Contents

Guest Editorial

magyar@tmit.bme.hu
szabo@tmit.bme.hu
klaus@nict.go.jp

The following six papers were selected from the rich content of the '**FuturICT 2009**' conference that was held in Budapest, Hungary, on 29-30 June 2009. The name connotes the First Hungarian-Japanese Joint Conference on Future Information and Communication Technologies.

In 2008, two Memoranda of Understandings (MoUs) were signed, creating a new framework of cooperation between the Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics (BME TMIT), the National Institute of Information and Communications Technology (NICT), and the Yokosuka Research Park (YRP) R&D Promotion Committee. Concrete information and researcher exchange started soon thereafter. One highlight of this international cooperation under the umbrella of the MoUs definitely was the FuturICT 2009 conference organized in Budapest during the Japan-Hungary Jubilee Year 2009.

2009 was a double anniversary for the Hungarian-Japanese diplomatic relations. We celebrated the 140th anniversary of the establishment of diplomatic relations between Japan and the Dual Monarchy of Austria-Hungary as well as the 50th anniversary of the resumption of bilateral diplomatic relations between Japan and Hungary.

The conference in the Jubilee Year was an outstanding opportunity to enhance mutual understanding between researchers and institutions, as well as strengthening the scientific relations between the two countries. It was jointly organized by BME TMIT, NICT and YRP. Interested participants working in the field of ICT were invited from Hungary and Japan. The conference secretariat was provided by the Scientific Association for Infocommunications, Hungary (HTE).

The focus of the conference was on the challenges in emerging fields such as networks of the future, ubiquitous computing, services and applications, and future wireless communications networks and technologies. The Guest Addresses were given by His Excellency Shinichi Nabekura, Ambassador of Japan to Hungary, Ábel Garamhegyi, State Secretary for International Economic Relations, Hungary, and Akira Terasaki, Vice-Minister for Policy Coordination, Ministry of Internal Affairs and Communications, Japan.

There were 154 registered participants (from 6 different countries), 24 presentations and 19 posters during the two days of the conference. Besides two keynote speeches, the program included three thematic sessions with several invited and short technical presentations. Also posters and demonstrations were on display during the two days.

The conference was supported by the National Office for Research and Technology (NKTH) Hungary, but it attracted sponsors (NEC, Panasonic, Sony) from the industry as well. NEC Eastern Europe Ltd. was a Gold Sponsor of the event. The program, abstracts, electronic versions of presentations and conference details are available on the conference web site: *http://www.futurict.org/*. A small photo gallery of the event has also been added.

Each thematic session of the conference is represented here by a set of two papers written by both Hungarian and Japanese authors resulting in a total of six papers.

"ETOMIC: A next generation experimental facility" (by *István Csabai, Attila Fekete, Péter Hága, Béla Hullár, Gábor Kurucz, Sándor Laki, Péter Mátray, József Stéger, Gábor Vattay)* describes a network traffic measurement platform with high precision GPS-synchronized monitoring nodes.

The paper of *Fumio Teraoka* and *Yunsop Han* deals with fast handover and fast failover mechanisms based on cross-layer collaboration among the link layer, the network layer and the transport layer.

Koji Zettsu, Yutaka Kidawara and *Yasushi Kiyoki* draw our attention to next generation web as collaboration media, rather than a traditional content-centered framework of information.

Tamás Gábor Csapó, Csaba Zainkó and *Géza Németh* provide a study of prosodic variability methods in a corpus-based unit selection TTS.

The paper "Fast-response inter-vehicle communications" *(Sadao Obana, Ryu Miura, Hiroyuki Yomo, Oyunchimeg Shagdar, Takashi Ohyama, Hideo Tsutsui, Michio Miyamoto, Eiji Takimoto, Yoshihisa Kondo, Jun Hasegawa and Suhua Tang)* gives a new perspective of supporting safe driving by exchanging information on vehicle location and status.

And last but not least, *László Toka, László Kovács* and *Attila Vidács* describe a novel dynamic spectrum sharing management scheme in which the allocation and the pricing of radio frequency bands are performed in a distributed manner.

We hope that the reader finds the contents of these six carefully selected papers interesting and that with the publication of this issue a much broader audience will be made aware of the on-going intensive information exchange and collaborative research activities between Hungary and Japan.

Gábor Magyar (BME TMIT)
Róbert Szabó (BME TMIT)
Werner Klaus (NICT)

IEEE VTS Connecting the Mobile World

VTC2010-Fall
OTTAWA
Connecting the Mobile World

2010 IEEE 72nd Vehicular Technology Conference

6–9 September 2010, Ottawa, Canada

CALL FOR PAPERS

This semi-annual flagship conference of the IEEE Vehicular Technology Society will bring together individuals from academia, government, and industry to discuss and exchange ideals in the fields of wireless, mobile, and vehicular technology. The conference will feature world-class plenary speakers, tutorials, and technical as well as application sessions.

General Co-chairs
Halim Yanikomeroglu, *Carleton University, Canada*
John Reid, *CATAAlliance, Canada*

Technical Program Chair
Xuemin (Sherman) Shen, *University of Waterloo, Canada*

Keynote and Plenary Chair
Hussein Mouftah, *University of Ottawa, Canada*

Panels Co-chairs
Lajos Hanzo, *University of Southampton, UK*
David Falconer, *Carleton University, Canada*

Tutorials Co-chairs
Hossam Hassanein, *Queen's University, Canada*
Zhisheng Niu, *Tsinghua University, China*

Workshops Co-chairs
Nirwan Ansari, *NJIT, USA*
Nei Kato, *Tohoku University, Japan*

Patronage & Exhibits Co-chairs
Jim Budwey, *ICTSGroup, USA*
Barry Gander, *CATAAlliance, Canada*

Finance Chair
Dennis Bodson, *IEEE VTS, USA*

Local Arrangements Co-chairs
Sreeraman Rajan, *DRDC, Canada*
Petar Djukic, *Carleton University, Canada*

Publicity Co-chairs
Chuang Lin, *Tsinghua University, China*
Dongmei Zhao, *McMaster University, Canada*

Technical Advisory Committee Chair
James M. Irvine, *University of Strathclyde, UK*

Conference Administrator
Jim Budwey, *ICTSGroup, USA*

We invite the submission of original, unpublished technical papers in the following areas

Transportation
Wireless Access
Wireless Networks
Antennas & Propagation
Ad-hoc & Sensor Networks
Transmission Technologies
Mobile Applications & Services
Vehicular Electronics & Telematics
Mobile Satellite & Positioning Systems
Cognitive Radio & Cooperative Communications
Multiple Antenna Systems & Space-Time Processing

Papers due
15 February 2010

Acceptances will be sent on
1 May 2010

Camera-ready papers due
1 June 2010

Photo courtesy Ottawa Tourism

Prospective authors are encouraged to submit a 5-page full paper (or a 2-page extended abstract including results) through the conference web site.

For more information, visit www.vtc2010fall.org

Fast-response inter-vehicle communications

SADAO OBANA, RYU MIURA, HIROYUKI YOMO, OYUNCHIMEG SHAGDAR, TAKASHI OHYAMA, HIDEO TSUTSUI, MICHIO MIYAMOTO, EIJI TAKIMOTO, YOSHIHISA KONDO, JUN HASEGAWA, SUHUA TANG

Adaptive Communications Research Laboratories,
Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan
obana@atr.jp

Keywords: ITS, Inter-vehicle Communications, Safe Driving, Media Access Control (MAC), CSMA, CDMA

Safe driving support is one of the most attractive and important applications of an inter-vehicle communication systems. Real-time and reliable exchange of status information on such as vehicle location, speed, sudden braking etc., among vehicles, is a key to offering prompt warnings to drivers in order to avoid fatal traffic accidents.
We have proposed a novel media access control (MAC) scheme based on code division multiple access (CDMA) technology, which offers fast response and high packet delivery ratio to meet the above requirements. This scheme is inherently robust to the hidden terminal problem and significantly outperforms the conventional MAC scheme, CSMA/CA, in the environment with high vehicle density. This paper introduces the proposed scheme, performance evaluation by simulation, and prototyping for field experiment. It also mentions future studies.

1. Introduction

More than 100 years have already passed since Carl Benz invented the gasoline automobiles. Automobiles have become an essential part of our dairy lives today. However, serious problems such as humans' death due to traffic accidents, loss of time due to traffic congestion and environmental pollution due to CO_2 emission, essentially remain unsolved.

Inter-vehicle communications that promptly exchanges status information such as on the vehicle location, speed, sudden braking etc. are expected to reduce the traffic accidents such as collisions among vehicles, and to reduce humans' death.

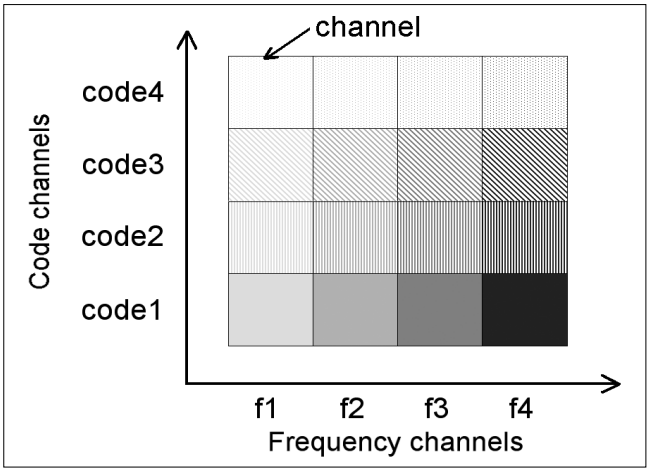
For inter-vehicle communications, much research [1-8] has been carried out and standardizations are now in process. Especially, IEEE is now going to standardize as 802.11p. Also in Japan, ITS Info-communications Forum, under the Ministry of Internal Affair and Communications of Japan, is now in the process to standardize specifications of inter-vehicle communication focusing on safe driving [9].

Communication schemes adopted in the most of the researches and the standardization on inter-vehicle communications, are based on the conventional media access control (MAC), i.e. carrier sense multiple access/ collision avoidance (CSMA/CA) [10], which is popularly used in wireless LANs.

However, CSMA/CA has a limit in transmission delay and packet delivery ratio, due to its control scheme based on the carrier sensing. For satisfying challenging requirements on supporting safe driving, we have proposed a new scheme called "Multi-carrier Multi-code Spread ALOHA (MM-SA)", which is based on code division multiple access (CDMA) [11] technology. This scheme significantly reduces the transmission latency to millisecond order and improves the packet delivery ratio among vehicles [12-17]. It also has inherent robustness to the increase of vehicle density as well as to the hidden terminal problem [18].

This paper introduces our research on novel communication technologies for safe driving, using wireless ad hoc network, that promptly exchanges the status information on the vehicles such as on the vehicle location, speed, sudden braking etc. The later parts of this paper consist of the following sections: Section 2 introduces the communication scheme of MM-SA; Section 3 shows the performance evaluation by computer simulation; Section 4 refers to the prototyping to evaluate the performance in actual environment. Finally, we summarize future studies.

Figure 1.
Example of Communication Channel Structure in MM-SA



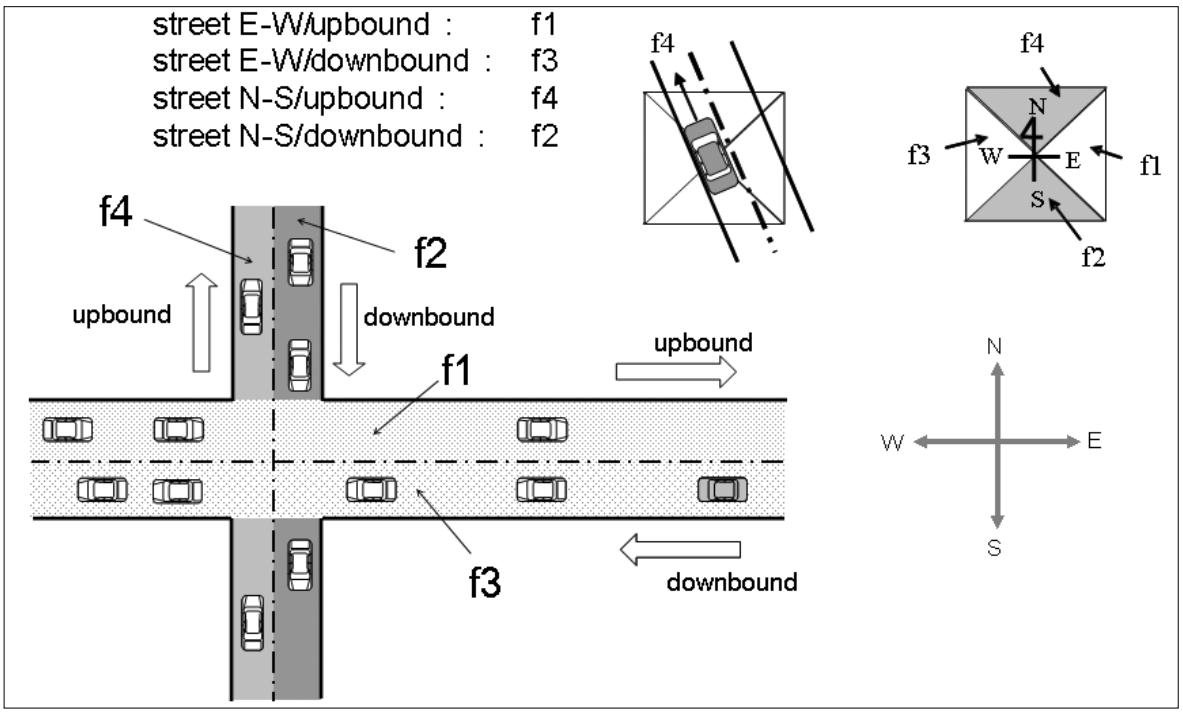


Figure 2. Frequency channel assignment example

2. Communication scheme of MM-SA

2.1 Features of MM-SA

MM-SA has the advantages of both FDMA (frequency division multiple access) and CDMA (code division multiple access). It allows prompt and reliable communication, even under the condition of high vehicle traffic density, by means of multiplexing in frequency channel domain and spreading code domain as shown in *Fig. 1*. Transmitted packet can be recovered regardless of packet collisions by using spreading code. Communication traffic is diffused to multiple frequencies.

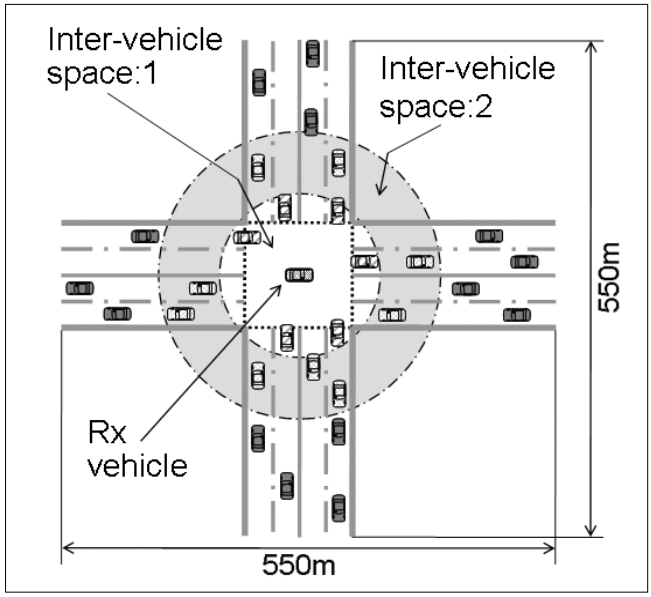
In the MM-SA, in order to reduce signal interference from other vehicles, frequency channel is autonomously determined by a vehicle itself, according to the vehicles' moving direction detected by GPS. Specifically, as shown in *Fig. 2*, assuming that four frequency channels are available, the moving direction in ±45° around the north is mapped to a channel, e.g., f3, and the moving direction in ±45° around the west is mapped to a channel, e.g., f2, and so forth.

However, this is the simplest case. In actual roads and intersections, topology or geometry of them is more complicated and in some cases, the same frequency channel may unfortunately be assigned to different roads. In such cases, frequency channels are determined by an additional control mechanism such as priority control etc.

2.2 Basic communication characteristics of CSMA/CA and CDMA

Basic communication characteristics of CSMA/CA and CDMA on which MM-SA scheme is based, for a single frequency channel (4.096 MHz bandwidth) in 5.8 GHz band, which are allocated for ITS wireless communications, are compared using Qualnet network simulator [19]. The parameters of CSMA/CA scheme are tuned to the specification defined by ITS Info-communications Forum. Transmission power (10 mW) and modulation type ($\pi/4$ shift QPSK) of CDMA are the same as the ones of CSMA/CA. The sensitivity levels of CSMA/CA and CDMA are -94.41 and -94.5 dBm, respectively. Spreading factor in CDMA is 7 and the contention window in CSMA/CA takes value from 1 to 256.

Figure 3. Simulation topology



Performance comparison targets the topology illustrated in *Fig. 3*, where vehicles are uniformly distributed on a 2-lane per direction. In the simulation, each vehicle (Tx nodes), except the one in the center of the intersection (Rx node), periodically generates status message of 140 bytes. Message generation cycle is 100 msec that are considered to be adequate for keeping track of

surrounding vehicles' status. *Table 1* shows the comparison on the average delay for different number of Tx nodes. *Fig. 4* shows average packet delivery ratio vs. inter-vehicle space, which indicates the relative position of Tx node to Rx node.

The number of Tx vehicles	100	200	300	400	500
CSMA/CA [ms]	4.9	72.4	133	168	193
CDMA [ms]	1.9	1.9	1.9	1.9	1.9

Table 1. Comparison in packet transmission delay

From Table 1 and Fig. 4, CSMA/CA obviously suffers from performance degradation in terms of transmission delay and packet delivery ratio, when the number of vehicles increases. The hidden terminal problem will further degrade the performance.

In the CDMA, while vehicles experience difficulty in receiving data from far vehicles, due to the near-far effect [20] which is the typical characteristic of CDMA, this feature enables prompt and reliable information exchange among near vehicles, without being affected by the number of vehicles. This implies that the degradation of the performance due to hidden terminals is quite small. In inter-vehicle collision avoidance, prompt and reliable information exchange among near vehicles is very important rather than that among far vehicles.

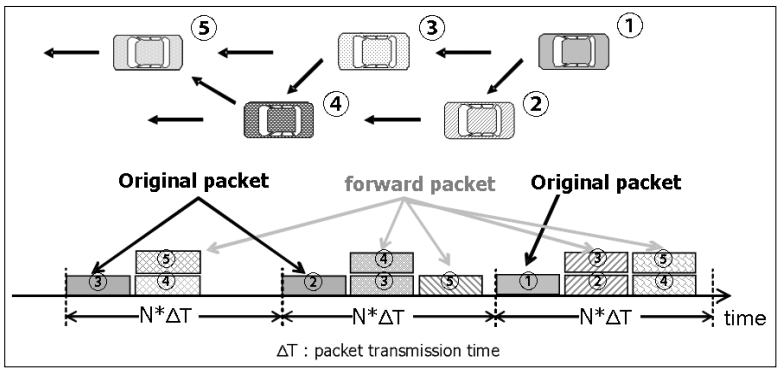
2.3 Packet forwarding scheme of MM-SA

In MM-SA, in order to enable vehicles to be aware of existence of far vehicles, packet forwarding scheme is adopted. In general, broadcast or flooding for packet forwarding causes, as known as the broadcast storm problem [21]. In order to avoid this problem, we have adopted the following strategies.

(1) To limit forwarding packets over a limited area determined by the vehicle's location (e.g. 100m in front and 10m in sides)?
(2) Not to forward packet with duplicated and/or outdated information.
(3) To adequately schedule packet transmissions, as illustrated in *Fig. 5*.

The key point of this scheduling is the control of vehicles'packet transmission time, in such a way that vehicles concurrently transmitting their own messages be located as far as possible from each other. This objective can be achieved by having each vehicle to generate its original message N*ΔT later than the message generation time at its adjacent vehicle in front. Here, ΔT is the packet transmission time and N is an integer constant. In our experience, N=3.

Figure 5. Packet forwarding scheme of MM-SA



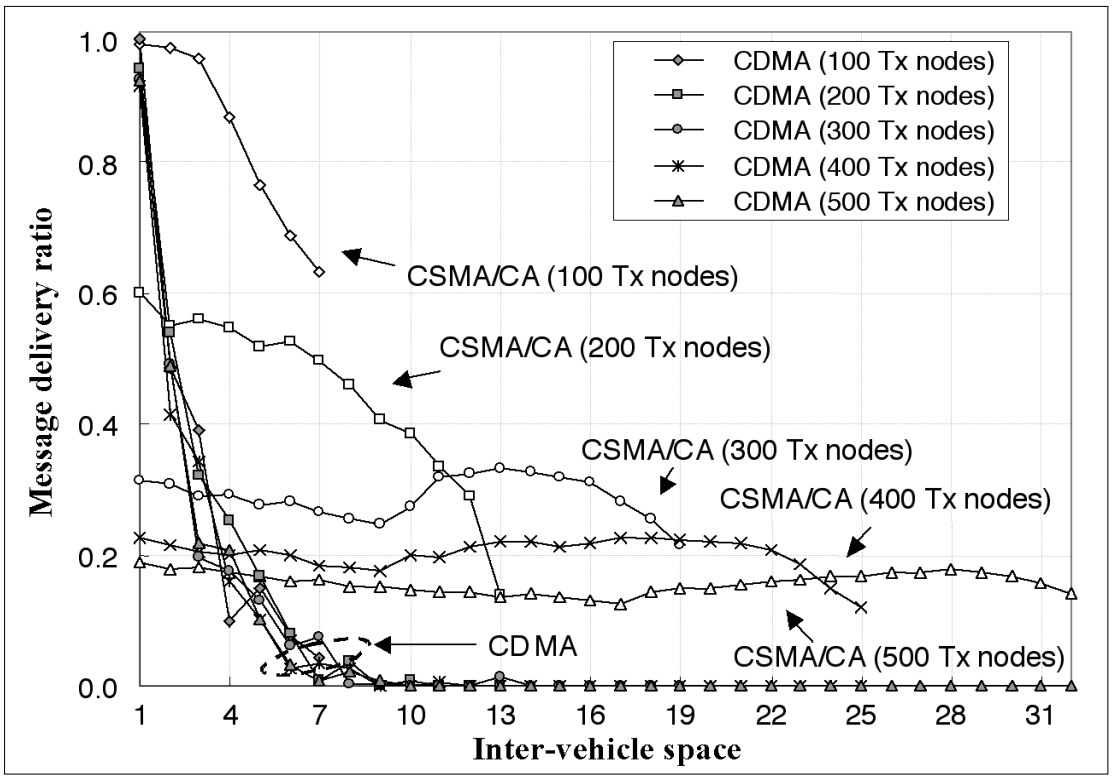


Figure 4. Comparison in packet delivery ratio

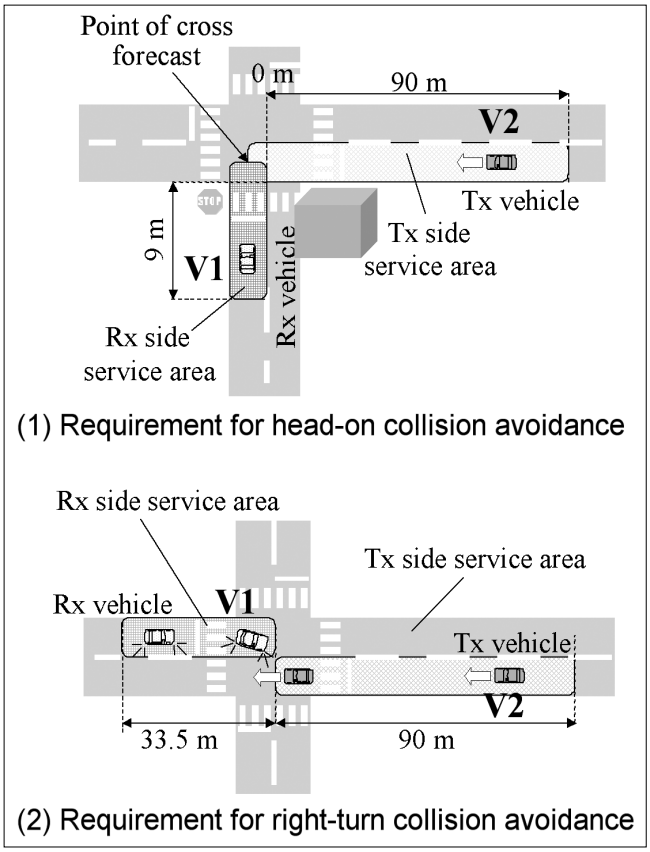
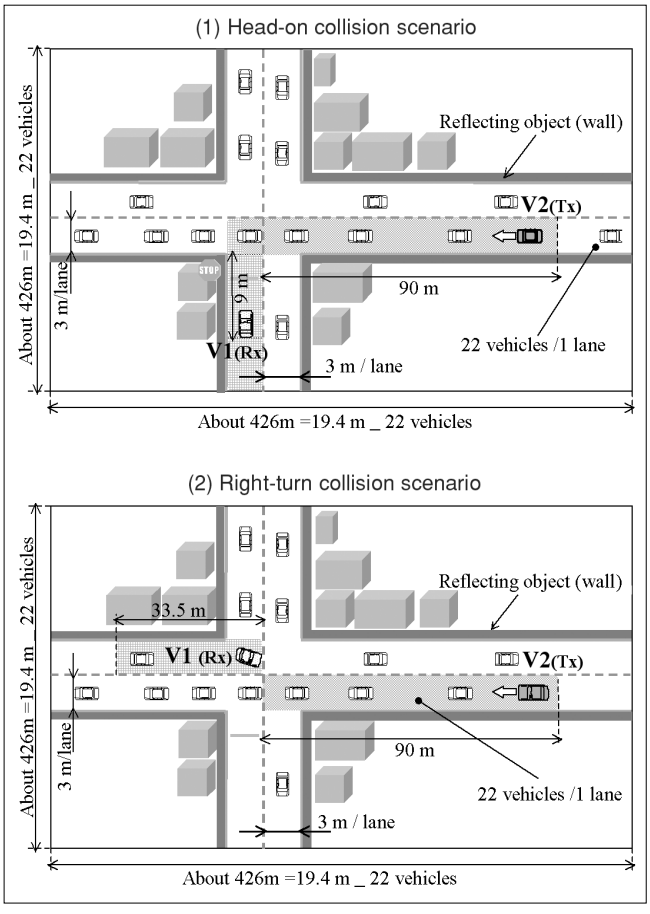


Figure 6.
Collision scenarios and ASV requirements

Figure 7.
Simulation conditions



3. Performance evaluation

3.1 Evaluation strategy

In Japan, the Advanced Safety Vehicle (ASV) program [22] assisted by automotive manufacturers and the Ministry of Land, Infrastructure, Transport and Tourism, has been defining the communication requirements for various safe driving scenarios. Among the scenarios, an especial emphasis has been put on intersection collisions that cause a large number of fatalities in each year. Some of the major scenarios of fatal traffic accidents are a head-on collision and a right-turn collision at an intersection. The performance of the MM-SA is evaluated, comparing with the CSMA/CA by simulation, in case of those collision scenarios based on the ASV requirements.

3.2 Collision Scenarios and ASV requirements

Fig. 6 (1) shows the head-on collision scenario. A vehicle V1 is going into an intersection and a vehicle V2 is going straight through the intersection. In this scenario, the driver of V1 can not recognize V2, which is blinded by a building at the corner. *Fig. 6 (2)* shows the right-turn collision scenario. In Japan, a vehicle must keep left on roads. A vehicle V1 is going to make a right-turn at an intersection and a vehicle V2 is going straight and through the intersection. In this scenario, the driver of V1 can not recognize V2, blinded by a large vehicle such as bus waiting to right-turn on the opposite direction.

In both collision scenarios, ASV requires that the message of V2 has to be received at V1 with larger than 80% of packet delivery ratio within 100ms from any location of V2, in the service area (Tx side service area) in the figures.

3.3 Simulation Conditions

Fig. 7 shows the simulation conditions. In the simulation, four frequency channels (4.096 MHz bandwidth for each) in 5.8 GHz band are used for channel allocation to roads. Locations of all the vehicles, other than V1 and V2, are determined randomly. Total number of vehicles around an intersection is 88. The inter-vehicle distance is 19.4 meter on average, assuming vehicles'speed of 70 km/h. Each vehicle periodically generates status message of 140 bytes with the period of 100 msec. It is assumed that there are walls along with both road sides, which represent buildings.

By simulations using Qualnet network simulator, the received power characteristic, the end-to-end packet delivery ratio and transmission delay from the vehicle V2 are measured at V1.

3.4 Received power characteristic in case of non-line-of-sight

Fig. 8 shows received power characteristic in head-on collision scenario, which is the case of non line-of-sight, by a simulation. In the simulation, ray-tracing is used for radio propagation model. The graph shows the

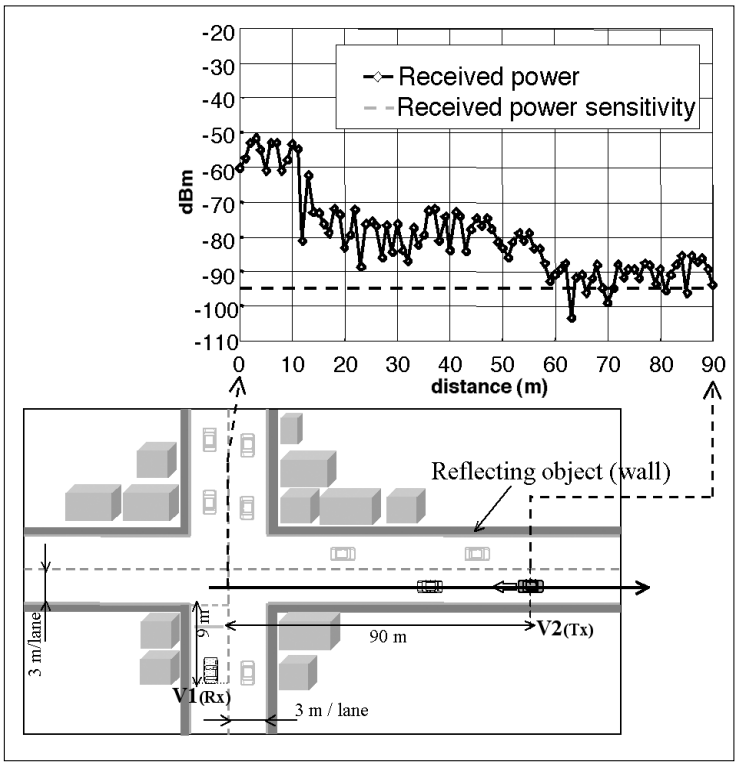


Figure 8.
Received Power Characteristic

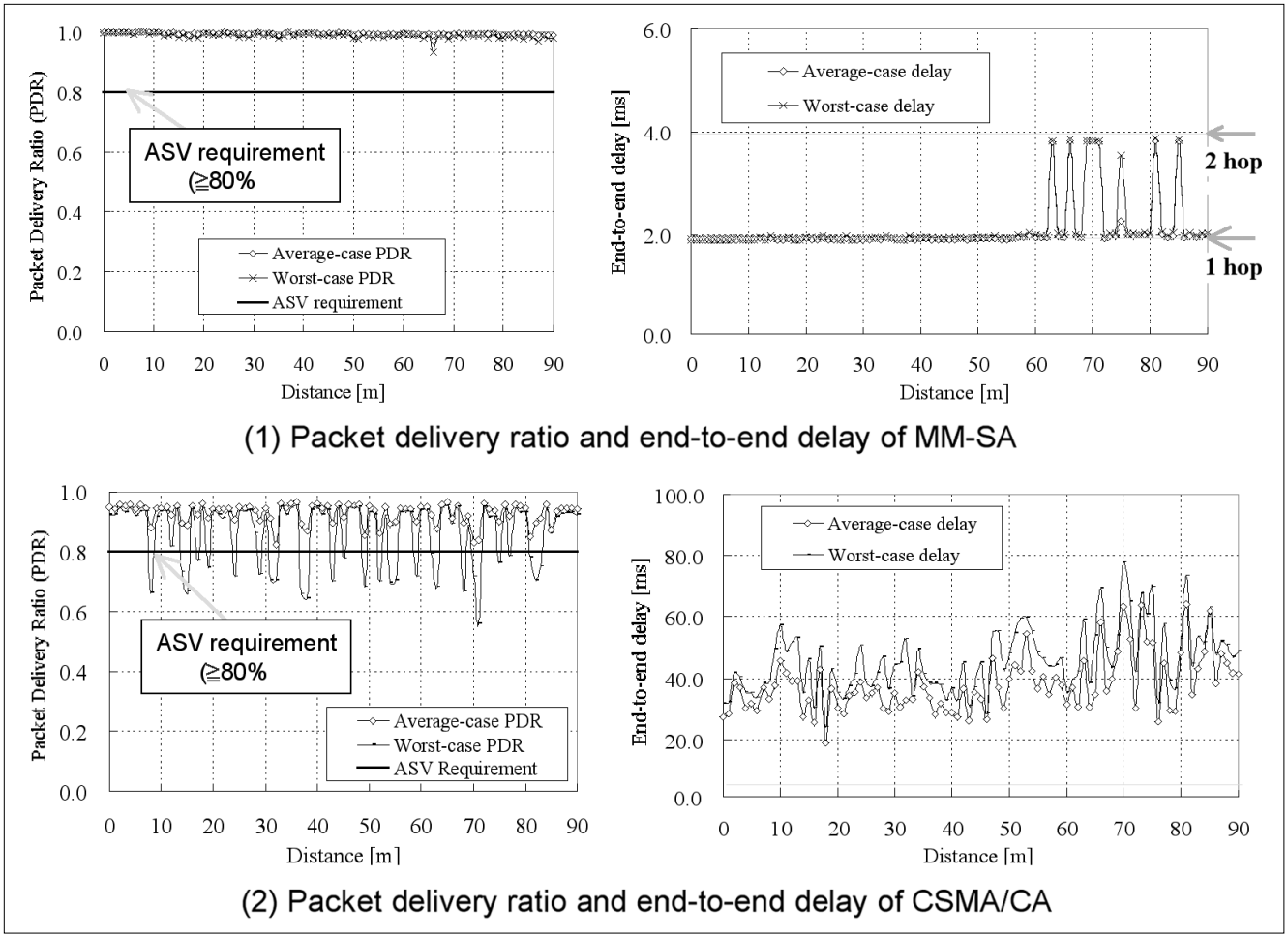
estimated signal strength of V2 received at V1, where the horizontal axis is location of V2 in the service area. In the figure, signal strength of V2 is partially below the sensitivity level. Obviously, neither MM-SA nor CSMA/CA schemes can satisfy the ASV requirements. This results that the packet forwarding is necessary, so that a vehicles' message can be disseminated over the service area.

3.5 Simulation results

Fig. 9 shows the packet delivery ratio and the end-to-end transmission delay in case of head-on collision avoidance. For fair comparisons, the frequency channel assignment rule and packet forwarding scheme of the MM-SA are also applied to CSMA/CA. The results show that in 88-vehicle scenario, the MM-SA achieves approximately 100% of packet delivery ratio and shows better performance than CSMA/CA, and also achieves significantly smaller end-to-end transmission delay (at most 4 msec). On the other hand, in CSMA/CA scheme, the end-to-end delay is much larger, taking approximately 20 to 80 ms.

Fig. 10 shows the packet delivery ratio and the end-to-end transmission delay in case of right-turn collision avoidance. The MM-SA also shows better performance.

Figure 9.
Simulation results in case of head-on collision avoidance



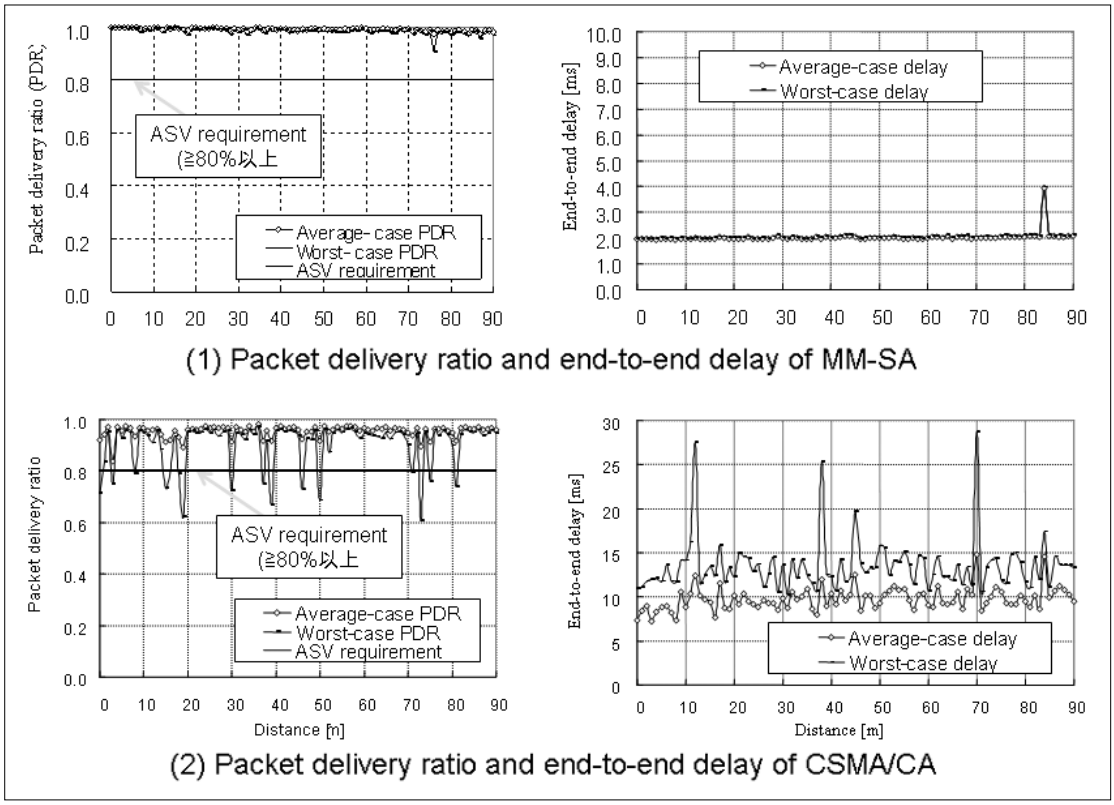


Figure 10. Simulation results in case of right-turn collision avoidance

4. Prototyping and field experiments

4.1 Prototype System

In addition to the evaluation of MM-SA scheme by simulation above, we have developed the prototype system shown in *Fig. 11*. *Table 2* shows the specification of the system. All the core technologies of MM-SA such as spread spectrum, frequency control, packet forwarding and transmission scheduling are implemented in the system. The system achieves the forwarding delay inside the system below 1 msec.

Figure 11. Prototype systems of MM-SA onboard unit

Table 2. Specification of Prototype Systems of MM-SA

Parameter	Specification
Frequency	5780, 5790, 5820, 5830MHz
Chip Rate	2.048Mcps
Bit Rate	585kbps
Spreading Factor	7
Modulation	π/4-shift QPSK
Data Detection	differential detection

4.2 Field experiments

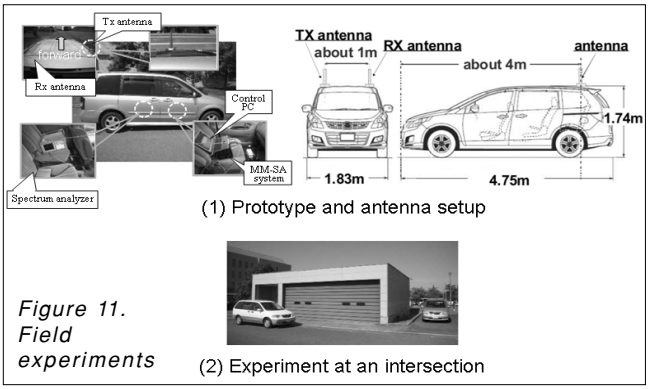
We are carrying out the field experiment, in order to evaluate the system performance in actual environment as shown in *Fig. 12*. Communication performance usually depends on the radio propagation environment, which significantly changes due to the surrounding obstacles and weather conditions. In this experiment, we measured the radio propagation and packet error rate characteristics under the non line-of-sight condition and the line-of-sight condition. The obtained data will be reflected to the computer simulation to improve its accuracy and reliability.

5. Conclusion

For inter-vehicle communications to assist safe driving, we have developed a novel scheme “MM-SA,” which is based on CDMA technology, achieving fast response in millisecond order and highly reliable transmission of vehicle’s status information.

We evaluated the effectiveness of MM-SA scheme by computer simulation based on Japanese ASV requirements, comparing with the existing and popularly used media access control, CSMA/CA. CSMA/CA suffers from performance degradation in terms of transmission delay and packet delivery ratio, when the number of vehicles increases and can not satisfy the ASV requirements. We also implemented the prototype system based on MM-SA and are carrying out the field experiment, to evaluate the system performance in actual environment.

In order to further make sure the effectiveness of MM-SA scheme, we plan to do the followings:



(1) Prototype and antenna setup

(2) Experiment at an intersection

Figure 11. Field experiments

(1) Field experiment on actual roads.
(2) Further performance evaluation in scalability (in larger numbers of vehicles) and in the other accident scenarios, comparing with the performance of IEEE 802.11p.
(3) Comparison with the other channel access schemes, e.g. timing synchronized CSMA and Distributed TDMA etc.

From the evaluations so far, we are convinced that MM-SA scheme is quite suitable for safe driving support by inter-vehicle communications.

Acknowledgments

This work was supported by the National Institute of Information and Communication Technology (NICT), Japan.

References

[1] Motegi S., Horiuchi H.,
"Relay control for data dissemination of spontaneous vehicular networks,"
IEEE ITS Telecommunications, pp.1098–1101, 2006.
[2] Kremer K.,
"Vehicle density and communication load estimation in mobile radio local area networks (MR-LANs),"
IEEE VTC'92, pp.698–704, 1992.
[3] Goldsmith A.,
"Wireless communications," Cambride UniP, 2005.
[4] Tonguz O.K., Wisitpongphan N., Parikh J.S., Bai F., Midalige P., Sadekar V.K.,
"On the broadcast storm problem in ad hoc wireless networks," IEEE BROADNETS, pp.1–11, 2006.
[5] Briesemeister L., Schafers L., Hommel G.,
"Disseminating messages among highly mobole hosts based on inter-vehicle communication,"
IEEE VTC 2005, Vol.1, pp.423–428, June 2006.
[6] Campelli L., Cesana M., Fracchia R.,
"Directional broadcast forwarding of alarm messages in VANETs,"
IEEE WONS2007, pp.72–79, January 2007.
[7] Yu S., Cho G.,
"A selective flooding method for propagating emergency messages in vehicle safety communications," IEEE ICHIT2006.
[8] Korkmaz G., Ekici E., Ozguner F.,
"An efficient fully ad-hoc multi-hop broadcast protocol for inter-vehicular communication systems,"
IEEE ICC2006, pp.423–428, June 2006.
[9] ITS Info-communications Forum,
"Guideline of the inter-vehicle communications system using 5.8 GHz: RC-005,"
http://www.itsforum.gr.jp/Public/J7Database/index.html
[10] ANSI/IEEE Std 802.11,
"Wireless LAN mediun access control (MAC) and physical layer (PHY) specifications (1999)."
[11] Glodsmith A.
"Wireless communications," Cambride UniP, 2005.
[12] Shagdar O., Ohyama T., Shirazi M.N., Yomo H., Miura R., Obana S.,
"Safety Driving Support using CDMA Inter-Vehicle Communications," The Journal of Inf. Proc., Vol.18, pp.1–15, January 2010.
[13] Yomo H., Shagdar O., Ohyama T., Miamoto M., Condo Y., Hasegawa J. Sakai T., Miura R., Obana S.,
"Development of a CDMA Inter-vehicle Communications System for Driving Safety Support,"
IEEE Wireless Communications, Vol. 16, No. 6, December 2009.
[14] Shagdar O., Ohyama T., Shirazi M.N., Tang S., Suzuki R., Miura R., Obana S.,
"Message Dissemination in Inter-Vehicle CDMA Networks for Safety Driving Support,"
IEEE VTC2009, April 2009.
[15] Yomo H., Miyamoto M., Shagdar O., Ohyama T., Shirazi M.N., Miura R., Obana S.,
"Supporting Safety Driving with Inter-Vehicle CDMA Networks under Realistic Accident Scenarios,"
IEEE ICC, Vehicular Networking & Applications Workshop, June 2009.
[16] Yomo H., Miyamoto M., Shagdar O., Ohyama T., Shirazi M.N., Miura R., Obana S.,
"Performance of CDMA and CSMA Based Inter-Vehicle Networks for Safe Driving,"
ITS World Congress 2009, September 2009.
[17] Shirazi M.N., Tang S., Shagdar O., Suzuki R., Obana S.,
"Performance Evaluation of Safety Applications over CDMA Vehicular Ad Hoc Networks,"
ITS World Congress, November 2008.
[18] Toh C.K.,
"Ad Hoc Mobile Wireless Networks – Protocols and Systems," PRENTICE HALL, 2002.
[19] Qualnet simulator,
http://www.scalable-networks.com
[20] Andrews J.F., Weber S., Haenggi M.,
"Ad Hoc networks: to spread or not to spread ?,"
IEEE Comm. Magazine, Vol.4, pp.84–91, 2007.
[21] Sze-Yao Ni, et al.,
"The Broadcast Storm Problem in a Mobile Ad Hoc Network," MOBICOM '99, pp.151–162, August 1999.
[22] ASV Programm,
http://www.mlit.go.jp/jidosha/anzen/01asv/index.html

Fast handover and fast failover mechanisms
based on cross-layer collaboration among the link layer, the network layer and the transport layer

FUMIO TERAOKA, YUNSOP HAN

Keio University/NICT, Yokohama, Japan
tera@ics.keio.ac.jp
hanstone@tera.ics.keio.ac.jp

Keywords: cross-layer collaboration, fast handover, fast failover, SCTP

This paper describes a fast handover mechanism in the network layer called L3-FHOX and a fast failover mechanism in the transport layer called SCTPfx. Both mechanisms are based on a cross-layer architecture called CEAL. CEAL enables the control information exchange between layers in a node with keeping the layering structure. L3-FHOX is an example of cross-layer collaboration between the link layer and the network layer. SCTPfx is an example of cross-layer collaboration among the link layer, the network layer and the transport layer. We implemented both mechanisms in FreeBSD. The entire handover time in L3-FHOX is approximately 10 msec plus the RTT between the mobile node and its location management server while the normal handover procedure in IPv6 takes more than 1 second. The failover time of SCTPfx is 122 usec plus the RTT between the two end nodes while the normal failover procedure in SCTP takes more than 31 seconds.

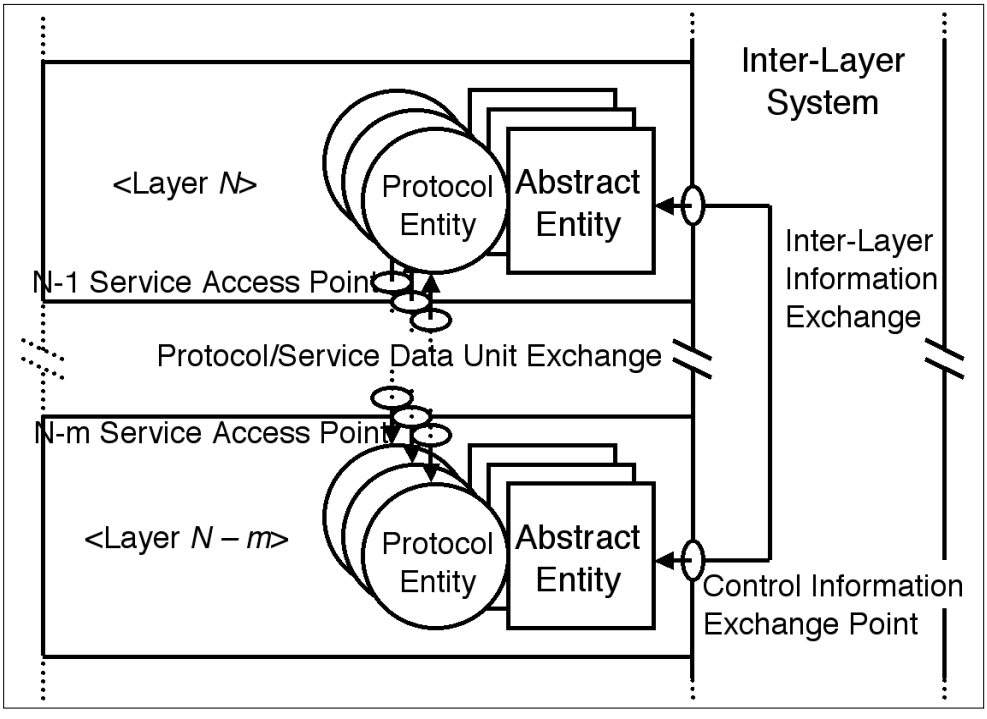
1. Introduction

The protocol layering model makes it easy to design protocols in a layer independently from other layers by concealing the details of a layer to other layers and providing abstract services to the upper layer. In some cases, however, a layer requires the information in other layers for efficient execution. Fast handover in the network layer and fast failover in the transport layer are typical examples that require the information in other layers for efficient execution. For fast handover in the network layer, the network layer requires the link layer information to predict handover and to select the most suitable access router to handover. For fast failover in the transport layer, the transport layer requires the lower layer information to know the failure event as soon as possible. Thus, the processing in a layer using other layer information is called *cross-layer collaboration* and the mechanism that enables the exchange of information between layers is called *cross-layer architecture*.

There are a lot of proposals that utilize cross-layer collaboration such as CLASS [1], ECLAIR [2], MobileMan [3], and Hydra [4]. We are proposing a cross-layer architecture called *CEAL* (Cross-layer control information Exchange between Arbitrary Layers) [5] and a fast handover mechanism in the network layer called *L3-FHOX* (L3-driven Fast HandOver mechanism based on X-layer arhitecture) [6] based on CEAL. In the transport layer, we are proposing a fast failover mechanism and a fast handover mechanism in SCTP [7]: SCTPfx [8] and SCTPmx [9], respectively.

This paper describes L3-FHOX as an example of collaboration between the link layer and the network layer and SCTPfx as an example of collaboration among the link layer, the network layer, and the transport layer.

Figure 1. Interaction model between layers in CEAL



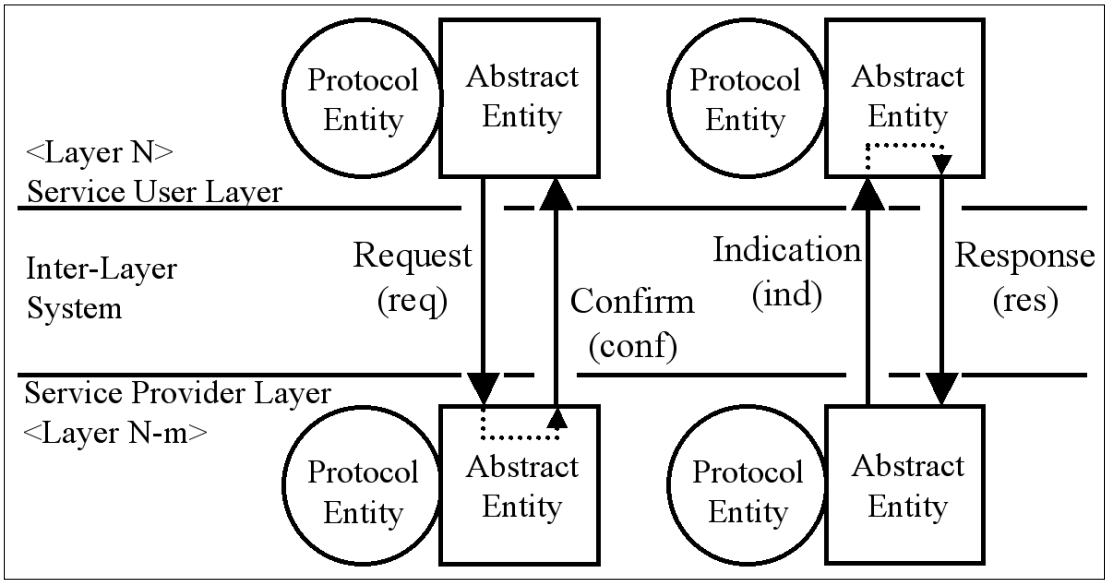


Figure 2. Four classes of primitives in CEAL

2. Overview of CEAL

In the OSI layering model, the *protocol entity* (PE) is the entity that processes a protocol. For cross-layer collaboration, the PEs in distinct layers must be able to exchange information. If a PE in a layer provides another PE in another layer with the protocol-specific information, each PE must be able to understand the information specific to all protocols. Therefore, CEAL introduces the *abstract entity* (AE) that transforms the protocol-specific information to the protocol-independent information as shown in *Fig. 1*. An AE is attached to a PE. CEAL also introduces the *inter-layer system* (ILS) that penetrates across all layers. Thus, the protocol-specific information of a PE is first transformed to the protocol-independent information by the AE attached to the PE; the protocol-independent information is carried to the AE attached to the PE that requires the information via the ILS.

In CEAL, each layer offers its services in the form of primitives. Four classes of primitives are defined as shown in *Fig. 2*. *Request* (req) is issued by the AE that wants to get the services or information from another AE, and *Confirm* (conf) is the acknowledgment of the request. *Indication* (ind) is the notification of the information to the AE that requested the service, and *Response* (res) is the acknowledgment of the indication. CEAL also defines three different usages of primitives. Type 1 is to provide the information in a layer to another layer immediately and consists of a Request and a Confirm. Type 2 is to notify another layer of events of a layer asynchronously. In Type 2, first a Request and a Confirm are exchanged; when the expected event occurs, an Indication is called. Type 3 is to control actions of a layer from another layer and consists of a Request and a Confirm.

CEAL defines nine L2 primitives as shown in *Tab. 1*. For example, *L2-PoAList* is used to acquire the list of available access points or base stations (i.e., points of attachment (PoA)). *L2-LinkStatusChanged* is used to receive a notification that the link status changed beyond the specified threshold.

3. L3-FHOX

3.1 Normal Handover Procedure in IPv6

Fig. 3 shows the normal handover procedure in IPv6. When the link status is getting worse, the link layer (L2) of the mobile node (MN) executes handover in the link layer (Fig. 3 (1)). This handover is called the L2 handover. However, the network layer (L3) of the MN does

Table 1. L2 primitives

Type	Name	Description
1	L2-LinkStatus	acquire link status
1	L2-PoAList	acquire list of points of attachment (PoA)
2	L2-LinkUp	notification that a link becomes available
2	L2-LinkDown	notification that a link becomes unavailable
2	L2-LinkStatusChanged	notification that the link quality changed beyond threshold
2	L2-PoAFound	notification that a new PoA is found
2	L2-PoALost	notification that a PoA disappeared
3	L2-LinkConnect	command to connect to specific PoA
3	L2-LinkDisconnect	command to disconnect

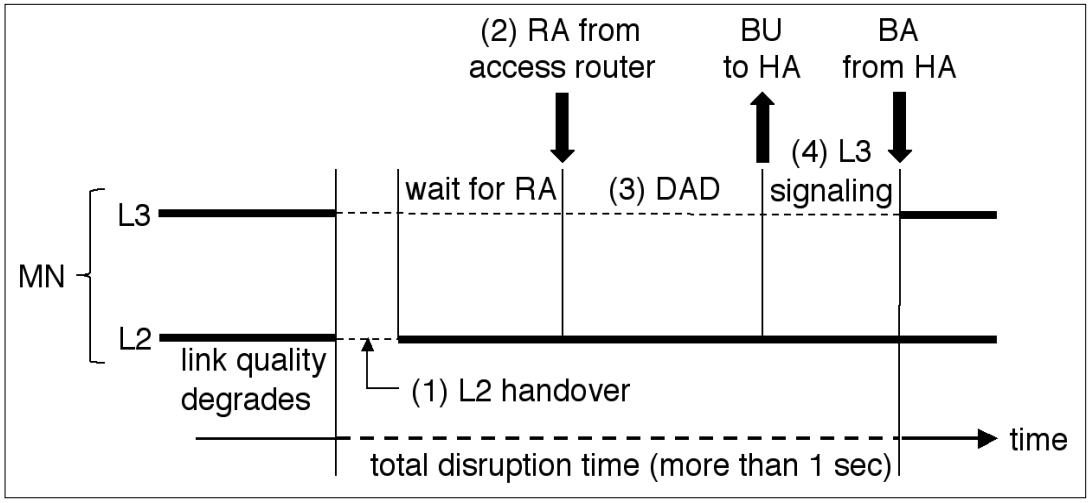


Figure 3. Conventional handover procedure

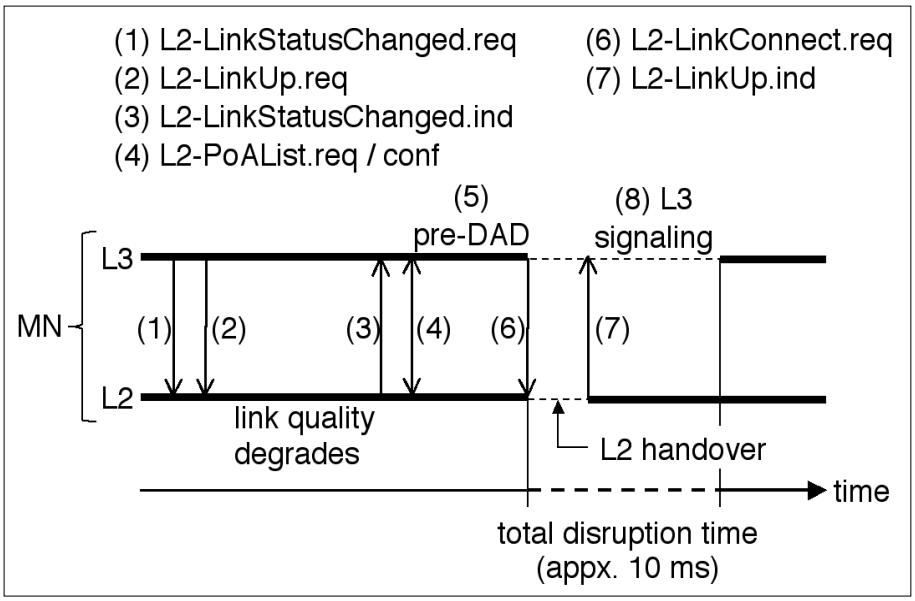
not perceive the *L2 handover*. In the network layer, the access router (AR) periodically sends the Router Advertisement message (RA). Upon receiving the RA, the L3 detects the L2 handover and starts the L3 handover procedure (Fig. 3 (2)). This wait time is one of the factors that make the entire handover time long. Next, the L3 generates a new IPv6 address and executes the duplicate address detection procedure (DAD) (Fig. 3 (3)). Since DAD makes use of timeout to detect duplicate addresses, it takes about 1 second. This is also one of the major factors that make the entire handover time long. Next, the L3 executes the L3 signaling, e.g., the exchange of the Binding Update message (BU) and the Binding Acknowledgement message (BA) between the mobile node and its home agent (HA) (Fig. 3 (4)). Thus, the entire handover takes more than 1 second.

3.2 L3-FHOX Procedure

L3-FHOX makes use of the L2 primitives defined in CEAL. *Fig. 4* shows the L3-FHOX handover procedure. First, the L3 of the MN issues *L2-LinkStatusChanged.req* and *L2-LinkUp.req* to the L2 to receive the notifications that the link status changes beyond the specified threshold and that the link becomes available again, respectively (Fig. 4 (1),(2)). When the link status is getting worse beyond the threshold, the L2 notifies the L3 of *L2-LinkStatusChanged.ind* (Fig.4 (3)). The L3 searches for the candidate access point for handover by exchanging *L2-PoAList.req/conf* (Fig. 4 (4)). Upon deciding the next access router (NAR), the L3 requests the pre-DAD to the NAR (Fig. 4 (5)). In the pre-DAD, since the NAR knows the all IPv6 addresses used in its cover area, it immediately responds to the MN. At this point, the L3 finishes the handover preparation. Next, the L3 issues *L2-LinkConnect.req* to the L2 to make the L2 start the L2 handover (Fig. 4 (6)). Upon finishing the L2 handover, the L2 notifies the L3 of *L2-LinkUp.ind* (Fig. 4 (7)). Next, the L3 starts the L3 signaling (Fig. 4 (8)).

As shown in Fig. 3, the disruption time due to the handover is the L2 handover time plus the L3 signaling time. In our experiment, the former is less than 10 msec in case of WiFi. The latter depends on the round trip time (RTT) between the MN and the HA; usually the order of 10 msec.

Figure 4. L3-FHOX handover procedure



3.3 Field Experiment of L3-FHOX

We implemented L3-FHOX in FreeBSD-5.4 and had a field experiment of L3-FHOX. On a circle road, we arranged eight access points / access routers each of which provides a distinct IPv6 subnet to the mobile node. The access point has a WiFi interface for the mobile node. The length of the circle road is approximately 1 km. The mobile node is on a car and sends real-time streaming data to the correspondent node (CN). The car runs at 40 km/h. The application used in this experiment is DVTS (Digital Video Transfer System) [10]. Although DVTS consumes approximately 35 Mbps bandwidth, we selected the half-

rate mode of DVTS due to the bandwidth limit of WiFi. LIN6 [11] is used as a mobility support network layer protocol, which is based on ID/Locator split architecture. As a result, the entire handover time is approximately 10 msec. Upon a handover, there is almost no bad effect on the play-backed movie on the CN.

4. SCTPfx

4.1 Normal Failover Procedure in SCTP

SCTP is a new transport layer protocol. It has several new features such as multihoming support and partial reliability. For multihoming support, SCTP can have multiple paths in a single association between two end nodes. Among multiple paths, SCTP uses a single path as the primary path for data communication and reserves other paths as secondary paths. If the primary path fails, e.g., due to crash of a router on the primary path, one of the secondary paths is engaged as the primary path. This procedure is called *failover*. Even if a failover occurs, the association is kept available. According to the specification of SCTP, SCTP detects path failure by five times of timeout of the ACK for the transmitted data. Similar to TCP, since SCTP employs binary back off to calculate the timeout value, it takes at least 31 seconds to detect path failure.

4.2 Fast Failover in SCTPfx

SCTPfx detects path failure as soon as possible by collaboration among the link layer, the network layer, and the transport layer. There are several possible causes of path failure. For example, the cable is accidentally unplugged from the end node, a router on the path crashes, and connectivity to the destination is lost due to routing change. Due to the page limit, this paper focuses only on the case that the cable is unplugged in an end node and the case that the connectivity is lost in the network core.

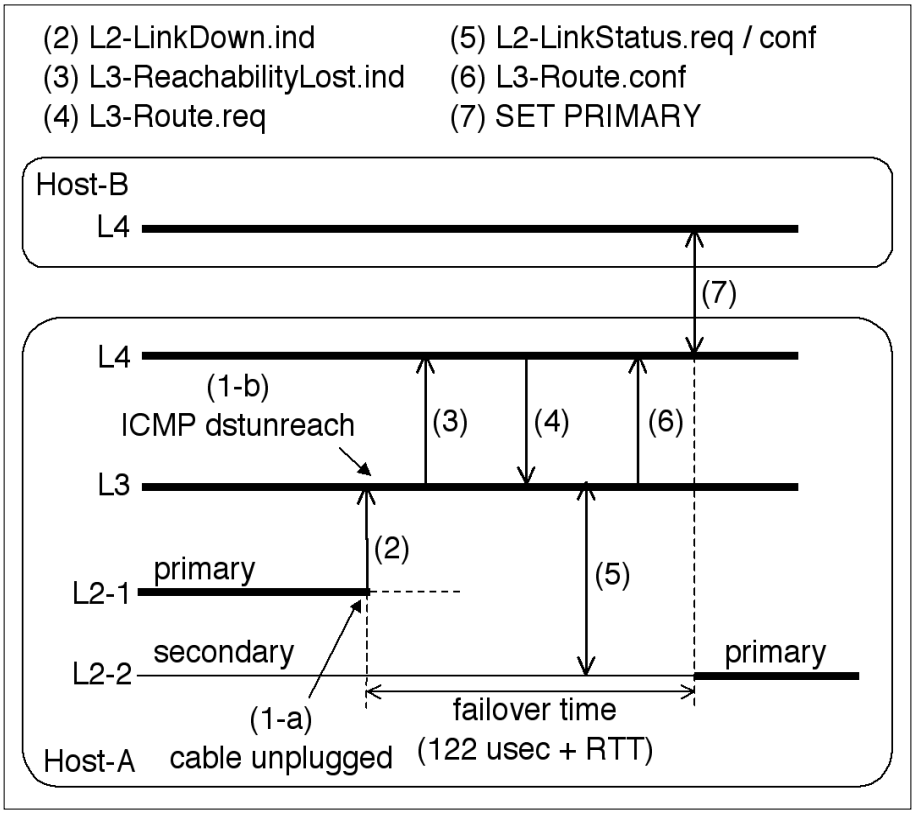
Fig. 5 shows the failover procedure in SCTPfx. In this example, the end node has two interfaces, L2-1 and L2-2. Suppose that L2-1 is the current primary path and L2-2 is the secondary path. In case that the cable is unplugged in the end node (Fig. 5 (1-a)), the L2 notifies the L3 of *L2-LinkDown.ind* (Fig. 5 (2-a)). Next, the L3 notifies the L4 of *L3-ReachabilityLost.ind* (Fig. 5 (3)). In case that the connectivity to the destination node is lost, e.g., the L3 of the end node receives the ICMP destination unreachable message (Fig. 6 (1-b)), the L3 notifies the L4 of *L3-ReachabilityLost.ind* (Fig. 5 (3)). The remaining procedure is the same for both cases. Upon receiving *L3-ReachabilityLost.ind*, the L4 (i.e., SCTP) issues *L3-Route.req* to the L3 to find an alternative route to the destination (Fig. 5 (4)). Upon receiving *L3-Route.req*, the L3 issues *L2-LinkStatus.req* to each interface and obtains the link information such as availability and the bandwidth (Fig. 5 (5)). Next, L3 selects the available routes to the destination and returns this result to the L4 by *L3-Route.conf* (Fig. 5 (6)). Upon receiving *L3-Route.conf*, the L4 selects the new primary path. Finally, the L4 of Host-A sends a *SET PRIMARY chunk* to the L4 of Host-B to switch the primary path (Fig. 5 (7)). Although SCTPfx also defines a fast recovery procedure, its description is omitted due to the page limit.

4.3 Evaluation of SCTPfx

Fig. 6 (on the next page) shows our test network. We implemented SCTPfx in the FreeBSD-6.1 kernel. We installed SCTPfx on two machines: Host-A and Host-B, both of which are IBM ThinkPad X40 with an Intel PentiumM 1.1 GHz CPU and 512 MB memory. At first, interface-1 of Host-A is the primary path. Next, we unplug the cable from interface-1, and then the fast failover procedure is executed. As a result, interface-2 becomes the new primary path.

In this environment, we measured the failover time. The result was 122 usec plus the RTT between Host-A and Host-B. This value is extremely small compared to the normal failover time (i.e., 31 seconds).

Figure 5. Failover procedure in SCTPfx



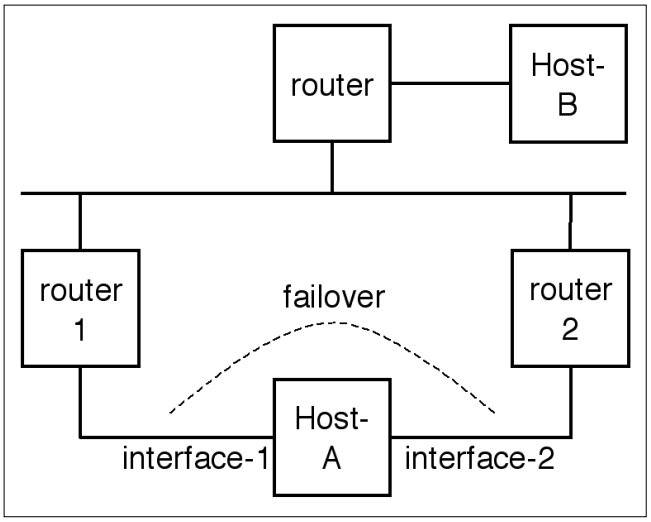


Figure 6. Test network of SCTPfx

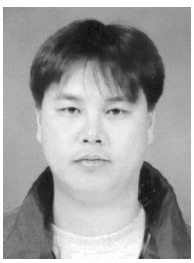
5. Conclusion

This paper shows effectiveness of cross-layer collaboration among the link layer, the network layer, and the transport layer by taking the fast handover mechanism in the network layer (L3-FHOX) and the fast failover mechanism in the transport layer (SCTPfx) as examples.

As the architecture of cross-layer collaboration, we are proposing CEAL. The current model of CEAL focuses only on cross-layer collaboration within a node. We plan to extend the current CEAL model so that it can deal with cross-layer collaboration between nodes.

Authors

FUMIO TERAOKA received a master degree in electrical engineering and a Ph.D. in computer science from Keio University in 1984 and 1993, respectively. He joined Canon Inc. in 1984 and moved to Sony Computer Science Labs., Inc. in 1988. Since April 2001, he is a professor of Faculty of Science and Technology, Keio University. He received the Takahashi Award of JSSST (Japan Society for Software Science and Technology) and the Motooka Award in 1991 and 1993, respectively. He also received the Best Paper Award in 2000 from IPSJ (Information Processing Society Japan). His research interest covers computer network, operating system, and distributed system. He was a board member of IPSJ from 2000 to 2002. He was a board member of JSSST from 2005 to 2009. He is a member of ACM, IEEE, JSSST, IPSJ, and IEICE.



YUNSOP HAN received his M.S. and Ph.D. degrees in Computer Science from Keio University, Yokohama, Japan, in 2005 and 2009 respectively. In September 2009, he joined The Keio University as a postdoctoral researcher. His research interests include wireless networks, cross-layer optimization, congestion control and transport protocols.

References

[1] Wang, Q., Abu-Rgheff, M.A., "Cross-Layer Signaling for Next-Generation Wireless Systems," In Proc. of IEEE WCNC 2003, pp.1084–1089, March 2003.

[2] Raisinghani, V.T., Iyer, S., "Cross-Layer Feedback Architecture for Mobile Device Protocol Stacks," IEEE Communications Magazine, Vol. 44, No. 1, pp.85–92, Jan. 2006.

[3] Borgia, E., Conti, M., and Delmastro, F., "Mobileman: Design, Integration, and Experimentation of Cross-Layer Mobile Multihop Ad Hoc Networks," IEEE Communications Magazine, Vol. 44, No. 7, pp.80–85, July 2006.

[4] Choi, S.H., Perry, D.E., Nettles, S.M., " A Software Architecture for Cross-Layer Wireless Network Adaptations," In Proc. of IEEE/IFIP WICSA 2008, pp.281–284, Februar 2008.

[5] Teraoka, F., Gogo, K., Mitsuya, K., Shibui, R., Mitani, K., "Unified Layer 2 (L2) Abstractions for Layer 3 (L3)-Driven Fast Handover," RFC 5184, May 2008.

[6] Gogo, K., Shibui, R., Teraoka, F., "An L3-Driven Fast Handover Mechanism in IPv6 Mobility," In Proc. of SAINT 2006 Workshop, pp.4–7, Januar 2006.

[7] Stewart, R., "Stream Control Transmission Protocol," RFC 4960, September 2007.

[8] Han, Y., Teraoka, F., "SCTPfx: A Fast Failover Mechanism Based on Cross-Layer Architecture in SCTP Multihoming," In. Proc. of AINTEC 2008, pp.113–122, November 2008.

[9] Han, Y., Teraoka, F., "SCTPmx: A Fast Handover Mechanism Using a Single Interface Based on a Cross-Layer Architecture," IEICE Trans. on Communications, Vol. E92-B, No. 9, pp.2864–2873, September 2009.

[10] DVTS Consortium, *http://www.dvts.jp/en/index.html*

[11] Ishiyama, M., Kunishi, M., Uehara, K., Esaki, H., Teraoka, F., "LINA: A New Approach to Mobility Support in Wide Area Networks," IEICE Trans. on Communications, Vol. E84-B, No. 8, pp.2076–2086, August 2001.

Developing next generation web as collaboration media

KOJI ZETTSU, YUTAKA KIDAWARA
National Institute of Information and Communications Technology, Japan
{zettsu, kidawara}@nict.go.jp
YASUSHI KIYOKI
Faculty of Environmental Information, Keio University, Japan
kiyoki@sfc.keio.ac.jp

Keywords: Global Knowledge Grid, service-oriented architecture, knowledge web

Beyond the semantic web era, digital contents are not the key assets of the web any more, but the knowledge acquired from the web contents plays an important role for users to understand situations, make decisions and/or take actions. Considering knowledge as a key asset, it is worthwhile to renovate the traditional content-centered framework of web lifecycle so as to be a knowledge-centered one. We introduce the Global Knowledge Grid, a distributed knowledge service computing environment based on the service-oriented architecture (SOA). It facilitates data-centric collaborations of different types of knowledge services: knowledge capture, knowledge association, and knowledge provision. An evolving network of knowledge is built by interconnecting heterogeneous knowledge sites over the global knowledge grid. The web browser is not restricted to hyper-document metaphor any more, but it facilitates adaptation of knowledge delivery by controlling contents and their screen layout for proactive navigation of user's information demands. Some typical applications are demonstrated to show how well our framework works.

1. Introduction

In today's networked society, knowledge-intensive work involves communication among communities of people and the social practices that occur in a particular context. Knowledge-intensive work engages in a large number of communication, coordination, and cooperation practices that cross the boundaries of organizations, countries, cultures, and/or disciplines. Beyond the semantic web era, digital contents are not, by themselves, the key assets of the web any more. Instead, the knowledge acquired from the digital contents plays an important role for users to know what is going on, make decisions, and take actions. Considering knowledge as a key asset, it is worthwhile to renovate the traditional content-centered framework of web lifecycle as a knowledge-centered one. Our intention is to change the nature of the web as an information-based resource to a knowledge-based resource.

The Global Knowledge Grid [1,2] is an integrated infrastructure for coordinating knowledge sharing and problem solving in distributed environments. The concept of knowledge grid was originally introduced as a parallel distributed knowledge discovery and data mining (PDKDD) environment, which uses the basic functions of a grid and defines a set of additional layers to implement the functions of distributed knowledge discovery [3]. We extend the original idea of knowledge grid with particular emphasis on the following concepts.

Harnessing collective intelligence

The web provides a platform for establishing networks made up of communities of people (or organizations or other social entities) connected by social relationships such as friendship, collaboration, or information exchange based on common interests. These web-supported social networks can be regarded as virtual communities.

From a sociological perspective, knowledge is considered to be socially constructed. Social processes influence the processes of generating and applying knowledge. As a consequence, knowledge can be described not as objective truth but as what a social system considers to be true. On the web, individuals are responsible for identifying knowledge sources according to their own demands on the basis of the collective intelligence [4] to be harnessed. The unit is shifting from documents to semantically coherent units of content representing, for instance, opinions, explanations, and interpretations.

A conceptualization provides a context in which knowledge elements can be uniformly organized on the basis of a specific common understanding. A conceptualization is never universally valid, but rather, it is valid for a limited number of users committing to that conceptualization. Web-supported social networks or virtual communities manage and analyze distributed knowledge. It becomes necessary to have a global platform for sharing and processing heterogeneous knowledge repositories across community boundaries. The "architecture of participation" becomes inclusive defaults for aggregating user data and building value as a side-effect of the ordinary use of the application. The mechanism that "users select for value" retains knowledge with a high-reuse rate but discards knowledge with a low reuse rate.

Virtual organizations on grid architecture

The web-based sharing that we are concerned with is not primarily web document exchange but rather direct access to data, software services, and other resources, as is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. Because of our focus on dynamic, cross-organizational sharing, Grid architecture complements, rather than competes with, existing web-based distributed environment. The real and specific problem that underlies the grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional *virtual organizations* (VO) [5].

This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form a VO.

Service-oriented knowledge

Service orientation is widely acknowledged for its potential to revolutionize the world of computing by abstracting from all resources as services in a service-oriented architecture (SOA) [6]. The service-oriented knowledge utilities model [7] stems from the necessity of providing knowledge and processing capabilities to everybody, thus promoting the advent of a competitive knowledge-based economy. Here, the knowledge services are instantiated and assembled dynamically, so that the structure and location of the knowledge services are changing at run-time on the basis of the user needs and issues. In addition, it leads to the idea that a fundamental change will be seen in the world of knowledge base so that knowledge bases move away from the traditional relational and object models, toward the *associative model* of knowledge elements, such as graph representations and triple stores.

2. System overview

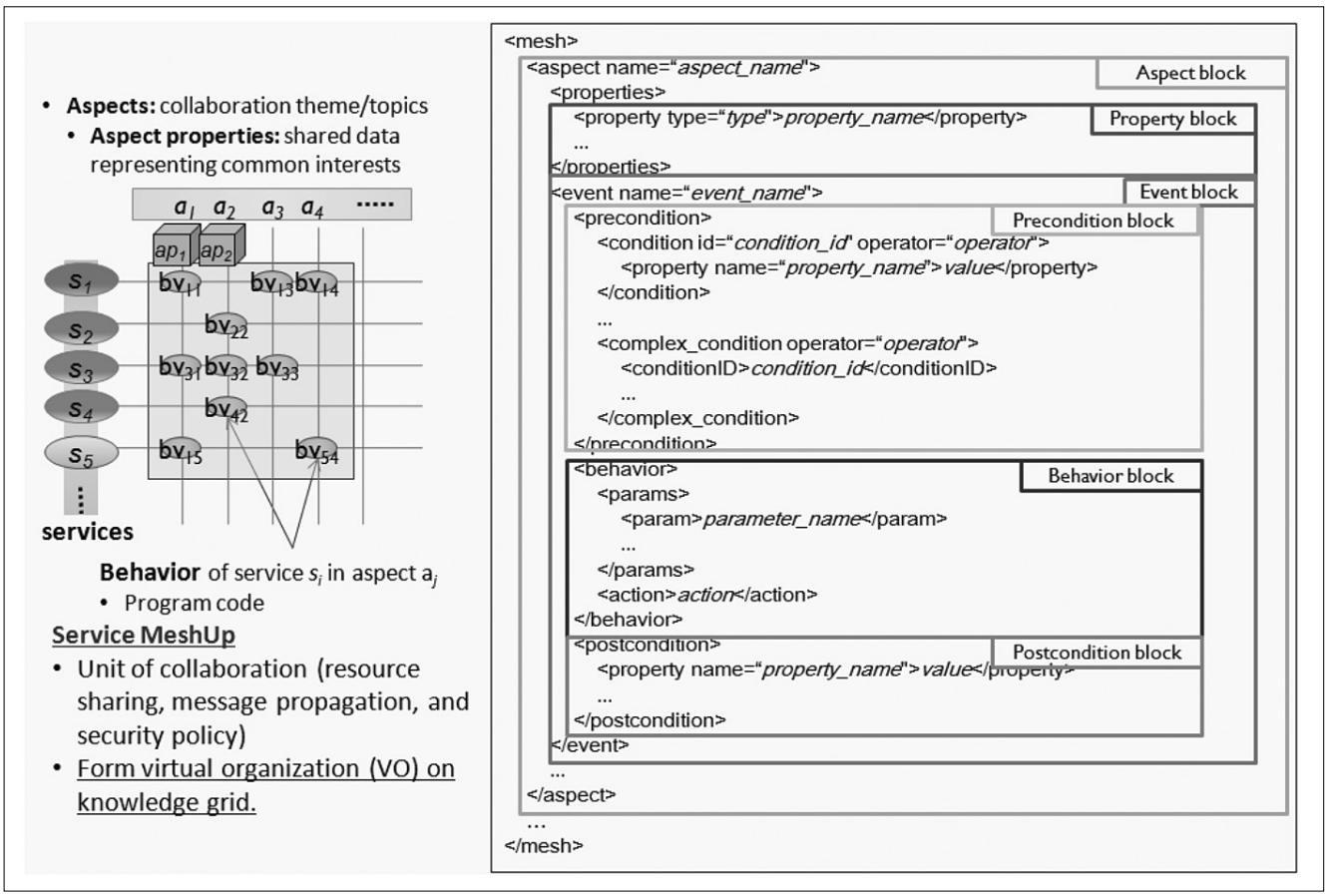
Knowledge services on the Global Knowledge Grid are classified into the following categories:

• **Knowledge Discovery Service:** It provides concept descriptions of an information source on the basis of the knowledge discovery approach by data mining, such as segmentation, classification, summarization, and ontological annotations.

• **Knowledge Association Service:** It produces various kinds of associations between multiple information sources on the basis of their concept descriptions provided by the corresponding knowledge discovery services.

• **Knowledge Delivery Service:** It is responsible for (visual) presentation and navigation of the results from knowledge discovery services and knowledge association services. It also handles user interfaces and interactions.

Figure 1. An example of Service MeshUp description



An application on the Global Knowledge Grid is realized by a specific collaboration of the abovementioned services. The Global Knowledge Grid provides the following service collaboration models:

• **Workflow model:** It designs a control flow of service execution. It is described using web service business process execution language (WS-BPEL) [8], a de facto standard for describing workflow-based service collaborations.

• **Data-centric model:** It designs the interests and behaviors of services with respect to shared data. It is described using *Service MeshUp description language*, our original description language for data-driven service collaborations.

In the rest of this section, we explain only the data-centric model. To realize the data-driven service collaborations, instead of the conventional service mashups we have proposed a novel approach called Service MeshUp.

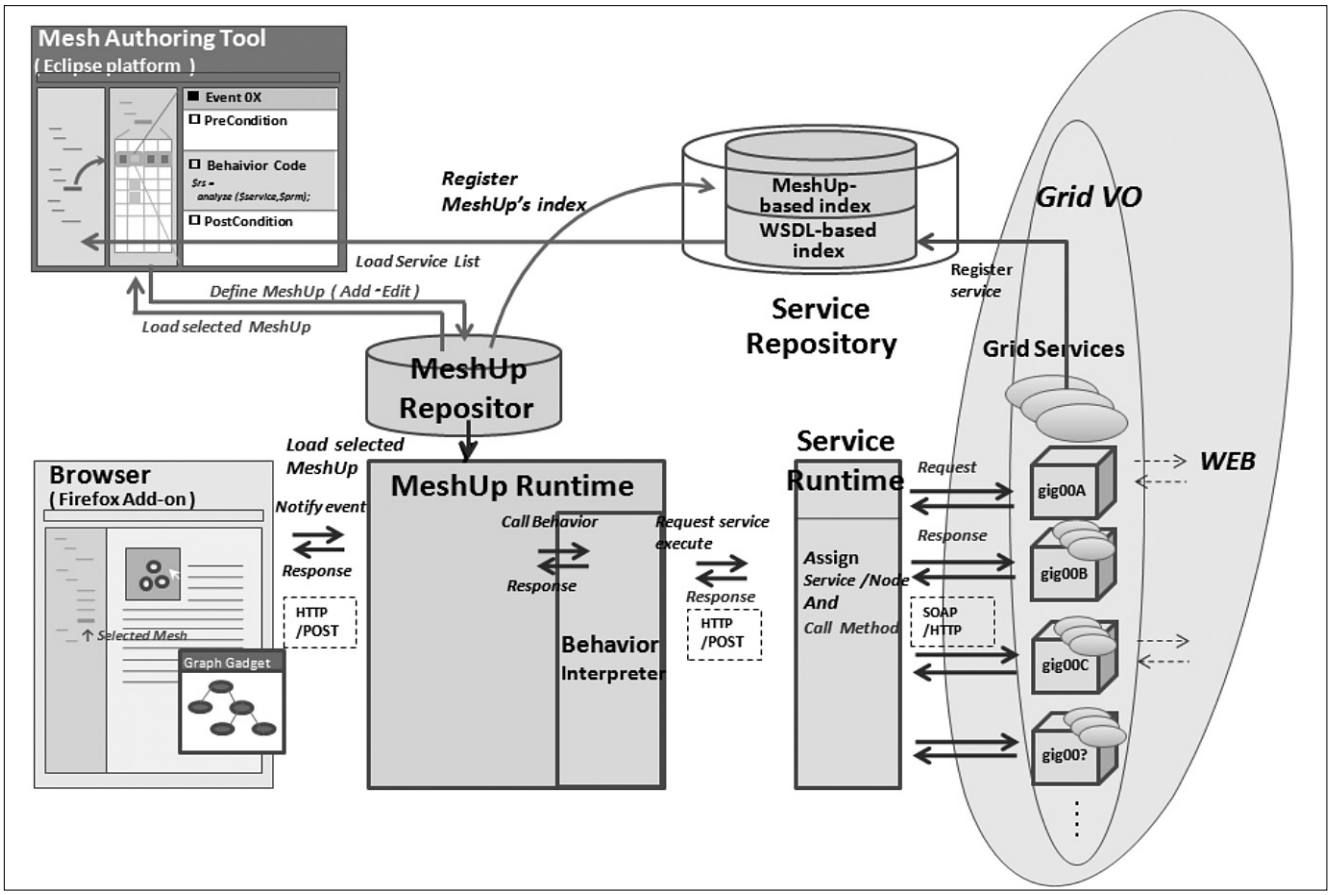
Fig. 1 shows an example of the *Service MeshUp* description. The Service MeshUp description consists of a set of *aspects*, each of which defines a single task or function in an application. Each aspect has its own properties *(aspect properties)* to be shared by the services. For each aspect, an application developer determines a set of services to be involved, and then, for each service, he/she defines *precondition, behavior,* and *postcondition* with respect to the aspect properties.

Here, the precondition describes the conditions for activating the service with respect to the aspect properties. For instance, the precondition becomes true when a specific aspect property is modified or has a specific value. Of course, Boolean composition of the conditions is also specifiable. The behavior defines what the service will do when it is activated. Basically, it invokes the functions of the corresponding service. The postcondition defines what will be done after the behavior. In most cases, the data obtained as the result of the behavior will be set to the aspect properties.

In contrast to conventional service mashup methods based on workflow model, our Service MeshUp aims at realizing service collaborations sharing "common interests," each of which is represented by the aspect properties. While both models are convertible with each other, our Service MeshUp model is more suitable for self-organization processes and ever-evolving processes such as building collective knowledge or monitoring situations than it is for transactional processes like business workflows.

The Global Knowledge Grid consists of three layers: grid layer, knowledge service layer, and application layer. In the grid layer, a computational grid network is constructed using de facto standard grid middleware Globus Toolkit [9]. In contrast with the conventional grid for building high-performance computer clusters, the Global Knowledge Grid nodes are distributed all over the world like traditional web servers. On individual grid

Figure 2. Component architecture of knowledge service layer



nodes, the knowledge services are developed independently and in parallel, like traditional web pages. The grid nodes are connected by an extensive secure network. The knowledge service layer is responsible for implementation, deployment, search, and collaboration of the knowledge services.

Fig. 2 illustrates in detail the knowledge service layer. The knowledge service layer consists of the following components:

• **Mesh runtime:** executes Service MeshUp descriptions to realize collaborations of the knowledge services.

• **Mesh repository:** stores and retrieves the Service MeshUp descriptions.

• **Service runtime:** invokes the knowledge services on local/remote grid nodes.

• **Service repository:** creates the catalog of the knowledge services, which have the mappings between service names (used in Service MeshUp description) and their end-point-references (used by Service Runtime for service invocations).

The Mesh Runtime interprets a given Service MeshUp description, and then resolves the end-point-references (EPRs) of the service names by looking up the Service Repository. According to the EPRs, the Mesh Runtime creates the VO consisting of the grid nodes where the services are deployed. The dynamically created VO provides a secure playground for sharing aspect properties and activating (i.e., invoking) the services. Once the VO is created, the Mesh Runtime initializes the aspect properties and starts the MeshUp process. While the MeshUp process is running, the knowledge services are invoked through Service Runtime.

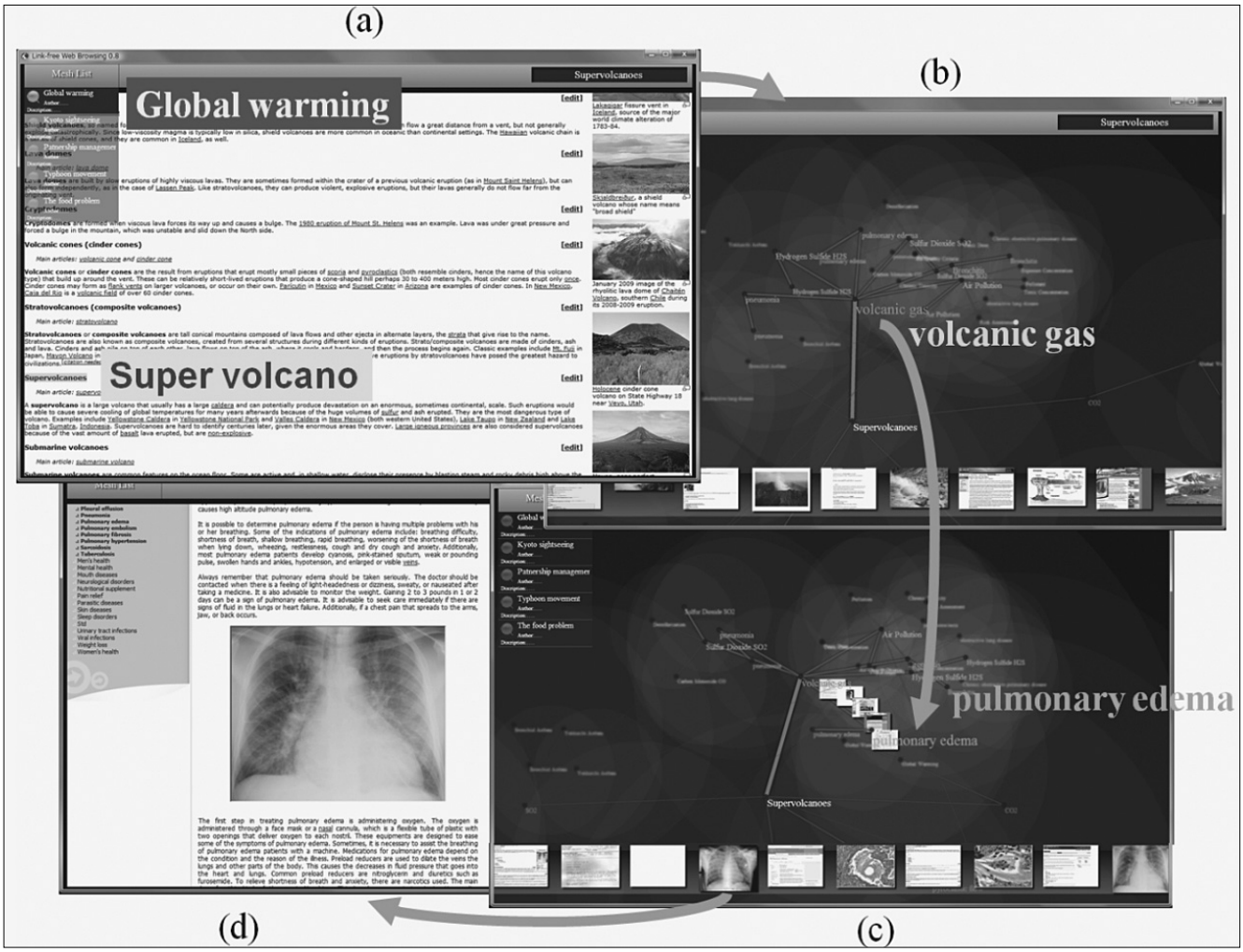
The Service Repository is managed in a decentralized manner. Once a new knowledge service is deployed to a grid node, the service information is stored in the local repository of the grid node first and then propagated to the Service Repositories on other grid nodes.

3. Application Example

Fig. 3 shows the application implemented on the Global Knowledge Grid. It allows users to browse thematically correlated web contents in a specific subject, rather than simply traversing hyperlinked web pages [10].

Here, let us suppose that when a user opens a web page to learn about volcanoes, a question ("What effect has global climate change had on us?") suddenly comes up in his/her mind. The user specifies the question to the

Figure 3. An application on Global Knowledge Grid for browsing thematically correlated web contents



browser, and then the browser shows a list of knowledge association services, each of which provides the context for leading him/her to a particular answer from the current page. Supposing that the user selects the knowledge association service "global warming," the browser first collects the relevant information from the content of current page (i.e., volcanoes) and then tries to describe the information by using, for instance, natural disaster concepts. It is done by the corresponding knowledge discovery service.

As the result, the user will discover concept terms like "volcanic gas" and "volcanic eruption". Supposing that the knowledge asset "global warming" is designed to find the possible effects to healthcare and ecosystem caused by the natural disasters related to global warming, the browser tries to discover the healthcare concepts and ecosystem concepts dependent on the natural disaster concepts, and then searches the web for information representing those concepts. The browser shows the causal relationships between the concepts using the semantic network layout, because it is considered the most direct way of visual navigation for causal relationships and is also specified to the knowledge asset "global warming" as the knowledge delivery method. It is done by the corresponding knowledge delivery service.

The user explores the causal relationships by following the edges of the semantic network. In addition, the information used for analyzing the causal relationships can be viewed by the corresponding edges. If the user clicks on the thumbnail of web information related to the concept node, the browser jumps to the web page containing the information, and then, the conventional web page browsing will be resumed.

4. Summary

Towards next generation web as collaboration media, the key concepts are:

(1) collaboration on demand with collective intelligence,
(2) grid architecture with emphasis on virtual organization, and
(3) everything as a service.

Our Global Knowledge Grid is an integrated infrastructure for coordinating knowledge sharing and problem solving in distributed environments, where three categories of knowledge services – knowledge capture, association and provision services –, collaborate on global-scale grid network. We explained the mechanisms for data-centric service collaboration, called Service Mesh Up, and dynamic VO construction to provide multi-institutional and secure playground for the service collaboration.

Currently, there are more than 400 services running on the Global Knowledge Grid. Our future work includes developing a service search engine for better performance and scalability.

References

[1] Zettsu, K., Nakanishi, T. Iwazume, M., Kidawara, Y., Kiyoki, Y., "Knowledge Cluster Systems for Knowledge Sharing, Analysis and Delivery among Remote Sites," Information Modeling and Knowledge Bases, Vol. XIX, IOS Press, pp.282–289, 2008.

[2] Zhang, R., Zettsu, K., Kidawara, Y., Kiyoki, Y., "A Decentralized Architecture for Resource Management of Group-based Distributed Systems," Journal of Frontiers of Computer Science in China (FCSC), pp.224–233, 2008.

[3] Cannataro, M., Talia, D., "The Knowledge Grid: Designing, Building, and Implementing an Architecture for Distributed Knowledge Discovery," Communications of the ACM, Vol. 46, No. 1, pp.89–93, 2003.

[4] Zettsu, K., Kiyoki, Y., "Towards Knowledge Management based on Harnessing Collective Intelligence on the Web," In Proc. of the 15th International Conference of Knowledge Engineering and Knowledge Management – Managing Knowledge in a World of Networks – (EKAW2006), Lecture Notes in Computer Science, Vol. 4248, pp.350–357, 2006.

[5] Foster, I., Kesselman, C., Tuecke, S., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," International Journal of High Performance Computing Applications, Vol. 15, No. 3, pp.200–222, 2001.

[6] Papazoglou, M.P., Georgakopoulos, D., "Service-Oriented Computing," Communications of the ACM, Vol. 46, No. 10, pp.24–28, 2003.

[7] Future of European Grids, Grids and Service-Oriented Knowledge Utilities, Next Generation Grids (NGG) Expert Group Report 3, 2006.

[8] Fu, X., Bultan, T., Su, J., "Analysis of Interacting BPEL Web Services," In Proc. of the 13th International Conference on WWW, pp.621–630, 2004.

[9] The Globus Alliance, *http://www.globus.org/*

[10] Nakanishi, T., Zettsu, K., Kidawara, Y., Kiyoki, "Approaching to Interconnection of Heterogeneous Knowledge Bases on a Knowledge Grid," In Proc. of The International Conference on Semantics, Knowledge and Grid (SKG 2008), pp.71–78, 2008.

General distributed economic framework for dynamic spectrum allocation

LÁSZLÓ TOKA, LÁSZLÓ KOVÁCS, ATTILA VIDÁCS

HSN Lab, Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics
{toka, kovacsl, vidacs}@tmit.bme.hu

Keywords: distributed spectrum management, dynamic spectrum allocation, game theory

We present our novel dynamic spectrum sharing management scheme in which the allocation and the pricing of radio frequency bands are performed in a distributed manner. We focus on a non-cooperative setting where the frequency leasers act for their own benefit, and we design the system policies in order to assure that the resulted allocation yields high spectrum utilization. We provide scalable and incentive-compatible allocation and pricing mechanisms on our physical radio interference model. Our evaluations prove that our distributed dynamic spectrum allocation scheme imposes high charges on frequency leasers that exclude others by their presence in terms of interference; therefore it is a suitable approach to reach efficient and flexible spectrum utilization.

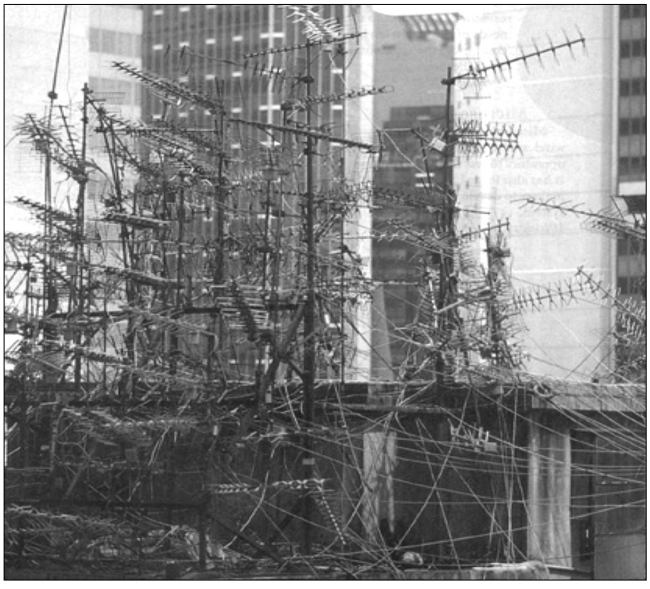
1. Introduction

Actual radio spectrum allocation is not efficient due to rigid regulation: it is access-limited (i.e., big player syndrome), and peak traffic planning and spectral re-usage restrictions cause temporal and spatial under-utilization since spectrum demands vary in time and space. Therefore the current radio spectrum allocation regulation results in sub-optimal spectrum utilization, and excludes many potential frequency exploitation opportunities. While new generation radio interfaces support flexible transmission frequencies and the convergence of telecommunications services makes actual restrictions seem out of date, recently presented dynamic spectrum allocation (DSA) solutions also lack the consideration of some key issues.

Bounding interactions among frequency leasers (e.g., noise, interference) must be taken into account with the required emphasis reflecting realistic relations, and the management framework must fulfill the basic requirements of general distribution of limited resources. We propose a distributed spectrum management framework to allocate frequencies for Wireless Service Providers (WSPs) dynamically with the goal of improving the efficiency of frequency utilization: we make the case of spatio-temporal DSA.

We build a self-organizing scheme in which the participants manage the allocation and pricing of spectrum, and the central authority only enforces our policies. The result is efficient frequency utilization while inciting the deployment of interference-tolerant technologies. Our framework takes into account the selfishness of WSPs (in the game theoretic sense) and provides a scalable allocation method.

We apply game-theoretic modeling to reflect WSP selfishness, and we build the mechanism design with



the goal of assuring desirable properties of allocating limited resources among participants. Our basic principles are the following: the overall spectrum utilization should be maximal, and in case of "conflict of interest" the frequency bands are allocated to those who "value" it the most.

2. Related work

In this section we direct the focus on papers that study the allocation and pricing aspects of DSA from the important body of research considering the management of DSA systems.

The seminal paper of Buddhikot [1] initiated a sequence of papers focusing on allocation and pricing. Their models assume a central spectrum broker that allocates spectrum licenses for short leasing times, and introduce the notion of interference *conflict graph*. The authors provide linear programming formulation of the spectrum allocation with feasibility constraints: maximal service vs. minimal interference, maximal broker revenue vs. max-min fairness. In [9] fast heuristic algorithms are proposed to perform the broker's central allocation by optimizing these metrics. The same authors give a general bidding framework in [8], where the broker strives to maximize its revenue.

Zheng's [2] introduces distributed algorithms to allocate spectrum by local coordination and collaborative sharing among users: selfishness is not taken into account. In [4] they switch to an auction-based allocation scheme in which the objective is maximizing the revenue. Their work in [12] highlights the weaknesses of the widely-employed interference modeling by pairwise conflict graph, and they show how to derive the latter from physical interference models. Zheng et al. return to auction-theory by proposing to implement the Vickrey-Clarke-Groves (VCG) mechanism [11,3,5] for spectrum allocation in [13].

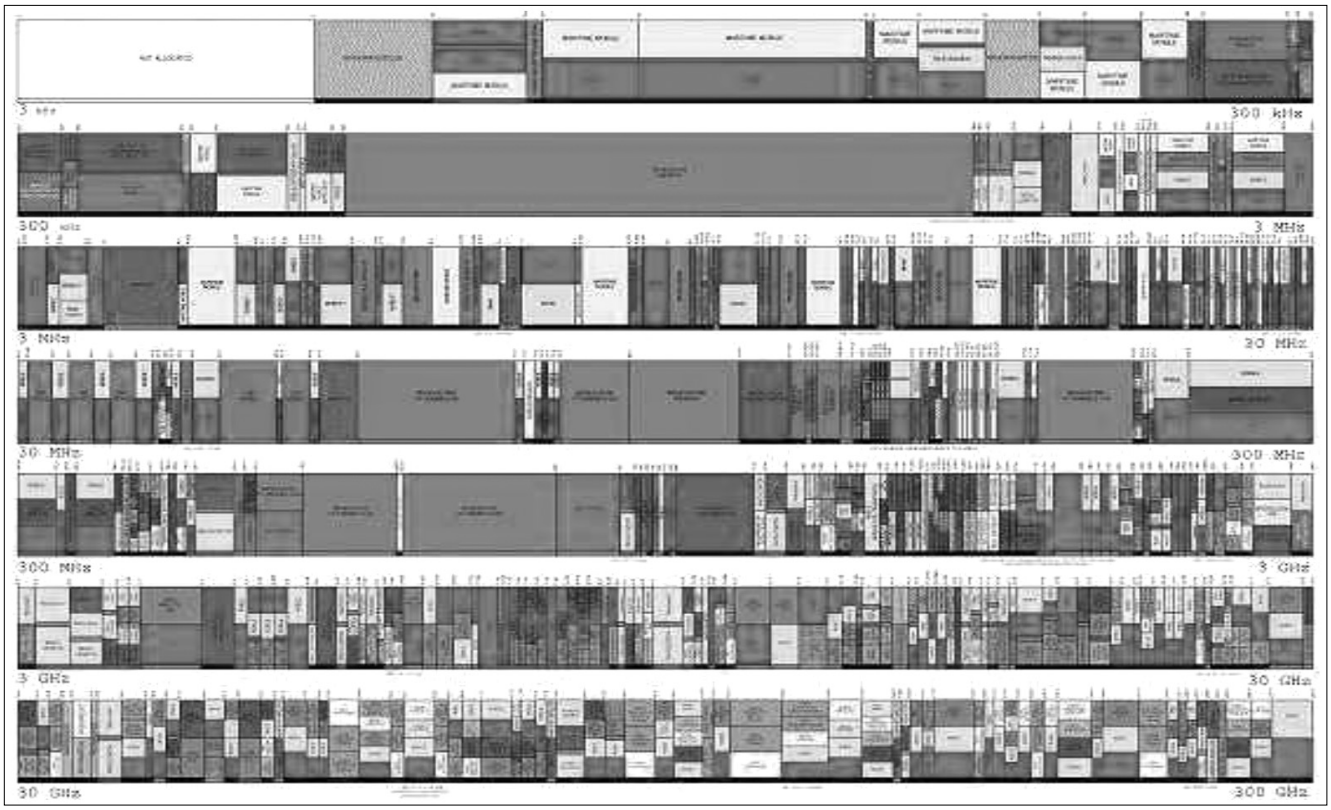
The DSA system presented in [6] also performs allocation and pricing by VCG mechanism. The authors propose a general spatio-temporal model with physical interference modeling. Frequency leasers participate in one-shot multi-bid auctions and obtain frequency usage rights for prices that maximize broker revenue or social welfare.

3. Spectrum allocation model

In our allocation framework the fairness receives new connotation, i.e., unlike the max-min fairness presented in [1,2] that assures a bit of the spectrum for every participant, in our model the one who can pay more gets the frequency band. This approach yields fairness towards the spectrum band itself because this latter is going to be exploited by the highest added utility-providing leaser. In this section we review our model's details: first, we present a simple way to describe participants' economic perception linked to spectrum usage, second, we argue on a general interference model, and at last our allocation and pricing schemes are introduced.

3.1 Node description

Our model's system participants are the possible frequency leasers that exploit radio bands within delimit-



able geographic zones, practically base stations of WSPs, called nodes. We model each node by its frequency band demand and its utility that describes its willingness to pay for acquired frequency shares. The utility is based on discount estimated incomes from the node's services. In order to model interference tolerance, we also define the "bearable" interference level for each node, i.e., the maximum cumulative interference level that the node can tolerate. Interference may occur if the same frequency is also used by other nodes, and it is defined as the maximal measured interference on the node's operating area. Details on interference are discussed in the next section.

3.2 Interference model

As the majority of related work, we assume that the radio spectrum can be divided into small, non-overlapping, homogeneous spectrum bands with pre-defined sizes. Our general physical model considers point-to-point signal attenuation formulas to take spatial and transmitting power parameters into account in order to establish the interference values among nodes (i.e., measured signal-to-interference-and-noise ratio). The central authority controls the applied radio technologies (e.g., coding), the types of radio transmitters and transmitting power levels of nodes. We refer the reader to [6] for more details on the physical interpretation of interference aspects.

Many prior works model inter-node coupling by conflict graph where nodes represent the frequency leaser base stations, and an edge exists between two given nodes if they cannot utilize the same frequency band without facing serious performance diminution due to high interference. This approach has been shown to have important drawbacks: as [12] argues, it is unable to model aggregated (cumulative) interference, moreover [6] reasons on the asymmetric nature of interference. With physical interference model, however, significant complexity is reached when optimizing spectrum allocation, which calls for scalable distributed DSA.

3.3 Distributed spectrum allocation and pricing

Distributing spectrum is harder than dividing other goods, mainly because of interference and tolerance. We introduce the notion of one-way buy-outs among nodes: if necessary because of inter-node jamming, disturbing leaser nodes can be excluded by disturbed nodes. The buy-outs are performed via auctions: nodes place bids for required frequency bands, and can bid against any actual license holder if inter-node interference overgrows their bearable limits. If multiple bidders are present for the same frequency band, a second-price (or Vickrey) auction is carried out, i.e., the highest bidder wins and pays the second bid. After nodes have made their bids to acquire necessary licenses, the winner's "second price" is divided between the former leaser and the authority.

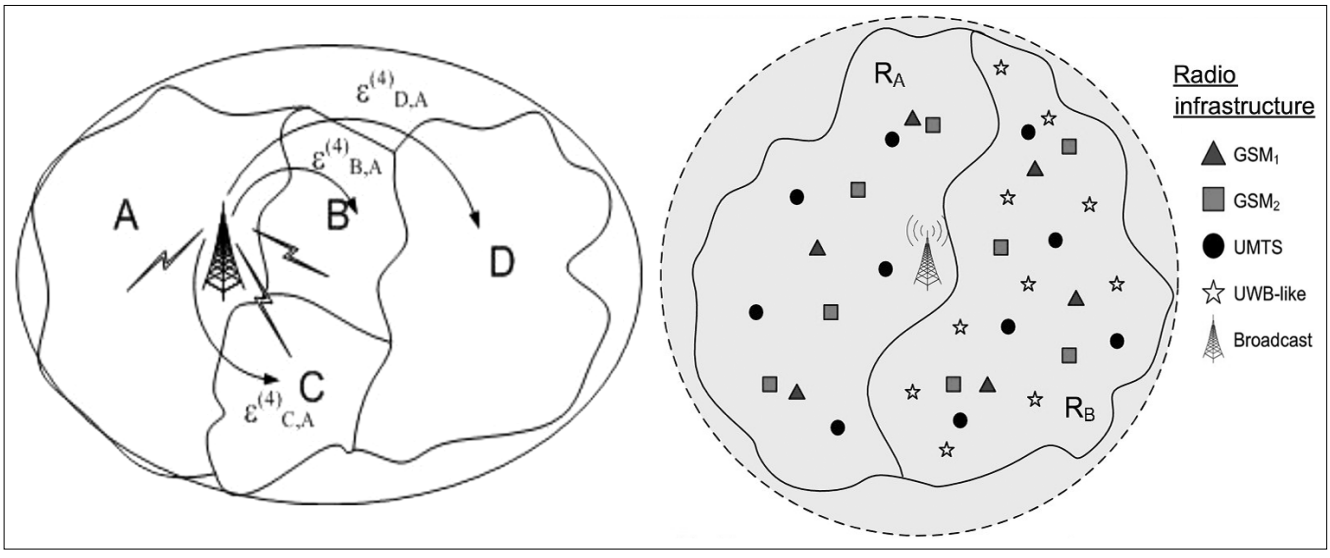
Two different types of spectrum allocation might occur:

Seamless: if a new node demands the use of a given frequency band on which its perceived interference is low, and it would not cause too much interference to other nodes. In this case, the node acquires the right to exploit the frequency, and does not have to pay anything in order to use the frequency.

Exclusive: at any allocation, if the interference experienced by any node is/becomes too high, then the disturbed node may buy out the node that causes (a part of) the interference. Exclusion happens in the form of a second-price auction: the buyer pays the second highest bid, the positive difference between the winner bid and the excluded node's prior payment, if any, is paid to the authority and the rest compensates the excluded node. Any node may voluntarily leave the spectrum by auctioning its actually leased spectrum band: the price of the frequency band is set among the bidders.

3.4 Distributed algorithm

We assume that nodes are autonomous and selfish, thus they try to maximize their payoffs. The payoff is, by definition, the *realized* valuation of the spectrum (the



difference of incomes and expenses), therefore selfish nodes strive to allocate the required size and quality frequency bands for the minimal occurring cost. The cost one might need to cover is due to the price of exclusion of other, interfering nodes. As the number of other nodes, that need to be excluded, grows, the cost increases as well.

In our algorithm, nodes perform two optimization steps *iteratively* until stable allocation is reached. At first, each node checks its allocated frequency band's size and the incurred interference against its demand and tolerance levels respectively. If no inconvenience is found, the current allocation is held. Otherwise, it positions its frequency band on the spectrum, so that, first, re-buying frequencies at which it has been previously excluded would cost the least possible, and second, the cost of exclusion of other nodes to assure that interference is kept below the required level would be minimal. Therefore the selfish strategy of each node is to buy out the cheapest interfering player set possible to assure own service quality at the cheapest frequency band.

Finding the optimal allocation of spectrum centrally is an NP hard problem in general. The main reason for this complexity is interference, therefore many approaches introduce simplifying models and apply heuristics. In our framework the allocation is optimized in a distributed way. For a brute-force exhaustive search, the algorithm requires a node to evaluate $|F|* \exp(|G|)$ values, where F stands for the number of frequency unit bands in the spectrum, and |G| denotes the cardinality of the group consisting such nodes that cause interference to it. Restricting the focus only to those nodes that cause significant interference, the algorithm complexity drops, however, we applied heuristics (greedy, simulated annealing, etc. optimization to find the cheapest band, conform to demand, with the cheapest necessary buyouts) in our simulations in order to further accelerate the algorithm.

4. Evaluation of distributed DSA

In this section we discuss the advantageous properties of the DDSA framework.

Our pricing rules incite truthful bidding of selfish nodes, i.e., nodes report their true utilities when bidding for spectrum in DDSA [10]. The well-known truthfulness property of Vickrey auctions makes bidding true utilities to be bidders' dominant strategies. A more detailed proof of incentive compatibility and truthful bidding in second-price auctions is shown in [7].

The DDSA framework supports the fairness idea behind VCG mechanisms in terms of efficient pricing. An interference-friendly node, that has high interference tolerance level and causes little interference to others, pays relatively less for the spectrum [10]. The implication of our valuation based allocation and pricing framework is that the iteration of re-allocations assures fairness despite the fact that exclusions are only unidirectional. Those nodes that cause important interference must hold high valuation because interfered nodes try to buy them out by engaging in auctions for the disturbing nodes' spectrum bands: service providers that cause high interference are punished with high costs. Also, this approach leads to an efficient spectrum allocation, where frequency bands are allocated to the most valuable leasers at the highest possible price.

Our distributed design makes the system more flexible in terms of possible re-allocation of spectrum at any time without a centrally announced or periodical auction. Furthermore, no central intelligence is needed for computing allocation and prices.

Generally, when selfish nodes' decisions drive the system, the outcome is suboptimal compared to the result of a central allocation based on full information. Since in numeric evaluations we arrived at similar outcomes (for simple simulations) to those of applying central DSA, but much faster, the distributed optimization seems efficient and scalable.

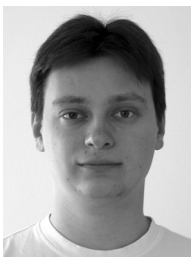
5. Conclusions

We proposed a general distributed DSA framework that offers a distributed mechanism design, well suited to practical employment issues. The applied model handles interference effects without any restricting assumptions. Through game theoretic modeling and mechanism design, we put the emphasis on the economic perspective.

We proposed distributed allocation and pricing schemes, and heuristic algorithms that provide scalable, efficient and incentive-compatible spectrum allocation.

Authors

LÁSZLÓ TOKA graduated in 2007 at Budapest University of Technology and Economics and received his MSc degree in Telecommunications on the Faculty of Electrical Engineering and Informatics. He also obtained the engineer diploma of Telecom Bretagne and Eurecom in France. He participated in the pre-doctoral education courses of the Networks and Distributed Systems Department at the University of Nice Sophia-Antipolis and received a research master degree. After graduating he enrolled as a PhD candidate at Telecom Paris and at Budapest University of Technology and Economics. His research domain is around economic modeling of distributed IT systems and networks.



LÁSZLÓ KOVÁCS was born in Pápa, Hungary, in 1981. He received the M.Sc. Degree from the Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, in 2004. From 2004 to 2007 he was a Ph.D. student at the Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics. From 2007 to 2008 he was working as a researcher at the same place. Since 2008 he is working for Ericsson Hungary Ltd. as a research fellow. His research interests are in the field of dynamic spectrum access networks and quality of service.

ATTILA VIDÁCS was born in Budapest, in 1973. He received the MSc and PhD Degrees from the Budapest University of Technology and Economics at the Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, in 1996 and 2001, respectively. During 1997, he worked as a visiting researcher at the Research and Development Center of the Nippon Telegraph and Telephone Corp., Tokyo, Japan. Currently he is Associate Professor at the Dept. of Telecommunications and Media Informatics, Budapest University of Technology and Economics. His research interests are in the field of teletraffic modeling and traffic engineering, sensor networking, and dynamic spectrum access networks.

References

[1] M.M. Buddhikot, K. Ryan,
"Spectrum management in coordinated dynamic spectrum access based cellular networks",
In Proc. of First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 8-11 November 2005, Baltimore, MD, USA, pp.299–307.

[2] L. Cao, H. Zheng,
"Distributed spectrum allocation via local bargaining",
In Proc. of Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON), 26-29 September 2005, Santa Clara, CA, USA, pp.475–486.

[3] E.H. Clarke,
Multipart pricing of public goods.
Public Choice, 11(1), September 1971.

[4] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, S. Suri,
"A general framework for wireless spectrum auctions"
In DySPAN, 17-20 April 2007, Dublin, Ireland, pp.22–33.

[5] T. Groves,
Incentives in teams.
Econometrica, 41(4):617–631, 1973.

[6] L. Kovács, A. Vidács, B. Héder,
Spectrum auction and pricing in dynamic spectrum allocation networks.
The Mediterranean Journal of Computers and Networks, Special issue on Recent Advances in Heterogeneous Cognitive Wireless Networks, 4(3):125–138, 2008.

[7] S. Sengupta, M. Chatterjee, S. Ganguly,
An economic framework for spectrum allocation and service pricing with competitive wireless service providers.
In DySPAN, 2007.

[8] A.P. Subramanian, M. Al-Ayyoub, H. Gupta, S.R. Das, M.M. Buddhikot,
Near-optimal dynamic spectrum allocation in cellular networks.
In DySPAN, 2008.

[9] A.P. Subramanian, H. Gupta, S.R. Das, M.M. Buddhikot,
Fast spectrum allocation in coordinated dynamic spectrum access based cellular networks.
In DySPAN, 2007.

[10] L. Toka, A. Vidács,
Distributed Dynamic Spectrum Management,
In 32nd International Conference of Telecommunications and Signal Processing, 2009.

[11] W. Vickrey,
Counterspeculation, auctions, and competitive sealed tenders.
The Journal of Finance, 16(1):8–37, 1961.

[12] L. Yang, L. Cao, H. Zheng,
Physical interference driven dynamic spectrum management.
In DySPAN, 2008.

[13] X. Zhou, S. Gandhi, S. Suri, H. Zheng,
eBay in the sky:
Strategy-proof wireless spectrum auctions.
In ACM MobiCom, 2008.

ETOMIC advanced network monitoring system for future Internet experimentation

István Csabai, Attila Fekete, Péter Hága, Béla Hullár, Gábor Kurucz, Sándor Laki, Péter Mátray, József Stéger, Gábor Vattay

Eötvös Loránd Univeristy, Budapest, Hungary
{csabai,fekete,haga,hullar,kurucz,laki,matray,steger,vattay}@etomic.org

Keywords: network monitoring, ETOMIC, traffic measurement, GPS

ETOMIC is a network traffic measurement platform with high precision GPS-synchronized monitoring nodes. The infrastructure is publicly available to the network research community, supporting advanced experimental techniques by providing high precision hardware equipements and a Central Management System. Researchers can deploy their own active measurement codes to perform experiments on the public Internet. Recently, the functionalities of the original system were significantly extended and new generation measurement nodes were deployed. The system now also includes well structured data repositories to archive and share raw and evaluated data. These features make ETOMIC as one of the experimental facitlities that support the design, development and validation of novel experimental techniques for the future Internet. In this paper we focus on the improved capabilities of the management system, the recent extensions of the node architecture and the accompanying database solutions.

1. Introduction

The European Traffic Observatory Measurement InfrastruCture (ETOMIC) was launched in 2004 [1,2]. It is targeted to provide the scientific community with an Internet measurement platform that is fully open and reconfigurable, extremely accurate and GPS-synchronized.

The ETOMIC system has been designed to allow researchers to perform any kind of active network measurement. The users are provided with a web-based graphical user interface for the definition of the experiments to run. Researchers may either choose from a number of built-in measurement scripts that cover the most popular measurement techniques, like traceroute or packet-pair experiments, or they can provide their own code for the experiments. To avoid conflicts in resource utilization each measurement has to be scheduled to exclusively reserve node resources for its execution. The node reservations are performed through the web-based user interface. The ETOMIC management kernel takes care of the software upload and experiment execution in a fully automated fashion.

After the successful duty of the measurement nodes since 2004 the renewal of the system components was necessary. In the Onelab project [3] we have extended the capabilities of the measurement hardware to match the current technological level and to incorporate the software evolution of the last years that are important from the perspective of network measuerments. The ETOMIC infrastructure now provides two ways of collecting experimental data. One possibility is when the researcher reserves and configures the measurement nodes and sets the parameters of the experiment through the Central Management System.

In this case, besides the original ETOMIC nodes, newly deployed enhanced measurement boxes can also be used for experimentation. To meet the requirements of high precision measurements the nodes are equipped with a DAG card (for the original nodes) or an ARGOS card (for the new generation nodes) to provide nanosecond-scale timestamping of network packets. Besides these nodes a third type of hardware compoment was also introduced, which is called Advance Probing Equipment (APE). APE is a low cost hardware solution developed to serve as a measurement agent for user applications: it provides a web service interface to conduct experiments.

This approach enables autonomic software compoments to automatically collect relevant network data from the ETOMIC system they rely on for their operation. As a consequence of a development in the system kernel the nodes of the PlanetLab platform [4] can also be used as measurement nodes by the ETOMIC system. The goal of this integration was to enable the federated usage of the high precision ETOMIC nodes and the numerous PlanetLab nodes.

To make it easier to handle and archive the huge amount of data collected by the ETOMIC platform we have created data repositories. There are two different interfaces for these data archives. The periodic measurements web interface can be used to poll automatically collected measurement data through pre-defined queries. As another approach, the Network Measurement Virtual Observatory (nmVO) [5] provides standard SQL database access to the user community. The nmVO provides a graphical user interface and a web service interface for the users to access raw and evaluated measurement.

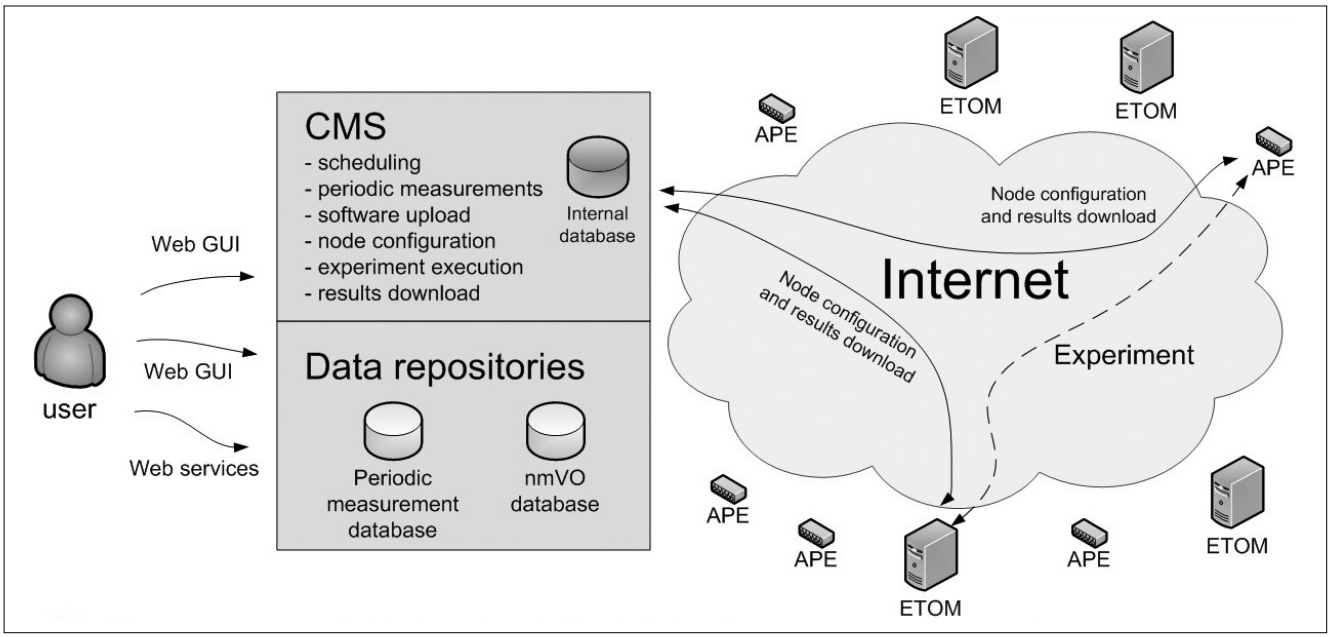


Figure 1. ETOMIC system architecture

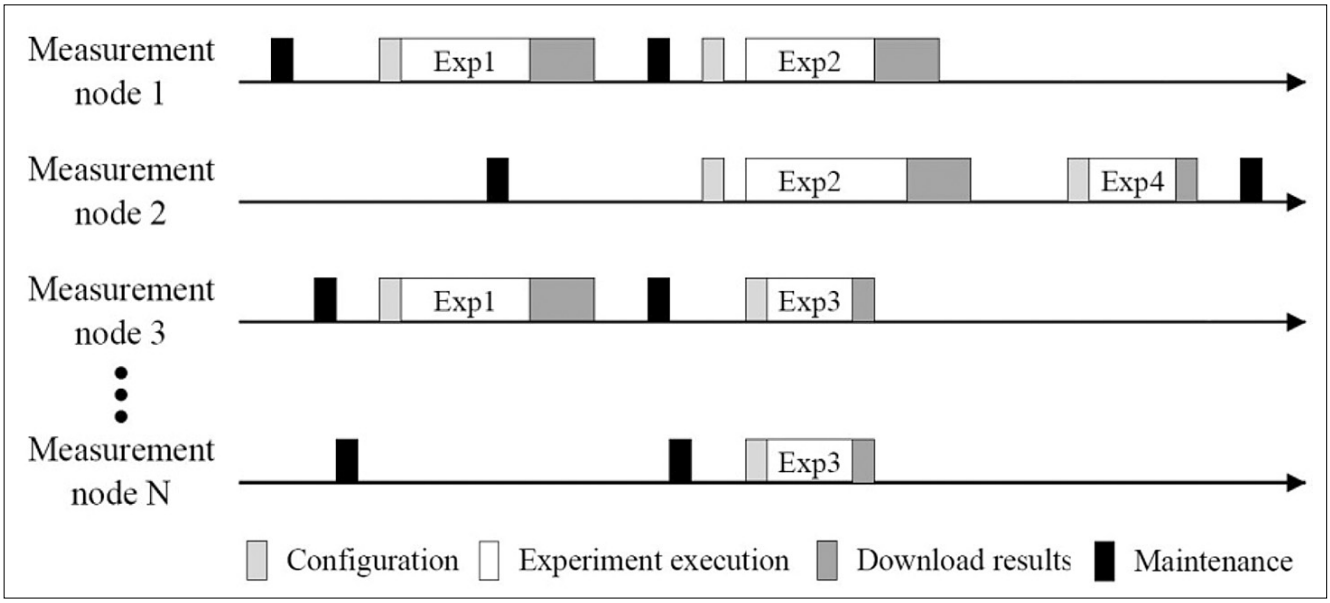
2. System architecture

The ETOMIC infrastructure is constituted of high precision measurement equipment modules placed hosted by European universities, research institutes and company laboratories. The clocks of the measurement nodes are synchronized via GPS signals, which allow not only packet round-trip time estimation, but also precise one-way delay measurements. The ETOMIC platform is very flexible, since researchers can develop and run any kind of active experiments.

A Central Management System (CMS) is in charge of system control, comprising not only the scheduling and execution of measurements experiments, but also system monitoring and configuration. The main software component of the CMS is called the *management kernel* which is running on a dedicated server computer. The kernel is responsible for scheduling tasks, deployment of user software to the measurement nodes, node configuration, experiment execution and the collection of measurement results from the nodes. The CMS provides a web-based graphical user interface where the researchers can configure the system and reserve system resources for their measurements. An internal database is attached to the kernel where the system and user level management information are stored. The results of a finished measurement are also collected and stored in this database until the user downloads them. The system components are depicted in *Fig. 1.*

As an important add-on for the original ETOMIC system the architecture has been extended with large capacity data repositories that are publicly available. The

Figure 2. Scheduling of experiments and maintenance tasks



system provides several different interfaces for these repositories through which the users can reach the collected datasets. The interfaces allow the users to run intelligent queries in order to filter and process the raw data on the server side.

3. Management kernel

3.1 Central Management System (CMS)

The ETOMIC management kernel constitutes the core of the Central Management System (CMS). It is in charge of user management, experiment scheduling and keeping the corresponding results in the temporary data storage.

In order to isolate the different measurements and to schedule experiments and maintenance tasks a calendar is used for each measurement node, as it is shown in *Fig. 2*. The management tasks can be divided into two branches: tasks that correspond to experiment definitions and tasks that correspond to the execution of the scheduled experiments.

3.2 Scheduling and calendar maintenance

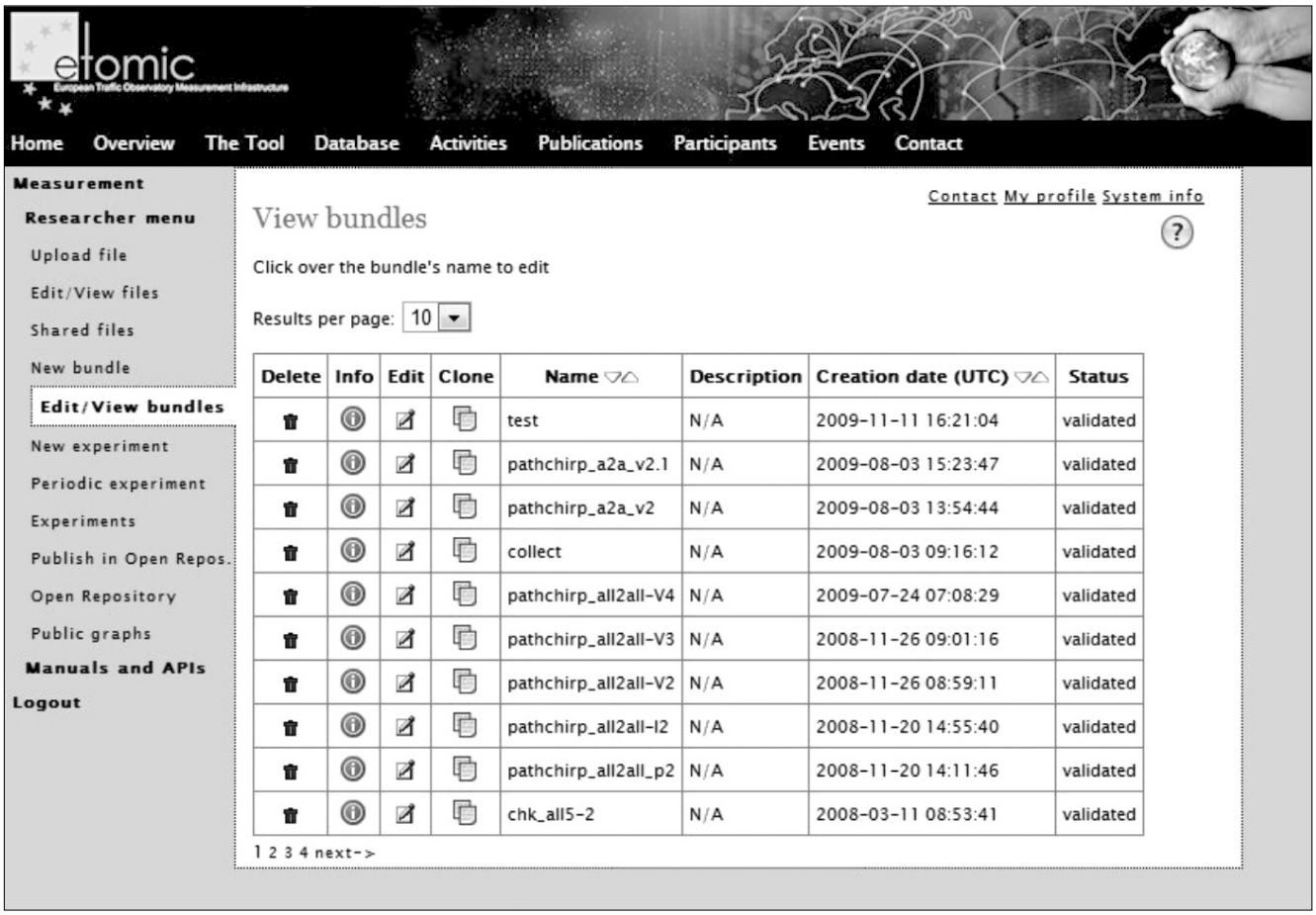
The researcher is expected to book measurement nodes for a certain time interval and to upload the applications neccessary for the measurement. The web interface is in charge of checking that the timeline for the experiment does not collide with any other previously registered measurement. In case of a successful resource reservation the CMS inserts the new experiment information into its internal database. The management kernel is continuously checking the internal database for new measurement requests. Once a new experiment has been defined and the deadline for execution approaches the management kernel uploads the measurement softwares and configures the nodes, executes the experiment and finally downloads the results to its temporary storage.

3.3 Web interface to CMS

Users are provided with a graphical interface for setting up the experiment beforehand. Then, the management kernel is in charge of experiment execution. An internal database is used to store all the necessary information to run the experiments. The stored information includes the applied softwares, external data files, the experiment results, the experiment status and the measurement node status.

Users can find manuals, the programming API and example codes here. The users can do the following operations: add a new program; upload the necessary data files; define the experiment bundle by scheduling the start and end times of the measurement; book ETOMIC time and reserve the measurement nodes; download the results; define periodic experiments with the repetition period; sharing files with other users.

Figure 3. List of experiment bundles on the web based graphical user interface



	ETOM w DAG	ETOM w ARGOS	APE	PlanetLab
platform	Intel server PC	HP server PC	Blackfin board	Variable
timestamping accuracy	60 ns	10 ns	100 ns	~10 ms
time synchronization	yes	yes	yes	no
GPS receiver	Garmin 35HVS	U-Blox LEA-4T	U-Blox LEA-4T	—
number of deployed nodes	18	20*	20*	~300**
user interface	web GUI	web GUI	web services	web GUI
				*under deployment, **under integration

*Table 1. Available measurement nodes in the ETOMIC system *under deployment, **under integration*

3.4 Integration of Planetlab's nodes

PlanetLab is a global platform for supporting the development of new network services. This platform is also used for network experiments. The nodes of the PlanetLab platform are accessed interactively via remote shell. This access method enables the CMS to use the PlanetLab nodes as its own nodes. Although the main hardware capabilities of the PlanetLab and ETOMIC nodes significantly differ, the large number of Planet Lab nodes makes them very attractive to the user community.

The capabilities of PlanetLab are not described in this paper, here we only note that the PlanetLab nodes are usually up-to-date server PCs without any hardware components specialized for network measurements. The slice based management of PlanetLab nodes allows multiple users to run experiments simultaneously in the same remote node at the same time, while the CMS takes care of the unique resource allocation. In spite of the basic differences of PlanetLab and ETOMIC the federated usage of the high precision ETOMIC nodes and the numerous PlanetLab nodes could lead to new ways of experimentation.

The software installed on ETOMIC nodes has been adapted to make the joint usage possible, using a slice of PlanetLab that is automatically renewed by the CMS. This makes the whole range of ETOMIC and PlanetLab remote nodes available through the ETOMIC web interface. The most important challenge for the integration was the synchronization of the clocks in nodes from both platforms as they use different reference signals with highly different precisions.

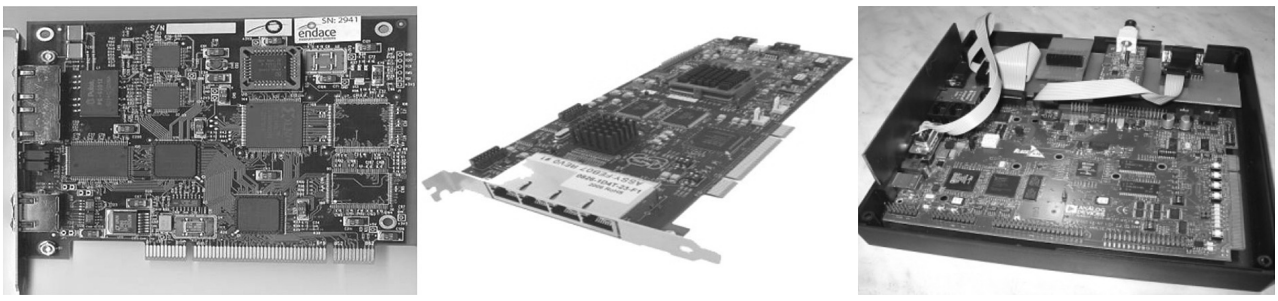
4. The measurement nodes

The nodes can be divided into two groups based on their hardware architecture. The ones that are built on server PC architecture are called *ETOM*. These nodes are accessible via the web-based graphical interface presented in Section 3.3. The ones that are based on a lightweight programmable board are called *APE*. The APE nodes are accessible via a web service interface. GPS receivers are connected to all types of measurement hardware to provide the precise time synchronization between the nodes and to provide the reference clock for the measurement cards.

Both of ETOM's and APE's hardware solutions is high-precision ones, due to the incorporated precision equipment (DAG and Argos cards) that are specifically designed to transmit packet trains with strict timing, in the range of nanoseconds.

APE is built on a development board with Blackfin processor. The board is manufactured by Analog Devices Inc. and has a number of different interfaces for hosting auxiliary hardware components that are responsible for specific network measurements tasks. Each measurement node is provided by two network interfaces:

Figure 4. Hardware components of ETOMIC responsible for the precision timestamping of the network packets. From left to right: the Endace DAG 3.6GE measurement card, the netFPGA based ARGOS measurement card and the APE node.



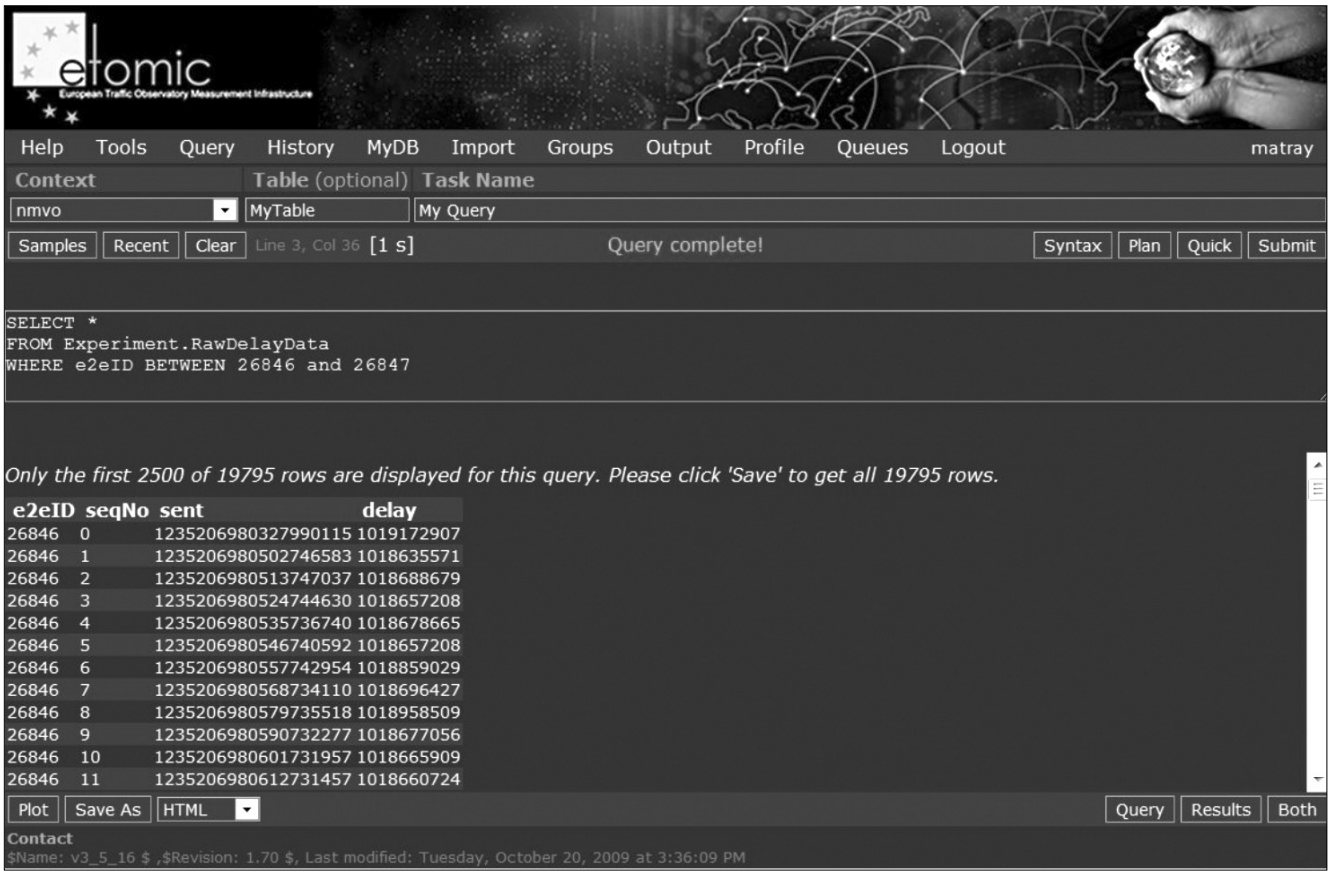


Figure 5. The web interface of the nmVO

a standard network interface card for management purposes (maintenance, software upload, data download) and an additional precision card for the probe traffic. The main feature of the APE nodes is that they provide measurement services which can be remotely called by user applications in an online fashion, without time slot reservation.

5. Accessing measurement data

5.1 Data repositories

For network measurements the collected raw data is traditionally stored in files in some standard (like traceroute dump, tcpdump) or custom formats. The files are then processed according to the research questions to be answered. Detailed analysis of complex networks requires large statistical samples. This requirement leads to substantial data size in case of measurements in high bandwidth networks, even if just a few parameters of the packets are recorded (like IP addresses, arrival time, protocol, size or delay).

Practically, measurements can produce dozens of megabytes at each monitoring node that sums up to hundreds of megabytes or even terabytes in multi-node experiment. Keeping only the results of the data analysis and discarding the raw data themselfes is not a good way to solve the data handling issues: measurement data gathered today cannot be reproduced in the future. Thus it is preferable to store the original datasets to allow further re-analysis and support the study of the long-term evolution of the network.

For these purposes we have created data repositories to store measurement data collected by means of the ETOMIC system. There are two different interfaces to reach raw and aggregated measurement data. The periodic measurements web interface can be used to poll data collected from automatic measurements in the system. The users can choose from pre-defined queries by selecting a given type of measurement, a set of nodes on which the measurement was conducted and a time frame. Long term data sets are available for one-way delay values, traceroute measurements and Paris traceroute measurements [6] data between all the ETOMIC node pairs. The results of these measurements are reachable through the periodic measurement repository of the ETOMIC website's Database / Open Repository menu [1].

5.2 Network Measurement Virtual Observatory (nmVO)

The basic concept presented in [5] is an approach to efficiently store and share research data. Beyond the simple data collecting and archiving functions it aims at providing easy-to-use analysis tools via both human and machine readable interfaces.

One of the main features of the nmVO is that it provides SQL access to the databases that are integrated under its framework, thus the users can edit and run their customized queries through either the web-based SQL interface or the web services interface. The main ad-

vantage of this solution is that the researchers can filter out the relevant information from the huge archives using server side processing. Hence, only the necessary datasets and results have to be downloaded from the server.

To sketch the nmVO principle through a possible application, consider a scenario in peer-to-peer overlay networks where management information is needed to optimize the routing between the peers. It would be unthinkable to use gzipped files for such real-time evaluation. On the contrary, the scenario is feasible if one turns to the nmVO to get the typical loss rate, the average delay on certain routes or the shortest path between the peers. This means that beyond the data itself, analysis tools are also needed to perform such data filtering and transformation queries efficiently. Using these stored procedures we can move the typical filtering and preprocessing tasks to server side.

The majority of the experimental data collected in the ETOMIC system is inserted into the data repositories and can be reached through both the nmVO web graphical user interface and the web service interface. The developed nmVO API is integrated into the APE nodes' software, so that all measurement data from the APE boxes are automatically copied into the nmVO data repository.

In addition to the historical raw data collections and the evaluated results of periodic measurements, also non-ETOMIC traceroute logs and topology data [7], one-way delay values [8], queueing delay tomography data [9], available bandwidth results [10], router interface clustering and IP geolocalization data [11] can be found in the archive.

The nmVO can be accessed through the ETOMIC website's Database / CasJobs Query Interface menu [1] and also via Web services for client applications.

6. Conclusions

In this paper we presented the enhanced ETOMIC network measurement infrastructure. We described the key components of the architecture and the new features of the Central Management System. The improved system kernel includes support for periodic measurements and the federated usage of the high precision ETOMIC nodes and the numerous PlanetLab nodes. Besides the kernel development novel hardware components have been developed and deployed. New lightweight measurement equipments have been installed that provide measurement services which can be remotely called by user applications via web services.

The system now also includes well structured data repositories to archive and share the experimental data. Periodic measurement data can be polled with customizable pre-defined queries, while the nmVO framework gives full SQL access to its archive. The recent developments make ETOMIC an easy to use experimental facitlity with versatile feature for network research.

Acknowledgements

The authors thank the partial support of the EU ICT OneLab2 Integrated Project (grant agreement No.224263), the EU ICT MOMENT Collaborative Project (grant agreement No.215225) and the National Office for Research and Technology (NAP 2005/ KCKHA005).

Authors

ISTVÁN CSABAI is a professor of physics at the Department of Complex Systems, Eötvös Loránd University. His research interest is quite wide, ranging from artificial intelligence through cancer research to cosmology. Although these disciplines are diverse, the way of handling of their inherent complexity and the necessary computational tools are often common. He has written his first paper about the dynamics of communication networks 15 years ago, recently he is working on the development of the ETOMIC network measurement system and the Network Measurement Virtual Observatory concept.

ATTILA FEKETE was born in 1975, Budapest. He received his MSc degree in physics from Eötvös Loránd University (ELTE) in 1999. After graduation he studied TCP dynamics in collaboration with ELTE Communication Networks Laboratory and Ericsson Research. Between 2003 and 2008 he worked in Collegium Budapest Institute for Advanced Studies on the modeling of complex networks. He received his PhD degree in physics from ELTE, where he is currently working as a teaching assistant, in 2009.

PÉTER HÁGA is an assistant professor at the Eötvös Loránd University. He received his PhD degree in physics in 2008. From 1999 he is a permanent member of the Communication Networks Laboratory and he is one of the founders of the ETOMIC measurement infrastructure. His research interest includes network measurement techniques, traffic modeling and adaptive protocols.

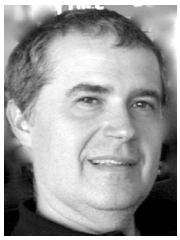
BÉLA HULLÁR received his MSc diploma in Computer Science in 2007 from the Eötvös Loránd University. He has been working in ELTE Communications Network Laboratory since then, as a PhD student. His research interests are in monitoring and measuring computer networks, developing the appropriate measurement infrastructure and its management system.

GÁBOR KURUCZ received his first MSc degree in Telecommunications and Media Informatics at 2003 from the Budapest University of Technology and Economics (BME). He earned his second diploma in Security of Information and Communication Systems at 2008 from BME. In 2006, he joined the Department of Physics of Complex Systems at ELTE. His research interests are in developing anetwork equipment with a precise timestamping module under the OneLab project.

SÁNDOR LAKI received his MSc degree in computer science from the Eötvös Loránd University, Budapest, in 2007. Currently he is working towards his PhD in computer science at the Department of Physics of Complex Systems. His primary research interest includes Internet measurement techniques, adaptive protocols and network modeling, especially IP geolocation and traffic classification.

PÉTER MÁTRAY received his M.Sc. degree in mathematics and computer science from Eötvös Loránd University, Budapest in 2005. Currently he is doing his PhD studies at the Department of Physics of Complex Systems. His attention is mainly focused on the database aspects of network measurements. Besides that, he is partly involved in active probing Internet measurements (especially network tomography) and visualization.

JÓZSEF STÉGER graduated as a biophysicist at the Faculty of Science at Eötvös Loránd University in 2001. Now works as an assistant lecturer at the Dept. of Physics of Complex Systems. He is carrying out research tasks within the study of communication networks, like the Internet. By using the measurement architecture introduced in the article he made queuing delay tomography experiments possible, which enables the group to notice congested hotspots within the network. He is an active member of international projects Moment and Onelab2.



GÁBOR VATTAY is a professor of physics of complex systems at the Faculty of Sciences of Eötvös University. He received his PhD in 1994 and his DSc title in 2003 in mathematical physics of chaotic systems. His interest is in the complex dynamics arising in natural and man-made systems. Since 1998 his interest shifted toward the dynamics of networked systems. He is the founder and director of the Communication Networks Laboratory established at Eötvös Loránd University by Ericsson Research since 2000 and the leader of the EU supported Internet measurement effort ETOMIC.

References

[1] The ETOMIC website,
http://www.etomic.org

[2] D. Morato, E. Magana, M. Izal, J. Aracil, F. Naranjo, F. Astiz, U. Alonso, I. Csabai, P. Haga, G. Simon, J. Steger, G. Vattay,
ETOMIC: A testbed for universal active and passive measurements. IEEE TRIDENTCOM 2005,
Best Testbed Award, pp.283–289, 23-25 Febr. 2005.

[3] OneLab – Future Internet Test Beds,
http://www.onelab.eu

[4] L. Peterson, T. Anderson, D. Culler, T. Roscoe,
A Blueprint for Introducing Disruptive Technology into the Internet. Workshop on Hot Topics in Networks, October 2002.

[5] P. Mátray, I. Csabai, P. Hága, J. Stéger, L. Dobos, G. Vattay,
Building a Prototype for Network Measurement Virtual Observatory. ACM SIGMETRICS – MineNet 2007,
12 June 2007, San Diego, USA.

[6] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, R. Teixeira,
Avoiding traceroute anomalies with Paris traceroute. Internet Measurement Conference, October 2006.

[7] S. Garcia-Jimenez, E. Magana, D. Morato, M. Izal,
Techniques for better alias resolution in Internet topology discovery. IFIP / IEEE IM 2009,
1-5 June 2009.

[8] A. Hernandez, E. Magana,
One-way Delay Measurement and Characterization. International Conference on Networking and Services (ICNS '07) , 2007.

[9] T. Rizzo, J. Stéger, I. Csabai, G. Vattay, P. Pollner,
High Quality Queueing Information from Accelerated Active Network Tomography. Tridentcom 2008,
Innsbruck, Austria, March 17-20, 2008.

[10] P. Hága, K. Diriczi, G. Vattay, I. Csabai,
Granular model of packet pair separation in Poissonian traffic.
Computer Networks, Vol. 51, Issue 3 , pp.683–698,
21 Februar 2007.

[11] S. Laki, P. Mátray, P. Hága, I. Csabai, G. Vattay,
A Detailed Path-latency Model for Router Geolocation.
IEEE TridentCom 2009 Conference,
6-8 April 2009, Washington D.C., USA, 2009.

A study of prosodic variability methods in a corpus-based unit selection text-to-speech system

TAMÁS GÁBOR CSAPÓ, CSABA ZAINKÓ, GÉZA NÉMETH

Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics
{csapot, zainko, nemeth}@tmit.bme.hu

Keywords: prosody variation, corpus-based TTS

This paper introduces the implementation and evaluation of a method to increase the prosodic variability of synthesized speech. Different generated prosody target versions were tested in a Hungarian corpus-based unit selection Text-To-Speech (TTS) system: the baseline prosody of the synthesizer, a rule-based prosody target and the prosody of the new method.
It is based on F0 database templates which are derived from natural sentence corpora. The corpora included that of domain specific TTS and annotated radio news. The listening test validation of the new method showed that the speech quality of the corpus-based TTS was improved. Our method was tested in a Hungarian system, and it can be extended to other European languages with fixed (e.g. Finnish) and varying (e.g. English) stress.

1. Introduction

The quality of Text-To-Speech (TTS) systems is judged on the basis of how successfully the generated synthetic speech approaches the features of human speech. The intelligibility of synthetic speech is close to that of human speech in state-of-the-art TTS systems. However, there seems to be a lack of variability in most speech synthesizers: they produce deterministically the same speech output for the same textual input, when it is repeatedly given to the system. This contradicts the variability of human speech.

The variation in human speech has been addressed by Chu et al. using a database containing two repetitions of 1000 recorded sentences in Mandarin [1]. They investigated the differences of prosody (e.g. intonation, rhythm) in the paired sentences and observed the invariant and variable parts of speech. It was measured that the two repetitions had wide variations in the mean F_0 and durations of syllables, while the meaning was the same. The rhythmic organization was more stable. The results show that the variability of human speech can be as large as half of the dynamic range of a speaker, which has to be considered in speech synthesis.

The common part of the corpus-based prosody generation approaches is that they try to associate properties (e.g. F_0) of recorded speech with the text to be synthesized. However, there are some differences in the methods and element sizes that are applied. In [2], a rule-based prosody model is complemented with a corpus-based module. In the data-driven part, the F_0 templates are as small as syllables from the corpus. [3] uses a similar method. The most important difference is the length of the F_0 templates: employing flexible-sized segments allows the modeling of both macro- and microprosody. In the corpus-based approach of [4], a linear regression statistical model produces the pitch contour of a sentence, based on word-sized items. The new feature of [5] is the use of Case-Based Reasoning. They show that a data-driven model can work with stress-group units as F_0 templates reasonably well. Besides the aforementioned corpus-based methods, some superpositional corpus-based intonation generation approaches can also be found in the literature.

In the paper of [6], three levels of intonation are derived from the speech database. Sentence-, phrase- and syllable-level prosody are hierarchically separated. [7] combines decompositional modeling with corpus-based pitch contour search. The pitch contours in the corpus are typically decomposed into phrase, accent and segmental perturbation curves. [8] introduces a corpus-based synthesis system by considering several candidate intonation contours.

The method presented in this paper uses phrase-long F_0 templates without any decomposition. Our prior work concentrated on the feasibility of a method in increasing the prosodic variability in speech synthesis [9]. As the results were rather promising, this first simple approach is further developed here. In this paper, Section 2 introduces the method that tries to mimic the variable nature of human declarative sentences in TTS systems. To pair the right pitch contour with the input text, a database of recorded speech samples is used in order to find more F_0 templates to the input. Variability is ensured by the random selection from the available intonation samples. Our hypothesis is that using this kind of speech samples, variability of human speech can be modeled in artificial systems. To prove this statement, the method is applied in a Hungarian corpus-based unit selection TTS.

In Section 3, a listening test is described that was carried out to evaluate the naturalness of the generated variable synthesized speech. The results of the test are described in Section 4, which show that the method

can generate equally natural but still different versions of sentences. The last section concludes the paper.

2. Methods

2.1 Generation of variable prosody

Fig. 1 shows the steps of our prosody generation approach used in the Hungarian system. When the system is given a raw textual input, first the sentence is partitioned into prosodic phrases. Then their syllabic and stress structures are automatically determined. Intonation is assigned in a separate step.

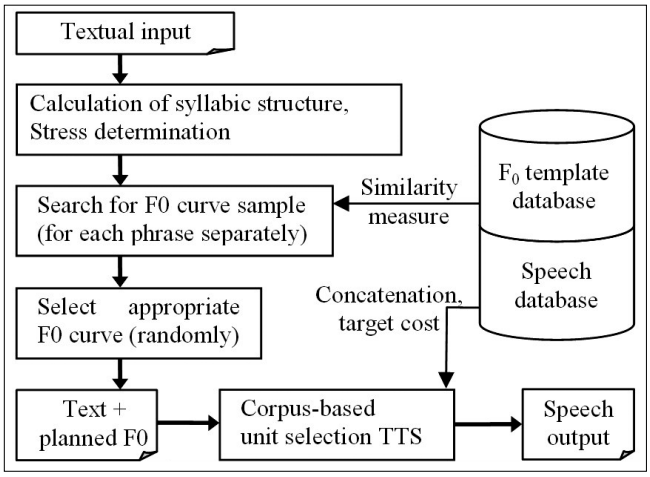


Figure 1.
Generation of variable prosody using a database of F_0 templates in a corpus-based unit selection TTS system

In this paper, the syllabic structure of a phrase is represented by the number of words in the phrase and number of syllables in each word plus their stress tags (e.g. in English a prosodic phrase might be like: "Weather warnings have been issued"; 5 words; "2+2+1+1+2" syllabic structure, plus stress marks on the words). The stress structure is specified with the rule-based method used in the Profivox TTS [10].

Based on this information target F_0 curves are searched in the database of natural sample phrases. The search is done using a similarity measure that will be discussed later. If the sample database is large and variable enough, more candidates with different F_0 curves are found. The variation of the system is realized in the next step, when one F_0 contour is selected randomly from the proposed ones. If the system is given repeatedly the same sentences or those with similar structure, this part ensures that the output speech will not always be deterministically the same, creating the desired variable prosody. The input text with the selected F_0 contour is forwarded to the corpus-selection module that tries to realize the proposed prosody during the unit selection. Note that as the system may use F_0-sample databases and speech corpora from different speakers, it is important to normalize F_0 targets, before sending them to the selection component.

2.1.1 Coverage ratio

It is not sure that an appropriate F_0 template can be found for a given input phrase. If no similar phrase is found, the rule-based prosodic model of the Profivox TTS is used. The hit rate depends on the similarity measure used and the size and variability of the database. In order to find out what degree of coverage can be reached with different similarity measures and corpora, a coverage ratio is defined. For a given input sentence, it refers to the length of the prosodic phrase for which F_0 samples are found divided by the total length of the sentence. The length is measured in number of syllables.

As an illustration in English, if the method finds a F_0 template only for the first of the two prosodic phrases in the *"A minor storm will brush the Northwest, resulting in showers."* sentence, the coverage ratio is *9/15 = 0.6*.

2.1.2 Similarity measure

Two different similarity measures are investigated in this study. They are based on the syllabic structures of the input phrases and F_0 template phrases from the database. The first one, the "exact" similarity measure means that the structures of the two phrases have to be exactly the same. The second measure, "similar" structure is less strict: the number of syllables of the longer words in the two phrases can differ in one syllable. This "loosening" in the similarity measure causes a higher coverage ratio, as discussed later. Besides the syllabic structure, the stress structure of the input and database phrases also have to be the same, in order to use the F_0 contour of the sample database.

2.1.3 Databases

The F_0 template databases were derived from several speech corpora. The textual version, phonetic transcripts, sound boundaries and measured F_0 curves were used to generate the database. The sentences were cut to phrases by the Profivox text processor [10], and for each phrase, the syllabic structures were calculated.

For each syllable, the mean F_0 was calculated. It was used when selecting an F_0 contour for the input text. The stress structure of the phrases was derived from

Table 1.
Corpora and attributes of each of them used in this study

Name	Sentences	Phrases	TTS	F_0-1	F_0-2	F_0-3
Weather	5239	13803	x	x	x	
Ph. rich	1941	3146	x		x	
Numbers	205	(205)	x			
Railway	682	1291	x		x	
Prompts	672	990	x		x	
Radio news	3651	8746				x

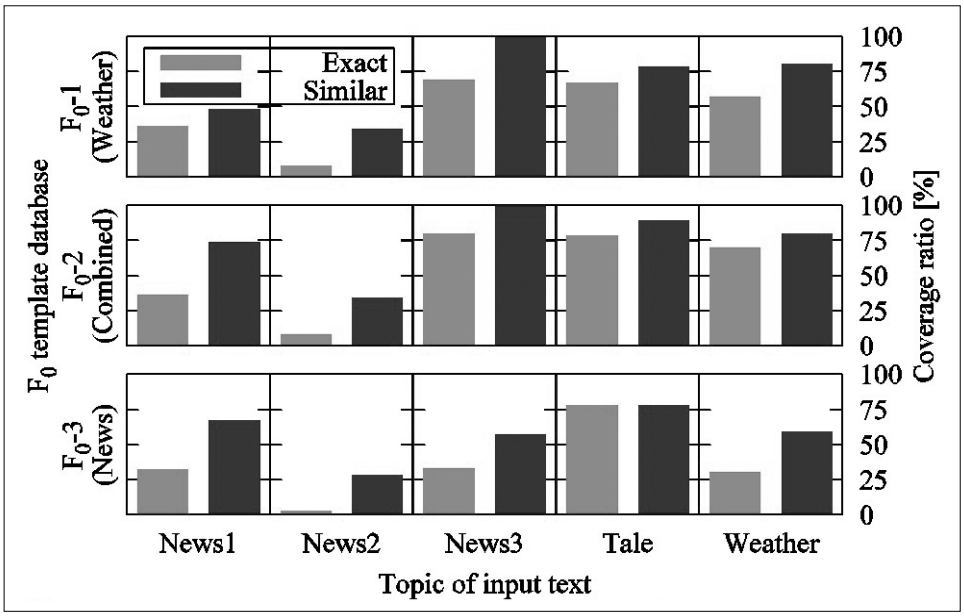


Figure 2.
Dependence of coverage ratio on similarity measure, topic of the input sentences and the domain of the F0 sample database

the textual transcripts. This solution introduces some errors because the written form does not have a one-to-one correspondence with the utterance. The development of an automatic method is in progress in order to determine the realized stress structure in the utterances of the corpora.

In this study six corpora were explored, as described in *Table 1*. The sentences of the first five corpora were read by a professional voice actress in a sound studio. The largest ("Weather") was built from weather forecast sentences [11]. The second largest corpus ("Ph. rich") contains phonetically rich and balanced sentences [12]. We also used sentences which were originally recorded for a number-to-speech synthesis application ("Numbers", [13]). The rest of it was produced for a railway station announcer system ("Railway") and other smaller, fixed inventory systems (the "Prompts" of e-mail reader systems). The last corpus ("Radio news") contains news spoken by three different announcers. Except the last one all other corpora were used in the limited domain corpus-based TTS.

Three different F_0 template databases were built for investigations of the prosody generation method: one from the "Weather" corpus itself (F_0-1), one combining four corpora of the same speaker (F_0-2) and one from the "Radio news" corpus (F_0-3).

We collected text samples in five topics: Public transport (News1), Economy (News2), Sport (News3), Tale and Weather forecasts. *Fig. 2* indicates the coverage ratios defined in 2.1.1, that can be reached with the different databases (Y axis) on input sentences from various topics (X axis). In order to illustrate which sample database is more appropriate for the variable prosody generation method, we conducted a simple test. For the different domains, we investigated up to 8 sentences to demonstrate the dependence of coverage ratios on the type of the databases.

It can be seen that the use of "F_0-2 (Combined)" database creates the highest coverage ratios, for all of the input sentences. The "Similar" similarity measure enhances the coverage ratio over "Exact", as expected. The low coverage value in "News2" sentences is due to the fact, that they contain extremely long phrases. As our method is based on the correspondence between the phrases of the input sentence and the database ones, F_0 patterns for these long input phrases were not easily obtainable.

2.2 Corpus-based unit selection TTS

The unit selection TTS that was used in our experiments is described in detail in [11]. The currently used speech databases contain sentences from several domains, as described in Section 2.1.3. The synthesizer can generate the prosody in two ways, depending on the type of the input sentence. If the sentence fits in the domain of the corpus, a simple prosody model is used, based on the relative position of words within a prosodic phrase. Because it is based on words, it will work properly only if most words of the input sentence are found in the corpus.

If the sentence is out of theme, there will not be enough whole words, which can determine the prosody. The prosody is undefined where whole words are missing. On those parts of sentences the F_0 values are determined only by the continuity criterion of the F_0 of the units, but this allows irregular prosody. In that case the F_0 generation method described in Section 2.1 is extremely useful, as it defines a target F_0 for the input sentence. In order to realize variability, always different but still natural F_0 curves are compared to the sentences with similar structure after each other. The obtained F_0 values are used in the target cost function of the TTS to follow the F_0 curve. Besides F_0 values the generated phoneme durations also are respected in target costs, but with less weight than F_0.

The words of the TTS corpus cover about 55% of the Hungarian texts [14]. It means that those words of a sentence that are out of domain are often missing on the

word level. Because the missing words are usually built from 4-5 syllables, in some cases as much as 60% of the synthesized sentence is determined by prosody rules for words that are concatenated from shorter (phone/diphone/triphone) units. In this case the overall prosody of the sentence is not determined by whole words of the corpus.

3. Experiments

Several sentences from different domains were collected in order to find out whether the TTS with the variable prosody generation method produces synthesized speech that sounds natural enough. A listening test was conducted to verify our hypothesis that the variability of human speech can be modeled using a TTS.

3.1 Test sentences

Only two sentences were chosen from each domain in order to decrease the duration of the listening test. The same domains were used as in the simple test of Section 2.1.3. We generated five versions for each sentence. The first is synthesized with a triphone based concatenative TTS (Profivox). It works with a rule based prosody module, and it can produce the waveform with the prescribed prosody. The second version is the original output of the corpus-based TTS, with its simple position-based prosody module (with minimal prosody modification). In that sentence the prosody is determined by the parameters of the units in the speech corpora. In the third version the corpus-based TTS uses the rule based prosody (of the Profivox algorithm) as the target F_0 curve. The last two versions are generated by the corpus-based TTS, but the target F_0s come from the variable prosody module which is described in Section 2.1.

During the collection of the different sentence versions we found that for one of the “News3” (Sport) sentences only a bad F_0 target was available. The end of that target is wrong, it contains an increasing end (in F_0) instead of a decreasing one. In spite of its incorrect prosody we inserted it in the test, in order to measure the tolerance of the audience for this type of error.

In the listening test we evaluated these sentences in two ways. Each sentence appeared in a Mean Opinion Score (MOS) test of the naturalness of prosody. To detect smaller differences between versions, we conducted a paired comparison test, too. In the paired comparison test, only sentences generated by the corpus-based TTS with different prosody modules were investigated. On the basis of a preliminary test the triphone-based concatenative TTS is definitely weaker than the other four versions of each sentence. The paired comparison test contained four versions of 10 sentences, making 60 pairs altogether. The MOS test contained five versions of the 10 sentences.

3.2 Listening test

A web based listening test was conducted to determine the naturalness and quality of synthetic sentences. 103 native speakers of Hungarian participated in the test with no known hearing loss. The results of 10 listeners were excluded from the evaluation because they either did not finish the test, or were found to respond randomly. Some of the excluded listeners reported playback difficulties. The remaining 93 listeners consisted of 67 male and 26 female testers having a mean age of 32 years. 49 listeners used head- or earphones while 44 testers listened to loudspeakers. The listening test took 38 minutes to complete, on average.

The test consisted of six parts. The first and the second part were used for another unrelated study. To lessen the load of the testers and to improve attention we cut in half both the MOS and the paired comparison tests. In the third and fifth parts the listeners compared the sentence pairs. In the fourth and sixth part, the sub-

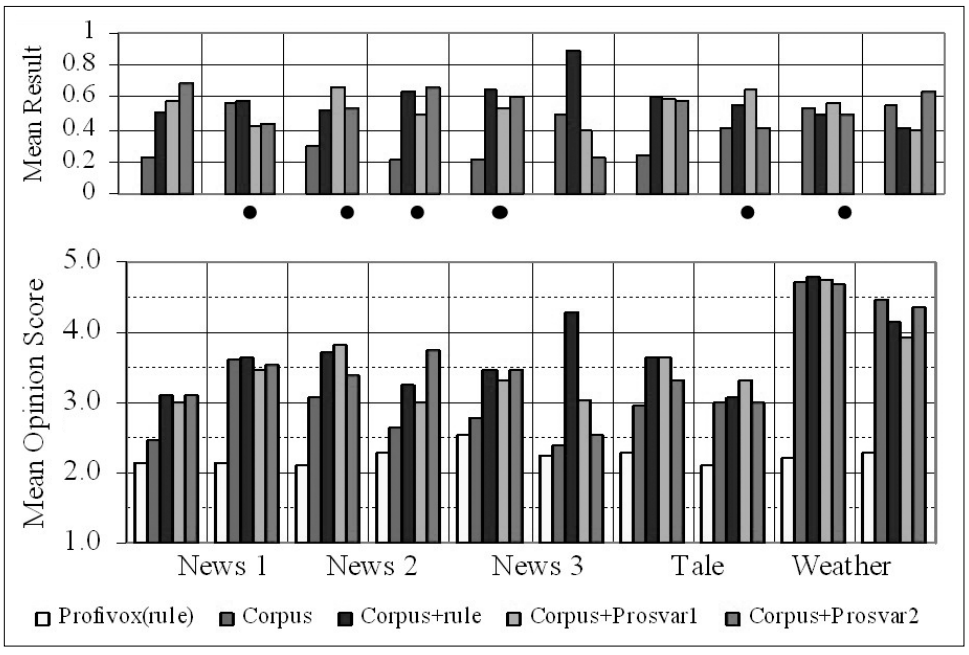


Figure 3.
Mean paired comparison results and Mean Opinion Scores of the sentences

jects used a 5-point scale to grade the quality of a sentence. The test was self-paced. The listeners had the option to replay a stimulus as many times as they wished, but they were not allowed to go back to a preceding stimulus, once they rated it. The playback order and the order of utterances in the paired comparisons were randomized individually for each listener.

4. Results

The bottom section of *Fig. 3* shows the results of the MOS test for all the 10 sentences under study. The means were calculated on the basis of the judgment of the listeners. The first column of each group shows the variants generated using a triphone rule-based TTS. It is rated between 2.1 and 2.5 MOS points. These results coincide with our earlier study [11].

The versions synthesized with the corpus-based TTS are divided into two parts: the first 8 groups were less natural than the last two groups according to the expectations. The sentences of the "News" and "Tale" area are out of domain for the synthesizer, the weather forecast sentences conform to the corpus. For "Weather" sentences changing the original prosody module of the corpus-based TTS does not show major improvements, at the second "Weather" sentence it caused rather a degradation. The worst sentences probably significantly differ from human expectations.

Except "Weather" sentences in all the other groups better results were reached with the modified, target-based prosody modules. The three versions of prosody targets (rule, Prosvar1, Prosvar2) scored nearly equally. The bars marked with dots show sentences when the "Radio-news" corpus was used as the F_0 template database for the prosody generation subsystem. These sentences were evaluated similarly to versions "Prosvar1" and "Prosvar2". In some cases, the variable prosody generation method failed to produce human-like utterances. The corpus-based TTS with the rule-based prosody method generates the best MOS score at the second sentence of the "News3" group. This sentence has a correct prosody and enough proper word units in the corpus. The "Prosvar1" and "Prosvar2" prosodies are incorrect in this case, as expected (described in Section 3.1). Subjects gave low scores, they did not accept declarative sentences with high F_0 values in the end.

The responses of the listeners in the paired comparison test were summarized in the top section of *Fig. 3*. For the two utterances in a pair, the results were calculated for each answer as follows: the more natural sentence was given a 1.0 score, the less natural one received a 0.0 score. If a listener could not hear any difference between them, a score of 0.5 were given for both variants in the pair. The averages of these values were calculated in each sentence group. The top section of *Fig. 3* shows the mean values for each variant. ANOVA tests were run for each sentence, Tukey-HSD post hoc tests showed the significant differences. In six cases of the eight non-weather sentences, the versions with prosody generated by the external module were significantly better ($p<0.05$).

Fig. 4 shows the summarized results of the paired comparison tests, averages for each F_0 generation method. The utterances generated by the corpus-based TTS without external prosody information (left column) are significantly ($p<0.05$) less natural than the other three cases, in which a target F_0 curve was given to the synthesizer. The mean results for the three different approaches of the variable prosody generation subsystem (right three columns) are not significantly different.

5. Summary and conclusions

We successfully integrated a new prosody generation method to a Hungarian corpus-based unit selection TTS system. It can provide variable prosody while increasing the quality of synthesized sentences when the input is outside the corpus domain. A listening test showed that in most cases, versions of the same sentence with different intonations were evaluated as equally natural, indicating that the variability of human speech can be applied into speech synthesis. The method can also be applied for increasing prosody variation even when the TTS works in a closed domain, but some quality degradation may occur then. Rule-based target prosody and the alternative sample-based prosody gave similar MOS values. It was found, that the method can

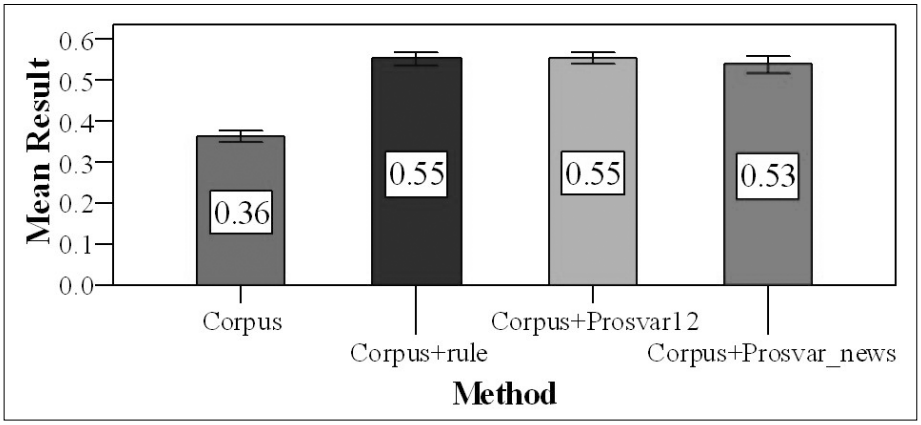


Figure 4.
Mean and 95% confidence interval of paired comparisons

work with various F_0 template databases. The use of F_0 targets from the "Radio news" did not significantly decrease the quality of synthesized speech over the "Combined" corpus.

Future work should address the construction and analysis of several F_0 template databases, in order to analyze the relationship of the realized variance in synthesized speech and the size, type and domain of speech databases. Our variable speech generation method can be extended to other languages. Languages with fixed stress (e.g. Hungarian, Finnish) are easier to handle in this system. The method can be used for languages with varying stress (e.g. English) as well, if we can determine the stress structure of the sentences based on textual input. Using other technologies (e.g. HMM) requires a different similarity measure. The results can be applied in improving the acceptability of long synthesized texts such as synthesized talking books.

Acknowledgements

The work was partially supported by the Hungarian National Office for Research and Technology (Teleauto project, OM-00102/2007) and by ETOCOM project (TÁMOP-4.2.2-08/1/KMR-2008-0007) through the Hungarian National Development Agency in the framework of Social Renewal Operative Programme supported by EU and co-financed by the European Social Fund.

Authors

CSAPÓ TAMÁS GÁBOR (1985): He obtained his MSc degree in Computer Science, major in Next Generation Networks at the Faculty of Electrical Engineering and Informatics of BME in 2008. In 2007, he was awarded with 1st prize of the National Conference of Scientific Student's Associations. Since 2008, he is a PhD student at the Speech Technology Laboratory of BME TMIT. Research fields: speech technology, speech synthesis.

ZAINKÓ CSABA (1976): obtained his MSc in Technical Informatics at the Faculty of Electrical Engineering and Informatics of in 1999. He is a member of Speech Technology Laboratory of BME TMIT. Research areas include speech synthesis, text processing, human computer interaction, multimodal information systems, user interfaces.

NÉMETH GÉZA (1959): He obtained his MSc in Electrical Engineering, major in Telecommunications at the Faculty of Electrical Engineering of BME in 1983. Also at BME: dr. univ, 1987, PhD 1997. He is the Head of the Speech Technology Laboratory of BME TMIT. Research fields: speech technology, service automation, multilingual speech and multimodal information systems, mobile user interfaces and applications.

References

[1] Chu, M., Zhao, Y., Chang, E.,
"Modeling stylized invariance and local variability of prosody in text-to-speech synthesis",
Speech Communication, Vol. 48, pp.716–726, 2006.

[2] Meron, J.,
"Prosodic unit selection using an imitation speech database", SSW4-2001, p.113, 2001.

[3] Raux A., Black, A.W.,
"A unit selection approach to F0 modeling and its application to emphasis", ASRU, pp.700–705, 2003.

[4] Saito, T.,
"Generating F0 contours by statistical manipulation of natural F0 shapes",
IEICE – Transactions on Information and Systems, Vol. E89-D, No. 3, pp.1100–1106, March 2006.

[5] Iriondo, I., Socoro, J.C., Alias, F.,
"Prosody modelling of Spanish for expressive speech synthesis", ICASSP, Vol. 4, pp.821–824, 2007.

[6] Dong, M., Lua, K.-T.,
"An example-based approach for prosody generation in Chinese speech synthesis", ICSLP, pp.303–307, 2000.

[7] van Santen, J., Kain, A., Klabbers, E., Mishra, T.,
"Synthesis of prosody using multi-level unit sequences", Speech Communication,
Vol. 46, No. 3-4, pp.365–375, 2005.

[8] Díaz, F.C., Banga, E.R.,
"A method for combining intonation modelling and speech unit selection in corpus-based speech synthesis systems", Speech Communication,
Vol. 48, No. 8, pp.941–956, 2006.

[9] Németh, G., Fék, M., Csapó, T.G.,
"Increasing prosodic variability of text-to-speech synthesizers", Interspeech, pp.474–477, 2007.

[10] Olaszy, G., Németh, G., Olaszi, P., Kiss, G., Gordos, G.,
"PROFIVOX – a Hungarian professional TTS system for telecommunications applications".
International Journal of Speech Technology,
No. 3/4, pp.201–216, 2000.

[11] Fék, M., Pesti, P., Németh, G., Zainkó, Cs., Olaszy, G.,
Corpus-Based Unit Selection TTS for Hungarian.
Proc. of Text, Speech and Dialogue, Brno, 2006.
pp.367–374.

[12] Hungarian Speech Database for Creation of Voice Driven Teleservices: Technical report EU INCO Copernicus Project, No. 977017.
European Commission Brussels, 2000.

[13] Olaszy, G., Németh, G.,
"IVR for Banking and Residential Telephone Subscribers Using Stored Messages Combined with a New Number-to-Speech Synthesis Method"
In: Gardner-Bonneau D. (ed.): Human Factors and Voice Interactive Systems. Kluwer, pp.237–256, 1999.

[14] Németh, G., Zainkó, Cs.,
"Multilingual Statistical Text Analysis,
Zipf's Law and Hungarian Speech Generation",
Acta Linguistica Hung. 49. (3-4), pp.385–405, 2002.

Quantum information theoretical based geometrical representation of eavesdropping activity on the quantum channel

LÁSZLÓ GYÖNGYÖSI, SÁNDOR IMRE

Department of Telecommunications, Budapest University of Technology and Economics
{gyongyosi, imre}@hit.bme.hu

Keywords: quantum cryptography, quantum cloning, quantum informational distance

Quantum cryptography is an emerging technology that may offer new forms of security protection, however the quantum cloning based attacks against the protocol will play a crucial role in the future. According to the no-cloning theorem, an eavesdropper on the quantum channel can not copy perfectly the sent quantum states. The best eavesdropping attacks for quantum cryptography are based on imperfect cloning machines. In our method we use quantum relative entropy as an informational distance between quantum states. We show a geometrical approach to analyze the security of quantum cryptography, based on quantum relative entropy and Laguerre Delaunay triangulation on the Bloch sphere. Using Laguerre diagrams, we can compute efficiently the radius of the smallest enclosing ball of quantum states on the Bloch sphere. We present a basically new method to derive quantum relative entropy based Delaunay tessellation on the Bloch ball and to compute the radius of smallest enclosing ball of balls to detect eavesdropping activity on the quantum channel.

1. Introduction

The security of modern cryptographic methods like asymmetric cryptography, relies heavily on the problem of factoring large integers. In the future, if quantum computers become reality, any information exchange using current classical cryptographic schemes will be immediately insecure. Current classical cryptographic methods are not able to guarantee long-term security. Other cryptographic methods, with absolute security must be applied in the future. Cryptography based on quantum theory principles is known as quantum cryptography. Using current network technology, in order to spread quantum cryptography, interfaces able to manage together the quantum and classical channel must be implemented [2]. Quantum cryptography provides new ways to transmit information securely, using the fundamental principles of quantum-mechanics. As classical cryptography uses and manipulates classical bits, quantum cryptography does the same with qubits to realize provably, absolute secure communication. In quantum cryptographic schemes, the secret information is not encoded directly into the quantum states, the qubits are used only to generate a secret cryptographic key, shared between two legal parties, called Alice and Bob. The main idea behind the quantum cryptographic protocols was the absolute secure key distribution, hence we rather call these cryptographic methods as Quantum Key Distribution (QKD) systems [2,7].

Using computational geometry, many complex high dimensional problems can be expressed with graphs and tessellation diagrams [6]. In our fundamentally new security analysis of quantum cryptography, we derive the fidelity of the eavesdropper's cloning machine from Laguerre-type Delaunay diagrams on the Bloch sphere. Using Laguerre diagrams, we can compute efficiently the radius of the smallest enclosing balls of mixed states on the Bloch sphere, and give the level of eavesdropping activity. The geometric interpretation of quantum states can be used to investigate informational distances between two different quantum states [5,6]. We compute the fidelity of the quantum cloning transformation using the classical algorithm presented by Badoui and Clarkson, and the Laguerre Delaunay triangulation on the Bloch sphere [11,13].

Our paper is organized as follows. First we discuss the basic facts about computational geometry and quantum information theory. Then we explain the main elements of our security analysis, and we show the application of our theory for the security analysis of eavesdropping detection on the quantum channel. Finally, we summarize the results.

2. Preliminaries

The security of QKD schemes relies on the *no-cloning* theorem [2]. Contrary to classical information, in a quantum communication system the quantum information cannot be copied perfectly. If Alice sends a number of photons $|\psi_1\rangle, |\psi_2\rangle, ..., |\psi_N\rangle$ through the quantum channel, an eavesdropper is not interested in copying an arbitrary state, only the possible polarization states of the attacked QKD scheme. To copy the sent quantum state, an eavesdropper has to use a quantum cloner machine, and a known *"blank"* state $|0\rangle$, onto which the eavesdropper would like to copy Alice's quantum state.

If Eve wants to copy the *i*-th sent photon $|\psi_i\rangle$, she has to apply a unitary transformation *U*, which gives the following result:

$$U\left(|\psi_i\rangle \otimes |0\rangle\right) = |\psi_i\rangle \otimes |\psi_i\rangle, \quad (1)$$

for each polarization states of qubit $|\psi_i\rangle$. A photon chosen from a given set of polarization states can be cloned perfectly only, if the polarization angles in the set that are distinct, are all mutually orthogonal [2,7]. The unknown non-orthogonal states cannot be cloned perfectly, the cloning process of the quantum states is possible only if the information being cloned is classical, hence the quantum states are *all orthogonal*. The polarization states in the QKD protocols are not all orthogonal states, which makes no possible to an eavesdropper to copy the sent quantum states [2].

Our goal is to measure the level of quantum cloning activity on the quantum channel, using fast computational geometric methods. The fidelity analysis of the eavesdropper's cloning machine indicates, how much the eavesdropper preserves the quality of the cloned quantum states. In our method, quantum informational distance plays an important role in the estimation of the fidelity of eavesdropper's cloning machine.

2.1 The communication model

In our method we measure the *informational theoretical* impacts of quantum cloning activity in the quantum channel. Alice's side is modeled by random variable $X=\{p_i=P(x_i)\}, i=1,...N$. Bob's side can be modeled by another random variable Y. The Shannon entropy for the discrete random variable X is denoted by $H(X)$, which can be defined as $H(X)=-\sum_{i=1}^{N} p_i \log(p_i)$, for conditional random variables, the probability of the random variable X given Y is denoted by $p(X|Y)$. Alice sends a random variable to Bob, who produce an output signal with a given probability.

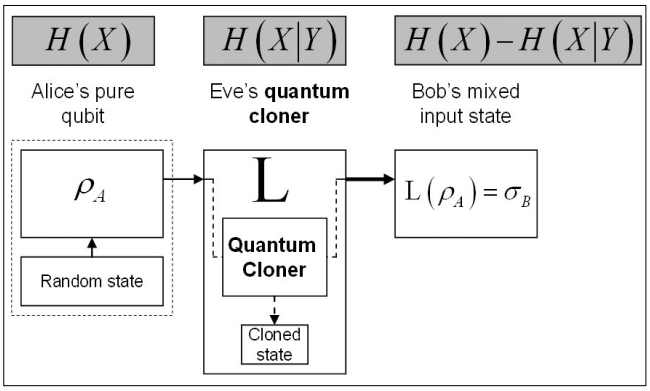
We analyze in a geometrical way the effects of Eve's quantum cloner on Bob's received quantum state. Eve's cloner in the quantum channel increases the uncertainty in X, given Bob's output Y. The informational theoretical noise of Eve's quantum cloner increases conditional Shannon entropy $H(X|Y)$, where

$$H\left(X|Y\right) = -\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} p\left(x_i, y_j\right)\log p\left(x_i|y_j\right), \quad (2)$$

Our geometrical security analysis is focused on the cloned mixed quantum state, received by Bob. The type of the quantum cloner machine depends on the actual protocol. For the four-sate QKD protocol (BB84), Eve chooses the phase-covariant cloner, while for the Six-state protocol she uses the universal quantum cloner (UCM) machine [7,9,10]. Alice's pure state is denoted by ρ_A, Eve's cloner modeled by an affine map $\mathcal{L}$, and Bob's mixed input state is denoted by $\mathcal{L}(\rho_A)=\sigma_B$. In our calculations, we can use the fact, that for random variables X and Y, $H(X,Y)=H(X)+H(Y|X)$, where $H(X)$, $H(X,Y)$ and $H(Y|X)$ are defined by using probability distributions $p(x)$, $p(x,y)$ and $p(y|x)$. We measure in a geometrical representation the information which can be transmitted in presence of an eavesdropper on the quantum channel.

In *Fig. 1* we illustrated Eve's quantum cloner on the quantum channel. Alice's pure state is denoted by ρ_A, the eavesdropper's quantum cloner transformation is denoted by $\mathcal{L}$. The mixed state received by Bob, is represented by σ_B.

Figure 1.
The analyzed attacker model and the entropies



In a private quantum channel, we seek to maximize $H(X)$ and minimize $H(X|Y)$ in order to maximize the radius r^* of the smallest enclosing ball, which describes the maximal transmittable information from Alice to Bob in the *attacked* quantum channel:

$$r^* = Max_{\{all\ possible\ x_i\}} H\left(X\right) - H\left(X|Y\right). \quad (3)$$

To compute the radius r^* of the smallest informational ball of quantum states and the entropies between the cloned quantum states, instead of classical Shannon entropy, we can use von Neumann entropy and quantum *relative entropy*.

Geometrically, the presence of an eavesdropper causes a detectable mapping to change from a noiseless one-to-one relationship, to a stochastic map. If there is no cloning activity on the channel, then $H(X|Y)=0$ and the radius of the smallest enclosing quantum informational ball on Bob's side will be maximal.

2.2 Geometrical representation of quantum states

A quantum state can be described by its *density matrix* $\rho\in\mathbb{C}^{d\times d}$, which is a $d\times d$ matrix, where d is the level of the given quantum system. For an n qubit system, the level of the quantum system is $d=2^n$. In our model, we use the fact, that particle state distributions can be analyzed probabilistically by means of density matrices.

A *two-level* quantum system can be given by its density matrices in the following way:

$$\rho = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix},\ x^2+y^2+z^2 \le 1, \quad (4)$$

where i denotes the complex imaginary $i^2=-1$.

The density matrix $\rho=\rho(x,y,z)$ can be identified with a *point* (x,y,z) in the 3-dimensional space, and a ball **B** formed by such points $\mathbf{B}=\{(x,y,z)\,|\,x^2+y^2+z^2\le 1\}$, is called Bloch ball. The eigenvalues λ_1,λ_2 of $\rho(x,y,z)$ are given by

$$\left(1 \pm \sqrt{x^2+y^2+z^2}\right)\Big/2, \quad (5)$$

the eigenvalue decomposition ρ is $\rho=\sum_i \lambda_i E_i$, where $E_i E_j$ is E_i for $i=j$ and 0 for $i\neq j$. For a *mixed* state $\rho(x,y,z)$, $\log\rho$ defined by $\log\rho=\sum_i(\log\lambda_i)E_i$.

In quantum cryptography the encoded pure quantum states are sent through a quantum communication channel. Using the Bloch sphere representation, the quantum state ρ can be given as a three-dimensional point $\rho=(x,y,z)$ in $\mathbb{R}^3$, and it can be represented by spherical coordinates

$$\rho=(r,\theta,\varphi), \tag{6}$$

where r is the radius of the quantum state to the origin, θ and φ represents the latitude and longitude rotation angles. Using the spherical coordinates, a three-dimensional point on the Bloch sphere $\mathcal{B}$, can be given by:

$$\begin{aligned} x &= r\sin\theta\cos\varphi,\\ y &= r\sin\varphi,\\ z &= r\cos\theta\cos\varphi. \end{aligned} \tag{7}$$

A mesh of the Bloch sphere $\mathcal{B}$ can be described as a number of points connected in some way by lines, the points and the lines of the mesh are referred to as edges and vertices.

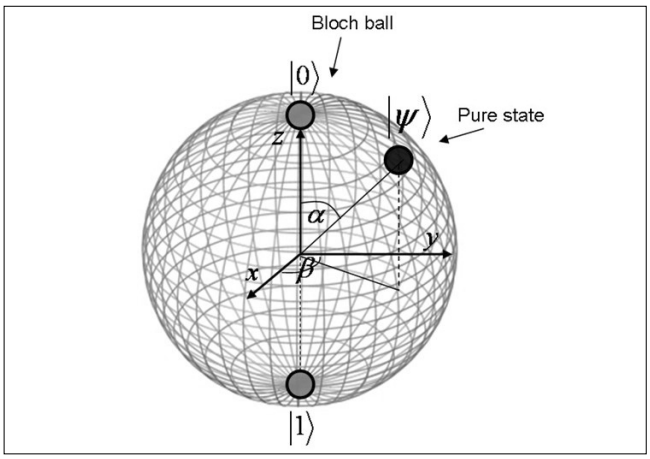
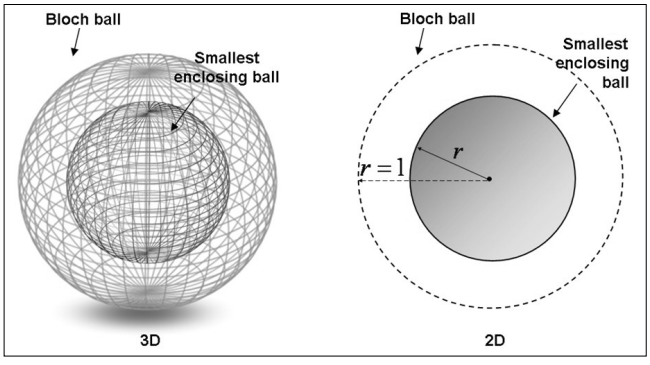


Figure 2. Mesh of Bloch sphere

Geometrically, the *pure* states are on the boundary of the Bloch ball $\mathcal{B}$, while the *mixed* states are inside the Bloch ball. In *Fig. 3* the pure states with unit radius are on the surface of the Bloch-sphere, while the mixed states with radius $r<1$ are contained inside the sphere.

Figure 3.
The effect of the eavesdropper's cloning transformation in geometrical representation



A pure state can be given by $|\psi\rangle=\alpha|0\rangle+\beta|1\rangle$ and the projector of the state is

$$|\psi\rangle\langle\psi| = \frac{1}{2}\left(\mathbf{1}+\hat{n}\cdot\vec{\sigma}\right),$$

where $\hat{n}$ is the Bloch vector, and it can be given by $\hat{n}=(2\mathrm{Re}(\alpha\beta^*), 2\mathrm{Im}(\alpha\beta^*), |\alpha|^2-|\beta|^2)$ [4,7].

2.3 Measuring distances between quantum states

In the proposed security analysis, the distance between quantum states is defined by the *quantum relative entropy* of quantum states. The relative entropy of quantum states measures the *informational distance* between quantum states. In our model, the informational distance between quantum states is computed by using their density matrices. The classical Shannon-entropy of a discrete d-dimensional distribution p can be given by

$$H(p)=\sum_{i=1}^{d} p_i \log\frac{1}{p_i} = -\sum_{i=1}^{d} p_i \log p_i\ . \tag{8}$$

The *von Neumann* entropy $S(\rho)$ of quantum states is a generalization of the classical Shannon entropy to density matrices [4,7]. The entropy of quantum states can be given by:

$$\mathrm{S}(\rho) = -Tr(\rho\log\rho). \tag{9}$$

The quantum entropy $S(\rho)$ is equal to the Shannon entropy for the eigenvalue distribution:

$$\mathrm{S}(\rho)=\mathrm{S}(\lambda)=-\sum_{i=1}^{d}\lambda_i\log\lambda_i, \tag{10}$$

where d is the level of the quantum system.

The relative entropy in classical systems is a measure that quantifies how close a probability distribution p is to a model or candidate probability distribution q [4,7]. For p and q probability distributions the *relative entropy* can be given by

$$D(p\|q)=\sum_i p_i\log_2\frac{p_i}{q_i}, \tag{11}$$

while the relative entropy between quantum states measured by

$$D(\rho\|\sigma)=Tr\left[\rho\left(\log\rho-\log\sigma\right)\right]. \tag{12}$$

The quantum relative entropy plays a key role in the description of the quantum state space. The quantum informational distance has some distant-like properties, however it is *not* commutative [4,7], thus $D(\rho\|\sigma)\neq D(\sigma\|\rho)$ and $D(\rho\|\sigma)\geq 0$ iff $\rho\neq\sigma$, and $D(\rho\|\sigma)=0$ iff $\rho=\sigma$. We note, that if σ has zero eigenvalues $D(\rho\|\sigma)$ may *diverge*, otherwise it is a finite and continuous function. The quantum relative entropy reduces to the classical Kullback-Leibler relative entropy for simultaneously diagonal matrices.

2.4 Quantum relative entropy

The *relative entropy* between *quantum states* can be described by a strictly convex and differentiable generator function **F** as:

$$\mathbf{F}(\rho) = -\mathrm{S}(\rho) = Tr(\rho\log\rho), \tag{13}$$

where –S is the negative von Neumann entropy function.

The *relative quantum entropy* $D(\rho\|\sigma)$ for density matrices ρ and σ can be given by generator function **F** in the following way:

$$D\left(\rho\middle\|\sigma\right)=\mathbf{F}\left(\rho\right)-\mathbf{F}\left(\sigma\right)-\left\langle\rho-\sigma,\nabla\mathbf{F}\left(\sigma\right)\right\rangle, \quad (14)$$

where $\langle\rho,\sigma\rangle=\mathrm{Tr}(\rho\sigma^*)$ is the *inner product* of quantum states, and $\nabla\mathbf{F}(\cdot)$ is the gradient.

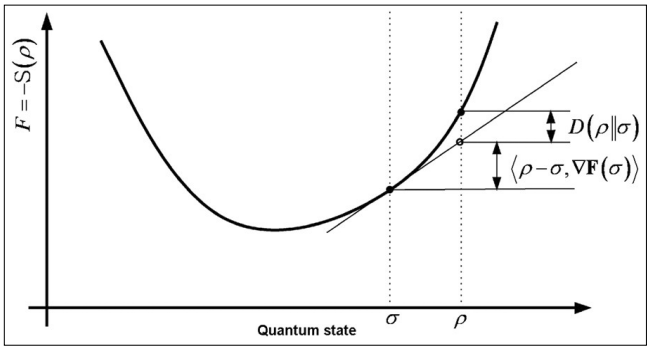


Figure 4.
Visualizing generator function as negative von Neumann entropy

The quantum relative entropy for general quantum state $\rho=(x,y,z)$ and mixed state $\sigma=(\tilde{x},\tilde{y},\tilde{z})$, with radius $r_\rho=\sqrt{x^2+y^2+z^2}$ and $r_\sigma=\sqrt{\tilde{x}^2+\tilde{y}^2+\tilde{z}^2}$ can be given by

$$\begin{aligned}D\left(\rho\middle\|\sigma\right)=&\frac{1}{2}\log\frac{1}{4}\left(1-r_\rho{}^2\right)+\frac{1}{2}r_\rho\log\frac{\left(1+r_\rho\right)}{\left(1-r_\rho\right)}\\&-\frac{1}{2}\log\frac{1}{4}\left(1-r_\sigma{}^2\right)-\frac{1}{2r_\sigma}\log\frac{\left(1+r_\sigma\right)}{\left(1-r_\sigma\right)}\left\langle\rho,\sigma\right\rangle\end{aligned} \quad (15)$$

where $\langle\rho,\sigma\rangle=(x\tilde{x}+y\tilde{y}+z\tilde{z})$. For a maximally mixed state $\sigma=(\tilde{x},\tilde{y},\tilde{z})=(0,0,0)$ and $r_\sigma=0$, the quantum informational distance can be expressed as

$$\begin{aligned}&D\left(\rho\middle\|\sigma\right)=\\&\frac{1}{2}\log\frac{1}{4}\left(1-r_\rho{}^2\right)+\frac{1}{2}r_\rho\log\frac{\left(1+r_\rho\right)}{\left(1-r_\rho\right)}-\frac{1}{2}\log\frac{1}{4}.\end{aligned} \quad (16)$$

The density matrices of quantum bits are represented by 3D points in the Bloch ball. If we compute the distance between two quantum states in the 3D Bloch ball representation, we compute the distance between two Hermitian matrices ρ and σ.

The eavesdropper's cloner transformation is modeled by an affine map, that maps quantum states to quantum states. Geometrically, the effect of the eavesdropper is to map the Bloch ball to a deformed ball. The cloning activity in the channel can be analyzed by the radius of the deformed Bloch ball, which can be computed by geometrical methods.

In our security analysis we use Delaunay tessellation, which is *symmetric* only for pure states, and *asymmetric* for mixed states. It can be proven, that for pure states the Delaunay diagram coincidences with Euclidean Delaunay diagram, but for *mixed* states the Delaunay diagram is asymmetric, hence it is not identical to Euclidean diagrams [16].

3. Eavesdropping activity on the quantum channel

In quantum cryptography the best eavesdropping attacks use the quantum cloning machines [7-9]. However, an eavesdropper can not measure the state $|\psi\rangle$ of a single quantum bit, since the result of her measurement is one of the single quantum system's eigenstates. The measured eigenstate gives only very poor information to the eavesdropper about the original state $|\psi\rangle$ [2,7].

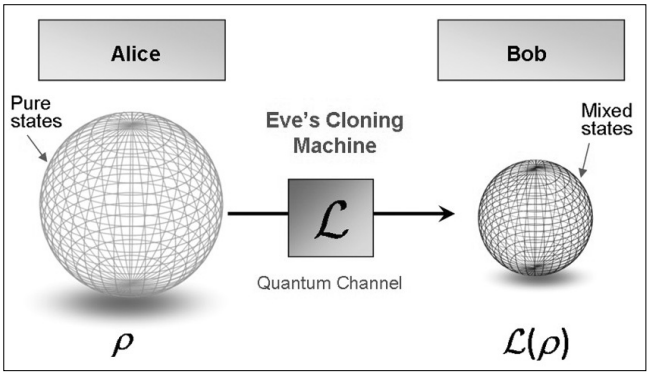


Figure 5.
The effect of quantum cloning attack on the sent pure quantum state

The process of cloning of *pure* states can be generalized as

$$\left|\psi\right\rangle_a\otimes\left|\Sigma\right\rangle_b\otimes\left|Q\right\rangle_x\rightarrow\left|\Psi\right\rangle_{abx}, \quad (17)$$

where $|\psi\rangle$ is the state in the *Hilbert space* to be copied, $|\Sigma\rangle$ is a *reference* state, and $|Q\rangle$ is the *ancilla* state [7]. A cloning machine is called *symmetric* if at the output all the clones have the same fidelity, and asymmetric if the clones have different fidelities [8,9].

The no-cloning theorem has important role in quantum cryptography, since it makes no possible to copy a quantum state perfectly. In 1996 Buzek and Hillery published the method of imperfect cloning, while the original no-cloning theorem was applied only to perfect cloning [2]. The asymmetric cloning machines have been discussed for eavesdropping of quantum cryptography in [10,15]. For attacks on some quantum cryptography protocol, it has been proven that the best strategy uses quantum cloning machines [7,9].

3.1 The smallest enclosing quantum-information ball

We would like to compute the radius *r* of the smallest enclosing ball of the cloned mixed quantum states, thus first we have to seek the center **c*** of the point set S. The set S of quantum states is denoted by $S=\{\rho_i\}_{i=1}^n$.

The distance function $d(\cdot,\cdot)$ between any two quantum states of S is measured by quantum relative entropy, thus the *minimax* mathematical optimization can be applied to *quantum relative entropy* based distances to find the center **c** of the set S. We denote the quantum relative entropy from **c** to the furthest point of S by

$$d\left(\mathbf{c},\mathcal{S}\right)=\max_i d\left(\mathbf{c},\rho_i\right). \quad (18)$$

Using the *minimax* optimization, we can *minimize* the *maximal* quantum relative entropy from **c** to the furthest point of $\mathcal{S}$ by

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} d(\mathbf{c}, \mathcal{S}). \quad (19)$$

In *Fig. 6* we illustrated the *circumcenter* **c*** of $\mathcal{S}$ for the Euclidean distance and for *quantum relative entropy* [1].

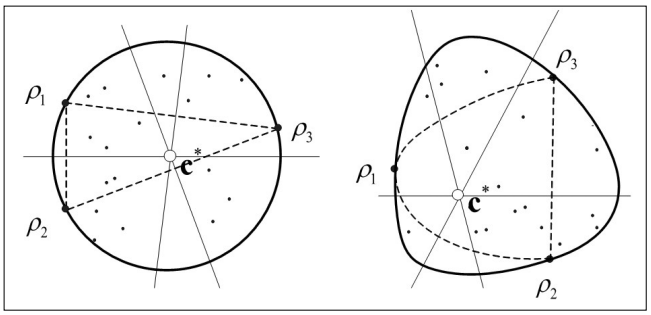


Figure 6.
Circumcenter for Euclidean distance ball and quantum relative entropy ball

In our fidelity analysis we assume, that the eavesdropper's cloning machine does a *linear transform* $\mathcal{L}$ that maps quantum states to quantum states. The eavesdropper's cloning transformation $\mathcal{L}$ is a trace-preserving, i.e. $Tr\mathcal{L}(\rho)=Tr(\rho)$, and completely positive map [1]. The informational theoretical effect of the *eavesdropper's cloning machine* is described the radius of the smallest enclosing quantum informational ball by r^*. The quantum informational theoretical radius equal to the *maximum* quantum informational distance from the center, and it can be expressed as:

$$r^* = \min_{\sigma \in \mathcal{S}(\mathbb{C}^2)} \max_{\rho \in \mathcal{S}(\mathbb{C}^2)} D\left(\mathcal{L}(\rho) \| \mathcal{L}(\sigma)\right). \quad (20)$$

In our procedure of computing smallest enclosing information ball, we use quantum Delaunay diagrams, because it is the *fastest* known tool to seek a center of a smallest enclosing ball of points.

4. Geometrical model of secure quantum communication

4.1 Properties of quantum cloners

The *maximal* fidelity of the eavesdropper's cloning machine is denoted by F_{Eve}. The parameter F_{Eve} represents the theoretical upper bound on the *cloning machine's fidelity* [1]. For example, if Eve uses *universal* quantum cloner, then the value of parameter F_{Eve} is independent of input quantum state $|\psi\rangle$, and the *fidelity* of her optimal quantum cloning machine is

$$F_{Eve} = \langle\psi|^{(in)} \rho^{(out)} |\psi\rangle^{(in)} = \frac{1}{2}(1+\eta), \quad (21)$$

where η is the *reduction* factor. The quantum cloning transformation optimal [8,9], if $\eta=2/3$, hence the maximal fidelity of optimal universal cloning is $F_{Eve}=5/6$, and the maximal radius of the cloned state is

$$r_{Eve}^{universal} = \frac{2}{3}. \quad (22)$$

The *quantum informational theoretical* radius can be defined as

$$r_{Eve}^{*universal} = 1 - \mathrm{S}\left(r_{Eve}^{universal}\right), \quad (23)$$

where S is the von Neumann entropy of corresponding quantum state with radius length $r_{Eve}^{universal}$.

In general, the universal cloning machine output state can be expressed by [7-9]

$$\rho^{(out)} = F_{Eve} |\psi\rangle_a \langle\psi| + (1 - F_{Eve}) |\psi_\perp\rangle_a \langle\psi_\perp|. \quad (24)$$

4.2 Asymetric phase-covariant quantum cloner

Asymmetric cloning has direct application to eavesdropping strategies in quantum cryptography. The best-known example of state-dependent quantum cloning machine is the *phase-covariant* cloning machine. Here, the states lie in the equator (x–y) of the Bloch sphere, thus the fidelity of the cloning will be independent of φ. The phase-covariant cloning machine has a remarkable application in quantum cryptography, since it is used in the optimal strategy for eavesdropping [8-10]. The importance of equatorial qubits lies in the fact that quantum cryptography requires these states rather than the states, that span the whole Bloch sphere [9].

In phase-covariant cloning, the transformations restrict for pure input states

$$|\psi_\phi\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\phi}|1\rangle\right), \quad (25)$$

where the parameter $\phi \in [0, 2\pi)$ represents the angle between the *Bloch vector* and the x-axis. These qubits are called *equatorial* qubits, because the z-component of their Bloch vector is zero. The *phase-covariant* quantum cloners [9] can clone arbitrary *equatorial* qubits, and they keep the quality of the copies same for all equatorial qubits. The *reduced* density operator of the copies at the output can be expressed as [9]

$$\rho^{(out)} = \left(\frac{1}{2} + \sqrt{\frac{1}{8}}\right)|\psi_\phi\rangle\langle\psi_\phi| + \left(\frac{1}{2} - \sqrt{\frac{1}{8}}\right)|\psi_{\phi,\perp}\rangle\langle\psi_{\phi,\perp}|, \quad (26)$$

where $|\psi_{\phi,\perp}\rangle$ is orthogonal to state $|\psi_\phi\rangle$. Thereby, the optimal fidelity of 1 to 2 phase-covariant cloning transformation is given by

$$F_{1\to2}^{phasecov.} = \frac{1}{2} + \sqrt{\frac{1}{8}} \approx 0.8535. \quad (27)$$

If Eve has a phase-covariant quantum cloner, then the maximal value of her *radius* r_{Eve}^{phase} is

$$r_{Eve}^{phasecov.} = 2\sqrt{\frac{1}{8}}. \quad (28)$$

The *quantum informational theoretical* radius r_{Eve}^{*phase} of the phase-covariant cloner can be defined as

$$r_{Eve}^{*phasecov.} = 1 - \mathrm{S}\left(r_{Eve}^{phasecov.}\right), \quad (29)$$

where S is the von Neumann entropy of corresponding quantum state with radius length r_{Eve}^{phase}. The phase-covariant quantum cloning transformation produces two

copies of the equatorial qubit with optimal fidelity. The phase-covariant cloning transformation without ancilla is a two-qubit unitary transformation, it can be given by $|0\rangle|0\rangle \rightarrow |0\rangle|0\rangle$ and $|1\rangle|0\rangle \rightarrow \cos\eta|1\rangle|0\rangle + \sin\eta|0\rangle|1\rangle$, where $\eta \in [0,\pi/2]$ is the shrinking parameter, which is related to the fidelity.

In *Fig. 7* we compared the information theoretical radiuses r^*_{UCM} and $r^*_{phasecov.}$ of the smallest enclosing quantum informational balls for idealistic UCM based attack and idealistic phase-covariant cloner based attack. The maximal distance states are denoted by ρ_{UCM} and $\rho_{phasecov.}$

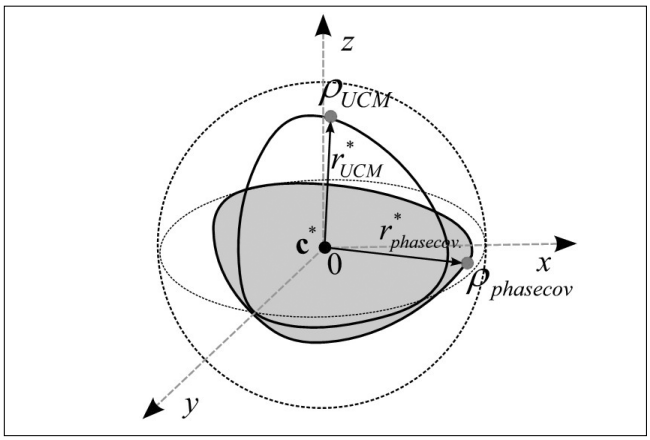


Figure 7.
Comparison of smallest enclosing quantum informational balls of idealistic UCM and phase-covariant based attack

The best quality of the two outputs simultaneously can be realized with an UCM. If an eavesdropper uses a phase-covariant cloner, one of the two outputs should have better fidelity, while the fidelity of second output will be lower.

4.3 Quantum cloning detection

In our model we derive the fidelity of the eavesdropper's cloning machine from the quantum informational theoretical radius r^* of the smallest enclosing quantum informational ball, and the theoretical upper bound on the *quantum informational theoretical* radius of the eavesdropper's cloning machine denoted by r^*_{Eve} [1].

As the first part of our theorem, for a *secure* quantum channel, the radius r^* of the smallest enclosing quantum information ball of *mixed states* has to be greater than r^*_{Eve}, thus

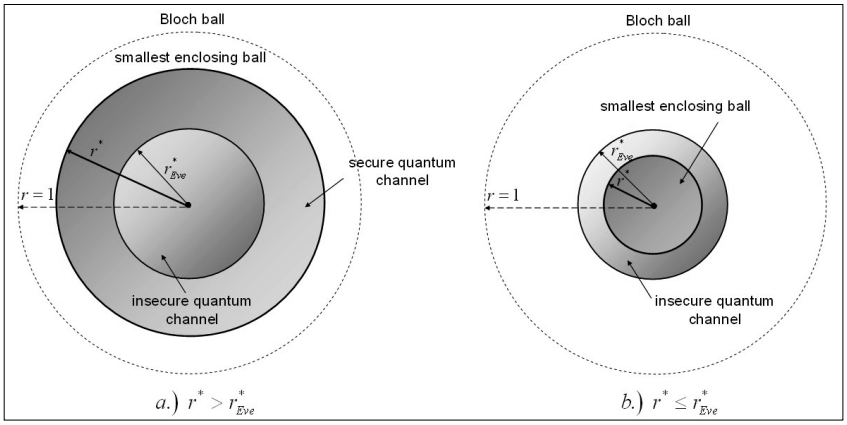
$$r^* > r^*_{Eve}. \tag{30}$$

As the second part, for an *insecure* quantum channel, the radius r^* is smaller than or equal to r^*_{Eve}, thus

$$r^* \leq r^*_{Eve}. \tag{31}$$

In *Fig. 8* we show the geometrical interpretation of our model [1].

Figure 8.
The radius of the smallest enclosing information ball for a secure (a) and insecure (b) quantum communication



In our security analysis, we use the spherical Delaunay tessellation to compute the quantum information theoretical radius r^*, since it can be simply obtained as an ordinary Euclidean Delaunay triangulation mesh. The quantum relative entropy based Delaunay tessellation of pure states is *identical* to the conventional spherical Delaunay tessellation, and it differs between mixed quantum states [6].

5. Tessellation on the Bloch sphere

5.1 Mathematical background

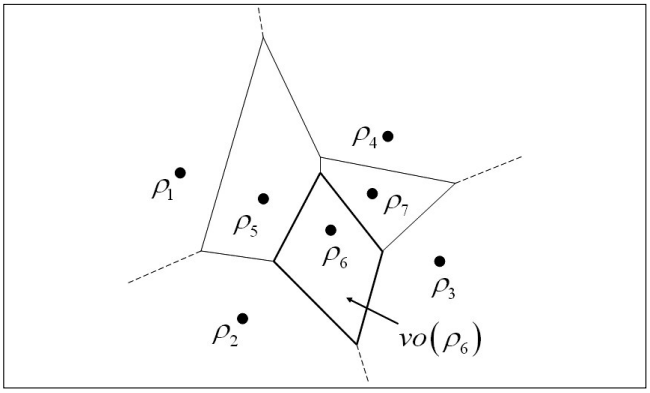
The dual of Delaunay diagram of a set of quantum states on the Bloch ball $\mathcal{B}$, is the division of the space into regions. The regions contain the part of the *quantum space* which is closer to that point than any other. Formally, for a given set of quantum states $\mathcal{S}=\{\rho_1,\rho_2,...\rho_n\}$ in $\mathbb{R}^d$, the Voronoi diagram $V(\mathcal{S})$ is the partition of $\mathbb{R}^d$ into *n polyhedral* regions, one for each quantum states ρ_i. These regions on the Bloch ball $\mathcal{B}$ are the *Voronoi cells*, denoted by $vo(\rho)$, containing the points in $\mathbb{R}^d$ which are closer to quantum state ρ than all other points.

Formally, the Voronoi cell $vo(\rho)$ for quantum state ρ and the set of quantum state $\mathcal{S}$ can be given by

$$vo(\rho) = \left\{ x \in \mathbb{R}^d \middle| d(x,\rho_i) \leq d(x,\rho_j) \in \mathcal{S}\{\rho_i\} \right\}, \tag{32}$$

where $d(\cdot)$ is the *distance* function. The Voronoi vertices are in the intersections of the *bisectors* or *boundaries*, as we illustrated it in *Fig. 9*.

Figure 9. An Euclidean tessellation on the Bloch ball



On the Bloch ball $\mathcal{B}$ every $vo(\rho)$ corresponds to a quantum state ρ, thus we have n Voronoi cells for n quantum states, and there are $O(n)$ vertices and edges [6].

5.2 Delaunay triangulation in the quantum space

We use the *Voronoi vertices* in our security analysis, since these vertices play a crucial role in the computation of Delaunay triangulation on the Bloch ball $\mathcal{B}$. The circumcenter of the given quantum states is the *circle* that passes through the quantum states ρ_1 and ρ_2 of the edge $\rho_1\rho_2$ and endpoints ρ_1, ρ_2 and ρ_3 of the triangle $\rho_1\rho_2\rho_3$.

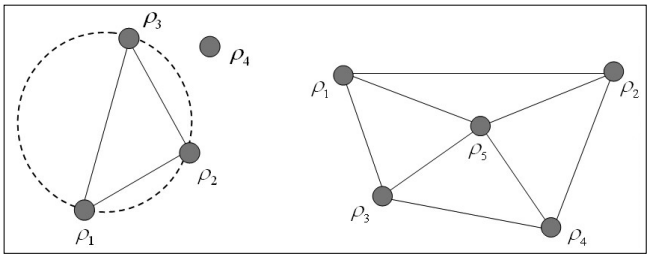


Figure 10.
The Delaunay triangulation of a set of quantum states

The triangle t is said to be *Delaunay*, when its circumcenter is *empty* [6]. The circle centered at a vertex **c**, gives an *empty* circumcenter for quantum states $\{\rho_1, \rho_2, \rho_3\}$. The Delaunay triangulation of a set of quantum states S, denoted by $Del(S)$, is unique, if at most three quantum states $\rho \in S$ are co-circular [5]. The *Delaunay triangulation* $Del(S)$ of a set of quantum states $S=\{\rho_1, \rho_2, ...\rho_n\}$ *maximizes* the *minimum* angle among all triangulation of the given set of quantum states.

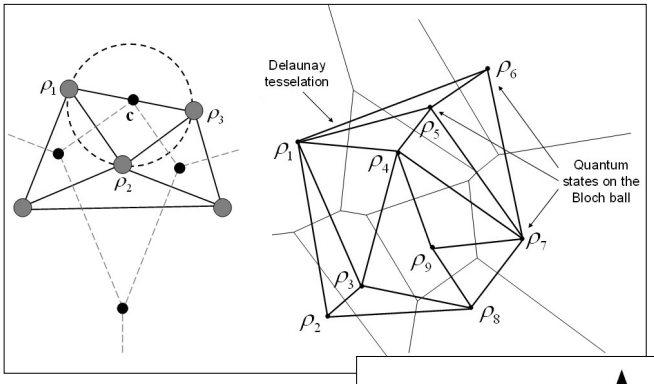


Figure 11.
The triangle of quantum states corresponds to the vertex c, which is the center of its circumcenter (a) and Delaunay tessellation on the Bloch sphere (b)

In our security analysis we use the fact, that the Voronoi diagram $V(S)$ of set of quantum states S, and the Delaunay triangulation $D(S)$ are dual to each other in Euclidean space, and in the quantum space with geodesic edges [6].

Using the Voronoi-Delaunay duality, every triangle $t \in Del(S)$ corresponds to a vertex $v \in V(S)$, and every edge $e(\rho,\sigma) \in Del(S)$ in the *Delaunay triangle* between two quantum states in S corresponds to the boundary edge between the Voronoi cells $vo(\rho)$ and $vo(\sigma)$.

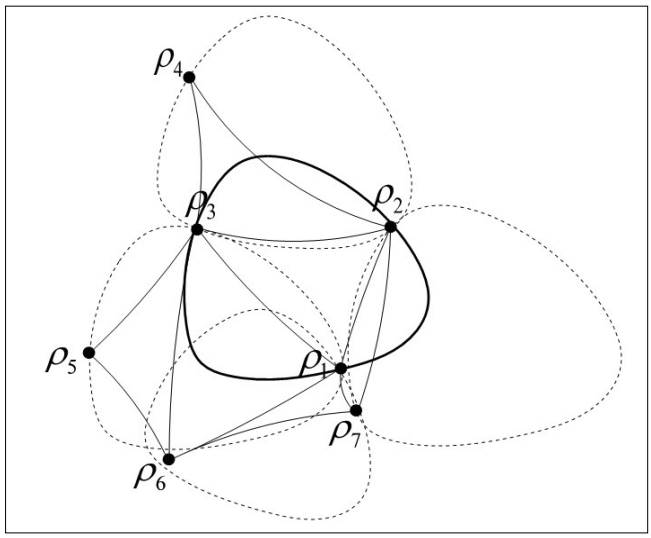


Figure 12. The empty ball property for quantum Delaunay triangulation

The quantum Delaunay diagrams between mixed states differ from Euclidean diagrams, as we have illustrated it in *Fig. 12*.

In *Fig. 13/a* we illustrated the dual-Delaunay diagram for pure quantum states, with unit length radiuses. The quantum diagram for pure states is equivalent to the ordinary Euclidean diagram on the Bloch-sphere.

In *Fig. 13/b* we illustrated the quantum diagrams for mixed states with radiuses $r_{\rho_{1,2,3,4}}<1$, in the Bloch ball representation. Since the quantum informational distance is asymmetric, we can define two types of diagrams. The first-type diagram is illustrated by bold lines, the dashed lines show the dual curved, second-type diagram.

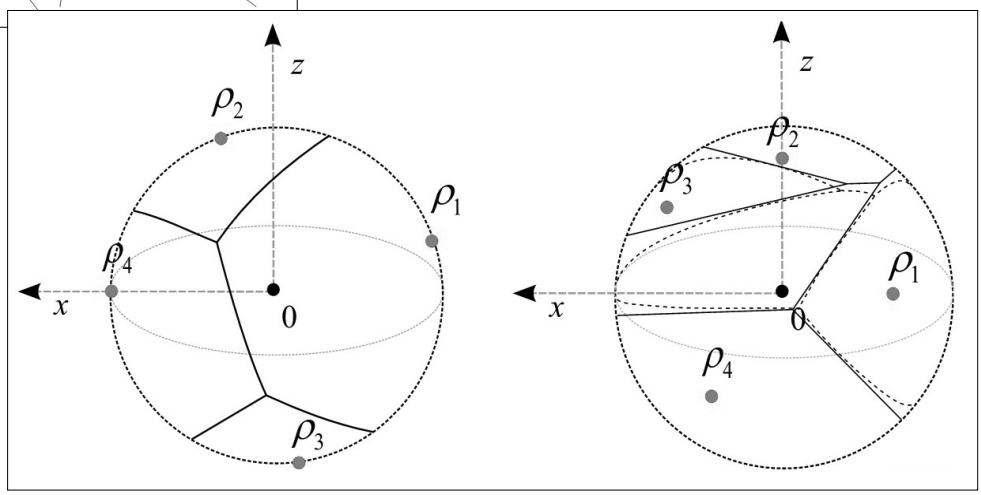


Figure 13.
Dual-Delaunay diagram for pure states (a) and for mixed states (b) on the Bloch ball. For mixed states, the quantum diagrams differ from the Euclidean diagram.

As we can conclude, the quantum diagrams of *pure* quantum states are *equivalent* to ordinary Euclidean diagrams. The quantum diagrams of mixed states with different radiuses are equivalent to quantum informational diagrams.

5.3 Laguerre diagram for quantum states

We use *Laguerre* Delaunay diagram to compute the radius of the smallest enclosing ball [6]. In generally, the *Laguerre* distance for generating quantum state x_i and with weight r_i^2 can be expressed as

$$d_L(\rho, x_i) = \|\rho - x_i\|^2 - r_i^2 . \quad (33)$$

The Delaunay diagram with respect to the *Laguerre* distance is called *Laguerre* Delaunay diagram. For the *Laguerre* bisector of two *three-dimensional* Euclidean balls $B(\rho,r_\rho)$ and $B(\sigma,r_\sigma)$ centered at quantum states ρ and σ, we can write the following equation [6]:

$$2\langle x, \sigma - \rho\rangle + \langle \rho, \rho\rangle - \langle \sigma, \sigma\rangle + r_\sigma^2 - r_\rho^2 = 0. \quad (34)$$

The bisector equation for the ordinary three-dimensional Euclidean Delaunay tessellation can be given by

$$2\langle x, \sigma - \rho\rangle + \langle \rho, \rho\rangle - \langle \sigma, \sigma\rangle = 0, \quad (35)$$

thus for *pure* quantum states, where $r_\sigma^2 = r_\rho^2$, the *quantum relative entropy* based *Delaunay tessellation* on the Bloch ball coincidences with the ordinary Euclidean distance based Delaunay tessellation [6]. On the *Laguerre* diagram, the center of the quantum informational ball can be described by the density matrix χ_i as [6]:

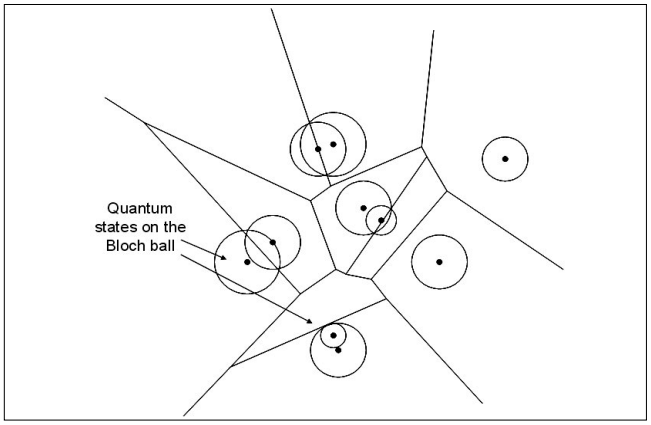
$$\nabla F_{\mathcal{B}}(\chi_i) = \log \chi_i = A_i \begin{pmatrix} \log \lambda_{i,1} & 0 \\ 0 & \log \lambda_{i,2} \end{pmatrix} A_i^*, \quad (36)$$

where (37)

$$A_i = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{x_i - iy_i}{\sqrt{x_i^2 + y_i^2}} \sqrt{\frac{r_i + z_i}{r_i}} & \frac{x_i - iy_i}{\sqrt{x_i^2 + y_i^2}} \sqrt{\frac{r_i - z_i}{r_i}} \\ \sqrt{\frac{r_i - z_i}{r_i}} & -\sqrt{\frac{r_i + z_i}{r_i}} \end{pmatrix}.$$

We illustrated the dual diagram of the *Laguerre* Delaunay tessellation in the Euclidean space in *Fig. 14*.

Figure 14.
Laguerre diagram for quantum states on the Bloch ball



The squared radius r_i^2 of the quantum state ρ_i on the Bloch sphere can be given by

$$r_i^2 = \quad (38)$$
$$\left\langle \nabla \mathbf{F}_{\mathcal{B}}(\rho_i), \nabla \mathbf{F}_{\mathcal{B}}(\rho_i) \right\rangle + 2\left(\mathbf{F}_{\mathcal{B}}(\rho_i) - \left\langle \rho_i, \nabla \mathbf{F}_{\mathcal{B}}(\rho_i) \right\rangle \right).$$

As we can conclude, for weight r_i^2, the *Laguerre* distance $d_L(\rho,x_i)$ can be interpreted as the square of the length of the line segment starting at ρ and tangent to the circle centered at x_i, with radius $\sqrt{r_i^2}$. Thus, the circle centered at x_i with radius $\sqrt{r_i^2}$ is the circle associated with x_i.

6. The proposed algorithm for quantum cloning detection

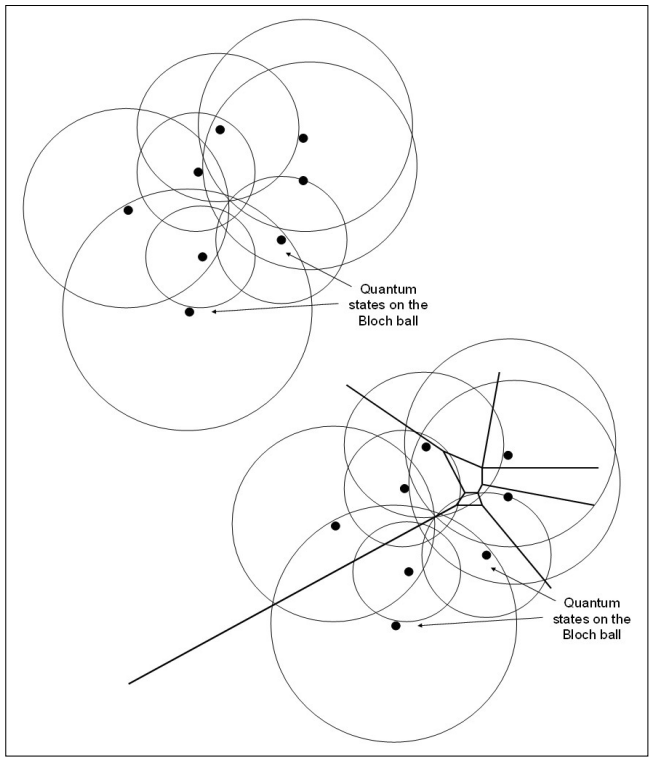
In our algorithm we present an effective solution to seek the *center* **c** of the set of smallest enclosing quantum information ball, using *Laguerre* diagrams.

Our geometrical based algorithm consists of two main steps:

1. *We construct Delaunay triangulation from Laguerre diagrams on the Bloch ball.*
2. *Seek the center of smallest enclosing ball.*

A Delaunay triangulation in the *d*-dimensional quantum space can be obtained by other methods, like a *paraboloid* in the *d+1* dimensional space expressed by $x_{d+1}=x_1^2+...+x_d^2$ and *tangent planes* at the points [1]. In this method we can use the fact, that *the lower envelope of the tangent planes* is a Delaunay diagram [3,17]. However, in this paper we show a more effective algorithm to compute Delaunay tessellation on the Bloch sphere $\mathcal{B}$.

Figure 15.
Tessellation on the Bloch ball obtained by Laguerre diagram



6.1 Construction of Delaunay triangulation from Laguerre diagrams

In our algorithm, we use the fact that the Delaunay tessellation can be computed by *Laguerre* diagrams, thus we can give the tessellation from the *Laguerre* diagram of a set of corresponding ball [5].

In *Fig. 15* we illustrated the Laguerre diagram on the Bloch ball, and the construction of Voronoi diagram.

We use the results proposed in [5], to construct the quantum relative entropy based dual diagram of the Delaunay tessellation, using the Laguerre diagram of the n Euclidean spheres of equations

$$\left\langle x-\rho_i', x-\rho_i' \right\rangle = \left\langle \rho_i', \rho_i' \right\rangle + 2\left(\mathbf{F}\left(\rho_i\right) - \left\langle \rho_i, \rho_i' \right\rangle\right),\ \left(i=1,\ldots,n\right), \tag{39}$$

where ρ_i and ρ_i' denote the first-type and second-type diagrams. In *Fig. 16* we show the ordinary triangulation of quantum relative entropy based Voronoi diagram.

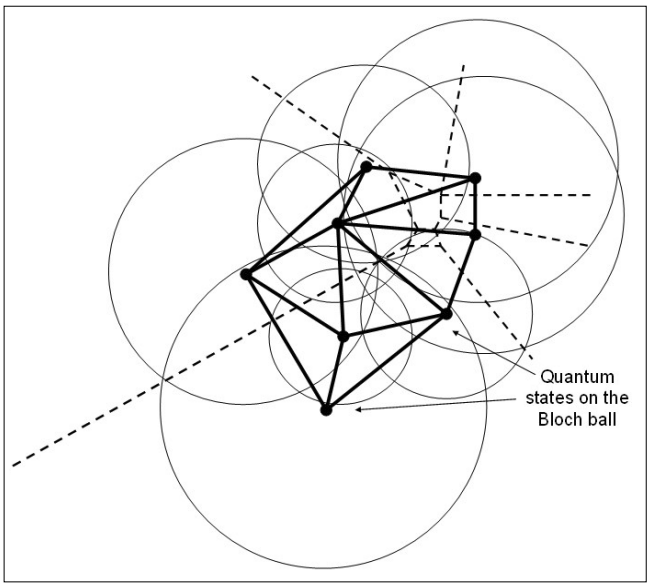


Figure 16. Ordinary Euclidean Delaunay triangulation

The *centers* of the Euclidean spheres are ρ_i, and $\rho_i = \rho_i'$, thus $r_i^2 = 0$. The *generator* function of the quantum relative entropy based diagram is the negative quantum entropy $\mathbf{F}(x) = \sum_i x_i \log x_i$, and the *gradient* $\nabla\mathbf{F}(x) = [\log x_1 \ldots \log x_d]^T$. On the *quantum relative entropy* based diagram, we map quantum state $\rho = [\rho_1 \ldots \rho_d]^T$ to a Euclidean ball of center $\rho' = [\rho_1 \ldots \rho_d]^T$ [5], with radius $r_\rho^2 = \sum_i (\log^2 \rho_i - 2\rho_i)$. The most important result of this equivalence, that we can construct efficiently quantum relative entropy based Delaunay triangulation on the Bloch sphere using Euclidean spheres, which can be calculated efficiently by fast algorithms [5].

6.2 Seek the center of the smallest enclosing quantum informational ball

In our security analysis we use an approximation algorithm from classical *computational geometry* to determine the smallest enclosing ball of balls using *core-sets*. We apply the approximation algorithm presented by *Badoui and Clarkson*, however in our algorithm, the distance measurement between quantum states is based on *quantum informational distance* [11,13]. To apply our approximation algorithm in eavesdropping activity detection, we use the $\mathcal{E}$-core set C for the *minimax* quantum information ball of set of quantum states S [11]. The $\mathcal{E}$-core set C is a subset of the set $C \subseteq S$, such for the *circumcenter* $\mathbf{c}$ of the *minimax* ball

$$d\left(\mathbf{c}, \mathcal{S}\right) \le \left(1+\mathcal{E}\right) r, \tag{40}$$

where r is the radius of the smallest enclosing quantum information ball of set of quantum states S. Our geometrical based eavesdropping detection can be computed very effectively, based on the fact that approximating algorithm can find the *radius* r of the smallest enclosing ball of balls in $O(dn/\mathcal{E}^2)$ time, with an $(1+\mathcal{E})$ approximation [11]. Moreover, in the applied approximation algorithm the *core-set* sizes are bounded by $2/\mathcal{E}$, independently of the dimension [13,14].

Quantum relative entropy based approximation

The approximating algorithm, for a set of quantum states $S = \{s_1,\ldots,s_n\}$ and *circumcenter* $\mathbf{c}$ first finds a farthest point s_m of ball set B, and moves $\mathbf{c}$ towards s_m in $O(dn)$ time in every iteration step. The algorithm does $\lfloor 1/\mathcal{E}^2 \rfloor$ iterations to ensure an $(1+\mathcal{E})$ approximation, thus the overall cost of the algorithm is $O(dn/\mathcal{E}^2)$ [11].

The main steps of our quantum relative entropy based algorithm are:

Algorithm

1. *Select* a random center $\mathbf{c}_1$ from the set of quantum states $\mathcal{S}$

$$\mathbf{c}_1 = s_1$$

for $\left(i = 1,2,\ldots,\left\lceil \frac{1}{\mathcal{E}^2} \right\rceil\right)$ **do**

2. *Find* the farthest point s of $\mathcal{S}$ wrt. *quantum relative entropy*

$$S \leftarrow \arg\max_{s' \in \mathcal{S}} D_F\left(\mathbf{c}_i, s'\right)$$

3. *Update* the circumcircle:

$$\mathbf{c}_{i+1} \leftarrow \nabla_F^{-1}\left(\frac{i}{i+1}\nabla_F\left(\mathbf{c}_i\right) + \frac{1}{i+1}\nabla_F\left(S\right)\right).$$

4. *Return* $\mathbf{c}_{i+1}$

We denote the set of n d-dimensional balls by $B = \{b_1,\ldots,b_n\}$, where $b_i = Ball(s_i, r_i)$, where S_i is the center of the ball b_i, and r_i is the radius of the i-th ball radius. The smallest enclosing ball of set $B = \{b_1,\ldots,b_n\}$ is the unique ball $b^* = Ball(\mathbf{c}^*, r^*)$ with minimum radius r^* and center $\mathbf{c}^*$, containing all the set $\{b_1,\ldots,b_n\}$. The smallest enclosing ball of a ball set, can be written as $\min_{\mathbf{c}} F_B(\mathbf{c})$, where $F_B(X) = d(X,B) = \max_{i \in \{1,\ldots,n\}} d(X,B_i)$, and the distance function $d(\cdot,\cdot)$ measures the relative entropy between quantum states [14]. The minimum ball of the set of balls is unique, thus the *circumcenter* $\mathbf{c}^*$ of the set of quantum states is:

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} F_B\left(\mathbf{c}\right). \tag{41}$$

In *Fig. 17* we illustrated the smallest enclosing ball of balls in the quantum space.

At the end of our algorithm, the radius r^* of the smallest enclosing ball $\mathcal{B}^*$ with respect to the quantum informational distance is equal to $\min_{\sigma \in \mathcal{S}(\mathbb{C}^2)} \max_{\rho \in \mathcal{S}(\mathbb{C}^2)} D\left(\mathcal{L}(\rho) \| \mathcal{L}(\sigma)\right)$.

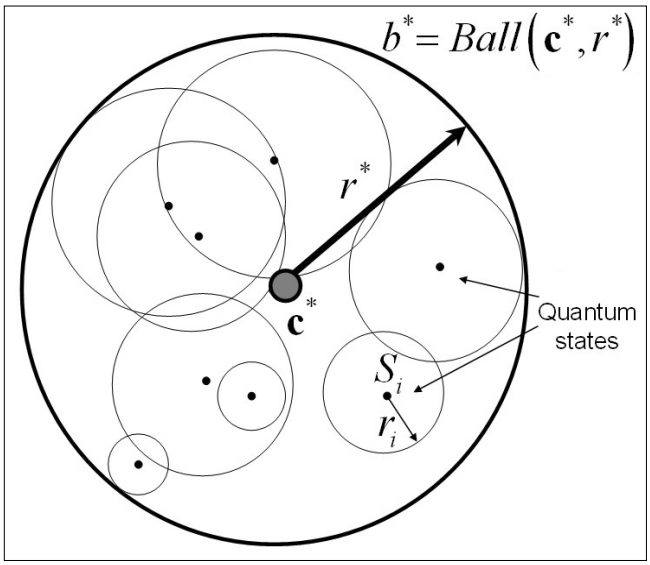


Figure 17.
The smallest enclosing ball of a set of balls in the quantum space

The *security* of the quantum channel is determined by our geometrical model with assumptions $r^* > r^*_{Eve}$, and $r^* \leq r^*_{Eve}$, as we have defined it in Eq. (37) and (38).

Finally, the approximated value of the fidelity parameter F_{Eve}, can be expressed as:

$$F_{Eve} = \langle \psi |^{(in)} \rho^{(out)} | \psi \rangle^{(in)} = \frac{1}{2}(1+r), \qquad (42)$$

where r can be derived from the quantum informational theoretical radius r^* by $r^* = 1 - S(r)$, where S is the von Neumann entropy.

In *Fig. 18* we compared the smallest quantum informational ball and the ordinary Euclidean ball (dashed-line) for a random set S of mixed quantum states. As we can conclude, the quantum states ρ_1, ρ_2 and ρ_3, which determine the Euclidean smallest enclosing ball, differ from the states of the quantum informational ball.

Figure 18.
The maximal distance states of the smallest balls are differing for quantum informational distance and Euclidean distance

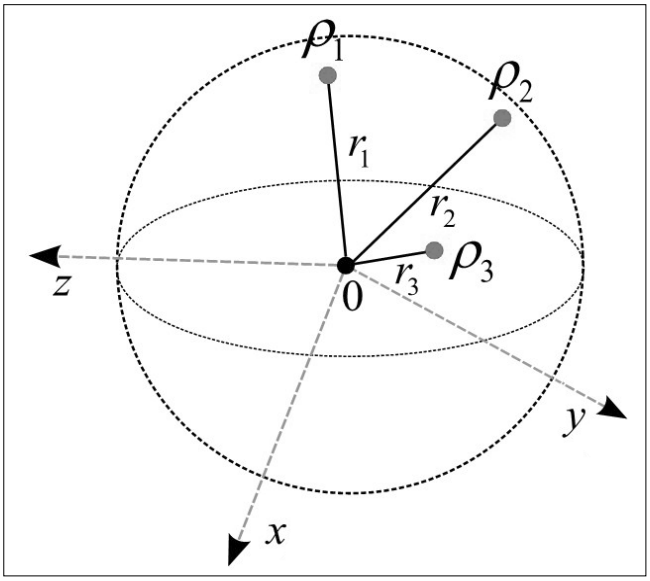
6.3 The computational complexity of the proposed algorithm

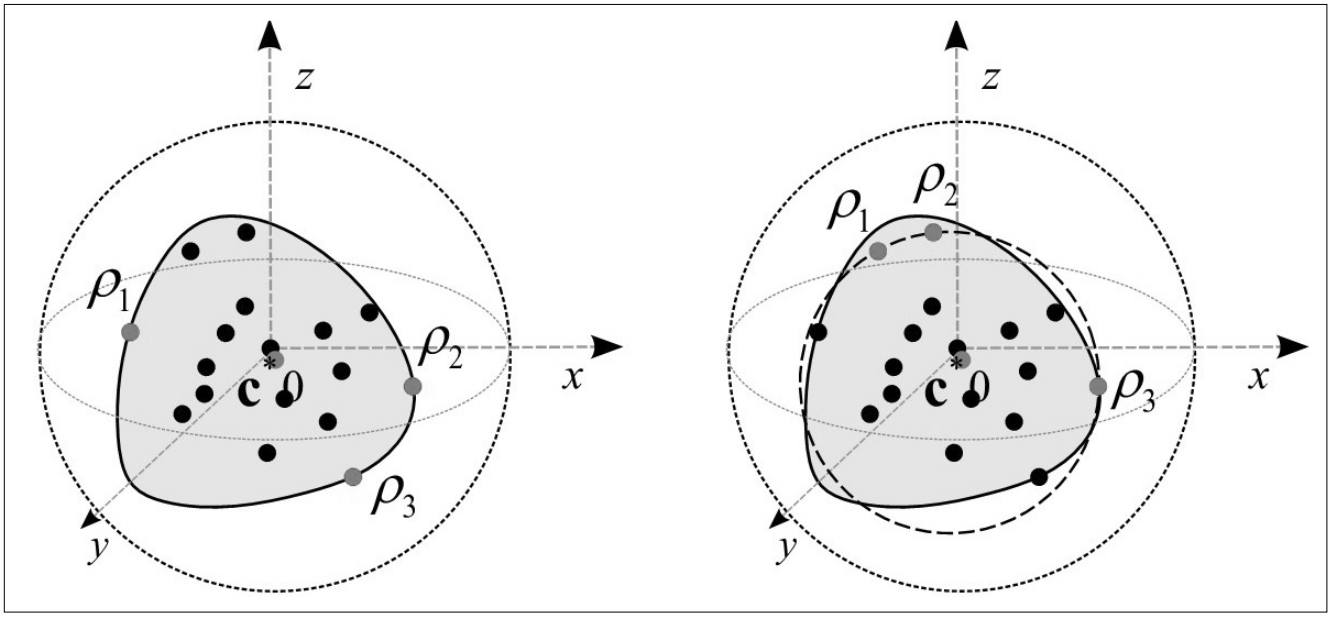
The quantum relative entropy based algorithm at the i-th iteration gives an $O(1+\sqrt{i})$-approximation of the real *circumcenter*, thus to get an $(1+\varepsilon)$ approximation, our algorithm requires

$$\mathcal{O}\left(\frac{dn}{\varepsilon^2}\right) = \mathcal{O}\left(\frac{d}{\varepsilon^2}\frac{1}{\varepsilon}\right) = \mathcal{O}\left(\frac{d}{\varepsilon^3}\right) \qquad (43)$$

time, by first sampling $n = 1/\varepsilon$ points. Based on the computational complexity of the smallest enclosing ball, the $(1+\varepsilon)$ approximation of the fidelity of the eavesdropper cloning machine can be computed in $O(d/\varepsilon^2)$ time. As *future work*, we would like to improve our method to get an $O(d/\varepsilon)$ time $(1+\varepsilon)$-approximation algorithm in quantum space.

Figure 19. Mixed quantum states in the Bloch ball





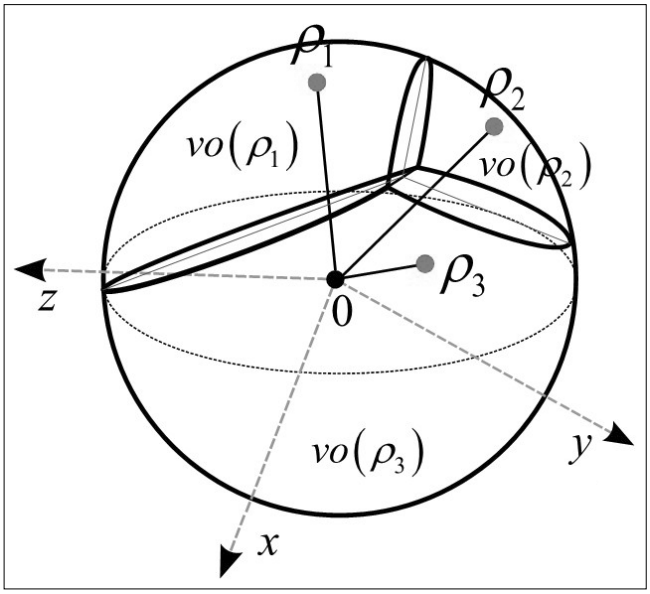


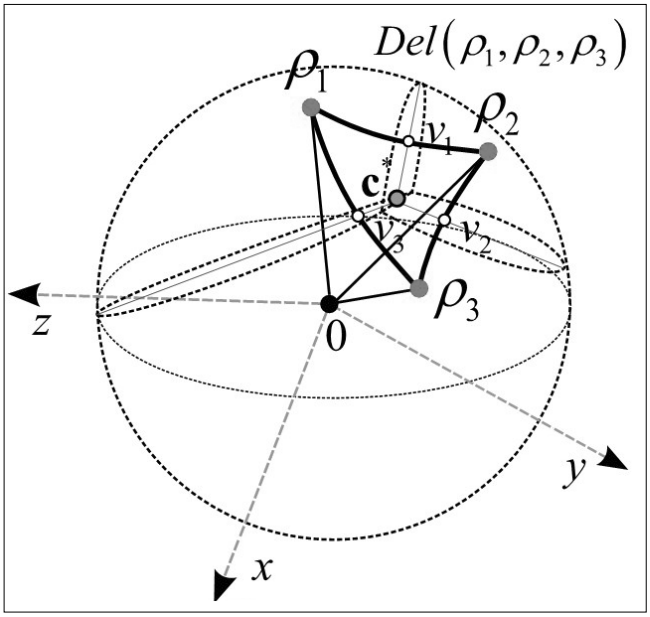
Figure 20.
Voronoi cells of quantum states on the Bloch ball

7. An illustrative example

Finally, we summarize the steps of our quantum relative entropy based cloning machine detecting algorithm. In the next example, we compute the smallest enclosing quantum information ball for three mixed quantum states. In *Fig. 19*, the *mixed* quantum states in the Bloch ball denoted by ρ_1, ρ_2 and ρ_3. The radius of the quantum states are denoted by r_1, r_2 and r_3.

First, we determine the Voronoi cells for the mixed quantum states. The Voronoi cells in the Bloch ball are denoted by $vo(\rho_1)$, $vo(\rho_2)$, and $vo(\rho_3)$. The distance between quantum states calculated with respect to quantum relative entropy.

Figure 21.
Delaunay triangle with respect to quantum informational distance



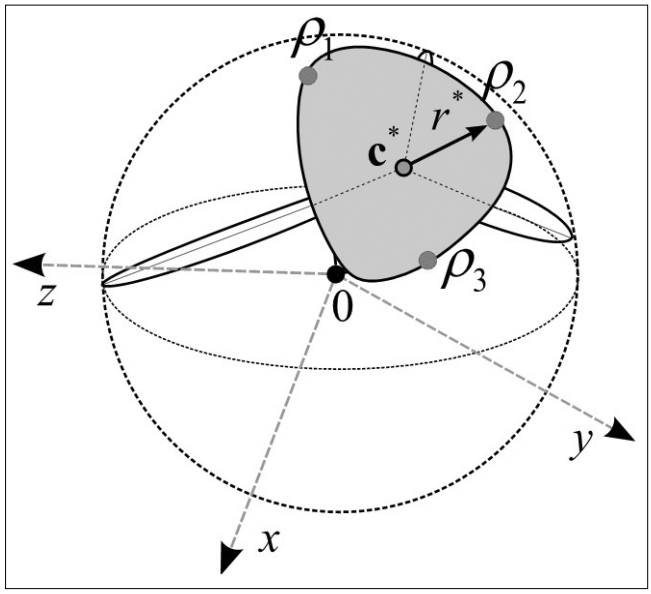
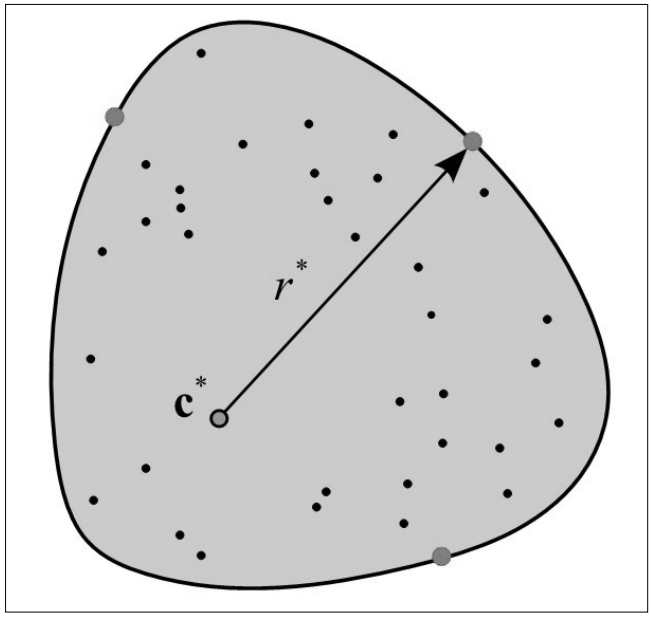


Figure 22.
The smallest enclosing quantum informational ball and its radius

In the next phase we compute the Delaunay triangulation with respect to quantum relative entropy. The quantum informational Delaunay triangle is distorted, according to the distance properties of quantum relative entropy.

In *Fig. 21* the quantum Delaunay triangle is denoted by $Del(\rho_1, \rho_2, \rho_3)$. The bisector points between the quantum states with respect to quantum relative entropy denoted by points v_1, v_2 and v_3. The bisectors intersect the center of the smallest quantum informational ball, denoted by $\mathbf{c}^*$. Finally, we get the radius r^* of the smallest enclosing quantum informational ball, centered at point $\mathbf{c}^*$. The distorted structure of the smallest enclosing quantum relative entropy ball is well observable in *Fig. 22*.

Figure 23.
The smallest enclosing quantum informational ball inside the Bloch sphere



In *Fig. 23* we show an example for a two-dimensional smallest enclosing quantum informational ball. This quantum relative entropy ball is a deformed ball, thus our approximation algorithm is tailored for quantum informational distance.

The center c^* of the smallest enclosing quantum informational ball differs from the center of an Euclidean ball.

In this given example, the center point is $\mathbf{c}^*(x,y) = (0.3287, 0.3274)$, and the radius r^* of the smallest enclosing quantum informational ball is $r^* = 0.4907$.

8. Conclusions

We showed a fundamentally new approach to measure the information theoretical impacts of quantum cloning on the private quantum channel. In our analysis the fidelity of the eavesdropper's cloning machine is numerically computed by tessellation on the Bloch sphere. In classical computational geometry Delaunay triangulations has an important role [4]. Using Delaunay tessellation on the Bloch sphere, the quantum *space* can be divided very efficiently.

We showed, that we can use efficiently *Laguerre* diagrams on the Bloch sphere, since the *Laguerre* diagrams are defined both on mixed and pure quantum states. We presented a novel approach to compute the relative quantum entropy, using an approximation method for the smallest enclosing ball of balls using core-sets. We presented an effective approximation algorithm to compute the informational fidelity using quantum information balls, equipped with quantum relative entropy as a distance measure.

As future work we would like to present a more effective algorithm to compute the eavesdropper's cloning machine, and make a deep study on our algorithm's convergence rate.

Authors

LÁSZLÓ GYÖNGYÖSI, Ph.D Student since 2008, Budapest University of Technology and Economics. He received the M.Sc. degree in Computer Science with Honors from the Technical University of Budapest in 2008. His research interests are in Quantum Computation and Communication, Quantum Cryptography and Quantum Information Theory.

SÁNDOR IMRE was born in Budapest in 1969. He received the M.Sc. degree in Electronic Engineering from the Budapest University of Technology (BME) in 1993. Next he started his Ph. D. studies at BME and obtained dr. univ. degree in 1996, Ph.D. degree in 1999 and DSc degree in 2007. Currently he is carrying his teaching activities as Head of the Dept. of Telecommunications of BME. He was invited to join the Mobile Innovation Centre of BME as R&D director in 2005. His research interest includes mobile and wireless systems, quantum computing and communications. Especially he has contributions on different wireless access technologies, mobility protocols and reconfigutable systems.

Acknowledgments

The authors would like to thank Dr. Katalin Friedl for useful discussions.

References

[1] L. Gyöngyösi, S. Imre,
Computational Geometric Analysis of Physically Allowed Quantum Cloning Transformations for Quantum Cryptography,
4th WSEAS International Conference on Computer Engineering and Applications, Section on Quantum Computing,
University of Harvard, Cambridge (Massachusetts), USA, 2010.

[2] S. Imre, F. Balázs,
Quantum Computing and Communications – An Engineering Approach,
Published by John Wiley and Sons Ltd, The Atrium, England, ISBN 0-470-86902-X (283 pages), 2005.

[3] L. Gyöngyösi, S. Imre,
Geometrical Estimation of Information Theoretical Impacts of Incoherent Attacks for Quantum Cryptography, International Review of PHYSICS, Print ISSN: 1971-680X, 2010.

[4] P.W. Lamberti, A.P. Majtey, A. Borras, M. Casas, A. Plastino,
Metric character of the quantum Jensen-Shannon divergence.
Physical Review A (Atomic, Molecular, and Optical Physics), 77(5):052311, 2008.

[5] F. Aurenhammer, R. Klein,
Voronoi Diagrams.
In J. Sack and G. Urrutia (Eds.): Handbook of Computational Geometry, Chapter V, pp.201–290. Elsevier Science Publishing, 2000.

[6] J.-D. Boissonnat, C. Wormser, M. Yvinec,
Curved Voronoi diagrams.
In J.-D.Boissonnat and M. Teillaud (Eds.): Effective Computational Geometry for Curves and Surfaces, pp.67–116, Mathematics and Visualization, Springer-Verlag, 2007.

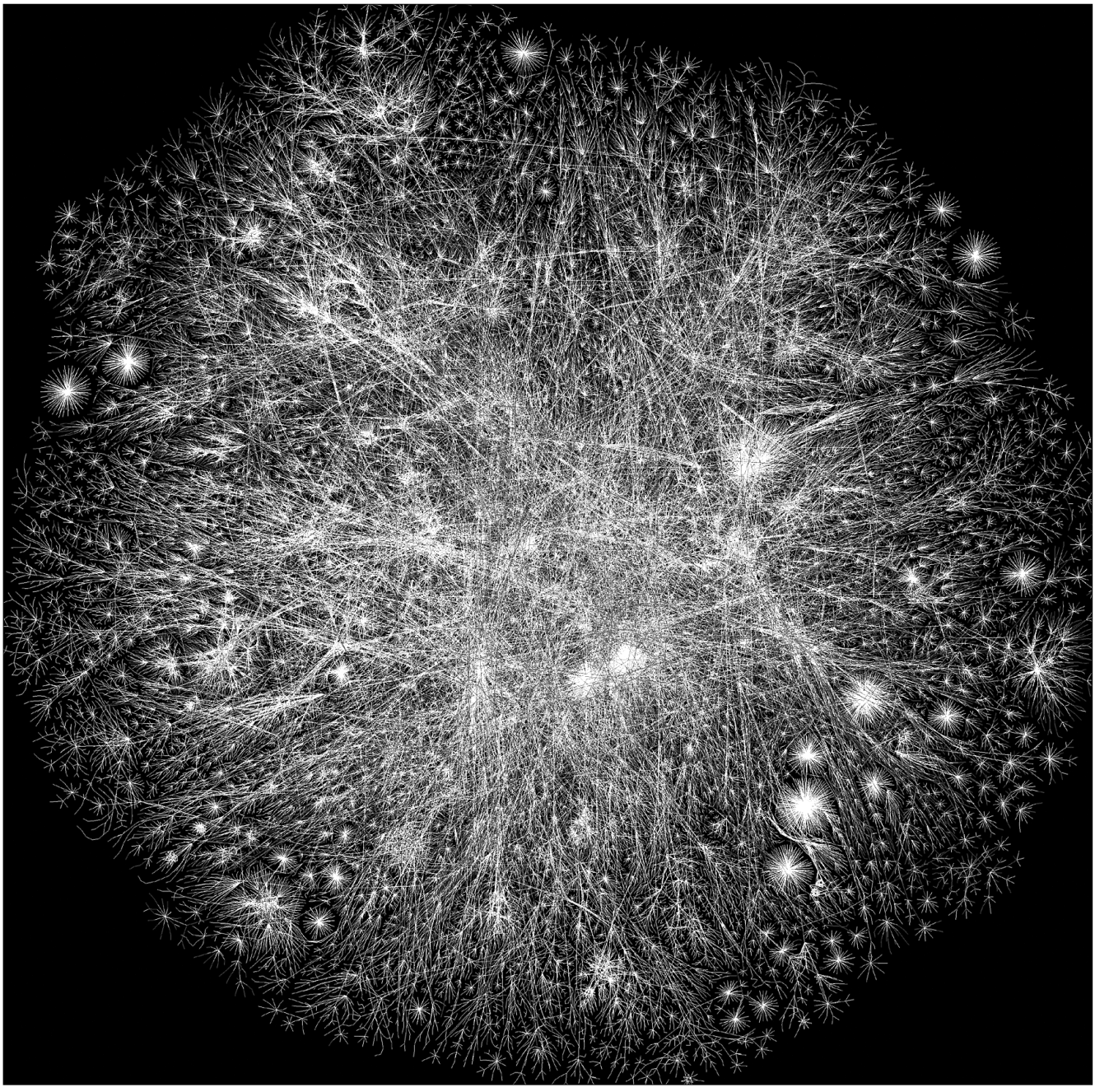
[7] M.A. Nielsen, I.L. Chuang,
Quantum Computation and Quantum Information.
Cambridge University Press, 2000.

[8] Cerf, N.J., M. Bourennane, A. Karlsson, N. Gisin,
Security of Quantum Key Distribution Using d-Level Systems,
Phys. Rev. Lett. 88, 127902., 2002.

[9] D'Ariano, G.M., C. Macchiavello,
Optimal phase-covariant cloning for qubits and qutrits,
Phys. Rev. A 67, 042306., 2003.

[10] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, V. Scarani,
Device-independent security of quantum cryptography against collective attacks,
Phys. Rev. Lett. 98, 230501., 2007.

[11] M. Badoiu, K.L. Clarkson,
Smaller core-sets for balls.
In Proceedings 14th ACM-SIAM Symposium on Discrete Algorithms, pp.801–802, 2003.
[12] R. Panigrahy,
Minimum enclosing polytope in high dimensions.
CoRR, cs.CG/0407020, 2004.
[13] M. Badoiu, S. Har-Peled, P. Indyk,
Approximate clustering via core-sets.
In Proceedings 34th ACM Symposium on Theory of Computing, pp.250–257, 2002.
[14] P. Kumar, J.S.B. Mitchell, A. Yıldırım,
Computing core-sets and approximate smallest enclosing hyperspheres in high dimensions.
In Algorithm Engineering and Experimentation, LNCS, Springer-Verlag, pp.45–55, 2003.
[15] M. Curty, N. Lütkenhaus,
Phys. Rev. A 69, 042321., 2004.
[16] S.-I. Amari, H. Nagaoka,
Methods of Information Geometry.
Translations of Mathematical Monographs, 191., AMS, Oxford University Press, Oxford, 2000.
[17] L. Gyöngyösi, S. Imre,
Quantum Divergence based Quantum Channel Security Estimation,
N2S'2009 Int. Conf. on Network and Service Security, Section on Quantum Cryptography and QKD, IFIP TC6 WG, IEEE France, Paris, June 2009.



Call for Papers

Prospective authors are invited to submit original research papers for publication in the upcoming issues of our Infocommunications Journal.

Topics of interests include the following areas:

Data and network security
Digital broadcasting
Infocommunication services
Internet technologies and applications
Media informatics
Multimedia systems
Optical communications
Society-related issues
Space communications
Telecommunication software
Telecommunications economy and regulation
Testbeds and research infrastructures
Wireless and mobile communications

Theoretical and experimentation research results achieved within the framework of European ICT projects are particularly welcome.

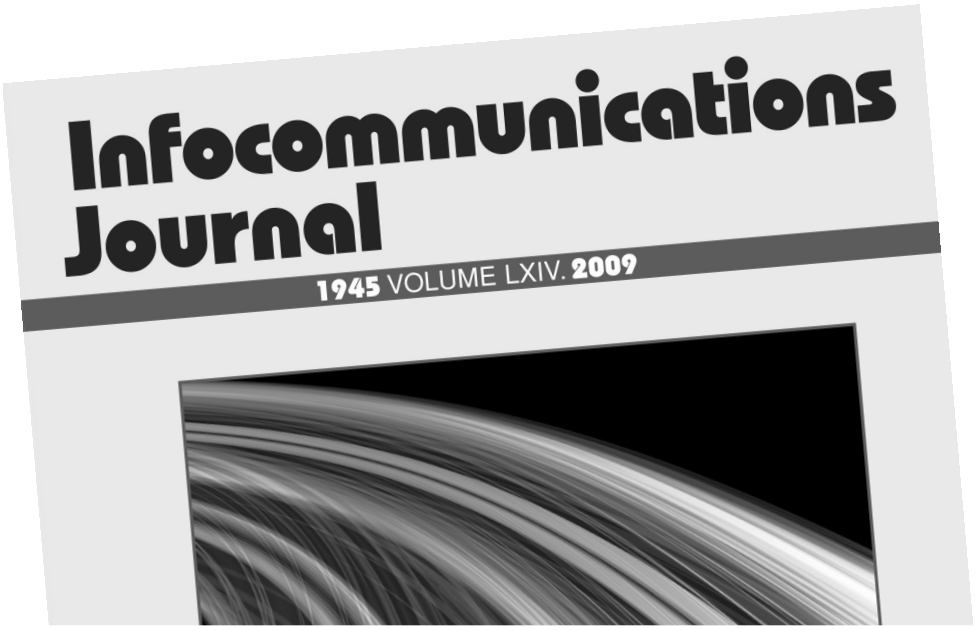
From time to time we publish special issues and feature topics so please follow the announcements. Proposals for new special issues and feature topics are welcome.

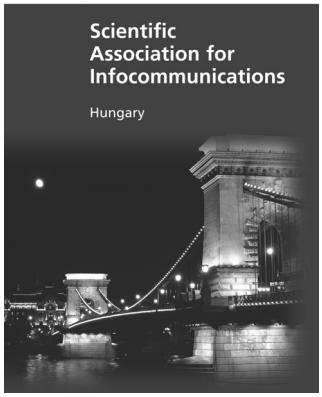
Our journal is currently published quarterly and the editors try to keep the review and decision process as short as possible to ensure a timely publication of the paper, if accepted.

As for manuscript preparation and submission, please follow the guidelines published on our website:
http://www.hiradastechnika.hu/for_our_authors

Authors are requested to send their manuscripts via electronic mail (preferably) or on a CD by regular mail to the Editor-in-Chief:

Csaba A. Szabó
Dept. of Telecommunications, Budapest University of Technology and Economics
2 Magyar Tudósok krt., Budapest 1117, Hungary
E-mail: szabo@hit.bme.hu





GUIDELINES FOR OUR AUTHORS

Format of the manuscripts

Manuscripts shall be submitted in MS Word. Authors are kindly requested not to use the formatting tools of Word such as automatic title formats etc. The parts of the papers should be numbered in a hierarchical way. Do not use more than three levels of hierarchy, ideally only two. Use 12 pts font size. Do not use two-column format.

Length of the manuscripts

The length of papers in printed format is ideally 4-6 journal pages which correspond to 12,000-16,000 characters (incl. spaces). Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of Introduction (the part of paper numbered by "1"), and Summary (the last numbered part), and several Sections in between. The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an Abstract and a few keywords.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript by their full names and affiliations (without postal addresses). An e-mail address of the authors (or at least one of them) is also included in the heading of the paper. No degrees or other titles of the authors are given.

One of the authors should provide his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue

Format of a reference:

[11] Boggs, S. A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, Vol. ET-19, No. 2, pp.87–92., Apr. 1984.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."

When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to send their manuscripts via electronic mail or on a CD by regular mail to the Editor-in-Chief:

Csaba A. Szabó
Department of Telecommunications, Budapest University of Technology and Economics
2 Magyar Tudósok krt., Budapest 1117, Hungary
E-mail: szabo@hit.bme.hu

Summaries • of the papers published in this issue

Fast-response inter-vehicle communications

Keywords: ITS, inter-vehicle communications, safe driving, media access control (MAC), CSMA, CDMA

Safe driving support is one of the most attractive and important applications of an inter-vehicle communication systems. Real-time and reliable exchange of status information on such as vehicle location, speed, sudden braking etc., among vehicles, is a key to offering prompt warnings to drivers in order to avoid fatal traffic accidents. We have proposed a novel media access control (MAC) scheme based on code division multiple access (CDMA) technology, which offers fast response and high packet delivery ratio to meet the above requirements. This paper introduces the proposed scheme, performance evaluation by simulation, and prototyping for field experiment. It also mentions future studies.

Fast handover and fast failover mechanisms based on cross-layer collaboration among the link layer, the network layer and the transport layer

Keywords: cross-layer collaboration, fast handover, SCTP

This paper describes a fast handover mechanism in the network layer called L3-FHOX and a fast failover mechanism in the transport layer called SCTPfx. Both mechanisms are based on a cross-layer architecture called CEAL. CEAL enables the control information exchange between layers in a node with keeping the layering structure. We implemented both mechanisms in Free BSD. The entire handover time in L3-FHOX is approximately 10 msec plus the RTT between the mobile node and its location management server while the normal handover procedure in IPv6 takes more than 1 second. The failover time of SCTPfx is 122 usec plus the RTT between the two end nodes while the normal failover procedure in SCTP takes more than 31 seconds.

Developing next generation web as collaboration media

Keywords: Global Knowledge Grid, SOA, knowledge web

Beyond the semantic web era, digital contents are not the key assets of the web any more, but the knowledge acquired from the web contents plays an important role for users to understand situations, make decisions and/or take actions. Considering knowledge as a key asset, it is worthwhile to renovate the traditional content-centered framework of web lifecycle so as to be a knowledge-centered one. We introduce the Global Knowledge Grid, a distributed knowledge service computing environment based on the service-oriented architecture (SOA). It facilitates data-centric collaborations of different types of knowledge services: knowledge capture, knowledge association, and knowledge provision. An evolving network of knowledge is built by interconnecting heterogeneous knowledge sites over the global knowledge grid.

General distributed economic framework for dynamic spectrum allocation

Keywords: distributed spectrum management, dynamic spectrum allocation, game theory

We present our novel dynamic spectrum sharing management scheme in which the allocation and the pricing of radio frequency bands are performed in a distributed manner. We focus on a non-cooperative setting where the frequency leasers act for their own benefit, and we design the system policies in order to assure that the resulted allocation yields high spectrum utilization. We provide scalable and incentive-compatible allocation and pricing mechanisms on our physical radio interference model. Our evaluations prove that our distributed dynamic spectrum allocation scheme imposes high charges on frequency leasers that exclude others by their presence in terms of interference; therefore it is a suitable approach to reach efficient and flexible spectrum utilization.

ETOMIC advanced network monitoring system for future Internet experimentation

Keywords: network monitoring, ETOMIC, traffic, GPS

ETOMIC is a network traffic measurement platform with high precision GPS-synchronized monitoring nodes. The infrastructure is publicly available to the network research community, supporting advanced experimental tecniques by providing high precision hardware equipements and a Central Management System. Researchers can deploy their own active measurement codes to perform experiments on the public Internet. These features make ETOMIC as one of the experimental facitlities that support the design, development and validation of novel experimental techniques for the future Internet. In this paper we focus on the improved capabilities of the management system, the recent extensions of the node architecture and the accompanying database solutions.

A study of prosodic variability methods in a corpus-based unit selection text-to-speech system

Keywords: prosody variation, corpus-based TTS

This paper introduces the implementation and evaluation of a method to increase the prosodic variability of synthesized speech. Different generated prosody target versions were tested in a Hungarian corpus-based unit selection Text-To-Speech (TTS) system: the baseline prosody of the synthesizer, a rule-based prosody target and the prosody of the new method. It is based on F0 database templates which are derived from natural sentence corpora. Our method was tested in a Hungarian system, and it can be extended to other European languages with fixed (e.g. Finnish) and varying stress.

Quantum information theoretical based geometrical representation of eavesdropping activity on the quantum channel

Keywords: quantum cryptography, quantum cloning, quantum informational distance

Quantum cryptography is an emerging technology that may offer new forms of security protection, however the quantum cloning based attacks against the protocol will play a crucial role in the future. According to the no-cloning theorem, an eavesdropper on the quantum channel can not copy perfectly the sent quantum states. In our method we use quantum relative entropy as an informational distance between quantum states. We show a geometrical approach to analyze the security of quantum cryptography, based on quantum relative entropy and Laguerre Delaunay triangulation on the Bloch sphere. We present a basically new method to derive quantum relative entropy based Delaunay tessellation on the Bloch ball and to compute the radius of smallest enclosing ball of balls to detect eavesdropping activity on the quantum channel.

Infocommunications Journal

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1055 Budapest, Kossuth Lajos tér 6-8, Room: 422
Mail Address: 1372 Budapest Pf. 451. Hungary
Phone: +36 1 353 1027, Fax: +36 1 353 0451
E-mail: info@hte.hu
Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Telecommunications
Tel.: +36 1 463 3261, Fax: +36 1 463 3263
E-mail: szabo@hit.bme.hu

Subscription rates for foreign subscribers:
4 issues 50 USD, single copies 15 USD + postage

Publisher: PÉTER NAGY • Manager: ANDRÁS DANKÓ

HU ISSN 2061-2079 • Layout: MATT DTP Bt. • Printed by: Regiszter Kft.