ICDT 2013

The Eighth International Conference on Digital Telecommunications

ISBN: 978-1-61208-262-2

April 21 - 26, 2013

Venice, Italy

ICDT 2013 Editors

Calin Vladeanu, Politehnica University of Bucharest, Romania

Alex Galis, University College London, UK

ICDT 2013

Foreword

The Eighth International Conference on Digital Telecommunications (ICDT 2013), held between April 21st-26th, 2013 in Venice, Italy, continued a series of special events focusing on telecommunications aspects in multimedia environments. The scope of the conference was to focus on the lower layers of systems interaction and identify the technical challenges and the most recent achievements.

High quality software is not an accident; it is constructed via a systematic plan that demands familiarity with analytical techniques, architectural design methodologies, implementation polices, and testing techniques. Software architecture plays an important role in the development of today's complex software systems. Furthermore, our ability to model and reason about the architectural properties of a system built from existing components is of great concern to modern system developers.

Performance, scalability and suitability to specific domains raise the challenging efforts for gathering special requirements, capture temporal constraints, and implement service-oriented requirements. The complexity of the systems requires an early stage adoption of advanced paradigms for adaptive and self-adaptive features.

Online monitoring applications, in which continuous queries operate in near real-time over rapid and unbounded "streams" of data such as telephone call records, sensor readings, web usage logs, network packet traces, are fundamentally different from traditional data management. The difference is induced by the fact that in applications such as network monitoring, telecommunications data management, manufacturing, sensor networks, and others, data takes the form of continuous data streams rather than finite stored data sets. As a result, clients require long-running continuous queries as opposed to one-time queries. These requirements lead to reconsider data management and processing of complex and numerous continuous queries over data streams, as current database systems and data processing methods are not suitable.

We take here the opportunity to warmly thank all the members of the ICDT 2013 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICDT 2013. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICDT 2013 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICDT 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of digital communications.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the charm of Venice, Italy.

ICDT Advisory Committee:

Constantin Paleologu, University Politehnica of Bucharest, Romania
Tomohiko Taniguchi, Fujitsu Laboratories Limited, Japan
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Abdulrahman Yarali, Murray State University, USA
Michael Grottke, University of Erlangen-Nuremberg, Germany

Javier Del Ser Lorente, TECNALIA RESEARCH & INNOVATION - Zamudio, Spain
Saied Abedi, Fujitsu Laboratories of Europe Ltd. (FLE), UK
Gerard Damm, Alcatel-Lucent, USA
Dan Romascanu, Avaya, Israel
Klaus Drechsler, Fraunhofer Institute for Computer Graphics Research IGD - Darmstadt, Germany

ICDT 2013

Committee

ICDT Advisory Chairs

Constantin Paleologu, University Politehnica of Bucharest, Romania
Tomohiko Taniguchi, Fujitsu Laboratories Limited, Japan
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Abdulrahman Yarali, Murray State University, USA
Michael Grottke, University of Erlangen-Nuremberg, Germany
Javier Del Ser Lorente, TECNALIA RESEARCH & INNOVATION - Zamudio, Spain
Saied Abedi, Fujitsu Laboratories of Europe Ltd. (FLE), UK
Gerard Damm, Alcatel-Lucent, USA
Dan Romascanu, Avaya, Israel
Klaus Drechsler, Fraunhofer Institute for Computer Graphics Research IGD - Darmstadt, Germany

ICDT 2013 Technical Program Committee

Saied Abedi, Fujitsu Laboratories of Europe Ltd. (FLE) - Middlesex, UK
Bilal Al Momani, Cisco Systems, Inc., Ireland
Antonio Marcos Alberti, INATEL - Instituto Nacional de Telecomunicações, Brazil
Abdullah M. Alnajim, Qassim University, Saudi Arabia
Maria Teresa Andrade, FEUP / INESC Porto, Portugal
Iosif Androulidakis, MPS Jozef Stefan, Slovenia
Regina B. Araujo, Federal University of São Carlos, Brazil
Khaled Assaleh, American University of Sharjah, United Arab Emirates
Andreas Aurelius, Acreo AB, Sweden
Anteneh Ayanso, Brock University, Canada
Francisco Barcelo-Arroyo, Technical University of Catalonia, Spain
Ilija Basicevic, University of Novi Sad, Serbia
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Abdelouahab Bentrcia, King Saud University – Riyadh, Kingdom of Saudi Arabia
Andrzej (Andrew) Borys, University of Technology and Life Sciences (UTP) – Bydgoszcz, Poland
Andi Buzo, University Politehnica of Bucharest, Romania
Lee-Ming Cheng, City University of Hong Kong, Hong Kong
Alberto Coen-Porisini, Università degli Studi dell'Insubria – Varese, Italy
Doru Constantin, University of Pitesti, Romania
Gerard Damm, Alcatel-Lucent, France
Klaus Drechsler, Fraunhofer-Institut für Graphische Datenverarbeitung IGD - Darmstadt, Germany
Roger Pierre Fabris Hoeffel, Federal University of Rio Grande do Sul, Brazil
Peter Farkas, FEI STU – Bratislava, Slovakia
Christophe Feltus, Public Research Centre Henri Tudor, Luxembourg
Gerardo Fernández-Escribano, University of Castilla-La Mancha, Spain
Mário Ferreira, University of Aveiro, Portugal

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Ontology for IP Telephony Networks

Ilija Basicevic, Miroslav Popovic, and Nenad Cetic
Department of Computer Engineering
Faculty of Technical Sciences
Novi Sad, Serbia
Email: ilibas@uns.ac.rs, miroslav.popovic@rt-rk.com, nenad.cetic@rt-rk.com

***Abstract*—Ontologies have already been used in computer science research in different fields. In this paper, we apply the concept to the area of IP telephony. The paper presents ontology for development of software for IP telephony networks. It is based on the analysis of SIP and H.323 protocols, as dominant signaling protocols in VoIP telephony of today. The ontology we developed covers both static aspects (structure of the network) and dynamic aspects (most importantly telephony sessions, but also other types of associations). The structure of the telephony network is modeled with classes that represent different types of nodes in the network: end user terminals and various types of infrastructure nodes. The dynamic aspects of network operation are modeled with a set of classes, the most important being those that represent different associations that are established and terminated in the network. The realized ontology can be used in development of frameworks for telephony applications, and for specification of common data format used by cooperating telephony applications.**

Keywords-Ontologies; Network communications; Internet Applications; VoIP; IP telephony; SIP; H.323

I. INTRODUCTION

Ontologies have already entered and claimed an important place in several fields of computing science. In this article, we present application of ontology in the field of IP telephony. The ontology is focused primarily on signaling aspect, but covers all aspects of operation of IP telephony networks. Computer telephony is a result of merging of computing and telephony domain. Traditionally the two domains were separated, and telephony applications restricted to circuit switched networks, e.g. Public Switched Telephone Network (PSTN). With the intensive development of Internet in the last three decades, it became logical that IP network (although a packet switched network) could be used for personal communications with acceptable quality, which gave birth to Voice over IP (VoIP). Today VoIP is used in great extent - for example majority of new installed Private Branch Exchange (PBX) lines are VoIP. In most of the published papers dealing with application of ontologies in telephony domain, only segments or certain details of used ontologies are presented. This paper provides a systematic overview of the ontology we developed. The overview begins with the development process. The method that has been applied in development of the ontology is described. Next, two important class hierarchies that embody structure and dynamics of IP telephone networks are presented, visually and in text. Also, the paper tries to position the developed ontology in the context of existing work in this area.

Section II describes shortly the method applied in development of the ontology. In Section III are presented classes dealing with the structure of the telephony network, which is a very important static aspect of the network functionality. The prevailing formats of protocol messages in IP telephony systems are shortly described in section IV. Section V presents classes that are responsible for dynamic aspect of the operation of telephone network. In the section VI are shortly presented some published papers on the use of ontologies in this application domain. Section VII contains an analysis on possible applications of the ontology, including the one already realized. Section VIII contains concluding remarks.

II. THE DEVELOPMENT OF IP TELEPHONY ONTOLOGY

The ontology that is presented in this paper has been developed using a combination of deductive and inductive approaches. In the conception phase of development, the deductive approach has been followed. As in other fields, published standards can be used as one of the starting points in definition of ontologies. Number of telephony standards is quite large. Each of many telephony standards in a way establishes its own taxonomy and usually defines most important terms. Thus the existing standards: most notably Session Initiation Protocol (SIP) [1], and in less extent H.323 [2] influenced very much the conception and elaboration phases in the development of the ontology. The importance we gave to SIP is due to its dominant place in the market.

The analysis of standards and telephony applications resulted in an informal description of ontology, and has been an input for the elaboration phase. The result of elaboration phase has been expressed in Unified Modeling Language (UML) and Java. The ontology has been the basis for the development of framework [3]. By building the framework we tested and further developed the ontology (see Fig. 1). In this manner we combined the deductive and inductive approach. This approach is similar to Forward-Lockstep Build-Test model of Helix-Spindle [4].

 ISBN: 978-1-61208-262-2

III. STRUCTURE OF TELEPHONY NETWORK

Structure of telephony network is in general case a graph consisting of nodes and links. Identification property of any network node is its network address. There are several types of nodes. The most important classification is into end points, directly accessed by end users who use the services of network, and infrastructure points, ”invisible” to end-users. For that reason class NetPoint has the following subclasses: EndPoint, RoutingPoint, LocServPoint and ConferencePoint (see Fig. 2). ConferencePoint presents infrastructure nodes that support conferencing feature (conference servers and media mixers). The majority of infrastructure nodes in a typical network are responsible for routing. Those are modeled by RoutingPoint. Examples of RoutingPoint are SIP [1] proxy and H.323 [2] gatekeeper. Essential property of a routing point is the administrative policy database and a routing table. Routing relies on address mapping which is the responsibility of LocServPoint (location service or registrar in SIP terminology). Network is administratively separated into segments called domains in SIP and zones in H.323. Usually, there is one location service and one routing point per telephony domain. Location service keeps mappings for endpoints in that domain. Essential property of a location service is the table of address mappings that map telephony addresses to network layer addresses. Fig. 3 presents a typical case of IP telephony communication.

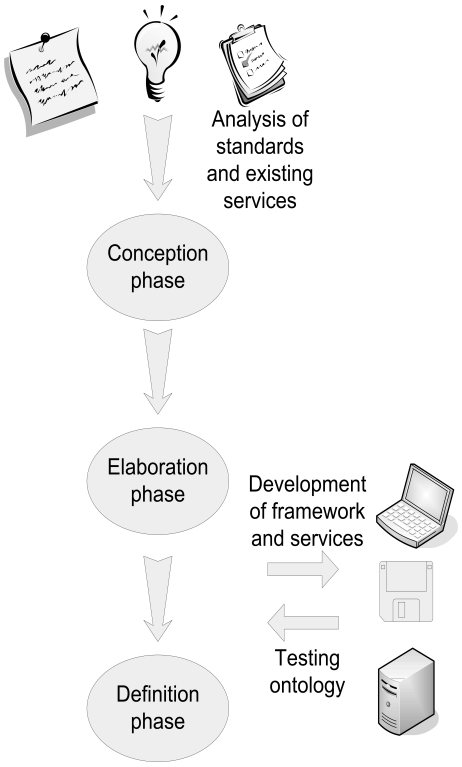


Figure 1. The development process

It is usually referred to as SIP trapezoid. There are two EndPoint instances (SIP user agents) shown in Fig. 3, one in each domain. In both domains there is RoutingPoint and LocServPoint - in the SIP case, the two (SIP proxy and SIP registrar) are often collocated.

An important class of telephony applications is the class of network side applications [5]- applications with logic residing at infrastructure nodes. For that reason ontology contains another class, subclassed from RoutingPoint RoutingPointExt. In SIP networks, network side applications are often built using back-to-back user agents which is a network element containing SIP client and server side user agent stack, controlled by one application. Essential property of endpoint is its configuration and the list of supported capabilities. The important property of EndPoint class is the active user currently logged in at the end point, modeled by the EndUser class. The characteristics of telephony services are such that there is only zero or one user active at any endpoint at any moment of time. The identification property of this class is the end user address, specified using the address scheme supported by telephony protocol of the network in a case. Examples are sip and sips address schemes, supported by SIP protocol or E.164 directory numbers [6], supported by H.323. While functions of endpoint are obvious - it is a user interface to end users that supports required functions (login/logout, authentication, call control, input/output of multimedia stream), the functions of infrastructure elements are more sophisticated and include:

- routing of messages
- enforcement of administrative policies
- bandwidth control
- location service
- authorization and authentication.

It can be seen that authentication is a support function required at all network nodes, both end points and routing points.

IV. PROTOCOL MESSAGES

All telephony operations are carried out as exchange of messages. Rigid property of ProtocolMessage is the message syntax. H.323 uses ASN.1 [7] while SIP uses UTF.8. H.225 [8] messages are based on Q.931 [9] syntax. The procedural view of telephony provides us with a list of most important operations including location service, capabilities query, session-related operations (establishment, modification, termination) and authentication. Each of listed operations assumes parameters. In case of session establishment and session modification, that parameter is proposed session description (in SIP case, it is presented in Session Description Protocol (SDP) [10] notation and in H.323 it is part of OpenLogicalChannel H.245 message [11]). In case of location query, this parameter is user address. But, modern protocols allow specification of many parameters of secondary importance - for instance, ringing tone that is used

 ISBN: 978-1-61208-262-2

to signal new call can be specified in session establishment in SIP.

V. Dynamic characteristics of the network

A. Associations

Speaking about dynamic relationships that appear during the operation of telephony network, one of the most important classes in this segment is Association. This class is identified with addresses of participating sides and identification. We consider the association of two sides first. Association is a subclass of feature class (Feature). Feature is any function provided by the network to end users or a primitive function that goes unnoticed by end users, serving as a building block for complex functions (services) that end users invoke. One rigid property of Feature is OperationInterface. Subclasses of Association class are: Session – for telephony session, RegAssociation - for registration association, PubSubAssociation – for publish/subscribe association, and Conference (see Fig. 4). Session class models peer-to-peer telephony communication. The notion of telephony session implies existence of media (audio/video) stream that typically conveys speech or a video phone call, thus a rigid property of the Session class is SessionDescription which includes negotiated characteristics of audio/video stream and some other parameters. Registration association is the prerequisite of location service. In case of RegAssociation, rigid property is address mapping. Publish/subscribe associations are used for event publishing and are building blocks for different types of complex services. Rigid property of subscribe/publish association is the specific item of interest, in other words, the topic of information that is published. Conference is a feature that requires existence of infrastructure support (media mixers). Conference server is node which contains signaling sessions to participating nodes and media mixer contains media sessions to participating nodes. The two do not have to be collocated. Conference is a subclass of Association that presents association of $n > 1$ points. Rigid property of this class is the list of participants. In its early form, SIP dialog was about multimedia session only but, with the introduction of publish/subscribe mechanism, the notion of dialog has been extended. At this point, we do not recognize the need to incorporate such a composite class in ontology because all operations can be described using the existing association classes.

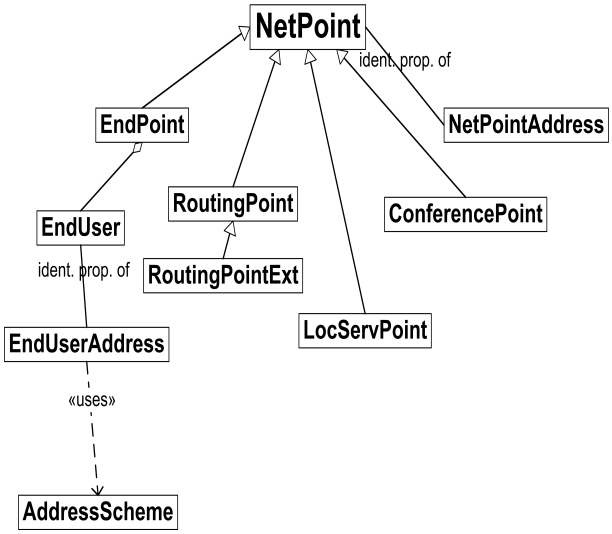


Figure 2. The NetPoint class and related classes

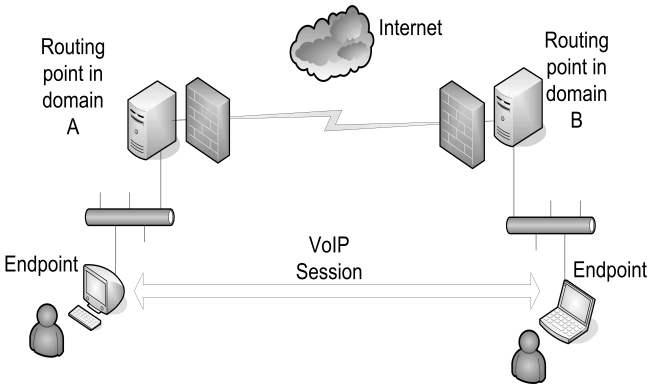


Figure 3. Typical IP Telephony communication path

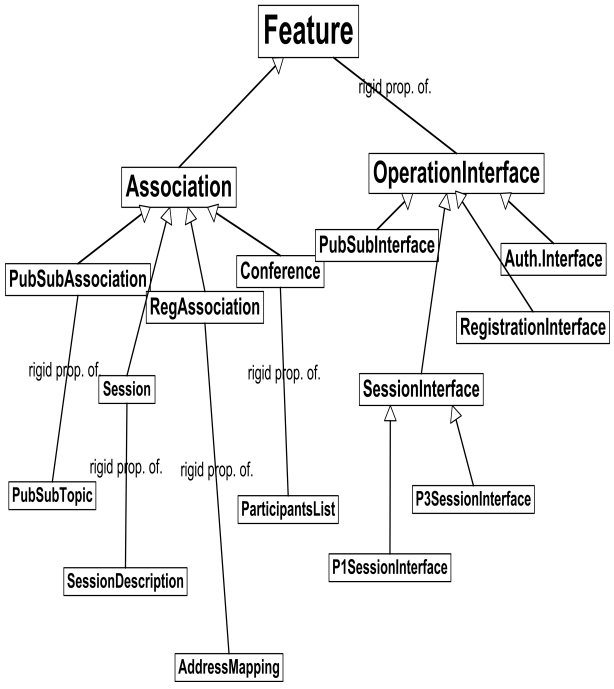


Figure 4. The Feature class and related classes

B. Interfaces

The telephony operations are provided by interfaces of our ontology. For each subclass of Association, there is

analogous interface, providing telephony operations typical for that class. The following are subclasses of OperationInterface: SessionInterface, PubSubInterface, RegistrationInterface and AuthenticationInterface. There are two instances of SessionInterface: P1SessionInterface and P3SessionInterface, one being for the first party, and the other for the third party call control of the telephony session, which is used in complex telephony applications - call centers etc.

Several assumptions are used. Terminate is a two-step transaction and Establish is done in three steps, as it is the case in three way handshake that is used in SIP protocol. The session negotiation supports exchange of three messages both in SDP [10] based offer-answer model and in H.245 [11] Open Logical Channel sequence - in the case of bidirectional channel. (In the case of two unidirectional channels H.245 assumes exchange of four messages in session negotiation.) Also, session modification is in our ontology realized in three steps reusing some of the messages of Establish transaction. This is again in accordance with SIP protocol and its re-INVITE mechanism. On the other hand, a protocol has been proposed where instead of two-step and three-step transactions, each message represents a transaction by itself [12]. Essential methods of PubSubInterface are: Subscribe, Accept and Notify. Subscribe/Publish association finds its use in many complex services. For example, in SIP environment, it is used for dissemination of presence and registration information, for publication of the result of transfer operation, and for third party control. In SIP protocol, the session transfer operation is complex feature that uses basic service session and publish/subscribe association as building blocks. For de-registration, we use SIP convention: register message with specific values of parameters (expiration time set to 0). For that reason, there is no explicit Deregister method in the interface.

C. Authentication

Session establishment as well as registration may require authentication. In our ontology, we assume stateless challenge-based solution as in SIP protocol. Authentication interface contains the following methods: AuthenticationRequired, AuthenticationRequest and AuthenticationResponse. However, the last two are not transferred as separate messages but as message parameters in messages like Establish or Register. For example, endpoint sends Establish with no user credentials. Routing point checks the administrative policy rules, determines that authentication is required and sends AuthenticationRequired back with the challenge included. Endpoint sends Establish again, this time with user credentials and computed response to the received challenge. This solution implies existence of some kind of public key infrastructure.

VI. Related Work

A discussion on the use of ontologies in telecom domain can be found in [21]. The work presented by Geneiatakis et al. [14] is an example of the use of ontology for VoIP security. It is the case of target-centric intrusion detection. Data model based on this ontology would provide a framework for cooperation of elements of distributed IDS system protecting the domain of telephony network. Also, cooperation of different IDS systems protecting different domains of telephony network can be based on the model. The ontology presented in this paper can also be used in security and, if extended with classes related to specific details of telephony protocols, it would cover wider range of detected attacks since it comprises a holistic view of IP telephony network.

Campbell et al. [15] presents specialization of APPEL policy language for call control in the scope of the Advanced Call Control Enhancing Network Technologies (ACCENT) project. APPEL encapsulates generic aspects of a policy language, but some specialized aspects related to call control are included too. In [15], parts of call control ontology used in ACCENT project, related to trigger, condition parameter and action classes are presented. Campbell et al. [15] presents segment of the call control ontology at the level of details, which complements the high level overview given in this paper. The Call Processing Language (CPL) scripts are used in many SIP based services. Devlic [16] presents context ontology for enhancement of CPL possibilities and the architecture of context aware VoIP system. Another system that uses context information and is developed using ontology is presented in [22]. The context related classes are not covered in the ontology presented in this paper. It is a potential topic for further work on the ontology.

Tetlow et al. [17] presents system architecture of an ontology based application server. Sources that provide input to the system are various metadata files (including application server configuration files). This information is parsed into semantic metadata - metadata in terms of the applied ontology. The semantic metadata and the ontology are inputs to the inference engine that is embedded in the application server. The engine is used either by core services or by tools which provide GUI interface. Akhmanov et al. [18] presents an ontology based system for prevention of Spam over Internet Telephony (SPIT) attacks in SIP networks. Anti-SPIT security policy is expressed as a set of rules, coded in SWRL language. Each rule contains two segments: conditions and actions. Applicable actions are: Allow, Block, Check Further. The rules are incorporated into the ontology model. An example of a rule that checks PRIORITY and FROM headers in a SIP message is given, but the details of the ontology are not.

 ISBN: 978-1-61208-262-2

VII. Applications of the Ontology

Presented ontology is used in the development of telephony framework [3]. The framework supports the following features:

- first-party call control operations,
- third-party call control operations,
- instant messaging,
- presence services,
- monitoring of the state of logged-in users and active sessions,
- network side applications.

Besides the classes described in this paper, the framework includes several other important classes. An object of SignalingDevice class interfaces the underlying session layer protocol. P3Server and P3Session classes implement third party call control. AppLEObserver is a parent of classes that provide support for monitoring, in the framework are used EndUserObserver and SessionObserver classes. Monitoring is based on publish-subscribe relationship which maps directly to SIP event notification mechanism. An object of FeatureMng class is responsible for feature interaction control (in collaboration with the classes that implement active features, and inherit the generic Feature class).

Certain design decisions regarding feature management are further explained in [13]. The session transfer and session redirection features are used as examples. Also in the paper are described two simple taxonomies of call control features. The first one groups features into primitive, derivative and composite. Basic call with interface for first party call control is a primitive feature. An example of a composite feature is conference. The relationship between composite and primitive feature is similar to aggregation ('has' relationship in Unified Modeling Language). In call processing derivative features can have read access to basic call data, but no feature has full (read and write) access to other feature's data. The second taxonomy distinguishes symmetric and asymmetric features. The former are implemented as peer-to-peer, and the later as client-server modules. Basic call is a symmetric feature, while session transfer and redirect are asymmetric features.

The ontology could be used in a framework as Multimedia Ontology-Driven Architecture (MODA) [18]. It would complement the ontologies used in [18] (ITU multimedia services ontology, IETF protocols ontology, software implementations ontology etc.). In network security, this ontology could be used for building data models, as in [14]. In that case the ontology should be extended with the work in [20]. In SIP environment, complex services are built as software applications residing on top of a SIP stack. Such ontology could be used to establish common data format for cooperation of different VoIP applications.

VIII. Conclusion and Future Work

The paper systematically presents important classes of a new ontology for IP telephony networks - by giving an overview of important aspects of network functionality and corresponding classes, while in most of the published papers only segments of used ontologies are presented. The ontology outlined in this paper is built around several key classes that are roots of class hierarchies. Those classes embody the most important qualities of the telephony network. The most important static aspect is the structure of telephony network. The corresponding class hierarchy starts with the class NetPoint. The dynamics of the telephone network are best described with appearance, and duration of telephony associations. The corresponding class hierarchy starts with the class Association. The most obvious association is the telephony session, or call, but there are also other important associations, which are not so obvious. An example is the publish/subscribe association, which is a building block for many complex services. Class Feature represents a generic telephony service.

Application of this ontology is possible both in design of telephony software (as in [3], [13]), and in data formats used for interoperability of different telephony applications and/or security applications in telephony systems. The support for context-based applications and potential application of ontology for multimedia systems beyond telephony are elements for future work.

Acknowledgment

This work was partially supported by the Ministry of Education and Science of the Republic of Serbia under the project III 44009-2.

References

[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, and R. Sparks, "SIP: Session Initiation Protocol", RFC 3261, Internet Engineering Task Force, 2002.

[2] ITU-T, "Visual Telephone Systems and Equipment for Local Area Networks Which Provide A Non-Guaranteed Quality Of Service", H.323, 1996.

[3] I. Basicevic, "Object-Oriented Framework for Development of Telephony Applications", International Conference on Digital Telecommunications, Colmar, France, 2009, pp.11–14.

[4] R. Kishore, H. Zhang, and R. Ramesh,, "A Helix-Spindle Model for Ontological Engineering", Communications of ACM, vol. 47, No. 2, 2004, pp. 69-75.

[5] I. Basicevic and M. Popovic, "Use of SIP in the Development of Telecom Services - A Case Study", The Journal of the Institute of Telecommunications Professionals, vol.2, Part 3, 2008, pp. 55-65.

[6] ITU-T, "The International Public Telecommunication Numbering plan", E.164, 2005.

[7] ITU-T, "Information technology - Abstract Syntax Notation One (ASN.1): Specification of basic notation", ITU-T Rec. X.680 —ISO/IEC 8824-1, 2002.

[8] ITU-T, "Call Signaling Protocols and Media Stream Packetization for Packet-Based Multimedia Communication Systems", H.225.0, 1998.

 ISBN: 978-1-61208-262-2

[9] ITU-T, “ISDN User-Network Interface Layer 3 Specification for Basic Call Control”, Q.931, 1998.
[10] M. Handley, V. Jacobson, and C. Perkins, “SDP: Session Description Protocol”, RFC 4566, Internet Engineering Task Force, 2006.
[11] ITU-T, “Control Protocol for Multimedia Communication”, H.245, 1998.
[12] P. Zave and E. Cheung, “Compositional Control of IP Media”, International Conference On Emerging Networking Experiments And Technologies, 2006.
[13] I. Basicevic, “Modular Design of Call Control Layer in Telephony Software”, International Journal Of Computer Science Issues, vol. 8, Issue 1, 2011.
[14] D. Geneiatakis and C. Lambrinoudakis, “An Ontology Description for SIP Security Flaws”, Computer Communications, vol. 30, 6, 2007, pp. 1367-1374.
[15] G. A. Campbell and K. J. Turner, “Ontologies to support Call Control Policies”, Proc. 3rd. Advanced International Conference on Telecommunications, 2007, pp. 5.1–5.6.
[16] A. Devlic, “Extending CPL with context ontology”, IMAC Workshop, Espoo, Finland, 2006
[17] P. Tetlow, J. Pan, D. Oberle, E. Wallace, M. Uschold, and E. Kendall, “Ontology Driven Architectures and Potential Uses of the Semantic Web in Systems and Software Engineering”, World Wide Web Consortium (W3C) working draft, 2006.
[18] S. A. Akhmanov, A. S. Chirkin, K. N. Drabovich, A. I. Kovrigin, D. Stelios, and G. Dimitris, “An Ontology-Driven antiSPIT Architecture”, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 26, 2010, pp. 189-198.
[19] E. Exposito, J. Gomez-Montalvo, and M. Lamolle, “Multimedia Ontology-Driven Architecture for Multimedia Systems”, Second International Conference on Information, Process, and Knowledge Management, eKNOW ’10, 2010, pp.7–12.
[20] VOIPSA, “VoIP Security and Privacy Threat Taxonomy”, Public Release 1.0, http://www.voipsa.org/Activities/taxonomy.php, accessed 05.2.2013.
[21] J. Frankowski, P. Rubach, and E. Szczekocka, “Collaborative Ontology Development in Real Telecom Environment”, 1st International Working Conference on Business Process and Services Computing, 2007, pp. 40-53.
[22] P. Gutheim, “An Ontology-Based Context Inference Service for Mobile Applications in Next-Generation Networks”, IEEE Communications Magazine, vol. 49, no. 1, 2011, pp. 60-66.

 ISBN: 978-1-61208-262-2

Rate Distortion Performance of H.264/SVC in Full HD with Constant Frame Rate and High Granularity

Martin Slanina, Michal Ries
Brno University of Technology
Technicka 12
Brno, Czech Republic
e-mail: slaninam@feec.vutbr.cz, ries@feec.vutbr.cz

Janne Vehkaperä
VTT Research Centre of Finland
Kaitoväylä 1
Oulu, Finland
e-mail: janne.vehkapera@vtt.fi

***Abstract*—In this paper, we provide an analysis of performance of the scalable extension of the H.264/AVC video codec. We assume a fixed display scenario at full HD resolution with a constant frame rate. The encoded bit stream consists of one Coarse Grain Scalability (CGS) and one Medium Grain Scalability (MGS) layer with three sublayers, allowing for creating four quality layers from the complete bit stream. Hierarchical coding is employed, which results in dyadic decomposition of temporal layers. In order to increase the granularity of quality layers, the packets are selectively dropped from appropriate quality and temporal layers. We provide a performance analysis of such approach, compare the rate distortion performance to a mainstream H.264 encoder and analyze the composition of the bit stream at the considered operation points. Our findings show that quality enhancement of temporal layers has different effect on the overall performance depending on which temporal layer is enhanced.**

Keywords-High definition video; quality of service; scalability; video coding.

I. INTRODUCTION

It has been pronounced in the recent years that the video traffic forms a significant share in the data carried over the Internet. In 2011, the share of Internet video was 51 % of all consumer Internet traffic and is expected to raise to 55 % in 2016 [1]. These high numbers do not include video exchanged through peer to peer sharing. Counting them in, the share of video on the overall Internet traffic is expected to reach 86 % in 2016.

There is a rapid increase in the number of broadband connections to the Internet. Quite naturally, as the throughput increases, the quality demands of the users increase as well. Advanced Internet video (stereoscopic 3D and high definition) is getting an important share on the overall Internet video traffic and is expected to reach 46 % of the consumer Internet video traffic in 2014 [2]. These numbers clearly show the importance of efficient online delivery of high quality video.

The bit rate of the video carried over the data network is one of the most important aspects related to quality. When the bit rate is too low, the visual quality of the decoded video images may be degraded. On the other hand, when the bit rate is too high, the risk of network having issues delivering the high amount of data increases, which may lead to video freezes or longer waiting times (usually with TCP-based connections) or loss of data and visual impairments (for UDP-based connections). The optimum bit rate is thus limited from both ends and depends, to a large extent, on the individual parameters of the user's connection.

In order to cope with the different bit rate requirements of the delivered video, the most common approach is to offer several copies of the same content, encoded at different bit rates, and let the user or the video player select one of the bit rates according to the available throughput. Such fixed bit rate approach is, however, inefficient in case the available network throughput is changing over the playback. One possible solution is to allow for switching among the different bit rates offered by the provider, implemented in e.g. HTTP Live Streaming technique [3] or DASH - Dynamic Adaptive Streaming over HTTP [4]. A common feature of the mentioned adaptive streaming techniques is that it is the *client* who decides on the bit rate to receive. For such decision, the client needs to be informed on the actual available bit rate and take fast action when the available bit rate decreases. The advantage of the Scalable Video Coding (SVC) [5] extension of the H.264/AVC standard is that parts of the transmitted data stream have different importance. Through dropping certain packets in case of insufficient transmission capacity, SVC offers graceful degradation by decreasing the quality of the decoded received stream. A nice description of the concepts used in SVC can be found in [5].

The aim of this paper is to present a scheme for dropping SVC packets with high bit rate/quality granularity and to compare its performance to the non-scalable H.264/AVC encoder using multiple streams. We focus on the full HD resolution and full frame rate of 25 frames per seconds, as such format is very promising for the current and future video delivery.

Throughout this paper, the rate distortion performance will be measured using the Peak Signal to Noise Ratio (PSNR). Even though PSNR is not a good quality metric in most cases as the correlation of its outputs with results of subjective experiments is poor, it is suitable in this setup as a performance indicator: as it has traditionally been used for quantifying the performance of coding tools and algorithms, it is extremely wide spread and well understood in the video

 ISBN: 978-1-61208-262-2

coding community.

This paper is organized as follows: First, the related work is mentioned in Section II. The used packet dropping scheme and the resulting SVC performance are described in Section III. Finally, the paper concludes in Section IV.

II. Related Work

The performance of the SVC encoding algorithm has been studied even before the standard was released in 2007. In [6], the authors find that the overall rate-distortion performance of the SVC is approximately 10 % worse compared to AVC with identical settings. In other words, SVC needs approximately 10 % more bit rate to achieve the same quality in terms of PSNR. In [6], one base layer and one enhancement layer in spatial or quality domain is used. SVC performance analysis based on subjective tests has been done in [7], confirming the previous results. In [6] and [7], one base layer and one enhancement layer are used in terms of quality and spatial scalability.

A common approach to selecting the layers in the scalable bit stream is selecting a certain combination of D (dependency), T (temporal) and Q (quality) parameters. Then, all the inferior layers are included in all three domains. For fixed display conditions, i.e. fixed frame rate and fixed spatial resolutions, one only gets as many bit rate/quality levels as there are Q levels in the original stream. We utilize dropping in different T layers in order to increase the bit rate/quality granularity.

The authors of [10] provide an analysis of different approaches to dropping layers in the SVC bit stream in the mobile environment. Through a subjective study, different impact on Quality of Experience is introduced by different scaling approaches and for different contents. Generally, spatial scaling is regarded worse compared to temporal and quality scaling, which leads to a recommendation that the quality layers and some temporal layers should be dropped first. This is in complete agreement with our approach, where no spatial scalability is included.

In [11] and [12], the authors analyze the impact of unstable transmissions on the perceived quality when a scalable video bit stream is transmitted. In this context, the unstable transmission leads to varying number of layers received in the sequence duration. It has been shown that quality of unstable videos is subjectively perceived close to the quality of stable videos. Furthermore, it was found that temporal scalability introduces severe degradation of perceived quality while quality scalability leads to best results. These results are the motivation to keep the video frame rate constant through all operation points in our experiment.

III. SVC Performance

A. Packet Dropping Scheme

This subsection presents the packet dropping scheme employed for the increased bit rate granularity. Let us first describe the structure of the full quality bit stream as shown in Fig. 1. Using the hierarchical B frames, the full frame rate of 25 fps can be decomposed in a dyadic structure down to 3.125 fps in case four temporal layers are used. In the lowest frame rate (T = 0), only the 1^{st} and 9^{th} frame are encoded. For each of them, several quality layers can be defined (Q = 1 to Q = 3). The frames at the lowest frame rate serve as the basis for prediction of frames in the higher frame rate, i.e. T = 1. At T = 1, the 1^{st}, 5^{th} and 9^{th} frame are available. Continuing such prediction leads to doubling the frame rate at each step, resulting in 25 fps at T = 3.

TABLE I. OPERATION POINTS.

T:	T = 0	T = 1	T = 2	T = 3	T = 4
Q:	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
OP1	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠
OP2	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ □
OP3	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ □ □
OP4	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ □ □ □
OP5	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □
OP6	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ □	⊠ □ □ □ □
OP7	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ □ □	⊠ □ □ □ □
OP8	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ □ □ □	⊠ □ □ □ □
OP9	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ □ □ □ □
OP10	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ ⊠ ⊠ ⊠ ⊠
OP11	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ ⊠ ⊠ ⊠ □
OP12	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ ⊠ ⊠ □ □
OP13	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ ⊠ □ □ □
OP14	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ □	⊠ □ □ □ □	⊠ □ □ □ □
OP15	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP16	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP17	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP18	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ ⊠ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP19	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ ⊠ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP20	⊠ ⊠ ⊠ ⊠ ⊠	⊠ ⊠ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP21	⊠ ⊠ ⊠ ⊠ ⊠	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP22	⊠ ⊠ ⊠ ⊠ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP23	⊠ ⊠ ⊠ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP24	⊠ ⊠ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP25	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □
OP26	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ ⊠ ⊠ ⊠ ⊠
OP27	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ ⊠ ⊠ ⊠ □
OP28	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ ⊠ ⊠ □ □
OP29	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ ⊠ □ □ □
OP30	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □	⊠ □ □ □ □

The common approach to dropping SVC data is illustrated in Fig. 2. In order to keep the highest frame rate (T = 3), all the inferior frame rates are kept in full quality (Q = 3). Obviously, this technique can reach only 4 operation points at the full temporal resolution.

In contrast, the approach that we use throughout this paper allows dropping quality layers even from lower temporal resolutions. As long as the base quality layer is kept in each temporal layer, the full frame rate can easily be reconstructed. Such approach is shown in Fig. 3. Note the obvious inconsistency in layer T = 1: Although in the inferior layer (T = 0) no quality enhancement data was kept, higher quality layers are present at T = 1. As there are no references used to the *specific* Q at the lower layer, prediction can still be done and enhanced at the current layer. Moreover, we will show that in some cases this scenario can lead to higher

 ISBN: 978-1-61208-262-2

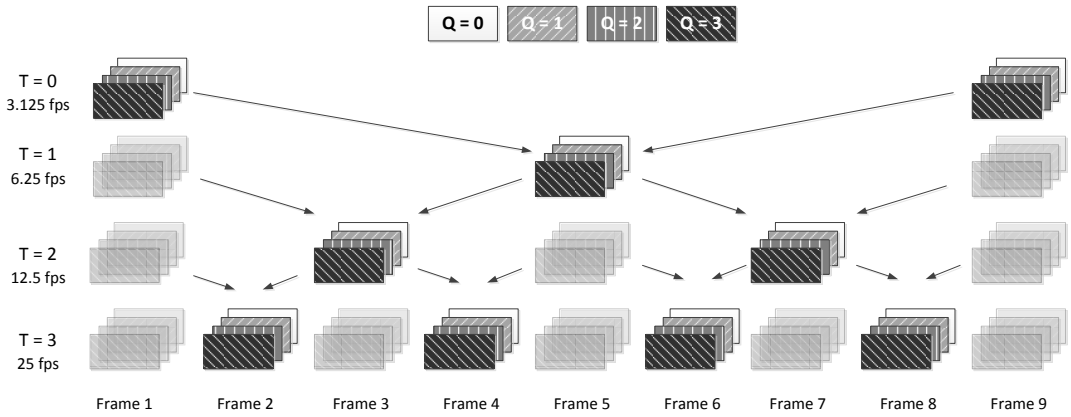


Figure 1. Layers in the SVC bit stream in full quality and full temporal resolution.

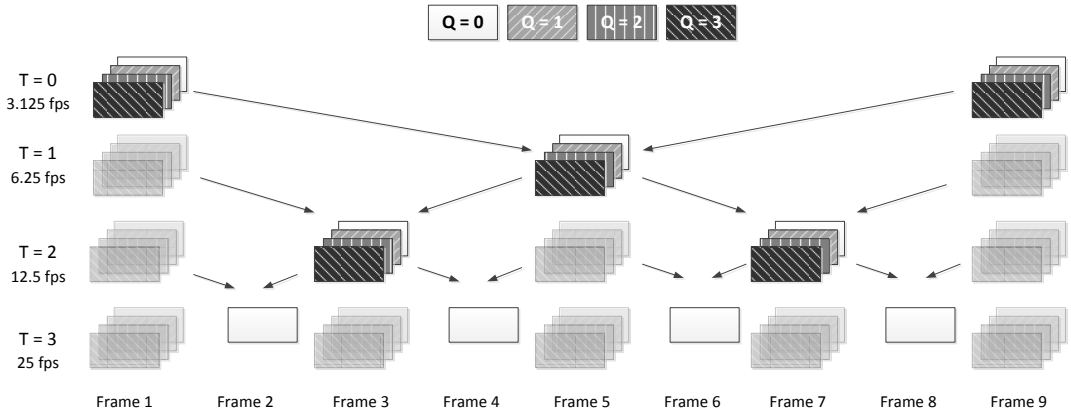


Figure 2. Layers in the SVC bit stream in case the highest temporal resolution is present with only the base quality layer.

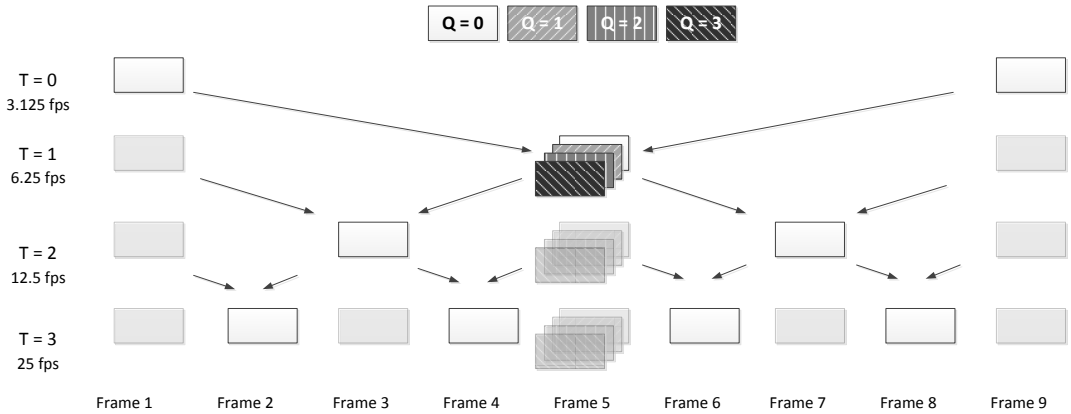


Figure 3. Layers in the SVC bit stream in case only the base quality layer is kept for several temporal layers.

 ISBN: 978-1-61208-262-2

quality at a given bit rate. All the operation points considered in this experiment are described by TABLE I. An operation point (OP) is defined by the subset of layers, employed in the respective sub-stream. Please note that OP25 and OP30 are made up from an identical set of sub-layers, which means one of them could be removed as redundant. However, we keep both of these operation points as they represent the lowest quality reached through different packet dropping strategies, which is useful in further descriptions and plots. In order to drop packets from an SVC bit stream, trace files were prepared and used with the bit stream extractor tool available with JSVM. The resulting streams were decoded with JSVM and compared to uncompressed originals in terms of mean PSNR of the luma component.

B. Test Setup

Two sequences were used in our experiment: The ”blue sky” sequence, which has midrange spatial and low temporal activity (a view of a sky through treetops with slow camera rotation) and the ”tractor” sequence, which has high spatial and temporal activity (a tractor working on the field with complex background and high motion). The sequences have been downloaded from [13].

The sequences were encoded with the JSVM encoder (Joint Scalable Video Model) version 9.19.13 [8]. As this version has no rate control mechanism implemented, we decided for a Quantization Parameter (QP)-based configuration and all the sequences were encoded with the same initial QPs. According to the complexity of each scene, the resulting bit rates and qualities differ.

The bit stream was formed by the component layers:

1) *The base layer,* identified by Q = 0. This layer was coded with the GOP size of 16 to allow efficient decomposition of frame rate. The IDR period was 32. The layer was coded using High profile, initial QP set to 48.
2) *CGS enhancement layer*, identified by Q = 1. The Coarse Grain Scalability (CGS) allows for enhancing the quality of a decoded picture by employing inter-layer prediction mechanisms, such as prediction of macroblock modes, motion parameters and residual prediction. [9]. The layer was encoded with initial QP set to 42.
3) *MGS enhancement layer*, identified by Q = 2 to 4. The Medium Grain Scalability (MGS) provides a finer granularity compared to CGS by splitting a given quality enhancement layer into several MGS layers. Basically, MGS divides the transform coefficients of each macroblock into multiple groups. In our configuration, we used 4 (most significant) coefficients for the lowest quality layer, 4 coefficients for the medium and 8 (least significant) coefficients for the highest quality MGS layer. The initial QP of the MGS enhancement layer was set to 30.

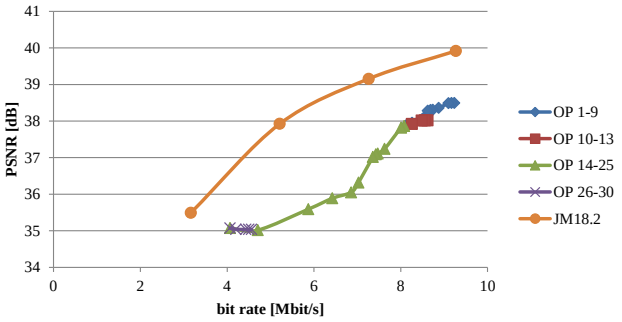


Figure 4. Bit rate vs. quality (PSNR) for the blue_sky sequence.

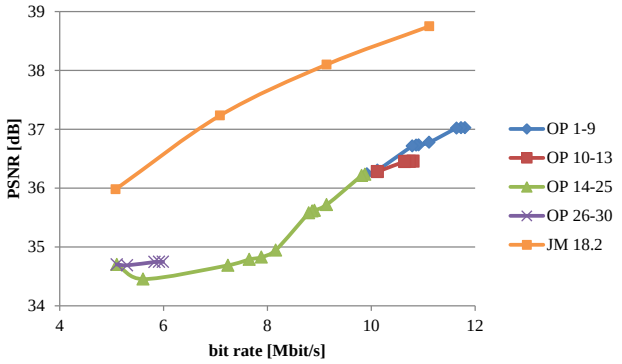


Figure 5. Bit rate vs. quality (PSNR) for the tractor sequence.

Furthermore, the sequences were encoded using the reference implementation of the H.264/AVC encoder [14] version JM 18.2. The motivation for using another encoder is to keep track of the overall performance of the scalable coding and the usability of the created bit streams. The AVC encoder was set to High profile, level 4.0, with hierarchical B-frames having the same structure as the B-frames in SVC.

C. Results

Figures 4 and 5 illustrate the average PSNR for 200 frames of each sequence along all the SVC operation points defined in TABLE I. Furthermore, a curve illustrating the AVC performance is shown. It can be observed that the coding performance of AVC is several dB better compared to SVC. On the other hand, the SVC offers flexibility in modifying a readily encoded bit stream such that the required bit rate can be easily altered with no need of transcoding.

An interesting phenomenon can be observed in the area of lower SVC bit rates for the ”tractor” sequence. Here, different qualities are achieved for the curve representing operation points 14-24, where the quality layers are dropped regularly from higher to lower temporal layers, and the curve representing operation points 26-29, where only the quality

 ISBN: 978-1-61208-262-2

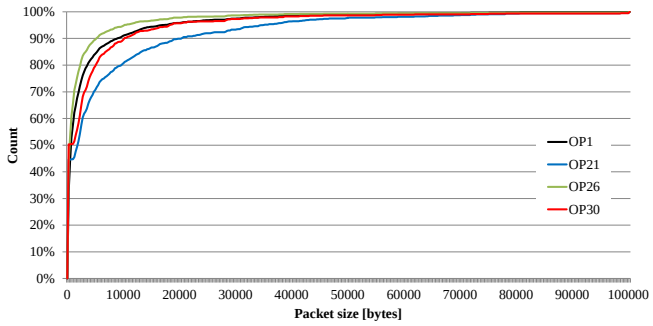


Figure 6. Cumulative distribution function for the blue_sky sequence.

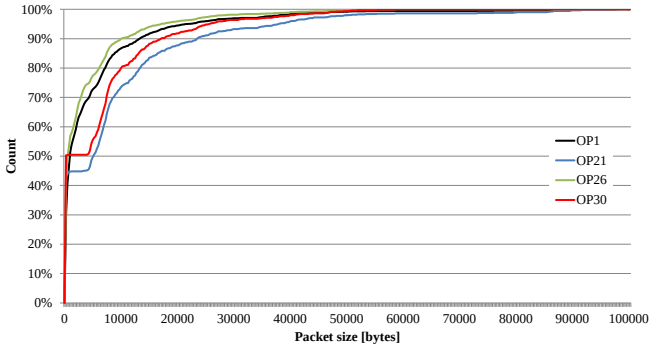


Figure 7. Cumulative distribution function for the tractor sequence.

enhancement data of the highest temporal layer are kept.

In OPs 26-29, the frames from the highest temporal layer are enhanced, which means quality enhancement information is present in every second frame of the sequence (see T = 3 in Fig. 1) - the number of frames in this temporal layer is higher than the number of frames in the inferior temporal layers. One would expect that high bit rate is required to present the frames of highest temporal layer at a higher quality (more quality layers). This might not be true in all cases since the frames at the highest temporal layers quite often do need lower bit rate than the frames at lower temporal layers (due to temporal prediction).

In order to understand the composition of the packets in the bit stream, we have plotted the cumulative distribution function of the packet sizes for both sequences in Fig. 6 and Fig. 7. In these CDFs, the operation point OP1 (full bit stream) is displayed using a black line, OP30 (lowest bit rate stream) is displayed with a red line, OP21 is drawn with a blue line and, finally, OP26 is drawn with a green line. The reason for selecting OP21 and 26 is that they follow different trends in Fig. 5 and use different parts of data for quality enhancement: While OP21 drops quality layers monotonously from highest to lowest temporal layers, OP26 uses quality enhancement in the highest temporal layer only. The cumulative distribution function plots show that the curve for OP 21 is generally below the other curves, which means that there is a low percentage of smaller packets in the bit stream. On the other hand, the curve for OP26 is clearly above all other curves, which means that the percentage of small packets in the bit stream is higher. An explanation for this is that in OP26, the quality enhancement information is present for a larger number of frames (higher temporal layer) using a lower amount of data in small packets.

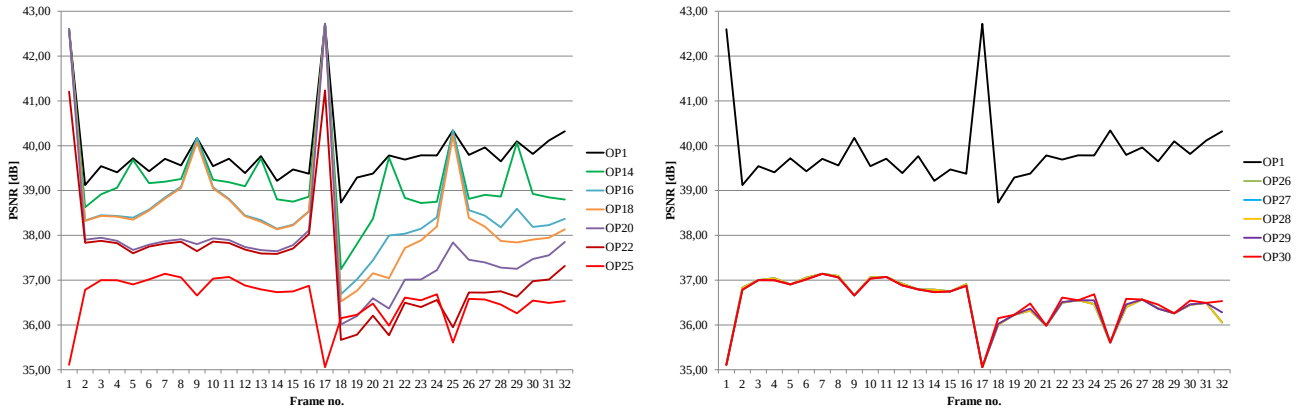
Still, the explanation for the increased PSNR in higher operation points in Fig. 5 is missing. For a better understanding, we have plotted the PSNR values in time for two GOPs in Fig. 8 and Fig. 9. In all plots, OP1 is drawn with black line, OP30 (equal to OP25) is drawn with red line. First, let us focus on Fig. 8 showing the ”blue_sky” sequence: Fig. 8a shows the behaviour for OPs 14-25. With the increasing operation point index, there is a monotonous decrease of quality for all frames. Sharp PSNR peaks can be observed at frames number 1 and 17, which are in fact the boundaries of a GOP (I frames). These peaks mean that the the PSNR of the frames in the lowest temporal layer is high in case all the quality enhancement information is available. When the enhancement information is discarded from the lowest temporal layer, the PSNR of I frames drops drastically. As a result, what we can observe for both extreme cases, being the highest and the lowest overall mean PSNR for OP1 and OP25, is that there is a significant quality fluctuation present within a group of pictures. This unwanted effect should be eliminated by the rate control mechanism of the encoder. Unfortunately, no rate control has been implemented in the JSVM reference encoder for scalable video coding so far. Fig. 8b shows the PSNR of frames within two GOPs of OPs 26-30. In this case, enhancing the higher temporal layers only brings no improvement over OP30.

An interesting thing can be observed in the plots for the ”tractor” sequence in Fig. 9. In contrast to high PSNR fluctuations appearing for OPs 14-25 (Fig. 9a), the enhancement of higher temporal layers only results in a smoother PSNR curve as shown in (Fig. 9b) It can be expected that the PSNR fluctuations are annoying for the user and their elimination is desirable. When properly configuring the encoder, enhancement of the higher temporal layers only should lead to a more equal improvement of the quality of all frames throughout the sequence.

IV. Conclusion

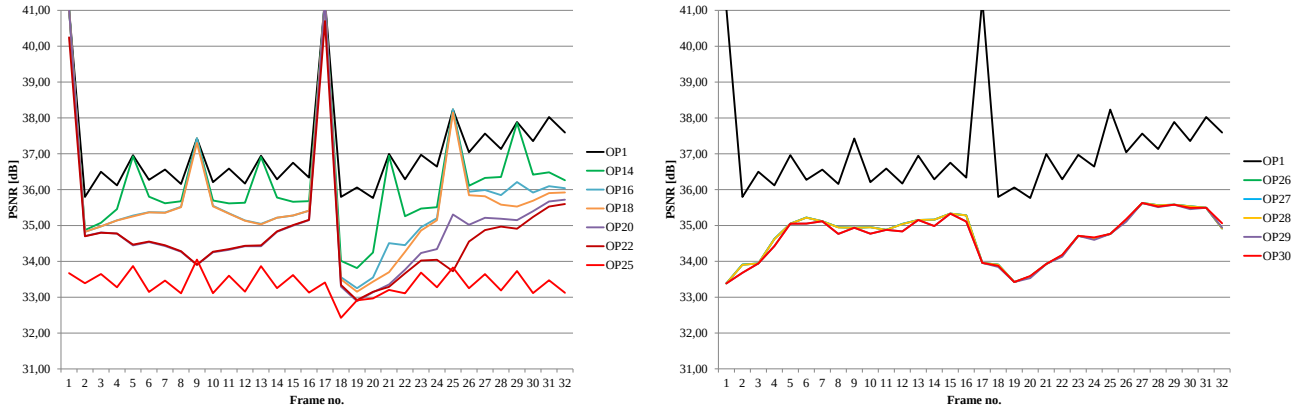
The usability of H.264/SVC for efficient adaptive video transmission depends, to a larger extent, on the desired usage scenario. In our experiment, the usage scenario was high definition video, whose application is very likely to be found in IP-based television or wireless multimedia services. An approach for increasing the quality granularity of a scalable bit stream was presented, dropping the quality enhancement packets from different temporal layers according to a defined scheme.

 ISBN: 978-1-61208-262-2



a) b)

Figure 8. The blue_sky sequence: PSNR per frame for operation points 1 and a) 14 to 25; b) 26 to 30.



a) b)

Figure 9. The tractor sequence: PSNR per frame for operation points 1 and a) 14 to 25; b) 26 to 30.

It has been shown that the overall performance of H.264/SVC in this scenario is considerably worse compared to the very well performing implementation of H.264/AVC. However, one has to keep in mind that the SVC reference encoder lacks efficient rate control, which degrades its coding efficiency. It can be expected that with other encoder implementations, the rate-distortion performance of the SVC encoder approaches AVC more closely.

Analyzing the cumulative distribution function of the packet sizes for the considered operation points has revealed the fact that quality enhancement data for higher temporal layers are carried in a higher number of smaller packets, which can be a useful information from the network point of view. It is also in a perfect agreement with the fact that with high temporal layer enhancement, a higher number of frames is enhanced by a smaller PSNR difference.

Finally, it has been shown that when the enhancement of the lower temporal layers only is kept, the quality of video frames in time tends to exhibit fluctuations in terms of PSNR. This effect does not appear when the enhancement of higher temporal layers only is employed. In order to exploit the impact of the quality fluctuations on the quality perceived by the consumer, the authors plan to conduct a subjective testing session. In case it is proved that removing these fluctuations would bring an overall improvement of the video quality, an optimization of the packet dropping strategy based on high temporal layer enhancement will be considered.

ACKNOWLEDGEMENT

The paper was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. The support of the project CZ.1.07/2.3.00/20.0007 WICOMT, financed from the operational program Education for competitiveness, is gratefully acknowledged. This work was supported by

the Czech Ministry of Education under project number LD11081.

REFERENCES

[1] Cisco Systems, Inc. Cisco Visual Networking Index: Forecast and Methodology, 2011-2016, San Jose, CA, USA: Cisco Systems, 2012.

[2] Cisco Systems, Inc. Cisco Visual Networking Index: Forecast and Methodology, 2009-2014, San Jose, CA, USA: Cisco Systems, 2010.

[3] D. Van Deursen, W. Van Lancker, and R. Van de Walle, "On media delivery protocols in the Web," In 2010 IEEE International Conference on Multimedia and Expo (ICME), 2010, pp. 1028 - 1033.

[4] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over internet," IEEE Multimedia, vol. 18, no. 4, 2011, pp. 62-67.

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Video Coding, vol. 17, no. 9, 2007, pp. 1103-1120.

[6] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, 2007, pp. 1194-1203.

[7] T. Oelbaum, H. Schwarz, M. Wien, and T. Wiegand, "Subjective performance evaluation of the SVC extension of H.264/AVC," In 15th IEEE International Conference on Image Processing, ICIP 2008, San Diego (CA, USA), 2008, pp. 2772-2775.

[8] SVC Reference Software (JSVM) [Online, retrieved: January, 2012] Available at: http://ip.hhi.de/imagecom_G1/savce/downloads/SVCReference-Software.htm

[9] R. Gupta, A. Pulipaka, P. Seeling, L. J. Karam, and M. Reisslein, "H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation," IEEE Transactions on Broadcasting, vol. 58, no. 3, 2012, pp. 428 - 439.

[10] A. Eichhorn and P. Ni, "Pick your Layers wisely - A Quality Assessment of H.264 Scalable Video Coding for Mobile Devices," In IEEE International Conference on Communications, ICC, Dresden (Germany), 2009, pp. 1 - 6.

[11] L. C. Daronco, V. Roesler, and J. V. de Lima, "Subjective Video Quality Assessment Applied to Scalable Video Coding and Transmission Instability," In Proceedings of the 2010 ACM Symposium on Applied Computing, SAC'10, Sierre (Switzerland), 2010, pp. 1898 - 1904.

[12] L. C. Daronco, V. Roesler, J. V. de Lima, and R. Balbinot, "Quality analysis of scalable video coding on unstable transmissions," Springer Multimedia Tools and Applications, 2011, DOI: 10.1007/s11042-011-0760-y.

[13] Xiph.org Video Test Media [Online, retrieved: June, 2012], Available at: http://media.xiph.org/video/derf/ .

[14] K. Suehring, The H.264/AVC Reference Software, version 18.2. [Online, retrieved: April, 2012]. Available at: http://iphome.hhi.de/suehring/tml/download/ .

 ISBN: 978-1-61208-262-2

Synthesis of MPEG-like Standard with Interval Multiset Estimates

Mark Sh. Levin
Institute for Information Transmission Problems
Russian Academy of Sciences
Moscow, Russia
Email: mslevin@acm.org

***Abstract*—The paper addresses modular modeling, design, and improvement of MPEG-like standard for multimedia information processing. Morphological (modular) system design and improvement are considered as composition of the standard elements (components) configuration. The solving process is based on Hierarchical Morphological Multicriteria Design (HMMD) approach: (i) multicriteria selection of alternatives for system components, (ii) synthesis of the selected alternatives into a resultant combination. Assessment of design alternatives is based on interval multiset estimates, assessment of compatibility between the design alternatives is based on ordinal scale. Improvement of the obtained solutions is examined as well (knapsack-like problem). Numerical examples illustrate the design process.**

***Keywords*-standard for multimedia information; combinatorial synthesis; combinatorial optimization; multiset.**

I. INTRODUCTION

In recent decades, the significance of multimedia information processing is increased (e.g., [1], [2], [7], [8]). A structural approach to modeling of MPEG-like standard for multimedia information has been presented in [14]. Multicriteria analysis of algorithms for processing of image sequences to reveal Pareto-efficient methods for some typical image sequences was studied in [4]. This work is a basis for on-line selection of the best processing algorithms for an input image sequence. In [12], an example for combinatorial synthesis of MPEG-like standard was described. This paper focuses on combinatorial synthesis of MPEG-like standard and its improvement with using interval multiset estimates of standard elements. The approach can be considered as a basis for on-line design and modification of the standards in multimedia information processing.

Morphological (modular) system design and improvement are considered as composition of elements of MPEG-like standard (components, e.g., rules, algorithms). Hierarchical Morphological Multicriteria Design (HMMD) approach is used for modular design with interval multiset estimates for assessment of design alternatives (DAs) for elements of standard. This composition method (with interval multiset estimates) has been suggested in [10] and was used in some synthesis works (e.g., [11], [13]). HMMD implements a multi-stage design framework and provides cascade-like design framework:

(1) Decomposition/partitioning of system and system requirements to obtain a hierarchical system model and a hierarchy of system requirements, which correspond to system parts/components,

(2) 'Bottom-Up' design process:

(i) multicriteria selection of design alternatives (DAs) for system components,

(ii) synthesis of the selected alternatives into a resultant combination.

The additional systems problem is examined: improvement of the obtained solutions (multiple choice problem).

Fig. 1 depicts a simplified scheme of the approach.

The numerical design examples involve hierarchical system structure of MPEG-like standard, DAs for system parts/components, estimates of DAs and their compatibility, Bottom-Up design process, analysis and improvement of the obtained system solutions. Assessment of DAs and their compatibility is based on expert judgment.

The structure of the paper is the following. Section II presents description of combinatorial synthesis with interval multiset estimates. In Section III, hierarchical modeling of the MPEG-like standard (including design alternatives for leaf nodes of the model) and combinatorial synthesis of four Pareto-efficient solutions (on the basis of HMMD) are described. Section IV presents improvement of the obtained solutions as selection of improvement actions (on the basis of multiple choice problem).

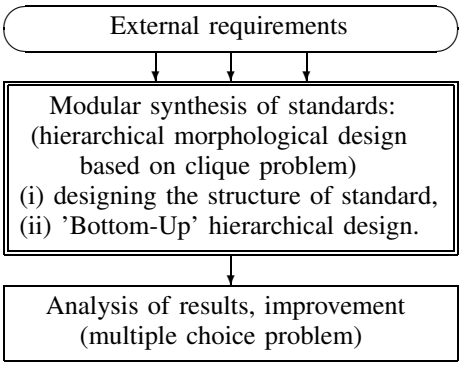


Figure 1. Simplified design framework

 ISBN: 978-1-61208-262-2

II. SYNTHESIS WITH INTERVAL MULTISET ESTIMATES

This section presents a compressed description of combinatorial synthesis with interval multiset estimates, which has been suggested in [10]. Close compressed materials are contained in ([11], [12],[13]).

The approach consists in assignment of elements $(1, 2, 3, ...)$ into an ordinal scale $[1, 2, ..., l]$. As a result, a multi-set based estimate is obtained, where a basis set involves all levels of the ordinal scale: $\Omega = \{1, 2, ..., l\}$ (the levels are linear ordered: $1 \succ 2 \succ 3 \succ ...$) and the assessment problem (for each alternative) consists in selection of a multiset over set Ω while taking into account two conditions:

1. Cardinality of the selected multiset equals a specified number of elements $\eta = 1, 2, 3, ...$ (i.e., multisets of cardinality η are considered);

2. "Configuration" of the multiset is the following: the selected elements of Ω cover an interval over scale $[1, l]$ (i.e., "interval multiset estimate").

Thus, an estimate e for an alternative A is (scale $[1, l]$, position-based form or position form): $e(A) = (\eta_1, ..., \eta_\iota, ..., \eta_l)$, where η_ι corresponds to the number of elements at the level ι ($\iota = \overline{1, l}$), or $e(A) = \{\overbrace{1, ..., 1}^{\eta_1}, \overbrace{2, ...2}^{\eta_2}, \overbrace{3, ..., 3}^{\eta_3}, ..., \overbrace{l, ..., l}^{\eta_l}\}$. The number of multisets of cardinality η, with elements taken from a finite set of cardinality l, is called the "multiset coefficient" or "multiset number" ([6],[15]): $\mu^{l,\eta} = \frac{l(l+1)(l+2)...(l+\eta-1)}{\eta!}$. This number corresponds to possible estimates (without taking into account interval condition 2). In the case of condition 2, the number of estimates is decreased. Generally, assessment problems based on interval multiset estimates can be denoted as follows: $P^{l,\eta}$. The assessment problem $P^{3,4}$ will be used in numerical examples.

In addition, operations over multiset estimates are used [10]: integration, vector-like proximity, aggregation.

Integration of estimates (mainly, for composite systems) is based on summarization of the estimates by components (i.e., positions). Let us consider n estimates (position form): estimate $e^1 = (\eta_1^1, ..., \eta_\iota^1, ..., \eta_l^1)$, . . ., estimate $e^\kappa = (\eta_1^\kappa, ..., \eta_\iota^\kappa, ..., \eta_l^\kappa)$, . . ., estimate $e^n = (\eta_1^n, ..., \eta_\iota^n, ..., \eta_l^n)$. Then, the integrated estimate is: estimate $e^I = (\eta_1^I, ..., \eta_\iota^I, ..., \eta_l^I)$, where $\eta_\iota^I = \sum_{\kappa=1}^n \eta_\iota^\kappa \quad \forall \iota = \overline{1, l}$. In fact, the operation $\biguplus$ is used for multiset estimates: $e^I = e^1 \biguplus ... \biguplus e^\kappa \biguplus ... \biguplus e^n$.

Vector-like proximity is considered as follows. Let A_1 and A_2 be two alternatives with corresponding interval multiset estimates $e(A_1)$, $e(A_2)$. Vector-like proximity for the alternatives above is: $\delta(e(A_1), e(A_2)) = (\delta^-(A_1, A_2), \delta^+(A_1, A_2))$, where vector components are: (i) δ^- is the number of one-step changes: element of quality $\iota + 1$ into element of quality ι ($\iota = \overline{1, l-1}$) (this corresponds to "improvement"); (ii) δ^+ is the number of one-step changes: element of quality ι into element of quality $\iota + 1$ ($\iota = \overline{1, l-1}$) (this corresponds to "degradation"). It is assumed: $|\delta(e(A_1), e(A_2))| = |\delta^-(A_1, A_2)| + |\delta^+(A_1, A_2)|$.

A median (aggregated) estimate (aggregation) for a set of initial estimates is defined as follows. Let $E = \{e_1, ..., e_\kappa, ..., e_n\}$ be the set of initial estimates. let D be the set of all possible estimates ($E \subseteq D$). Thus, the median estimates ("generalized median" M^g and "set median" M^s) are: $M^g = \arg\min_{M \in D} \sum_{\kappa=1}^n |\delta(M, e_\kappa)|$; $M^s = \arg\min_{M \in E} \sum_{\kappa=1}^n |\delta(M, e_\kappa)|$.

A brief description of combinatorial synthesis (HMMD) with ordinal estimates of design alternatives is the following ([9], [10]). An examined composite (modular, decomposable) system consists of components and their interconnection or compatibility (IC). Basic assumptions of HMMD are the following: *(a)* a tree-like structure of the system; *(b)* a composite estimate for system quality that integrates components (subsystems, parts) qualities and qualities of IC (compatibility) across subsystems; *(c)* monotonic criteria for the system and its components; *(d)* quality of system components and IC are evaluated on the basis of coordinated ordinal scales. The designations are: (1) design alternatives (DAs) for leaf nodes of the model; (2) priorities of DAs ($\iota = \overline{1, l}$; 1 corresponds to the best one); (3) ordinal compatibility for each pair of DAs ($w = \overline{1, \nu}$; ν corresponds to the best one). Let S be a system consisting of m parts (components): $R(1), ..., R(i), ..., R(m)$. A set of design alternatives is generated for each system part above. The problem is:

Find a composite design alternative $S = S(1) \star ... \star S(i) \star ... \star S(m)$ *of DAs (one representative design alternative* $S(i)$ *for each system component/part* $R(i)$, $i = \overline{1, m}$ *) with non-zero compatibility between design alternatives.*

A discrete "space" of the system excellence (a poset) on the basis of the following vector is used: $N(S) = (w(S); e(S))$, where $w(S)$ is the minimum of pairwise compatibility between DAs, which correspond to different system components (i.e., $\forall R_{j_1}$ and R_{j_2}, $1 \le j_1 \ne j_2 \le m$) in S, $e(S) = (\eta_1, ..., \eta_\iota, ..., \eta_l)$, where η_ι is the number of DAs of the ιth quality in S. Further, the problem is described as follows:

$$\max\ e(S), \quad \max\ w(S), \quad s.t.\ \ w(S) \ge 1. \qquad (1)$$

As a result, we search for composite solutions, which are nondominated by $N(S)$ (i.e., Pareto-efficient). "Maximization" of $e(S)$ is based on the corresponding poset. The considered combinatorial problem is NP-hard and enumerative solving schemes or heuristics are used.

In the article, combinatorial synthesis is based on usage of multiset estimates of design alternatives for system parts. For the resultant system $S = S(1) \star ... \star S(i) \star ... \star S(m)$ the same type of the multiset estimate is examined: an aggregated estimate ("generalized median") of corresponding multiset estimates of its components (i.e., selected DAs). Thus,

 ISBN: 978-1-61208-262-2

$N(S) = (w(S); e(S))$, where $e(S)$ is the "generalized median" of estimates of the solution components. The modified problem is:

$$\max\ e(S) = M^g = \arg\min_{M \in D} \sum_{i=1}^{m} |\delta(M, e(S_i))|,$$
$$\max\ w(S),$$
$$s.t. \qquad w(S) \geq 1. \qquad (2)$$

Here, enumeration methods or heuristics are used ([9], [10]).

The basic multiple choice problem is (e.g., [3], [5]):

$$\max \sum_{i=1}^{m} \sum_{j=1}^{q_i} c_{ij} x_{ij}$$
$$s.t. \sum_{i=1}^{m} \sum_{j=1}^{q_i} a_{ij} x_{ij} \leq b; \ \sum_{j=1}^{q_i} x_{ij} \leq 1, \ i = \overline{1, m};$$
$$x_{ij} \in \{0, 1\}. \qquad (3)$$

In the case of multiset estimates of item "utility" $e_i, i \in \{1, ..., i, ..., m\}$ (instead of c_i), an aggregated multiset estimate as the "generalized median" is used. The item set is: $A = \bigcup_{i=1}^{m} A_i$, $A_i = \{(i, 1), (i, 2), ..., (i, q_i)\}$. Boolean variable $x_{i,j}$ corresponds to selection of the item (i, j). The solution is a subset of the initial item set: $S = \{(i, j) | x_{i,j} = 1\}$. The problem is:

$$\max\ e(S) = \max\ M =$$
$$\arg\min_{M \in D} \sum_{(i,j) \in S = \{(i,j) | x_{i,j} = 1\}} |\delta(M, e_{i,j})|,$$
$$s.t. \quad \sum_{i=1}^{m} \sum_{j=1}^{q_i} a_{ij} x_{i,j} \leq b; \quad \sum_{j=1}^{q_i} x_{ij} = 1;$$
$$x_{ij} \in \{0, 1\}. \qquad (4)$$

In the paper, a greedy heuristic is used.

III. Structure of Standard and Synthesis

Hierarchical modular structure of MPEG-like standard has been examined in [14]. A simplified hierarchical model of MPEG-like standard was analyzed in [14]. Here, a modification of the above-mentioned hierarchical model is considered (Fig. 2):

0. MPEG-like standard $S = A \star B$.

1. General part $A = C \star B \star D \star E \star F$:

1.1. Applied layer (videotelephony, videoconferencing, digital broadcast, digital storage media, etc.) C: bit rate 64 kbit/s ... 2 Mbit/s $C_1(1, 3, 0)$; bit rate 4...80 Mbit/s $C_2(2, 2, 0)$; bit rate 24...1024 Mbit/s, Web, security for applications $C_3(0, 4, 0)$.

1.2. Time and picture quality mode $D = X \star Y$: *1.2.1.* Time mode X: delay $X_1(0, 2, 2)$, real time, low delay $X_2(1, 3, 0)$; *1.2.2.* Picture quality Y: low $Y_1(0, 3, 1)$, good $Y_2(3, 1, 0)$, variable $Y_3(1, 3, 0)$.

1.3. Format $E = U \star V$: *1.3.1.* Resolution U: low $U_{1,2,1}(3)$, high $U_{0,2,2}$; *1.3.2.* Color decomposition V: basic $V_1(1, 2, 1)$, high profile $V_2(0, 2, 2)$.

1.4. Basic operations $F = T \star P \star M$: *1.4.1.* Transformation T: basic mode $T_1(1, 3, 0)$, Dolby Digital $T_2(1, 2, 1)$; *1.4.2.* Playback/features P: basic mode $P_1(1, 3, 0)$, with scalability $P_2(1, 2, 1)$; *1.4.3.* Streaming (video, audio, synchronization, streaming data, testing, control) M: basic mode $M_1(1, 3, 0)$, media objects $M_2(1, 2, 1)$, real-time streaming $M_3(0, 3, 1)$.

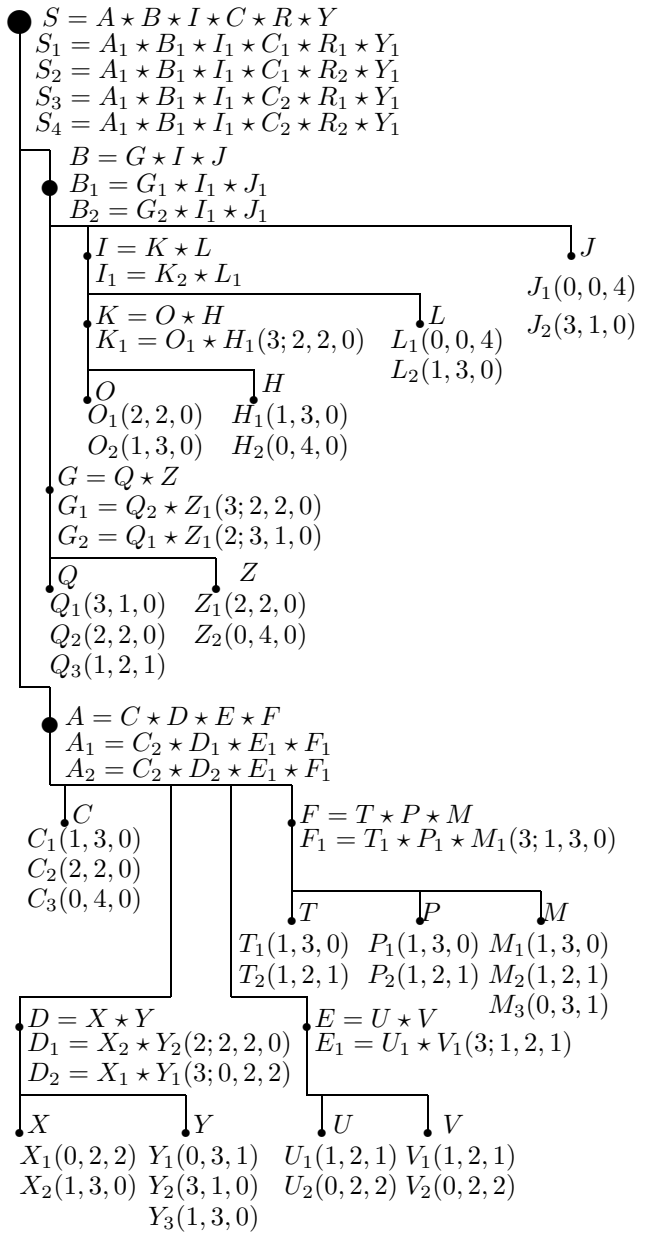


Figure 2. Considered structure of MPEG-like standard

2. Coding/compression $B = G \star I \star J$.

2.1. Basic components $G = Q \star W \star Z$: *2.1.1.* Video coding (methods for transformation into digital codes) Q: variable word length coding of coefficient structure $Q_1(1)$, VLC tables for DCT (non-linear) $Q_2(2)$, flexibility of coding for audio/videa $Q_3(2)$; *2.1.2.* Motion estimation (vector, etc.) Z: $-1024...1023$ pixel (for half) $Z_1(3)$, $-2048...2047$ pixel (for full) $Z_2(1)$.

2.2. Principles and structure $I = K \star L$:

2.2.1. Principles $K = O \star H$: *2.2.1.1.* Block decomposition O: 16 $times16$ (macroblock) and 8×8 (block) $O_1(2, 2, 0)$, object-based (VOB) $O_2(1, 3, 0)$; *2.2.1.2.* Scanning H: progressive scan (zigzag) $H_1(1, 3, 0)$, alternative $H_2(0, 4, 0)$.

2.2.2. Structure (basic processing scheme, extended processing scheme, 'open structure' including transcoding) L: basic mode $L_1(1)$, separation of motion and texture data $L_2(2)$.

2.3. Algorithms J: simple $J_1(2)$ complicated $J_2(1)$.

Interval multiset estimates of DAs are presented in Fig. 2 (in parentheses, expert judgment, illustrative character).

The following abbreviations are used hereinafter: Dolbi Digital (format Dolbi Digital), VLC (Variable-Length Coding), DCT (Discrete Cosine Transform),

Compatibility estimates are presented in Table 1 and Table 2 (expert judgment).

The obtained intermediate composite DAs for subsystems are the following: (1) $D_1 = X_2 \star Y_2$, $N(D_1) = (2; 2, 2, 0)$; (2) $D_2 = X_1 \star Y_1$, $N(D_2) = (3; 0, 2, 2)$; (3) $E_1 = U_1 \star V_1$, $N(E_1) = (3; 1, 2, 1)$; (4) $F_1 = T_1 \star P_1 \star M_1$, $N(F_1) = (3; 1, 3, 0)$; (5) $G_1 = Q_2 \star Z_1$, $N(G_1) = (3; 2, 2, 0)$; (6) $G_2 = Q_1 \star Z_1$, $N(G_1) = (2; 3, 1, 0)$; (7) $K_1 = O_1 \star H_1$, $N(K_1) = (3; 2, 2, 0)$. Fig. 3 illustrates quality of intermediate composite DAs for subsystems D, F, G, K.

TABLE I. COMPATIBILITY ESTIMATES

	Y_1	Y_2	Y_3
X_1	3	2	2
X_2	2	2	2

	V_1	V_2
U_1	3	2
U_2	2	3

	P_1	P_2	M_1	M_2
T_1	3	2	3	1
T_2	1	3	1	3
P_1			3	2
P_2			2	3

TABLE II. COMPATIBILITY ESTIMATES

	Z_1	Z_2
Q_1	2	1
Q_2	3	2
Q_3	1	1

	H_1	H_2
O_1	3	2
O_2	2	3

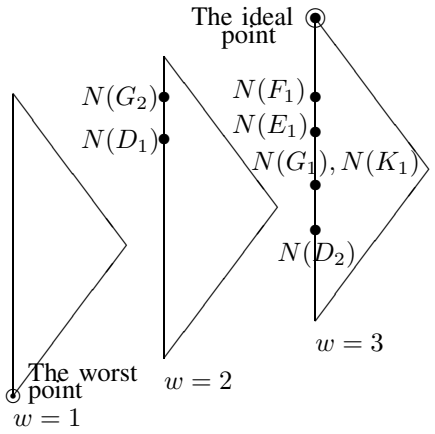


Figure 3. Quality for D,F,G,K

Further, combinations at the next higher level of the hierarchical system model are as follows:

(a) $A_1 = C_2 \star D_1 \star E_1 \star F_1$, (b) $A_2 = C_2 \star D_2 \star E_1 \star F_1$,

(c) $B_1 = G_1 \star I_1 \star J_1$, (d) $B_2 = G_2 \star I_1 \star J_1$.

Finally, four alternative combinations for the designed system are obtained:

(1) $S_1 = A_1 \star B_1 = (C_2 \star D_1 \star E_1 \star F_1) \star (G_1 \star I_1 \star J_1) = C_2 \star (X_2 \star Y_2) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (Q_2 \star Z_1) \star (O_1 \star H_1) \star L_1 \star J_1$;

(2) $S_2 = A_2 \star B_1 = (C_2 \star D_2 \star E_1 \star F_1) \star (G_1 \star I_1 \star J_1) = C_2 \star (X_1 \star Y_1) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (Q_2 \star Z_1) \star (O_1 \star H_1) \star L_1 \star J_1$;

(3) $S_3 = A_1 \star B_2 = (C_2 \star D_1 \star E_1 \star F_1) \star (G_2 \star I_1 \star J_1) = C_2 \star (X_2 \star Y_2) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (Q_1 \star Z_1) \star (O_1 \star H_1) \star L_1 \star J_1$;

(4) $S_4 = A_2 \star B_2 = (C_2 \star D_2 \star E_1 \star F_1) \star (G_2 \star I_1 \star J_1) = C_2 \star (X_1 \star Y_1) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (Q_1 \star Z_1) \star (O_1 \star H_1) \star L_1 \star J_1$.

Note, the initial set of possible solutions includes 82944 combinations.

IV. IMPROVEMENT/RECONFIGURATION

Generally, system improvement process can be based on the following methods (e.g., [9]):

(i) improvement of a system component (element),

(ii) improvement of compatibility between system components,

(iii) change a system structure, for example, extension of the system by addition of system components/parts.

Here, the system improvement (or reconfiguration) process is briefly presented as a combination of improvement actions by elements and improvement actions by compatibility. The illustrative improvement actions are presented in Table 3.

 ISBN: 978-1-61208-262-2

TABLE III. BOTTLENECKS AND IMPROVEMENTS

Composite DAs	Bottlenecks		Improvement actions
	DA	IC	
$D_1 = X_2 \star Y_2$	X_2		$(1,3,0) \Rightarrow (4,0,0)$
$D_1 = X_2 \star Y_2$		(X_2, Y_2)	$2 \Rightarrow 3$
$D_2 = X_1 \star Y_1$	X_1		$(0,2,2) \Rightarrow (2,2,0)$
$D_2 = X_1 \star Y_1$	X_1		$(0,2,2) \Rightarrow (4,0,0)$
$D_2 = X_1 \star Y_1$	Y_2		$(3,1,0) \Rightarrow (4,0,0)$
$E_1 = U_1 \star V_1$	U_1		$(1,2,1) \Rightarrow (2,2,0)$
$E_1 = U_1 \star V_1$	U_1		$(1,2,1) \Rightarrow (4,0,0)$
$E_1 = U_1 \star V_1$	V_1		$(1,2,1) \Rightarrow (2,2,0)$
$E_1 = U_1 \star V_1$	V_1		$(1,2,1) \Rightarrow (4,0,0)$
$G_1 = Q_2 \star Z_1$	Q_2		$(2,2,0) \Rightarrow (4,0,0)$
$G_2 = Q_1 \star E_2$		(E_3, Z_1)	$2 \Rightarrow 3$
$G_2 = Q_1 \star Z_1$	Q_1		$(3,1,0) \Rightarrow (4,0,0)$
$G_2 = Q_1 \star Z_1$	Z_1		$(2,2,0) \Rightarrow (4,0,0)$
$K_1 = O_1 \star H_1$	O_1		$(2,2,0) \Rightarrow (4,0,0)$
$K_1 = O_1 \star H_1$	H_1		$(1,3,0) \Rightarrow (4,0,0)$

Further, the following improvement actions are examined (binary variables $\{y_{i,j}\}$ are used):

(1) Two versions for X_2: $y_{1,1}$ (none), $y_{1,2}$ ($(1,3,0) \Rightarrow (4,0,0)$);

(2) Two versions for (X_2, Y_2): $y_{2,1}$ (none), $y_{2,2}$ ($2 \Rightarrow 3$);

(3) Three versions for X_1: $y_{3,1}$ (none), $y_{3,2}$ ($(0,2,2) \Rightarrow (2,2,0)$); $y_{3,3}$ ($(0,2,2) \Rightarrow (4,0,0)$);

(4) Two versions for Y_2: $y_{4,1}$ (none), $y_{4,2}$ ($(3,1,0) \Rightarrow (4,0,0)$);

(5) Three versions for U_1: $y_{5,1}$ (none), $y_{5,2}$ ($(1,2,1) \Rightarrow (2,2,0)$), $y_{5,3}$ ($(1,2,1) \Rightarrow (4,0,0)$);

(6) Three versions for V_1: $y_{6,1}$ (none), $y_{6,2}$ ($(1,2,1) \Rightarrow (2,2,0)$), $y_{6,3}$ ($(1,2,1) \Rightarrow (4,0,0)$);

(7) Two versions for Q_2: $y_{7,1}$ (none), $y_{7,2}$ ($(2,2,0) \Rightarrow (4,0,0)$);

(8) Two versions for (Q_1, Z_1): $y_{8,1}$ (none), $y_{8,2}$ ($2 \Rightarrow 3$);

(9) Two versions for Q_1: $y_{7,1}$ (none), $y_{7,2}$ ($(3,1,0) \Rightarrow (4,0,0)$);

(10) Two versions for Z_1: $y_{10,1}$ (none), $y_{10,2}$ ($(2,2,0) \Rightarrow (4,0,0)$);

(11) Two versions for O_1: $y_{11,1}$ (none), $y_{11,2}$ ($(2,2,0) \Rightarrow (4,0,0)$);

(12) Two versions for H_1: $y_{12,1}$ (none), $y_{12,2}$ ($(1,3,0) \Rightarrow (4,0,0)$).

Table 4 contains binary variables (y_{ij}), improvement actions and their estimates (illustrative, expert judgment). As a result, the improvement problem is (q_j equals the number of corresponding versions):

$$\arg\min_{M \in D} \sum_{(i,j) \in S = \{(i,j) | y_{ij} = 1\}} |\delta(M, e_{ij})|,$$

$$s.t. \quad \sum_{i=1}^{12} \sum_{j=1}^{q_i} a_{ij} y_{ij} \le b; \quad \sum_{j=1}^{q_j} y_{ij} = 1;$$

$$y_{ij} \in \{0,1\}. \qquad (5)$$

Some examples of the improvement solutions are (improvement problem corresponds to certain solution, i.e., S_1, S_2, S_3, S_4):

(1) S_1, $b = 41$: $y_{1,2} = 1$ (X_2, improvement 1), $y_{4,2} = 1$ (Y_2, improvement 1), $y_{8,2} = 1$ ((Q_1, Z_1), improvement 1); $y_{9,2} = 1$ (Q_1, improvement 1);

$S_1 \Rightarrow \widetilde{S_1} = C_2 \star (\widetilde{X_2} \star \widetilde{Y_2}) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (\widetilde{Q_1} \star Z_1) \star (O_1 \star H_1) \star L_1 \star J_1$;

(2) S_3, $b = 28$: $y_{1,2} = 1$ (X_2, improvement 1), $y_{4,2} = 1$ (Y_2, improvement 1), $y_{8,2} = 1$ ((Q_1, Z_1), improvement 1);

$S_3 \Rightarrow \widetilde{S_3} = C_2 \star (\widetilde{X_2} \star \widetilde{Y_2}) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (Q_1 \star Z_1) \star (O_1 \star H_1) \star L_1 \star J_1$;

(3) S_4, $b = 43$: $y_{3,2} = 1$ (X_1, improvement 1), $y_{9,2} = 1$ (Q_1, improvement 1), $y_{10,2} = 1$, (Z_1, improvement 1);

$S_4 \Rightarrow \widetilde{S_4} = C_2 \star (\widetilde{X_1} \star Y_1) \star (U_1 \star V_1) \star (T_1 \star P_1 \star M_1) \star (\widetilde{Q_1} \star \widetilde{Z_1}) \star (O_1 \star H_1) \star L_1 \star J_1$.

TABLE IV. ESTIMATES OF IMPROVEMENTS

Improvement actions	Multiset estimate e_{ij}	Cost (a_{ij})
$y_{1,1}$ (X_2, None)	$(0,0,4)$	0
$y_{1,2}$ (X_2, Improvement 1)	$(4,0,0)$	10
$y_{2,1}$ ((X_2, Y_2), None)	$(0,0,4)$	0
$y_{2,2}$ ((X_2, Y_2), Improvement 1)	$(4,0,0)$	14
$y_{3,1}$ (X_1, None)	$(0,0,4)$	0
$y_{3,2}$ (X_1, Improvement 1)	$(2,2,0)$	10
$y_{3,3}$ (X_1, Improvement 2)	$(4,0,0)$	29
$y_{4,1}$ (Y_2, None)	$(0,0,4)$	0
$y_{4,2}$ (Y_2, Improvement 1)	$(4,0,0)$	9
$y_{5,1}$ (U_1, None)	$(0,0,4)$	0
$y_{5,2}$ (U_1, Improvement 1)	$(2,2,0)$	17
$y_{5,3}$ (U_1, Improvement 2)	$(4,0,0)$	30
$y_{6,1}$ (V_1, None)	$(0,0,4)$	0
$y_{6,2}$ (V_1, Improvement 1)	$(2,2,0)$	12
$y_{6,3}$ (V_1, Improvement 2)	$(4,0,0)$	24
$y_{7,1}$ (Q_2, None)	$(0,0,4)$	0
$y_{7,2}$ (Q_2, Improvement 1)	$(4,0,0)$	15
$y_{8,1}$ ((Q_1, Z_1), None	$(0,0,4)$	0
$y_{8,2}$ ((Q_1, Z_1), Improvement 1)	$(3,1,0)$	9
$y_{9,1}$ (Q_1, None)	$(0,0,4)$	0
$y_{9,2}$ (Q_1, Improvement 1)	$(4,0,0)$	13
$y_{10,1}$(Z_1, None)	$(0,0,4)$	0
$y_{10,2}$(Z_1, Improvement 1)	$(4,0,0)$	20
$y_{11,1}$(O_1, None)	$(0,0,4)$	0
$y_{11,2}$(O_1, Improvement 1)	$(4,0,0)$	18
$y_{12,1}$(H_1, None)	$(0,0,4)$	0
$y_{12,2}$(H_1, Improvement 1)	$(4,0,0)$	22

 ISBN: 978-1-61208-262-2

V. Conclusion and Future Work

The paper describes a hierarchical approach to modular design of MPEG-like standard. Note, this material is a preliminary one and is based on illustrative estimates of design alternatives for elements of the standard (and estimates of compatibility between the design alternatives). The described illustrative solving schemes can be considered as prototype frameworks for real-world applications.

In the future, it may be prospective to consider the following research directions:

1. Consideration of other design and improvement (adaptation) problems in analysis and generation of communication standards;

2. Special simulation research to analyze various versions of communication standards;

3. Usage of AI techniques; and

4. Usage of the described application and design approach in engineering/CS education.

Acknowledgment

The author would like to thank the partial support of the work by grant of the Ministry of Education and Science of Russian Federation "QoS provisioning methods for the access to broadband multimedia services in wireless self-organized networks" (application no. 2012-1.2.1.-12-000-2006-009).

References

[1] L. Chiariglione, "MPEG and multimedia communications," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 2, 1997, pp. 5–18.

[2] C. E. Fogg, D. J. Le Gall, J. L. Mitchell, and W. B. Pennebaker, Eds., *MPEG Video. Compression Standard.* Boston: Kluwer Academic Publishers, 2002.

[3] M. R. Garey and D. S. Johnson, *Computers and Intractability. The Guide to the Theory of NP-Completeness*. San Francisco: W. H. Freeman and Company, 1979.

[4] E. Kamensky and O. Hadar, "Multiparameter Method for Analysis and Selection of Motion Estimation Algorithms for Video Compression," *Multimedia Tools and Applications*, vol. 38, no. 1, 2008, pp. 119–146.

[5] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems.* Berlin: Springer, 2004.

[6] D. E. Knuth, *The Art of Computer Programming. Vol. 2, Seminumerical Algorithms.* Mass: Addison Wesley, Reading, 1998.

[7] P. Kuhn, *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation.* Boston: Kluwer Academic Publishers, 1999.

[8] D. Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications," Comm. of the ACM, vol. 34, no. 4, 1991, pp. 47–58.

[9] M. Sh. Levin, *Composite Systems Decisions.* New York: Springer, 2006.

[10] M. Sh. Levin, "Multiset Estimates and Combinatorial Synthesis," Electronic preprint. 30 pp., May 9, 2012. http://arxiv.org/abs/1205.2046 [cs.SY]

[11] M. Sh. Levin, "Combinatorial Synthesis of Communication Protocol ZigBee with Interval Multiset Estimates," Proc. 4nd Int. Congress on Ultra Modern Telecomm. & Control Systems (ICUMT-2012), IEEE Press, Oct. 2012, pp. 29–34.

[12] M. Sh. Levin, "Modular Approach to Communication Protocol and Standard for Multimedia Information (Survey)," Information Processes, vol. 12, no. 4, 2012, pp. 389–399 (in Russian).

[13] M. Sh. Levin, "Towards Combinatorial Evolution of Composite Systems," Expert Systems with Applications, vol. 40, no. 4, 2013, pp. 1342–1351.

[14] M. Sh. Levin, O. Kruchkov, O. Hadar, and E. Kaminsky, "Combinatorial Systems Evolution: Example of Standard for Multimedia Information," Informatica, vol. 20, no. 4, 2009, pp. 519–538.

[15] R. R. Yager, "On the Theory of Bags," Int. J. of General Systems, vol. 13, no. 1, 1986, pp. 23–37.

 ISBN: 978-1-61208-262-2

Peer to Peer Location Sharing

Dmitry Namiot
Lomonosov Moscow State University
Faculty of Computational Math and Cybernetics
Moscow, Russia
dnamiot@gmail.com

Manfred Sneps-Sneppe
Ventspils University College
Ventspils International Radio Astronomy Centre
Ventspils, Latvia
manfreds.sneps@gmail.com

Abstract— **This paper describes several new models for sharing location information without disclosing identity data to some third party server. All our services proposed in this research share the common background idea. It could be described as a safe location sharing. It combines server-side (centralized) location information with the locally based distributed identity information. In this distributed data store, identity info is always saved locally. The proposed approach eliminates one of the biggest concerns for location based systems adoption – privacy. This article describes our approach as well as several generic implementations.**

Keywords-location; privacy; lbs; mobile; HTML5; geo coding; social networks

I. INTRODUCTION

It is a well-known fact that the question "Where are you?" is one of the most often asked during the communications. More than 600 billion text messages per year in the US ask "Where are you?" [1]. A huge amount of services use location exchange as a key feature. Location plays a basic role in context-aware applications. The classical definition [2] describes context as location, identities of nearby people and objects, and changes to those objects. As per N. Hristova [3], context-related information can consist of user profiles and preferences, their current location, the type of connection that to the mobile network, the type of wireless device being used, the objects that are currently in the user's proximity, and/or information about their behavioral history. Actually, most of the authors define context awareness as complementary element to location awareness, whereas location may serve as a determinant for resident processes. By this reason, all the context-aware applications are linked to location exchange.

Location, as one of the most widely adopted sensor readings of a modern smart phone, is probably the first attribute (candidate) to share for mobile users. The typical applications are well known and include for example geo-tagged context, friend-finder, recommendation systems, turn-by-turn navigation, etc.

In location-based service (LBS) scenarios we can describe the following actors [4]:

- Intended recipient. For example, the service company, friends, parents, etc. This usually involves the use of a service provider that offers to forward your location to the intended recipient. Actually, it is a main goal for our research. How can we deal with intended recioients without telling too much data to service providers? The final goal is to create a safe location sharing system without explicit centre for all circulated data.
- Service provider. For example, Google providing you with the Latitude application or Yelp provides a restaurant recommendation system for near-by places. In contrast to the intended recipient, users usually do not have a primary goal of letting the service provider know their location – it is a by-product of getting a restaurant review or staying in touch with friends.
- Infrastructure provider. The typical example is a mobile operator. While self-positioning systems such as GPS can work without an infrastructure provider, mobile phone users are often implicitly located in order to provide communication services (for example, route phone calls or emergency communications).
- Unintended recipients. For example, we can mention accidental recipient, illegal recipient and law enforcement.

In the most cases, by describing various LBS, we assume that for a given system, the infrastructure provider needs to be trusted. In other words, the need for sharing location data with infrastructure providers is always a non-discussable topic. In the same time, Palen [5] argues that privacy is not simply a problem of setting rules and enforcing them. It is rather an ongoing and organic process of negotiating boundaries of disclosure, identity, and time. Authors suggest genres of disclosure for managing interpersonal privacy, which are "socially-constructed patterns of privacy management," as a sort of design pattern approach to support the development of privacy-sensitive applications. Examples might include creating and managing accounts at shopping Web sites, taking appropriate photographs at social events, exchanging contact information with a new acquaintance, etc. [6].

This paper summarizes our efforts in safe location sharing. The rest of the paper is organized as follows. Section II contains an analysis of existing projects devoted to privacy in location sharing applications. In Section III, we consider our Geo Messages service and related applications. In Section IV, we describe our WATN application..

II. LOCATION SHARING AND PRIVACY

In the most cases, location sharing is implemented as the ability for the mobile user (participants) write down (save)

 ISBN: 978-1-61208-262-2

own location info in the some special place (e.g., special mobile application).

But, it means of course that user must be registered in this service or deploy some specially downloaded application. What is more important here – everyone who needs this information must use the same service too [1]. This chicken-egg problem is a typical for many LBS applications. The service is useless until many users register there, but there are no reasons to register due to lack of useful information from userless service.

One of the biggest concerns for all location-based services is user's privacy. Despite the increased availability of these location-sharing applications, we have not seen yet wide adoption for the most of them. It has been suggested that the reason for this lack of adoption may be users' privacy concerns regarding the sharing and use of their location information.

For example, the widely cited review of social networks practices [7] concluded, that location information is preferably shared on a need to know basis, not broadcast.

Participants were biased against sharing their location constantly, without explicit consent each time their location is requested. This suggests that people are cautious about sharing their location and need to be reassured that their private information is only being disclosed when necessary and is not readily available to everybody.

The key point for any existing service is some third party server that keeps identities and locations. We can vary the approaches for sharing (identity, locations) pairs but we could not remove the main part in privacy related concerns – the third part server itself.

As it is mentioned in [8], peer opinion and technical achievements contribute most to whether or not participants thought they would continue to use a mobile location technology. In this connection, Hong [6] suggests the following end-user and application developer requirements divided into four high-level groups:

- A decentralized architecture, where as much personal information about an end-user is captured, stored, and processed on local devices owned by that end-user.
- A range of mechanisms for control and feedback by end users over the access, flow, and retention of personal information, to support the development of pessimistic, optimistic, and mixed-initiative applications.
- A level of plausible deniability built in.
- Special exceptions for emergencies.

Actually, we will present below some mix (mashup) of decentralized and server-based architectures.

One possible solution is using peer-to-peer location sharing. The easiest way to apparently "solve" location privacy problems is to authorize (or do not authorize) manually or automatically the disclosure of location information to others. But, we should see in the same time the other privacy related problem that is not eliminated. Your location will be disclosed to (saved on) some third party server. For example, you can share location info in Google Latitude on "per friend" mode, but there is still some third party server (Google) that keeps your location and your identity. In other words, we still should have some trusted source. This source keeps all information.

Typically we have now two models for location sharing in services. At the first hand, it is some form of passive location monitoring and future access to the accumulated data trough some API; it is Google Latitude, for example. The first possible problem with this approach is privacy. There is some third party tool that constantly monitors my location and what is more important – saves it on the some external server. The second problem is the shorted life for handset's batteries.

Another model for location sharing is check-in. It could be an active (e.g., Foursquare), when user directly writes down his current location or passive (e.g., Twitter), when location info could be added to the current message. A check-in is a simple way to keep tabs on where you've been, broadcast to your friends where you are, and discover more about other people in your community. But, here we can see not only privacy related issues, when all my friends (followers) can see my location. We have here also a noise related issue too - my location info could be actually interested for the physical friends only. For the majority of followers my location info (e.g., Foursquare's statuses in my Twitter's time line) is just a noise [9].

There is another interesting solution associated with check-ins. At the first hand, technically check-ins could be customized [10]. But, after the customization, we can make the next step also and create a new form of check-ins that is disconnected from the social stream. We can create a new type of check-in records and separate them from rest of stream. It means that we will provide a separate database that just contains a list of accounts from the social network being concentrated (at this moment!) nearby some place. It is a temporal database, check-in records could be changed constantly and it does not contain the social stream itself – just IDs (e.g., nick names) for accounts confirmed check-ins. Figure 1 illustrates the new check-in form.

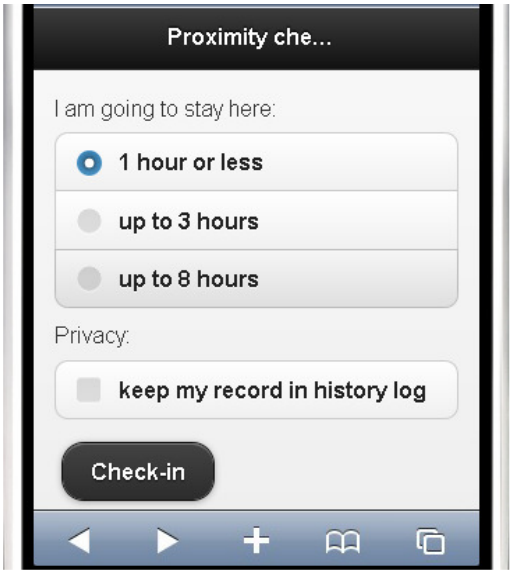


Figure 1. External check-in

 ISBN: 978-1-61208-262-2

It is one of the features for the upcoming new SpotEx [11] version. This temporal database lists people who are currently nearby some selected point (Wi-Fi access point in Spotex). And more important, that this database is visible only to those of people who are nearby themselves too. It is some form of temporal geo-limited location sharing [12].

Lets us describe some existing approaches in LBS development that targets the privacy.

One of the most popular methods for location privacy is obfuscation [13]. Obfuscating location information lowers its precision, e.g., showing only street or city level location instead of the actual coordinates, so that the visible (within our system) location does not correspond to the real one. For example, in Google Latitude we can allow some of the users get our own location info on the city level only. Sometimes even the random noise could be added to the real location data [14]. But, once again – it is just a visible location. The central point (points) for such a system can have all the information.

Some papers prefer the term spatial cloaking and describe it as the most commonly used privacy-enhancing technique in LBS. In spatial cloaking algorithm, mobile and stationary users can entertain location-based services without revealing their exact location information. The main idea is that before requesting any location-based service, the mobile user will form a group from his peers via single-hop communication and/or multi-hop routing. Then, the spatial cloaked area is computed as the region that covers the entire group of peers. [15].

Another popular approach in the area of location privacy is "k-anonymity" [16]. As per this approach, the actual location is substituted by a region containing at least k − 1 other users, thus ensuring that a particular request can only be attributed to "1 out of k" people. Of course, this approach has the disadvantage that if the region contains too few people, it has to be enlarged until it contains the right number of people. But, in general k-anonymity protects identity information in a location-oriented context [17]. In the same time, the group-composing algorithm is complex and the member peers are dynamic. The big question here is the level of implementation. Who is responsible for this anonymity? In other words, what kind of data do we have inside of our system – anonymous location info right from the moment data being put into our system or it is just a view and our data saved internally in the raw formats?

Of course, the deployment of location privacy methods depends on the tasks our system is going to target. For example, obfuscating location information in case of emergency help system could not be a good idea. But, from other side many geo-context aware applications (e.g., geo search) can use approximate location info.

Also we need to highlight the role of identity in LBS. It looks like combining identity with location info is just an attempt for delivering more targeted advertising rather then the need of the services themselves. It is obvious, for example, that local search for some points of interests (e.g., café) should work for the anonymous users too.

More traditionally, peer-to-peer LBS refer to the way sharing information is traversed over the network [18]. For example, the peer-to-peer k-anonymity algorithm has several steps: select a central peer who will act as an agent for the group, next, the central peer will discover other k-1 different peers via single-hop or multi-hop to compose the group, and finally find a cloaked region covering all locations that every peer may arrive.

In our article, we are using "peer-to-peer" term at the first hand for highlighting the target party for the location-sharing request. It is "another peer" directly, rather than the central server (data store).

In terms of patterns for LBS, this approach targets at the first hand such tasks as 'Friend finder' and the similar. In other words, it is anything that could be linked to location monitoring.

III. GEO MESSAGES

Our idea of the signed geo messages service (geo mail, geo SMS) based on the adding user's location info to the standard messages like SMS or email. Just as a signature. So with this service for telling somebody 'where I am' it would be just enough to send him/her a message. And your partner does not need to use any additional service in order to get information about your location. All the needed information will be simply delivered to him as a part of the incoming message.

It is obviously peer-to-peer sharing and does not require any social network. And it does not require one central point for sharing location with by the way. Our location signature has got a form of the map with the marker at the shared location. And what is important here – the map itself has no information about the sender and recipient. That information exists only in the message itself. The map (marker) has no information about the creator for example. That is all about privacy [9].

Signed geo message service offers a mobile web mashup that lets users add a signature with geo information to the standard messages (SMS, email). As any other signature, our signature is just a text. And this text simply contains a dynamically generated link that leads either directly to the mobile map or to some landing page where mobile map is a part of it. And that mobile map (visible area) shows the current position of the sender. It is where the name of service is coming from – signed geo messages or geo signatures. The Figure 2 illustrates this approach. This figure illustrates Geo Mail client. Geo Messages approach has been implemented as a series of web mashups. All of them are based on the same principles. For example, Geo Mail mashup detects user's location (via W3C Geo location, supported in HTML5 browsers) and lets user choose the format of the signature (static or dynamic map, just a pair of latitude, longitude). By the similar principles we can share location info via SMS (it could be based on SMS URI scheme), Facebook Messages, Twitter's direct messages, etc.

The interesting moment that should be mentioned here, is the technical ability to implement such approach on the level of SIM-card. In other words, it could be provided as a standard feature by telecom operators. As it is described in [9], it could be implemented via Smart Cards Web Server Servlet that requests local info. This servlet can perform

 ISBN: 978-1-61208-262-2

proactive command for getting Location Information (MCC, MNC, LAC and Cell Identity). Such approach let us provide web mashups for smart cards. Obtained location info could be processed by the external web service. External service will convert Cell Id data into location data (latitude, longitude) and use obtained information for signing outgoing requests.

Figure 2. Geo Mail client

Because we can use Geo Messages as web mashups, there is no need to download (install) some applications in order to share location info. Of course, there is a common problem with mobile users that are not aware about Geo Email at all. One viral trick suggests modifications for Geo Mail, so the modified version sends a link to Geo Mail application instead of the question 'Where are you?'. For example, a signature for Geo Mail can include a link to Geo Mail itself. It lets the recipients answer to the question 'Where are you?' just by the opening Geo Mail link.

The next step is Geo Mail integration with Address Book on mobile. This feature at thi moment could be implemented via mobile application only. Currently, it exists as Android app. We are looking a way for creting portable version for it.

IV. WATN

Geo Messages approach works and really eliminates the problems with identity revealing. It provides a good solution for 'Where Are You Now?' question, but in the same time is not very convenient for the constant monitoring with several participants. It is simply not very convenient to jump from one message to another. WATN ('Where Are They Now') [1] application solves this problem. It provides a new peer-to-peer service that solves the privacy issues and lets you deal with several location-related feeds (location peers) simultaneously.

In Geo Messages approach the standard header for messages (e.g., 'From' header field for email, phone number for SMS, etc.) has been used for the identification. There is no own identification in Geo Messages. For multiple location-related feeds, we need to identify (distinguish) them by some way.

WATN provides own identification scheme, but in the same time it separates locations and identity. Actually, it is the basic idea behind WATN. In other words, rather than use one server that keeps all our data in some centralized system (like Google's Lattitude), we will switch to some mashup of distributed and centralized architectures.

We can separate location info and identity data just in three steps:

a) assign to any participant some unique ID (just an ID, without any connection to the personality)

b) save location data on the central server with links to the above-mentioned IDs

c) keep the legend (descriptions for IDs, who is behind that ID) locally

In this case, any participant may request location data for other participants from third party server (as per sharing rules, of course), get data with meaningless IDs and map them against locally saved database with names. With such replacement we can show location data in the "natural" form (replace meaningless IDs with mnemonic names). And in the same time, our server (third party server for our users) is not aware about IDs decrypting.

Obviously, that in this model each client keeps own legend info. And because our clients are not aware about each other and there are no third party servers that know all registered clients. It means, obviously, that in this model the same ID may have different legends. Technically, each client can assign own name (nick) for the same ID. Our social graph saves information (links between participants) using the above-mentioned meaningless IDs only. And the human readable interpretation for that graph can vary from client to client.

The next basic moment is the implementation for our distributed system. The local sub-part has been based on HTML5 standards. There is so called local storage specification [19]. As per W3C documents, HTML5 web storage is local data storage, web pages can store data within the user's browser.

The concept is similar to cookies, but it's designed for larger quantities of information. Cookies are limited in size, and mobile browser can send them back to the web server (use them as a part of HTTP requets). Web Storage is more secure and faster and our data is not included with every server request, but used only when asked for. It is also possible to store large amounts of data, without affecting the website's performance. The data is stored in key/value pairs, and web pages can only access data stored by them. In other words, Web Storage follows to standard same-origin policies.

Storage is defined by the HTML5 standards as this:

interface Storage {

 ISBN: 978-1-61208-262-2

```
    readonly attribute unsigned long length;
    DOMString key(unsigned long index);
    getter DOMString getItem(DOMString key);
    setter creator void setItem(DOMString key, DOMString value);
    deleter void removeItem(DOMString key);
    void clear();
};
```

The DOM Storage mechanism is a means through which string key/value pairs can be securely stored and later retrieved for use. The goal of this addition is to provide a comprehensive means through which interactive applications can be built (including advanced abilities, such as being able to work "offline" for extended periods of time).

User agents must have a set of local storage areas, one for each origin. User agents should expire data from the local storage areas only for security reasons or when requested to do so by the user. In our projects Local Storage keeps identification data. It is a typical key-value system. Key is user's ID, and value is user's name (nick, alias).

Now we are ready to describe the whole algorithm. Each user automatically obtains own ID. For the first time visitors, web service generates a new ID, for the returning users ID will be extracted from user's local storage. As soon as ID is obtained, user can share his location information. There is no registration system, so "share location" requests could be addressed to any person with valid email address. Technically, "Share my location" request is just an ordinary email with the link to WATN service. This link contains an ID for the person who is going to share location. It is an ordinary email (or SMS) and WATN service is completely unaware about the target address. Actually, it is one of the main features. The communications in WATN are completely separated from the service.

As soon as the email is received, the recipient can open the shared link. It is a new request (hit) to WATN service. As any other request to WATN it caused user's ID detection (assign a new one or select existing from local storage – see above). It means that for such kind of requests ('share location') the service collects two IDs. One of them is user's ID, and the second one is passed as a parameter. It shows who is sharing location info now.

As the next step for processing 'share location' request, WATN service asks recipient for two things:

a) accept (do not accept) request
b) assign a new name (nick name, alias) for ID from accepted request

This name choise is completely up to the recipient. Conceptually, it is based on the fact that 'share location' request was opened right from some message (email, SMS). So, the target party in 'share location' request (email's recipient) is aware about the context. Simply, he knows who is an author. He can see email's header ('From' inormation) or phone number (address book info) for SMS. Based on this information, the recipient can choose some name for this correspondent. It is illustrated in Figure 3.

Here is a small additional technical trick. Of course, any name could be assigned. But for names, that corresponds to Twitter's (Facebook's) account it is possible to pull profile picture also. Note, that we cannot use some services like Gravatar, because the email address is unknown for WATN service.

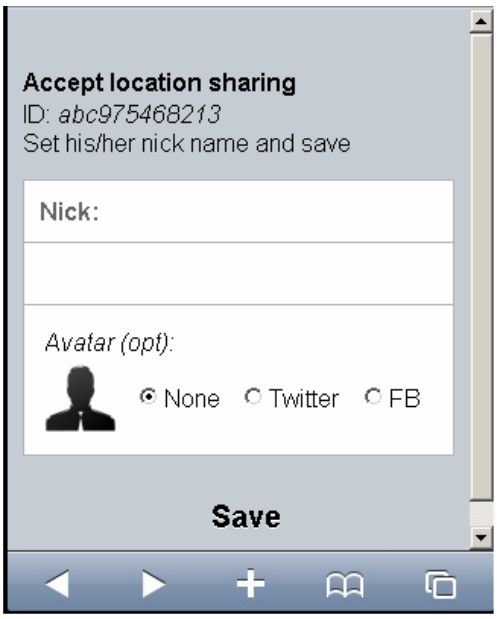


Figure 3. Share location

It explains the process of collecting (filling) local identification database. Each user simply assigns own names (nicks) to IDs provided in 'share with you requests'.

In the same time, if 'share location' request is accepted, we can fix this fact on the server. We have two IDs and can use them for creating social graph record: ID_1 shares location info with ID_2

This information is centralized, but it completely dumb. Our server is completely unaware about the names behind IDs. 'Share location' processing works like a typical two phase commit in distributed database. It saves social graph info on the server and saves identification info locally. It is how our distributed database within WATN is organized. It has centralized store for social graph (who is sharing location info with whom) and local store with identity data.

Now we can go back to the whole algorithm. Lets us return to the first step. As soon as any user hits WATN web site, mashup detects his ID. Having ID, we can pull data from the social graph. Actually, WATN engine returns extracted social graph info as JSON array. At the first hand, this array contains a list of IDs for participants who shared location info with the given user. Simultaneously, mashup records a new check-in (saves location info) for the current user. This information is anonymous again and contains meaningless ID, time and geo coordinates. Check-ins let us detect the current (more precisely – last known) location for the each participant. It let us add location into to the above mentioned extraction from the social graph. So, WATN

 ISBN: 978-1-61208-262-2

engine returns to the client IDs and locations. This JSON array could be mapped (on the client side) against locally saved identity info. This mapping replaces IDs with saved names (nicks). And this information could be simply visualized (Figure 4):

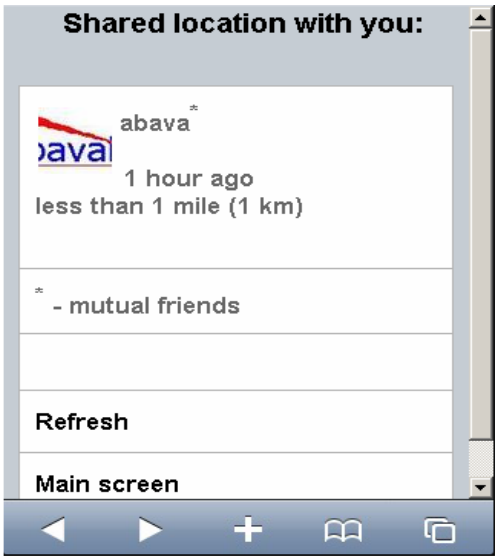


Figure 4. WATN information screen

On this screen ID for the user has been replaced with the mnemonic name (abava) and avatar fom Twitter.

V. CONCLUSION

This paper discussed several new models for sharing location information without disclosing identity data to some third party server. We described existing approaches as well as propose several new implementations. All proposed services share the common background idea about the separation the location and identity data. They could be described as a safe location sharing. It combines server-side (centralized) location info with the locally-based distributed identity info. In this distributed data store identity info is always either saved locally or borrowed from the external services (e.g., messages). The proposed approach eliminates one of the biggest concerns for location based systems adoption – privacy.

REFERENCES

[1] D. Namiot and M. Sneps-Sneppe, "Where Are They Now – Safe Location Sharing. A New Model for Location Sharing Services, Internet of Things", Smart Spaces, and Next Generation Networking, Lecture Notes in Computer Science, 2012, Vol. 7469/2012, pp. 63-74, DOI: 10.1007/978-3-642-32686-8_6

[2] G. Schilit and B. Theimer, "Disseminating Active Map Information to Mobile Hosts", IEEE Network, 8(5) (1994) pp. 22-32

[3] N. Hristova and G. M. P. O`Hare, "Ad-me: Wireless Advertising Adapted to the User Location, Device and Emotions," in Thirty-Seventh Hawaii International Conference on System Sciences (HICSS-37), 2004, pp. 1-10

[4] M. Scipioni, "A privacy-by-design approach to location sharing." Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM, 2012, pp. 580-583

[5] L. Palen and P. Dourish, "Unpacking "Privacy" for a Networked World", CHI Letters, 2003. 5(1): pp. 129-136.

[6] J. Hong and J. Landay, "An Architecture for Privacy-Sensitive Ubiquitous Computing", MobiSys'04, Jun. 6–9, 2004, Boston, Massachusetts, pp. 177-189

[7] D. Wagner, M. Lopez, A. Doria, V. Kostakos, I. Oakley, and T. Spilitiopoulos, "Hide and seek: location sharing practices with social media", Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, September 07-10, 2010, Lisbon, Portugal, pp. 55-58

[8] J. Tsai, P. Kelley, P. Drielsma, L. Cranor, J. Hong, and N. Sadeh ,"Who's viewed you?: the impact of feedback in a mobile location-sharing application", Proceedings of the 27th international conference on Human factors in computing systems, April 04-09, 2009, Boston, MA, USA, pp. 2003-2012

[9] D. Namiot, "Geo messages", Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2010 International Congress pp. 14-19 DOI: 10.1109/ICUMT.2010.5676665

[10] D. Namiot and M. Sneps-Sneppe "Customized check-in Procedures ", Smart Spaces and Next Generation Wired/Wireless Networking Lecture Notes in Computer Science, 2011, Volume 6869/2011, pp. 160-164, DOI: 10.1007/978-3-642-22875-9_14

[11] Y. Daradkeh, D. Namiot, and M. Sneps-Sneppe, "Spot Expert as Context-Aware Browsing", Journal of Wireless Networking and Communications, vol.2, N. 3, 2012, pp. 23-28

[12] D. Namiot, "Context-Aware Browsing -- A Practical Approach", Next Generation Mobile Applications, Services and Technologies (NGMAST), 2012 6th International Conference on, pp. 18-23, DOI: 10.1109/NGMAST.2012.13

[13] M. Duckham and L. Kulik. A FORMAL model of obfuscation and negotiation for location privacy. In Proceedings of Pervasive 2005, Munich, Germany, 2005. Springer. pp. 152–170

[14] J. Krumm, "Inference attacks on location tracks", In Proceedings of the Fifth International Conference on Pervasive Computing (Pervasive), volume 4480 of LNCS, Springer-Verlag, 2007, pp. 127–143

[15] C. Chow, M. Mokbel, and X. Liu, "A peer-to-peer spatial cloaking algorithm for anonymous location-based service", GIS '06 Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems, 2006, pp. 171 – 178 DOI:10.1145/1183471.1183500

[16] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In MobiSys '03: Proceedings of the 1st international conference on Mobile systems, applications and services, New York, USA, 2003., pp. 31–42

[17] M. Langheinrich, "Privacy in ubiquitous computing", Ubiquitous Computing, CRC Press, 2009, pp. 95–160.

[18] J. Xu, J. ZpP, M. Xu, and N.Zheng, "Mobile-Aware Anonymous Peer Selecting Algorithm for Enhancing Privacy and Connectivity in Location-Based Service", e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on Nov. 2010 pp. 172 – 177 DOI: 10.1109/ICEBE.2010.32

[19] M. Casario, P. Elst, C. Brown, N. Wormser, and C. Hanquez, "HTML5 Solutions: Essential Techniques for HTML5 Developers", 2011, pp. 281-303, DOI: 10.1007/978-1-4302-3387-9_11

 ISBN: 978-1-61208-262-2

Estimating Retransmission Timeouts in IP-Based Transport Protocols

Stan McClellan
Ingram School of Engineering
Texas State University
San Marcos, TX USA
stan.mcclellan@txstate.edu

Wuxu Peng
Dept. of Computer Science
Texas State University
San Marcos, TX USA
wuxu@txstate.edu

***Abstract*—This paper analyzes the algorithm used for estimating retransmission timeouts in connection-oriented IP-based transport protocols, such as the Transmission Control Protocol (TCP) and the Stream Control Transmission Protocol (SCTP). The estimation algorithm uses historical values of the round-trip time to estimate future round-trip delays, and so creates a maximum waiting time before triggering retransmission attempts. The purpose of the analysis is to question / validate some of the fundamental assumptions used in the estimation algorithm. The conclusion of the analysis is that the algorithm is somewhat mismatched to the application area. Alternative algorithms are discussed, and potential modifications are presented.**

***Keywords*-SCTP; retransmission timeout; round-trip time; RTT; RTO; Jacobson algorithm; Chebyshev approximation; parameter estimation; upper bound.**

I. INTRODUCTION

IP-based transport protocols such as the Transmission Control Protocol (TCP) [1] and the Stream Control Transmission Protocol (SCTP) [2], [3] estimate maximum round-trip times using data from prior successful transmissions. The purpose of this estimation process is to create a triggering mechanism for retransmission procedures when transmissions are lost or seriously delayed. Estimation of the maximum round-trip time is performed via the Jacobson algorithm [4], which is codified in several IETF RFC's, including RFC 6298 [5]. The Jacobson algorithm has an interesting basis in fundamental theory, but suffers from some performance issues due to a mismatch between the theory and the application area. Performance issues related to the Jacobson algorithm and other retransmission procedures have been noted and addressed in several alternative approaches, including [6]–[11]. This paper discusses the Jacobson algorithm, the theory which motivates it, and several alternative algorithms including a new approach which is a modified form of Jacobson. Section II describes the estimation process and its use in establishing timeouts for retransmission procedures. Related work in retransmission optimization and timeout boundaries is also summarized. In Section III, the parameters of the existing algorithm are analyzed, and in Section IV an alternative approach is presented based on similar theoretical concepts, and achieving improved results. Performance results based on implementation and simulation are summarized in Section V, and Section VI concludes the paper. Note that much of this work is presented in the context of SCTP, but is also applicable to TCP since the timeout estimation processes are identical.

II. RETRANSMISSION MECHANISMS

When an SCTP sender transmits a unit of data, called a *chunk*, it also initializes a *retransmission timer* with an estimated value of the round-trip time (RTT). The value of this timer is the *retransmission time-out* (RTO). When an acknowledgment arrives, the timer is cancelled. If the timer expires before an acknowledgment arrives, the chunk may be retransmitted. The value of RTO is calculated from observed / actual values of RTT using the *Jacobson Algorithm*, which is detailed in Section III. A too optimistic retransmission timer may expire prematurely, producing *spurious timeouts* and *spurious retransmissions*, reducing a connection's effective throughput. On the contrary, a retransmission timer that is too conservative may cause long idle times before lost packets are detected and retransmitted. This can also degrade performance [6]. So, the difficulty lies in finding an algorithm which has a solid theoretical basis, is not computationally expensive, and can predict RTT with enough bias to minimize retransmission events and waiting time *simultaneously*.

Performance issues related to retransmission procedures, including alternatives to the Jacobson Algorithm, have been noted and addressed several times in the literature. Much work has been focused on late retransmission and other optimizations of the overall retransmission scenarios [12], [13]. Many authors approach this problem with a "holistic" or overall perspective on the retransmission procedures where RTT estimation contributes to triggering these procedures. Other authors specifically address the estimation of RTT, and propose completely new algorithms. However, the Jacobson Algorithm is deeply rooted in the fabric of connection-oriented IP transport protocols, and its basis in fundamental theory is well-established. In following subsections, we briefly summarize several important approaches to retransmission and RTO estimation.

 ISBN: 978-1-61208-262-2

A. Holistic approaches

Holistic or overall approaches to improving retransmission performance typically address the state machines surrounding the retransmission process, including the estimation algorithm which may be used to trigger these procedures. Examples of holistic approaches include *Early Fast Retransmit* (EFR) [10], *Early Retransmit* (ER) [11], and *Window-Based Retransmission* (WB-RTO) [14] as well as protocol-specific techniques such as *Thin Streams* [9], [10], [15].

Early Fast Retransmit (EFR) is an optional mechanism in FreeBSD which is active whenever the congestion window is larger than the number of unacknowledged packets, and packets remain to be sent. When the RTO timer expires and the entire congestion window is not used, EFR retransmits all packets that could have been acknowledged [10]. The *Early Retransmit* (ER) algorithm [11] suggests that a mechanism should be in place to recover lost segments when there are too few unacknowledged packets to trigger Fast Retransmit. The Early Retransmit algorithm reduces waiting time in four specific situations [10]. The *Window-Based Retransmission Timeout* (WB-RTO) [14] asserts that timeout mechanisms based solely on RTT estimates lead to unnecessary retransmissions and unfair resource allocation, and proposes to schedule flows on the basis of their contribution to congestion. *Thin Streams* [9], [10], [15] optimizes throughput for "thin streams" which are often used in control applications, and often depend on SCTP for transport. When stream characterization is accurate, throughput can be improved by adapting specific sections of the retransmission procedures to match flow characteristics.

B. Alternative estimation algorithms

Alternative estimation algorithms address specific performance issues which have been noted in the Jacobson Algorithm. These issues may be related to overshoot in the estimated value, spurious behaviors for certain traffic characteristics, or inefficient bounding computations. In some cases, heuristic state-machine approaches are also included because of complexities associated with the retransmission process. Examples of alternative or modified RTT estimation algorithms include *Peak-Hopper* [7], *Eifel* [6], and *Weighted Recursive Median* (WRM) [8].

The *Eifel* approach [6] notes a particular style of erroneous performance in the Jacobson algorithm, and adapts the algorithm in several ways to compensate for this performance oddity. As a result, Eifel eliminates unnecessary retransmissions which can result from spurious RTO violations. Similar to Eifel, the *Peak-Hopper* algorithm [7] also observes that the Jacobson algorithm responds inappropriately to certain fluctuations in RTT. This behavior produces "spikes" in RTO values because the algorithm does not distinguish between positive and negative variations. The modification proposed in [7] reduces this effect for a wide range of cases, and the findings in [9] concur. However, this solution results in higher average RTO values than the RFC approach [5], which can be a problem [10]. The *Weighted Recursive Median* (WRM) algorithm [8] redefines RTT estimation from a signal processing standpoint. WRM is effective, but tends to be computationally expensive even in a recursive form, which is a problem for per-packet operations.

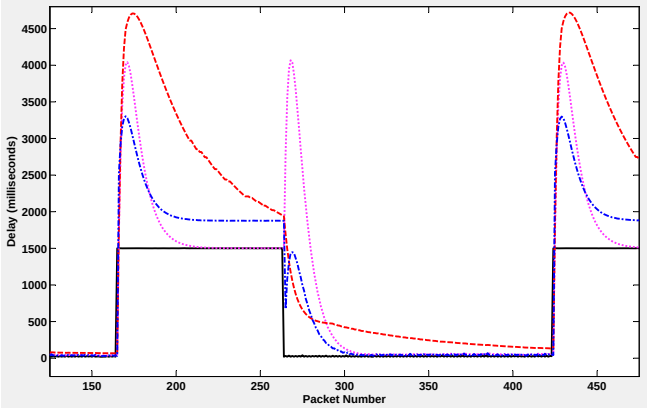


Figure 1. Relative performance of RTO estimators to large discontinuities in RTT. Estimators shown are the Jacobson Algorithm (dotted line), the Eifel Algorithm (dashed line), and the Modified Jacobson Algorithm (dash-dot line). The common RTT sequence driving all algorithms is shown as a solid line.

C. Other considerations

The remainder of this paper addresses the estimation process for the maximum RTT, or the value which establishes the RTO timer. When the RTO timer has expired, retransmission procedures commence, and may include various conditionally executed processes. We do not address those processes, or the overall retransmission procedure. In most cases, we use the Jacobson and Eifel estimation algorithms for comparison because they are widely accepted or implemented.

For reference, the performance of the Jacobson and Eifel algorithms are shown in Figure 1 along with the modified version of the Jacobson Algorithm which is discussed in detail in Section IV. The traces in the figure are all driven by a common RTT sequence created using our testbed of systems with modified networking stacks, as described in more detail in Section V. In the figure, the quiescent sections of the RTT sequence ($t < 170$ms, $260\text{ms} < t < 430$ms) have a fairly low mean, with similarly low standard deviation. This is typical of modern, high-speed networks. Also note that the RTT sequence has abrupt increases ($t \approx 170$ms, $t \approx 430$ms) followed by a stable period ($170\text{ms} < t < 260$ms), and then an abrupt decrease ($t \approx 260$ms). Note that Jacobson and Eifel both "overshoot" after the abrupt positive discontinuity in RTT. However, at the second discontinuity which is abrupt but negative, Eifel corrects downward, whereas Jacobson again corrects upward. This is a primary beneficial feature

 ISBN: 978-1-61208-262-2

of the Eifel approach. Unfortunately, the tuning of the algorithm creates a larger deviation or "upper bound" for the RTT sequence than is available through the Jacobson Algorithm.

Abrupt changes in RTT such as those shown in Figure 1 cause the RTO timer to expire, resulting in binary backoff of the timer (BEBO) and retransmission procedures. The RTO values associated with BEBO and retransmissions are not shown, because they do not affect future estimated values of RTO. Post-processing of the RTT estimate to enforce minimum values for RTO is also not considered here. In the RFCs for TCP and SCTP, a hard-coded minimum value for RTO is specified as 1000 milliseconds [1]–[3], [5], whereas the minimum value of the Eifel algorithm is defined as "RTT + 2 $\times$ ticks" (or 200 msec) [6]. The definition of the minimum RTO in TCP-Lite is typically 2 times the clock granularity, which is often taken as 500 msec [3], [6]. This minimum procedure is secondary to the estimation algorithm, and often completely replaces the RTT estimate. For example, in a broadband network with large minimum RTO, the RTT estimate can be orders of magnitude below the minimum, resulting in long wait-times for triggering retransmission. So, in all comparisons here we disable the enforcement of a minimum RTO and focus on the performance of the estimation algorithms.

III. The Jacobson Algorithm

The Jacobson Algorithm, originally proposed in [4], uses the Chebyshev Bound [16] to produce a reasonable value of RTO, or the maximum time the sender will wait for an acknowledgement. After exceeding RTO, a transmission is declared lost and retransmission procedures commence. Interestingly, Jacobson noted the poor performance of the algorithm, since loads higher than 30% resulted in retransmission of packets that were only delayed in transit (i.e., not lost) [4]. This behavior is also noted in later literature, including [6], [7], [15].

The specific computations for the bounding procedure are driven by estimates of the mean (μ) and the standard deviation (σ) of an assumed RTT distribution. The estimates of μ and σ are not conventional parameter estimates, as iid samples from a population. Rather, $\hat{\mu}$ and $\hat{\sigma}$ are *predicted* using prior RTT values. Jacobson estimates $\hat{\mu}$ and $\hat{\sigma}$ from observed values of RTT, and then computes an "overbound" for RTO using Chebyshev. This is the same as saying "we waited a reasonable time ($\hat{\mu}$) and *then* some ($K\hat{\sigma}$), but the ACK didn't come back, so the packet must have gotten lost". This RTO calculation is invoked by the sender for each unique transmission. As such, it is optimized for integer arithmetic and all coefficients are diadic.

Regardless of the specific assumptions or optimizations, Jacobson computes the RTO threshold as

$$x_{thr} = \hat{\mu}_{n+1|n} + K \cdot \hat{\sigma}_{n+1|n} \tag{1}$$

where x_{thr} is the RTO threshold or "overbound" for RTT, K is the number of "standard deviations past the mean," and $\hat{\mu}_{n+1|n}$ and $\hat{\sigma}_{n+1|n}$ are the estimates (predictions) of the mean and standard deviation of the RTT distribution for the next iteration (subscript $n+1$) given some data up to the current time (subscript n).

A. Jacobson and Chebyshev

The Chebyshev bound (2) is a universal bound applicable even for unknown distributions. Chebyshev shows that the probability of the random variable occurring outside a range around the mean (μ) depends on the standard deviation (σ).

$$\Pr\left[|X-\mu| \geq \epsilon\right] \leq \frac{\sigma^2}{\epsilon^2} \tag{2}$$

For RTT estimation, the Jacobson Algorithm uses a fixed offset from the mean $(K\sigma)$ [2], [3]. This simplifies the bound and allows a direct computation of the "violation probability". With $\epsilon = K\sigma$ and $K = 4$ in (2),

$$\Pr\left[|X-\mu| \geq K\sigma\right] \leq \frac{1}{K^2} \tag{3}$$

and the fixed, double-sided "violation probability" is $\frac{1}{16} = 0.0625$. However, the RTO timeout is single-sided because the timeout algorithm is only concerned with the case where $X > \mu$. So for a symmetric distribution and $K = 4$, the RTO timeout will be exceeded roughly 3% of the time, with the assumption that the RTT values are reasonably iid.

B. Predicting μ and σ

Rather than using conventional parameter estimation algorithms requiring storage of historical values and more complex calculation, Jacobson estimates μ and σ using simple *prediction algorithms* as in (4). These algorithms rely on current values of the quantities (subscript n) as well as the measured RTT value (x_n). Interestingly, the use of *filtered* (predicted) values for μ and σ implicitly contradicts the assumption of a valid distribution in (2). However, this approach gains computational efficiency and reflects the nonstationary nature of RTT values.

$$\begin{aligned} \hat{\mu}_{n+1|n} &= \hat{\mu}_n + \alpha(x_n - \hat{\mu}_n), \text{ and} \\ \hat{\sigma}_{n+1|n} &= \hat{\sigma}_n + \beta\left(|\hat{\mu}_n - x_n| - \hat{\sigma}_n\right). \end{aligned} \tag{4}$$

When viewed as a time-series prediction or filtering algorithm, it is clear that the relations in (4) are single-pole, lowpass IIR filters (predictors).

IV. The Modified Jacobson Algorithm

The use of Chebyshev to bound the retransmission timeout is reasonable, since it provides a "target probability" for the timeout calculation. However, the Markov bound [16] is also applicable for RTT estimation since it explicitly uses knowledge of the positivity of the random variable, as in (5) which is valid when $f_X(x) = 0$ for $x < 0$ and $\alpha > 0$.

$$\Pr\left[X \geq \alpha\right] \leq \frac{\mu}{\alpha}. \tag{5}$$

 ISBN: 978-1-61208-262-2

Using $\alpha = \mu + K\sigma$ in (5) produces an expression similar to (2). However, the Markov formulation results in a *variable* probability for RTO violation. This concept is particularly unappealing for small, relatively stable values of RTT, since the overbound RTO might have a high probability of violation, which would create spurious retransmissions and a large amount of unwanted network traffic. Recall that in the Chebyshev case, the choice of $\epsilon = K\sigma$ produced a fixed violation probability around 3%. The violation probability of the Markov bound can also be fixed as in (3) if $\alpha = 32\mu$:

$$\Pr[X \geq 32\mu] \leq \frac{\mu}{32\mu} = \frac{1}{32} = \frac{1}{2K^2}. \quad (6)$$

Unfortunately, a bias of 32 times the mean would not produce a viable estimate of RTT, and the overbound RTO would be extremely loose. Therefore, the Markov bound *alone* is not be a reasonable choice to estimate the RTO timeout. However, a *combination* of Markov and Chebyshev approaches seems to produce an effective estimator.

Combining an estimate of σ as in the Jacobson Algorithm with a *biased* estimate of μ as in the Markov bound results in a formulation that retains the Chebyshev structure but improves certain performance aspects. So, we use a slightly revised version of (1) and call this approach the *Modified Jacobson Algorithm*,

$$x_{thr} = \mathcal{A} \cdot \hat{\mu}_{n+1|n} + \mathcal{B} \cdot \hat{\sigma}_{n+1|n}. \quad (7)$$

In (7), the standard deviation estimator $\hat{\sigma}_{n+1|n}$ is identical to the estimator (4) used in the Jacobson Algorithm. However, the multiplier for $\hat{\sigma}_{n+1|n}$ is *reduced* (i.e. $\mathcal{B} = 2$ whereas $K = 4$). Further, the mean estimator $\hat{\mu}_{n+1|n}$ is replaced with the current value of RTT (x_n), and is *biased* in the spirit of a Markov estimator. In this case, we choose $\mathcal{A} = 1.25$ as a reasonable bias term, whereas Jacobson uses $\mathcal{A} = 1.0$.

Note that the bias term $\mathcal{A}$ for $\hat{\mu}_{n+1|n}$ in (7) is *not equivalent* to the use of gain in the prediction of $\hat{\mu}_{n+1|n}$ in (4). The structure of the prediction filter for $\hat{\mu}_{n+1|n}$ causes delay in the formulation of the overbound, which is problematic. There are no coefficients for the prediction filter which will simultaneously improve delay and maintain stability in the estimation of μ, and incorporating gain in the prediction does not improve the estimate. These undesirable effects are completely eliminated in the Modified approach with the use of x_n as the estimator of μ. This adjustment allows for the use of a Markov-like bias term $\mathcal{A}$ and significantly enhances the performance of the Modified algorithm.

Several factors must be specifically noted for the Modified Algorithm. First, dependence on the variance of the RTT sequence is preserved via $\hat{\sigma}_{n+1|n}$ and the Chebyshev-like formulation. Some dependence on σ must be maintained in the estimation procedure for cases where the RTT values exhibit significant variability. However, the multiplier $\mathcal{B}$ can be different (smaller) than in Jacobson. This reduced dependence on σ mediates undesirable "overshoot" which is problematic in Jacobson, and has been addressed heuristically in Eifel. Refer to abrupt changes in RTT as shown in Figure 1 for examples.

Secondly, dependence on the mean of the RTT sequence is preserved via the use of x_n for $\hat{\mu}_{n+1|n}$, and a bias is incorporated via the Markov-like formulation for cases where $\sigma \to 0$. Some dependence on μ is important, since this allows isolation of the variability. However, in cases where $\sigma \to 0$, Jacobson tends to "settle" directly onto RTT, leading to heuristic modifications including static minimum values which override the Jacobson estimates. This undesirable behavior of Jacobson is clearly evident in Figure 1.

Thirdly, explicit dependence on both μ and σ is retained via the hybrid Markov/Chebyshev formulation which *biases* the estimate higher and reduces the need for secondary minimum computations. Also, the prediction structure for $\hat{\mu}$ and $\hat{\sigma}$ is preserved, which is an important consideration.

Finally, the computational complexity of the Modified Algorithm is essentially the same as the original Jacobson Algorithm. Elimination of the prediction structure for $\hat{\mu}$ and the use of a bias term along with a simplified multiplier for $\hat{\sigma}$ results in an algorithm with the same operational complexity and no heuristic conditional logic steps.

V. Performance results

To validate algorithm performance, we constructed a "real-world" test environment which pairs client and server computers with modified network stacks via a controllable network infrastructure. In the network testbed, a client system transmits data to a server system using a specially-constructed user-space application which can vary overall payload length and SCTP chunk size, as well as other parameters such as the number of test iterations. The network stack of the client systems also implement user-selectable timeout estimation algorithms and record important parameters for each transmission. Parameters recorded by the client's network stack include the timestamped, per-chunk values of the actual (measured) RTT, estimated RTO, $\hat{\mu}$, $\hat{\sigma}$, and so on. As described in Section II-C, post-processing of the RTO estimate to enforce minimum values for RTO is disabled since we are investigating the effect of estimation algorithms.

Additionally, the network devices and SCTP server are modified to introduce algorithmically controllable delays in acknowledgements of chunks and delays in delivery of various classes of network traffic. This feature results in an ability to introduce specific "delay profiles" which duplicate other known results (as in Figure 2 [6]) or randomize the round-trip time of the network. Using trace data collected directly from the network stacks of the client & server computers, we were also able to create simulations of system performance which have been cross-checked for accuracy against the delay and estimation performance of the actual systems. All performance data described in this paper was produced using our "real-world" testbed, and has

 ISBN: 978-1-61208-262-2

Table I
ESTIMATION PERFORMANCE GATHERED USING NETWORK TESTBED AND SIMULATION. PERFORMANCE IS MEASURED AS MEAN ABSOLUTE ERROR (MAE) IN MILLISECONDS.

Delay profile	Figure	Jacobson	Eifel	Modified
Eifel ramp	Figure 2	1731	2091	1577
Quiet/Spikes	Figure 3	39.81	140.7	43.29
Delay burst	Figure 4	11.31	41.79	2.00

Table II
ESTIMATION PERFORMANCE GATHERED USING NETWORK TESTBED AND SIMULATION. PERFORMANCE IS MEASURED IN TERMS OF TIMEOUT EVENTS (TO) PER 10,000 PACKETS TRANSMITTED.

Delay profile	Figure	Jacobson	Eifel	Modified
Eifel ramp	Figure 2	1	1	1
Quiet/Spikes	Figure 3	99	51	79
Delay burst	Figure 4	378	82	12

been incorporated into accurate simulations of the estimation algorithms.

The performance of the Modified Jacobson Algorithm is shown in the context of various RTT sequence characteristics or "delay profiles" in Figure 2, Figure 3, and Figure 4. Figure 2 reproduces an important delay profile from literature describing the Eifel algorithm (Fig. 6 of [6]). Figure 3 contains the delay profile for a real, quiescent network with 100 msec average delay and short, artificially induced delay spikes. Figure 4 contains the delay profile for a long-term delay burst on an otherwise quiescent network with 80 msec average delay.

Quantitative assessment of algorithm performance gathered from a large number of packet transmissions is shown in Table I and Table II for each of the delay profiles described by Figure 2, Figure 3, and Figure 4. The data in Table I is presented in terms of Mean Absolute Error (MAE) in milliseconds between the RTO estimate and the actual RTT value for the same packet transmission, according to (8). Table II shows the number of induced timeout events (TO) for each estimation algorithm, and is measured in number of events per 10,000 packets transmitted.

$$\text{MAE} = \frac{1}{N} \sum_{N} |x_{thr} - x_n| \qquad (8)$$

Using our network testbed, we duplicated the RTT sequence in Figure 2 from [6]. This "delay profile" clearly shows the improvement of Eifel over Jacobson, particularly at the termination of the "ramp" sequences. Note that Jacobson overshoots *upward* even though the RTT sequence has rapidly declining values. Eifel compensates for overshoot, but at the expense of slightly higher bias from the RTT sequence. Note that the Modified approach mimics Jacobson during ramp ascension, and also compensates for the overshoot at ramp termination. Modified has a smaller

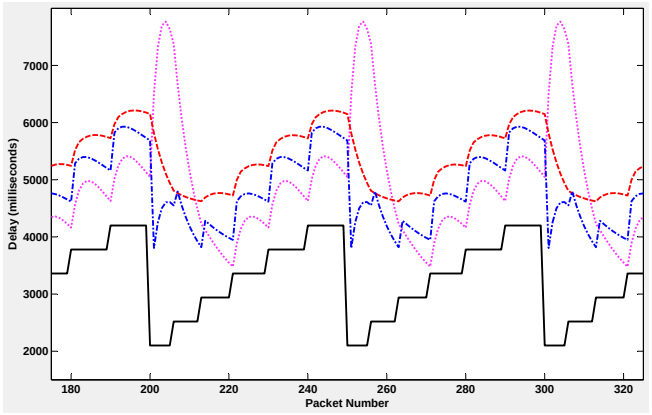


Figure 2. Performance of RTO estimators for the "ramp" RTT sequence described in [6]. Jacobson (dotted), Eifel (dashed), and Modified (dash-dot). The common RTT sequence driving all algorithms (solid).

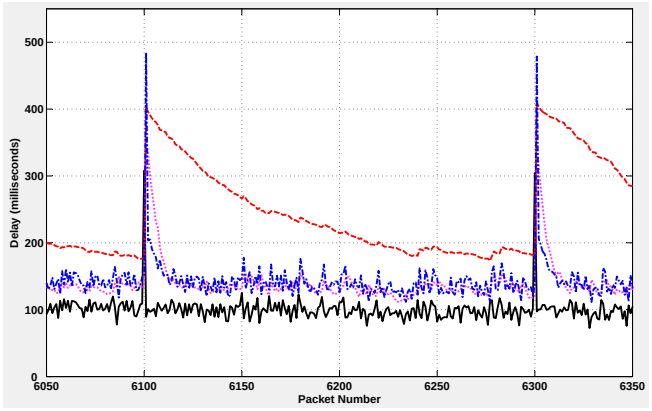


Figure 3. Performance of RTO estimators for a quiescent network with short, artificially induced delay spikes. Jacobson (dotted), Eifel (dashed), and Modified (dash-dot). The common RTT sequence driving all estimation algorithms (solid).

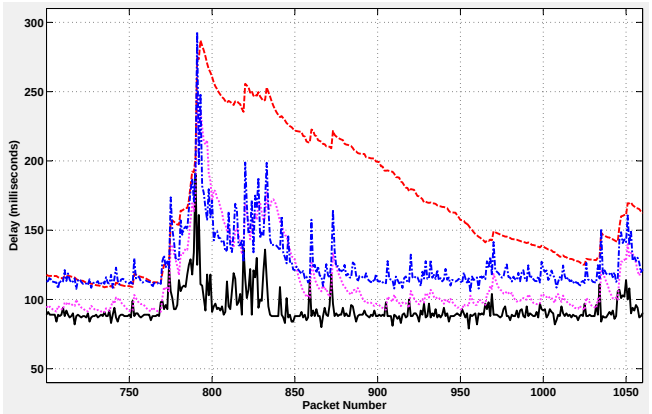


Figure 4. Performance of RTO estimators for a quiescent network with a large, naturally-occurring multi-packet delay burst. Jacobson (dotted), Eifel (dashed), and Modified (dash-dot). The common RTT sequence driving all algorithms (solid).

bias than Eifel. Interestingly, each algorithm induces a single timeout event, but the Modified Jacobson algorithm does so with an MAE 154 msec smaller than Jacobson, and 514 msec smaller than Eifel. Such a significant difference in MAE can translate to a smaller task-completion time in cases where a small differential in timeouts is encountered.

A delay profile from a relatively quiescent network with artificially induced delay spikes is shown in Figure 3. The network exhibits an average delay of around 100msec, and has artificially induced delay spikes which are typical of an unstable link. Note that with this network profile, delay spikes occur every 200 packets, forcing a timeout. After each delay spike, the estimation algorithms recover in very different manners: Modified and Jacobson fall quickly toward the quiescent RTT sequence, while Eifel decays very slowly, creating a relatively large wait-time for over 100 subsequent packets. As a result, the MAE for the Eifel algorithm with this delay profile is more than 3 times larger than the other estimation algorithms while still creating only 35% fewer timeout events. Regardless, each estimation algorithm creates a very small proportion of timeout events relative to the number of packets transmitted.

A delay profile from another network test is shown in Figure 4. In this figure, the network exhibits an average delay during quiescent periods of around 80 msec. However, between packets 750 and 850 a large, naturally occurring, correlated delay burst is observed which disrupts the estimation algorithms. Note that Modified and Jacobson both fall quickly after individual, large delay spikes. However, Modified maintains a larger offset during the quiescent period between packets 900 and 1000 due to the specific bias for $\hat{\mu}$. The failure of Jacobson to maintain a bias during periods of low RTT variance ($\sigma \rightarrow 0$) is responsible for many timeout events, with Jacobson inducing *30 times more timeouts* than Modified, and *almost 5 times more* than Eifel. Eifel again exhibits a bias which is significantly larger than Jacobson or Modified, with an MAE *20 times larger* than Modified, and *4 times larger* than Jacobson. Also note that Jacobson tracks RTT fairly well, but has large, positive overshoot when RTT drops suddenly. Modified compensates for the Jacobson "overshoot" problem in cases where the RTT sequence drops suddenly (cf. packet 800 & 840).

VI. Conclusion and Future Work

This paper analyzes the methods for computing and using RTT and RTO estimates in IP-based transport protocols such as SCTP and TCP. The theoretical basis of the Jacobson Algorithm is discussed, and an alternative approach is presented which retains the fundamentally sound theoretical basis and operational structure of the algorithm, but improves the performance over other well-known techniques without introducing heuristic modifications. Future work involves the continued optimization of the modified algorithm as well as investigation into the effects of variable minimum bound for the RTO timer.

References

[1] J. Postel, "Transmission control protocol," RFC793, Sep. 1981.

[2] R. Stewart *et al.*, "Stream control transmission protocol," RFC 2960, Oct. 2000.

[3] R. Stewart, "Stream control transmission protocol," RFC4960, Sep. 2007.

[4] V. Jacobson, "Congestion avoidance and control," in *Proc. Comput. Commun. Review*, ser. SIGCOMM'88. New York, NY: ACM, Aug. 1988, pp. 314–329.

[5] V. Paxson, M. Allman, J. Chu, and M. Sargent, "Computing TCP's retransmission timer," RFC6298, June 2011.

[6] R. Ludwig and K. Sklower, "The Eifel retransmission timer," *Comput. Commun. Review (SIGCOMM)*, vol. 30, pp. 17–27, July 2000.

[7] H. Ekstrom and R. Ludwig, "The Peak-Hopper: A new end-to-end retransmission timer for reliable unicast transport," in *Proc. 23rd Joint Conf. of the IEEE Computer & Communications Societies (INFOCOM 2004)*, vol. 4, Nov. 2004, pp. 2502–2513.

[8] L. Ma, G. Arce, and K. Barner, "TCP retransmission timeout algorithm using weighted medians," *IEEE Sig. Proc. Letters*, vol. 11, pp. 569–572, June 2004.

[9] J. Pedersen, C. Griwodz, and P. Halvorsen, "Considerations of SCTP retransmission delays for thin streams," in *Proc. 31st IEEE Conf. on Local Computer Networks*, Tampa, FL, Nov. 2006, pp. 135–142.

[10] A. Petlund, P. Beskow, J. Pedersen, E. S. Paaby, C. Griwodz, and P. Halvorsen, "Improving SCTP retransmission delays for time-dependent thin streams," *Multimedia Tools and Applications*, vol. 45, pp. 33–60, Oct. 2009.

[11] M. Allman, K. Avrachenkov, U. Ayesta, J. Blanton, and P. Hurtig, "Early retransmit for TCP and stream control transmission protocol (SCTP)," RFC5827, Apr. 2010.

[12] M. Allman and V. Paxson, "On estimating end-to-end network path properties," in *Proc. Conf. Appl., technol., arch., and protocols for comput. commun.*, ser. SIGCOMM '99. New York, NY, USA: ACM, 1999, pp. 263–274.

[13] L. Coene and J. Pastor-Balbas, "Telephony signaling transport over stream control transmission protocol (SCTP) applicability statement," RFC4166, Feb. 2006.

[14] I. Psaras and V. Tsaoussidis, "Why TCP timers (still) don't work well," *Computer Networks*, vol. 51, pp. 2033–2048, Nov. 2007.

[15] J. Pedersen, "Evaluation of SCTP retransmission delays," Master's thesis, University of Oslo Department of Informatics, May 2006.

[16] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York, NY: McGraw-Hill, 1991.

 ISBN: 978-1-61208-262-2

Performance Analysis of PCFICH and PDCCH LTE Control Channel

Jiri Milos, Stanislav Hanus
Dept. of Radio Electronics
Brno University of Technology
Brno, Czech Republic
Email: milos.jiri@phd.feec.vutbr.cz,
hanus@feec.vutbr.cz

***Abstract*—Control channels play a key role in the evaluation of mobile system performance. Authors have evaluated the performance of control channels implementation in a Long Term Evolution (LTE) system. The paper deals with simulating a complete signal processing chain for the Physical Control Format Indicator Channel (PCFICH) and Physical Downlink Control Channel (PDCCH) in an LTE system, Release 8. We implemented a complete signal processing chain for downlink control channels as an extension of the existing MATLAB LTE downlink simulator. The paper shows results of the PCFICH and PDCCH control channel computer performance analysis in various channel conditions. The presented results can be compared with the performance of data channels.**

Keywords-Long Term Evolution; Physical Control Format Indicator Channel; Physical Downlink Control Channel; MATLAB; link level simulator;

I. INTRODUCTION

Link level simulators are typically focused in the performance analysis of traffic channels [1], [2]. For an overall system performance and comparison with real network deployment it is necessary to include control channels in the simulations. Control channels are typically designed with more robust forward error correction and modulation than traffic channels. In the case of the 3GPP LTE system, the Orthogonal Frequency Division Multiple Access (OFDMA) access method and robust scrambling of control channel information is used [3], [4].

Our motivation is to implement control channels as an extension of the existing MATLAB LTE downlink simulator developed at Vienna University of Technology. After completion of the model of control channels, we can look into the influence of the propagation environment to overall throughput of the LTE system, using computer simulation.

The paper is organized as follows. First the LTE system is described briefly. The next part brings a short description of the signal processing chain of surveyed control channels. Simulation and their results are mentioned in the third part and it is summarized in the conclusion.

II. LTE PHYSICAL CHANNELS

The LTE physical layer in a downlink includes a triplet of control channels [5], [6]. The list of LTE control channels in downlink and their control information is shown in Table I.

TABLE I
LIST OF LTE CONTROL CHANNELS - DOWNLINK

Control Information	Corresponding Physical Control Channel
CFI	Physical Control Format Indicator Chan. (PCFICH)
HI	Physical Hybrid-ARQ Indicator Channel (PHICH)
DCI	Physical Downlink Control Channel (PDCCH)

The above-mentioned control physical channels are not used to transfer information from higher layers [7]. They are not associated with transport channels. This article deals with signal processing and Bit error rate (BER) in the Physical Control Format Indicator Channel and Physical Downlink Control Channel, performance analysis of the Physical Hybrid-ARQ Control Channel is not included.

A. Physical Control Format Indicator Channel (PCFICH)

Via this channel, the Control Format Indicator (CFI) is transmitted. The value stored in CFI determines the number of resource elements in the resource grid (in time domain) carrying the data of the PDCCH control channel. It determines a PDCCH control area in each subframe in downlink. PCFICH is transmitted in the first OFDM symbol in the resource grid [3].

The CFI parameter takes values 1, 2 or 3 only. The block diagram of PCFICH signal processing is shown in Fig. 1. The first operation in the signal processing chain is channel coding. A bit sequence of 32 bits in length is assigned to each value in CFI, according to Table II.

A codeword of 32 bits in length is scrambled with a pseudo-random scrambling sequence, which is unique for

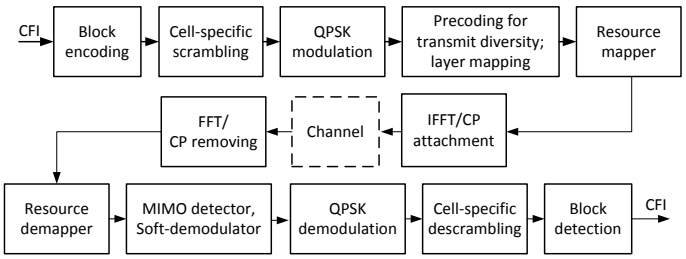


Figure 1. PCFICH processing chain.

 ISBN: 978-1-61208-262-2

TABLE II
PCFICH CHANNEL CODING

CFI	CFI codeword [b_0, b_1, …, b_{31}]
1	[01101101101101101101101101101101]
2	[10110110110110110110110110110110]
3	[11011011011011011011011011011011]
4 (reserved)	[00000000000000000000000000000000]

each cell. The scrambled bits are modulated by QPSK modulation and this block of complex-value symbols is mapped into ν-layers, depend on the number of transmitting antennas, where $\nu = \{1, 2, 4\}$. In the case of one transmitting antenna, layer mapping is not used. In the case of one transmitting antenna, transmit diversity precoding is not provided [8]. In the case of two or four transmitting antennas, it is necessary to provide precoding for transmit diversity.

Complex symbols for each transmitting antenna are grouped to quaternary symbols, so-called symbol quadruplets, which are mapped to defined positions in the resource grid [9]. Furthermore, a IFFT operation with symbols in the resource grid is performed and cyclic prefix (CP) is inserted.

After passing through the channel, CP is removed and FFT is performed. In a resource demapping block, PCFICH symbols are collected. When using more than one transmitting antenna, MIMO detection is provided. Complex-value symbols are demodulated and descrambled by pseudo-random sequention which is the same as in the transmitter. The last part of the signal processing chain is a CFI codeword detection block.

B. Physical Downlink Control Channel (PDCCH)

Physical Downlink Control Channel is the most important control channel in a downlink. It supports signalling for data channels in downlink and uplink. Via this channel Downlink Control Information (DCI) is transmitted.

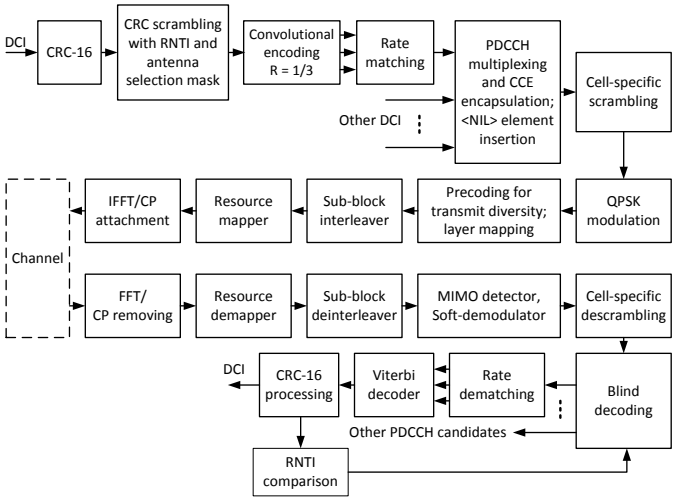


Figure 2. PDCCH processing chain.

PDCCH supports various formats of DCI messages [10]. They contain information about resource scheduling for downlink and uplink, Transmit Power Commands (TPC), etc.

The block diagram of the PDCCH signal processing chain is shown in Fig. 2. Individual DCI messages of different formats are channel encoded. Cyclic redundancy check (CRC-16) of 16 bits in length is added to the DCI message [4]. Afterwards, the CRC is scrambled with a Radio Network Temporary Identifier (RNTI) value and antenna's mask if needed. The next block in the processing chain is a convolutional encoder with coding rate R=1/3. In the rate matching block, interleaving is provided and decreases the length of the encoded DCI message.

These operations with all DCI messages are provided in parallel and these messages come into a PDCCH multiplexing block. The encoded DCI messages are encapsulated into so-called Control Channel Elements (CCE). These elements are mapped into individual PDCCH formats. The PDCCH frame is scrambled in the same way as in the case of PCFICH.

Scrambled bits are modulated by QPSK modulation and then form a block of complex-value symbols which are mapped into ν-layers and precoding for transmit diversity is performed. These symbols, split into ν-parallel streams and mapped to the quadruplets of symbols are interleaved using Free-quadratic Permutation Polynomial technology (QPP). When the interleaving is performed, quadruplets are mapped into defined positions in the resource grid. A distribution of PDCCH quadruplets is given by a value in CFI.

Further, complex-valued symbols are transferred into the time domain by using an IFFT and CP is added. After passing through the channel, on the receiving side, CP is removed and FFT is performed. In the resource demapper block, a matrix of PDCCH symbols and a matrix of corresponding channel coefficients $\mathbf{H}^{-1}_{\text{PDCCH}}$ are collected. Reverse interleaving is performed with both of these matrices. Modified symbols lead into the MIMO detector and soft-demodulator. A vector of demodulated bits is descrambled by the same cell-specific pseudo-random sequence as in the transmitting side and blind decoding is performed.

III. SIMULATIONS

Performance analysis of the LTE link level simulator is usually presented only for traffic channels [1]. The fundamental performance analysis of control channels is mentioned in [11] for PCFICH and PHICH. Similar problem examines [12], which presents PDCCH performance analysis results in AWGN channel model.

As was mentioned, the Link level simulator was used as a basic physical layer model of LTE. The simulator was developed at the Vienna University of Technology [1]. The model was extended by adding control channels. A performance analysis was done; the physical layer BER

 ISBN: 978-1-61208-262-2

TABLE III
SIMULATION PARAMETERS

Parameters	Description
Frame structure	FDD
System bandwidth	1.4 MHz
Cyclic prefix	Normal and extended
Antenna configuration	1×1, 2×1, 2×2, 4×1, 4×2
Subcarrier spacing	15 kHz
Channel estimation method	Perfect
Receiving algorithm	Soft-sphere decoder (SSD)

performance of PCFICH and PDCCH and their results are presented in this section. The main parameters of the simulation system are listed in Table III.

The simulator works with an FDD frame structure only. Due to a system bandwidth of 1.4 MHz, there were 6 resource element blocks (RB). In the case of using the CP with normal duration, there are 7 OFDM symbols in one slot (the first OFDM symbol in the slot has greater duration than the others [3]). In the case of using the CP with extended duration there are 6 OFDM symbols in one slot. The Soft-sphere receiving algorithm was used. Used fading channel models and their specification are listed in Table IV and Table V. All used fading channel models are included in the basic LTE simulator.

TABLE IV
PARAMETERS OF PEDESTRIAN B AND VEHICULAR A CHANNEL MODEL

Pedestrian B		Vehicular A	
Av. Power [dB]	Delay [ns]	Av. Power [dB]	Delay [ns]
0.0	0	0.0	0
−0.9	200	−1.0	310
−4.9	800	−9.0	710
−8.0	1200	−10.0	1090
−7.8	2300	−15.0	1730
−23.9	3700	−20.0	2510

TABLE V
PARAMETERS OF TYPICAL URBAN AND RURAL AREA CHANNEL MODEL

Typical Urban				Rural Area	
Av. Power [dB]	Delay [ns]	Av. Power [dB]	Delay [ns]	Av. Power [dB]	Delay [ns]
−5.7	0	−17.4	1349	−5.2	0
−7.6	217	−19.0	1533	−6.4	42
−10.1	512	−19.0	1535	−8.4	101
−10.2	514	−19.8	1622	−9.3	129
−10.2	517	−21.5	1818	−10.0	149
−11.5	674	−21.6	1836	−13.1	245
−13.4	882	−22.1	1884	−15.3	312
−16.3	1230	−22.6	1943	−18.5	410
−16.9	1287	−23.5	2048	−20.4	469
−17.1	1311	−24.3	2140	−22.4	528

TABLE VI
COMPARISON OF THE SIMULATION RESULTS OF MINIMAL SNR FOR BER EQUAL TO 10^{-4} IN PCFICH CHANNEL - NORMAL CP

Antenna configuration	Minimal SNR value in dB for different channel model type				
$N_{TX} \times N_{RX}$	AWGN	Ped B	Veh A	TU	RA
1×1	0.8	9.6	11.4	8.5	16.2
2×1	1.0	5.8	5.7	4.8	8.5
2×2	−2.2	0.0	0.0	−0.2	1.5
4×1	0.8	3.8	3.5	3.0	4.7
4×2	−2.2	−1.0	−0.4	−1.0	−0.3

TABLE VII
COMPARISON OF THE SIMULATION RESULTS OF MINIMAL SNR FOR BER EQUAL TO 10^{-4} IN PCFICH CHANNEL - EXTENDED CP

Antenna configuration	Minimal SNR value in dB for different channel model type				
$N_{TX} \times N_{RX}$	AWGN	Ped B	Veh A	TU	RA
1×1	0.7	10.5	10.5	8.6	13.5
2×1	0.8	5.5	6.6	4.6	8.6
2×2	−2.3	−0.2	1.0	0.5	1.7
4×1	0.7	3.8	3.7	3.6	4.6
4×2	−2.2	−1.0	−0.7	−0.5	−0.2

A. Performance Analysis of the PCFICH

The results of the PCFICH control channel transmission simulation and subsequently, analysis of BER depending on the Signal to Noise Ratio (SNR) for all antenna configuration and models of used transmission channels are listed in this section. The BER is calculated from the two- bit CFI values at the beginning and at the end of the transmission chain. The PCFICH BER curves sorted according to the available antenna configurations and both types of CP are shown in Fig. 3 to 7.

The summarized results of PCFICH BER in configuration with normal CP are shown in Table VI. The summarized results of PCFICH BER in configuration with extended CP are shown in Table VII. Here we see values of SNR at which the BER in PCFICH channel reaches the reference level of 10^{-4}. In the presented figures, observe the BER curves for configuration with normal CP length and extended CP length are without marked differences. The

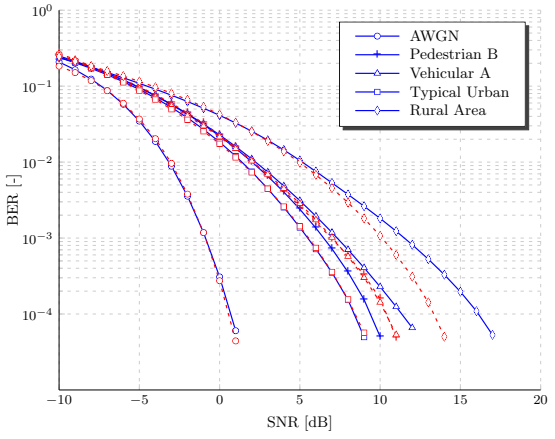


Figure 3. BER of the PCFICH for 1 transmitting and 1 receiving antenna (CP: normal - blue solid line, extended - red dotted line).

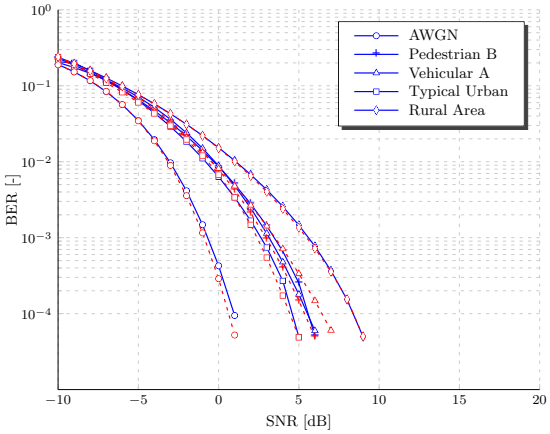


Figure 4. BER of the PCFICH for 2 transmitting and 1 receiving antenna (CP: normal - blue solid line, extended - red dotted line).

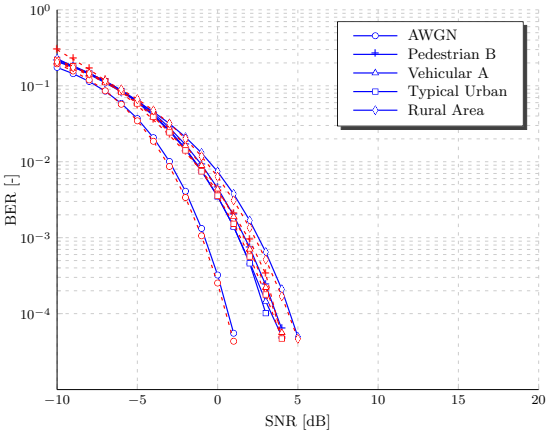


Figure 6. BER of the PCFICH for 4 transmitting and 1 receiving antenna (CP: normal - blue solid line, extended - red dotted line).

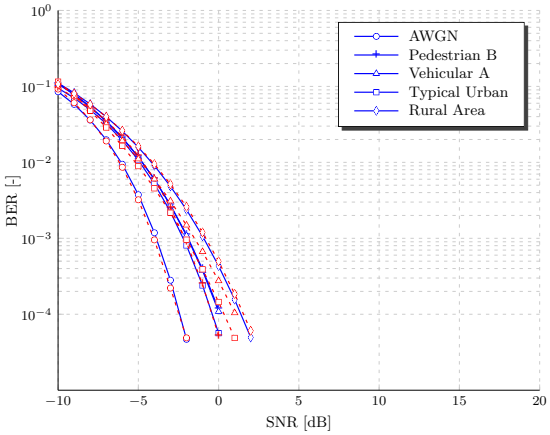


Figure 5. BER of the PCFICH for 2 transmitting and 2 receiving antennas (CP: normal - blue solid line, extended - red dotted line).

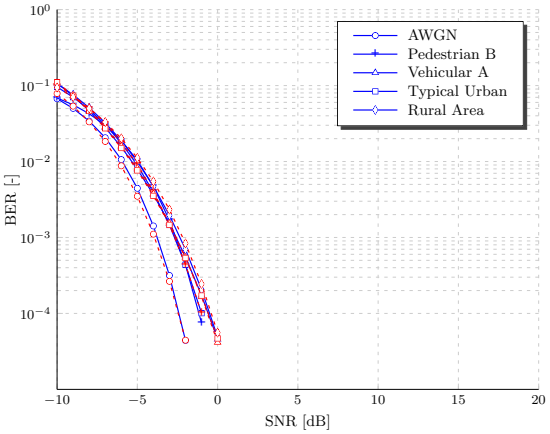


Figure 7. BER of the PCFICH for 4 transmitting and 2 receiving antennas (CP: normal - blue solid line, extended - red dotted line).

figure configured with 1 transmitting and 1 receiving antenna shows the greatest difference in the case of the rural area channel model. In this case, for extended CP, the coding gain is 2.7 dB.

B. Performance Analysis of the PDCCH

The results of the PDCCH control channel transmission simulation and subsequently, analysis of BER depending on the SNR for all antenna configurations and models of used transmission channels are listed in this section. The BER is calculated from the DCI value in bits (format 0 - set as the test case) at the beginning and at the end of the transmission chain. The PDCCH BER curves sorted according to the available antenna configurations are shown in Figs. 8 to 12.

The summarized results of PDCCH BER configured with normal CP are shown in Table VIII. The summarized results of PDCCH BER configured with extended CP are shown in Table IX. Here we see values of SNR at which BER in the PDCCH channel reaches the reference level of 10^{-4}. In Fig. 8, we can observe that simulations fail in all models of fading channels and both types of CP. The BER curves

TABLE VIII
COMPARISON OF THE SIMULATION RESULTS OF MINIMAL SNR FOR BER EQUAL TO 10^{-4} IN PDCCH CHANNEL - NORMAL CP

Antenna configuration $N_{TX} \times N_{RX}$	Minimal SNR value in dB for different channel model type				
	AWGN	Ped B	Veh A	TU	RA
1×1	5.9	–	–	–	–
2×1	6.0	13.6	13.4	13.0	16.2
2×2	2.9	6.2	6.8	5.9	8.5
4×1	5.9	15.2	14.8	16.0	15.1
4×2	2.7	7.0	7.0	7.9	6.8

TABLE IX
COMPARISON OF THE SIMULATION RESULTS OF MINIMAL SNR FOR BER EQUAL TO 10^{-4} IN PDCCH CHANNEL - EXTENDED CP

Antenna configuration $N_{TX} \times N_{RX}$	Minimal SNR value in dB for different channel model type				
	AWGN	Ped B	Veh A	TU	RA
1×1	5.8	–	–	–	–
2×1	5.8	12.6	13.8	13.8	15.8
2×2	2.9	6.7	6.6	6.7	7.8
4×1	5.7	–	13.3	–	13.9
4×2	2.9	7.8	6.7	7.8	6.4

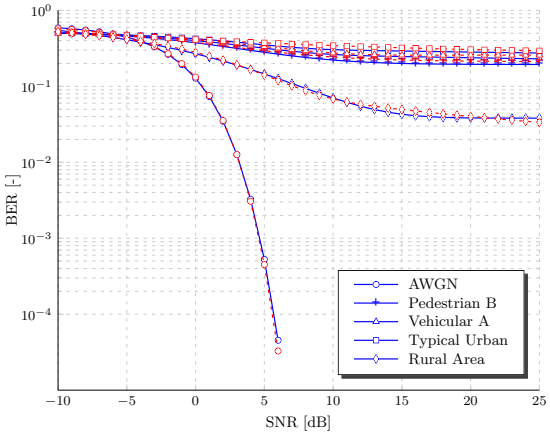


Figure 8. BER of the PDCCH for 1 transmitting and 1 receiving antenna (CP: normal - blue solid line, extended - red dotted line).

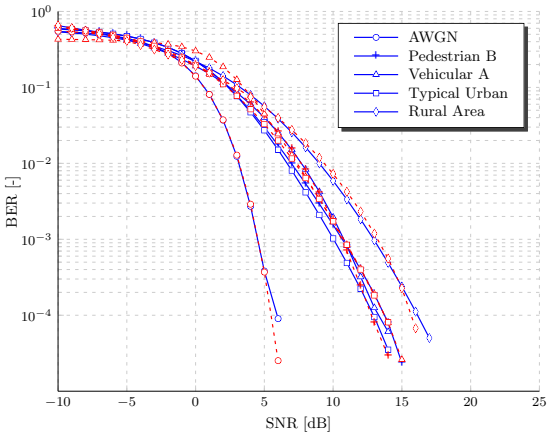


Figure 9. BER of the PDCCH for 2 transmitting and 1 receiving antenna (CP: normal - blue solid line, extended - red dotted line).

reach the reference value of BER in the case of the AWGN channel model only. In this case the convolutional encoding aborts. In Fig. 11, extended CP, BER for Pedestrian B and TU channel models reach BER= 10^{-3} only. This abnormality is caused by ISI.

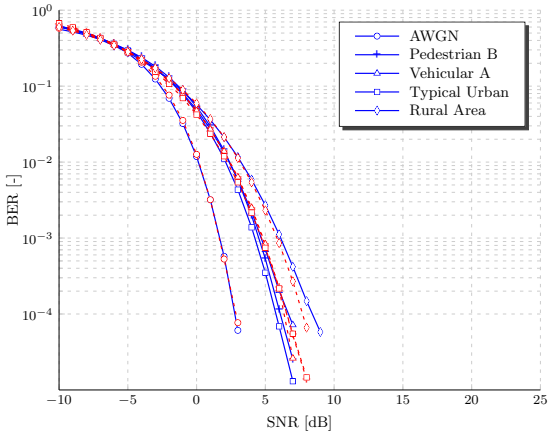


Figure 10. BER of the PDCCH for 2 transmitting and 2 receiving antennas (CP: normal - blue solid line, extended - red dotted line).

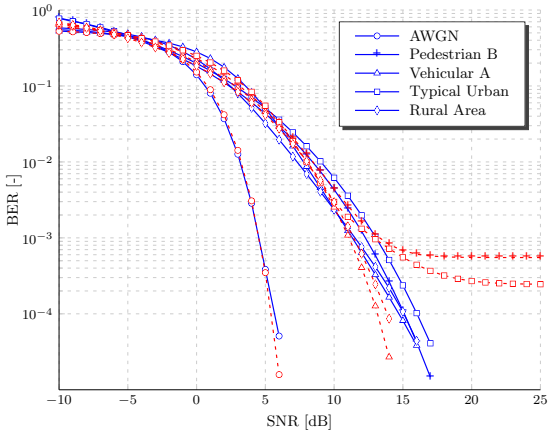


Figure 11. BER of the PDCCH for 4 transmitting and 1 receiving antenna (CP: normal - blue solid line, extended - red dotted line).

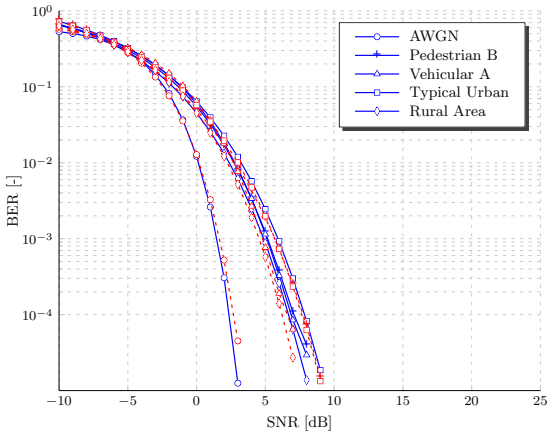


Figure 12. BER of the PDCCH for 4 transmitting and 2 receiving antennas (CP: normal - blue solid line, extended - red dotted line).

 ISBN: 978-1-61208-262-2

IV. Conclusion

In this paper, the PCFICH and PDCCH LTE downlink control channels were analyzed and simulated. The paper is focused on studying two selected control channels due to their functions being linked together. The performance of these two control channels was analyzed in AWGN channel model and in basic types of fading channels. As we can see from the presented graphs, the PCFICH BER is markedly lower than the PDCCH BER. It is given by the size of transmitted data and used channel coding. While in the case of the PCFICH channel, a two-bit CFI codeword is carried and this codeword is spread with SF $= 16$, whereas in the case of the PDCCH, many more bits are transmitted. In particularly bad cases of some fade channels, the used convolutional encoder fails, despite the use of other error protection techniques.

The BER curves for normal or extended CP length are without marked differences. In the case of PCFICH analysis, usage of transmitting diversity brings expected results. In the case of PDCCH analysis, results from the antenna configuration with one transmitting and one receiving antenna (SISO mode) are absolutely unusable in all types of fading channels for normal and extended CP length. However, these are only results from computer analysis. In a real LTE system, the SISO mode is very usable. In the case of four transmitting and one receiving antenna for extended CP length, the BER curves for Pedestrian B and Typical Urban channel model reach a BER value of 10^{-3} only. It is caused by ISI in this computer simulation.

As a further phase of the downlink control channel analysis, simulations impacting BER in control channels on overall throughput of the LTE system will be performed.

Acknowledgment

The authors would like to thank the support of the project CZ.1.07/2.3.00/20.0007 WICOMT, financed from the operational program Education for competitiveness, SIX project no. CZ.1.05/2.1.00/03.0072 and SYWIC project LD11081 in frame of COST IC 0906, are gratefully acknowledged.

References

[1] LTE downlink link level simulator, [Online]. Available: http://www.nt.tuwien.ac.at/ltesimulator/ [retrieved: Nov. 2012].

[2] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long term evolution physical layer," in Proc. 17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, Scotland, August 2009, pp. 1471–1478, [Online] Available: http://publik.tuwien.ac.at/files/PubDat_175708.pdf [retrieved: Jan. 2013].

[3] Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation, 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 36.211 Version 8.9.0, Dec. 2009.

[4] Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation, 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 36.212 Version 8.8.0, Dec. 2009.

[5] H. Holma and A. Toskala, LTE for UMTS: OFDMA and SC-FDMA based radio access. 1st. ed. New York: John Wiley and Sons, 2009. 433 p. ISBN 0470994010.

[6] M. Pesavento and W. Mulder, "LTE tutorial, part 1: LTE basics," in Proc. Femto Forum Plenary, Reading, United Kingdom. June 2010, pp. 1–46.

[7] R. Love, R. Kuchibhotla, and A. Ghosh, "Downlink control channel design for 3GPP LTE," in Proc. Wireless Communications and Networking Conference, WCNC 2008. Apr. 2008, pp. 814–818, ISBN 978-1-4244-1997-5.

[8] K. Il Gyu, H. Youngnam, K. Young Hoon, and B. Seung Chan, "Transmit diversity and multiplexing methods for 3G-LTE downlink control channels," in Proc. 64th Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE, Sept. 2006, pp. 1–4, ISBN 1-4244-0062-7.

[9] P. Hosein, "Resource allocation for the LTE physical downlink control channel," in Proc. GLOBECOM Workshops, 2009 IEEE. Dec. 2009, pp. 1–5, ISBN 978-1-4244-5626-0.

[10] Evolved universal terrestrial radio access (E-UTRA); physical layer procedures, 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 36.213 Version 8.8.0, Sept. 2009.

[11] S. J. Thiruvengadam and L. M. A. Jalloul, "Performance analysis of the 3GPP-LTE physical control channels," in Proc. EURASIP Journal on Wireless Communications and Networking, Jan. 2010, pp. 1–5, Article ID 914934.

[12] Jialing Liu; R. Love, K. Stewart, and M.E. Buckley, "Design and Analysis of LTE Physical Downlink Control Channel," in Proc. 69th Vehicular Technology Conference, 2009. VTC-2009 Spring. IEEE, April 2009, pp. 1–5.

 ISBN: 978-1-61208-262-2

Performance Analysis of Stereo Matching Using Segmentation Based Disparity Map

Arti Khaparde, Apurva Naik, Manini Deshpande, Sakshi Khar, Kshitija Pandhari, Mayura Shewale
Department of Electronics and Telecommunication
Maharashtra Institute of Technology
Pune, Maharashtra, India
artikhaparde@gmail.com, apurva.naik@gmail.com, deshpande.manini@gmail.com

***Abstract*— Stereo vision has been studied extensively due to its usefulness in many applications like 3D scene reconstruction, robot navigation, etc. Rather than finding out the disparity between two original stereo images, various segmentation techniques are used to segment the images and the disparity between the resulting segmented images is calculated. The comparison between the disparity of the original stereo image pair and that of the segmented image pair is done on the basis of compression ratio and Peak-Signal to Noise Ratio (PSNR), which is calculated for image quality measurement. Segmentation techniques like Mean Shift Algorithm, K-means Algorithm and Particle Swarm Optimization (PSO) are used and their results are compared on the basis of subjective and objective parameters. The experimental results show that PSO based 3D image reconstruction gives a good compromise between subjective quality and compression ratio.**

Keywords-Disparity Map; PSNR; Mean Square Error (MSE); Compression ratio; Particle Swarm Optimization.

I. INTRODUCTION

Stereo Image matching is one of the core research areas in Computer Vision and Digital Photogrammetry. Technological developments in stereo image matching have advanced from the primitive area based cross-correlation technique to more and more precise feature-based and area-based matching [9]. Stereo allows us to recover information from the given two images about a three dimensional location of objects, which does not exist in any single image. The main goal of stereo image matching is to recover depth information from the given two or multiple images. In order to recover depth information the stereo images should be brought into point-point correspondence. Correspondence points are the projections of a single point into the three dimensional scene. The difference between the locations of these two correspondence points is known as parallax or disparity, which is a function of position of the point in the scene, orientation and physical characteristics of the camera. So, disparity can be used as constraint for matching. Although, feature-based techniques are more accurate, but they generate sparse disparity maps [1, 2, 4, 8]. Hence, in this paper, we are proposing an area-based segmentation method, as it generates a dense disparity map for 3D reconstruction.

The paper is organized as follows: Section 1 gives brief introduction, Section 2 deals with segmentation techniques used; Section 3 gives the disparity estimation algorithm used for analysis, Section 4 with results while analysis and conclusions are given in Section 5.

II. SEGMENTATION ALGORITHMS

For some applications, such as image recognition or stereo vision, whole images cannot be processed, as it not only increases the computational complexity, but it also requires more memory. In literature, number of image segmentation algorithms like K-means [3], mean-shift [4], etc., have been proposed and extensively applied to stereo vision. The algorithm assumes that disparity values vary smoothly in those regions and that depth discontinuities only occur on region boundaries. Purely pixel-based methods are insufficient to express information of the image. The human identifies the objects by analyzing features of the objects such as color, texture and shape. Thus, segmentation-based stereo matching algorithm should be used. Segment-based methods have attracted attention due to their good performance on handling boundaries and texture less regions. They are based on the assumption that the scene structure can be approximated by a set of non overlapping planes in the disparity space and that each plane of target image is coincident with at least one homogeneous color segment in the reference image. Segment based methods perform well in reducing the ambiguity associated with texture less regions and enhancing noise tolerance. The computational complexity is reduced due to much larger segments. Matching becomes much easier even in the presence of noise, intensity variation and slight deviations in segmented area. Noise tolerance is improved by aggregating over pixels with similar colors. One major reason is that small segments may be insufficient for estimating surfaces like slanted planes, while large segments may contain segmentation errors which can affect the accuracy of disparity estimation. Similar colors in image do not always mean similar disparity [1-3]. Mostly, all segment-based stereo matching algorithms employ mean-shift segmentation technique. Taking into consideration all the above points, Particle Swarm Optimization based segmentation algorithms are compared with existing techniques like K-means and Mean Shift Segmentation. Then, disparity estimation was carried out on these segments to do the subjective and objective analysis.

A. Mean Shift Algorithm

Mean shift is a nonparametric iterative algorithm or a nonparametric density gradient estimation using a generalized kernel approach. Mean shift is one of the most powerful clustering techniques [4].

 ISBN: 978-1-61208-262-2

Given n data points x_i, i = 1,… ,n on a d-dimensional space R^d, the multivariate kernel density estimate obtained with kernel $k(x)$ and window radius h is

$$f(x) = \frac{1}{nh^d}\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right) \quad (1)$$

For radially symmetric kernels, it suffices to define the profile of the kernel $k(x)$ satisfying

$$k(x) = c_{k,d}k(\|x\|) \quad (2)$$

Where $c_{k,d}$ is a normalization constant which assures $k(x)$ integrates to 1. The modes of the density function are located at the zeros of the gradient function $\nabla f(x) = 0$.

$$\nabla f(x) = \frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(x_i - x)\, g\left(\left\|\frac{x-x_i}{h}\right\|\right) \quad (3)$$

Mean Shift is

$$m(x) = \frac{\sum_{i=1}^{n} g(x-x_i/h)x_i}{\sum_{i=1}^{n} g(x-x_i/h)} - x \quad (4)$$

B. K-means Algorithm

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem [5]. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early grouping is done. At this point, we need to re-calculate k new centroids as bary centers of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, we may notice that the k centroids change their location step by step until no more changes are done. In other words, centroids do not move any more. Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_i^{(j)} - c_j\right\|^2 \quad (5)$$

where $||x_i^{(j)} - c_j||$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centers. The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

C. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy, inspired by social behavior of bird flocking or fish schooling. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. In each iteration, each particle is updated by following two "best" values. The first one is the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called X_{best}. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the population. This best value is a global best and called g_{best}. After finding the two best values, the particle updates its velocity and positions with following equation (6) and (7).

$$V_R = W * V_R + rand_1 * \left(r_1 * (X_{best} - X_R)\right) + rand_2 * \left(r_2 * (g_{auxR} * g_{best} - X_R)\right) \quad (6)$$

$$X_R = X_R + V_R \quad (7)$$

where,

W is initial weight,

$r_1 = r_2 = 0.8$,

X_R is randomly generated using maximum and minimum gray level values of image and is updated for N particles

g_{aux} is initialized to 1 and

$$g_{best} = \frac{\sum probR[1{:}X_R(j,1)] * \sum_1^{X_{R(j,1)}} XprobR[1{:}X_R(j,1)]}{\sum probR[1{:}X_R(j,1)] - \left(\sum_1^{Lmax}(probR(1{:}Lmax))^2\right)} \quad (8)$$

whose threshold value is kept equal to -10000

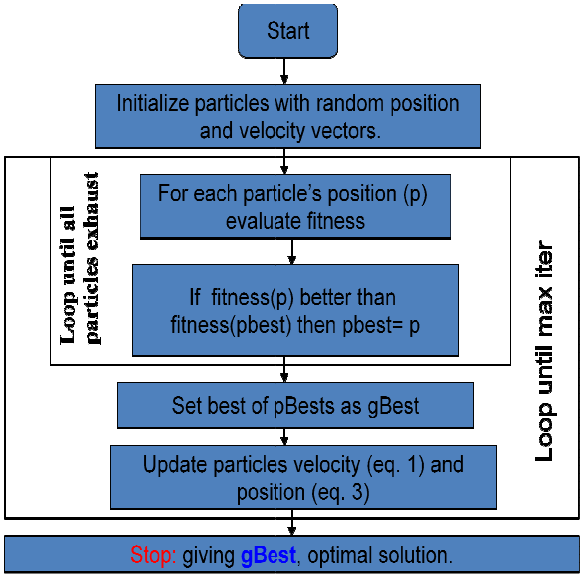


Figure 1. Flowchart of Particle Swarm Optimization

III. DISPARITY ESTIMATION

Disparity refers to the difference in image location of an object seen by the left and right eyes, resulting from the eyes' horizontal separation (parallax). The brain uses disparity to extract depth information from the two-dimensional retinal images in stereopsis. In computer vision,

disparity refers to the difference in coordinates of similar features within two stereo images as given by Bait et.al [8].

In the proposed work, disparity estimation is done in two steps: Disparity Computation and Disparity Optimization. Disparity is computed by finding the cost of matching point $I_l(x,y)$ in the left image to point $I_r(x,y,d)$ in the right image using Sum of Squared Differences (SSD), where matching cost is equal to square of difference of intensity values of pixels at disparity d and can be given as follows.

$$SSD(x,y,d) = \sum_{x,y\in W}\left(I_l(x,y) - I_r(x,y,d)\right)^2 \quad (9)$$

Disparity is computed as shown in Fig. 2. As a result, we get three sets of disparity cost. Optimization of these 3 sets is done by using winner-take-all method. This method inspects the cost associated with each disparity set via window centered on pixel. Disparity with smallest aggregated cost is selected and given as estimated disparity map.

Disparity estimation was done for four sets of image pair as follows

1. Left and right original images
2. Left and right segmented images using mean shift algorithm
3. Left and right segmented images using K-means
4. Left and right segmented images using PSO

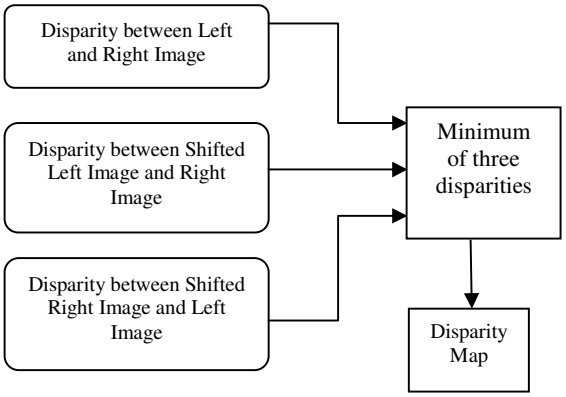


Figure 2. Process of Disparity Estimation

IV. RESULTS

Fig. 3 gives the flow chart for the proposed work for the reconstruction of 3D images. The segmentation stage is skipped when disparity is estimated for the original stereo image pair. It shows that disparity is estimated on the segmented image pair and reconstructed 3D images are also from segmented image pairs. These reconstructed 3D images and disparity obtained are used for both subjective and objective analysis. Fig. 4 illustrates the output of various stages of algorithm for image 'aloe', which is one of the image pairs from Middlebury dataset [10].

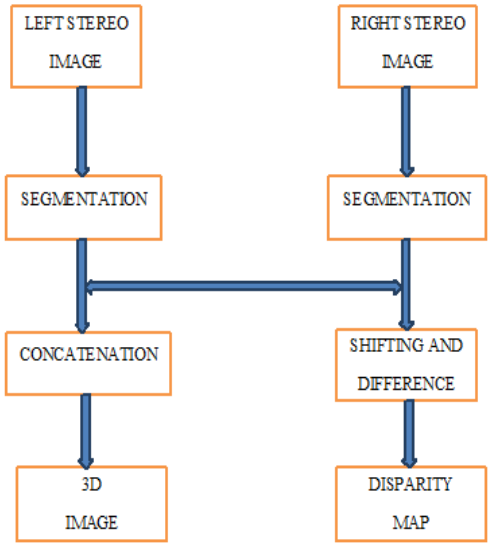


Figure 3. Proposed work flow

Figure. 4(a)

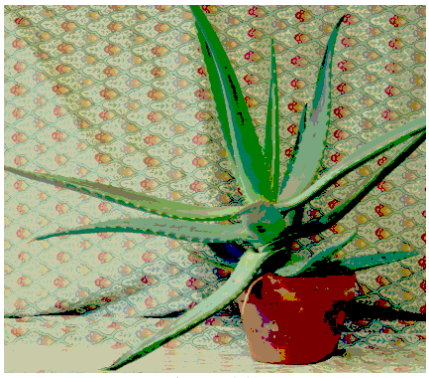


Figure. 4(b)

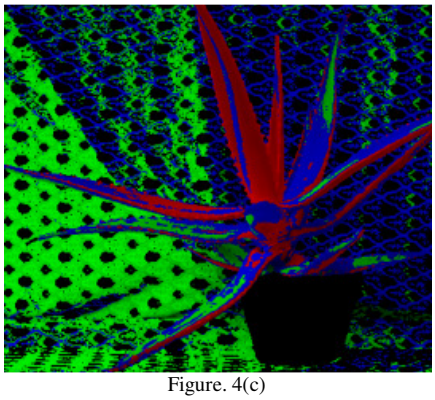


Figure. 4(c)

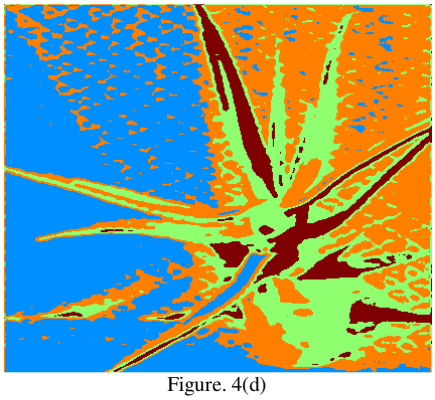


Figure. 4(d)

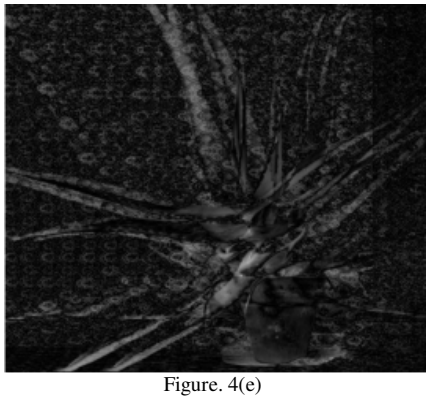


Figure. 4(e)

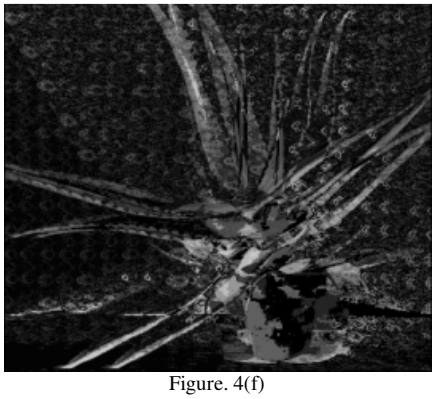


Figure. 4(f)

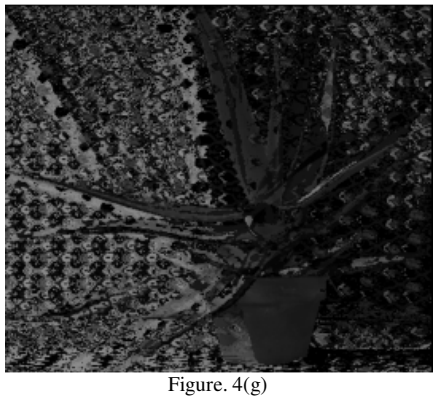


Figure. 4(g)

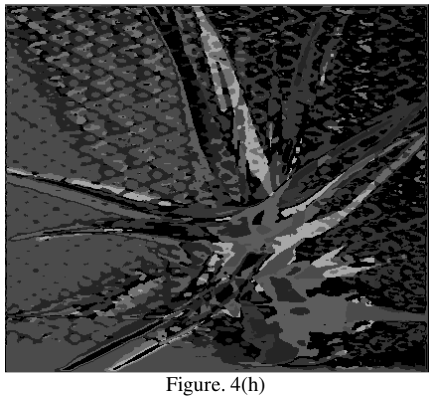


Figure. 4(h)

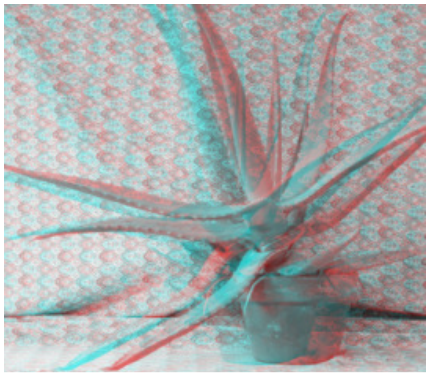


Figure. 4(i)

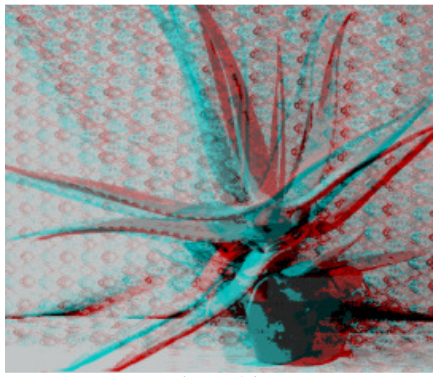


Figure. 4(j)

 ISBN: 978-1-61208-262-2

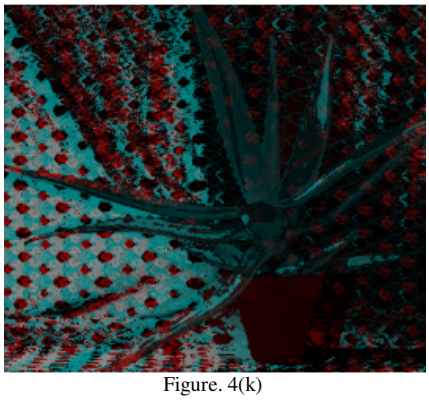


Figure. 4(k)

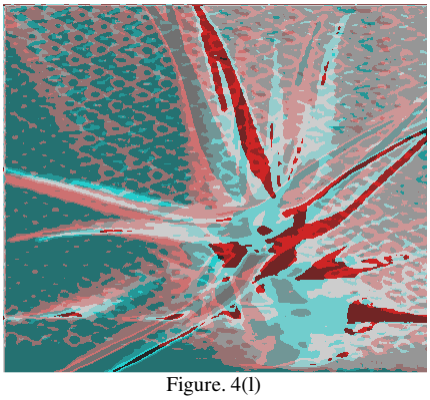


Figure. 4(l)

Figure 4. Implementation Results on 'Aloe' image

Fig. 4(a) represents the original left view from a stereo image pair. Fig. 4(b) is the segmented left image using PSO, Fig. 4(c) is the segmented left image using K-means, and Fig. 4(d) is the segmented left image using Mean Shift algorithm. Similar segmented images were obtained for the right view of stereo pair. Original stereo pair and segmented images pair were given as input for disparity estimation one-by one.

Fig. 4(e) is disparity obtained using original stereo pair while Fig. 4(f), Fig. 4(g) and Fig. 4(h) are disparities obtained from segmented images, using PSO, K-means Algorithm and Mean Shift Algorithm respectively.

Fig. 4(i) is the reconstructed 3D image obtained from original image pair. Fig. 4(j), Fig. 4(k) and Fig. 4(l) are the reconstructed 3D images obtained from PSO algorithm, K-means algorithm and Mean Shift algorithm, respectively. These images were tested on 100 subjects for subjective analysis.

V. CONCLUSION AND FUTURE WORK

All the algorithms were implemented on Middlebury database and were compared for the performance parameters like PSNR, compression ratio and number of depth levels extracted. Performance parameters for 12 such images are given here for reference.

Fig. 5 gives comparison for PSNR of disparity estimation using various segmentation techniques where vertical axis represents the PSNR in db with respect to disparity estimation of original image. It was observed that PSNR for all the three segmentation techniques is almost same.

Fig. 6 shows the plot of percentage compression on its vertical axis for different images. The compression ratio for PSO was calculated on an average of 50 % and for K-means was calculated on an average of 57% for the Middlebury database [10]. Even though the compression ratio of reconstructed 3D image based on PSO segmentation technique is less as compared to K-means, it can be seen that the subjective quality of PSO based 3D reconstructed images, (Fig. 4(j), Fig. 4(k) and Fig. 4(l)), gives much better vision than K-means and mean shift algorithms, which was found to 85% good and comparable to the 3D reconstruction using original images.

Fig. 7 shows the comparison of number of estimated depth levels for different algorithms. Number of Depth levels estimated using PSO segmentation is almost same to the number of depth levels estimated from original image. This may be one of the reasons why the subjective analysis for 3D image reconstructed using PSO segmentation provides better results compared to other two segmentation techniques.

PSO algorithm retains the original information of the image even after segmentation. PSO based segmented images provide better disparity estimation, with good number of estimated depth levels. The subjective quality of 3D images obtained using PSO technique is also better. In future, it can be one of the valuable segmentation techniques for color stereo vision. This paper also laid a good foundation for the study of improvement in the compression ratio using techniques like Functional Swarm Optimization (FSO) and Darwinian PSO (DPSO). The proposed algorithms not only approximate the boundaries of interest but may reduce the Computational complexity.

REFERENCES

[1] Z. Wang and Z. Zheng, "Region Based stereo Matching Algorithm Using Cooperative Optimization", Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 23-28 June 2008, Page(s): 1 – 8

[2] V. Borisagar and M. Zaveri, "A Novel Segment-based Stereo Matching Algorithm for Disparity Map Generation", 2011 International Conference on Computer and Software Modelling IPCSIT vol.14 (2011) © (2011) IACSIT Press, Singapore.

[3] A. Klaus, M. Sormann, and Karner, "K-Segment-Based Stereo Matching Using Belief Propagation"Pattern Recognition, 2006. ICPR 2006. 18th Internatonal Conference, Page(s):15-18

[4] D. Comaniciu and P. Meer, "Mean shift:A Robust Approach Toward Feature Space Analysis", Pattern Analysis and Machine Intelligence, IEEE Transactions 2002, Page(s):603-619

[5] K. Javed, "The Behaviour of k-Means:An Empirical Study" , Electrical Engineering 2008. ICEE 2008, Second International Conference on 25-26 March, Page(s):1-6

[6] P. Ghamisi, M. Couceiro, J. Benediktsson, and M. Ferreira, "An Efficient Method for segmentation of Images based on Fractional

 ISBN: 978-1-61208-262-2

Calculus and Natural Selection Expert Systems with Applications", An International Journal,Volume 39, Issue 16, November 2012.

[7] F. Bergh and A. Engelbrecht, "A study of Particle Swarm Optimization Particle Trajectories", Information Sciences: An International Journal, Volume 176, Issue 8, April 2006.

[8] T. Bait, D. Boudaoud, B. Matsuzewski, and L. Shark, "An Efficient Feature Based Matching Algorithm for Stereo Images", Geometric Modelling and Imaging, 2006, Page(s):195-202

[9] J. Joglekar and S. Gedam, "Area Based Stereo Image Matching Technique Using Hausdorff Distance And Texture Analysis", International Journal on Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 1, January 2012.

[10] www.vision.middlebury.edu/stereo/dataset, date: 23rd Feb 2013

Figure 5.Plot of PSNR of three segmentation techniques

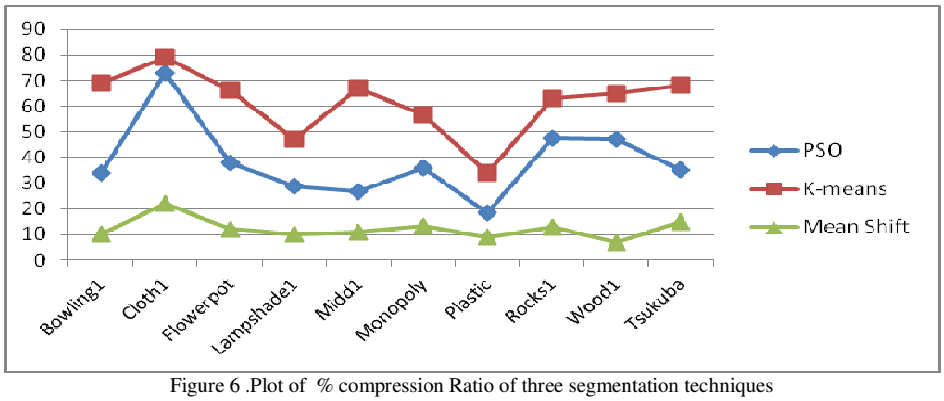


Figure 6 .Plot of % compression Ratio of three segmentation techniques

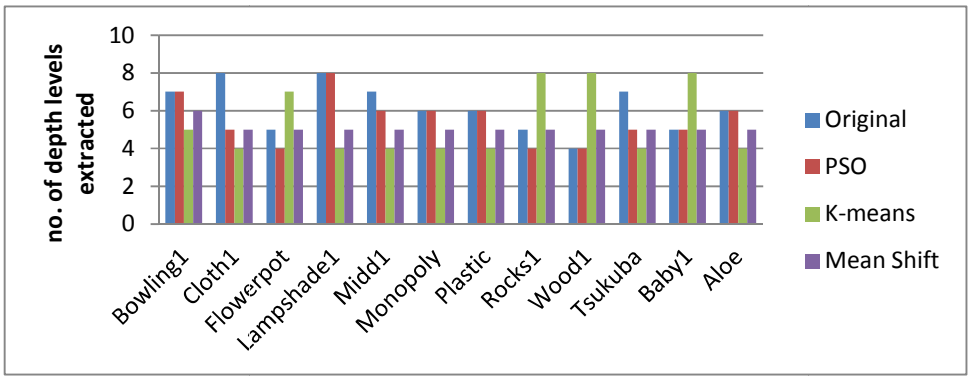


Figure 7.Plot of estimated depth levels of three segmentation techniques

G/G/c/c Simulation Model for VoIP Traffic Engineering with non-Parametric Validation

Imad AL Ajarmeh
James Yu
School of Computing and Digital Media
DePaul University
Chicago, USA
{iajarmeh, jyu}@cdm.depaul.edu

Mohamed Amezziane
Dept. of Mathematical Sciences
DePaul University
Chicago, USA
mamezzia@depaul.edu

***Abstract*— The widespread of Voice over IP systems necessitate finding suitable and modern traffic engineering models that can help design cost-efficient systems and study their performance under various conditions. In this paper, we provide a new VoIP simulation suite that consists of a parametric simulator based on Non-homogeneous Poisson Process call arrival model, and a non-parametric simulator based on real traffic data. Our simulators are validated against real call data obtained from multiple offices of a production VoIP carrier network. The simulation results show that our simulator can provide up to 28% better resource utilization than the legacy Erlang B model. Our simulator can also help carriers dynamically allocate network bandwidth to meet various traffic demands.**

Keywords- VoIP traffic engineering; simulatin; VSIM; modeling; G/G/c/c; non-parametric.

I. INTRODUCTION

The majority of the previous work in Voice over IP (VoIP) traffic engineering and modeling is based on the exponential approximations for call arrival rate and call holding time [3][16][17][18]. The exponential approximation allows finding analytical solution for the traffic queuing model but the approximation might be too aggressive that it will result in poorly engineered systems. The Erlang B model was introduced several decades ago to solve the phone system traffic queuing problem. This model is based on the traffic intensity of the busiest hour in the busiest week of the year (Busy Season Busy Hour: BSBH). BSBH traffic is assumed constant throughout the entire year and its arrival rate is modeled as a Poisson/exponential distribution. This assumption makes traffic calculations easier but using a constant call arrival rate for the entire year causes inefficient resource utilization.

We proposed using a Non-homogeneous Poisson Process (NHPP) for the call arrival rate in our previous work [1]. In NHPP modeling call arrival rate is a non-constant function of time. Whereas in the legacy Erlang B approach calls are assumed to arrive according to a Poisson process with a constant arrival rate. Therefore, using NHPP helps avoid the approximation and assumption errors associated with a constant arrival rate over the whole engineering period. Our NHPP model development was based on real call data extracted from a production VoIP carrier network. Examining the arrival data, we constructed a model that describes the variation of call arrival rates during a week since traffic patterns were observed to be repeated weekly. It is common in statistical analysis to model the logarithm of λ(t) instead of λ(t) itself for *count* data. Such transformation would guarantee that the estimate of the intensity function is always non-negative. Our model takes into consideration the daily arrival patterns and has the time-dependent intensity function of:

$$\boldsymbol{log}[\boldsymbol{\lambda}(\boldsymbol{t})] = \boldsymbol{\mu} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{k_o}} [\boldsymbol{\alpha_i}\,\mathbf{sin}(\boldsymbol{i\omega_o t}) + \boldsymbol{\beta_i}\,\mathbf{cos}(\boldsymbol{i\omega_o t})] + \sum_{\boldsymbol{j=1}}^{\mathbf{6}} \boldsymbol{\gamma_j I_j(t)} \qquad (\mathbf{1})$$

where: $\boldsymbol{\lambda(t)}$ is a function of time (t).

$\boldsymbol{I_j(t)}$ is day Indicator function where $\boldsymbol{j}$ is the day of the week. The value of $\boldsymbol{I_j(t)}$ is 1 if the time $\boldsymbol{t} \in \boldsymbol{j}$ and 0 otherwise. $\boldsymbol{k_o}$ is the number of harmonics in the model. $\boldsymbol{\mu}$ represents the model central tendency without daily effects. $\boldsymbol{\gamma_j}$ is the effect of day *j* and represents the difference between $\boldsymbol{\mu}$ and the mean number of calls for day *j*. $\boldsymbol{\alpha_i}$, and $\boldsymbol{\beta_i}$ are the contribution of the *ith* harmonic to the model.

We used maximum likelihood estimation to fit the proposed λ(t) to the actual call arrivals. In addition, we used likelihood ratio test and Wald's test in order to verify the significance of the model and its parameters. All the statistical test results verify that call arrivals is best fit by a NHPP rather than a constant Poisson process. We provide the detailed statistical analysis in [1]. In this paper, we introduce a comprehensive VoIP simulation suite (VSIM). VSIM consists of a G/G/c/c simulation model. According to Kendall's notation the G/G/c/c is a queuing system where calls are assumed to arrive according to a general distribution (G) and have a service time that follows another general distribution (G), the system has a limited

 ISBN: 978-1-61208-262-2

number of servers/channels (c) and no waiting queue (maximum number of calls in the system equals the number of servers c). VSIM includes a NHPP call arrival rate G/G/c/c simulator and also a non-parametric simulator based on real traffic data. The simulation models are validated against traffic data collected from an operational VoIP network.

The remaining of the paper is organized as follows: section II contains a description of the telecommunication system simulation approaches, section III gives detailed description of VSIM, section IV describes our approach in verifying the correctness of VSIM algorithms, in section V we show our process in validating VSIM results, and in section VI we present some results of VSIM simulations.

II. TELECOMMUNICATION SYSTEM SIMULATION

In telecommunication traffic engineering, it is always preferable to find analytical solutions for the queuing and traffic problems. However, the analytical solution might involve too many approximations in order to fit the data into exponential or other probability distribution functions. Such approximations will result in inaccurate engineering results. Simulation approach offers accurate and flexible model construction and validation, and can be used whenever analytical solutions are not practical [11].

Simulation models can be discrete or continuous. Discrete event simulations are suitable for problems where variables change in discrete time fashion. On the other hand, continuous simulations are suitable for problems in which the variables might change continuously [8]. Discrete event simulations are suitable for telecommunication network queuing problems since the events happen on discrete times [9]. Using discrete event simulators has been becoming more popular during the past few years because such simulations can help solving sophisticated problems, which are impossible to be solved using analytical approaches [10][11]. In addition, the availability of low-cost powerful computers and capable simulation packages makes the simulation-based solutions more accurate, capable, and easier to implement.

With the rapid increase of VoIP residential, enterprise, and carrier deployments, researchers realized the need for modern traffic simulation models that can be used in studying and designing reliable and cost-efficient VoIP networks. In [4] VoIP traffic sources were modeled as on-off sources with exponentially distributed of on-and-off times. In [6] and [7] the authors used Markov modulated Poisson process (MMPP) traffic model to analyze VoIP performance for wired and wireless networks. In [12] [13] [14] [15] the authors provide VoIP traffic performance and evaluation simulation tools that focus on the packet performance without taking into consideration the distribution of calls arriving at the system, which will have significant impact on the packet performance and QoS design. In this work we go a further step by providing two VoIP traffic simulators based on modeling the calls arriving at the system. The first one is a parametric model and uses a NHPP to represent the time-dependent call arrival rate, and the second one is non-parametric and uses the real traffic data to simulate the system behavior

III. VOIP TRAFFIC SIMULATION MODEL (VSIM)

VSIM is part of a larger traffic engineering system that starts by collecting call data from a production network. The collected data is processed and then fed into the NHPP parameter estimation model. NHPP model parameters are passed to VSIM to be used in G/G/c/c engine. For non-parametric simulation we skip NHPP model estimation process and feed the processed call arrival and call holding time data directly into VSIM. Figure 2 illustrates a high-level design for such traffic engineering system, and Figure 1 shows a sample of the collected call data. Each raw carries information for one call with some proprietary data (trunk group name and switch name) blanked. The remaining fields are: record type (STOP means a completed call), start date and time, end date and time, and call duration in 100s of seconds.

Examining the traffic patterns in the collected data, we notice large variation in the arrival rate. For example, at one second we might receive 10 calls and at the next second we might receive no calls. This variation is smoothed if we average the traffic data over longer time intervals. Figure 3 shows the raw traffic data for 1s, 10s, and 3600s averages, along with the generated NHPP model. The figure illustrates the accuracy and significance of the generated model. This accuracy has been established through the extensive mathematical and statistical analysis we provided in [1] and [2]. The accuracy of the input NHPP model will result in accurate simulation results as proven in section VII

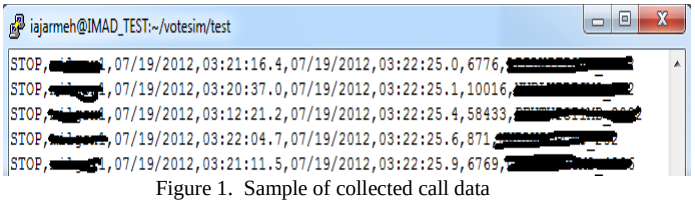


Figure 1. Sample of collected call data

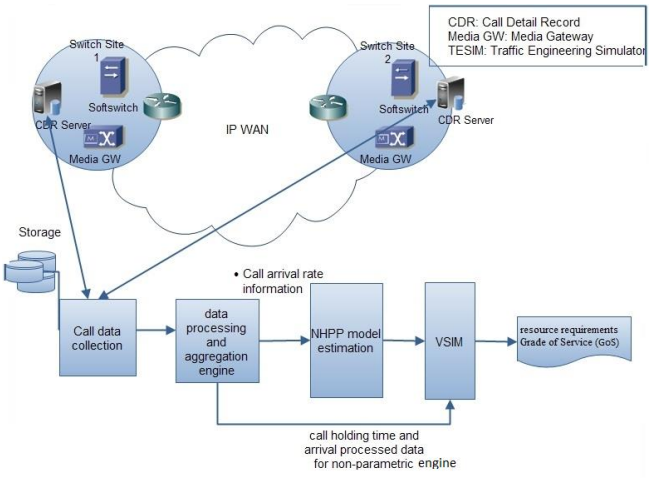


Figure 2. Traffic engineering system

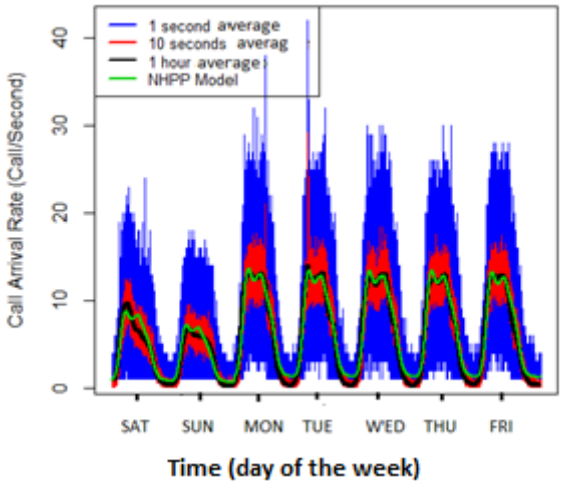


Figure 3. Call arrival data analysis and modeling

The VoIP simulation model (VSIM) estimates:

- resource requirements (IP trunk size) given a certain traffic pattern and a target blocking probability (Grade of Service: GoS), and
- GoS (blocking probability) given a certain traffic pattern and a known amount of call resources (IP trunks).

Figure 4 shows a sample of VSIM output.

VSIM is a flexible Java-based tool developed using CSIM for Java library [4], and therefore it can easily be ported to different computing platforms. VSIM can be used to estimate the trunk group size, generate GoS reports, and perform what-if analysis for VoIP networks.

VSIM is composed of two different simulation engines; the first is a parametric G/G/c/c simulator and the second is a non-parametric simulator. Both are discrete event simulators in which the VoIP system is modeled as chronological sequence of call arrivals and terminations. In the G/G/c/c engine we model the call arrival rate using the time-dependent function shown in (1), model parameters are estimated based on the collected data sample.

This call arrival function is used to generate random variables for call inter-arrival times. Once a call is generated, the simulation code polls a random call holding time from a list of real holding times. The simulation engine allocates a trunk for the duration of simulated call holding time. A separate thread is created for each call so that we can collect statistics for each individual call and trunk.

```
imad@linux18:/home/VSIM
Execution time: 22.826 S
Simulated Time: 604800.0 S
Simulated Time Offset: 0.0 S
Number of IP Trunks = 200
Holding Time data file = data/ht-nycgsx17-Mar-w24
NHPP parameters file = data/nhpp-nycgsx17-mar-w24
Blocked calls= 1635
Completed calls= 233587
Blocking Probability = 0.006950880444856348

Output is saved to GGK.out
```

Figure 4. VSIM sample output

Once the simulated call time is over, the completed calls counter is incremented by one and the trunk will be released back to the trunk pool.

The same procedure is repeated for the next calls until the pool of trunks is depleted. Once all trunks are busy we increment the blocked calls counter for each call that arrives while no trunks are available. Figure 5 illustrates the internal VSIM algorithm

The non-parametric simulator follows the same algorithm with the exception that we poll the inter-arrival time variable from a real data file rather than using a NHPP function to generate it.

A. Parametric VSIM G/G/c/c simulator

Parametric simulation is done by collecting traffic data and then developing statistical models that best approximate the collected data. Model parameters are estimated based on the data sample and then these parameters are used in the simulation. We developed a G/G/c/c simulation model for VoIP traffic engineering. The model consists of a loss multi-server queuing system with waiting queue length equal to zero (blocked calls are cleared from the system). The implementation of general call arrival rate and general call holding time in the simulator allows for arbitrary distributions and that increases the flexibility and usability of our simulation model. The examples given in this paper focus on modeling call arrival rate as NHPP using a generalized linear model that captures the variability in call arrival rate with respect to time. NHPP model parameters are estimated based on the real traffic data extracted from the production VoIP network under study.

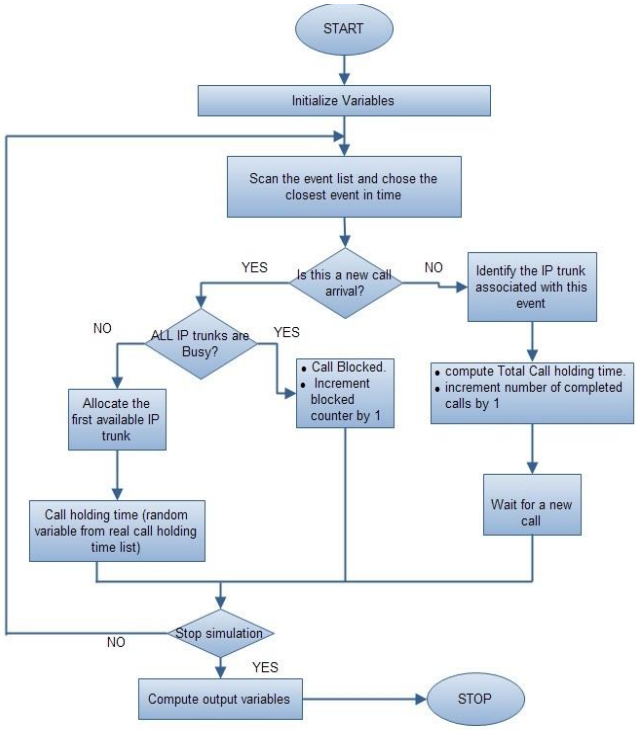


Figure 5. VSIM algorithm

B. Non-parametric VSIM simulator

In addition to the parametric G/G/c/c simulation engine implemented in VSIM, we also provide another non-parametric simulation engine. Non-parametric simulation is achieved throughout replaying the real traffic data without generating statistical models or estimating parameters. In other words, we use actual observations in the simulation rather than generating random variables from a statistical distribution. Therefore, non-parametric simulations are easier and tend to yield more accurate results since no modeling approximations are involved. This type of simulation is preferred when the data sample is large.

We use the non-parametric simulator in order to verify the correctness of our G/G/c/c model and to validate its results. The input of our non-parametric simulator is real call information that consists of call inter-arrival time and call holding time. The simulator will regenerate the calls based on the given data, and we can study the system and compute the required resources and GoS.

IV. VSIM MODEL VERIFICATION

It is important to verify the correctness of any simulation model before applying it to real-life problems. Simulation verification should cover the simulation engine algorithms as well as the random variables generated from the simulator statistical models. Therefore, we split VSIM model verification into two steps; the first is discussed in (*A*) and aims to verify the correctness of the NHPP random variables generated and used by the simulator. The second step is discussed in (*B*) and aims to verify the correctness of the simulation algorithms, timers and procedures

A. Internal simulation random variables

We instrumented VSIM and obtained the call arrival rate generated by the model based on the implemented NHPP linear model. This call rate is used as an internal input to the G/G/c/c simulation algorithm. Figure 6 shows the internally generated NHPP random variable along with the corresponding inter-arrival time against the actual traffic data

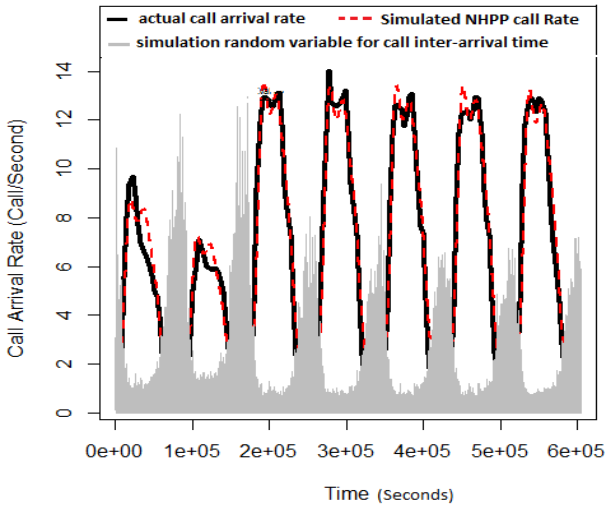


Figure 6. Simulated call arrival rate

As illustrated in Figure 6, the results of this verification process indicate that the NHPP model function used to generate call arrival random variables is correct and accurate.

B. VSIM simulation engine algorithms

A common approach to verify the correctness of a new simulation model is by comparing it to other well-established simulators. Unfortunately, we could not obtain any G/G/c/c model against which we could verify our work; therefore, we used the M/M/c/c special case of our model and compared the simulated result to the calculated results based on the Erlang B model. Our M/M/c/c utilizes the same simulation algorithms as our G/G/c/c and the only difference is that we use exponential distributions for both call arrival rate and call holding time. The goal is to verify the correctness of our simulation clock and algorithms. The validation of simulation results will be discussed in the next section. In this process we used different data samples each one consists of one week of traffic; an example of the results is shown below:

Mean call holding time = 185 second
Busy Hour Traffic (BHT) = 12.6 call/ second

Using the Erlang B calculator we need around 2520 trunks in order to carry this traffic without blocking (Blocking probability nearing zero). Using the M/M/c/c simulator with the same traffic parameters we found the required number of trunks to be 2515 for the same blocking probability. Comparable results were obtained for all the samples under test. These results verify the correctness of our VSIM simulation code and algorithms.

V. VSIM SIMULATION MODEL VALIDATION

The most important aspect of any simulator is that it should produce valid, correct and dependable results. The best approach to establish the validity of a simulator results is by comparison to real data. Therefore, we obtained the real resource utilization (number of simultaneous calls) from the production system for the period of time corresponding to the call arrival and holding time data used to develop the models. We used this data to validate VSIM simulation results. Different traffic samples and different simulation runs were used and all results agree with data obtained from the real network. In addition, we also used our nonparametric VSIM simulator to replay the same data samples and the results agree with those obtained from the system and those obtained from G/G/c/c simulation.

TABLE I. shows an example of the actual trunks obtained from the system compared to VSIM simulated output

TABLE I. SIMULATED VS ACTUAL IP TRUNK REQUIREMENTS

Number of Required IP Trunks	**Maximum Call Load (Pr[B] ≈ 0)**
Actual (Observed)	1807
G/G/c/c (simulated)	1936
Non-Parametric (Simulated)	1810

 ISBN: 978-1-61208-262-2

It can be seen from TABLE I. that both simulation models yield satisfactory results although the non-parametric model is a little better. The reason is because we don't have any modeling or estimation approximations for the non-parametric case. We used multiple data samples and executed multiple simulation runs and all the results are similar and indicate high accuracy of our VSIM for both G/G/c/c and non-parametric while the latter shows a little better results.

VI. RESULTS AND ANALYSIS

VSIM G/G/c/c simulator is based on using a function of time to model call arrival rate, and therefore VSIM can provide the resource requirements as a function of time as well. This function is important for system design, analysis and requirement studies, especially for converged networks where voice and data ride the same IP infra-structure. This function is also available for VSIM non-parametric simulator because we have real call information that depends on the time. Figure 7 illustrates sample resource functions (number of required IP trunks Vs. Time) generated by VSIM along with the corresponding simultaneous calls observed in the actual system (real data validation). The figure shows the effectiveness of VSIM as demonstrated by its ability to compute the required system resources (IP trunks) as a function of time accurately. The resource time function provided by VSIM can be utilized for dynamic resource allocation scheme in which resources are allocated for different applications based on the actual or expected demand. Such scheme helps achieve better resource utilization and hence better engineering and cost reduction. It is important to notice that Erlang B and M/M/c/c models suggest a linear relation between blocking probability and system capacity (maximum number of simultaneous calls). However, our VSIM G/G/c/c and non-parametric simulators suggest a non-linear relation as seen in Figure 8.

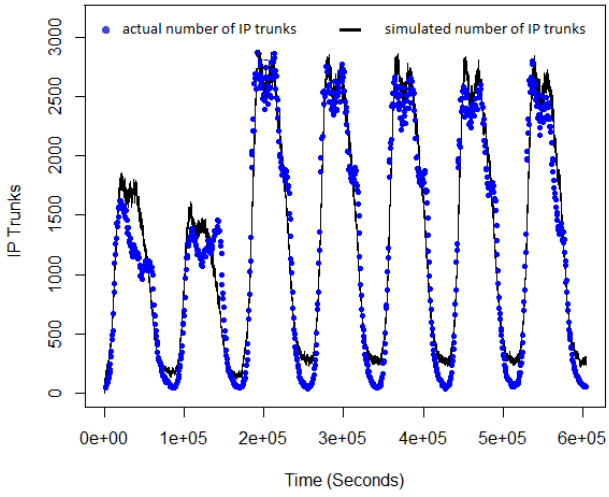


Figure 7. Number of IP trunks as a function of time (resource time function)

The figure shows identical match between the results obtained from the calculated Erlang B and the simulated M/M/c/c. In order to use Erlang B and M/M/c/c, we need to compute the average call arrival rate of the busiest hour and the average call holding time. For the results shown in Figure **8** and based on the data sample, we used 10.6 call/second for the arrival rate and 183.01 seconds for call holding time. Also, the figure shows close match between the G/G/c/c and non-parametric simulator. These results verify the correctness and validity of our procedure and modeling process. The deviation between the straight line calculated by the traditional Erlang B model and the curve generated by VSIM is significant and can affect the design and engineering decisions for the system. For example, if we want to build a switching system with 1600 maximum simultaneous calls (system capacity), the Erlang B approach suggests that the blocking probability will be 0.19 (P.19) while the VSIM G/G/c/c model results in a blocking probability of 0.02 (P.02). The difference between these two approaches is significant in the telecom world. VSIM nonparametric approach for the same data sample results in a blocking probability of 0.006 (P.006). Using the same example, we found that in order to achieve blocking probability of 0.01 (P.01), we will provision 1665 IP trunks using VSIM G/G/c/c model, or provision 1550 IP trunks using the VSIM non-parametric simulator. On the other hand we will provision 1991 trunks if we engineer the system using the Erlang B model. Therefore, we can see that using the VSIM model can save 28% of the resources over Erlang B at the P.01 blocking probability. Furthermore, Figure 8 suggests that we can achieve better than 28% resource saving if higher blocking probabilities is desired.

We conducted many simulation runs using different data samples collected from different switch offices located in different cities. Similar results were obtained throughout this study. Figure 9 shows another example where the data is collected from a different office with more trunks. It shows almost identical relation between the blocking probability and system resources.

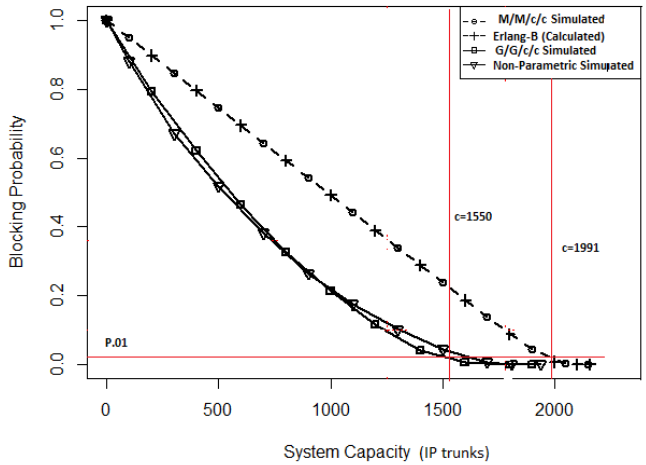


Figure 8. Blocking probability Vs system capacity

 ISBN: 978-1-61208-262-2

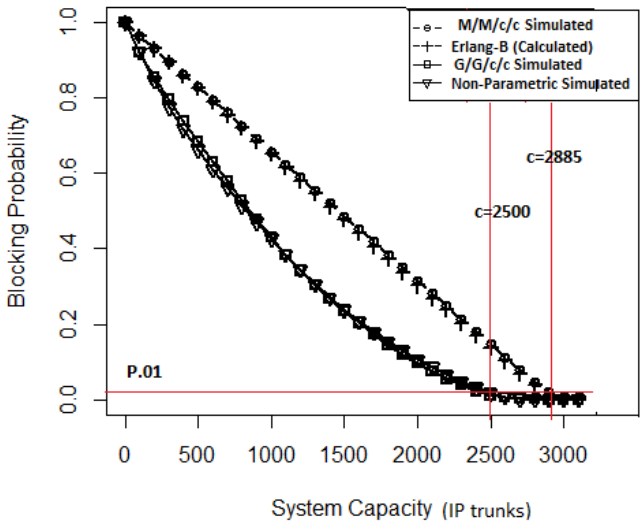


Figure 9. Blocking probability Vs system capacity example 2

For the results shown in Figure 9 and based on the data sample, we used 14.0125 call/second for the arrival rate and 204.186 seconds for call holding time

The graph suggests that we need 2500 IP trunks in order to carry the offered traffic load at blocking probability of 1%. On the other hand using the Erlang B model we will need 2885 IP trunks in order to achieve the same GoS. This sample shows that VSIM model could yield a saving of 15% of network resources in this office.

VII. CONCLUSION

One disadvantage of using complex traffic models such as NHPP is that an analytical solution is not feasible. With the availability of powerful computers, it makes the simulation approach feasible and effective, and hence, we can obtain accurate results. We provided two simulators: the first is based on NHPP call arrival rate and the second is based on non-parametric data. VSIM is capable of solving the traffic modeling problem for modern VoIP systems accurately using arbitrary and complex traffic models or by using the raw traffic information without estimation of parameters. Our results are validated against real data collected from multiple offices of a production VoIP carrier network. We observed that the non-parametric simulator results are more accurate. Real traffic data proves that using VSIM could save up to 28% of the resources over the Erlang B model or other exponential-based models

REFERENCES

[1] I. Al Ajarmeh, J. Yu, and M. Amezziane, "Framework of applying a non-homogeneous Poisson process to model VoIP traffic on tandem networks," 10th WSEAS International Conference on Informatics and Communications, Taipei, Taiwan, August 2010.

[2] I. Al Ajarmeh, J. Yu, and M. Amezziane, "Modeling VoIP traffic on converged IP networks with dynamic resource allocation," International Journal of Communications, ISSN: 1998-4480, Issue 1, Volume 4, 2010, pp. 47-55

[3] K. Salah and A. Alkhoraidly, "An OPNET-based simulation approach for deploying VoIP," International Journal of Network Management, Volume 16 Issue 3, May 2006, pp. 159-183

[4] http://www.mesquite.com/documentation/documents/CSIM_for_Java-UserGuide.pdf [retrieved: Feb, 2013]

[5] V. Rakocevic, R. Stewart and R. Flynn, "VoIP network dimensioning using delay and loss bounds for voice and data applications," Techincal Report, 2008.

[6] H. Lee and D. Cho, "VoIP capacity analysis in cognitive radio system," IEEE Communications Letters, VOL. 13, NO. 6, June 2009

[7] J-W. So, "Performance analysis of VoIP services in the IEEE 802.16e OFDMA system with inband signaling," IEEE Trans. Veh. Technol., vol.57, no. 3, May 2008, pp. 1876–1886

[8] A. M. Law and W. D. Kelton, "Simulation modeling and analysis," third ed. New York: McGraw-Hill, 2000

[9] O. Özgün and Y. Barlas, "Discrete vs. continuous simulation: when does it matter?," Proceedings of the 27th International Conference of The System Dynamics Society, July, 2009, Albuquerque, NM, USA

[10] G. Yuehong, Z. Xin, Y. Dacheng, and J. Yuming, "Unified simulation evaluation for mobile broadband technologies," IEEE Communications Magazine, vol. 47, Issue 3, March 2009, pp. 142-149

[11] N. I. Sarkar and S. A. Halim, "A review of simulation of telecommunication networks: simulators, classification, comparison, methodologies, and recommendations," Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), March Edition, 2011

[12] M. Bohge and M. Renwanz, "A realistic VoIP traffic generation and evaluation tool for OMNeT++," International Workshop on OMNeT++, Marseille, France, March 2008

[13] A. Bacioccola, C. Cicconetti, and G. Stea, "User-level performance evaluation of VoIP using ns-2," NSTools'07, October 22, 2007, Nantes, France

[14] J. D. Gupta, S. Howard, and A. Howard, "Traffic behavior of VoIP in a simulated access network," International Transactions on Engineering, Computing and Technology, 18, ISSN 1305-5313, 2006, pp. 189-194

[15] R. S. Naoum and M. Maswady, "Performance evaluation for VOIP over IP and MPLS," World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 3, 2012, pp. 110-114

[16] S. Shin and H. Schulzrinne, "Experimental measurement of the capacity for VoIP traffic in IEEE 802.11 WLANs," INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, May 2007

[17] T. D. Dang, B. Sonkoly, and S. Molnár, "Fractal analysis and modeling of VoIP traffic," 11th International Telecommunications Network Strategy and Planning Symposium, Austria 2004

[18] M. J. Fischer and D. M. Bevilacqua Masi, "Modeling overloaded voice over Internet Protocol systems," The Telecommunications Review 2006, pp. 94 - 106

 ISBN: 978-1-61208-262-2

Investigating Image Processing Based Aligner for Large Texts

Application for Under-Resourced Languages

Andi BUZO, Horia CUCU, Corneliu BURILEANU
SpeeD Laboratory
University "Politehnica" of Bucharest
Bucharest, Romania
e-mail: {andi.buzo, horia.cucu}@upb.ro, cburileanu@messnet.pub.ro

***Abstract*—Speech annotation is a costly and time consuming process because it requires high accuracy. Lightly supervised acoustic modeling solves this problem by making use of approximate transcriptions of speech recordings. In under-resourced languages, the speech recordings are not transcribed entirely and the accuracy of the transcription is poor. In this case, it is necessary an additional segmentation step. We propose a segmentation method that uses image processing techniques in order to spot a text island into a larger one. We also investigate on the effect of several tuning parameters on the method's accuracy.**

Keywords-Lightly Supervised Acoustic Modeling; Text Alignment; Under-resourced Languages; Image Processing.

I. INTRODUCTION

A. Motivation

State-of-the-art Automatic Speech Recognition (ASR) systems are based on large annotated speech data and text corpora in order to train accurate acoustic model and language model. The acoustic model requires hundreds or thousands of hours of annotated speech recordings, while the size of needed text corpora exceeds 1 terra words. The acquisition of such data is a time consuming and costly process. In particular, speech annotation requires higher attention because the speech label must represent the exact speech content. Reformulation of phrases, acronyms, and numbers will degrade the acoustic model accuracy.

Under-resourced languages are characterized by lack of speech and text databases, phonetic dictionaries, lexicons, tools and language expertise. For such languages, web data may represent an important source. However, web data are highly heterogeneous. Speech coding may be different and the sampling frequency may vary from one recording to another. The text content of the speech may not be exactly the same to the one uttered in the recording. Such transcriptions are referred to as loose transcriptions. Non-uniform text formatting should also be addressed through a normalization stage and, for particular languages, even a diacritics restoration stage.

Lightly supervised acoustic model training is largely used in the literature in order to obtain annotated speech databases from loose transcriptions. It is based on an ASR with fair performance that decodes the speech recordings. Then, the ASR output hypothesis is aligned to the loose transcription typically by Dynamic Time Warping (DTW) based techniques. But, the usual scenario in under-resourced languages is the one in which the speech transcription is found in a large text or only a small part of a long recording is transcribed. In this case, the DTW based techniques are not able to align the ASR hypothesis to the text transcription. Hence, a new step must be provided in order to spot the ASR hypothesis within the transcription.

We have proposed such an aligning method in [1]. In this paper, we investigate the effect of the tuning parameters that help increase the alignment accuracy. We also investigate the selection of the confidence scoring that determines the correctly aligned group of words. The rest of the paper is organized as follows: Section II describes the proposed method, Section III presents the experimental setup and the obtained results, while conclusions are given in Section IV.

B. Related Work

In the literature, there are several approaches that overcome the lack of resources barrier. One approach is by adapting the acoustic model of a highly-resourced language with little data from an under-resourced language [2]. Another approach is by building multilingual acoustic and language models as in [3]. Other approaches are focused on obtaining annotated data with little effort. The most common one is lightly supervised acoustic model training.

The first steps in lightly supervised acoustic modeling are made by Moreno et al. in [4]. In this paper, the authors look for confidence text islands called anchors that are supposed to be correctly alignments of ASR hypotheses to loose transcriptions. DTW is used for the alignment, while the confidence text islands are considered to be the ones where the density of the matched words is higher. The anchors are used for segmenting the transcription, so that in a further iteration a specific language model may be used in each segment. The granularity of the transcription is increased at each iteration until convergence is reached. Similar methods are also used in [5] and [6] with small variations in choosing the confidence scoring method. In these methods, strong emphasis is put in biasing the language model towards the transcriptions, which boosts the ASR accuracy.

Another way of matching the ASR hypothesis and the loose transcription is by forced alignment. The main problem with forced alignment is that it assumes that the transcription

 ISBN: 978-1-61208-262-2

represents exactly the speech content, which is not the case for loose transcriptions. In [7], Hazen overcomes this drawback by adding a phone loop branch to the decoding network. This extra branch is supposed to absorb the errors that may occur in the transcription, similar to the way out-of-vocabulary words are treated. In [8], the authors solve the same problem by training a *garbage model*.

These approaches are efficient if the size of the ASR hypothesis is comparable to the one of the transcription, otherwise they fail to align them. In [9], the authors attempt to resolve the issue of spotting a text island into a larger text within the context of lightly supervised acoustic modeling. They use the driven decoding algorithm as first iteration where confidence scores are calculated for the recognition hypotheses. These scores are used for language model rescoring in the second iteration.

In the proposed approach, we aim at resolving two main issues with DTW based techniques. The first is the incapability that DTW sometimes has in aligning texts with considerable different sizes. The repetitive words such as common nouns and verbs, prepositions, conjunctions, etc., combined with the poor accuracy of the ASR (especially for under-resourced languages) may create an optimal path that does not necessarily pass through the correct transcription. The second issue is the amount of memory needed for aligning large texts. It may exceed 20 GB if the text has more than 2000 words.

We overcome these issues by proposing a segmentation step whose purpose is to spot the ASR hypothesis (the short text) into the loose transcription (the large text). We achieve this by processing the ASR hypothesis-to-transcription match matrix as a grayscale image. The $m(i,j)$ element of the matrix is 1 if word i of the ASR hypothesis is identical to word j of the loose transcription and 0 otherwise. The matching segment should represent a diagonal line in the image (matrix), thus the method aims at emphasizing this line. This method proved to obtain better results than the DTW based techniques in terms of correctly spotting the target text within a larger one [1].

II. METHOD DESCRIPTION

A. Lightly Supervised Acoustic Modeling Overview

Lightly supervised acoustic modeling makes use of the loose transcriptions, which approximate the speech content of recordings. Such transcriptions are easier and cheaper to create or to obtain from the Internet. The ASR output hypotheses are obtained by decoding the speech recording using an ASR system with fair accuracy. Intuitively, identical word sequences in the ASR hypothesis and in the transcription suggest that these particular sequences of words are uttered in the respective speech segments. Hence, an alignment of the ASR hypothesis to the transcription is needed in order to localize such regions. Once identified, these segments represent annotated speech that can be used for training a new acoustic model.

In under-resourced languages, lightly supervised acoustic modeling raises supplementary difficulties. First, the accuracy of the ASR is lower, which means that the hypothesis will be more different from the speech content and thus more different from the loose transcription. In this case, the alignment is harder to achieve. Second, the available resources may be highly heterogeneous either from the speech coding or from the text formatting point of view. Third, in many cases only a part of the speech content is transcribed. Under these conditions, the problem consists in *spotting and aligning* a shorter text within a larger similar text. DTW based techniques sometimes fail to achieve this because of the word sequences that are repeated in the texts.

B. The Proposed Method

In this approach, we build a matrix whose elements $m(i,j)$ are 1 if the i^{th} word of the ASR hypothesis is identical to the j^{th} word of the loose transcription and 0 otherwise. Further on, we regard this matrix as a black and white image. If we build such a matrix to align one text to itself, we will obtain an image similar to Fig. 1. The matrix will be *a square one* (NxN, in this example N=60) and its diagonal represents the perfect match of the text's words. The other dots represent the repetitive words in the text. In the general case of aligning two different texts, if the two texts that are going to be aligned are very similar, then the diagonal line will be obvious and it will be easy to spot within the image.

The data collected for under-resourced languages are far from the ideal case. Fig. 2 illustrates the matrix (NxM, in this example N=1700 and M=4200) for the alignment of an ASR hypothesis and a loose transcription. In this image, the diagonal is not obvious because it is "lost" among the high number of dots (representing repetitive words). Fig. 3 shows a zoomed section from Fig. 2 (the white box). It can be observed that the diagonal line is not continuous as it was in Fig.1. This happens because the ASR hypothesis and the loose transcription are not identical. The purpose of our detection method is to make the diagonal line more obvious by using image processing techniques.

The processing steps are the following:

1. ***Repetitive words removal***. The repetitive words whose number exceeds a threshold over the lines, are removed (the whole line is reset to 0). It is possible that these words also occur in the diagonal line, but they do not contribute in emphasizing the diagonal line. On the contrary, they make it less obvious. This condition is expressed mathematically by formula (1):

$$\forall i = \overline{1,N} \quad S_i = \sum_{j=1}^{M} m(i,j) > th_{occ} \quad \Rightarrow \quad \forall j \quad m(i,j) = 0 \qquad (1)$$

where $m(i,j)$ represents the element from the row i and column j of the matrix, N is the number of lines, M is the number of columns, S_i is the number of occurrences of the i^{th} word from the ASR hypothesis in the loose transcription, and th_{occ} is the threshold. The result of this processing step is shown in Fig. 4.

2. ***Resolution reduction***. The resolution reduction is made by dividing the matrix into disjunct FxF squares and by calculating the density of each square as shown by equation (2):

$$d(k,l) = \sum_{i=Fk+1}^{F(k+1)} \sum_{j=Fl+1}^{F(l+1)} m(i,j) \qquad (2)$$

 ISBN: 978-1-61208-262-2

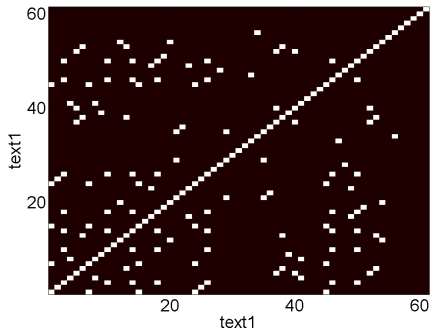


Figure 1. The matrix when aligning one text to itself

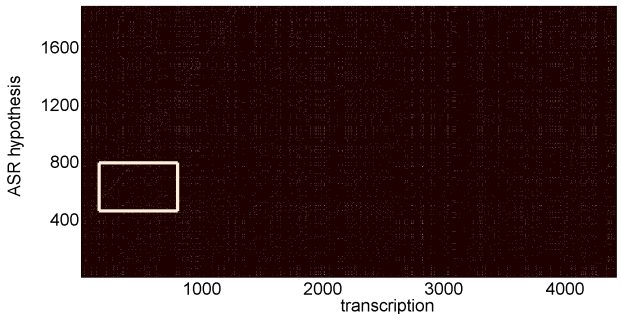


Figure 2. The full matching matrix

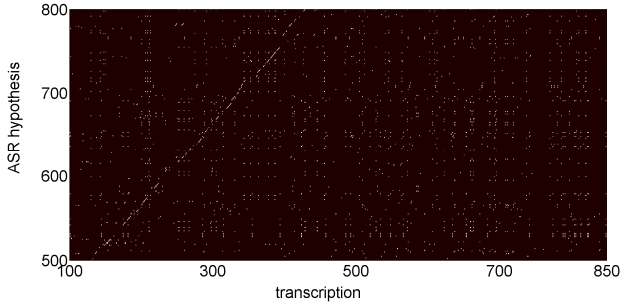


Figure 3. A zoomed region of the matching matrix

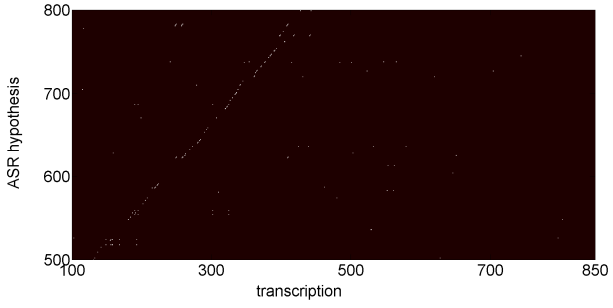


Figure 4. A zoomed region of the matrix after step 1

where $d(k,l)$ is the element of the new matrix d, while F is the factor of resolution reduction. The choice of F has an important impact on the spotting accuracy and its effect is shown by the experimental results (see Section III).

This processing step has a double role. First, it emphasizes the diagonal line because it is expected that the rest of the left dots are spread in an irregular way over the image (see Fig. 5). Second, thanks to the lower resolution, the computational load will be lower in the following steps.

3. ***Wiener filtering***. A 2D Wiener filtering is a low pass filter that reduces the additive noise power. The removal of the repetitive words at step 1 reduces the correlation among dots making their distribution more noisy. Without step 1 the Wiener filtering will not be successful. The filtering is performed pixel-wise according to equation (3) [10]:

$$\hat{d}(k_1,k_2) = \mu + \frac{\sigma^2 - \upsilon^2}{\sigma^2}(d(k_1,k_2) - \mu) \quad (3)$$

where:

$$\mu = \frac{1}{K^2}\sum_{l_1-K/2}^{l_1+K/2}\sum_{l_2-K/2}^{l_2+K/2} d(l_1,l_2) \quad (4)$$

$$\sigma^2 = \frac{1}{K^2}\sum_{l_1-K/2}^{l_1+K/2}\sum_{l_2-K/2}^{l_2+K/2} d(l_1,l_2) - \mu^2 \quad (5)$$

The mean (μ) and the variance (σ^2) are calculated within a square KxK matrix in the neighborhood of each pixel. υ^2 is the average of all variances. K determines how much the dots or the contours are faded, thus it must be chosen carefully.

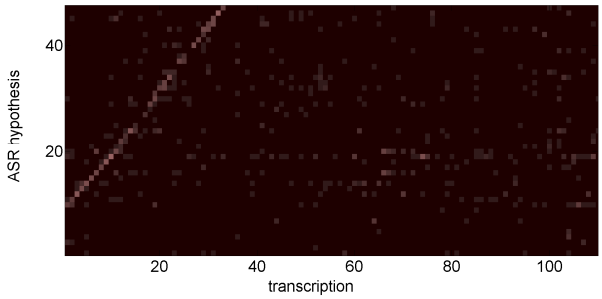


Figure 5. The full matrix after resolution reduction

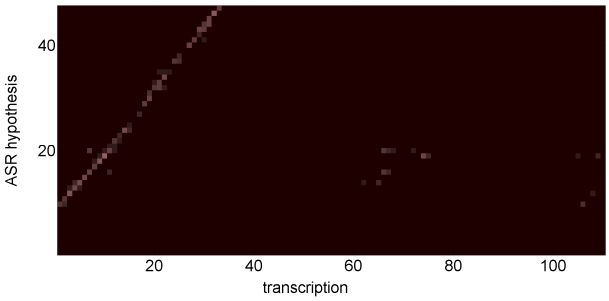


Figure 6. The matrix after Wiener filtering (step 3)

 ISBN: 978-1-61208-262-2

The experimental results in Section III show how the method efficiency depends on K. The result of this processing step is shown in Fig. 6.

4. ***Rough alignment***. At this step, we spot the ASR hypothesis within the loose transcription using intercorrelation. We do this by calculating the maximum values over the lines (max_l) and over the columns (max_c). It is expected that vectors max_l and max_c are highly correlated. The index for which the cross-correlation of the two vectors obtains the maximum value, provides the offset of the alignment for the ASR hypothesis and loose transcription.

C. Fine Alignment and Confidence Scoring

After the rough alignment, the word alignment is made by using a standard DTW technique that is used also for the calculation of Word Error Rate (WER). After the words are aligned a confidence score is needed in order to determine whether the matched words are indeed uttered in the speech content. Because the performance of the ASR system is not high we cannot count on the output likelihood. Instead, we make the assumption that longer matched words and longer matched word sequences are more probable to represent the content of the speech recording. The acceptance rule for the matched words is the following: words longer than m characters and sequences longer than n words are considered correct matches. Thresholds m and n strongly depend on the performance of the ASR system and they are determined empirically (see Section III).

III. EXPERIMENTAL RESULTS

A. Experimental Setup

Romanian language is considered to be an under-resourced language because there are no annotated speech resources available, neither free nor paid ones. Hence, the database used for lightly supervised acoustic modeling is collected from Romanian news websites. However, the collected speech recordings are not transcribed entirely. In other cases, the transcription of a short speech recording is found within a larger text. In both cases, segmentation is needed before word alignment. The accuracy of the transcriptions is not high. There is a WER of 30% compared to the exact speech content. The errors consist in several tags, special number formats, re-phrasing of speech content, misspellings, etc. The database used for experiments contains 100 hours of broadcast news.

The experiments are made with an ASR system in the Romanian language [11], [12]. The acoustic model is trained with 70h of read speech from 80 speakers. It is based on 3-state Hidden Markov Models, while the number of senones is 4000. Each state's output is modeled by a Gaussian Mixture Model with 64 densities. The vector of speech features has 39 elements (13 Mel Frequency Cepstral Coeficients + Δ + $\Delta\Delta$). The language model is 3-gram based and it is trained with 170 million of words. The ASR system obtains 20% WER for read clean speech and 59% WER for broadcast news.

B. Refinement of Method's Parameters

The purpose of these experiments is to determine the optimal values for the tuning parameters described in Section II. These parameters are the threshold of removal (th_{occ}) for the occurrences of repetitive words, the resolution reduction factor (F) and the size of the neighborhood in Wiener filtering (K). By varying these parameters a huge number of combinations can be obtained. Because the segmentation algorithm is time consuming (20h of run for 100h of speech), we have decided to maintain two parameters fixed while varying the third. The metric used for the accuracy of the segmentation method is the number of hours of the extracted annotated speech.

Table I shows how the method's efficiency varies with th_{occ}. The results show that there is an optimal value for th_{occ}. Indeed, th_{occ} cannot be too high, otherwise the remaining dots from step 1 will not have a noisy distribution, but rather a well defined pattern. In this case, the Wiener filtering is not efficient. On the other hand, th_{occ} cannot be too low either because words from the diagonal line will also be removed making it more subtle.

The effect of parameter F over the method's performance is shown in Table II. The results confirm that higher values of F make the diagonal line more obvious and easier to spot leading to a larger amount of extracted speech. This happens because the density of the dots along the diagonal line is higher than the density of the other areas that contain sporadic dots. However, F cannot be too high because for some transcriptions the number of words is smaller and this imposes a limit on the resolution reduction.

The effect of parameter K over the method's performance is shown in Table III. A higher K leads to a greater fading of the dots and a greater smoothening of the contours (the diagonal line). This contributes in making the diagonal line more obvious among the dots. For K>5 no further improvement is obtained. For experiments, we will choose K=5 as the optimal value because higher values for K involves more computations.

C. Annotated Speech Extraction

For the rough alignment of the ASR hypotheses and the loose transcriptions we have used the optimal values determined in Section III.B: th_{occ}=3, F=40, K=5. Once the texts are aligned, word alignment is performed based on a DTW based technique. In order to evaluate the performance of the system for several values of m and n described in Section II.C, we have verified in each case 10 minutes of annotated speech resulted from word alignment process. The results are shown in Table IV. The number in each cell indicates the percentage of speech segments that are erroneously annotated. Errors in speech annotation lead to inaccurate acoustic models. Hence, m and n are chosen so that no errors occur in the automatic annotation process (m=8 and n=4).

 ISBN: 978-1-61208-262-2

TABLE I. THE EFFECT OF TH_{OCC} IN SPEECH EXTRACTION

th_{occ}	***F*=20, *K*=3**
	No. extracted speech hours
2	4,2
3	6,2
4	6,1
5	5,9
6	5,7
7	5,2

TABLE II. THE EFFECT OF *F* IN SPEECH EXTRACTION

F	th_{occ} **=3, *K*=3**
	No. extracted speech hours
10	5,5
20	5,7
30	5,9
40	6
50	5,9

TABLE III. THE EFFECT OF *K* IN SPEECH EXTRACTION

K	th_{occ} **=3, *F*=40**
	No. extracted speech hours
2	5,9
3	6
4	6,1
5	6,2
6	6,2
7	6,2

TABLE IV. THE ERROR PERCENTAGE FOR DIFFERENT EXTRACTION RULES

n \ ***m***	***5***	***6***	***7***	***8***
2	3.0%	2.9%	2,1%	1.1%
3	2.0%	1.7%	1.5%	1.4%
4	0.9%	0.7%	0.3%	0.0%
5	0.5%	0.4%	0.3%	0.0%

IV. CONCLUSIONS

Lightly supervised acoustic modeling is a cost-effective method of obtaining annotated speech data. However, traditional DTW-based techniques sometimes fail to align large texts. The proposed method overcomes this issue by introducing a segmentation step based on image processing. Its purpose is to emphasize a diagonal line, which represents the correctly matched words, among other dots that represent the repetitive words.

The method is more efficient if the dots have a noisy distribution, thus the repetitive words whose occurrence number exceeds a threshold are removed. The experimental results also proved that there exist optimal values for other two tuning parameters: the resolution reduction factor and the size of the Wiener filtering neighborhood. By these refinements, the method outperforms the traditional DTW based aligning methods.

REFERENCES

[1] A. Buzo, H. Cucu, and C. Burileanu, "Text spotting in large speech databases for under-resourced languages," ICASSP: Proceedings of the Acoustics, Speech, and Signal Processing. Canada, *submitted paper*, 2013.

[2] V-B. Le and L. Besacier. "First steps in fast acoustic modeling for a new target language. Application to Vietnamese," ICASSP: Proceedings of the Acoustics, Speech, and Signal Processing. USA, pp. 821 - 824, 2005.

[3] D. Yu, L. Deng, P. Liu, J. Wu, Y. Gong, and A. Acero, "Cross-lingual speech recognition under runtime resource constraints," ICASSP: Proceedings of the Acoustics, Speech, and Signal Processing, Taiwan, pp. 4193-4196, 2009.

[4] P. J. Moreno, C. Joerg, J.-M. Van Thong, and O. Glickman, "A recursive algorithm for the forced alignment of very long audio segments," in Proceedings of ICSLP, USA, pp. 2711-2714, 1998.

[5] L. Lamel, J. Gauvain, and G. Adda, "Investigating Lightly Supervised Acoustic Model Training," ICASSP: Proceedings of the Acoustics, Speech, and Signal Processing, USA, pp. 189-192 vol. 1, 2001.

[6] H. Y. Chan and P. Woodland, "Improving broadcast news transcription by lightly supervised discriminative training," ICASSP: Proceedings of the Acoustics, Speech, and Signal Processing, Canada, pp. 737-740 vol.1, 2004.

[7] T. J. Hazen, "Automatic alignment and error correction of human generated transcripts for long speech recordings," Proceedings of Interspeech, USA, pp. 1606-1609, 2006.

[8] M. H. Davel, C. Van Heerden, N. Kleynhans, and E. Barnard, "Efficient harvesting of Internet audio for resource-scarce ASR," Proceedings of Interspeech, Italy, pp. 3154-3157, 2011.

[9] B. Lecouteux, G. Linares, J.F. Bonastre, and P. Nocera, "Imperfect transcript driven speech recognition," Proceedings of Interspeech, USA, pp. 1626-1629, 2006.

[10] Lim, Jae S., "Two-Dimensional Signal and Image Processing", Englewood Cliffs, NJ, Prentice Hall, 1990, pp. 536-540.

[11] H. Cucu, L. Besacier, A. Buzo, and C. Burileanu, "Enhancing Automatic Speech Recognition for Romanian by Using Machine Translated and Web-based Text Corpora," Proceedings of SPECOM 2011, Rusia pp. 81-88, 2011.

[12] H. Cucu, L. Besacier, C. Burileanu, and A. Buzo, "Investigating the Role of Machine Translated Text in ASR Domain Adaptation: Unsupervised and Semi-supervised Methods," in the Proceedings of the 2011 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), USA, pp. 260-265, 2011, ISBN: 978-1-4673-0366-8.

 ISBN: 978-1-61208-262-2

New Considerations for Accumulated ρ-Cross Power Spectrum Phase with Coherence Time Delay Estimation

Radu-Sebastian Marinescu[*,#], Andi Buzo[*], Horia Cucu[*], Corneliu Burileanu[*]
[#]Reasearch & Development, Rohde & Schwarz Topex
[*]SpeeD Laboratory, University Politehnica of Bucharest
Bucharest, Romania
radu-sebastian.marinescu@rohde-schwarz.com, {andi.buzo, horia.cucu}@upb.ro, cburileanu@messnet.pub.ro

***Abstract*—Time delay estimation (TDE) remains an important research issue because of its several approaches and large field of digital signal applications. As a solution for this topic, in this paper, we continue the evaluation of the recently proposed *accumulated ρ-cross power spectrum with coherence* TDE method. The experimental results confirm that the method is faster and more accurate than the previous separated variants. Another key finding is that the TDE based on accumulation of cross-power spectrum is at least twice as accurate as the TDE based on time domain averaging.**

Keywords-Time Delay Estimation; Accumulated ρ-Cross Power Spectrum with Coherence

I. INTRODUCTION

As technology evolved, more and more applications demanded a solution for time delay estimation. For echo canceling, acoustics, radar and sonar localization, seismic and medical processing, pattern detection and speech enhancement, scientists are still looking for better solutions. The variety of time delay estimation (TDE) applications, implementation aspects and proper constraints inhibit the design of a unique solution. Instead, various approaches have been developed based on application specific aspects.

The numerous proposed methods are based mainly on the *generalized cross-correlation* (GCC), *least mean square* (LMS) adaptive filtering and *adaptive eigenvalue decomposition* (EVD). Each category has its advantages making it optimal for specific applications. The large family of adaptive filtering methods [1-7] achieves very high accuracy, but, despite the variety of optimized variants, the adaptation time it is too long in some applications. A faster solution, proven to be efficient in audio applications from reverberant environment, is represented by EVD [8].

But, the most popular TDE methods, which do not need any adaptation time, are based on the generalized cross-correlation, initially proposed in 1976 by Knapp and Carter [9]. They have also presented a particular GCC weighting function named Cps-m. Based on this work, gradually, over time, multiple variations of the GCC weighting function were proposed: ROTH and SCOT [10], Eckart [11], Phase Transform (PHAT) or Cross-power Spectrum Phase (CSP) [12], [13], Wiener [14], HT (ML) [15], *accumulated CSP* (*acc-CSP*) [16], *ρ-CSPC* [17], HB [18]. For the majority of them, a review and a comparison, based on the root mean square deviation of the estimated delay and mean value, were presented in [19]. In [20], we proposed two new methods: *acc-ρCSPC* and *acc-ρCSP*, which benefit from the higher accuracy, which characterized the *ρ-CSPC* [17] and the lower computational load and robustness of the *acc-CSP* method [16]. *Acc-ρCSPC* and *acc-ρCSP* outperform previous methods in computation time, because of the accumulation of cross-power spectrum phase in frequency domain. This leads to only one Inverse Discrete Fourier Transform (IDFT) for any number of accumulated frames used. Also, in [20], it is shown that the first method *(acc-ρCSPC)* generally has a higher accuracy than *(acc-ρCSP),* but, in specific conditions, the second method achieves practically the same accuracy as the *acc-ρCSPC* at a lower computational load.

In this work, we continue to evaluate *acc-ρCSPC* over previous methods. We show that, for multiple frames estimations of the time delay, results based on accumulating cross-power spectrum in frequency have, in general, at least twice the accuracy compared to the normal results obtained by time averaging.

This paper is organized as follows. The presentation of the TDE problem and recently proposed solutions are included in Section II. In Section III, we provide the experiments and discussion about the results. Finally, the conclusions are reserved for Section IV.

II. TIME DELAY ESTIMATION AND EVALUATED METHOD

In several applications, we are confronted with two (or sometimes more) signals, $y_1(t)$ and $y_2(t)$, delayed and noisy versions of the same source signal $x(t)$. The time delay estimation tries to find the relative delay between these signals. Over the years, a large variety of approaches was proposed for TDE, but the most popular methods are based on the cross-correlation between the two signals.

In 1976, Knapp and Carter introduced, in [9], the so-called generalized cross-correlation (GCC), which adds a filtering function:

$$R^{g}_{y_1y_2}(t) = \int_{-\infty}^{\infty} \Psi(f) \cdot G_{y_1y_2}(f) \cdot e^{j2\pi ft} df \qquad (1)$$

 ISBN: 978-1-61208-262-2

where $\Psi(f)$ represents a general frequency weighting function and Gy_1y_2 is the cross-power spectrum. The introduction of the weighting function takes advantage of some characteristics of the source and noise, emphasizing different spectral information [13]. Thus, the value that maximizes the general cross-correlation function represents the estimated time delay.

A. TDE based on Cross-Power Spectrum Phase

A popular derivation of GCC is represented by the CSP. This method does not require any a priori knowledge of noise or source, making this approach independent of the input waveform characteristics, unless signals are strictly narrowband [13]. It has a large area of applications and it was shown to be an efficient technique for time delay estimation [12] [13] [16].

The weighting function Ψ_g for CSP is computed as follows:

$$\Psi(f) = 1/\left|G_{y_1y_2}(f)\right| \tag{2}$$

The TDE with CSP uses an analysis window, which is usually split in several smaller frames. Then the CSP is computed for every frame. The final result is then calculated as the average of all frame estimates. This last operation is done in time domain. Due to the fact that response time is important in almost all TDE applications, the focus should be on two factors: processing time and window length. The use of a larger frame leads to a higher accuracy rate for correct CSP estimation, but the downside is the increasing computing time.

As a solution to the above circumstances, the *acc-CSP* method was proposed by Matassoni and Svaizer [16]. The importance of the method is that it estimates time delay by averaging CSP over all frames, in the frequency domain. This way, it is shown that the processing time decreases. This is explained, because the cross-power spectrum phase is accumulated over multiple frames, remaining only one Inverse Fast Fourier Transform (IFFT) to be computed after the last accumulation. In frequency domain it can be expressed as follows:

$$G_{acc-CSP}(f) = \sum_{k=1}^{K} \frac{G_{y_1y_2,k}(f)}{\left|G_{y_1y_2,k}(f)\right|}, \tag{3}$$

where K represents the number of accumulated frames.

The *acc-CSP* method proposes the *accumulation scheme* of cross power spectrum in frequency domain, increasing the computation speed. The previous methods compute the TDE as the average of all partial estimated delays of each frame from the analysis window. This way, for K frames, the number of total Fast Fourier Transform (FFT) operation is equal to 3xK, because two FFT are used to transform the signals from time to frequency domain, and then one IFFT is used on the cross power spectrum to return in the time domain, for each frame. Instead, the accumulation scheme is faster because it does not calculate any partial TDEs. Because the cross-power spectrum averaging is computed in frequency domain, only one estimate will result, for any number of frames. This way, only one IFFT is needed for the final estimation and 2xK FFTs for time to frequency transformations. This leads to a total number of 2xK + 1 FFT for the *accCSP* method, which is less than the 3xK FFT needed by previous methods.

Beside the reduced computational complexity, the *acc-CSP* method enhances the estimation by intrinsic integration for fixed delay during the analysis window.

In [16], it was also shown that the computing time decreases for *acc-CSP* compared with CSP, at the cost of accuracy degradation. Separate from this, an accuracy improvement for the CSP method was proposed by Shen and Liu [17], with a modified GCC weighting function $\Psi(f)$ as in the following expression:

$$\Psi_{y_1y_2}(f) = \frac{1}{\left|G_{y_1y_2}(f)\right|^{\rho} + \min\left[\gamma^2_{y_1y_2}(f)\right]} \tag{4}$$

where $\gamma^2_{y_1y_2}(f)$ is the signal's coherence function:

$$\gamma^2_{y_1y_2}(f) = \frac{\left|G_{y_1y_2}(f)\right|^2}{G_{y_1}(f)\cdot G_{y_2}(f)} \tag{5}$$

The tuning parameter ρ (with values between 0 and 1) is a whitening parameter, which discards the non-speech portion of the signals (below 200Hz) [17][21]. To reduce errors for relatively small energy signals, the minimum of the coherence function was added in (5).

B. Accumulated ρ-Cross-Power Spectrum Phase Methods

Combining *acc-CSP* and *ρ-CSPC* methods, we proposed the new *accumulated ρ-Cross Power Spectrum Phase with Coherence* (*acc-ρCSPC*) in [20], defined as follows:

$$G_{acc-\rho CSPC}(f) = \sum_{k=1}^{K} \frac{G_{y_1y_2,k}(f)}{\left|G_{y_1y_2,k}(f)\right|^{\rho} + \min\left[\gamma^2_{y_1y_2,k}(f)\right]} \tag{6}$$

This way, it is possible to take advantage of both methods. Its effectiveness was proven by experimental results from [20], which showed a better accuracy even for low signal-to-noise ratios (SNR).

The new approach, summarized by (6), leads to faster computations compared to previous methods. This is due to the fact that it uses the accumulating scheme, which can also provide better results in unfavorable conditions for smaller frame sizes. Beside this, emphasis of speech regions from the spectrum is achieved by the whitening parameter (ρ), which reduces, at the same time, the impact of noise outside

 ISBN: 978-1-61208-262-2

the speech region. For parts of the signal with small energy, the addition of the minimum coherence function limits the effect of a very small denominator.

For a faster computation for applications where relatively small energy signals are not encountered, the minimum coherence function can be omitted from (6). This leads to *accumulated ρ-Cross Power Spectrum Phase* (*acc-ρCSP*) proposed in [20], as follows:

$$G_{acc-\rho CSP}(f) = \sum_{k=1}^{K} \frac{G_{y_1 y_2,k}(f)}{\left|G_{y_1 y_2,k}(f)\right|^{\rho}} \quad (7)$$

This method is faster than *acc-ρCSPC* because of the omission of the coherence calculation in equation (7). The experimental results from [20] confirm the utility of this method for proper ρ and above conditions.

III. Experimental Results and Discussions

A. Experimental Setup

In this work, we present a further evaluation and discussions for our recently proposed *acc-ρCSPC* method in [20]. The *acc-ρCSPC* and previous *CSP*, *accCSP* and *ρCSPC* methods were implemented in Matlab. The Noizeus [22] database corpus was used as the main database experiments. It contains 30 sentences (produced by three male and female speakers at 8 kHz) corrupted by 8 different real-world noises (suburban train, babble, car, exhibition hall, restaurant, street, airport and train station noise), from the AURORA database [23] at 4 different SNRs (0, 5, 10 and 15 dB).

All possible combinations of noise types were used, which results in a number of $C_2^8 = 28$ variants and all 4 different SNR levels. The whitening parameter ρ was set to 0.73 after the calibration as in [20]. The frame overlap factor was chosen 25%.

The metric used in these experiments is the accuracy, which is defined as the ratio between the number of perfectly estimated delays and the total number of estimations performed. This metric is the most relevant one for applications where the exact delay estimation is required.

B. Further evaluations for acc-ρCSPC method

In Table I, we show the accuracy improvement resulting from the combination of *CSP* and *ρCSPC* methods. The frame size was set to 1024 samples and we artificially introduced 5 delay values (5, 10, 25, 50 and 100 ms). Taking into account the combinations described in the previous subsection, the total number of test pairs was 28x30x4x5=16800. The evaluation was performed for *acc-ρCSPC* and previous *CSP* and *ρCSPC* methods, for 4 and 8 frames.

TABLE I. ACCURACY COMPARISON

No. Frames	Estimated scheme	Method accuracy [%]		
		CSP	*ρCSPC*	*acc-ρCSPC*
4	Average in time	23.0	34.0	N/A
	Accumulation scheme	N/A	N/A	92.8
8	Average in time	12.5	24.2	N/A
	Accumulation scheme	N/A	N/A	99.9

The N/A cells from Table I correspond to the following cases. Firstly, for the proposed *acc-ρCSPC* method the accumulation of the cross power spectrum over multiple frames is used, so accuracy results are not available for a final averaging in time domain. Secondly, *CSP* and *ρCSPC* use the averaging in time and, for them, the accumulation scheme cannot be used.

Table I shows that *acc-ρCSPC* outperforms the previous methods on the accuracy tests. We observe that the accuracy for *acc-ρCSPC* increases with the number of frames, while for the others it decreases. The differences between *averaging in time* and *accumulated cross power spectrum in frequency domain* result from the fact that, in frequency domain, the accumulation keeps the spectral information over multiple frames. This way, it is maintaining correlation between the frames. On the other hand, the accuracy of averaging in time domain decreases if more frames are used. This is explained because, with the increasing number of frames, the probability of a false estimation is also increasing.

The dependency of *acc-ρCSPC* accuracy over the SNR levels and delay variations is presented in Fig. 1. For this experiment, the frame size was set to 512 samples, resulting in a frame of 64 ms for an 8 kHz sampling frequency. The average accuracy was computed for all 16800 sentences combinations.

Fig. 1 shows that the method can achieve a high accuracy rate of more than 90%, even for delays of 78% of the frame size (50 ms delay for a frame size of 64 ms) at 15 dB SNR. This is an important aspect because most of the GCC methods provide reasonable results for delays up to 60-70% of the frame size.

Also, Fig. 1 shows that, for delays longer than 50% of the frame size, the influence of the SNR level affects much more the method's accuracy. For delays up to 50% of the frame size, the difference between accuracies on various levels of SNR remains almost the same.

A comparison between proposed *acc-ρCSPC* and previous *acc-CSP* methods is presented in Fig. 2. For this evaluation the same configuration as for the above experiment was used.

 ISBN: 978-1-61208-262-2

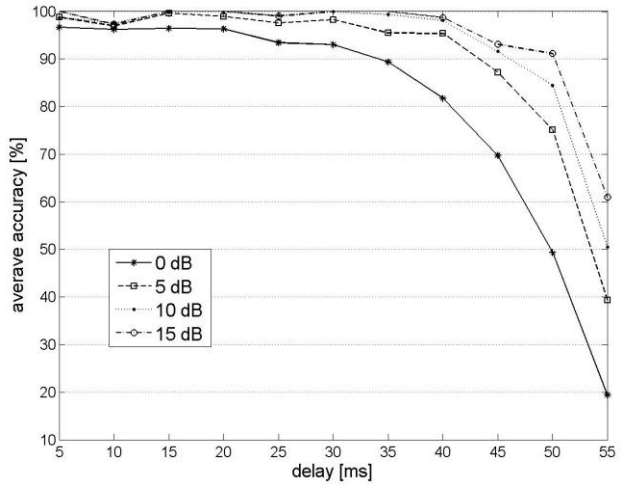


Figure 1. The influence of SNR and delay over the *acc-ρCSPC* accuracy

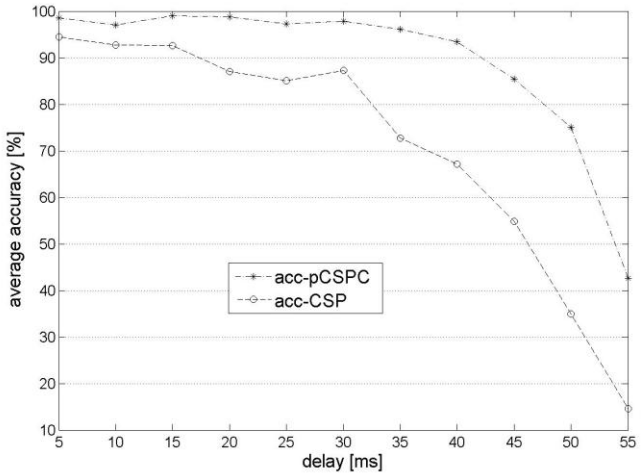


Figure 2. Comparison between acc-CSP and *acc-ρCSPC*

The results from Fig. 2 confirm the effectiveness of the *acc-ρCSPC* method. The important difference between the methods is due to the proposed mixture of previous *ρCSPC* and *acc-CSP* techniques. In Fig. 2 we observe that, for a delay of 78% from the frame size (50/64 ms), the average accuracy of *acc-ρCSPC* is twice as for *acc-CSP*.

The nonmonotonic characteristics from Fig. 1 and Fig. 2 are explained by the use of the limited database. For larger databases with much more signal combinations, we expect the accuracy characteristic to change to a monotonic shape. But, even with the actual obtained characteristics, it is easy to conclude about the performance of the methods.

The results from this work propose the acc-*ρCSPC* for TDE applications where the accuracy of estimation and the response time are important demands.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we continued the evaluation of the proposed TDE method, *accumulated ρ-Cross Power Spectrum Phase with Coherence* (*acc-ρCSPC*). The experiments were performed on the standard *Noizeus* database. The obtained results showed that the new combination of *acc-ρCSPC*, based on previous *acc-CSP* and *ρCSPC*, offers higher accuracy rate and faster computational speed.

The accumulating cross power spectrum phase, which is performed in the frequency domain, leads to an accuracy that is more than twice than the accuracy obtained with the previous methods, which average the final results in the time domain. Also, the accuracy of the *acc-ρCSPC* increases with the number of frames, while for previous methods that use the average in time it decreases as the number of frames grows.

The *acc-ρCSPC* outperforms other methods, with an accuracy rate over 90%, even for delays that are longer than 75% of the frame size. Its' accuracy remains almost stable by the SNR variations for delays which are smaller than 50% of the frame size.

The results from this work propose the *acc-ρCSPC* in TDE applications where the accuracy of estimation and the response time are important demands. It can be efficiently implemented to provide solution for realigning noisy signals in applications such as speech enhancement, echo canceling, seismic and medical processing, radar and sonar localization, and pattern detection.

Future work will involve development of *acc-ρCSPC* and acc-*ρCSP* applications for the VoIP environment. Specific analysis will also involve methods characterization for different system implementation.

REFERENCES

[1] B. Widrow and S.D. Strearns, "Adaptive signal processing", Penitence-Hall, ISBN 0130040290, USA, 1985.

[2] S.N. Lin and S.J. Chern, "A new adaptive constrained LMS time delay estimation algorithm", Signal Processing, Volume 71, Issue 1, Nov. 1998, pp. 29-44.

[3] A.W.H. Khong and P.A. Naylor, "Efficient Use Of Sparse Adaptive Filters", In Proceedings of ACSSC '06, Fortieth Asilomar Conference on Signals, Systems and Computers, article ID 10.1109/ACSSC.2006.354982, Nov. 2006, pp. 1375-1379.

[4] R.A. Dyba, "Parallel Structures for Fast Estimation of Echo Path Pure Delay and Their Applications to Sparse Echo Cancellers", In Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems, article ID 10.1109/CISS.2008.4558529, Mar. 2008, pp. 241-245.

[5] D. Hongyang and R.A. Dyba, "Efficient Partial Update Algorithm Based on Coefficient Block for Sparse Impulse Response Identification", In Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems, article ID 10.1109/CISS.2008.4558527, Mar. 2008, pp. 233-236.

[6] D. Hongyang and R.A. Dyba, "Partial Update PNLMS Algorithm for Network Echo Cancellation", In Proceedings of ICASSP 2009, IEEE International Conference on Acoustics, Speech and Signal Processing, article ID 10.1109/ICASSP.2009.4959837, Apr. 2009, pp. 1329-1332.

[7] K. Sakhnov, E. Verteletskaya, and B. Simak, "Partial Update Algorithms and Echo Delay Estimation," Communications – Scientific Journal of the University of Zilina, Zilina – Slovakia, vol. 13, no. 2, Apr. 2011, pp. 14-19.

 ISBN: 978-1-61208-262-2

[8] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization", J. Acoust. Soc. Am. Volume 107, Issue 1, Jan. 2000, pp. 384-391.

[9] C. Knapp and G.C. Carter, "The Generalized Correlation Method for Estimation of Time Delay", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 24, issue 4, Aug. 1976, pp. 320-327.

[10] D.H. Youn, N. Ahmed, and G.C. Carter, "On the Roth and SCOTH Algorithms: Time-Domain Implementations", In Proceedings of the IEEE, vol. 71, issue 4, 1983, pp. 536-538.

[11] Q. Tianshuang and W. Hongyu, "An Eckart-weighted adaptive time delay estimation method", IEEE Transactions on Signal Processing, vol. 44, issue 9, Sep. 1996, pp. 2332-2335.

[12] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique", Proceedings of ICASSP, Australia, Apr. 1994, pp. 273-276.

[13] M. Omologo and P.Svaizer, "Use of the crosspower-spectrum phase in acoustic event location", IEEE Transactions on Speech Audio Process, May 1997, pp. 288-292.

[14] V. Zetterberg, M.I. Pettersson, and I. Claesson, "Comparison Between Whitened Generalized Crosscorrelation and Adaptive Filter for Time Delay Estimation", In Proceedings of TS/IEEE, OCEANS, vol. 3, article ID 10.1109/OCEANS.2005.1640117, Sep. 2005, pp. 2356 – 2361.

[15] K.W. Wilson and T. Darrell, "Learning a Precedence Effect-Like Weighting Function for the Generalized Cross-Correlation Framework", IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, issue 6, Nov. 2006, pp. 2156-2164.

[16] M. Matassoni and P. Svaizer, "Efficient Time Delay Estimation Based on Cross-Power Spectrum Phase", European Signal Processing Conference (EUSIPCO), Florence - Italy, Sep 2006.

[17] M. Shean and H. Liu, "A Modified Cross Power-Spectrum Phase Method Based on Microphone Array for Acoustic Source Localization," IEEE International Conference on System, Man and Cybernetics, San Antonio, TX, USA, Oct. 2009, pp. 1286 – 1291.

[18] Y. Sun and T. Qiu, "The SCOT Weighted Adaptive Time Delay Estimation Algorithm Based on Minimum Dispersion Criterion", In Proceedings of the ICICIP Conference on Intelligent Control and Information Processing, Aug. 2010, pp. 35-38.

[19] K. Sakhnov, E. Verteletskaya, and B. Simak, "Echo Delay Estimation Using Algorithms Based on Cross-correlation," Journal of Convergence Information Technology, Volume 6, Number 4, Apr. 2011, pp. 1 – 11.

[20] R.S. Marinescu, A. Buzo, H. Cucu, and C. Burileanu, Fast Accurate Time Delay Estimation Based on Enhanced Accumulated Cross-Power Spectrum Phase, ICASSP 2013, "unpublished"

[21] D.V. Rabinkin, R.J. Renomeron, A. Dahl, J.C. French, J.L. Flanagan, and M.H. Bianchi, "A DSP Implementation of Source Location Using Microphone Arrays", The Journal of the Acoustical Society of America, Volume 99, Issue 4, Apr. 1996, pp. 2510-2527.

[22] NOIZEUS: A noisy speech corpus http://www.utdallas.edu/~loizou/speech/noizeus/ [retrieved: Feb, 2013]

[23] AURORA database, http://www.elda.org/article52.html [retrieved: Feb, 2013]

 ISBN: 978-1-61208-262-2

Real-Time Modeling in Pervasive Mining

Jose Perez
Electrical Engineering Department
Universidad de Santiago de Chile
Santiago de Chile, Chile
e-mail: jose.perezb@usach.cl

Miguel Alfaro
Industrial Engineering Department
Universidad de Santiago de Chile
Santiago de Chile, Chile
e-mail: miguel.alfaro@usach.cl

Ismael Soto
Electrical Engineering Department
Universidad de Santiago de Chile
Santiago de Chile, Chile
e-mail: ismael.soto@usach.cl

Francisco Cubillos
Chemical Engineering Department
Universidad de Santiago de Chile
Santiago de Chile, Chile
e-mail: francisco.cubillos@usach.cl

Pei Xiao
Centre for Communication Systems Research
University of Surrey, UK
Guilford, United Kingdom
e-mail: p.xiao@surrey.ac.uk

***Abstract*—This paper introduces an integrated design for real-time plant modeling at a copper mine, using Support Vector Machines (SVM), Neural Networks, and new technologies such as Machine-to-Machine (M2M) and Cloud Computing with an Android client. This solution was designed for a plant inside a copper mine which cannot tolerate interruption and for which their in situ modeling, in real time, is an essential part of the system to control aspects such as instability by adjusting their corresponding parameters without stopping the process.** *(Abstract)*

Keywords: mining; modeling; support vector machines; pervasive; machine-to-machine; cloud; neural networks (key words)

I. INTRODUCTION

Several dramatic accidents that have occurred in the mining sector have turned the spotlight on safety and the need for stricter supervision in mining operations. In this respect there is an urgent need to deploy new technologies in order to enhance the safety of miners and processes inside the mine. Several efforts have been made in this direction and the trend is to automate miner´s tasks as much as possible. This may be done by using new generation telecommunications [1] and computer technologies for accessing resources inside the mine. Enhancements of automation and remote operations, helped by mobile devices applications [2] will allow employees to supervise remotely the production tools and to have access to applications on servers outside the mine without leaving it.

In this work, we apply the concepts introduced above to solve the plant monitoring problem (dynamic system) for control and optimization purposes, using modeling tools. In particular, we use a virtual sensor model to estimate in real-time the value of a physical parameter which cannot be measured directly without stopping the process. The knowledge of this value on-site, in real-time, is indispensable to control aspects such as instability and optimization of the whole process. It will also contribute to enhance mine safety.

The proposed system includes the machine-to-machine (M2M) telecommunications platform, a server in the cloud that runs a modeling method (based on neural network, support vector machines, or other) and a mobile client inside the mine.

This is an ongoing project, and at this stage, it is still under development. Machine-to-machine communications must be supported by operators, but this issue has not been solved yet. We are currently working on the development of software interfaces for the mobile client, and on the server applications located in a private cloud.

The rest of the paper is organized as follows: Section II describes briefly the proposed general communications architecture. Section III introduces some terms and concepts related to the modeling of dynamic systems used in this work; Section IV presents a practical modeling application related to on-line/real-time estimation of states in a complex Semi-Autogenous Grinding process (SAG) in copper mining; and Section V concludes the paper.

 ISBN: 978-1-61208-262-2

II. SYSTEM DESCRIPTION

We present a design that integrates modeling tools and new communications technologies at a copper mine, in order to monitor, control and optimize a particular process. In order to deploy the solution, we have considered semi-autonomous grinding. The model will be useful to estimate, using a virtual sensor and in real time, the filling level parameter of the grinding, since it is very important to control and optimize the process.

Using new communications technologies, the plant operator as well as the supervisors may know the filling label value at the same time and anywhere inside or outside the mine. This brings up pervasive mining, a system with wider coverage, higher communication efficiency, better fault-tolerance, and anytime anywhere availability.

The proposed design considers M2M communications, a client-server application with the server in the cloud, and an Android client inside the mine (Fig. 1).

In order to employ real-time modeling, a prior suitable, liable plant model must be selected. The process for determining this model is performed as follows:

- A set of initial real data (e.g., 2500 samples) are sent directly from the plant towards the server in the cloud. Here, the data are divided into three data sets, to train (e.g., 500 samples), validate (e.g., 1000 samples), and test (e.g., 1000 samples) the candidate models, which the operator has defined previously.
- Tools that may be used for constructing the plant's models include Neural Networks, Support Vector Machines (SVM) or any other suitable tools.
- The server sends the calculated models to the operator, who can choose the best by comparing the forecasting errors. At the same time, the graphics and related results are sent to the monitoring screen, located anywhere inside or outside the mine, and where the whole process is controlled.
- Once the operator has chosen the best modeling tool (in our case a variation of support vector machines), it will be used for real-time modeling, in order to monitor, control and optimize the process.

A. M2M Communications and Architectures

Fig. 2 illustrates the high-level M2M communications architecture. In this scenario multiple connectivity options are available to serve machine-to-machine (M2M) applications requiring connectivity between end devices. The M2M client device can connect with the server (M2M server) directly through a WAN connection, M2M gateway, or an aggregation point [3]. In our design, our server modeling application will be installed in this server inside the cloud.

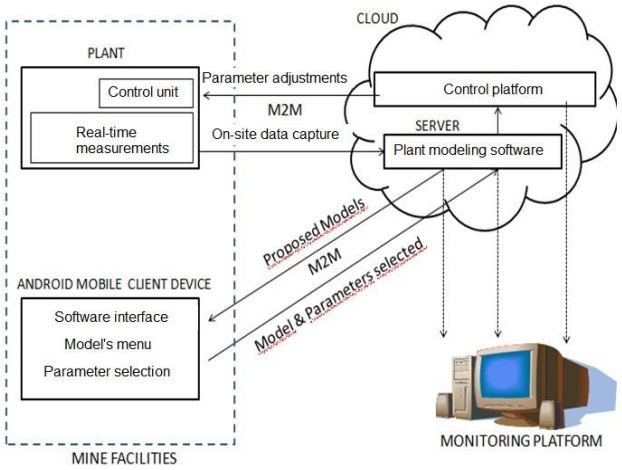


Figure 1. Modeling architecture

III. MODELING IN DYNAMIC SYSTEMS

A. Modeling tools

In order to get models for dynamic systems in production environments, several adaptive methods using approximation techniques have been developed. For instance, in [4], the authors used the Feed-Forward type Neural Networks (FFNN) to solve the problem of on-line identification of complex processes. As an important result, they encountered that FFNN converge satisfactorily in a few iteration cycles, showing better prediction capacity than recursive algorithms.

In [5] an observer model based on wavelet transform together with neural networks was successfully applied to solve the state observation problem of a dynamic system, when the dynamic model contains uncertainties or it is completely unknown.

In [6] the problem of on-line model identification for multivariable processes with nonlinear and time-varying dynamic characteristics has been solved using two online multivariable identification approaches with self-organizing neural network model structures. Two adaptive Radial Basis Function (RBF) neural networks have been defined, and the dynamic model is generated autonomously using the sequential input-output data pairs in real-time applications.

B. Model Performance

Once we had a set of the plant's real data, we divided it into three groups: to train, validate and test the model. However, when these goals are hampered by a lack of reliable experimental data it is necessary to construct a set of good quality, hypothetical data. In this sense an important tool to construct these data is the Lorentz model, which we used at the beginning of our work.

 ISBN: 978-1-61208-262-2

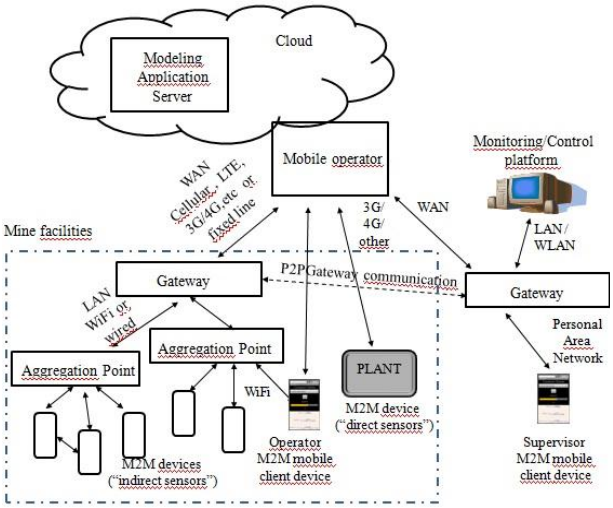


Figure 2. High-level M2M system architecture

This study will show a brief description of the Lorentz model that reveals the behavior of a dynamic system, and the related Forrester model which graphically represents a hypothetical plant that is being studied.

Data constructed from the Lorentz model are very valuable since they may represent the behavior of a plant inside a mine having chaotic outputs, and it is useful as a reference and for comparing purposes. If we have a deterministic, chaotic behavior, then reliable forecasting is possible and controlling the process in the mine becomes easier.

C. Influence Diagram

To study the behavior of dynamic systems, causal diagrams can be used to outline all the elements of a problem without going into the mathematical details in the possible model. An influence or causal diagram (also known as Forrester's diagram) represents influence relations that exist between the elements of a system, and therefore provides information about the structure.

As an example, Fig. 3 depicts the Forrester model corresponding to a plant whose behavior can be represented by the set of differential equations (1) related to the Lorentz model of a plant.

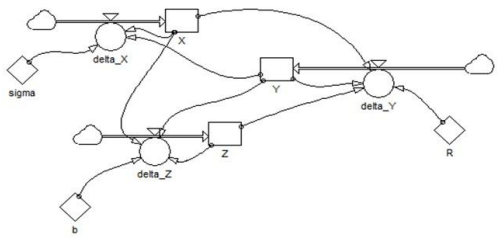


Figure 3. Forrester diagram for Lorentz model

A Forrester diagram shows the identified level, flow and auxiliary variables of a system. The level variables describe the states of the systems; flow variables refer to the input and output flows which influence the level; and auxiliary variables can determine a level or rate variable or another auxiliary variable, but it is not itself a level or rate variable of direct interest to be solved within the model. A Forrester model has a direct correspondence with the differential equations that define the relationship among the elements of the dynamic systems.

D. Lorentz Model

In 1963 Lorentz abstracted three differential equations that can be used to test ideas in nonlinear dynamics. These equations are

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= -xz + rz - y \\ \dot{z} &= xy - bz \end{aligned} \tag{1}$$

This set of equations (Lorentz model) has been applied to the comprehension of complex processes, obtaining generic models to make continuous simulations, in particular related to complex production system behaviors. As a result, it allows identifying regulation mechanisms in the presence of disturbances. However, a system having such deterministic behavior may result unpredictable yielding chaotic solutions because of their sensitivity on initial conditions and settings of parameters σ, r and b. In particular, the Lorenz attractor is a set of chaotic solutions [7] of the Lorenz system which, when plotted, resemble a butterfly or figure eight, as depicted in Figs. …

Due to the nonlinear terms, (1) cannot be solved analytically. As an example, if we choose σ = 10; r = 28; b = 2.67, and initial conditions $x_0 = 0$; $y_0 = 1$; $z_0 = 0$, and solve using Euler´s numerical solution, for 50,000 iterations the three-dimensional plot is the same as that shown in Fig. 4, and its projection on planes XY, YZ, and XZ, is depicted in Fig. 5.

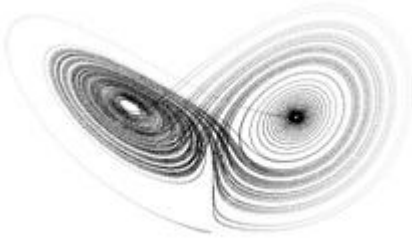


Figure 4. Lorentz attractor

 ISBN: 978-1-61208-262-2

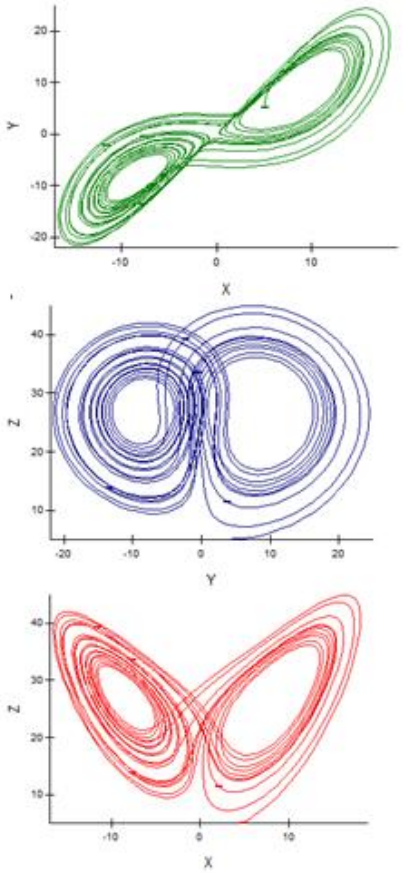


Figure 5. Lorentz chaotic attractors projections

$$\begin{aligned} \dot{x} &= 10(y-x) \\ \dot{y} &= -xz + 28z - y \\ \dot{z} &= xy - 2.67z \end{aligned} \qquad (2)$$

$$x_0 = 0;\ y_0 = 1;\ z_0 = 0$$

E. Client Mobile Device Interface

Whichever modeling tool is used (e.g., neural network, support vector machine, etc.), the process can be automated in a pervasive context. To that end, this design considers the use of a modern mobile communications platform, cloud computing, a mobile client device on-site by the plant, a monitoring platform inside or outside the mine, and a server in the cloud.

In the mobile client device, several software interfaces must be implemented to support the proposed solution [8]. An interface must be implemented for data transmission between the application server side in the private cloud and the M2M platform from the service provider [9]. This interface consists of two parts; one is to deploy a web service which provides local services for the remote M2M platform and is responsible for receiving real-time data from the Android client side; the other is to develop the service software which runs in the communication server background, with the functions of commands sending, data parsing, storage and query to the database. Two modules are considered; the first is to request services from the M2M platform, including platform login and sending command, and the second is the background processing of data which is called by local web service for parsing data and storing into the database. Another interface is responsible for getting data directly from plant sensors in the mine to the server in the cloud using the M2M infrastructure. Finally, a third interface must communicate any event, related to both mobile client and the plant inside the mine, to the screen at the monitoring and control platform.

F. Collecting Data and Request

From Fig. 1 we note that data from the plant's measurements are transmitted directly to the server in the cloud. In order to perform modeling and control over the plant, the process will be controlled by the operator, accessing the server in the cloud through the client program in the Android mobile device, allowing the operator in real-time, without leaving the site, to control the modeling process. Table 1 represents data from the plant inside the mine received by the server in the cloud. At this point no real data are collected, so they correspond to the solution of equation (2).

TABLE 1. COLLECTING DATA

Time (ms)	X	Y	Z
1	0	1	0
2	0.02	0.998	0
3	0.03956	0.997124	3.99E-05
…	…	…	…
1998	-8.493289	-8.793423	26.608915
1999	-8.499291	-8.799466	26.616371
2000	-8.505295	-8.805386	26.623996

G. Getting the Model

To identify and get a model for the plant's behavior, the operator using the application interface in the mobile client device asks the server to compute the model for the dynamic system under study (Fig. 6), specifying the modeling method.

The server will reply showing the requested model (Fig. 7). Several models can be tested until a suitable one is selected.

 ISBN: 978-1-61208-262-2

H. Selecting the Mobile Client

According to Gartner's statistics, until November 2012 the market share of different operating systems for mobile devices was 72.4% for Android, 13.9% for iOS, 5.3% for Research In Motion, 3.0% for Bada, 2.6% for Symbian, 2.4% for Microsoft, and 0.4% for others.

Android is an operating system designed specifically for mobile devices [11]. It runs on the Linux kernel. The Android Software Development Kit (SDK) provides the tools and Application Programming Interfaces (APIs) necessary to develop applications using Java.

Applications written in Java can be compiled to be executed in a Dalvik virtual machine, which is a specialized virtual machine implementation designed for mobile device use. Other interesting characteristics of Android are the capability for reusing and replacing components, and the availability of a number of handset layouts, adaptable to larger, VGA, 2D graphics library, 3D graphics library based on OpenGL ES 1.0 specifications, and traditional Smartphone layouts.

Hence, to implement our mobile client we used Android because of its open nature, widespread use and the portability of the code.

I. Monitoring Platform Screen

During the entire process every event is displayed on a monitor screen (Fig. 8) at the control and monitoring platform, located either inside or outside the mine. The information displayed on the screen includes commands, data sent by the plant's real-time measurement system, request and replay messages between mobile client and server, numerical and graphical results from modeling, and real-time plant's response. These enable supervisors or other expert employees to supervise the automated control and tools from a remote operations centre.

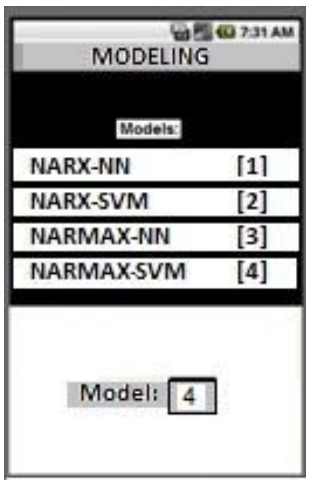


Figure 6. Android client interface

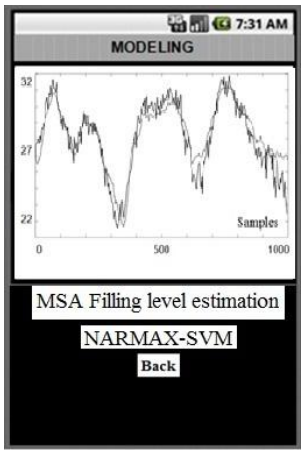


Figure 7. Model selection

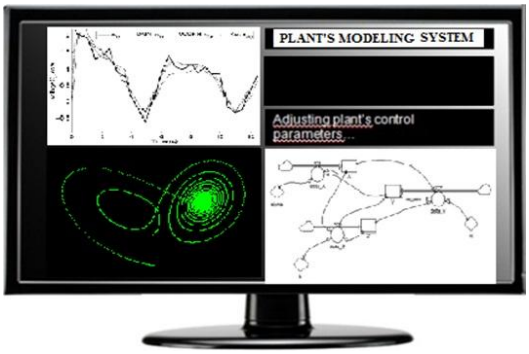


Figure 8. Monitoring screen

IV. MODELING APPLICATION

A. Software Sensor based on a NARMAX - Support Vector Machine Model for Semi-Autogenous Grinding

The estimation of states in complex processes such as the Semi-Autogenous Grinding process (SAG) in copper mining is an important and complex task due to difficulties in measuring some relevant variables directly online and in real time. In [12] the authors present interesting modeling results using Nonlinear Autoregressive Moving Average with Exogenous Input (NARMAX) and Support Vector Machines (SVM), when acting as estimators of state variables for a SAG milling operation. They propose a simple and original methodology to develop NARMAX models made with SVMs. In terms of the milling process, NARMAX-SVM provides a useful tool for estimating the value of the filling level parameter, which cannot be measured using readily available tools. The results show the predictive power of NARMAX-SVM models over those made of Neural Networks (NN). NARMAX-SVM has a significantly lower mean square error (MSE) than all other models.

The effective modeling results from NARMAX-SVM may be available to the operator through his mobile device,

 ISBN: 978-1-61208-262-2

and for those on the monitoring platform as well as their mobiles devices anywhere inside the mine.

In what follows we will explain shortly how NARMAX-SVM works.

B. Virtual sensor structure

The application of virtual sensors to the SAG milling process described in [12] is to estimate on-line and in real time the values of the variable "Level". This is a significant variable for the grinding process, whose values are very difficult to measure directly, in real time and offline.

The NARX model (Fig. 9), proposed to implement the virtual sensors, uses as inputs the previous values of the variable "Level" (the one to be estimated), and an exogenous variable, the "Pressure on the mill shaft breaks". This is a variable easy to measure online and in real time, and is related to the variable of interest.

In the NARMAX model (Fig. 10), in addition to the same inputs used in the NARX model, the previous errors committed by the model are used.

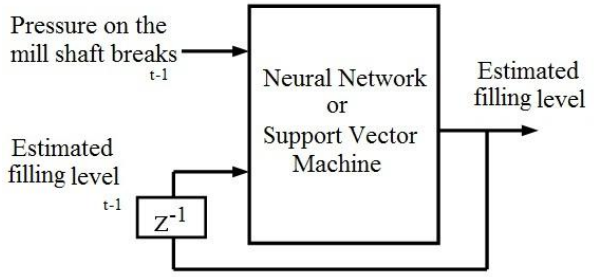


Figure 9. NARX virtual sensor structure

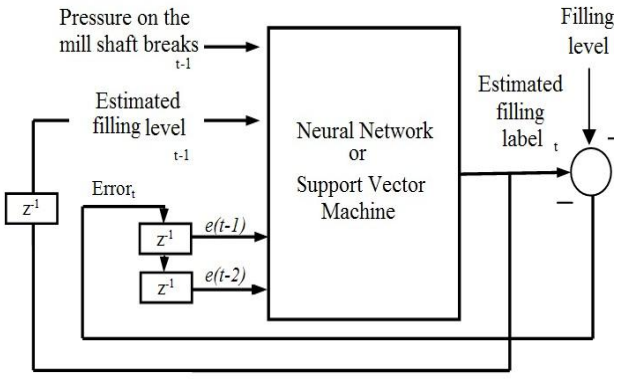


Figure 10. NARMAX virtual sensor structure

SVM and neural network models were trained with 500 samples and validated with 1000 samples. A third set of 1000 samples (test set) was used to get the final performance indices shown in Table 2. Each sample has the fill level and the pressure during breaks at time t-1 as inputs, and the filling level at time t as the output (the models are of first order).

Once identified, the four models obtained for estimating the filling level of the SAG mill (NARX and NARMAX using SVM and neural networks, respectively), their prediction capability for Multiple Step Ahead (MSA) forecast was evaluated. The estimation error was quantified using the mean square error (MSE) of Matlab. As a result, we can see that SVM implementations perform better than neural network cases.

TABLE 2. MSA FORECASTING MEAN SQUARE ERROR

	NARX	NARMAX
NN	3.5889	1.0773
SVM	1.0256	0.4424

D. Forecasting Results

Figs. 11 a) and b) show the estimation of the variable filling level (%) obtained with the NARX and NARMAX models, respectively, using SVM in MSA forecasting for the test data set.

From these results we can see that the NARMAX model performs better than NARX when both act as MSA predictors. NARMAX type models, though requiring a more complex identification procedure, consider previous prediction errors. Moreover, the models implemented using SVM significantly outperform those made using neural networks.

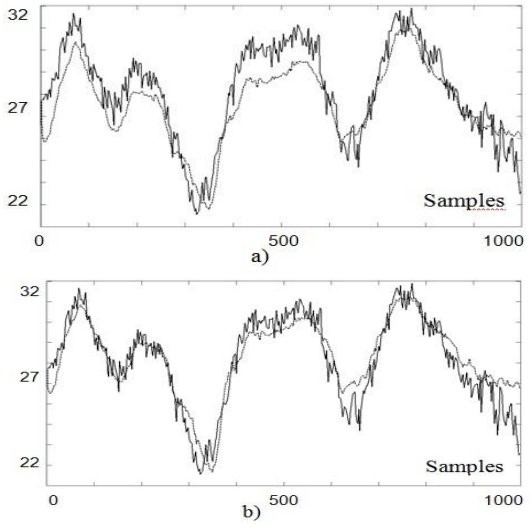


Figure 11. MSA Filling level estimation
a) NARX-SVM; b) NARMAX-SVM

 ISBN: 978-1-61208-262-2

V. CONCLUSIONS

This solution has advantages in many aspects. The machine-to-machine based communications mode with the server in the cloud and Android client inside the mine brings up pervasive mining, a system with wider coverage, higher communication efficiency, better fault-tolerance, and anytime anywhere availability.

This solution may be applied for any plant inside a mine for which their modeling in situ, in real time, allows to control aspects such as instability by adjusting their parameters without stopping the process.

In order to deploy the proposed design, this study has considered two modeling tools: NARX and NARMAX, which have been combined with Support Vector Machines (SVM) and Neural Networks (NN) to implement dynamic models. The purpose of the resulting models is to act as state estimators for the variable filling level of a semi-autogenous grinding process.

The proposed solution responds to the modeling needs, and also to the forecasting of plants functioning inside a mine. It simplifies the monitoring process and contributes to better control and enhanced safety.

The system proposed may represent a valuable design that helps to perform stricter supervision, set up safer work conditions for the miners, and deploy new technologies to enhance miner safety and improve processes inside the mine.

ACKNOWLEDGMENT

The authors acknowledge with thanks the financial support of the “Center for multidisciplinary research on signal processing” (Conicyt/ACT1120 project) and USACH/Dicyt project 06.917S.

REFERENCES

[1] N. Radio, Y. Zhang, M. Tatipamula, and V. Madisetti, «Next-Generation Applications on Cellular Networks: Trends, Challenges, and Solutions,» *Proceedings of the IEEE , vol.100, no.4,* pp. 841-854, 2012.

[2] E. Mancini, G. Wainer, K. Al-Zoubi, and O. Dalle, «Simulation in the Cloud Using Handheld Devices,» de *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on*, pp. 867-872, 2012.

[3] S. Pocuca, D. Giljevic, «Machine to machine (M2M) communication impacts on mobile network capacity and behaviour,» de *MIPRO, 2012 Proceedings of the 35th International Convention* , pp. 607-611, 2012.

[4] F. Cubillos, G. Acuna, «Simulation Studies of On-Line Identification of Complex Processes with Neural Networks,» *Lecture Notes in Computer Science,Volume 3972,* pp. 808-814, 2006.

[5] I. Chairez, «Wavelet Differential Neural Network Observer,» *Transactions on Neural Networks,* pp. 1439-1449, 2009.

[6] K. Salahshoor, A. S. Kamalabady, «On-line Identification of Nonlinear Multivariable Processes using self-generating RBF Neural Networks,» *Asian Journal of Control, Vol. 12, No. 5,* pp. 626-639, 2010.

[7] M. D. Alfaro, J. M. Sepúlveda, and J. A. Ulloa, «Forecasting Chaotic Series in Manufacturing Systems by Vector Support Machine Regression and Neural Networks,» *International Journal of Computers Communications & Control,* pp. 8-17, 2012.

[8] Y. Yang, H. Ye, and S. Fei, «Design of communication interface for M2M-based positioning and monitoring system,» *Electronics, Communications and Control (ICECC),* pp. 2624-2627, 9-11, 2011.

[9] W. Sun, M. Song, «A general M2M device model,» de *Web Society (SWS), 2010 IEEE 2nd Symposium*, pp. 578-581, 2010.

[10] C. Wang, W. Duan, J. Ma, and C. Wang, «The research of Android System Architecture and application programming,» *Computer Science and Network Technology (ICCSNT), 2011 International Conference on,* vol. 2, pp. 785 - 790, 2011.

[11] W. Hu, D. Han, A. Hindle, and K. Wong, «The build dependency perspective of Android's concrete architecture,» de *Mining Software Repositories (MSR), 2012 9th IEEE Working Conference on* , Zurich, Switzerland, pp. 128-131, 2012.

[12] G. Acuna, M. Curilem, and F. Cubillos, «Development of a Software Sensor based on a NARMAX-Support Vector Machine Model for Semi-Autogenous Grinding,» *Revista Iberoamericana de Automatica e Informatica Industrial,* p. unpublished, 2011.

 ISBN: 978-1-61208-262-2