swarm) of nodes that maintain the formation by exchanging periodically their sensed movement information. Application domains include cooperative surveillance, mapping unknown areas [8], disaster control, automated highway [5].

For simplicity, we assume that there exist a *leader* whose trajectory has to be copied by the *followers*. Compared to the related work in the robotics community, we specifically utilize only low-cost, low-resolution inertial sensors, such as accelerometers and magnetic compasses.

2 Solution Overview

We showed in previous work that sensor nodes can become aware of being *together* by observing that they share a common context [7], through the use of three-axial accelerometers. However, for maintaining a certain trajectory in the field, the information provided by accelerometers is not enough. It has to be complemented by heading information with respect to a reference system, for example the Earth magnetic field. Therefore, the two parameters used by our system are:

- *Velocity* v, determined by integrating the acceleration samples a over the decision time interval $[t_1; t_2]$:

$$v(t_2) = v(t_1) + \int_{t_1}^{t_2} a(t)dt \tag{1}$$

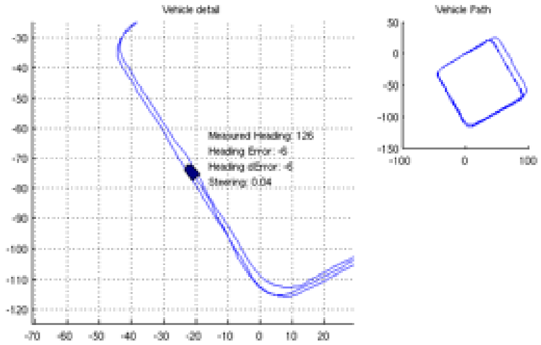


Fig. 1. Simulated behavior of the fuzzy logic controller running on the follower.

– *Heading* ϕ , computed from the magnetic field intensity H measured by the compass [1] :

$$\phi = arctan(H_y/H_x) \tag{2}$$

At each time step, the leader updates its speed and heading, and broadcasts this information over the radio. The follower compares the incoming data to its own measurements and tries to adapt to the trajectory of the leader. For this purpose, the follower runs a *fuzzy logic controller* with the following inputs and outputs:

– *Inputs*
  - The velocity error (leader velocity - follower velocity) VE and the variation of the velocity error ΔVE
  - The heading error (leader heading - follower heading) HE and the variation of the heading error ΔHE

– *Outputs*
  - The throttle, which controls the acceleration/deceleration.
  - The steering, which controls the orientation.

Figure 1 shows the simulated behavior of the fuzzy logic controller when trying to follow a rectangular trajectory.

3 Demonstration

For demonstration we use two remote controlled toy cars (see Figure 2) on which we mount two sensor nodes based on the Ambient μNode 2.0 platform [1]. The sensor nodes are equipped with the MSP430 microcontroller, three-axial digital

[1] For computing the X and Y magnetic field components, tilt compensation has to be applied. The tilt angles (roll and pitch) are determined from the static acceleration information.

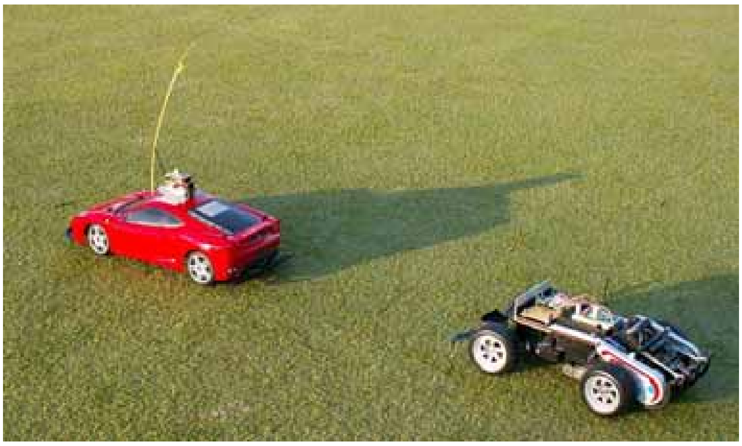


Fig. 2. The leader (left) and the follower (right) cars, equipped with sensor nodes and 3-D accelerometers and magnetic compasses.

accelerometer and magnetic compass on the SPI interface, 10kB of RAM and a radio transceiver with 100 kbps maximum data rate. The accelerometer is sampled at 160Hz, while the compass, due to constructive limitations, is sampled at only 16Hz. The communication is based on a simple scheduled access scheme imposed by the leader node. For logging, we use a third sensor node that acts as passive listener and forwards all incoming data packets to a PC via the serial port.

The leader car is driven using its original remote control. On the follower car, the original engine controller is modified to accept PWM throttle and steering commands from a second MSP430 microcontroller. This microcontroller and the one on the sensor node are connected through the I^{2}C interface. The sensor node on the follower receives the velocity and heading from the leader node, updates the inputs of the fuzzy controller as explained in Section 2, and finally sends the throttle and steering commands via the I^{2}C interface. The implementation of the fuzzy controller is based on the results reported in [6].

One of the major challenges of inertial measurement systems is the accumulation of measurement errors through acceleration integration. Without external reference, such as GPS, the error can grow unbounded in time. To prevent that, the follower node detects stationary periods, by analyzing the variance of the acceleration samples within a sliding time window, and resets its velocity to zero at those instances.

Figure 3 depicts data logged from a field test with the two cars. The movement sequence is as follows: drive backwards and turn, stop, drive forward and turn, stop, drive backwards and turn, stop. We see the velocity and heading computed by both the follower and leader, as well as the result of a simulated follower. In addition, the actual throttle and steering commands on the follower are also plotted. We observe that the follower copies quite closely the movement of the leader and also matches the result of the simulated follower.

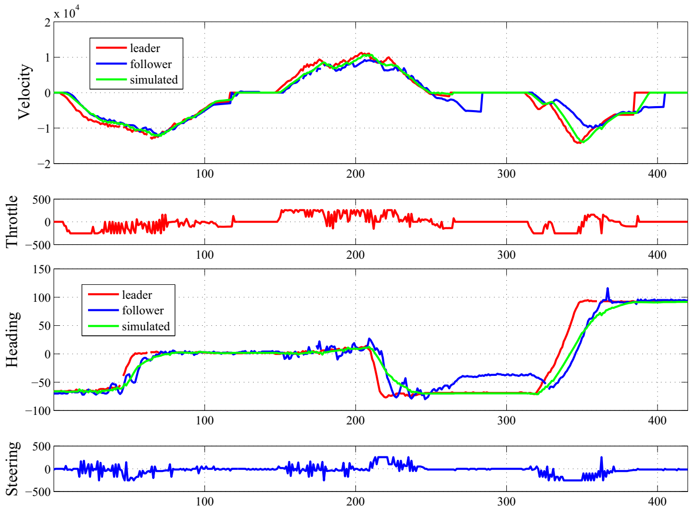


Fig. 3. Results of one field test.

4 Conclusions

We presented a solution for distributed movement coordination of wireless sensors and actuators, using simple inertial sensors and low-power radio communication. The entire inference and control are performed on resource-constrained sensor nodes running a lightweight fuzzy logic engine. For future work, we plan to evaluate the performance of our solution (in terms of distance and heading errors) in a larger scale setting.

References

1. Ambient Systems. http://www.ambient-systems.net.
2. I. Akyildiz and I. Kasimoglu. Wireless sensor and actor networks: Research challenges. *Ad Hoc Networks Journal*, 2(4):351–367, 2004.
3. M. Marin-Perianu et al. Decentralized enterprise systems: A multi-platform wireless sensor networks approach. *IEEE Wireless Communications*, 14(6):57–66, 2007.
4. E. Gaura and R. Newman. Wireless sensor networks: The quest for planetary field sensing. In *Local Computer Networks (LCN)*, 2006.
5. D. Hayat. Traffic regulation system for the future automated highway. *Systems, Man and Cybernetics*, 3, 2002.
6. M. Marin-Perianu and P. J. M.Havinga. D-FLER: A distributed fuzzy logic engine for rule-based wireless sensor networks. In *International Symposium on Ubiquitous Computing Systems (UCS)*, pages 86–101, 2007.
7. R. S. Marin-Perianu, M. Marin-Perianu, P. J. M. Havinga, and J. Scholten. Movement-based group awareness with wireless sensor networks. In *5th International Conference on Pervasive Computing (Pervasive)*, pages 298–315, 2007.
8. C. Stachniss F.E. Schneider W. Burgard, M. Moors. Coordinated multi-robot exploration. *IEEE Transactions on Robotics*, 21(3):376–386, 2005.

RFID Tag Pseudonyms with Efficient Reading and Scalable Management

C. Tutsch[1], A. Soppera[2], T. Burbridge[2], M. Aigner[1]

[1]TU Graz, Rechbauerstr. 12, 8010 Graz, Austria
christian.tutsch@student.tugraz.at, Manfred.Aigner@iaik.tugraz.at
[2]BT Research, Adastral Park, Martlehsam, Ipswich, UK
{andrea.2.soppera,trevor.burbridge}@bt.com

Abstract. In today's RFID system, a tag can be read without the knowledge of the owner. The interrogator can record the tag ID and correlate multiple occurrences of a certain tag to perform tracking through space and time. While this can be a powerful feature, unwanted tracking of tags can violate corporate confidentiality and customer privacy. We address these issues through a multi-layer RFID pseudonym protocol. We present an implementation of a secure pseudonym design that relies on standard symmetric cryptography to authenticate the tag. The protocol also provides an additional layer to deal with the fundamental problem key management. An authorised reader can determine the tag identity locally. Other readers may apply to a trusted authority. We also implement a strong notion of tag ownership and allow read delegation. This design has been fully implemented on ISO 18000 3M1 HF system.

Keywords: RFID, security, AES, pseudonyms, authentication, confidentiality, anti-counterfeiting

1. Introduction

Future systems will need to provide way for readers to authenticate RFID tags and prevent unwanted situations such as the surreptitious tracking of objects and people through time and space. Unfortunately current tags have a minimalist design with no security.

Lack of resources for basic cryptography poses in itself an intriguing research challenge [4]. Some very promising approaches in terms of computational complexity, communication overhead and ease of implementation can be subsumed as "pseudonym approaches" [2, 5-8]. These approaches are characterized by the fact that the tag identity is never transmitted. The tag responds with a different seemingly random pseudonym to every query. This token, or pseudonym, can be generated using a standard encryption scheme such as AES. These pseudonym schemes rely on a symmetric key shared between the reader and the tag but the fact that this secret key is also specific to the tag leads to a paradox. If a reader does not know which tag is interrogating it cannot know which key to use to authenticate the tag.

Our motivation is to develop a new pseudonym scheme that allows a fast lightweight routine read operation, yet is still scalable to large numbers of tags. We achieve this by dividing the tag functionality into separate reading and management functionality.

2. Demonstration

We show the use of our pseudonym scheme for the authentication of the tag and the encryption of the identifier. We first demonstrate that a reader without any knowledge of the tag secrets receives what appears to be a random identifier that changes with every read attempt. This occurs as a result of the tag encrypting the random nonces with a secret read key. We prove that once a reader is accepted by the community of the tag owner, and is delegated the appropriate read key; it can use its local key store to decode the pseudonym and obtain the identifier for the tag. In this manner, new participants can be introduced to participate in authorised supply chain operations. At any point the tag owner may change the read key, effectively revoking the access rights of the current participants.

In the next stage of the demonstration we show the concept of ownership. As we have explained, the owner of a tag has the rights to manage the access to the tag through changing and distributing the read key. As the legal owner of a supply chain product changes throughout the product lifecycle, the party who controls the access to the tag should also change. We show that anyone who currently possesses the tag may apply for ownership of the tag. This removes the requirement for the previous and new owner of the tag to negotiate the tag handover directly.

The applicant for ownership applies to a unique trusted authority (Trusted Centre) for the tag, presenting their credentials, along with an encrypted challenge received from the tag. This challenge is used by the Trusted Centre to uniquely identify the tag, and to compute the correct response that will allow the change of ownership to proceed. We show that a successful application will result in the tag changing its ownership key, which is also shared with the new owner. This ownership key is used by the new owner to immediately change the read key of the tag, removing the previous owner from any remaining influence over the tag.

We also discuss the problem of scalable key management. The set of current read keys possessed by a reader is likely to be limited to the number of tags owned by a small number of supply chain participants. This enables the singulation of the correct read key to be performed feasibly by a linear search. The read key is not used by the Trusted Centre since we desire that the read key may be changed locally by the current owner without problems of synchronisation. Instead the Trusted Centre shares a third category of secrets with the tag that are never released to other participants in the supply chain. We suggest that a tree of keys is used, following the work of Molnar and Wagner [7]. This enables scalable identification of the tag at the expense of a larger data exchange between the tag and reader.

This demonstration is supported by the use of several reader devices which participate in attempting to read the tag and participate in the transfer of ownership process with the Trusted Centre which is implemented as a web service. The tag itself is implemented using the HF DemoTag developed by the Technical University of Graz, using an AES encryption algorithm to implement the pseudonym protocol. The use of AES provides a standardised well-understand basis for the security of the rest of the system.

3. System

ISO 18000 3M1 tags [1] do not feature any security mechanisms. EPC C1 Gen2 tags [3] feature simple password control that is vulnerable to eavesdropping and brute force attacks. However, the standards allow a security layer to be built above the anti-collision protocol. In the pseudonym approach the tag ID or EPC in the existing protocols is replaced by the pseudonym.

All steps necessary to extend the standard compliant protocols can be done after the anti-collision sequence has been finished by using Custom Commands. Both standards reserve some command codes for optional, custom and proprietary commands. The HF system implemented consists of the components depicted in Fig.1:

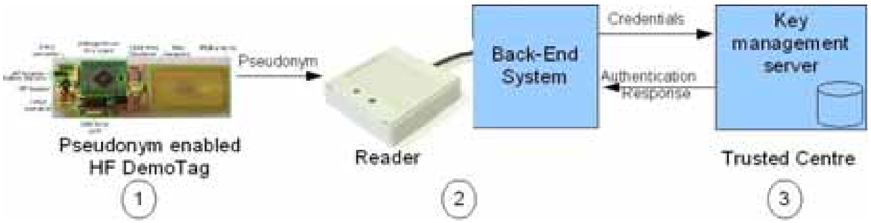


Fig. 1. Pseudonym tag communicating with reader and Trusted Centre

The pseudonym scheme implemented in the demonstrator is designed to fulfill a number of criteria. The scheme separates the read key from an ownership key to allow tag owners to control the access to the tag. The read operation is lightweight in terms of the data transmitted and the computation overhead. This enables the read operation to be completed with a minimum duration and number of communication cycles and low power requirements. However, for the Trusted Centre we desire an operation that is more scalable with a large number of tags, and does not require synchronisation whenever the read key is revoked.

To cope with these requirements, we adopted a scheme with separate read, ownership and Trusted Centre keys. These keys are used in the following operations:

1. Identification - The tag uses a short random number and a so-called read key kr, shared only between tag and reader, to produce a pseudonym short enough to replace the ISO 18000 tag identifier, for example the 64 bit ISO 18000 part 3 mode 1 ID or the ISO 18000 part 6 mode C EPC (64, 96 or 256 bits in future). The reader performs a linear search through its key space to identify the tag.
2. Take Ownership - A bit string based on the tree of secrets used by Molnar and Wagner is forwarded to the Trusted Centre. It includes a new secret ownership key kt protected by a secret kx that is derived from the tree of secrets, and a nonce nt.

auth = r | Fk0(r) | Fk01(r) | Fkx(kr) | Fkx(nt)

If the Trusted Centre allows the operation the new ownership key is transmitted to the reader and the nonce nt is used to commit the new ownership key on the tag.

3. Change Read Key - The tag will generate a new read key kr and nonce nr protected with the ownership key kt. The owner stores the new read key and commits the new key using nr.

auth = r | Fkt(r) | Fkt(kr) | Fkt(nr)

4. Conclusions

Our demonstration has presented a new pseudonym scheme that allows efficient reading operations while retaining the ability to operate with large volumes of tags. We have shown how the division of read and ownership functions within the tag can enable the simple delegation of the ability to read the tag and the ability to revoke access rights. The use of a Trusted Centre enables the transfer of tag ownership using a scalable tree of secrets. We believe that this scheme can enable business operation where secure tags are required that travel across multiple dynamic partners in supply chain operations.

Acknowledgements. This work is partly funded by the European Commission within the Sixth Framework Programme (2002-2006) projects BRIDGE (Building Radio frequency IDentification solutions for the Global Environment), IP Nr. IST-FP6-033546

References

[1] ISO 18000. Automatic Identification - Radio Frequency Identification for Item Management – Communications and Interfaces. 2001 - 2007

[2] Ari Juels. Minimalist Cryptography for low-cost RFID Tags. 2003

[3] EPC Global. Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz - 960 MHz. 2004

[4] Ari Juels. Security and Privacy in RFID systems: A Research Survey. IEEE Journal on Selected Areas in Communications, 24(2), pages 381–394. 2006 http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1589116.

[5] Shingo Kinoshita, Fumitaka Hoshino, Tomoyuki Komuro, Akiko Fujimura, and Miyako Ohkubo. Nonidentifiable Anonymous-ID Scheme for RFID Privacy Protection. 2003

[6] Anna Lysyanskaya, Ronald L. Rivest, Amit Sahai, and StefanWolf. Pseudonym Systems. 1999 citeseer.ist.psu.edu/lysyanskaya99pseudonym.html.

[7] David Molnar and David Wagner. Security and privacy in library RFID: Issues, practices, and architectures. In ACM CCS. 2004

[8] David Molnar, Andrea Soppera, and David Wagner. A Scalable, Delegatable Pseudonym Protocol Enabling Ownership Transfer of RFID Tags. In Selected Areas in Cryptography, pages 276–290. 2005

Stop Tampering of Products (SToP) – Integrated Processes for Product Authentication with Special Consideration of Mobile Phones

Harald Vogt[1], Carsten Magerkurth[2], Ali Dada[2], Jens Müller[1],
Nina Oertel[1], Felix Graf von Reischach[3]

[1] SAP AG, SAP Research, CEC Karlsruhe, Vincenz-Priessnitz-Strasse 1, 76131 Karlsruhe, Germany
[2] SAP (Schweiz) AG, SAP Research, CEC St. Gallen, Blumenbergplatz 9, CH-9000 St. Gallen, Switzerland
[3] SAP (Schweiz) AG, SAP Research, CEC Zürich, Kreuzplatz 20, CH-8008 Zürich, Switzerland
{harald.vogt, carsten.magerkurth, ali.dada, jens.mueller, nina.oertel, felix.graf.von.reischach} @ sap.com

Abstract. This demonstration illustrates variations of remote product authentication using the Product Verification Infrastructure developed in the EU project SToP. It demonstrates the interplay between in-the-field authentication using an NFC enabled mobile phone and the handling of failed authentication attempts in a back-office environment. For product authentication, RFID tag validity and visual packaging details of products in a store are checked, with the equipment allowing unobtrusive operation. Corresponding events are indicated and handled using back-office tools that integrate seamlessly in common work environments.

Keywords: Anti-Counterfeiting, NFC enabled mobile phone, Product Verification

1 Introduction

Counterfeiting of products has become a serious problem worldwide. Affecting almost all industries, it jeopardizes the well-being of consumers, diminishes the value of brands, and has an overall negative economic impact. Several initiatives have been put in place to fight against counterfeiting on a global scale and new technologies for security tagging are continuously developed in order to stay ahead of counterfeiters. Among others, RFID technology, holographic labels, fluorescent ink, and seals are widely used. Other means for anti-counterfeiting include legal prosecution and consumer education. To support the demands of different industries and use cases, measures need to be developed that integrate and combine these diverse means in a unified way.

1.1 The SToP Product Verification Infrastructure

The European research project SToP (Stop Tampering of Products) aims at developing secure, reliable, and cost effective approaches for product authentication, incorporating RFID technology (see also www.stop-project.eu).

A major result of the SToP project is a system called Product Verification Infrastructure (PVI). This system will provide mobile and stand-alone applications and devices with proper services and the related data to easily verify the authenticity of products, particularly in the pharmaceutical, luxury goods, aviation, and security document industries. Fig. 1 shows the different modules of the PVI.

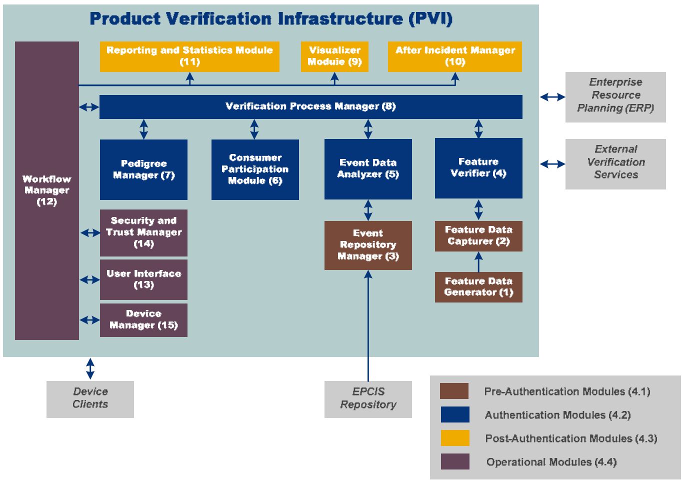

Product Verification Infrastructure (PVI)
Reporting and Statistics Module (11)
Visualizer Module (9)
After Incident Manager (10)
Verification Process Manager (8)
Workflow Manager (12)
Pedigree Manager (7)
Consumer Participation Module (6)
Event Data Analyzer (5)
Feature Verifier (4)
Security and Trust Manager (14)
User Interface (13)
Device Manager (15)
Event Repository Manager (3)
Feature Data Capturer (2)
Feature Data Generator (1)
Enterprise Resource Planning (ERP)
External Verification Services
Device Clients
EPCIS Repository
Pre-Authentication Modules (4.1)
Authentication Modules (4.2)
Post-Authentication Modules (4.3)
Operational Modules (4.4)

Fig. 1. The Product Verification Infrastructure (PVI).

The PVI supports various processes related to product authentication and the specific demonstration at IOT illustrates various variations of a remote product authentication, i.e. the usage of an NFC enabled mobile phone to check for RFID tag validity and visual packaging details of products in a store. Additionally, the demonstration also illustrates corresponding effects of the remote authentication attempts at a back office client that is used for after-incident-management and product data administration.

2 Remote Authentication Demonstration

With the demonstration we are addressing the different stakeholders such as back office workers and potential “field investigators”. Consequently, we demonstrate the complexity of the domain and solution to prove to the audience that we deliver a prototype that has external validity and is actually useful.

2.1 Scenario

The demonstration revolves around a mystery shopper scenario that shows the interplay of activities in the field and responsive actions at a back office site. The story starts with a mystery shopper who uses a mobile phone within a store in order to authenticate products. Depending on the selection of different items that will be physically brought to the demonstration site, the authentication succeeds, cannot be decided, or fails and the potential actions triggered at the manufacturer's back office are demonstrated. By using a mystery shopper scenario we circumvent potential problems of e.g. an end user scenario (direct link to the manufacturer, privacy, installation of mobile software, etc.), but are still able to show the rationale of the overall authentication process and the appropriateness of the two user interfaces. The scenario is illustrated in fig. 2.

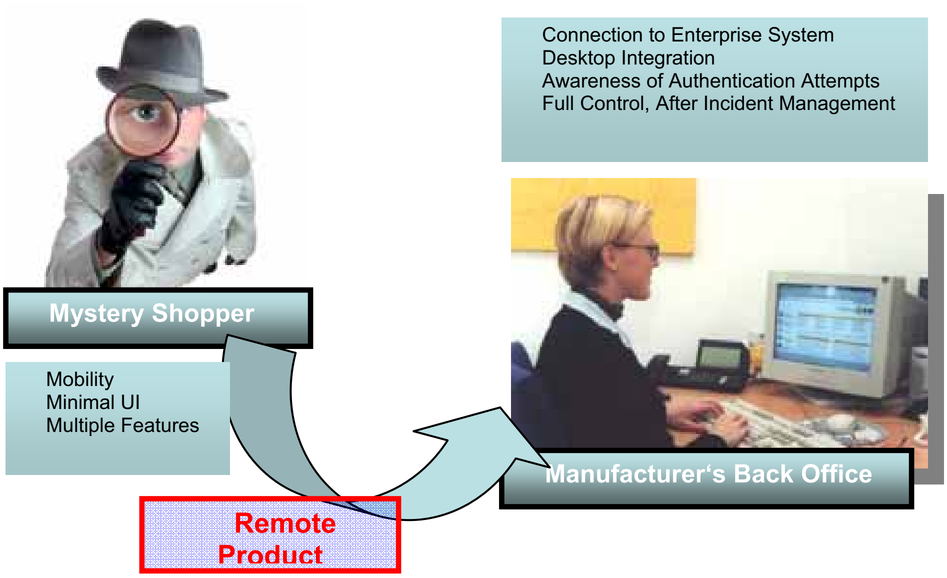

Connection to Enterprise System
Desktop Integration
Awareness of Authentication Attempts
Full Control, After Incident Management
Mystery Shopper
Mobility
Minimal UI
Multiple Features
Remote Product
Manufacturer's Back Office

Fig. 2. The Demo Scenario.

The mobile interface is optimized for quick and intuitive use, only providing information that is absolutely necessary as well as guiding the user through the process, if multiple steps are necessary for the authentication of a given item.

Correspondingly, the back office interface is optimized for a maximum amount of available information for each item/ authentication result. Furthermore, it facilitates triggering respective actions and features a scalable, non-disruptive UI that follows an ambient display approach.

2.1 Demonstration Setup

The physical setup consists of:

- NFC enabled mobile phone
- Back Office PC (laptop)
- Multiple tagged items
- Internet connection for the laptop

2.1 Process Variations

The demonstration runs through multiple variations of the basic mystery shopper process, i.e. the mystery shopper picks up an item in the store and tries to authenticate it. The corresponding effects of this authentication attempt are demonstrated at the back office laptop for each attempt.
The chain of variations and concrete process steps include:

1. Item is Genuine
    1. NFC: Start of the mobile application via touching an item
    2. NFC: Touch the item again, display the tag ID
    3. NFC: Click on "Verify"
    4. NFC: Receive and display the result
2. Item is Genuine: Show Details and Systray
    5. NFC: Touch the same item again and click on "Verify"
    6. PC: Demonstrate the temporary change in the systray icon
    7. NFC: Receive and display the result, click on "Details" to show more information
3. Suspicious Item: Complete Information
    8. NFC: Touch a suspicious item and click on "Verify"
    9. NFC: Receive and display the result ("unknown" with a closeup photo of a package detail: "does this part of the package look like this?")
    10. NFC: Perform visual 2nd authentication step by confirming the detail
4. Genuine Item: Demonstrate Dispatcher GUI and Item View
    11. NFC: Touch the same item again and click on "Verify"
    12. NFC: Receive and display the result (now "genuine")
    13. PC: Manually open (and close, and re-open) the Dispatcher GUI
    14. PC: Launch the Item View via the Dispatcher GUI and explain the checking history of the item
5. Fake Item: Initiate After Incident Management
    15. NFC: Touch a fake item and click on "Verify"
    16. NFC: Receive and display the result
    17. PC: Automatically open the Dispatcher GUI, open the Item View and send email to a third person to take care of the fake incident.

Synchronized Secrets Approach for RFID-enabled Anti-Counterfeiting

A. Ilic[1], M. Lehtonen[1], F. Michahelles[1], E. Fleisch[1,2]

[1]ETH Zurich, Information Management, CH-8092 Zurich, Switzerland
[2]University of St. Gallen, Institute for Technology Management, CH-9000 St. Gallen, Switzerland
{ailic,mlehtonen,fmichahelles,efleisch}@ethz.ch

Abstract. In today's global marketplace, brand owners need techniques for guaranteeing the authenticity of their products. By linking physical products with digital identities through RFID, secure and automatic authentication checks can be used to prevent counterfeit products from entering the licit distribution channel. However, cryptographic RFID tags seem still too expensive to be used on a broad scale. In this demo, we want to show how standard low-cost RFID tags can be used for anti-counterfeiting in a non-conventional but efficient way. We use a method that detects desynchronization when cloned tags are introduced in a protected channel and thus helps to prevent the further distribution of the counterfeit products.

Keywords: anti-counterfeiting, RFID, clone detection, synchronized secrets

1 Introduction: The Challenge of Anti-Counterfeiting

Product counterfeiting is an ever increasing problem that affects trademark and brand owners, governments, as well as consumers [1]. Though some aspects of the problem are often considered relatively harmless by the public, such as purchasing of fake designer handbags from street vendors, the intellectual property rights and investments of licit businesses must be protected. Furthermore, in more dangerous forms of product counterfeiting, the fake products are injected into the licit distribution channel and sold as genuine articles [2]. While potentially risking the health and safety of consumers, in this way the counterfeiters can sell their articles in higher price for higher profits. As a result, the licit supply chains need to be protected from counterfeit products. The emerging electronic pedigree in the pharmaceutical industry (e.g. [3]) is a prominent example of measures towards this objective.

Radio-frequency identification (RFID) is an emerging Automatic Identification technology. RFID systems comprise tags that are attached to products, interrogators that read and write data on tags, and back-end systems that store and share data. RFID has recognized potential in anti-counterfeiting [4]. Probably the most common approach to authenticate an RFID-tagged product is to cryptographically authenticate the transponder. However, cryptographic tags have cost and performance disadvantages due to their additional hardware and processing time requirements. In addition, cryptographic RFID tags remain computationally limited and are vulnerable to different tag cloning attacks, such as, cryptanalysis and reverse-engineering [5] and side-channel attacks [6]. As a result, cryptographic tags do not seem to deliver best possible trade-offs between cost, security, and performance today.

In our approach, we do not attempt to prevent tag cloning but instead we try to detect the cloned tags. Our approach is less expensive than cryptographic tag authentication in terms of tag price and tag processing time, and it provides a high level of security where the genuine products are repeatedly read in a high rate. The limitation of our approach is that in certain conditions the system will trigger a false alarm for the genuine product and thus another level of inspection is needed to ascertain the product's origins. However, we make use of the fact that RFID tags will be deployed anyway and our approach can be implemented with minimal hardware overhead that is some bytes of rewritable memory.

2 Demo

Our application scenario is the following: A manufacturer of pharmaceutical products inserts tags to individual articles at the manufacturing site. These products are distributed through multiple steps to a hospital or a pharmacy and they are authenticated throughout the supply chain to detect counterfeit products (cf. Fig. 1).

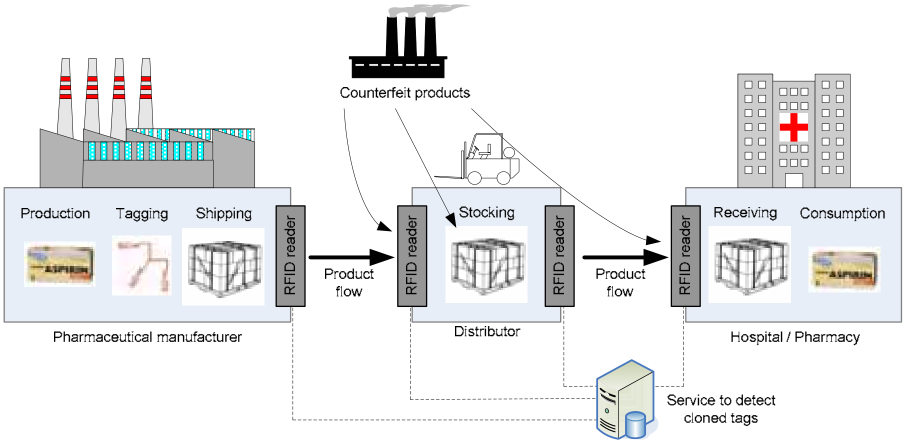


Fig. 1. Simplified view of an RFID-enabled pharmaceutical supply chain

In our demo scenario we relate to the practical context of the EU-funded project BRIDGE and show a part of an actual pharmaceutical supply chain. In this demo, the IoT conference visitors will see two Supply Chain Stations representing a pharmaceutical wholesaler and a retailer. As a Supply Chain Station we refer to the combination of a computer running the clone detection client, and an attached RFID reader (cf. Fig. 2). In addition, we set up a Supply Chain Station representing a malicious supplier that is able to clone tags and affix them to counterfeit products that are injected in the previously described two-stage supply chain.

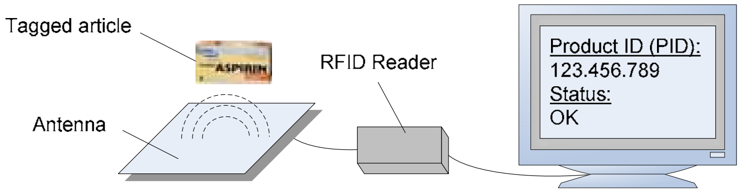
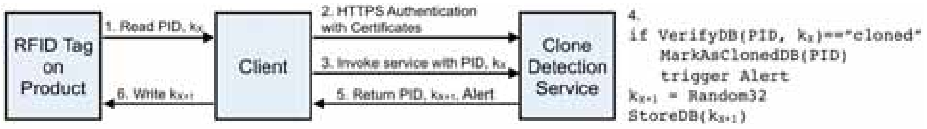


Fig. 2. Illustration of the hardware set-up of one Supply Chain Station

3 System

Our system bases on a web service for clone detection through synchronized secrets (e.g. [8]). We assume that all genuine articles are equipped with an RFID tag, which is checked at least once at every stage of the supply chain. A counterfeiter is therefore forced to equip also fake products with RFID tags. Following the Service-oriented Architecture paradigm, our solution provides a service to verify and update the synchronized secrets of tags. The same secret k_X is stored on both the tag's memory and the backend database. On every web service invocation, a new random secret k_{X+1} is generated and updated in both, the backend database and the tag's memory. If a genuine tag's identifier (PID) and synchronized secret are copied to a fake tag which is affixed to a counterfeit article, and the counterfeit article is injected into the supply chain, the backend will detect a desynchronization, i.e. a tag with invalid secret, as soon as both the genuine and counterfeit article are read. The backend service triggers an alarm upon desynchronization and a further verification, based on other authentication techniques, can be conducted to ascertain which article is the cloned one. A sequential illustration of the clone detection protocol flow, with pseudo code for the most relevant backend steps, is depicted on Fig. 3.


RFID Tag on Product
Client
Clone Detection Service
1. Read PID, k_X
2. HTTPS Authentication with Certificates
3. Invoke service with PID, k_X
5. Return PID, k_{X+1}, Alert
6. Write k_{X+1}
4.

```
if VerifyDB(PID, kX)=="cloned"
    MarkAsClonedDB(PID)
    trigger Alert
kX+1 = Random32
StoreDB(kX+1)
```

Fig. 3. Illustration of the clone detection protocol flow with pseudo code in the backend

The backend web service is implemented in PHP using a relational MySQL database to keep track of the synchronized secrets. The communication interface bases on XML_RPC and uses https as an encrypted transportation protocol. To demonstrate the clone-detection in practice, we implemented a user interface in Java that represents an organization's software that invokes the back end web service. To communicate with RFID readers and RFID tags, the client uses the open source EPC Network implementation Accada [7]. Each time a tag is read (by an RFID reader), the client invokes the web service's clone detection method, which verifies the synchronized secret through a lookup in the backend database. If clones are detected the user receives a visual alert together with the information about the other suspicious products and their last reading locations (if known). Also, as it is not safe to say whether the product with the desynchronized secret is the fake one, the database marks the product ID number as “cloned”. Next time a product with marked ID number is read, the user will receive automatically a warning that the product might be subject to counterfeiting.

4 Conclusions

Our demo presents a simple but effective method for detecting cloned RFID tags in an anti-counterfeiting application. In contrast to cryptographic RFID tags, our approach is more cost-effective as it can be deployed with low-cost tags. Our clone detection protocol bases on a desynchronization detection mechanism that triggers an alert when a cloned product is injected into the licit supply chain. The limitation of the presented approach is that it alone cannot prove which of the suspicious products is the counterfeit one and which is the genuine one. We demonstrate, however, that in practice already the awareness of counterfeits and the knowledge about their most recent locations can be used to effectively deter counterfeiting.

Acknowledgements. This work is partly funded by the European Commission within the Sixth Framework Programme (2002-2006) projects BRIDGE (Building Radio frequency IDentification solutions for the Global Environment), IP Nr. IST-FP6-033546, and SToP (Stop Tampering of Products), Nr. IST-FP6-034144.

References

1. Organization for Economic Co-operation and Development (OECD): The Economic Impact of Counterfeiting. (1998)
2. Lehtonen, M., Al-Kassab, J., Graf von Reischach, F., Kasten, O., and Michahelles, F.: Problem-Analysis Report on Counterfeiting and Illicit Trade. Deliverable D5.1 of EU-BRIDGE Project, July 2007.
3. EPCglobal Inc.: Pedigree Ratified Standard. Version 1.0, January 5th, 2007. Available: http://www.epcglobalinc.org/standards/pedigree/pedigree_1_0-standard-20070105.pdf
4. U.S. Food and Drug Administration (FDA): Combating Counterfeit Drugs. February 2004. Available: http://www.fda.gov/oc/initiatives/counterfeit/report02_04.html
5. Bono, S., Green, M., Stubblefield, A., Juels, A., Rubin, A., Szydlo, M.: Security analysis of a cryptographically enabled RFID device. 14th USENIX Security Symposium (2005).
6. RFIDJournal: EPC Tags Subject to Phone Attacks. News Article, February 24, 2006. Available: http://www.rfidjournal.com/article/articleview/2167/1/1/
7. Floerkemeier, C., Roduner, C., and Lampe, M.: RFID Application Development With the Accada Middleware Platform. IEEE Systems Journal, Volume 1, Issue 2 (2007)
8. Ilic, A., Michahelles, F., and Fleisch. The Dual Ownership Model: Using Organizational Relationships for Access Control in Safety Supply Chains. IEEE International Symposium on Ubisafe Computing (UbiSafe-07), Niagara Falls, Ontario, Canada (2007)

Tomu-R: Real-Time Query for Massive Sensor Databases

Keisuke Kanai[1], Hiroki Ishizuka[2], Yoh Shiraishi[3], Yoshito Tobe[1]

[1]Department of Information Systems and Multimedia Design, Tokyo Denki University
[2]Department of Information and Media Engineering, Tokyo Denki University
[3]Center for Spatial Information Science, The University of Tokyo
Email: {ksk, isi}@u-netlab.jp, siraisi@csis.u-tokyo.ac.jp, yoshito_tobe@osoite.jp
[Yoshito Tobe is co- affiliated with CREST, Japan Science and Technology Agency]

Abstract.
Recently, technologies for massive sensor databases have been developed along with the dissemination of large-scale sensor networks. Data processing for such databases is hardly acceptable to be utilized in real-time applications since it consumes much time to process a query. Base on this background, we have built Tomu-R, a system which enables an incremental response to a user's query for processing massive sensor data in real-time. Tomu-R divides a query issued an application to assure the real-time performance based on the predicted processing time of the query. Tomu-R continues publishing divided queries to a massive sensor database incrementally until user's utility is satisfied. In this demonstration, we show the behavior of Tomu-R for a navigation application.

Key words: Real-time systems, Database systems, Sensor Networks

1. Introduction

The dissemination of sensor networks is transforming the real-world into a computing platform. Technologies for massive sensor databases [1, 2] have been paid much attention as the size of sensor networks increases. In contrast, various applications of sensor networks are currently being developed. Some of these applications such as an ambient route navigation using a large number of sensor data require processing in real-time. Conventional technologies of data processing has not considered being able to adapt a massive sensor database. To solve the problem, we propose an incremental response to a user's query called Tomu-R for processing massive sensor data in real-time. Tomu-R divides user's query according to the number of predicted tuples which are enabled to finish processing in user's deadline and continues publishing divided queries to a massive sensor database incrementally until user's utility is satisfied. To predict the number of tuples that can be processed in real-time, we consider two characteristics related to data processing for a massive sensor database. First, a user usually specifies at least one attribute for a query such as a sensed time, sensed position and sensed value whenever the user publishes a query to a sensor database. To predict the number of tuples for the query, Tomu-R maintains the distribution of all data in

the database regarding such three attributes. Second, every sensor intermittently inserts sensed data into the database. Tomu-R needs to observe the data distribution of the database as accurately as possible.
In this demonstration, we explain the architecture of Tomu-R and its prototype implementation and describe a demonstration scenario using a navigation application for downtown Tokyo.

2. System Architecture

The architecture of Tomu-R indicated in Figure 1 consists of profiling and processing functions. The profiling function maintains the data distribution of a sensor database based on three attributes: a sensed time, position, and value by observing streams of sensed data. First, an application publishes a normal query *q(T, U, SQL)* including time *T*, utility *U* and a statement of SQL *SQL* to Tomu-R. We define the utility *U* as follows: N_α and n_i represent the number of predicted tuples for the normal query and the number of tuples for the *i* th divided query, respectively.

$$U = \frac{\sum_i n_i}{N_\alpha} \qquad \text{(exp 1)}$$

Next, Tomu-R predicts the processing time T_α of the normal query in the profiling function and calculates the number of predicted tuples which finishes processing while *T* comparing *T* to T_α in the processing function. Finally, the processing function divides the normal query according to the number of predicted tuples and continues publishing divided queries incrementally to the massive sensor database until ***q(U)*** is satisfied.

3. Prototype Implementation and Experiments

We implemented a prototype of Tomu-R. The prototype was implemented as a query optimizer installed on PostgresSQL8.1.2. The profiling and processing functions were written in Java and the size amounted to approximately 1500 lines. In addition, we show its effectiveness through experiments with utilizing data set from real deployed sensor networks. Each experiments leverage approximately 110 thousand datasets which were measured during 3rd to 8th in August, 2007 at the UScan[3] system. Moreover, sensed value in attributes shows temperature because UScan measures temperature data aiming to analyzing urban heat island. In our experiments, the query included specified time pressure from 0s to 60s and published per second in the following scenario.

Exp 1: Query specified an interval of sensed time
Example of SQL statement is SELECT * FROM data WHERE time BETWEEN '2007-08-03 00:00:00' AND '2007-08-04 00:00:00';

Exp 2: Query specified a width of sensed value

Example of SQL statement is SELECT * FROM data WHERE value BETWEEN 36 AND 40;.

Exp 3: Query specified a sensed area

Example of SQL statement is SELECT * FROM data WHERE (latitude BETWEEN 139.761 AND 139.763) AND (longitude BETWEEN 35.692 AND 35.693);.

Figure 3 shows the result of each experiment. Additionally, the y-axis of the graph indicates a processing time of a first divided query. Since all queries could ensure user's deadline, we demonstrated that Tomu-R was useful for a massive sensor database system in real-time.

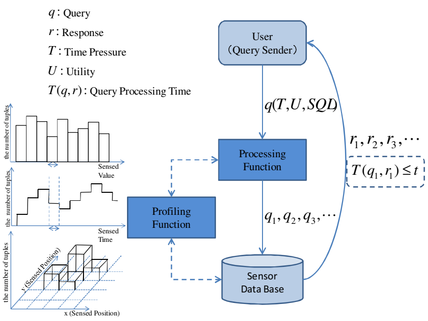


Fig. 1. Architecture of Tomu-R

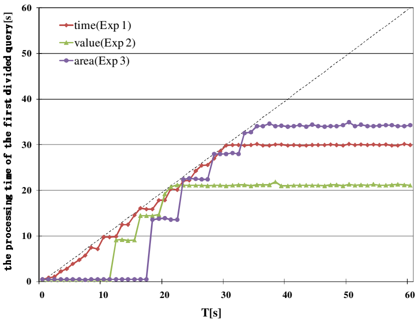


Fig. 2. Result of experiment

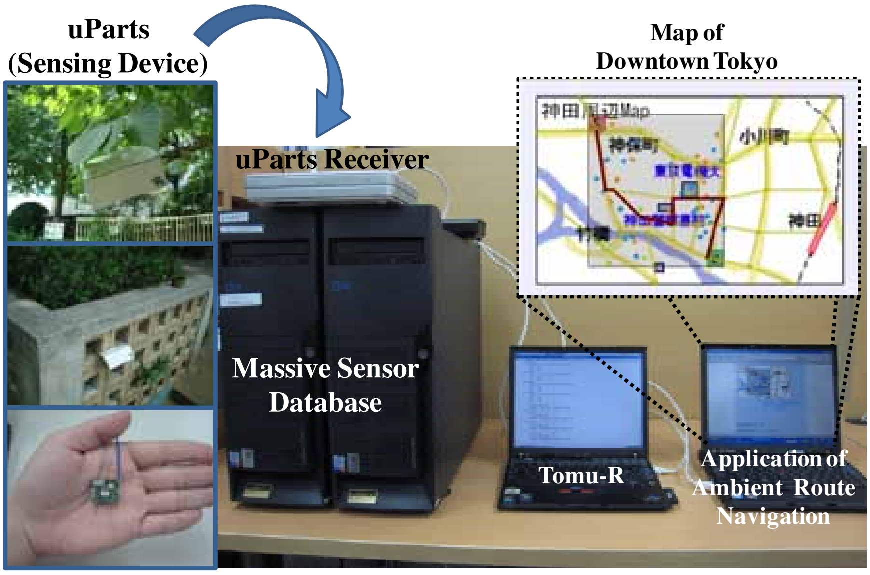


Fig. 4. Demonstration scenario

4. Demonstration Descriptions

We assume applications that can support our urban lives. One can find an ambient walking route and an "oasis" spot, a windy and low-temperature location in hot summer. To enable such applications for ambience, weather-related sensors need to be placed in a fine grained manner. The sensor database of such applications needs to retain a large amount of sensor streams. Therefore, a normal query processing is in nature useless since it consume much time when a user retrieves an ambient walking route from the huge database with its required time. To retrieve the route for a user in a real-time fashion, we utilize our proposed system that enables real-time processing by dividing a query. In this demonstration, our system prioritizes data around a user’s location instead of processing the whole data. Consequently, the user always acquires an ambient route around user's location incrementally.

Figure 4 shows our demonstration scenario. We have a system called UScan by which temperature in fine resolution was measured by uParts [4] in downtown Tokyo. Our demonstration assumes application of ambient route navigation utilizing massive sensor data that has been gathered from UScan.

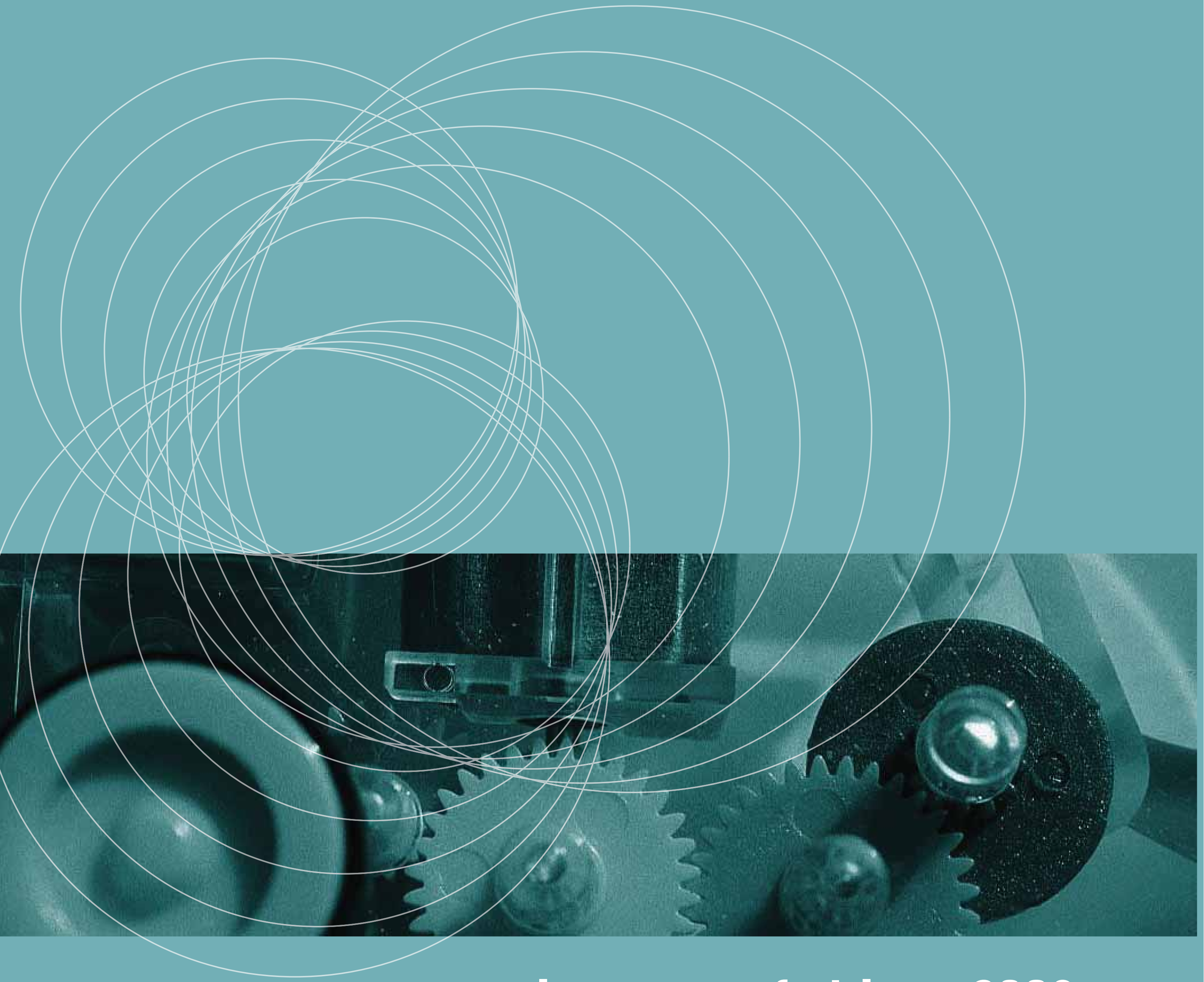
5. Conclusion

In this demonstration, we proposed an incremental response to a user’s query called Tomu-R for processing massive sensor data in real-time. Tomu-R maintains the data distribution of a sensor database based on three attributes: a sensed time, position, and value by observing streams of sensed data, and divides user query using the data distribution. We implemented a prototype of Tomu-R and experimented in three scenarios. The result of experiment shows our system is available as data processing of a massive sensor database in real-time.

Reference

1. Daniel J. Abadi, Donald Carney, Ugur Cetintemel, Mitch Cherniack, Christian Convey, Sangdon Lee, Michael Stonebraker, Nesime Tatbul, and Stanley B. Zdonik. Aurora : a new model and architecture for data stream management. VLDB Journal, 12(2), (2007)
2. Hideyuki Kawashima. Kraft: A real-time active dbms for signal streams. In Proc of the International Conference on Networked Sensing Systems, (2007)
3. Takahiro Ono; Hiroki Ishizuka; Kanoko Ito; Yasuyuki Ishida; Shohei Miyazaki; Oru Mihirogi; Yoshito Tobe ,UScan: Towards Fine-Grained Urban Sensing, International Workshop on Real Field Identification(RFId2007)
4. M. Beigl, A. Krohn, T. Riedel, T. Zimmer, C. Decker, M. Isomura.The uPart Experience: Building a wireless sensor network. In IPSN, 2006.

INTERNATIONAL CONFERENCE
FOR INDUSTRY AND ACADEMIA
MARCH 26-28, 2008 / ZURICH



www.internet-of-things-2008.org

Official Carrier
swiss
Swiss International Air Lines

SPONSORED BY >

SAP

SIEMENS

METRO Group
Future Store Initiative

Google

ambient

BRIDGE

IBM

ERCIM