JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2013

(Contents Continued on Back Cover)

ISSN 1509-4553 on-line: ISSN 1899-8852

Warsaw 2013

Circulation: 300 copies

Sowa – Druk na życzenie, www.sowadruk.pl, tel. 22 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

In this volume of the 1/2013 *Journal of Telecommunications and Information Technology* we have collected a set of thirteen papers presenting a broad range of topics related to telecommunication networks. The topics cover subjects ranging from various applications of telecommunications networks and recommendations as to the implementation of new services and protocols, to analytical and simulation methods for modelling telecommunications networks and systems.

This issue is opened by the article *Wireless at the "Connected Games": How the London 2012 Olympic and Paralympic Games will utilize the latest Wi-Fi Technology*, the best paper of the 51st FITCE Congress held in Poznań in 2012. Peter Leonhardt, the author of the article, describes Wi-Fi deployment across the London 2012 Olympic Games venue – one of the largest and most dense Wi-Fi deployment in the world. The article presents the adopted strategy in the implementation of the Wi-Fi network, effective analysis of different technologies in Wi-Fi networks and the Wi-Fi traffic patterns observed during the most popular Olympic events.

The second article *Connecting for Surgery: The Belgian Use Case on the Legal Aspects of the Digital Operating Room*, the authors, Niels Vandezande, Griet Verhenneman, and Jos Dumortier, address the problem of protecting patients' privacy and securing their health information following the threats resulting directly from the implementation of telecommunication technology in medicine. The introduction of the concept of the Internet of Things has led to solutions termed as "digital operating room", according to which different devices will become interconnected and will create store and exchange patients' data. The authors of the article attempt to provide some answers to whether the currently used automation of the flow and the storage environment of patient's data violate the European framework on data protection.

Practical dimensions of the implementation of the IP network in the datacentre infrastructure of the company Capgemini are presented in Marco van der Pal's article *IPv6 Preparation and Deployment in Datacenter Infrastructure – A Practical Approach*. The author gives a detailed description of particular stages in the implementation of the IPv6 protocol, pointing out that the IPv6 is not solely a networking technology. The presented conclusions on the cooperation between the Pv6 and the IPv4 networks, the ongoing evolution of standards for IPv6 and the influence of the implementation of IPv6 on existing operating systems and software, may help facilitate the process of migration towards IPv6 in many companies and organizations.

The present issue of the journal (*JTIT*) also includes articles that focus on advances in analytical and simulation methods for modelling telecommunications systems. The first article in this section of articles, entitled *The Goals and Benefits of Network Modelling in a Commercial Environment*, attempts to define and classify the areas of implementation for modelling and simulation tools. Edward Smith, the author of the article, describes a number of academic solutions as well as commercial applications, juxtaposing their performance levels and providing a comparative evaluation. The evaluation of modelling tools is accompanied with an important remark of the author that 85% of the results presented in literature cannot actually be independently confirmed by other researchers due to the lack of precise description of parameters for a given model application.

One of the modelling tools described by E. Smith, OPNET discrete event simulation tools, is then used in the following article *Reliable and High QoS Wireless Communications over Harsh Environments* to evaluate the efficiency of the proposed algorithm for improvement of multicast transmission in wireless networks. The authors of the article, Josu Bilbao, Aitor Calvo, Igor Armendariz, and Pedro Crespo, have managed to achieve a significant improvement in the QoS parameters for multicast transmission through the use of inter-node collaborative schema and Network Coding.

Simulation tool, i.e., ns2, has been also used by Veronica Palma and Anna Maria Vegni to evaluate vehicular network performances in the dissemination of safety and emergency messages. The studies carried out by the authors have made it possible to assess throughput and delays values for different traffic scenarios and for both Vehicle-to-Vehicle (V2V) communications and for Vehicle-to-Infrastructure (V2I) communications. The authors of the article *On the Optimal Design of a Broadcast Data Dissemination System over VANET Providing V2V and V2I Communications "The Vision of Rome as a Smart City"* proceed to consider the results of their research study within the framework of recommendations for optimum deployment of relay nodes in order to maximize network performance while keeping low the economic requirements.

In the article *Innovative Method of the Evaluation of Multicriterial Multicast Routing Algorithms*, simulation studies are also used for the evaluation of the effectiveness of multicriteria multicast routing algorithms. The article proposes a new routing algorithm: Aggregated MLARAC (Multi-dimensional LAgragian Relaxation based Aggregated Cost algorithm). To evaluate its effectiveness, a comparative study of the effectiveness in constructing trees in the function of the number of receivers for networks with different topologies and size has been carried out. The authors of the article, Krzysztof Stachowiak and Piotr Zwierzykowski, propose a new original method for the evaluation of multicriteria multicast routing algorithms and demonstrate that the proposed Aggregated MLRAC algorithm is the most effective when chosen for a small number of criteria.

An analytical approach to modelling of telecommunications systems is presented in the article *Recurrent Method for Blocking Probability Calculation in Switching Networks with Overflow Links*. The article proposes an analytical method for modelling multi-service switching networks with overflow links. The method worked out by the authors makes it possible to quickly determine values of the internal and the external blocking probabilities. The results obtained in the study clearly indicate a possibility of a significant reduction of the value of internal blocking for all classes of streams of offered traffic following an application of overflow links.

In the article *Blind Estimation of Linear and Nonlinear Sparse Channels*, Kristina Georgoulakis presents a Clustering Based Blind Channel Estimator for the so-called zero pad channels. The proposed algorithm uses an unsupervised clustering technique for the estimation of data clusters. Clusters labeling is performed by a Hidden Markov Model of the observation sequence appropriately modified to exploit channel sparsity. The algorithm achieves a substantial complexity reduction compared to the fully evaluated technique. In conjunction with a Parallel Trellis Viterbi Algorithm for data detection, the proposed algorithm can lead to the reduced complexity without performance reduction.

Juliusz L. Kulikowski in his article *Hidden Context Influence on Pattern Recognition* considers the influence of hidden additional information concerning the circumstances of input data acquisition on the quality of decisions based on the data. An analogue to the intuition influencing natural decision making is indicated in the article. The problem of contextual information in decision making based on Bayes rule, on reference data sets in various applications as well as of scene analysis by numerous examples is also illustrated.

The last three articles discuss the results obtained from processing of measurement data for telecommunications systems. Jakub Sobczak in his article *Traffic Analysis in the Network of a Local Voice over Internet Protocol Operator* presents the results of measurements obtained in a network of one of the operators of the Voice over IP service. The study was aimed at gathering information on distributions describing the intensity of calls and holding times of connections that represent residential and business subscribers to the VoIP service. Thus obtained statistical data, along with information on voice codecs used (determining demanded throughput and bit rates) allowed the author to present his own conclusions regarding the choice of method for dimensioning of VoIP networks.

The paper *Diversity of Temporal and Territorial Social Penetration Rates of Information Technology in Europe* by Beata Ziewiec presents the diversity of social penetration rates of information and communication technologies (ICT) among selected European countries according to European statistics on diverse ICT indicators. The data considered cover the 2006–2010 time range and was obtained from the Eurostat portal. The following ICT indicators were analyzed: percentage of households or corporation with broadband access to the Internet (HHBAI), percentage of individuals who are regularly using the Internet (IRUI), percentage of individuals who ordered goods or services over the Internet (IOGSI). These indicators were analyzed, inter alia, in terms of delays or advances (in years) as compared to the averages in EU.

In the last article entitled *Performance and Limitations of VDSL2-based Next Generation Access Networks*, George Heliotis, Lowell-Panayotis Dimos, Ioannis Kordoulis and George Agapiou present the results of feasible bit rates in VDSL2-based access networks. The measurements involved 3 VDSL2 operational profiles for different scenarios that differed in crosstalk values and impulse noise as the two principal sources of degradation in high-rate transmission systems. The obtained results will be then used by the authors to improve the quality of transmission in access networks.

Mariusz Głąbowski
Guest Editor

Paper

Wireless at the "Connected Games": How the London 2012 Olympic and Paralympic Games Utilized the Latest Wi-Fi Technology

Peter Leonhardt

British Telecom, London, England

Abstract—The London 2012 Olympic and Paraolympic Games drew together tens of thousands of people in the form of athletes, organisers, media, VIPs, public and many more groups and individuals. With the growth in smart phones and tablets coupled with the ever expanding volume and range of content accessed via Aps and browsers, this huge volume of people expected connectivity from their mobile devices during their time on the Olympic sites. It was BT's role in the games as the official communications services partner to deliver Wi-Fi connectivity across the venues, covering the range of users from the public, athletes, organisers and ticket scanners. This paper examines this state of the art Wi-Fi solution.

Keywords—high density Wi-Fi, Olympics, paralympics, Wi-Fi, wireless.

1. Introduction

The London 2012 Olympic Games was coined the "Connected Games", as never before there were so many systems, user groups and individuals associated with the Olympics demanding network connectivity to each other and the wider world. British Telecom (BT) as the official communications services partner to the London 2012 rolled out a national network to support the games, overlaid with different services for different requirements. Wi-Fi was just one such service, but a service which was very high profile in terms of being visible to how well it was performing due to the personal connectivity it provides, as well as being an ever growing area due to the increasing numbers of smart phones, tablets and the ever expanding content and services available online.

The paper examines the London 2012 Wi-Fi, looking at the requirements and the nature of the deployment including High Density Wi-Fi, and then examines its operation and how it performed.

2. The Requirements

The Wi-Fi requirements for the London 2012 Olympics were demanding and unique. The solution had got to be on a national level, providing a centralised Wi-Fi deployment to both sporting and non-sporting venues, 27 in total. The solution had to be able to scale up and scale down fast, as venues were built, used for a test event, or the real event, and then taken away, in some cases over a matter of days. The solution had to be secure, and centrally managed, monitored and operated from the London 2012 Technical Operations Centre. The Wi-Fi radio frequencies also had to be carefully planned with the London Organising Committee of the Olympic and Paralympic Games (LOCOG), to make sure different systems did not interfere with one another. This also included games systems such as for fencing and taekwondo which had wireless transmitters operating in

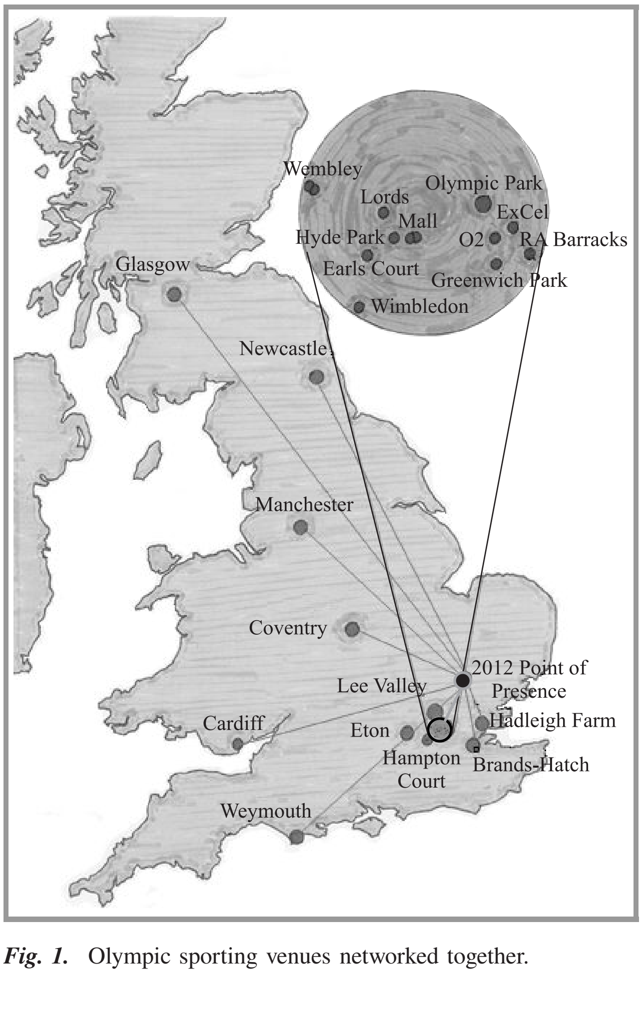


Fig. 1. Olympic sporting venues networked together.

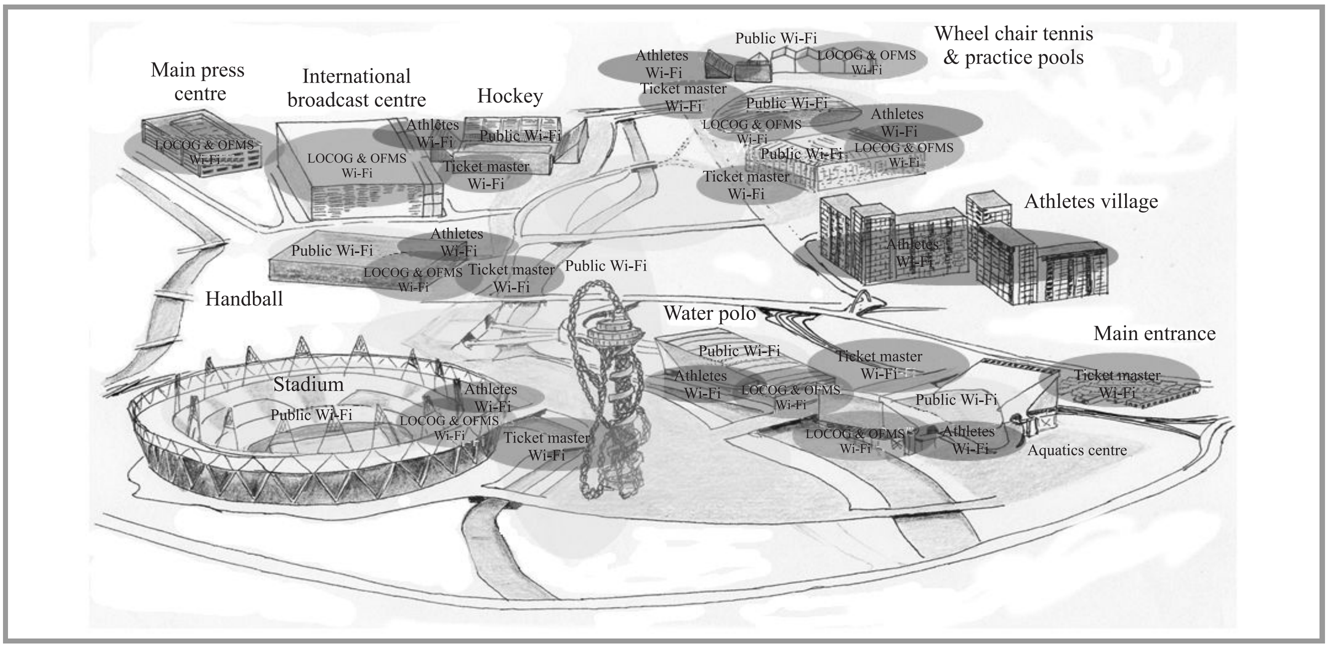


Fig. 2. Wi-Fi services rolled out across the Olympic Park.

the Wi-Fi spectrum range built into the competitors clothing to relay hits back to the scoring system.

The solution also had to be reliable and resilient. All the venues were dual homed to BT's resilient points of presence dedicated for the games, and the wireless traffic once taken into the wired network via the access points on the venue where sent over this extensive backhaul network which linked all the venues together. Figure 1 shows just the sporting venues which were linked together, providing the backbone for the wireless network. The solution also had to cater for multiple user groups which are outlined in the below section.

3. User Groups

Public. The public increasingly wanted and expected Wi-Fi to be available for their general use to get to the internet, to had the ability to quickly upload photos and get to social media sites such as Facebook and twitter to share their experiences at the games.

LOCOG. LOCOG required a Wi-Fi service within the admin areas of venues, both sporting and non-sporting for general use. This service was also be purchased for use by the press, or other associated groups or individuals.

Machines (ticketing). A Wi-Fi service was required for the mobile ticket scanners which were operated by ticket master at the entrance areas.

Athletes. As part of the general facilities for athletes, a good Wi-Fi service was required to be available for them both within sporting venues and the Olympic Village.

Mobile Data Offload. Samsung had partnered with the Olympic organisers and had produced an Olympic Family range of mobile phones which were used by organisers and VIPs on the park and in venues. These OFMS (Olympic Family Mobile Service) phones required data offload via Wi-Fi when in designated areas.

4. Wi-Fi Deployment Across the Olympic Park

The different user groups required different Wi-Fi set-ups depending on physical location and logical configuration. Broadly speaking, the Wi-Fi was split into "front of house" services and "back of house" services. Front of house services were in areas where the public would be to watch the games or were in transit, so these Wi-Fi services would include public, OFMS, and ticket master. The back of house services were mainly for use in areas the public would not be, but athletes, organiser and press would, these Wi-Fi services included LOCOG, OFMS and athlete's Wi-Fi.

The coverage areas across the venues were clearly dependent on where the user groups would require, or desire access to Wi-Fi, and therefore different areas had different wireless services rolled out. This applied to all venues, both on and off the park. The diagram below shows the main coverage areas specifically for the park so as to provide an overview of how the services were laid out in differing logical patterns, but all were sharing the same centralised physical infrastructure.

The most significant deployment in terms of differences in physical and logical configuration was the public Wi-Fi located in the seating bowls of the venues. This was High Density Wi-Fi, and will be explained in more detail in the following section.

5. High Density Wi-Fi

High Density Wi-Fi is a form of Wi-Fi deployment which is designed to provide good Wi-Fi coverage to groups of users in a densely packed group, for example, seating next to one another in a sports stadium. Standard Wi-Fi deployments will fail to provide good service in such environments once there are significant numbers of Wi-Fi clients, as the access points and their associated radio channels will become congested by the volume of connections. Simply adding in more conventional access points into the environment will not necessarily make things better, as the radio footprint of additional access points will overlap with others and can in turn cause even more congestion.

The principle of High Density Wi-Fi (HD Wi-Fi) is that it is engineered to create lots of small radio footprints so as to be able to cater for pockets of clients in localised areas of the stadium. This means the radio traffic between the client and access point can be kept to lower volume and power, thus to avoiding the access point becoming swamped by too many clients and minimising radio interference from surrounding access points and their associated client traffic. Figure 3 below illustrates the way HD Wi-Fi is designed to cover seats in a stadium seating bowl.

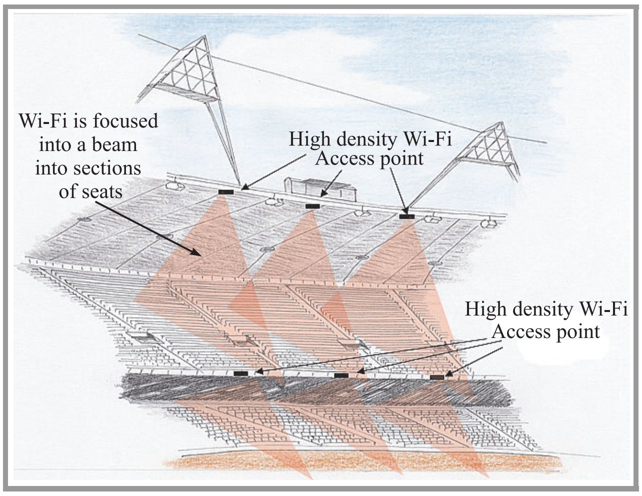


Fig. 3. High Density Wi-Fi in the stadium.

There are four areas of physical and logical configuration detail which gives HD Wi-Fi access points their smaller radio footprint:

Antenna. A directional antenna is used to focus the radio waves from the access point onto a specific area, in a similar way the lens of a torch is used to focus light. For the 2012 deployment, the Cisco 3502P access point combined with the Cisco AIR-ANT25137NP-R grayling antenna was used. This can be seen in Fig. 4.

Power. The power of the access point is reduced to a level where it can adequately cover its designated area, but not much further. This helps to avoid its signal over spilling into other access point's coverage areas.

Fig. 4. Cisco 3502P AP with grayling antenna.

Minimum Supported Data Rate. The further away the client, the less strong the signal, thus the lower the data rate which can be achieved by the client to the access point. Therefore the minimum supported data rate of the access point is raised to prevent clients which are further away, outside of the access points catchment area connecting.

Coverage Threshold. The minimum coverage threshold is reduced to make it more stringent, only allowing clients with a good Receive Signal Strength Indicator to connect to the access point. This helps keep clients local to the access point.

The small radio footprints means the access points must be designed and deployed carefully so as to avoid black spots but also not overlap too much with each other. In total, 176 HD Wi-Fi access points have been deployed in the seating bowl of the olympic stadium. When mounting access points within the infrastructure of a venue, it is necessary to work with the features of the venue so as to find the best mounting positions. In the case of the hockey stadium, there is no roof or over-hanging gantry, so the access points were mounted underneath the seats pointing upwards. A photo showing an access point in this deployment is shown below in Fig. 5.

Once the access points have been positioned physically, and the antennas angled to cover the desired areas of seating or other high density area, then the solution is optimised via testing and then tuning of the configuration parameters above.

One of the challenges with the public Wi-Fi in the seating bowl is how to test it, to accurately simulate a crowd of people for testing. The short answer is there is no precisely accurate way to simulate a crowd of people, due to the random nature of user behaviours and different individual's technology. So BT organised test events to create a controlled crowd of users, who could then be monitored and their experiences fed back for further optimisation. Rehearsals in the stadium also were taken advantage of for the testing and monitoring of the performance of the solution. Once the install and optimisation is completed, it is then

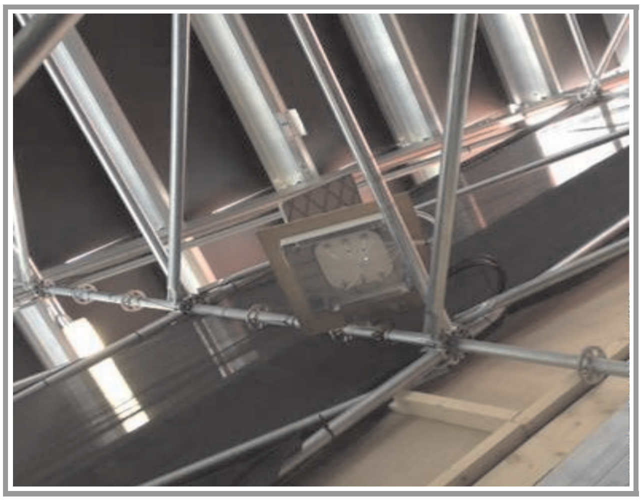


Fig. 5. Access point with external antenna directed upwards under the seats in the hockey venue.

necessary to police and monitor the radio spectrum in the Wi-Fi space for sources of interference which could disrupt the service.

6. Policing of the Wi-Fi Spectrum

As Wi-Fi uses unlicensed radio spectrum in the 2.4 GHz and 5 GHz ranges, it is subject to all kinds of interference. This can come in the forms of other Wi-Fi access points, as well as many other non-Wi-Fi forms such as wireless cameras, Bluetooth devices, and also anything that can transmit radio waves as a side effect of their main job such as microwave ovens.

In a distributed Wi-Fi network the scale and breadth of the Olympic Games, it is impossible to prevent all the sources of interference. However, it is important to pick up on the service affecting sources of interference, identify them quickly, and also have the political measures and planning in place to deal with them.

A good example of interference was during the Wi-Fi optimisation of the hockey stadium, it was noticed that the Wi-Fi access was being intermittently knocked out. Upon investigation the source of the interference was tracked down to a crane operating nearby. The crane had a wireless camera on its boom which relayed pictures using the 2.4 GHz spectrum to the a TV screen in the drivers cab. Every time the crane was operated, the signal was so powerful it knocked out the ability of the Wi-Fi in the hockey stadium to operate. In this example, once the source was identified, it was not a problem as the crane would only be there before games time. But it does show the importance of being able to pin down sources of disruption so as to be able to understand and highlight the cause, and asses and mitigate any impact it may have on service.

Early on in planning for the Olympics, it was necessary to work with LOCOG to understand what measures were being taken to regulate within the Olympic venues the radio spectrums, so as to be able to plan the successful deployment of Wi-Fi. With the large numbers of press, TV and broadcasting bodies descending on the park just before the games, there was huge scope for transmission equipment to disrupt the Wi-Fi services. Equipment of this kind therefore had to be approved by LOCOG for use, and plans on how to deal with sources of disruption formulated.

The BT Wi-Fi solution was monitored centrally from London 2012's Technical Operations Centre by BT, so any sources of disruption could be picked up pro-actively. A team on the ground with mobile spectrum analysers could then be dispatched to the location of the problem, and the exact cause could then be pinned down. How to deal with the cause depends on what it was, but the important thing was finding it so as to be able to rule out other technical problems for the service disruption.

7. Results – the Wi-Fi in Operation

The 2012 Wi-Fi performed well across all services during the games. The LOCOG service was used during the years and months running up to games, mainly from the organising offices for the general day to day build and preparation during the games, predictably the other services overtook the LOCOG Wi-Fi as most popular, with the public BT Wi-Fi being the most utilised, followed by the athletes Wi-Fi. The ticket master Wi-Fi was built for consistent support for the mobile ticket scanners, which it provided throughout the games.

The pattern of utilisation is clearly shown below in Fig. 6, with the athletes Wi-Fi client counts during the games, with the number of clients ramping up over the preceding days of the games, peaking during the middle, and then tailing off to the end. This pattern was reflected across all the services apart from ticket master which was consistent with the number of scanners throughout.

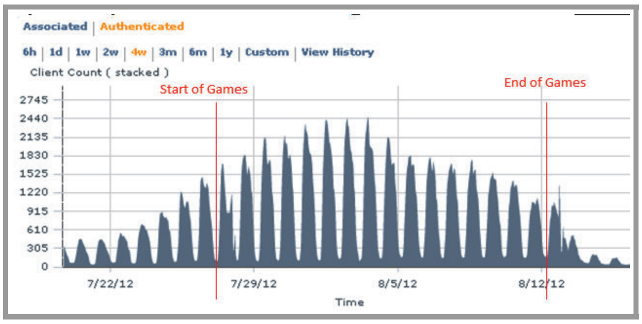


Fig. 6. Client count for the athlete's Wi-Fi service.

Out of the 2.4 GHz and the 5 GHz Wi-Fi spectrum, the 2.4 GHz was most heavily used. This is because the majority of handheld Wi-Fi devices only support 2.4 GHz, such as the iPhone. As a result, in areas of dense utilisation, the 2.4 GHz had slower speeds, where are the 5 GHz devices (such as iPads and the Samsung OFMS phones) had better performance as there was less congestion in the air.

The HD Wi-Fi performed as designed in the seating bowls of the venues it was installed, creating smaller Wi-Fi footprints, being able to deal with dense crowds better than conventional Wi-Fi.

The traffic patterns during popular events showed an interesting trend. It was originally thought that during popular events, upload traffic out the venue where the event was being held would be the most predominant traffic on the 2012 Wi-Fi solution, with people uploading pictures, messages and videos of the event to the internet. In reality however, the predominant traffic trend during a popular event such as the 100 meter final was in the download direction. When the download traffic was drilled into, it was seen to be destined for other venues with people in, but not the venue where the popular event was being held. So in the case of the 100 meter final, the traffic spike was in the basketball and handball arenas. Figure 7 shows the traffic spike into the basketball arena when the 100 meter final was being run in the stadium.

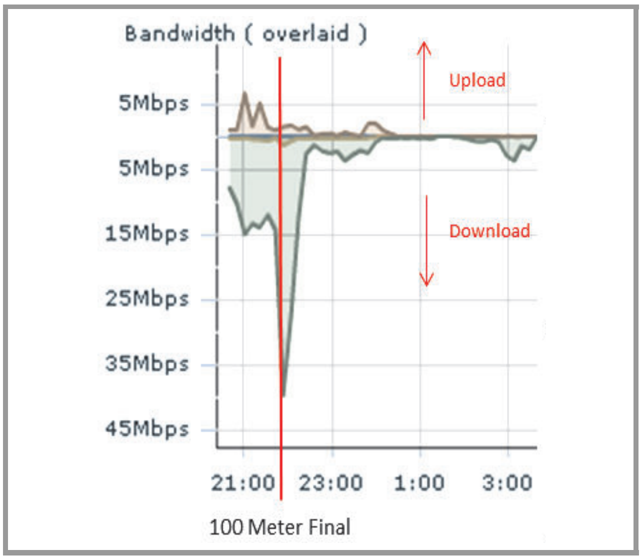


Fig. 7. Wi-Fi download spike into the basketball arena during the 100 meter final.

The reason for this is that spectators in the other venues were streaming the live footage and results of the popular event to their mobile devices, taking a short break from watching the event they had tickets to, to not miss out on the big events happening nearby.

8. Conclusions

The extensive use of Wi-Fi at the London 2012 Olympics by multiple user groups shows that as would be expected, the demand for wireless communications is strong. This is driven by the growth in smart phones and tablets, and the ever growing availability of content, and interactive social media. It is therefore expected that this demand for wireless connectivity and bandwidth will continue to grow and be demanded at other such events which draw people together.

In order for a Wi-Fi deployment on this scale and with such diversity to be successfully deployed, including HD Wi-Fi, the strategy should follow the below sequence:

Design. This includes site surveys to identify the best positions for access points, and the backhaul infrastructure.

Build & Deployment. The backhaul infrastructure should be built first, and then the access points should be commissioned last.

Optimisation & Testing. This is an especially critical phase for HD Wi-Fi. During optimisation the parameters outlined in section V are tuned so as create the optimum Wi-Fi cell sizes, with the best physical locations and angles of the antennas tweaked and adjusted as necessary. Continual testing is required to make sure and confirm what effects the changes are having in the overall performance of the Wi-Fi. Test events should be conducted to see how the Wi-Fi performs in realistic crowded situations.

Policing and Monitoring. After optimisation and testing has settled on the set-up of the Wi-Fi solution, then it needs to be policed. This means the radio spectrum needs to be monitored for sources of interference to the Wi-Fi, and if appropriate, acted upon. What is acted upon depends on what has been agreed with the organisers (LOCOG in the case of London 2012), but the main thing is being able to pin down any radio interference issues to their cause, so appropriate action if required is an option.

In areas congested with Wi-Fi clients, devices which operated within the 5 GHz channels performed better. This is due to less congestion, as less device types operate on 5 GHz frequencies, and there are more channels available in this spectrum range. As time passes however, it is expected more devices will support 5 GHz, so this advantage they have at the moment will diminish with the growth in 5 GHz devices.

Wi-Fi traffic patterns observed during the most popular Olympic events reveal that the most noticeable trend was downloads to the venues where people were present, but the popular event was not taking place. This was due to people streaming live video of the event taking place elsewhere to their mobile devices. This is probably a phenomenon unique to the Olympics, or other such multi venue sporting events.

Conventional Wi-Fi using a normal omnidirectional antenna does not operate effectively in areas with dense numbers of clients such as the stadium seating bowl. High Density Wi-Fi provides a better service by providing smaller but more frequent wireless cell sizes thus cutting back on the volume of clients associated to an AP, keeping the traffic local and minimising congestion.

Peter Leonhardt after graduating from the University of York, has worked through progressive roles in British Telecom, winning various awards for his work with the London 2012 project and major business customers along the way. He was a lead IP Consultant for BT during the build London 2012, a LAN/WAN Lead Consultant phase during the games itself, and is now a consultant for BT's Rapid Response Unit. He has completed a M.Sc. in business and Telecoms from UCL, achieving a Distinction, and won the award for the best paper at the 2012 FITCE Congress in Poznań, Poland. He is also involved in projects developing appropriate communication systems for developing countries, assisting with projects in Africa.

E-mail: peter.leonhardt@bt.com
British Telecom
Department BT London 2012
London, England

Paper

Connecting for Surgery: The Belgian Use Case on the Legal Aspects of the Digital Operating Room

Niels Vandezande, Griet Verhenneman, and Jos Dumortier

Interdisciplinary Centre for Law & ICT – iMinds, Faculty of Law, KU Leuven, University of Leuven, Leuven, Belgium

Abstract—Telecommunications technology is making its way into operating rooms by new developments in e-health. However, conflicts arise with existing legal principles regarding data protection. This paper deals with key elements of the interactions between data protection and evolution in e-health. The scope will be the digital operating room, where different health services and activities converge through networked technology, raising a number of privacy-related issues. For instance, the patient's health records and tools for recording surgical procedures could be integrated within the same platform, potentially leading to sensitive personal data linkage. Also the possible duration and reason of storage of surgical recordings, is a matter that remains largely unresolved in current practice. First, this paper will analyze the data exchanges of the digital operating room. As these will include personal patient data, it must be assessed whether and how the European framework on data protection can apply. Second, the regulatory regime of the manufacturers of the devices of the digital operating room will be analyzed. Can the current legal framework relating to e-health provide for suitable regulation for such devices? Drawing from experience gained in research projects, this paper aims to provide practical answers to often theoretical questions.

Keywords—data protection, e-health, privacy, telecommunications.

1. Introduction

The nature of the Internet has undergone a number of drastic changes in the last few years. One of the most notable of such changes is the rise of what is referred to as the *Internet of Things*. This phenomenon can be described as being the pervasive and ubiquitous interconnecting of all kinds of everyday objects or rather: "things" that can interact with each other and cooperate to achieve common goals [1]. Relying on the use of new communication technologies including radio-frequency identification (RFID) and near-field communication (NFC) the *Internet of Things* aims to support data exchanges in the global supply chain, facilitating several aspects of daily life [2]. On a global scale, the *Internet of Things* could help track the movement of the many goods that are being transported every day, potentially supporting the identification of counterfeit goods. On a more personal level, a smart refrigerator could detect when it is running out of milk and add this item to the groceries list. This growing number of interconnections between ever more devices will of course result in further growth of the number of services that is already provided online. Also, the amounts of data exchanged in networked environments can be expected to increase exponentially.

Also in the field of healthcare, different kinds of services are gradually moving towards networked environments. In a first wave of e-health solutions, paper records were digitalized into electronic health records that could be shared between healthcare providers. A next step is to use the Internet as a medium for the delivery of healthcare services. As diagnostic services in the fields of pathology and radiology, for instance, are already widely delivered over the Internet, this step is very much unfolding right now. This use of the Internet for telemedicine purposes is only expected to grow, with trials already taking place in fields like nursing, pharmacy and even surgery.

The delivery of such telemedicine services requires the tools and devices that are equipped for use in a highly networked environment in which many services are provided over a distance. Especially in the field of telesurgery, this evolution requires an update to the regular equipment found in the current layout of operating rooms. In what can be regarded as a digital operating room, several types of services such as the provision of health data and the monitoring of vital statistics and different components such as cameras, wires and surgical tools will have to be integrated into a single device that is equipped to handle the data flow with which it will interact.

This convergence of different e-health services, relating to the use, transfers and storage of data can, however, also raise questions with regards to the protection of such data. Especially when the patient's health data is involved, it will have to be assessed what measures need to be taken to protect the patient's privacy. This paper aims to address the specific privacy issues that arise in the context of a digital operating room from a legal point of view. First, it will have to be addressed whether the current legal framework regarding data protection can be applied to the technological developments of the digital operating room. This analysis will shed further light on how the services expected from future e-health developments can be affected by this legal framework. In secondary order, this paper will address the regulatory regime applicable to the manufacturers of the devices of the digital operating room. More specifically, it will be analyzed whether these manufacturers can be subjected to the stricter regulations that generally apply to medical devices.

2. Applicability of the Data Protection Framework

The data exchanges envisioned within the digital operating room will be performed in a sensitive environment. The health records of patients will be handled, a number of persons will execute divergent tasks and a high degree of trust is bestowed onto the proper functioning of the devices that enable the use, exchange and storage of all data generated in the performance of the activities of the digital operating room. The collusion of these different factors raises a number of legal questions with particular regard to the protection of the patient's personal data. Therefore, it will first have to be assessed whether the current legal framework regarding data protection can be applied to the activities of the digital operating room.

Within the European Union, the legal framework relating to the protection of personal data relies on Directive 95/46/EC, also known as the *Data Protection Directive* [3] and the national implementations thereof by the Member States. This directive applies to the processing of personal data, whereby personal data needs to be understood as being any information relating to an identified or identifiable natural person. Important for the determination whether a piece of information constitutes personal data, is therefore whether this information can identify a natural person or at least make him identifiable. A person will generally be considered to be identified when his identity can be confirmed immediately so that he can be singled out from a group. This is the case with, for instance, public sector identification numbers, such as the identification number of a national identity card, which are supposed to be uniquely assigned to one citizen only. As a result, every one of such identification numbers will directly and solely identify one citizen.

The concept of identifiability, however, may raise more discussion. According to the *Data Protection Directive*, a person is identifiable when he "*can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*". In this case, the citizen will not be identified immediately, but a number of elements, or a combination thereof, may lead to the indirect identification of the citizen. In such case, he will be considered to be identifiable [4]. As this broad definition encompasses an enormous amount of different elements and factors of information, it is clear that there is only a limited amount of information that could not be considered as personal data.

When information relates to a citizen, but cannot in any possible way be traced back or linked to that citizen, this information is considered to be anonymous [5]. It should be noted, however, that there is only a very limited amount of data that is truly anonymous. Even when certain information has been depersonalized or encoded, it could still contain information that could be traced back to the citizen, thus making the citizen identifiable and therefore making the information personal data.

In the context of the digital operating room, this could have implications for the recordings made during surgeries. In, for instance, an endoscopic procedure, the recorded images will mainly show a patient's abdominal and pelvic cavity. It is clear that these images alone will in most cases not be able to lead to a direct identification of the patient, unless the images show a specific disease or deviation that is so rare and recognizable that it leads to a specific patient. In that sense, the images recorded are generally not able to directly identify a patient. However, they could still make a patient identifiable, if they can be linked to other information that could lead to the effective identification of the patient. For instance, if the recorded images are stored or in one way or the other linked to the patient's electronic health record, they could be traced back to a specific patient. Also if the filename of the recorded images or the metadata stored in it makes any referral to information that could identify the patient, the recorded images will be considered as personal data under the European Union legal framework regarding data protection.

It can therefore be found that caution should be paid to the broad scope of application of the current legal framework on data protection. Even when particular data does not seem to identify a specific patient, the accompanying metadata and the way in which the data is handled, could still make the patient identifiable. Such would lead to the conclusion that personal data is processed in the digital operating room and that therefore the *Data Protection Directive* will be applicable to such data flows.

3. Recording, Storage and Later Use of Data

One of the main aspects of future e-health developments is that the use of pervasive and ubiquitous network infrastructures in the digital operating room will lead to a substantial growth in the amount of data that is generated during surgery. Such data will then also become more easily subjected to storage thereof in centralized servers. When stored, data could potentially be used at a later stage, for a variety of purposes. For instance, laparoscopic images recorded during surgery could potentially serve educational purposes. Alternatively, such images could also be stored in the patient's electronic health record. Vital statistics of the patient during the course of a surgery could also be added to that same electronic health record to provide a more complete account of how the procedure went.

These aspects, however, all hold considerable concerns under the applicability of the legal framework on data protection. In the following, it will be analyzed how the applicability of the *Data Protection Directive* influences the recording, storage and later use of personal data in the context of the digital operating room.

3.1. Recording Data

With regards to the recording of data including vital statistics and surgical images during surgical procedures, the first question to be answered is whether the patient's specific consent is needed to this end. Specific consent for the recording of data during a surgical procedure could be found unnecessary because it is assumed to be part of the surgery itself, to which the patient already consented. Such idea of "one consent fits all" should of course be treated with care as consent principally needs to be specific [6]. The patient can therefore not be assumed to have given his consent to the recording unless he was already clearly informed on this when giving his consent to the procedure and if the recording would fit within the scope of the purpose of the procedure itself.

This strict interpretation of consent, however, does not accommodate the importance of research in advancing medical practice. Medical research – including personal data processing – can benefit public health by identifying patterns of diseases and finding new treatments [7]. To facilitate such later research, it could be argued to have the patient provide a broad consent aimed at providing a legal ground for future research. But as such future research may not yet be designed or performed for months or years to come [8], it becomes difficult to provide the specific information required by the patient in order to provide his informed consent. Indeed, the broader and more general the information given to the patient, the less informed his consent will be, thus no longer satisfying this requirement in personal data processing [9]. As a result, broad consent, despite its importance for the medical and scientific community, can be considered as problematic from a legal point of view.

Specifically asking consent for surgical recordings could also be found unnecessary if the patient is considered to be unrecognizable on the recordings made by, for instance, an endoscopic camera. However, as indicated before, there are other ways in which such information could make the patient identifiable, such as when the recording's metadata could be linked to the patient's unique health record. Such would still qualify the recordings as personal data, thus requiring consent or another justification ground before being allowed to be processed. It should therefore be stressed that consent is in principle required for the recording of data during any surgical procedure and that such consent cannot just be assumed from the patient's general consent to the surgical procedure in itself. As genuinely anonymous data is extremely rare, it would be advisable to seek specific consent for the recording of surgical procedures.

This means that the patient needs to be informed on the purposes of such processing, the duration of the storage thereof, etc. Given the benefits of an integrated approach, the patient's consent to the recording could be given at the same time as his consent to the surgical procedure in general, but needs to be clearly differentiated thereof.

3.2. Storage of Data

In past times, operating rooms could be considered as isolated islands, where during a surgical procedure nothing could get in or leave. The use of networked equipment in the digital operating room will provide a direct and constant connection to the outside network, thus providing opportunity to send and receive information in realtime. One possibility that can be envisioned here is the direct recording of surgical images and the storage thereof on the hospital's network. Such storage is a practice already found in hospitals today [10] and should therefore be addressed from a legal point of view.

First, this practice raises questions with regards to the duration of the storage of what can be considered as being the patient's personal data. As can be found in the legal provisions of the data protection directive, personal data collected for processing cannot be stored longer than necessary for achieving the purposes for which they were collected. This means that personal data storage needs to have a clearly defined end-point, after which the data needs to be deleted. Apart from the specific purposes of the processing, the end-point of health data storage will also be determined by other factors. For instance, Belgian law requires patients' files to be stored for thirty years since the last contact between the healthcare professional and patient [11].

An important factor in the usefulness of such data storage is the advancement of the medical state of the art. As surgical practices and procedures are continuously improving, it would not make sense to use a particular recording of a procedure for a long period of time, as it will eventually show outdated practices and procedures. It would therefore seem advisable to predetermine a specific duration for the storage of recorded surgical procedures.

On another note, the storage of what can be regarded as personal data also requires the implementation of specific measures aimed at safeguarding the security and confidentiality of such data. Security can generally be understood to include a number of aspects [12]. Integrity, for instance, ensures that processed information remains accurate and that no unauthorized modifications are made. Also, availability ensures that the data is readily accessible and usable. Additionally, one can refer to data origin authentication, which guarantees the origins of the data and non-repudiation, which ensures that actions committed cannot be denied by their performers [13].

The general obligation to ensure personal data security requires the data controller to ensure an appropriate level of security taking into account the current technical state of the art, the cost of implementing such measures, the nature of the personal data to be protected and potential risks. As the digital operating room includes the processing of health data, the appropriate level of security should be considered to be high. While the current non-digital OR also includes the processing of health data, it is precisely the advanced degree of interconnection between different devices in and outside of the OR that makes the Digital OR a more risk-bearing environment. Patient records are no longer physically transported from a secured archive to the OR when required, but can be consulted electronically at all times from anywhere in the hospital. Also the higher data flow resulting from the convergence and interconnection of equipment that is currently still used "offline" will augment the risk potential, for instance in terms of data breaches.

The *Data Protection Directive* calls for technical measures of security, which includes the physical protection of the personal data by ensuring that non-authorized people cannot get access to this data [5]. This is, however, an important problem in hospital settings, as most areas are open for public access and mobile devices are often not properly stored. Physical data security would therefore in this context also require a change in attitude of the actors involved. Therefore, more purely technical measures are also to be considered, such as protecting the devices and applications by encryption and passwords. Such would ensure that unauthorized people cannot get access to the personal data, even if they would get physical access to the devices containing or being able to access such data. More organizational measures include raising staff awareness and responsibility with regards to data security. As an obligation of means, the data controller is bound to deliver his best efforts rather than a specific result and must therefore demonstrate that he delivered the effort that another diligent controller would have delivered under the same circumstances.

Confidentiality requires the data controller to limit the access and processing competencies of the actors under his authority [5]. This duality requires the personal data to be off limits for unauthorized persons, but also holds that authorized persons cannot be given unrestricted access. In general, access to the personal data must be restricted to what the properly authorized persons need to know for performing their respective duties. For access provision, a regular authentication procedure can be followed [14]. This includes registration of the authorized persons, after which they can present their identification. Such identification can be made by information known only to the user such as passwords or by tokens only held by the user such as an identification card. Following the authentication verifying that the claimed identity is real the person will be authorized and granted access. Such authorization could be leveled, ensuring that a particular user is only granted access for as much as his role demands. Actors executing higher demanding roles will be given higher levels of access rights. Categorizing the patient's personal data can be useful in developing a modular access matrix. Logging and tracing mechanisms can be used to verify whether appropriate access levels were given and whether only properly authorized users accessed the data corresponding to their level of demand.

Additionally, with regards to education, employees should be instructed on their applicable organizational security policies and the importance thereof [5]. Given the particular status of healthcare work, employees should not only be instructed on general data protection requirements, but also on requirements stemming from their status as health professionals. One requirement is that health data must be obtained at the patient and can only be processed under the responsibility of health professionals, unless otherwise consented to. Also, given the importance of the networked infrastructure of the digital operating room, it is important to ensure that these networks are adequately secured in order to guarantee the security and integrity of the data transferred over them.

3.3. Later Use of Data

One of the main reasons to store data is to preserve the possibility of using such data at a later stage. According to the *Data Protection Directive*, personal data can only be used for the specific purposes for which it was collected. As a result, personal data collected for a specific and justified purpose cannot be used at a later stage for purposes that are irreconcilable with the purposes for which the data was first collected.

To judge whether the original and subsequent purposes of the data processing are reconcilable, all relevant factors need to be taken into account, in the first place the data subject's reasonable expectations. The difference between original use and later secondary use needs to be stressed in the context of the digital operating room as well. If, for instance, a surgical procedure is recorded for a specific purpose, then later use of those images will have to be reconciled with the original purposes for which the procedure was recorded. If such secondary use cannot be reconciled with the original purposes, the secondary use will have to be treated as a new processing, thus requiring the fulfillment of all data protection requirements such as consent, purpose statement, etc.

Further processing of data for historical, statistic or scientific purposes is principally not considered to be irreconcilable and will therefore be allowed, be it under specific conditions. To make this matter more concrete, the Belgian use case will be presented as an example of how the further processing of personal data can be regulated. Note, however, that this regulation may differ across the European Union. The reason for this is that there are no harmonizing legal instruments on this matter, apart from the *Clinical Trials Directive* [20].

In the Belgian use case, the Royal Decree of 13 February 2001, executing the *Belgian Data Protection Act*, deals with the concept of further processing for historical, statistic or scientific purposes [15]. In general, Article 3 of the Royal Decree prefers that anonymous data is used. As such data cannot be linked back to a specific data subject, it is by definition no personal data and therefore can be processed further. If anonymous data cannot suffice to satisfy the purposes of the processing, Article 4 of the Royal Decree calls for the use of encoded data. This is data that can be linked to a specific data subject, but only by means of a code. Only when also encoded data does not satisfy the purposes of the processing, Article 5 allows the use of non-encoded personal data.

Note that there are three scenarios imaginable [16]. If personal data is primarily collected for historical, scientific or statistic purposes as original purposes of the processing, the use of this data for these historical, statistic or scientific purposes is no secondary use and therefore all sorts of specific national regulations relating to the issue of further processing such as the Belgian Royal Decree will not apply. If the data is collected for other purposes and used in secondary order for historical, scientific or statistic purposes that are reconcilable with the original purposes, the Royal Decree will also not apply as there is not incompatibility between the original purpose and the purpose of the secondary use for historical, statistic or scientific purposes. The Royal Decree only applies when data is collected for specific purposes and later used secondarily for historical, scientific or statistic purposes that are not reconcilable with the primary purposes.

4. Device Manufacturer Regulations

Apart from addressing the main concerns resulting from the application of the principles of the *Data Protection Directive* to the developments of the digital operating room, this paper also aims to look at this matter from the perspective of the manufacturer of the devices that make up such digital operating rooms. In the following, it will be analyzed to what regulatory regime such devices and their manufacturers are subjected.

While general healthcare regulations are mostly aimed at establishing the rights and responsibilities of patients, medical professionals and medical institutions such as hospitals, one should not forget about the legal position of the manufacturers of the many products that enable or facilitate the provision of healthcare, including the devices and applications that will play a role in the digital operating room. The reason why these product manufacturers are typically not included in general healthcare regulations is that they normally do not directly engage in contracts with the patient. Medical professionals or institutions engage with product manufacturers through contracts spanning from regular sales of goods contracts to elaborate service contracts that are mostly governed by standard contract law. Direct contact between patients and product manufacturer is generally only found in certain cases of the manufacturer's liability for faulty products.

However, certain sectors apply specific rules to manufacturers that aim to bring products on the market in that sector. Especially in the healthcare sector, one can understand the need to preserve certain standards of quality. Surgical scalpels and hypodermic needles need to be fully sterile, monitoring and diagnostic equipment needs to be reliable, etc.

At the level of the European Union, a number of directives provide the basic legal framework that needs to ensure a high level of quality of medical devices in order to guarantee the protection of human health and safety. Such directives provide basic lists of requirements that need to be met before medical devices can be put on the market. When devices are marketed, they must also bear the CE mark as a proof of certification, although self-certification is possible in certain cases. Devices are divided over four categories (I, IIa, IIb and III) according to the risks their use poses to the patients. The criteria used for such classification take into account the invasiveness of the device, the intended duration of its use, whether the device is active or passive, etc. [17].

While such requirements listed do provide a basic idea of what one should be able to expect from a compliant medical device, there are virtually no technical details included. For instance, it is stated that devices delivered in a sterile state must be manufactured and sterilized using an appropriate and valid method, yet apart from a reference to standards developed by standardization bodies such as the International Organization for Standardization (ISO) it is left open to interpretation by the Member States to further define such method. As a result, Member States need to incorporate and further develop these requirements in their national legal system. A Competent Authority reporting to the Minister of Health will be formed in all Member States to monitor the adoption and application of these principles.

In Belgium, the Federal Agency for Medicines and Health Products (FAMHP) evaluates, approves, follows and controls the requests for clinical trials for medicines and health products. This agency follows medical devices and medicines from their R&D phase to their introduction on the market and performs inspections to ensure the quality of these devices and medicines. Every product manufacturer aiming to bring a medical device or medicine to the Belgian market will therefore have to apply to the FAMHP. To this end, medical devices are defined as any instrument, equipment, material or other article used on its own or jointly, including software required for it to function correctly, which is intended by the manufacturer to be used on humans for the purposes of diagnostic, prevention, control, treating or diminishing an illness, an injury or a handicap, of studying, replacing or modifying part of the anatomy or a physiological process and of controlling conception and whose principal intended action in or on the human body is not obtained by pharmacological or immunological means or by metabolism but whose function can be assisted in such a way [18]. This includes accessories specifically intended by its manufacturer to be used with a device to enable the use of that device in line with the instructions of the manufacturer of the device.

Given this broad definition, taken literally from the European Union directive, the scope of the regulatory competence of the FAMHP spans from the simplest of tools such as tongue depressors to much more complex diagnostic and monitoring devices and computer systems. The devices envisioned in the digital operating room and their accessories may therefore also have to comply with the existing regulations applied by the FAMHP.

Taken that a digital operating room could be defined as including the development of technologies for central external monitoring equipment and the network infrastructure that enables image distribution and collaboration, it will have to be assessed whether such can fall under the scope of the FAMHP. Here, Article 7 of the Royal Decree of 18 March 1999 refers to system manufacturers as "*all natural or legal persons reassembling devices with a CE marking, depending on their destination and the limitations of use granted by their manufacturers, in order to launch them as a system or a kit*". Such systems are subject to a mere notification and do not need to go through the whole certification procedure. However, if the system contains components that do not carry the CE mark or if they are used in a manner incompatible with their originally intended use, the system is considered as a separate medical device, thus subject to the standard procedure. As the digital operating room would integrate different medical devices into a central hub, such hub could be considered as a system. The envisioned network infrastructure for image distribution and collaboration will also be integrated in this hub and will be used in collaboration with medical devices, thus becoming part of the medical system. As the central hub in itself will not directly come into contact with the patient, it could be seen as a "Class I medical device".

Looking at the Belgian use case, one will have to refer to the Act concerning *Experiments on the Human Person* [19]. While this act is the Belgian implementation of the so-called Clinical Trials Directive [20], the Belgian legislator has chosen to expand the scope of the directive from "*clinical trials, including multi-centre trials, on human subjects involving medicinal products*" towards every type of experiment involving human subjects with the goal to expand knowledge on medical practices. As a result, every test, study or research involving human subjects that is aimed at expanding knowledge on the practices of health professions will be subjected to the scope of this act. Given this broad definition, one will have to assess whether trials concerning the digital operating room hub or other applications would constitute an experiment under the scope of the Act concerning *Experiments on the Human Person*.

While tests should be understood as referring to medicinal products, studies and research also apply to non-medicinal trials. However, nor the act, nor the preparatory works provide a clear definition of these trials. The act does, however, refer to medical devices [19]. Like trials involving medicinal products, studies and research focusing on medical devices should receive a positive advice from an ethical committee and from the Minister of Health. More concretely, it could be argued that one should follow the procedure stated in the Royal Decree on medical devices, which leads to notification to the FAMHP, as discussed before.

Two other conditions that need to be fulfilled for the application of the Act concerning *Experiments on the Human Person* include the goal to expand knowledge on medical practices and the involvement of human subjects. If the experiment is aimed at advancing the state of the art in medical practice, then the condition of knowledge expansion will be fulfilled. The condition of human involvement is fulfilled as soon as the experiment physically involves a born and living human subject. The mere processing of his personal health data, for instance, will not lead to the application of this act. When the Act concerning *Experiments on the Human Person* applies, the human subject participating in the study or research will have to grant his written prior informed consent [19]. He also enjoys specific protection, such as that that experiment needs to abide by the proportionality principle that risks and benefits need to be weighed off against each other, etc. Further responsibilities and liabilities are imposed on the promoter.

5. Practical Consequences

While the previous sections discuss the more theoretical aspects of this matter, the question remains what this means in practice. How are professionals in the telecommunications sector affected by the advent of telesurgery practices? Which dangers need to be heeded when engaging toward the implementation of a Digital OR solution? This section will summarily consider the practical implications of the evolutions discussed here. First, it is reminded that device manufacturers must comply with European and national legislation in order to deliver medical devices. Second, data protection concerns must be taken into account. Third, potential liability issues need to be minded.

5.1. Device Regulations

The manufacturers of the devices, tools and applications of the Digital OR must assess whether their product can constitute a medical device according to existing legislation in this field. As this legislation is not very much harmonized at the level of the European Union, it must be ensured that both the scarce European legislation in this field – for instance concerning the requirement to bear the CE mark – and the applicable national legislation of the Member States are complied with. In most cases, this will entail a submission to the competent national agencies concerned with monitoring medical devices and medicines. Only when the applicable rules and procedures are complied with and authorization – where required – is obtained, the medical devices can be offered to customers in the healthcare sector.

5.2. Data Protection

As noted before, the devices of the Digital OR are becoming more interconnected, meaning that devices that used to perform their tasks isolated from other devices are increasingly becoming part of a network of data exchanges. Even a simple heart rate monitor could be modified to record its readings – or anomalies in particular – and store them on the centralized hospital network. The result of this is that the data flows in the Digital OR are very likely - or even certain - to involve personal data processing operations. These data flows and the subsequent use of that data must therefore comply with the requirements of European and national data protection legislation.

More concretely, this means that it is important to determine who will serve as the data controller to that personal data processing operation, as such data controller will hold the final responsibility over the processing. This data controller will have to determine the purposes of the processing, the duration of storage, ensure that no data excessive to the purposes is processed, etc. Another pivotal element to a fair and lawful processing of personal data is that of the legitimate justification ground. While specific justification grounds do exist for use in a medical context – for instance in case of medical urgency – the patient's consent will undoubtedly serve as the most important justification ground. Health data is considered to be sensitive personal data and the processing thereof must therefore comply with stricter regulations. At European level, for instance, it is stipulated that consent for the processing of sensitive personal data must be explicit. National implementations of this provision, however, may differ. In Belgium, for instance, written consent is required [3]. Another duty of the data controller is to ensure that the patient's rights as a data subject are respected and that proper notification is made to the competent national *Data Protection Authority*.

The data controller is defined as the party to the processing that decides the means and purposes of that processing. Within the context of the Digital OR, this will generally be the surgeon, or even the hospital. While recent evolutions make it difficult to apply static concepts such as that of data controller to complex data processing operations, it is clear that this role belongs to a medical professional and principally not to the manufacturers of medical devices. However, the fact that the hospital and the health professionals principally share the burden of the task of data controller does not mean that other parties, such as the device manufacturers do not need to mind data protection rules. If these manufacturers become involved in performing the processing on behalf of another party, they could still be considered as processors. If, for instance, a medical device assists in the processing of personal data, its service provider could be viewed as a processor if his device only serves as a means for the processing. And if the device requires additional data to be processed, it could even be viewed to determine the purposes of the processing as well, thus leading to its service provider becoming a (joint) controller [4]. Device manufacturers are therefore advised to clearly define their role within the personal data processing operations their devices will become involved in.

5.3. Liability

The manufacturers of the devices of the Digital OR will also have to mind potential liabilities for their products. Under general contract law, these manufacturers are bound to a duty of conform delivery, meaning that their products need to be without visible or hidden flaws and that they must live up to the expectations of the product agreed upon. Especially in a medical context, devices will need to demonstrate a high degree of reliability.

Outside of the strict contractual framework, product manufacturers can also be held liable for damages caused by their faulty products. This product liability can be considered as an objective liability, as it does not require a fault on the manufacturer's behalf. The party suffering damages will only have to prove that those damages were caused by a fault in the product. Given the extra-contractual nature of this liability, it serves as a means for patient to direct a claim for compensation for damages sustained directly to the product manufacturers, as they will generally not have entered into a contractual bond with this party.

By converging different services into fewer devices, the Digital OR is a much more complex environment. Device manufacturers will need to adapt to these complexities and ensure that their products are compliant to the standards expected in the medical sector.

6. Conclusion

Technological developments such as the *Internet of Things* will soon make their way into hospitals worldwide. In what can be referred to as the digital operating room, different devices will become interconnected and will create, store and exchange data on a larger scale than has ever been possible before. Such data flows can, however, also pose concerns with regards to the patient's privacy. To this end, this paper has first analyzed the applicability of the current legal framework on data protection to the data flows that can be found in such digital operating room. Here, the focus was put on the concept of identifiability. As the *Data Protection Directive* requires the data subject to the identified or to be reasonably identifiable in order for data to be considered as personal data, it is precisely this concept of identifiability that can determine the true scope of the notion of personal data. Indeed, in this context it was found that data that is often considered not to identify a patient and thus to be anonymous can still be used to lead to the identification of a particular patient, when coupled with other data such as metadata or when linked to the patient's health record. The applicability of the *Data Protection Directive* should therefore always be assumed, given the broad spectrum of its applicability. In particular, this paper focused on the recording, storage and later use of data in the digital operating room. Here, it was found that such data recording principally requires additional specific consent from the patient. Also the storage is bound to particular requirements, such as that of limited storage duration and the adoption of specific security and confidentiality measures. When data is stored for future use, it needs to be ensured that such secondary use can be reconciled with the primary purposes for which the data was collected. Finally, with regards to the status of the manufacturers of the devices of the digital operating room, it was found that such devices can fall under the specific status of medical equipment, which means that they may have to comply with a number of specific requirements following from the sensitive nature of such equipment.

Acknowledgements

This research was performed as part of the *Telesurgery project*, funded by iMinds and the Flemish government agency for Innovation by Science and Technology (IWT).

References

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey", *Comp. Netwo.*, vol. 54, pp. 2787–2805, 2010.

[2] R. Weber, "Internet of things – Need for a new legal environment?", *Comp. Law Secur. Rev.*, vol. 25, pp. 522–527, 2009.

[3] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, *Official Journal* L 281 of 23 November 1995, pp. 31–50. In Belgium, this directive has been implemented as: Act of 8 December 1992 on the protection of the privacy against the processing of personal data, *Belgian State Gazette*, 18 March 1993 (Data Protection Act).

[4] C. Kuner, *European Data Protection Law*. Oxford: University Press, 2007.

[5] D. De Bot, *Verwerking van Persoonsgegevens*. Antwerpen: Kluwer, 2001.

[6] Article 2(h) Directive 95/46/EC.

[7] Academy of Medical Sciences, *Personal data for public good: using health information in medical research*, London, Academy of Medical Sciences, 11–12, 2006.

[8] T. Caulfield, R. Brown, and E. M. Meslin, "Challenging a Well Established Consent Norm?: One Time Consent for Biobank Research", *J. Inte. Biotechnol. Law*, vol. 4, pp. 69–74, 2007.

[9] V. Arnason, "Coding and Consent: Moral Challenges of the Database Project in Iceland", *Bioethics*, vol. 18, pp. 41–42, 2004.

[10] S. Elprama, P. Duysburgh, and A. Jacobs, "TeleSurgery D1.3.1 – Draft Desk and Key Observational Informant Research on Everyday Surgical Practices", *IBBT*, 2011.

[11] Article 1, 3 Royal Decree of 3 May 1999 defining the minimum requirements of the medical record, as defined in article 15 of the Act on hospitals coordinated on 7 August 1987, *Belgian State Gazette*, 30 July 1999.

[12] B. Van Alsenoy *et al.*, "GINI D3.1 – Legal Provisions for Deploying INDI services", pp. 29–30, 2011 [Online]. Available: www.gini-sa.eu

[13] S. Pearson and A. Charlesworth, "Accountability as a way forward for Privacy Protection in the Cloud", in *Proc. First Int. Conf. Cloud Computing*, Beijing, China, Dec. 2009, pp. 131–144.

[14] X. Huysmans, "Identity Management for eGovernment – Deliverable 1.2 Conceptual Framework", *IBBT*, 2005.

[15] Royal Decree of 13 February 2001 executing the Act of 8 December 1992 on the protection of the privacy regarding the processing of personal data, *Belgian State Gazette*, 13 March 2001.

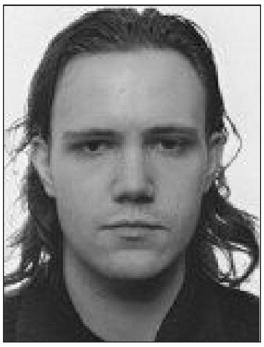
[16] J. Vandendriessche, "De verwerking van persoonsgegevens voor historische, statistische en wetenschappelijke doeleinden", *TBBR*, vol. 9, pp. 534–543, 2006.

[17] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, *Official Journal* L 169 of 12 July 1993, pp. 36–40.

[18] Act of 25 March 1964 on medicines, *Belgian State Gazette*, 17 April 1964.

[19] Act of 7 May 2004 concerning Experiments on the Human Person, *Belgian State Gazette*, 18 May 2004.

[20] Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use, *Official Journal* L 121 of 1 May 2001, pp. 34–44.



Niels Vandezande is a Legal Researcher at the Interdisciplinary Centre for Law & ICT – iMinds, Faculty of Law, KU Leuven-University of Leuven, Belgium. He obtained his Master of Laws degree in 2009 at the same university, focusing on ICT-law, international law and corporate law. He currently conducts research in the fields of data protection, privacy, security & trust and public information availability. For his Ph.D., he researches the legal value of electronic information.
E-mail: niels.vandezande@law.kuleuven.be
ICRI – KU Leuven
Sint-Michielsstraat 6 (B3443)
3000 Leuven, Belgium

Griet Verhenneman is a Research Associate at the Interdisciplinary Centre for Law and ICT (ICRI), KU Leuven since 2007. She is working on different Belgian and European research projects with a strong focus on eHealth. She is preparing a Ph.D. on the patient's right to privacy and autonomy in a changing healthcare environment, an environment based on disease management techniques supported by ICT tools. She is also a part-time professional support lawyer at Time.Lex law firm where she deals with eHealth and privacy protection.
E-mail: griet.verhenneman@law.kuleuven.be
ICRI – KU Leuven
Sint-Michielsstraat 6 (B3443)
3000 Leuven, Belgium

Jos Dumortier is a Professor in Information Technology Law at the Faculty of Law, KU Leuven since 1989 and the founder and director of the Interdisciplinary Centre for Law and ICT since its start in 1990. He is also the co-founder and partner of Time.Lex, a Brussels-based law firm specializing in commercial, ICT and intellectual property law. Prof. Dumortier has been closely involved in the drafting of the Belgian personal data protection law and works regularly as an expert for the European Commission. With his research team in Leuven and his law firm in Brussels he has authored many European studies in the field of information security and data protection law.
E-mail: jos.dumortier@law.kuleuven.be
ICRI – KU Leuven
Sint-Michielsstraat 6 (B3443)
3000 Leuven, Belgium

Paper

IPv6 Preparation and Deployment in Datacenter Infrastructure – A Practical Approach

Marco van der Pal

Generic Services – Network Infrastructure Services, Capgemini Netherlands B.V., Utrecht, The Netherlands

Abstract—This article describes the experiences with the initiative to introduce IPv6 into Capgemini's datacenter environment, and to be more specific, the part of the project known as Phase 1: the preparation before actually doing so. Phase 1 comprises of training, testing and research of the IPv6 protocol and its features with the purpose to better understand the consequences of the introduction of IPv6 in a datacenter environment. It was a specific choice to not deploy IPv6 in a production environment, and to build a dedicated test environment first (Proof of Concept). This test environment would accommodate most basic features of IPv6 to safely prepare us for the actual deployment. The technical results of the IPv6 experience were documented in a structured way, useable for future reference. Test results were also used as input to develop Capgemini best practices for IPv6 deployment.

Keywords—best practices, defining showcases, deployment conclusions, guidelines for deployment, setting up a test environment.

1. Introduction

Two years ago some network engineers of Capgemini Infrastructure Outsourcing Services in the Netherlands (IOS-NL) put their heads together to challenge themselves with the implementation of IPv6. They wondered what it would take to implement IPv6 in their IT infrastructure and decided to create a business case. Their main goal was – of course – to address the technical aspects of IPv6.
When thinking of the implementation of IPv6 one would primarily consider it as being a technical challenge. It probably is addressing functionalities. However there seem to be many more obstacles on the road, and these are considered as important as the technical.
IP connectivity in the Capgemini IOS-NL datacenter infrastructure is solely based on IP version 4 (IPv4). IPv4 has been de facto standard for internet communications all over the world for tens of years. IP address space is limited however, and we have reached its boundaries. IPv6 was developed to overcome the IPv4 address space limitations and other shortcomings. Capgemini IOS-NL initiated a project "Introduction IPv6" to investigate IPv6 and to learn about it.
The project will encompass many phases until the goal – datacenter and customer networks fully IPv6 – will be reached. Phase 1 was intended as a learning phase to familiarize ourselves with the numerous aspects of the protocol itself, as well as to experience the consequences of the introduction of IPv6 within our existing datacenter infrastructure. This first phase is also called the Proof of Concept (PoC) Introduction IPv6. Another reason was that this would show us the best approach to safely implement the protocol in a production environment. In Phase 2 we will thus introduce IPv6 in a part of the datacenter network (the edge) "for real". The experience of Phase 1 gives guidelines to several aspects of the implementation of IPv6, among which:

- IPv6 address planning,
- IPv6 assessment of existing infrastructure,
- best practices,
- security,
- IPv4 and IPv6 coexistence,
- migration paths.

2. IPv6: Phase 1

2.1. Background

IPv6 was introduced to overcome the limitations of IPv4: insufficient address space, lack of integrated security features, complex NAT constructions and more.
However, IPv4 and IPv6 are not compatible and can't talk to each other directly. Although the protocols resemble each other at first sight, their philosophy differs. Even more complex is that IPv4 and IPv6 will be running simultaneously and parallel to one another for years. These and other reasons have lead to the initiative to carefully explore the ins-and-outs of the IPv6 protocol in a separate environment, before even attempting to bring IPv6 into production.

2.2. Aim

The Proof of Concept aims to fulfil a number of major goals:

- to gain hands-on experience with implementing IPv6 in a data center environment,
- to gain the knowledge and confidence to be able to build a IPv6 enabled data center,

– to build a representative IPv6 test environment for further development, testing, knowledge transfers and training,

– to develop a new consulting service to customers, known as “IPv6 Audit”, to assess the client’s infrastructure for its readiness to adopt IPv6.

2.3. Starting Points

The starting points for the Proof of Concept Introduction IPv6 are:

- The introduction of IPv6 should be a collaboration between different Capgemini business units and disciplines:
  - IOS, Capgemini Infrastructure Outsourcing Services; the owner of the Capgemini datacenter infrastructure;
  - ITS, Capgemini Infrastructure Transformation Services, providing consultancy to clients;
  - APPS, Capgemini Application Services; provider of complex software applications like Oracle and SAP.
- To be able to examine all features and functionalities of the IPv6 protocol the project is (technically) multidisciplinary. The following solution teams are involved:
  - networking
  - Unix/Linux,
  - Microsoft,
  - applications,
  - consultancy.
- The technical aspects of IPv6 are investigated within a separate test (PoC) environment. The only allowed shared component is remote access to the PoC environment. This ensures that all testing does not in any way affect the customer’s production environment.
- Most of the used hardware in the test environment consists of surplus devices, thus available on short notice and resulting in low costs.
- The software used for the demo environment is either for testing or evaluation purposes or is open source.
- The IPv6 test environment is physically placed in a dedicated and easily accessible test room, and will remain available until further notice. This means that modifications to the test environment can be made without using formal procedures.
- The test environment is primarily virtualized to minimize the physical set of hardware.
- The showcases in the test environment have been defined as those that represent the majority of real life situations, but is not intended to cover all possible situations.

3. Goals for the Proof of Concept

3.1. What We Want to Achieve

The primary goal of the PoC Introduction IPv6 is to gain knowledge, expertise and experience of the IPv6 protocol and its features:

- Basic knowledge can be achieved by theoretical study and through training. The members of the IPv6 project team have had a generic 5 day IPv6 training.
- Expertise can be achieved by combining theory and practice, and testing in e.g. workshops.
- Experience can be achieved only by taking theory into practice, and by trying or running into uncommon situations as well. Troubleshooting is an essential part of gaining experience.

In order to gain experience with a complex concept like IPv6 one of the goals was building a dedicated test environment, specifically aimed at working with IPv6 in all its (technical) aspects.
To be able to continuously learn about IPv6 and its features several showcases have been defined. These showcases reflect real-life infrastructure environments, and for these a test environment was built.
Currently we are working on Phase 2 of the project: deploying IPv6 in the edge environment of the Capgemini IOS datacenter infrastructure. As a preparation secondary goals have been defined as follows:

– developing a service to assess an infrastructure for readiness to implement IPv6,

– develop a deployment plan to roll out IPv6 in Capgemini IOS datacenter infrastructure.

The outcome of the Phase 2 will be a recommendation for deployment of IPv6 in Capgemini’s datacenter infrastructure in a variety of possible solutions.

3.2. Setting up a Typical IPv6 Environment

To be able to test with IPv6 in a safe manner a separate test environment was built. The idea of having this dedicated test environment is also that it will be used for additional testing (e.g. new features), and for workshops and training purposes. To be able to test the relevance of IPv6 in a typical IT infrastructure, the following key areas have been defined:

– datacenter infrastructure,

– WAN connectivity,

– Internet connectivity,

– office infrastructure,

– IPv6 security.

Each of these areas consists of one or more of the following systems and functionalities that all together cover the most relevant scenarios:

- server systems,
- routers and switches,
- firewalls and general security,
- operating systems,
- client systems,
- Web services and applications,
- Internet connectivity,
- DNS,
- DHCP,
- IPAM,
- IPv4-to-IPv6 and coexistence.

Different scenarios have been defined to represent real life situations. These scenarios have been defined in showcases, which all together discuss most of the relevant key areas of IPv6, according to the needs of the implementation in Capgemini's datacenter infrastructure.

3.3. Showcases

Before we set up a test environment we first identified the key functionalities of IPv6 to investigate. These topics are considered to cover most relevant aspects of real life scenarios. The identified topics were bundled to address as many topics as possible in a single showcase.
These showcases demonstrate the capabilities of IPv6 and its basic functionality. The IPv6 showcases will teach us the similarities and differences with its predecessor IPv4.
The showcases have been defined to represent the majority of functions we're using in the Capgemini datacenters, as well as those that we expect to see in client environments.
An example of a defined showcase is to hosting an IPv6 enabled website in a Demilitarized Zone (DMZ).
One of the first services we expect to be requested by clients is to host an IPv6 enabled website. In our showcase we simulate both the home user who will surf to an IPv6 enabled website as the service provider hosting the IPv6 enabled website. In a variation on this environment we also try to simulate a corporate environment where the client is an office user instead of a home user. Corporate environments usually contain proxy servers, so we include that as well.
The following functionality will be implemented (Fig. 1):

- an IPv6 enabled home user with an IPv6 capable browser,
- DNS functionality to point the client to the website,
- IPv6 Internet connectivity for both the client and the web server,
- routing, subnetting and firewalling functionality for IPv6 at the datacenter side,
- a (simulated) WAN connection with IPv6 capability to connect the corporate client to the datacenter,
- an IPv6 enabled web server,
- an IPv6 enabled proxy server.

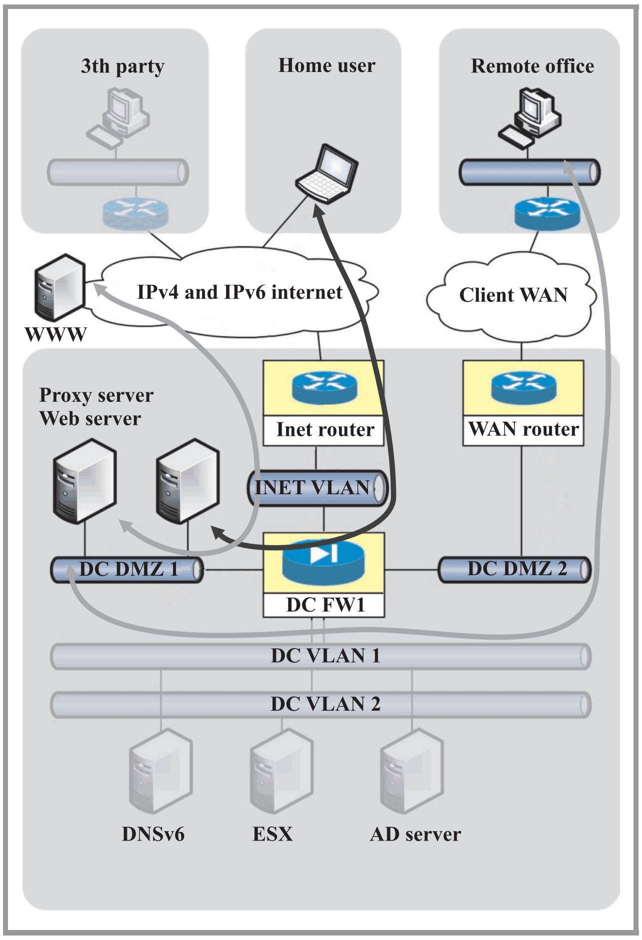


Fig. 1. Showcase example of Capgemini datacenter.

The expected results of this showcase are:

- Demonstrate that a home user can browse using IPv6 to a website hosted in Capgemini datacenter environment.
- Demonstrate this by either using a website which shows which IP address the connection is coming from, or by showing log files. In the alternative case where a proxy server is involved, show the logging of the proxy server as well.

One of the advantages of this approach is that multiple combinations of IPv4 and IPv6 functionality can be tested within one showcase. We accomplished this by pre-provisioning a set of VLANs within every area (Remote office, DMZ, production VLANs, etc.) according to the following VLAN plan.

Table 1
VLAN overview

VLAN types		Global IPv6 Address
1	IPv4 only	
2	IPv4 + IPv6 ULA	
3	IPv4 + IPv6 Global	Manual
4	IPv4 + IPv6 ULA + IPv6 Global	SLAAC
5	IPv6 ULA	
6	IPv6 Global	SLAAC + DHCPv6
7	IPv6 ULA + IPv6 Global	DHCPv6

When testing a specific scenario it is easy to look at the same from a slightly different perspective by just moving the (VM) system to another VLAN. This will give extra meaning to the way IPv4 and IPv6 coexist in e.g. a dual-stack environment (Table 1).

3.4. Documenting Knowledge and Experiences

Since the primary goal of the project was to gain knowledge and experience with IPv6 and its features, it was of high importance to document as much as possible. This doesn't mean however to rewrite existing documentation from vendors. Every showcase was built upon vendor documentation for a specific subtopic of the entire showcase. The intention of our documentation was in bringing it all together. That would result in the following:

Design Principles. When proposing and building a (client) solution it is advised to follow design principles. These will make sure that the solution fits in a broader range of Capgemini concepts and adaptations, and will thus increase a successful delivery.

Best Practices. Every showcase would give specific results in specific situations: these were gathered to form a set of best practices, substantiated with the reason why to use it in that situation.

Do's and don'ts. When deviating from best practices it is of major importance to know what another solution would lead to. The do's and don'ts gives handles in these situations.

The documentation principle was summarized in the following statement *If it isn't documented, it didn't happen!*

To make sure all testing in the several showcases was documented in a uniform way, a test protocol was used. This protocol guides you through a series of steps to document the following:

- To document feature testing in a unified and consistent way.
- To guide the test owner through the deployment of every feature the following must be documented per test:
  - which feature is tested,
  - what are the test requirements,
  - explain starting position,
  - explain expected results,
  - describe test results with, e.g., screenshots, configurations, etc.,
  - describe unexpected behaviour,
  - describe adjustments (if any),
  - what are the advantages/disadvantages of the solution,
  - conclusion (in terms of usability, manageability, scalability, future proof, difficulty, etc.).
- To be able to use the test results in future deployments and/or to recreate the test for future use.
- As input for a Capgemini blueprint for IPv6 in the datacenter
- As input for the development of in-house trainings.

To make sure all documentation was gathered and easily accessible a Wiki was set up. The Wiki contains all project documentation, technical details about the test environment, test results, best practices, etc.

3.5. Knowledge Transfer to Capgemini Staff

The IPv6 protocol is not solely a networking technology, hence the reason to put together a project team with members from different technology areas. However, specific IPv6 knowledge can be relevant in one team only, or in multiple teams with a different main point.
The idea is to gather specific IPv6 knowledge for all technology areas, with generic theoretical knowledge for everyone, and more specific knowledge for individual teams or groups of people. The result would be a set of IPv6 technology training modules, of which some are mandatory for some technology team, and other modules would be optional.
The knowledge and experience gained in the project and in future projects will be used to develop a training program specifically aimed at the target audience. Theory and practice will be combined with use of the test environment to visualize the different topics.
In this way theory and practice will accumulate to maximize the learning effect. Individuals interested to gain more expertise on the subject are eligible to use the existing test environment. One of the conditions would be that any new experiences would be documented and shared.

4. Conclusions

The experience and the lessons learned in the proof of concept have lead to the following major conclusions:

IPv6 is a very complex protocol. It differs from IPv4 in many areas, most important its philosophy on the use of IP addressing functionalities. Another important aspect is the compatibility issues that arise when IPv6 wants to communicate to IPv4 and vice versa. All possible solutions to overcome this issue have their technical limitations and/or are limited on, e.g., scalability or manageability.

IPv6 is evolving. New technologies are being developed, from which many address the compatibility issues between IPv4 and IPv6, primarily aiming at migrating towards IPv6. Also standards have changed in the last decade, which means that different implementations for same functionality coexist, possibly resulting in different behaviour.

IPv6 addressing is very much different from that in IPv4. Therefore IP Address Management (IPAM) and DNS are most important, not only in a normal datacenter or office environment, but moreover in, e.g., Cloud environments, where provisioning and control are key.

While working on IPv6 even more IPv6 topics will show to be interesting. It is advised to investigate these other topics as well, or at least examine them for relevancy. Besides of that, and as stated: IPv6 is evolving, so it may be relevant to keep an eye on other new (migration) technologies as well.

IPv6 is gaining popularity in the internet community. There may be other arguments that will require speeding up deployment of IPv6 in the Capgemini datacenter environment. A phased approach will give insight in deploying IPv6 in a real-life environment, and will pave the way for gradually deploying IPv6 across the entire datacenter infrastructure. This approach gives the time to learn and gain experience, even before they are sorely needed.

The project Introduction IPv6 was an infrastructure project only, mainly focusing on network, systems and OSs. This enabled us to investigate some of the IPv6 topics, but some have not had the attention they need. It is suspected that many applications and scripts have IPv4 hard-coded in the source code. It is thus to be expected that most or all of these applications will not work in an IPv6-only environment. If it is possible to work around these incompatibilities or whether the source code needs an update is unknown. It is highly advised to cooperate with the applications team to further investigate the possibilities with applications like these.

If a client requests to deploy IPv6 in the network infrastructure, one needs to know the possibilities and support of IPv6 in the current infrastructure systems and applications. Is IPv6 supported at all, and if so, which features do work, and which don't? Also make sure these features work in conjunction with one another, and be sure to examine the subtle distinctions of a vendor stating that their implementation supports IPv6. What are the demands on the current infrastructure design? What is the expertise of the IT staff? What home-made applications are used? And also what would it need to migrate towards IPv6? Expert consultants are needed to address these questions. It is therefore advised to investigate the possibilities to develop new services that can address these client demands: an IPv6 Audit to assess the network infrastructure and related services, and IPv6 Migration Services to accompany a client to successfully deploy IPv6 in the network infrastructure.

Because of the complexity and philosophy of IPv6 it is highly recommended to already start with setting up a training program to technical staff. Make sure that the training is fit for the attending people, e.g., a Windows admin does not require IPv6 knowledge on OSPF. A properly set up program will make sure that IPv6 philosophy is already in the mind of people and that the complexity will not be underestimated when running into IPv6 in real-life. Another aspect is that non-technical staff should be made aware of IPv6. Think of increased complexity and management efforts, the migration process and corresponding timeline, IPv6 support in (new) hardware and software, procurement, etc.

The PoC "Introduction IPv6" gave us a peek only at the possibilities of IPv6 and its features.
This peek was enough to experience some very basic functions of IPv6, and gave us enough ideas about complexity and possible issues. However this was only the first phase of a complete project with many phases to reach the ultimate goal of having a datacenter and customer networks full IPv6 deployed.
Finally, do not underestimate the efforts to be taken to deploy IPv6. For a good understanding of the possible obstacles on the road, read the paper of some engineers of Google who are migrating their enterprise network towards IPv6 [1].

References

[1] H. Babiker, I. Nikolava, and K. K. Chittimaneni, "Deploying IPv6 in the Google enterprise network. Lessons learned" [Online]. Available: http://static.usenix.org/events/lisa11/tech/full_papers/Babiker.pdf

Marco van der Pal is a network Consultant with Capgemini in The Netherlands. He has over 15 years of networking experience. He is member of the datacenter network team, and as such responsible for a network stretching 3 datacenters. He is working on network architecture, design and implementations. Marco van der Pal started his career as a network engineer, primarily on the equipment of Cisco Systems, but also on 3Com, NBase, Juniper and HP Networking. He achieved CCIE certification on Routing & Switching in 2000, and is now CCIE Emeritus. Mr. Van der Pal has been working on IPv6 for 3 years.
E-mail: marco.vander.pal@capgemini.com
Generic Services – Network Infrastructure Services
Capgemini Netherlands B.V.
Utrecht, The Netherlands

Paper

The Goals and Benefits of Network Modelling in a Commercial Environment

Edward Smith

British Telecom, Bracknell ATE, Bracknell, United Kingdom

Abstract—Modelling is a technique that can be used in the development and sizing of networks. It involves building a mathematical representation of a physical system and is often computer based, allowing system behaviour to be investigated based on the variation of key parameters. This paper will describe some applicable tools for network modelling and consider a range of applications. Whilst modelling and simulation tools are used extensively to design and size ICT solutions, each tool tends to be targeted at a specific range of applications and a specific user community. It is shown that different tools expect different levels of expertise from their user communities. The benefits of using such tools are described, as are some of their limitations. Modelling and simulation provide a cheaper and faster approach to trying new ideas than building prototypes or test networks. They provide a valuable guide to designing networks and can predict operational outcomes.

Keywords—emulation, modelling, performance, service quality, simulation.

1. Introduction

This paper will look at modelling as a technique for use in the development and sizing of networks. It will examine what the basic objectives and goals of modelling are, how the approach delivers benefits and identify the broad principles behind modelling. This will be illustrated using specific applications.

An investigation of the broad applications of modelling in communications systems is given and a description of applicable tools for network modelling provided. A range of applications for these tools will then be considered. This will be followed by some general observations and conclusions about the use of these techniques.

Mathematical modelling has a long history; one example of a mathematical representation of a physical system is the use of the techniques developed the Marāgha School to evaluate geocentric models of the Solar System and their later use by Copernicus to develop the heliocentric model [1]. It is applied in a number of circumstances such as modelling: complex systems based around differential equations [2], financial systems [3] and weather systems [4]. The role of computer model can be summed up as depending on: "thousands and thousands of repetitive calculations based on explicit and deterministic rules. Just what a computer is good at" [4].

Within the networking field, network simulation can be performed either by mathematically evaluating the interaction between the network entities or capturing and playing back data gathered from a production network. This allows the behaviour of the network to be monitored as network parameters are varied in a controlled manner [5].

2. Applications of Modelling

This section of the paper will consider three examples of where modelling has been successfully applied to problems associated with ICT propositions, these are:

- Workforce Management applications for contact centre implementations,
- modelling for investment decisions in network service propositions,
- support of Service Level Agreement (SLA) management in Information and Communications Technology (ICT) proposals.

2.1. Workforce Management

Workforce Management packages are specialist simulation tools used to forecast future workload and plan staffing schedules. More general process modelling tools such as SIMPROCESS [6], used to model business process re-engineering projects, have also been used to model the principal activities of a contact centre. Using raw data (call volumes, service times, routing data, schedules and financial data) collected from live systems to build the models, they were able to characterise each customer contact by what the customer wants. What if scenarios were modelled for example; to identify the impact of variation of staffing levels on call abandonment rate. It was found that quite small variations in input conditions can have a significant effect. The benefits of the approach are that it gives: a deeper systems understanding, a view of the impact of variations in inputs and it allows bottlenecks to be identified making process improvement possible.

Workforce Management uses data captured from a contact centre's Management Information applications, to model overall performance and hence ensure that sufficient agents are provided to handle the projected customer calling patterns. The goal is to provide best quality of service, using optimal staffing at the lowest cost. Historical data is obtained from the ACD System (Automatic Call Distribution) and real time information from sources such as CTI (Computer Telephony Integration) and diallers [7], [8].

Workforce management packages are usually used by large organisations, but less so by smaller organisations [9]. It is argued that [10] the demand for computer simulation analysis is driven by increasing traffic complexity and the move to skills based routing; the change in operations due to mergers and acquisition activity, business volatility, outsourcing and multiple channels and cheaper, faster desk based computing, combined with specialised contact centre applications.

The key driver for investing in this type of capability is that personnel account for 60 to 70% of a contact centre's costs and effective management of this resource is consequently a priority [11].

2.2. Modelling Investment Decisions

Models have been developed to forecast the profitability of network services, based on the underlying infrastructure and the subscriber population and by linking this to network performance and customer satisfaction [12]. Service quality is expected to be driven by network QoS, network availability and customer care, all of which can be modelled and used to identify the gap between perceived performance and the customer's expectations. Service profitability was found to be dependent on: price, the customer's subjective view of performance, the number of customers willing to re-purchase the service and the new customers the service provider can attract.

Customer satisfaction is a major driver of a customer's intention to repurchase which is a Baynesian decision process. It is subjective, whereas quality is an objectively measurable attribute specified as the sum of service quality and perceived quality. Customer satisfaction is modelled through the interaction of perceived utility and expectation and is a perception function. The profitability of a service is given by revenue generating potential after costs and service penalties have been subtracted.

It is observed that that network upgrades are normally carried out based on past experience and rule of thumb and that there is a gap between network planners and business analysts, who have a very coarse understanding of how improved network performance leads to future revenue generation [12]. The model relates infrastructure to profit as a function of performance, customer behaviour and market dynamics. Rather than maximising service quality, it is more important to ensure that the service quality meets customer expectations.

2.3. Modelling Service Characteristics

Most broadband implementations have no scope for re-routing in the network for consumer broadband access lines and therefore physical intervention is often required to fix faults [13]. Such faults are often fixed by moving the customer from one port to another, however some DSLAMs (DSL Access Multiplexers) are located at unmanned sites due to equipment and operational costs. These require a visit to fix, creating an issue of man power availability, which may be offset by optimising operative travelling time.

An algorithm has been developed to identify optimum travelling time, thereby reducing operational costs [13]. This ranks the relationship between decision alternatives and setting performance thresholds for several operational parameters. Resolution priority may be given to some high priority clients. Inputs into the algorithm are: the customer category (related to the value of the aggregate business from that customer), prospective penalties, service criticality, type of service and any existing service complaints. Prospective penalties are seen to increase with increasing Quality of Service requirement, bit rate, service type (i.e., VoIP is more sensitive than IPTV which is more sensitive than the Internet). In addition the algorithm makes use of key cost drivers: material cost (which includes operative travelling time), human cost, penalty payments, lost revenue and bad publicity.

The results of the algorithm may be viewed by a number of service restoration options: faults that have exceeded the critical threshold, locations where there are clusters of faults, locations where specific services are clustered, locations with more affected services, locations with more unavailable services affected by a customer complaint.

The possibility of routing around link failures and commencing re-configuration before a fault occurs, thus preventing packet loss, has been considered [14]. An approach based around machine learning from historical activity and correlation with real time events is used. Network failures have a clustering property which is used in the analysis, as was confirmed by looking at a tier one ISP for 8 weeks. It is possible to associate the failure of other links with that of a single link failure, using historical data to formulate the rules for prediction. However this identifies the symptoms and not root causes of failure.

Frequent failures and therefore clusters that fail within a short time of one another can be identified. It is possible to monitor alarms and use a decision engine to trigger the network reconfiguration process, which initiates changes to routing metrics which force the network to re-converge away from the projected failure. The level of confidence and therefore applicability of rules changes with time and consequently it is possible to disable a rule when the associated level of confidence drops below a threshold value. The approach can be validated using NS 2 modelling.

3. Tools For Network Modelling

Network modelling tools can operate at a number of levels, one example of a tool for modelling networks at the physical level, to aid the sales process is British Telecom's (BT)

SPEED tool, which is designed to both improve the speed at which customer data network quotes can be delivered and to reduce the level of expertise required to provide them [15]. It incorporates best practice design principles and makes use of a catalogue of the up to date product and service offerings with their current tariff and discount structures. One of its major differentiators is the sophistication of the business logic, its associated data and the expertise that is modelled internally.

The tool captures requirements in the form of site location information, which is plotted onto a GUI (Graphic User Interface) derived map, allowing traffic flows, CPE (Customer Premise Equipment), link aggregations, PoP (Point of Presence) locations, backup arrangements, service requirements and contract details to be incorporated into the model's driving parameters. The appropriate routers are identified down to sub component level and appropriate pricing generated. Not only does the package deliver pricing and design material, it also generates appropriate bid governance documentation.

SPEED has been a key component of BT's arsenal of sales tools and has been seen as a considerable success in shortening the sales process, reducing the need for specialist staff and standardising solutions. Although typically used for networks of a few hundred sites, it can handle networks containing several thousand nodes.

A number of authors have looked at the accuracy of sophisticated modelling tools, when compared with actual model operation [16]–[20]. Simulations may be defined as any action that mimics reality [16] and computer based models execute a set of rules to mimic a system. The major independent variable which drives most models is time and this is underpinned with other variables known as dependent variables that are functions of time. It is the nature of these variables as discrete or continuous, which define the two major forms of model.

The dependent variables, associated with a continuous model, are known as state variables which are expressed in the form of differential equations that ideally can be solved by analytical means, but often have to be evaluated using simulation techniques. Discrete event simulation models are based on a system containing objects or entities, with the state of the system being defined by in terms of the numeric values associated with those entities. The simulation is then often formulated in terms of an overall process, comprising a number of activities, which are triggered by specific events. Typical modelling entities of such simulation tools are shown in Fig. 1.

One application of modelling tools is the validation of network designs and the validation of network configurations. One tool aimed at providing the ability to simulate network designs in order to optimise capacity and performance, could emulate up to 2000 routers and 1500 subnets and constructed the network model from router configuration files adding traffic analysis and device utilisation modelling to the network configuration and routing functions through additional software. Traffic data could be input from RMON (Remote Network Monitoring) probes, router accounting, MIB (Management Information Base) information and Netflow.

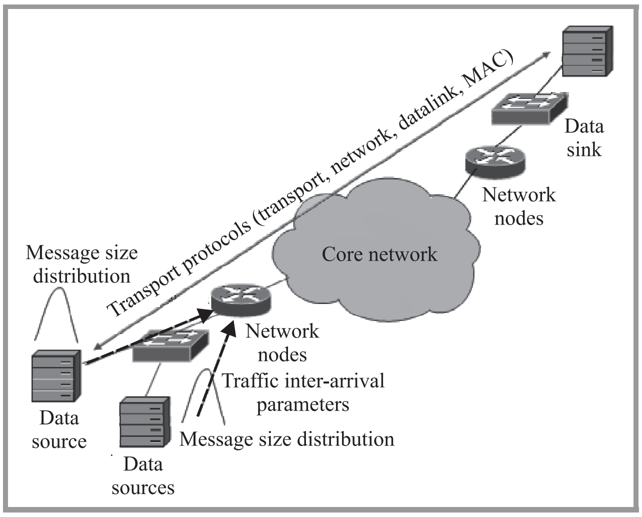


Fig. 1. Typical modelling entities and parameters.

GNS 3 is a graphical network simulator, distributed under a GNU license, which can run under Windows or Linux, can emulate various Cisco devices running a native IOS (Internetwork Operating System) version and is made of three components:

– the Cisco IOS emulator, which will execute a native IOS binary image,

– a user friendly text based layer,

– a graphical environment, which allows model networks to be built.

The capabilities of the tool are limited by the power of the PC that runs the emulator, but sizeable networks can be built of 20 to 30 routers, allowing the tool to be used extensively for training purposes and for preparation for certification exams. Networks can be built that can transmit a thousand packets per second, which is much less powerful than a modern router that could transmit up to a thousand times more. The current version of the tool will accommodate a wide range of routers, but not the whole Cisco range and does not appear to support the very latest models. Once the network has been built and set running, it takes 100% of the PC CPU.

The graphical tool allows the user to select a router or firewall and to configure that device with the desired interface cards. Links between devices can then be created, using the graphical user interface, to allow a network topology to be built. The user interface into the simulator is provided using a TN3270 emulator, with all commands available on the IOS image being used, being available to the user. There are a number of options available to allow the emulation of user traffic and packet capture from the emulated network can be achieved using Wireshark software.

One tool widely used for modelling both networks and protocols is the open source NS 2 package [20]. This is a discrete event simulator that provides functionality for performing: protocol design, traffic studies and protocol comparison. Tools like this overcome the scale limitations and cost of the test laboratory and can provide a more in-depth understanding than traditional analysis can. This package has a simulation core supported by visualisation tools, traffic and topology generators and a trace analysis capability. These are underpinned by a capability to support a wide variety of traffic models, applications, protocols, routing techniques, queuing mechanisms and physical media.
Models are constructed using a series of steps that start with creating the event scheduler, building the network, computing the routes, creating the traffic (inserting errors if required and tracing. Traffic can be created using traces taken from a real network.
Other widely deployed modelling tools are the commercially available COMNET III and OPNET modelling tools. These are discrete event simulators and are widely described in the literature [16]–[21]. Such tools typically provide a graphical user interface for building network models and setting up the model parameters and can model the operation of a network and its protocols and the scheduling of applications on end systems. A model typically tries to incorporate several characteristics of a design problem and when an optimal solution is obtained, the values of the decision variables can be used to evaluate design choices.
Whilst both COMNET III and OPNET offer very sophisticated modelling tools, these are predominantly used by core network designers and researchers, who specialise in network modelling. Those designing customer networks rarely possess traffic projections that are sufficiently accurate to warrant the deployment of these sophisticated modelling tools, indeed many networks are designed based on either the bandwidths deployed on the networks they are expected to replace or on the customer's best estimate of the traffic levels to be carried.

4. Applications of these Tools

A survey of the recent literature suggests that much of the work with modelling tools is focussed in the area of radio networks, particularly Mobile Ad Hoc Networks (MANETs). However, the technique is being applied to a number of other significant areas such as:

– the modelling of the impact of Intrusion Detection and Distributed Denial of Service attacks,

– modelling of video transmission across Wireless LAN infrastructures,

– simulating QoS models and traffic congestion scenarios,

– evaluating the packet loss probability in voice over IP networks,

– examining traffic bandwidth constraints in networks which deploy DiffServ Traffic Engineering techniques (DS-TE).

Modelling tools are not only used for design and research purposes, they can also be effective teaching tools, allowing the student to carry out hands-on experiments, without recourse to the costly provision of extensive teaching networks. This allows essential networking concepts to be taught to students.
It is sometimes desirable to model the impact that adding a new application to an existing network has on that network. This requires the network planner to identify the individual transactions that make up the interactions between the user and the new applications and superimpose these on a model of the existing network.
The network model is typically constructed by:

- Using packaged software to build a model of the switches and routers deployed in the target network or a simplified representation of that network.

- Loading appropriate live device configurations onto that model.

- Setting the utilisation of that network based on peak and average link utilisation figures extracted from the live network. This information may be loaded from files prepared using standard SNMP polling packages.

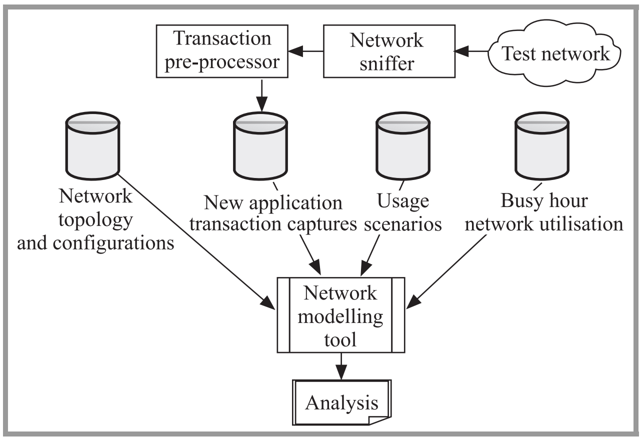


Fig. 2. Transaction profiling.

The transaction profile of the new application is normally captured from a test network using LAN capture software such as Sniffer or Wireshark LAN data capture software on a suitable portable device. Once captured, this information is fed in to another specialist application to provide a set of transaction representations that may be fed into the network modelling tool.
This coupled with estimates of overall transaction volumes, including numbers of simultaneous users, allows the network manager to see the impact of the new application on overall network performance.

BT Algorithmica is a division within BT that is concerned with network modelling and optimisation. This team can develop a set of algorithms to provide an exact fit to any network problem under investigation. They have modelled performance of BT's 21st Century UK network and the global MPLS (Multiprotocol Label Switching) platform.
The tool set is based on three components delivered by OPNET, which form layer zero of the modelling capability, to this Algorithmica adds a layer of sophisticated software libraries, scenarios and a SDK (Systems Developers Kit). The *scenario bank* can be used to generate the network model. The tool is being used in for example: the cell planning of WiMAX networks, 21CN infrastructure modelling and the designs for Next Generation Access.

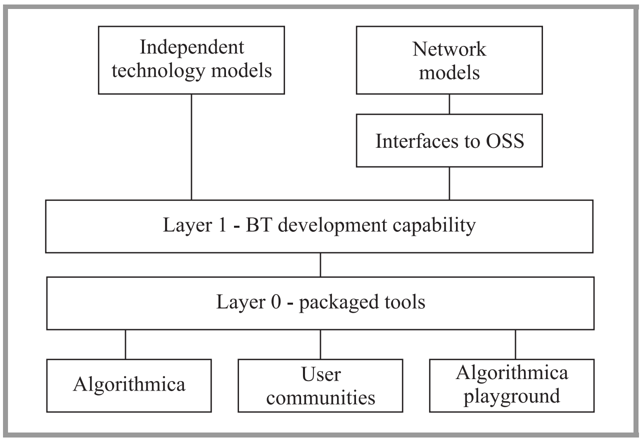


Fig. 3. Overview of the BT Algorithmica Suite.

Network simulations are built using:

– a network model built from existing router configurations,

– traffic based on MIB data collected using an industry standard SNMP MIB poller,

– a transport layer built from the inventory and topology of BT's underlying SDH and wavelength layer core networks in Europe.

Where overseas networks have been modelled, this can look at choices such as: whether infrastructure should be self-built or bought in from a third party, what technology and suppliers should be used, tariffing decisions and PoP location.
Increasingly network customers are seeking network performance guarantees and hence modelling of this is a concern for network providers. It has been demonstrated that queuing delays for such traffic systems can be evaluated [22] by assessing a G/M/1 queue with a Pareto based input. The analytical average queuing delay of the self-similar traffic can be compared to the delay from a simulated model, to obtain a useful model for delay prediction [23]. In this case the self-similar model was generated by loading a packet rate and packet size derived from trace data into the OPNET modelling tool to generate traces for comparison with the live data.
The OPNET package provides traffic simulation tools that simplify traffic analysis. These have been used in a number of studies of self-similar traffic, for example to simulate a priority based queuing system for DSCP (Differentiated Services Code Point) based QoS models [24] or to generate traffic traces which mimic live traffic [25]. These studies suggest that in moderate and high utilisation models, provisioning based on traditional methods may be over optimistic and may result in under estimating network resources like buffers and CPU capacity when trying to provide QoS guarantees (i.e., queuing delays, buffer overflow probabilities and packet loss rates) to the corresponding class traffic. They further show that the timescale for observation does not change the behaviour of the traffic pattern, whilst multiplexing makes peaks worse and uncontrolled heavy peaks cause traffic loss and jitter [25].
On some occasions the behaviour of protocols, which have not been implemented in commonly available network devices needs to be tested. One field where this is common is in the simulation of protocols associated with MANETs where protocols can more readily be emulated using a network simulation tool than they can be created with a routing device. In these cases the implementation of routing protocols, in radio networks where the network nodes themselves are capable of moving relative to each other, can be emulated [26].

5. Observations and Conclusions

This paper has looked at modelling as a technique for use in the development and sizing of networks. It has examined the basic objectives and the broad principles of modelling, illustrating how the approach delivers benefits by reference to specific applications.
The tools described cover a number of academic studies and direct commercial applications. It is arguable that the most widely deployed commercial tools are work force management and other contact centre dimensioning software, which looks at how the available work force, expected voice traffic arrival patterns and call patterns affect contact centre performance. These facilities whilst set up initially by network specialists are often run by the business managers responsible for running the contact centre.
Tools such as SPEED are frequently run by large numbers of sales engineers, who have significant network design expertise, but little in the way of formal modelling skills. These tools isolate the designer from the complexities of the core network and instead focus on providing a costed network based on a VPN (Virtual Private Network) core, which is treat essentially as a black box with fixed performance characteristics. Application dimensioning tools tend to be used by a smaller number of specialists, who have more modelling knowledge and are more interested in the characteristics of the client VPN, than they are the core network. Such capabilities are usually used to model the

impact of a new application on the existing network or to track down application performance issues.

Large scale discrete network simulation tools tend to be used widely as a research capability or, by some network providers in a commercial environment, as detailed design tools. Such tools are both powerful and versatile but tend to require support from small teams of design staff with more specialist modelling knowledge.

Whilst network simulation tools provide a powerful means of analysing system performance ahead of design or to compare alternatives, they do have their limitations [27]. Errors in simulations or improper data analysis can produce misleading results. In some cases different packages produce different results; hence it is important that practitioners and researchers document which parameters and which modelling tool are used, in order that others can replicate their results. It is important to validate simulation models against the real world. It has been suggested that 85% of test results in the published literature cannot be repeated by other authors, because key parameters of the model used are not reported.

Further issues pointed out [27] were:

– the statistical significance of the results collected,

– results should only be collected when the model has reached steady state,

– the simulation assumptions should always be clearly documented,

– the appropriate selection of traffic flows,

– where a given factor is varied, a sensitivity analysis should be performed.

In a more commercial context: a number of communications equipment providers produce models for sizing their more complex offerings. These can come with limited explanation of how they determine dimensioning and are quite capable of delivering results, which have unwelcome cost implications, impact on project delivery or are counter intuitive. Further the engineer has limited information on either the track record or accuracy of such models, as a predictor of real behaviour. This can pose a problem to the designer, who may end up having to explain the results delivered by such models, without having the detail needed to provide the justification required by an unhappy customer.

In the final analysis modelling provides a cheaper and faster approach to trying new ideas than building prototypes of test networks. It provides a valuable guide and can predict other outcomes; it does however have its limitations.

References

[1] J. Al-Khalali, *Pathfinders: The Golden Age of Arabic Science*. London: Penguin, 2010.

[2] I. Peterson, *The Mathematical Tourist*. New York: W. H. Freeman, 1988.

[3] B. B. Mandelbrot and R. L. Hudson, *The Misbehaviour of Markets*. London: Profile Books, 2008.

[4] I. Stewart, *Does God Play Dice?* London: Penguin, 1997.

[5] http://en.wikipedia.org/wiki/Network_simulation (accessed 14-6-2012).

[6] O. Tanir and R. J. Booth, "Call Center Simulation in Bell Canada", in *Proc. Winter Simul. Conf.*, Phoenix, AZ, USA, 1999, pp. 1640–1646.

[7] http://www.cisco.com/en/US/solutions/collateral/ns340/ns394/ns165/ns45/ns14/net_brochure09186a00800a3296.pdf (accessed 14-6-2012).

[8] https://www.vtrenz.net/imaeds/ownerassets/681/IEX_TotalView_Central_3.12Version.pdf (accessed 14-6-2012).

[9] W. Durr, "Spreadsheets and Workforce Management – An odd couple" [Online]. Available: http://verint.com/contact_center/resources/files/256/Odd_Couple_US.pdf?dt=wp (accessed 14-6-2012)

[10] V. Mehrotra and J. Fama, "Contact Centre Simulation Modelling: Methods, Challenges and Opportunities", in *Proc. Winter Simul. Conf.*, New Orleans, LA, USA, 2003, pp. 135–143 [Online]. Available: http://www.informs-sim.org/wsc03papers/016.pdf (accessed 14-6-2012).

[11] http://www.aspect.com/performanceedge/brochures/PerformanceEdge_Workforce-Management_Brochure.pdf

[12] J. Xiao, and R. Boutaba, "Assessing Network Service Profitability Modelling from a Market Space Perspective", *IEEE/ACM Trans. Netw.*, vol. 15, pp. 1307–1319, 2007.

[13] M. J. Alarcón, D. Andina, J. B. Grau, and J. Torres, "Service Assurance in fixed broadband networks", in *Proc. 49th FITCE Conf.*, Santiago De Composella, Spain, 2010, pp. 20–24.

[14] V. Kirthikeyin and D. Nank, "Link Failure Prediction in IP Networks", *J. Institute of Telecom. Prof.*, vol. 5, part 3, pp. 33–41, 2011.

[15] C. J. Tyrrell, "Custon Network Design and Pricing Tools", *BT Technol. J.*, pp. 30–35, vol, 18, no. 2, 2000.

[16] Y. Zheng and S. Akhtar, *Networks for Scientists and Engineers*. New York: Oxford University Press, 2002, pp. 379–406.

[17] http://www.opnet.com/?gclid=CI2smsecuJwCFc0B4wod318WnA (accessed 14-6-2012).

[18] http://privatewww.essex.ac.uk/ fleum/weas.pdf (accessed 14-6-2012).

[19] http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=478790&isnumber=10228 (accessed 14-6-2012).

[20] http://www.isi.edu/nsnam/ns/ns-tutorial/ipam_0203_john.pdf (accessed 14-6-2012).

[21] M. A. Rahman, A. Pakštas, and F. Z. Wang, "Review: Network Modelling and Simulation Tools", *Simul. Model. Pract. Theory*, vol. 17, pp. 1011–1031, 2009.

[22] J. Gordon, "Pareto Process as a Model for Self Similar Packet Traffic", in *Proc. IEEE Global Telecomm. Conf. GLOBECOM'95*, Singapore, 1995, vol. 3, pp. 2232–2236.

[23] Y. G. Kim and P. S. Min, "On the Prediction of Average Queuing Delay with Self Similar Traffic", in *Proc. IEEE Global Telecomm. Conf. GLOBECOM'03*, San Francisco, USA, 2003, pp. 2987–2991.

[24] Z. Quan and J.-M. Chung, "Impact of Self Similarity on Performance Evaluation in Differential Service Networks", in *Proc. 45th Midwest Symp. Circ. Syst. MWSCAS 2002*, Tulsa, OK, USA, 2002, pp. II-326–II-329.

[25] P. Leys, J. Potemans, B. Van den Broeck, J. Theunis, E. Van Lil, and A. Van de Capelle , "Use of Raw Packet Generator in OPNET" [Online]. Available: www.esat.kuleuven.ac.be/telematic/networking/opnetwork02_pieter.pdf

[26] Y. Zhu and T. Kunz, "MAODV Implementation for NS-2.26" [Online]. Available: http://kunz-pc.sce.carleton.ca/thesis/maodv-ns-2.26.pdf

[27] T. R. Andel, and A. Yasinsac, "On the Credibility of Manet Simulations", *IEEE Comp.*, vol. 39, no. 7, pp. 48–54, 2006.

Edward Smith is a consulting engineer with British Telecom, working with major clients across a wide range of technologies. He has been with BT since 1982 developing his IT skills with United Biscuits handling real time process control systems and Elida Gibbs, developing transaction processing middleware and data communications solutions. He holds B.Sc. and Ph.D. degrees from the University of Leicester and a Post Graduate Certificate in Commercial Management from De Montefort University. He is a Fellow of the British Computer Society, a Chartered Information Technology Practitioner, a Chartered Engineer and a TOGAF accredited Enterprise Architect.
E-mail: edward.a.smith@bt.com
British Telecom
Bracknell ATE
Market St
Bracknell, Berkshire, RG12 1JG, United Kingdom

Paper

Reliable and High QoS Wireless Communications over Harsh Environments

Josu Bilbao[a], Aitor Calvo[a], Igor Armendariz[a], and Pedro Crespo[b]

[a] *Communications Department, IK4-Ikerlan, Arrasate-Mondragon, Spain*

[b] *Electronics and Communications Department, CEIT and Tecnun University of Navarra, San Sebastian, Spain*

Abstract—One of the most challenging research fields in which research community has taken a very active role is focused on trying to bring the features of wireless networks into line with the traditional wired solutions. Given the noisy and lossy nature of the wireless medium, it is more difficult to provide a comparable Quality of Service (QoS) and Reliability over wireless networks. This lack of reliability avoids the use of wireless solution in scenarios under harsh environment and mission-critical applications. In this paper we propose an inter-node collaborative schema with the aim of improving the achievable QoS level for multicast streaming, through the use of Network Coding and the algebra it is based on. We also present an implementation of the described algorithm on the OPNET discrete event simulation tool. Experimental results highlighting the performance achieved by the proposed algorithm and its improved efficiency as compared to other solutions are described.

Keywords—cooperative wireless networks, multicasting management techniques, next-generation networks, opportunistic wireless, quality and performance evaluation, QoS, reliability and performance modelling, reliability of networks, wireless and mobile networks.

1. Introduction

During the last decade, the research field related with improvements on the QoS (Quality of Service) of communications infrastructures for both in-home, in-building and WAN networks has been a very active topic ([1]–[3] and [4] describe good surveys on it). The research community has produced a large number of protocols and mechanisms enabling the use of networks with shared resources to provide time-sensitive services and distribute applications requiring a high QoS [3], [5]–[7].

Wireless communications have been introduced for applications such as web browsing or other services with low QoS requirements. In fact, the lower layers of the OSI model [8] for wireless environments do not offer a link quality comparable to that of wired networks, and they require the use of large buffers to compensate the effects of packet loss [9]. One of the most challenging research fields for the Future Internet therefore focuses on providing the same QoS level for a wireless link as for a wired link, facilitating the deployment of high-demand services in wireless networks, despite their noisy and lossy nature.

A new research branch emerges from new innovative Information Theory field based on network coding that has recently appeared, known as the *"Modern Theory of Communication"* [10], [11]. Our paper proposes a solution derived from the algebraic principles of network coding [12], and describes a basic implementation of the algorithm on a discrete event simulation platform. We propose a collaborative and opportunistic coding framework for this purpose.

This paper describes a collaborative algorithm between wireless nodes and an innovative queue management system that provides an increase in the quality of service and reliability of wireless networks as compared to already existing solutions. The broadcast nature of wireless networks makes this algorithm particularly effective, and considerably reduces the reliability difference with respect to wired solutions. This research paper describes an example of multicast streaming (e.g., for applications such as SVC, Scalable Video Coding) in a noisy and time-variable environment. The proposed algorithm increases both QoS and stability with regard to temporal variations in the environment, typical of the wireless nature of the links.

The rest of the paper is organized as follows. In the following paragraph (Section 2) the main fundamental algebraic principles are described, over which the algorithm proposed in this paper is based on. Afterwards, in Section 3 we describe the algorithm and in Section 4 we analyze the different configuration parameters that have been designed for it. In Section 5 the design and algorithms of the different nodes integrating the proposed protocol are described so that the complete communication network is defined. In the same section, the details of the used simulation environment and the generation of the harsh environment are analyzed.

Section 6 describes the experimentation process and the results of the performance measurement, highlighting the achieved milestones. Afterwards, Section 7 shows the comparison of the achieved performance among the proposed new algorithm versus the classical solutions for the multicast transmission over harsh environments. Finally, we describe the conclusions of the obtained results and future research lines are glimpsed.

2. Algebraic Fundamentals

2.1. Transmitted Information (vs) Sent Packets (Information Representation)

Lets consider an acyclic directed graph $G = (V, E)$ with unit capacity edges, a source node $s \in S$, and a multicast transmission to a set of receiver nodes $T \subseteq V$. R. L. Li *et al.* [13] and Koetter and Médard [12] proved that multicast capacity is achievable if linear network coding is used in a directed graph network, whereas using routing schemas it is generally impossible to guarantee that multicast capacity can be achieved and finding the best possible routing schema is computationally NP-hard.

The proposed algorithm is based on the linear combination of information units (as an abstract representation of packets) [14], [15]. Each *"original"* packet (ingress to the combining node) is considered to be an independent variable in a system of equations [13]. This being the case, the packets can be transmitted *"systematically"*, so that they are not combined with other packets, and only the original variable is sent. However, an interesting alternative is to send equations with these variables – in other words, linear combinations of the original incoming packets – to the networked nodes [16], [17]. As it is a linear system, it can be represented as a matrix of equations, whose elements are the original packets and the coefficients used to generate the linear combinations among them. Each combined packet can be represented as [18], [19]:

$$P_{lc} = \alpha \cdot P_{in_1} + \beta \cdot P_{in_2}, \tag{1}$$

where $\alpha, \beta \in F_{q=2^8}$.

Being P_{in_1} and P_{in_2} the original incoming packets to be redistributed, α and β the coefficients from a Galois finite field F_q with a field size of ($q = 2^8$), so that the finite field is defined as F_{2^8}, and P_{lc} the outgoing packet as a result of linear combination of incoming packets. It must be taken into account that P_{lc} will be in general the concatenation of several symbols (*bytes*) obtained multiplying incoming symbols by random coefficients.

The complete representation of each system is therefore determined by the extended matrix, made up of the equation coefficients and the payload of the packets, whose bytes would be the free terms of each equation. By basic algebra, if a node obtains sufficient equations to complete the rank of the matrix, it has sufficient information to decode the whole system of equations, and consequently the original packets (information).

2.2. Galois Fields

Galois fields are the basis for coding and collaboration algorithms. Galois fields are sets of finite numbers with operations such as the sum and the product, and linearity properties. Galois fields are frequently used in areas such as cryptography and coding theory, and they have been implemented in the field of network coding since its creation [20]. The finite fields habitually used are of size 2^8 [21], which is very useful in computing as each element can be represented in exactly one byte.

2.3. Galois Matrices

Galois matrices are matrices whose elements belong to a Galois field. In this research work, we will use Galois matrices to perform linear operations over packets. The proposed network coding algorithm takes the original packets as variables in a system of equations, which can be represented by a matrix. This means that this system can be solved if a sufficient number of equations are available. As linear operations can be performed on the matrices, a Gauss-Jordan reduction can be applied to them and their rank can be calculated by receiver nodes. Moreover, the matrix can be even further reduced, and it can be determined which variables can be solved, even if the whole rank is not available. The number of these variables is referred to as *"solvable"*, and it determines the level of information decoded by a node.

2.4. Transmission Matrix

Generalizing the outgoing linearly combined packet generation statement, we have already mentioned that we can consider that the transmission is defined by a system of equations, also known as the *"Transmission Matrix"*. Let's consider P_{in_i} as the i-th original packet, we can consider a bucket of K original packets, being K the *"Generation Size"*, the following linear system equation defines the generation of the linearly combined outgoing packets P_{lc_i}:

$$\begin{bmatrix} P_{lc_1} \\ P_{lc_2} \\ P_{lc_3} \\ \dots \\ P_{lc_K} \end{bmatrix} = \begin{pmatrix} \xi_1(e_1) & \xi_2(e_1) & \dots & \xi_K(e_1) \\ \xi_1(e_2) & \xi_2(e_2) & \dots & \xi_K(e_2) \\ \xi_1(e_3) & \xi_2(e_3) & \dots & \xi_K(e_3) \\ \dots & \dots & \dots & \dots \\ \xi_1(e_K) & \xi_2(e_K) & \dots & \xi_K(e_K) \end{pmatrix} \cdot \begin{bmatrix} P_{in_1} \\ P_{in_2} \\ P_{in_3} \\ \dots \\ P_{in_K} \end{bmatrix}. \tag{2}$$

The K-by-K matrix of $\xi_j(e_i)$ elements corresponds to the coefficients randomly chosen in F_q, where $\xi_j(e_i)$ represents the random coefficient multiplying the symbol P_{in_j} incoming from the (e_i) artificial edge to s node [10], [18]. The generalized expression for Eq. (1) is therefore:

$$P_{lc_i} = \sum_{j=1}^{K} \xi_j(e_i) \cdot P_{in_j}. \tag{3}$$

The Transmission Matrix is therefore compounded of the linear combinations generated by multiplying the incoming packet information bytes and the random coefficients in the defined Galois finite field of F_q. It will be necessary to send K packets P_{lc_i}, being $\{i = 1, \dots, K\}$ and K the generation size, to send the whole original information.

3. Principles of the Algorithm

3.1. Algorithm's Basic Operation

Given that the algorithm is located at OSI-Layer 2, also known as Link Layer, everything proceeding from the higher layer is considered to be encapsulated in a Layer-2 payload. As we have mentioned, each *"original"* packet is associated with a variable in a system of equations. This association is made by the transmitter node. As the total information to be transmitted will be potentially greater than the size of this system of equations, we propose that several systems of equations may co-exist throughout the transmission, not necessarily at the same time. Each of them must be identified with respect to the others. To designate each system of equations, we will call each one a different *"generation"*, being K the generation size or number of equations of the linear system. When performing operations such as adding or multiplying *"variables"* (packets), these operations are performed directly on the packet payload, taking each of the bytes as elements of the Galois field of size 2^8.

3.2. Packet Format

In order to provide the basic functionalities of the algorithm, a specific packet format has been defined (Fig. 1). The fields of the proposed packet format are:

Generation ID. A unique identifier of the system of equations (also known as generation). It is defined by the original transmitter of the packets.

Generation Size. The size of the system of equations (number of independent variables, or original packets involved in each generation). We have previously defined K as the number of equations in the linear system or the generation size.

Remaining. Number of remaining equations to be transmitted by the original transmitter. The receivers thus know the remaining amount of equations to be transmitted for the generation to be completed. Being P_{lc_i} the i-th current packet, remaining field will be equal to $K-i$, being K the generation size.

Type. The type of packet transmitted. In the present version of the algorithm there are two types of packet:

- *None*: packet transmitted by the original transmitter. These are combinations of original packets, not recombination of them.
- *Collaboration*: collaboration packets sent by intermediate network nodes. These are recombination of the packets received by the node.

3.3. Source Coding Mechanism

One of the most simple mechanisms for obtaining greater reliability in the wireless link is the use of coding in the source node, also known as source coding. The possibility of linearly combining the packets provides a very straightforward mechanism for this coding. The idea consists of the source node sending a system of equations containing the information corresponding to the whole original packet generation. The system of equations transmitted by the source may be composed of the minimum number of equations or may be over-determined. This over-determined linear system takes into account added packets with redundant information, so that it generates a system of equation

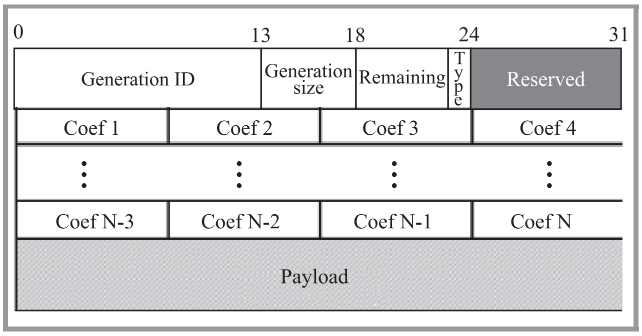


Fig. 1. Proposed packet format description.

of size ψ, being $\psi > K$. This system can be expressed by means of a matrix, and its structure affects the quality of reception by the nodes.

3.4. Inter-Node Network Coding Collaboration Schema

The *"Remaining"* field of the packet informs the nodes of how many more packets of the current generation will be sent by the transmitter. This packet field can be used as the trigger for the collaboration mechanism. To make use of the broadcast characteristics of the wireless medium, after a minimum threshold for the remaining field has been exceeded, the receiver nodes in turn transmit one or more collaboration packets, linearly combining those they have received from the generation in progress. As the information is multicast, one packet can provide information to several nodes at the same time. This collaborative concept is the basis of the network coding [17], [21], and although more sophisticated mechanisms may be designed, even this simple version provides added reliability to the network.

In Fig. 2, a simple example of the collaborative schema is depicted. On the upper side, it can be observed that the transmitter broadcasts 2 packets (P_1 and P_2). Both P_1 and P_2 are the result of linearly combining 2 "original" information packets (x and y), so that $P_1 = x+2y$ and $P_2 = 2x+y$. Therefore, the coefficients to generate P_1 have been $\xi_1(e_1) = 1$, $\xi_2(e_1) = 2$ and $\xi_1(e_2) = 2$, $\xi_2(e_2) = 1$ to generate P_2. However, due to the harsh environment (interference created by the jammer node) and the lossy wireless medium, R_2 and R_4 lose one packet each node. R_2 node lost packet P_2, whereas node R_4 lost packet P_1. Nodes R_1 and R_3 received the whole information transmitted by the transmitter (P_1 and P_2), therefore they can collaborate with the rest of the network to complete the information distribution to the remaining nodes. In the proposed algorithm, nodes that completed the whole information reception will be able to collaborate, in case they are configured to do so, in the redistribution of the information that some nodes have missed.

In the example shown in Fig. 2, as the packet that nodes R_2 and R_4 miss is a different one, transmitter would need to retransmit 2 packets in classical transmission schema. However, in the proposed algorithm, nodes that have received the whole information can collaborate with linear

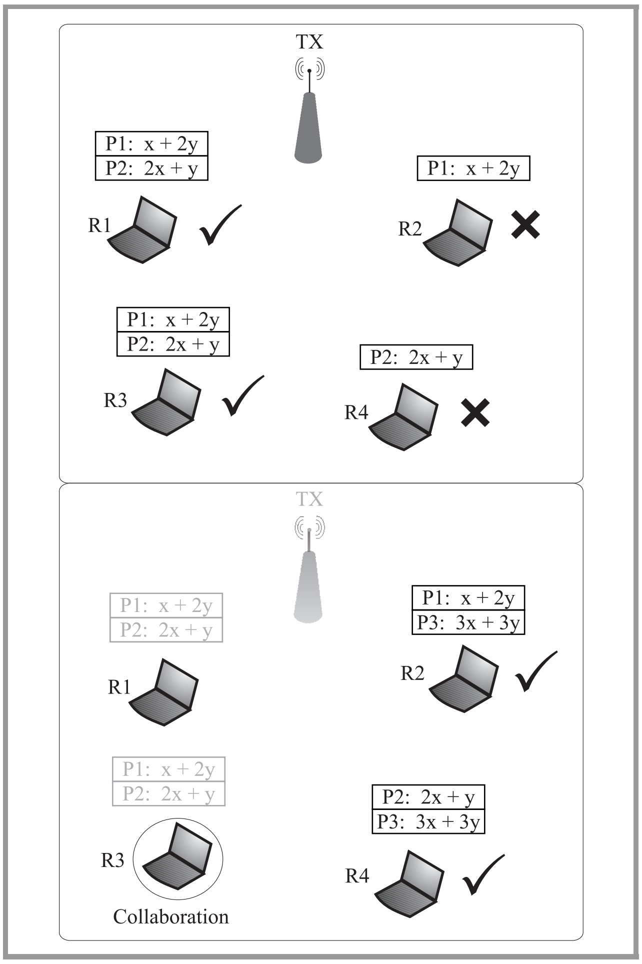


Fig. 2. Collaborative schema. Transmitter sends original packets (upper side) and some packets are lost due to harsh environment. Complete reception with collaborative schema (lower side)

combinations of received packets. Hence, as can be observed in the lower side of Fig. 2, node R_3 acts as a collaborative node and sends a packet (P_3) generated as a linear combination of the packets it has received (P_1 and P_2), for example: $P_3 = P_1 + P_2$, it therefore generates $P_3 = 3x + 3y$. This single packet helps nodes R_2 and R_4 to complete the whole original information reception after just 1 packet collaboration.

4. Algorithm Configuration Parameters

It can be observed that the proposed algorithm described up to now has a variety of configurable parameters, some of them vital for improving its performance, and also resulting in improved QoS and reliability.

4.1. Source Coding Structure

The source coding mechanism provided by the algorithm is defined as the linear combination of packets of the same generation, and can be fully represented by the matrix known in advance as the "transmission matrix".

Basic (*fundamental*) transmission matrix: by elementary algebra, even on a perfect scenario with no packet loss, the minimum matrix to be transmitted must have at least as many equations as variables, and it must also have the maximum matrix rank (equal to the generation size). These are the minimum conditions required for the information to be decoded by the receiver node(s).

We now introduce the concept of *"Coding Density"*, which is the number of different coefficients other than zero of each of the source-generated equations (Fig. 3). It can vary from density 1 (systematic coding) to a density equal to the size of the system of equations (dense coding). In other words, the coding density represents the number of original packets that have been combined to generate an outgoing packet. To ensure that the transmission matrix rank will always be the maximum, the main diagonal elements are always other than zero.

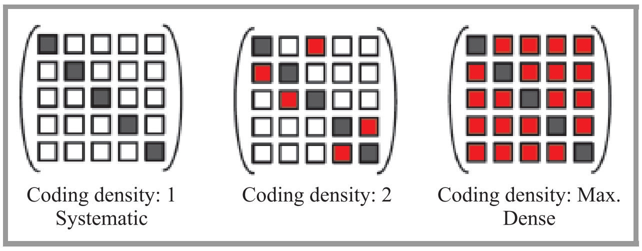


Fig. 3. Coding density options for the transmission matrix.

If the transmission matrix is created using systematic coding, the variables can be solved as soon as they are received, which can be an advantage on some time-sensitive scenarios. By the other side, increasing the coding density improves the probability of a single packet being able to help in the reconstruction of the original information when several nodes have lost different packets.

4.2. Redundancy Blocks

All the packets exceeding the minimum equation system size (rank of the system of equations or generation size) are considered as redundancy packets. A variable number of redundancy blocks can be configured. Each of these blocks can be configured in size and structure, in a similar way to the basic transmission matrix (Fig. 4).

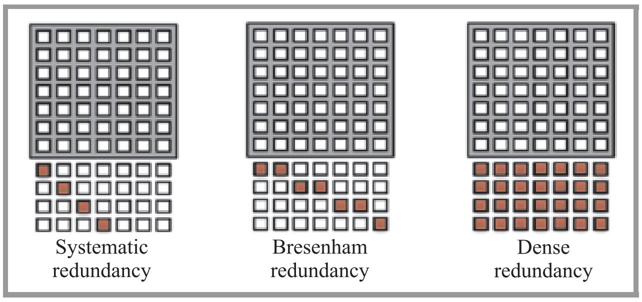


Fig. 4. Redundancy schema options.

Systematic redundancy: this is equivalent to repeating a limited number of packets, as only the main diagonal of a redundancy block is filled. This gives us a single non-zero coefficient in each redundant equation.

Random redundancy: the matrix is filled with variable density, in the same way as the basic matrix, starting by filling the main diagonal. In this case the density defines the number of non-zero coefficients of a given equation.

Bresenham redundancy: this is based on Bresenham's line-drawing algorithm. In this way we obtain the minimum number of non-zero coefficients in each equation, but all the variables are present in the redundancy block.

4.3. Collaboration

All the receiver nodes can be configured to send collaboration messages to help the rest of the receivers. The main parameters are:

- **Collaboration threshold**: this determines the value that needs to be reached by the *remaining* field for a node to decide to send its collaboration message (see packet format in Fig. 1).

- **The number of collaboration messages** sent by each node can be configured, from zero (only transmitter node is able to send packets) to an arbitrary number.

5. Simulation Environment

In the following paragraph we describe the simulation environment implemented to analyze the behaviour of the proposed algorithm, and we study in detail the differences in the QoS achieved by the different mechanisms in study.

5.1. Simulation Tool

A discrete event simulator (OPNET) has been chosen for implementing the proposed algorithm and for testing its performance. This is an event driven network simulator, widely used in the industry and academia to analyze the performance of a diverse range of communication networks.

5.2. Wireless Environment Simulation

OPNET has a specific high performance software package to simulate wireless networks and the effects of the wireless broadcast nature. It is based on the implementation of a series of stages (pipelines) which model each step of a wireless transmission/reception, from the antenna gains in transmission and reception, transmission and propagation delays, as well as wireless medium effects. For the purposes of this paper we have simplified the pipeline cascade, maintaining the pipelines that are the most critical (Fig. 5) for correctly representing a real environment.

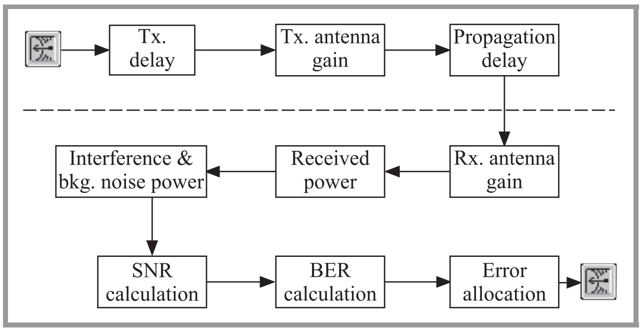


Fig. 5. Wireless pipeline stages.

5.3. Simulated Losses

Packet losses have been simulated in the OPNET environment by means of *interfering nodes*, also known as *jammer nodes*. These jammer nodes have been designed to generate *noise packets* with configurable duration, power and temporal distribution for the noise packet inter-arrival. The implementation of these customized jammer nodes is essential to define a controlled harsh environment.

5.4. Algorithm Design for Simulated Nodes

Three types of nodes have been implemented in the aforementioned discrete event simulation tool:

- **Bridge/transmitter node**: this is the only node with two interfaces: a wired Ethernet interface and a radio interface for the wireless link. This bridge node takes the Ethernet frames that come from an standard Ethernet source, encapsulates them into packets with the previously described format and groups them into generations (systems of equations composed of set of packets whose number is defined by the generation size), running the described algorithm. This node is in charge of source coding mechanism, and therefore the transmission of the transmission matrix. Algorithm 1 describes the pseudo-code of the bridge/transmitter node's process.

Algorithm 1: Bridge/Transmitter node process pseudo-code

```
Insert packet into buffer.
if (# linearly indep. packets in buffer ≥ Gen. Size) then
    Create transmission matrix.
    for i = 0 to transmission matrix rows do
        Send packet or equation
    end for
end if
```

- **Receiver node**: all the receiver nodes are potential destinations for the multicast streaming. They attempt to solve the systems of equations, and, if they are thus configured, to send collaboration packets. These nodes are therefore able to combine packets. Algorithm 2 describes the pseudo-code of the receiver node's process.

Algorithm 2: Receiver node process pseudo-code

```
Receive packet from RF interface.
Read packet coefficients.
Insert coefficients into reception Matrix (Φ)
if (Rank of Matrix Φ has been increased) then
   insert packet into reception buffer
   if (Rank of received matrix == max. rank) then
      information received
      if (collaboration not sent) then
         for i = 1 to collaboration messages do
            send random collaboration packet
         end for
         set status as 'collaboration sent'
      end if
   end if
else
   delete packet
end if
```

- **Sniffer node**: to view the functioning of the algorithm, a simple sniffer node has been implemented, which prints the data of all the packets in the wireless medium into a file. This node provides very valuable information of the protocol performance for an off-line analysis.

6. Algorithm Performance Measurement

A test-bed scenario has been used for comparative measurements and performance analysis of different algorithms. Some of the previously described parameters have been varied, taking the degree of information received as the result. As the algorithm is geared to a multicast streaming, the average of the information received has been measured at all the receiver nodes, or, equivalently, the perceived loss rate has been calculated taking into account the average results over all the receiver nodes.

6.1. Generation Size Variation

The size of the transmitted system of equations has been varied. The transmission matrix is constructed using a basic dense coding matrix and a redundancy block of the same size, also dense. $2 \cdot N$ packets are thus transmitted, where $N = K$ is the generation size and where the coefficients of all the packets are a value other than zero.

It is observed (in Table 1) that the reception quality, and consequently the reliability of the network as a whole, is improved as the generation size is increased. This is due to the fact that there is a greater mathematical probability of receiving any 32 packets from 64 than any 2 from 4. However, the computational complexity (see the execution time required) of solving the coefficient matrices on decoding is of the type $\theta(n^3)$, being n the size of the transmission matrix (see the increasing required execution time as an indirect indicator of the computation complexity in Table 1).

Table 1
Traffic loss rate variation with the generation size

Generation size	2	4	8	16	32
Traffic loss rate [%]	1.79	1.08	0.47	0.15	0.05
Execution time [s]	16.3	17.9	22.4	38.7	119

This effect, together with a larger generation size, involves transmission of more coefficients, and a greater header overhead in each packet means a compromise value must be found. In the following simulations, the generation size of $K = 8$ is taken as the compromise value.

6.2. Coding Density Variation

In order to analyze the effect of the coding density variation, a total number of $2 \cdot N$ (where $N = K$) packets are sent again, but varying the density of their equations, i.e., the number of non-zero coefficients they have, in both the basic transmission matrix block and the redundancy block appended to it. It should be observed that the case of a redundancy with density 1 is identical to repeating each packet twice.

Table 2
Traffic loss rate variation with coding density

Coding density	1	2	3	4
Traffic loss rate [%]	2.30	0.76	0.44	0.44
Coding density	5	6	7	8
Traffic loss rate [%]	0.47	0.49	0.48	0.47

From a density greater than 2, the variations are mainly due to statistical variance. It can be seen that repeating packets is less efficient than sending dense coding, as with dense coding any 8 packets are valid, but when packets are repeated at least one of the two copies of each packet is required. The dense coding will be taken as standard, as apart from being one of the most efficient options it is the most simple schema to be implemented as no special precautions need to be taken to generate the maximum rank matrix.

6.3. Redundancy Type Variation

To compare the different methods of sending redundancy appended after the transmission matrix, different simulations have been made, sending from 2 to 6 redundancy packets using different methods to generate the redundancy block.

A larger redundancy size clearly implies a higher quality level at receiver nodes. It is interesting to observe its variation according to the type of redundancy in each case. In all cases, dense redundancy is more effective. This is an effect of the broadcast nature of the wireless medium. If

Table 3
Traffic loss rate variation with different redundancy type

Type	Systematic	Density 2	Bresenham	Dense
Redundancy size ×2 [%]	7.86	7.04	6.20	6.26
Redundancy size ×4 [%]	6.06	4.61	3.98	2.87
Redundancy size ×6 [%]	4.17	2.77	3.13	1.00

the packets are combined, they can potentially help more receiver nodes and the QoS of the network considerably improves. Medium density and Bresenham solutions yield intermediate results.

7. Comparison with Existing Algorithms

After having tested the behaviour of the proposed algorithm in a variety of configurations, the configuration considered optimum has been selected for comparison with classical schemas in order to improve the QoS of a wireless network. To make a fair comparison, both the classical solution and the new algorithm should generate the same average use of the wireless medium. The comparative scenario contemplates the following alternatives:

- **Traditional/classical solution**: sending standard packets, without confirmation. The medium is used once for each packet transmitted.
- **Repetition of packets**: two copies of each packet are sent, without any coding or collaboration of any type. The medium is used twice for each transmitted packet.
- **Source coding**: dense source coding, with dense redundancy and generation size equal to 8, is used. The medium is therefore used 16 times for each 8 packets.
- **Collaborative algorithm**: each "relay" node in Fig. 6 sends two collaboration packets with a dense redundancy (coding density 8). In total, the medium is used 8 + 4 (*dense redund.*) + 2 (*collab.*) + 2 (*collab.*) = 16 times per each 8 packets.

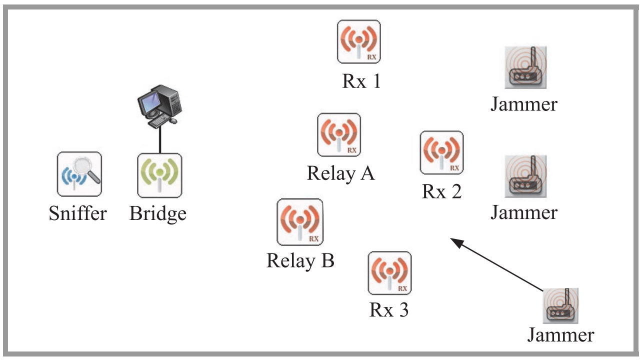


Fig. 6. Benchmark scenario with (moving) interfering nodes.

7.1. Temporal QoS Variation

An *interfering node* (see Fig. 6) is brought up to the network, which worsens the average quality of reception interfering packets sent by the bridge/transmitter or relay nodes. The interfering or jammer node travels a distance of 50 meters in the 60 seconds the simulation lasts. The reception quality of the R_{x3} node worsens considerably. This also entaile worsening the average network behaviour, and consequently its reliability as can be observed in Fig. 7.

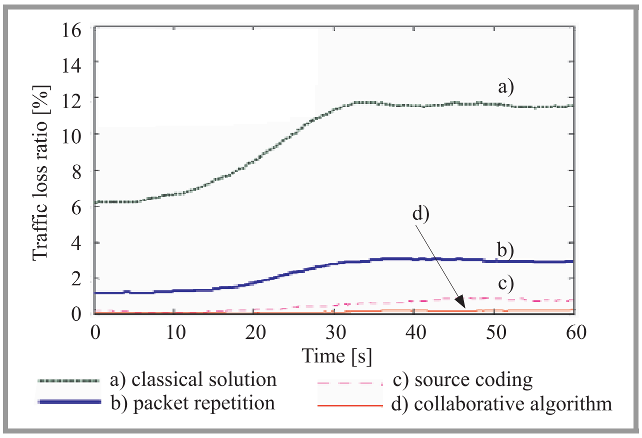


Fig. 7. Traffic Loss Ratio measurement (average among all nodes).

The traditional solution represents the result of a *gross* use of the wireless link, and thus has the worst quality of all. Repeating the packets notably increases this quality, but even higher quality (lower traffic loss rate ratio) is yielded by Galois finite field coefficient-based source coding. The mathematical bases for this improvement have already been presented, and they are clearly shown in this example. Among all the solutions, inter-node collaboration shows the best performance, as it is based on the network coding principle: one single help packet can serve to increase the information received at all nodes at the same time. The network QoS achieved with the proposed schema remains more stable than in the other cases, on the same scenario.

7.2. Interference Power Variation

The second comparison scenario proposed consists of maintaining the protocol configuration stable throughout each simulation run, varying the power of the interference nodes between each run. However, the PER (Packet Error Rate) value is represented on the horizontal axis, given that it is a more objective measurement than the *gross* quality of the wireless link. The environment thus becomes increasingly harsher. For fairer comparison, single repetition of packets (solution b in Fig. 8) is used, together with the protocol configuration that appears to achieve the best results (solution d), as both generate the same use of the wireless medium.

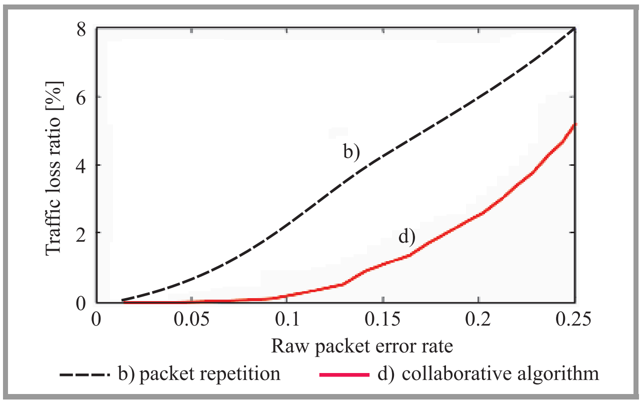


Fig. 8. Traffic loss ratio with raw packet error rate (average among all nodes).

Again, the result of combining the coding and collaboration among nodes using network coding mechanisms achieves clearly better performance than automatic packet repetition. This proofs that the proposed algorithm provides higher network reliability and reaches higher QoS and stable link behavior compared to classical solutions.

8. Conclusions and Future Lines of Work

We have analysed the theoretical bases of network coding and Galois arithmetic-based source coding. We proposed a simulation environment with a simple coding and inter-node collaboration algorithm to improve the network performance under harsh environments. We have tested the influence of several algorithm's main parameters, selecting the solution that optimizes the performance and network overall behavior.

Using the OPNET network simulator and its wireless simulation package, we have designed and implemented the proposed algorithm with a series of solutions and configurable parameters in order to increase the network reliability, demonstrating that the combination of the coding and collaboration mechanisms is the algorithm providing the most stable reliability, regardless of how harsh the medium is.

The presented algorithm increases the QoS and improves its immunity to changes or variations in the network. However, the collaboration mechanism is somewhat basic. We are currently working on more sophisticated mechanisms, based on adaptivity of the nodes collaborating in the network. The aim of all this work is to fully optimise the wireless link's efficiency, seeking the best possible QoS with the same bandwidth use.

Acknowledgements

This work was supported in part by the funding of the Government of the Basque Country (Spain) for Strategic Research projects through the Etortek framework program and MOVITIC (IE11-315) project.

Authors would like to thank FITCE reviewers and JTIT editorial Advisory Board, as well as Professor Pedro Crespo from CEIT-Tecnun University of Navarra and Professor Mariusz Glabowski from Poznań University of Technology for their valuable comments to the text.

The author would like to thank to Professor Médard, Professor Yeung and Professor Crespo for their inspiring work.

References

[1] J. B. Ugalde and I. A. Huici, "Convergence in digital home communications to redistribute IPTV and high definition contents", in *IEEE Consumer Commun. Netwo. Conf. CCNC*, Las Vegas (NV), USA, 2007. pp. 885–889.

[2] L. Chen and W. B. Heinzelman, "A survey of routing protocols that support QoS in mobile ad hoc networks", *IEEE Network*, vol. 21, no. 6, pp. 30–38, 2007.

[3] W. Bin and J. C. Hou, "Muticast routing and its QoS extension: problems, algorithms and protocols", *IEEE Network*, vol. 14, no. 1, pp. 22–36, 2000.

[4] X. Masip-Bruin *et al.*, "Research challenges in QoS routing", *Comp. Commun.*, vol. 29, no. 5, pp. 563–581, 2006.

[5] T. B. Reddy *et al.*, "Quality of service provisioning in ad hoc wireless networks: A survey of issues and solutions", *Ad Hoc Netw.*, vol. 4, no. 1, pp. 83–124, 2006.

[6] M. Médard and A. Sprintson, *Network Coding: Fundamentals and Applications*. Waltham (MA), USA: Academic Press, 2011.

[7] J. Bilbao *et al.*, "Formulation and Methodology for the analysis of viability of Communication Technologies in high QoS requirements multimedia flow redistribution (HDTV) in the Extended-Home environment", in *Proc. IEEE Int. Symp. Broadband Multim. Syst. Broadcasting BMSB*, Bilbao, Spain, 2009.

[8] A. S. Tanenbaum, D. J. Wetherall, *Computer Networks*, 5th Edition. Boston (MA), USA: Pearson, 2010.

[9] ITU-D Study Group 2, Question 16/2. Handbook "Teletraffic Engineering", Geneva, January 2005.

[10] J. Bilbao, I. Armendariz, and P. Crespo, "Disruptive mechanism in the QoS provision", in *Proc. FITCE 2010*, Santiago de Compostela, Spain, Sept. 2010.

[11] C. Fragouli and E. Soljanin, *Network Coding Fundamentals*. Boston-Delft: Now Publishers, 2007.

[12] R. Koetter and M. Médard, "An algebraic approach to Network Coding", *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, 2003.

[13] S.-Y. R. Li, R. Yeung, and N. Cai, "Linear Network Coding", *IEEE Trans. Informa. Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[14] R. Ahlswede, N. Cai, R. Li, and R. W. Yeung, "Network information flow", *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[15] T. Ho and D. Lun, *Network Coding: An Introduction*. Cambridge, UK: Cambridge University Press, 2008.

[16] T. Ho *et al.*, "A random linear network coding approach to multicast", *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[17] M. Médard *et al.*, "On coding for non-multicast networks", in *Proc. 41st Ann. Allerton Conf. Commun. Contr. Comput.*, Monticello (Il), USA, 2003.

[18] P. A. Chou and Y. Wu, "Network coding for the internet and wireless networks", *IEEE Sig. Proces. Mag.*, vol. 24, no. 5, pp. 77–85, 2007.

[19] J. Bilbao, A. Calvo, I. Armendariz, and P. Crespo, "On high QoS constraints over shared resource networks", in *Proc. 7th ACM/IEEE Symp. Architec. Netwo. Commun. Syst. ANCS 2011*, Austin (TX), USA, 2011, pp. 221–222.

[20] T. Yoon and J. Park, "FPGA Implementation of Network Coding Decoder", *IJCSNS Int. J. Comp. Sci. Netw. Secur.*, vol. 10, no. 12, pp. 34-39, 2010.

[21] R. Yeung, *Information Theory and Network Coding*. Springer, 2008.

Josu Bilbao graduated in 2003 obtaining the Telecommunication Engineer degree from the Faculty of Engineering of Bilbao, University of the Basque Country. He received a M.Sc. in Communications and Control from University of the Basque Country in 2008. He is a Ph.D. student at University of Navarra. In 2003 he was awarded with a research fellow in QoS constraint networks field. Since 2005 he is a researcher and project manager at IK4-Ikerlan. He has actively collaborated in the consecution of the national research agenda of Spain (eNEM) as member of the management group. His research interest focuses on High QoS constraint networks, network simulation and reliable communications. He is an IEEE Comsoc, IEEE ITS, FITCE and ACM-Sigcomm member.

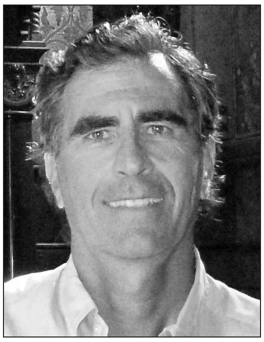
E-mail: jbilbao@ikerlan.es
Communications Department, IK4-Ikerlan
J.M. Arizmendiarrieta 2
20500, Arrasate-Mondragon, Spain

Aitor Calvo was graduated from the Faculty of Engineering of Bilbao, University of the Basque Country and obtained the Telecommunication Engineer degree in 2010. He was granted a research scholarship from IK4-Ikerlan in 2010. Since 2011 he is a researcher at the Communications Department in IK4-Ikerlan. His research interest are focused on reliable communications and discrete event network simulation tools.

E-mail: acalvo@ikerlan.es
Communications Department, IK4-Ikerlan
J.M. Arizmendiarrieta 2
20500, Arrasate-Mondragon, Spain

Igor Armendariz graduated in computer science in June of 1989 from Universidad del Pais Vasco UPV/EHU and obtained a Ph.D. degree from the same University in 1996. He had a scholarship award from January 1991 until December 1994 at CEIT (Centro de Estudios e Investigaciones de Gipuzkoa). He worked as a researcher in the same centre from January 1995 until November 1996 and as a teacher in the computer science faculty of the UPV/EHU from December 1996 until September 2000. Since 2000 he has been working in the communications department of IK4-Ikerlan. He has wide experience working in analogue electronics projects. He has collaborated actively in the national research agenda of Spain (eNEM) and the NEM initiative and is member of IEEE. He has also developed several national and international research projects as well as industrial product development projects related with multimedia distribution over Home Networks.

E-mail: iarmendariz@ikerlan.es
Communications Department, IK4-Ikerlan
J.M. Arizmendiarrieta 2
20500, Arrasate-Mondragon, Spain



Pedro M. Crespo received the engineering degree in Telecommunications from Universidad Politécnica de Cataluña, Barcelona, Spain, and the M.Sc. in applied Mathematics and Ph.D. in Electrical Engineering from University of Southern California, in 1983 and 1984, respectively. From September 1984 to April 1992, he was a member of the technical staff in the Signal Processing Research group at Bell Communications Research, NJ, USA, where he worked in the areas of data communication and signal processing. He actively contributed in the definition and development of the first xDSL (Digital Subscriber Lines transceivers) prototypes. From May 1992 to August 1999 he was a district manager at Telefónica Investigación y Desarrollo, Madrid, Spain. From 1999 to 2002 he was the technical director of the Spanish telecommunication operator Jazztel. At present, he is a professor at University of Navarra and the director of the Electronics and Communications department at the R&D center CEIT, San Sebastian, Spain. Dr. Crespo holds seven patents in the areas of signal processing and wireless communications. Dr. Crespo is a Senior Member of IEEE, and he is a recipient of the Bell Communication Research's Award of excellence. His research interests are in the area of communications and information theory with a focus on wireless communications and sensor networks.

E-mail: pcrespo@ceit.es
Electronics and Communications Department
CEIT and Tecnun University of Navarra
Manuel de Lardizabal 15
20018, San Sebastian, Spain

Paper

On the Optimal Design of a Broadcast Data Dissemination System over VANET Providing V2V and V2I Communications "The Vision of Rome as a Smart City"

Veronica Palma and Anna Maria Vegni

Department of Engineering, University of Roma Tre, Rome, Italy

Abstract—In different this paper we present the performance evaluation study of a simple broadcast data dissemination technique in new emerging Vehicular Ad hoc Networks (VANETs). Differently from the traditional Mobile Ad hoc NETworks (MANETs), VANETs require particular routing protocols, due to the high dynamism of the network topology, and to different traffic and mobility patterns. For safety and emergency message applications, broadcast data dissemination is an important key factor in VANETs. However, the design of optimal deployment of relay nodes in different network scenarios allows to enhance system performance. At this aim, this work analyzes vehicular network performances in terms of throughput and delays, for different traffic scenarios, exploiting both inter-vehicular communications, as well as the availability of fixed network infrastructure.

Keywords—broadcast protocol, IEEE 802.11, V2V, V2I, Vehicular Ad hoc Networks.

1. Introduction

Vehicular Ad hoc NETworks (VANETs) are a particular class of Mobile Ad hoc NETworks (MANETs), where mobile nodes are vehicles moving at different speeds and forming dynamic topology network scenarios [1]. VANETs provide data communications among nearby vehicles via Vehicle-to-Vehicle (V2V) protocol, in the support of Internet access (e.g., web browsing, instant messaging, online gaming, data sharing, etc.), as well as a large variety of safety applications (e.g., assisted braking, controlled safety distance, warning message delivery, etc.).

However, due to the variable nature of such networks, mainly due to dynamic vehicle speed and different mobility and traffic scenarios (e.g., urban, rural and highway), connectivity is time-varying and may cause intermittent and delayed packet delivery. Moreover, in totally-disconnected scenarios (i.e., in highways during nightly hours), data delivery has to rely on available network infrastructure (also called Road Side Units) or, in the case of no availability, on satellite connectivity links [1], [2].

Leveraging previous aspects, it is evident that connectivity issues still represent an open issue in vehicular networks [3]. Ongoing efforts are aimed at enabling inter-vehicle communications supported by existing network infrastructure, in order to provide seamless connectivity and efficient data propagation even in sparse-traffic scenarios. Intelligent Vehicular Ad hoc Networking (InVANET) has defined a smart novel way of using vehicular connectivity by integrating on multiple wireless technologies, such as 3G cellular systems, IEEE 802.11p, and IEEE 802.16e, for effective Vehicle-to-Infrastructure (V2I) communications [4]. Also, V2V and V2I communication technology has been developed as part of the Vehicle Infrastructure Integration initiative [5], which considers the network infrastructure as composed by several Road Side Units (RSUs), equipped with a 5.9 GHz Dedicated Short Range Communication (DSRC) transceiver (i.e., for communications between vehicles and RSUs), and a GPRS interface (i.e., to forward messages to the backbone networks).

In such heterogeneous network environments, protocols for data dissemination and delivery still represent a challenge. It is then evident that the decision on connectivity switching among different "short-lived" links[1] (i.e., V2V, and V2I) should be taken by each vehicle based on main vehicular parameters (i.e., speed, traffic, mobility, type of service, and so on). As an alternative, hybrid vehicular communication protocols (i.e., V2X) represent a viable solution to opportunistically exploit the nature of the vehicular network (i.e., traffic and mobility pattern, availability of fixed relay nodes, etc.) [6], [7].

In this paper, we provide a performance evaluation of a simple broadcast protocol for packet dissemination in VANETs, for different traffic and mobility patterns, ranging from highways with sparse traffic to a very congested traffic environment, where both V2V and V2I connectivity are provided. The main aim is addressed to evaluate the feasibility analysis of a vehicular network in a real highway scenario, considering a portion of highway in Rome (Italy). An effec-

[1] In VANETs, the availability of connectivity links is affected by mobility and clusters formation. V2V and V2I links occur in unpredictable fashion, resulting as intermittent and short-lived. Opportunistic links are exploited for packet transmissions.

tive deployment of RSUs has been also investigated in order to enhance performance especially in low traffic density scenarios. Considerations on optimal RSUs deployment in the vehicular network are taken in order to maintain acceptable network performance, while limiting implementation costs.

This paper is organized as follows. In Section 2 we investigate previous related works on data dissemination protocols in VANETs. Section 3 introduces important issues related to broadcast routing for different scenarios, ranging from a sparse-traffic scenario, with lack of connectivity, to a fully congested scenario. Considering the most simple broadcast approach for data delivery in VANETs, in Section 4 we assess extensive simulation results for different traffic, mobility, and communication modes. Finally, conclusions are drawn in Section 5.

2. Related Work

In the last years, several data dissemination techniques suitable for VANETs have been proposed. Routing algorithms are based on particular vehicular communication protocols (i.e. V2V, V2I and hybrid) and analyze how messages are propagating in VANETs (i.e., message propagation distance and end-to-end delivery delay). The main issue related to a vehicular network is that it lacks of connectivity due to quick disconnections, variable mobility of vehicles and rapidly changing network topology. VANETs suffer from a reliable data delivery specially in sparse-traffic, and totally disconnected scenarios, where vehicle density is very low, and null, respectively [8]. In these scenarios, there is a direct relationship between the amount of packets, which can be successfully received by a vehicle, and the traffic patterns and vehicle speed.

Data dissemination represents a challenge specially in commercial applications (e.g., Internet access, video-on-demand, advertising dissemination, etc.). In entertainment applications, where data flows are larger w.r.t safety applications, message dissemination should be efficient in order to reconstruct a whole data flow from a limited number of received messages. At this aim, the potentiality of network coding protocols for data dissemination has been largely exploited [9]–[12]. This approach can provide a rapid sharing of real-time messages, particularly suitable for comfort applications. As an instance, the use of Fountain codes has been demonstrated to provide efficient and reliable vehicular communications even in high dynamic networks [11], [12]. In [10] the authors propose VANETCODE, a content distribution scheme assuming the content as divided into smaller blocks, which are linearly encoded by vehicles.

The use of hybrid communication protocols has been considered by Cataldi *et al.* in [12]. The proposed scheme is I2V2V, where vehicles can communicate both with network infrastructure (i.e. I2V) and other neighboring vehicles (i.e. V2V), providing a cooperating approach between vehicles, since messages are delivered from the infrastructure to a set of relay vehicles, and then directly to the destination vehicles. This method improves the speed of data delivery in an end-to-end connection, also due to the use of rateless codes providing data reconstruction in a fast way with low overhead.

The use of multi-hop protocols has been exploited in many works for the analysis of message propagation in VANETs, and for a variety of communication modes (i.e., V2V, V2I and V2X). In [13], Resta et al. deal with multi-hop V2V emergency message dissemination through a probabilistic approach. The authors derive lower bounds on the probability that a vehicle correctly receives a message within a fixed time interval. Similarly, Jiang *et al.* [14] introduce an efficient alarm message broadcast routing protocol, and estimate the receipt probability of alarm messages sent to vehicles. Finally, the use of a vehicular grid together with an infrastructure has been largely discussed [15], [16], where benefits of using the opportunistic infrastructure placed on the roads are analyzed.

In this paper we focus on a traditional broadcast protocol in order to assess the feasibility analysis of a vehicular network, and validate network performance in real traffic scenarios. In this study, the vehicular environment is assumed to allow vehicles to communicate both in V2V and V2I modes.

3. Data Dissemination in VANET

Routing in VANETs is an emerging issue due to high mobility of nodes and the dynamic network topology. A VANET is characterized by very short-lived links and then lacks of knowledge about neighborhoods (i.e., vehicular density can change in different areas of the same network). Due to typical features of VANETs, traditional routing protocols designed for MANETs cannot always be suitable. In general, broadcast techniques are frequently used in VANETs for data sharing, traffic, weather and emergency applications. However, different traffic regimes[2] can have impact on data dissemination performances, as summarized as follows [8]:

- *Sparse* traffic condition (i.e., low vehicular density [veh./km]),
- *Dense* traffic condition (i.e., medium vehicular density [veh./km]),
- *Congested* traffic condition (i.e., high vehicular density [veh./km]).

The first scenario, which is very troublesome for conventional routing protocols, considers a limited number of vehicles on the road [8]. It is very typical of night hours, where the traffic density is very low and data dissemination from a source (i.e., a vehicle attempting to broadcast messages) to other relay vehicles is difficult to occur, due

[2] Notice that the traffic density varies heavily depending on the specific road, the time of day, etc. [17]

to the out of the transmission range of the source from the receiver node. Moreover, there might be no cars within the transmission range of the source in the opposite lane either. Under such circumstances, routing and broadcast techniques become a challenging task.

In the second scenario i.e., dense traffic regime, the vehicular density is not uniform in all the vehicular grid (i.e., some vehicles can have a low number of neighbors, while other nodes are moving in a high vehicular density area). The nodes inside the networks do not experience the same topology knowledge. In this case, some vehicles will have to apply a *broadcast suppression* algorithm[3], while some others will have to *store-carry-and-forward* the message in order to preserve the network connectivity.

The third case represents a congested scenario, that is when the vehicular traffic density is above a certain threshold (i.e., > 70 veh./km). Several consecutive cars will share the same wireless medium leading to an excessive number of the same safety message, and then there will be a strong increase of packet collisions and medium contentions among vehicles attempting to communicate. This problem is also referred to as *broadcast storm problem*, which occurs when the traffic density is above a certain value (e.g., when vehicles are in congested traffic scenarios, like during a rush hour).

In order to alleviate this problem, several solutions have been proposed [18], [19], mainly based on decisions for packet (re)-transmission i.e., when and how a safety message should be (repeated) delivered. As a solution, selective broadcast or multicast strategies seem more applicable than either unicast routing or flooding, for the requirement of limitation of broadcast storm problem. Indeed, broadcasting to selected vehicles provides a high overhead without increasing the success rate substantially. Several solutions have been made to introduce intelligence to the basic broadcast concept and make it more selective and, thus, more efficient in its resource usage.

In this paper we consider a broadcast data dissemination technique for all three different traffic regimes, assuming vehicles are moving in a real highway scenario. Extensive simulation results will highlight the design of a vehicular network, providing effective network performance, expressed in terms of throughput and packet delay.

4. Simulation Results

In this section we describe our simulation scenario, where vehicles can communicate via V2V, as well as V2I, for different traffic conditions. The aim of the following tests is to analyze how the increase of vehicles as well as the presence of RSUs in the same area, can affect the network performance, in terms of throughput and delay. In fact, if there is a large gap between two vehicles[4], a packet could be lost or discarded. At this aim, the number of received packets can be increased by exploiting the presence of fixed nodes (i.e., RSUs), which are able of relaying packets inside the vehicular network and then, enlarging the coverage area. We remind that RSUs may generally be either stand-alone devices that communicate only with vehicles via wireless communication, or may be interconnected via a backbone network, or via a mesh network of RSUs [17]. Figure 1 depict the schematic of the use of RSUs to extend the network coverage. In Fig. 1(a) a source vehicle attempts to transmit a packet forward along the route. Packet loss occurs due to a large inter-vehicular gap w.r.t the source transmission range, and no RSU can act as relay node toward a destination vehicle. In Fig. 1(b) the source vehicle transmits a packet to a RSU acting as relay node. The packet is received by the destination vehicle, due to the presence of a RSU along the route, which extends the source vehicle's transmission range.

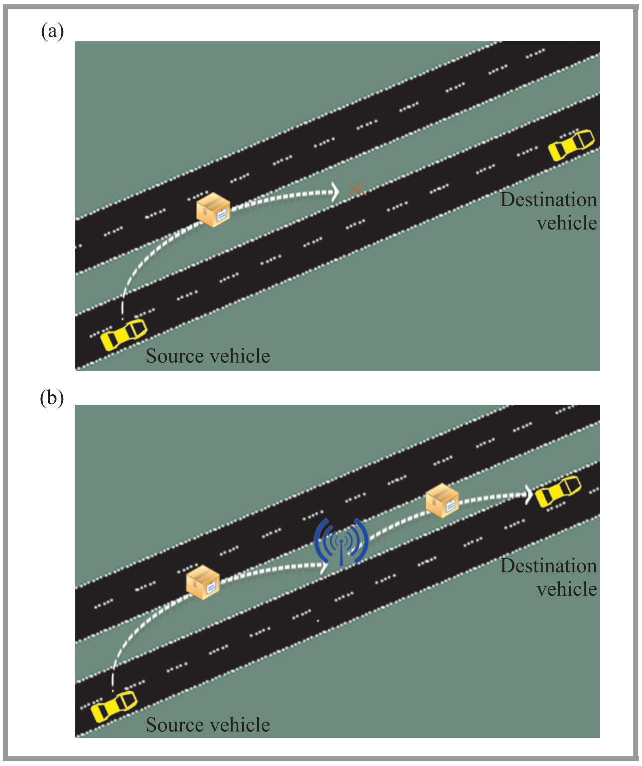


Fig. 1. Schematic of the use of RSUs in a vehicular network to avoid packet losses: (a) packet loss, (b) effective reception.

The network model has been simulated using ns2 [20], and the mobility traces have been generated by SUMO tool [21]. In particular, the used mobility trace is based on a existing highway map from the city of Rome (Italy). Figure 2 depicts the portion of a very well known highway in Rome, called GRA (Grande Raccordo Anulare)[5], which has been used for our simulations.

Different environment configurations have been simulated, varying the mobility traffic level (i.e., from sparse to dense

[3] A broadcast suppression algorithm is used for limiting the number of copies of a message, which can be rebroadcast several times.

[4] The inter-vehicle distance is higher than the source transmission range (i.e., typically > 125 m, depending on the technology.)

[5] Literally, "Great Ring Junction" that is a toll-free, ring-shaped orbital motorway, encircling the city of Rome.

Fig. 2. Map of a portion of GRA highway in Rome (Italy).

traffic scenarios), and the presence of network infrastructure (i.e., RSUs positioned near the routes). Two sets of simulations have been performed, such as:

- *V2V-oriented*, which is aimed to evaluate V2V communications only, and no RSU is considered;

- *V2I-oriented*, which is aimed to evaluate not only V2V, but also V2I communications, due to the increasing number of RSUs.

In both cases, the following three scenarios, depicting real vehicular traffic, have been simulated:

- *Smooth-flowing – sparse – traffic scenario*, where traffic is assumed to be disconnected, as typical of night hours. In particular, we simulate 90 vehicles driving at variable speed, as typical of highway environments.

- *Dense traffic scenario*, i.e. the inter-vehicle distance is almost small, but connectivity is not always guaranteed. Basically, we simulate a high number of vehicles up to 180.

- *Rush hours scenario*, i.e. with high vehicular density (i.e., 300 vehicles moving inside the vehicular area), as typical of rush traffic hours.

Figure 3 depicts the three simulated traffic conditions. We assume that each vehicle is equipped with IEEE 802.11n transceivers, allowing to act as a mobile relay node, and a GPS receiver, allowing information on vehicle's localization. More in detail, in our simulations we consider the design specifications of RSUs and the on-board equipment for V2I and V2V communications, according to Savari Networks [22]. We consider the Savari MobiWAVE On Board Equipment (OBE), which is mounted on-board and is comprised of the main devices for safety applications (e.g., the Vehicle Awareness Device, the Automotive Safety Device, and more others), and the StreetWAVE RSU that is a fixed wireless gateway that can be mounted on a road side traffic pole, working according to ITS applications. Figure 4 (a) and (b) depict the Savari MobiWAVE OBE, and the StreetWAVE RSU, respectively [22].

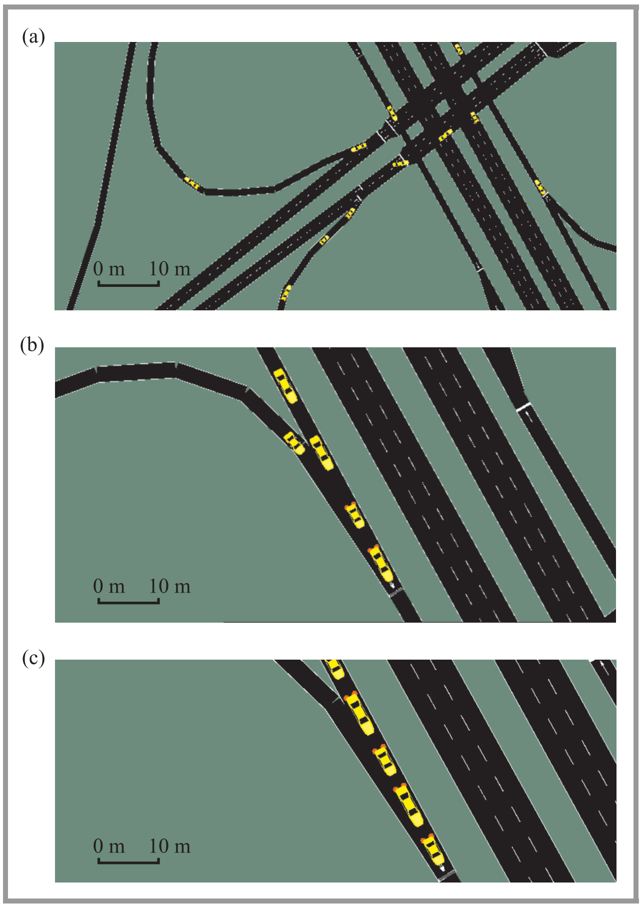


Fig. 3. The three typical simulated traffic conditions inside the vehicular network. (a) Case 1: smooth-flowing traffic scenario, (b) Case 2: dense traffic scenario, and (c) Case 3: congested traffic scenario.

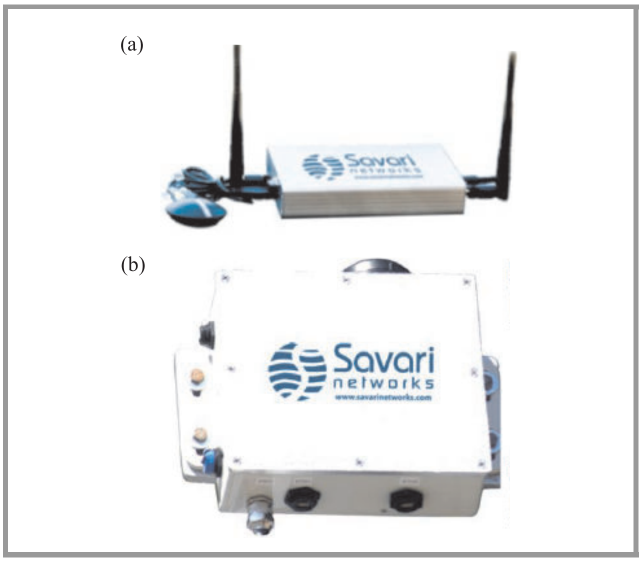


Fig. 4. Savari devices for RSU and on-board unit, [22]. (a) Savari MobiWAVE OBE, and (b) StreetWAVE RSU.

All vehicles are equipped with the Savari MobiWAVE OBE, featuring of a IEEE 802.11*p* network interface card, a highly accurate GPS receiver and a 5.9 GHz DSRC radio. Packets are generated with a constant generation rate, and are transmitted according to a fixed data rate.

On the RSUs' deployment in the vehicular network, four configurations have been considered. In the first case, no RSUs have been included in the scenario; in the second configuration, a dense deployment of RSUs has been introduced (i.e., with a non-uniform RSUs gap), while in the third and fourth cases, RSUs are separated respectively at 1 and 2 km each other. In Table 1, we summarize the main parameters used for all the scenarios.

Table 1
Simulation parameters setup

Parameter	Value
Network Simulator	NS2
Traffic Simulator	SUMO
Number of vehicles	$[90, 300]$
Vehicular area	14 km^2
Simulation time	300 s
Vehicle and RSU MAC	IEEE 802.11*p*
RSU range	$[1000, 2000]$ m
Data Transmission Flow	UDP
Transmission range	300 m
Data rate	6 Mbit/s
Propagation model	Free Space

The simulation results have been compared in all the scenarios, in terms of network performance through the following metrics:

– *Throughput*, as the total amount of data transmitted from the source to destinations in a unit period of time,

– *End-to-end delay*, as the total latency experienced by a packet routed from a source vehicle to a destination node inside the network.

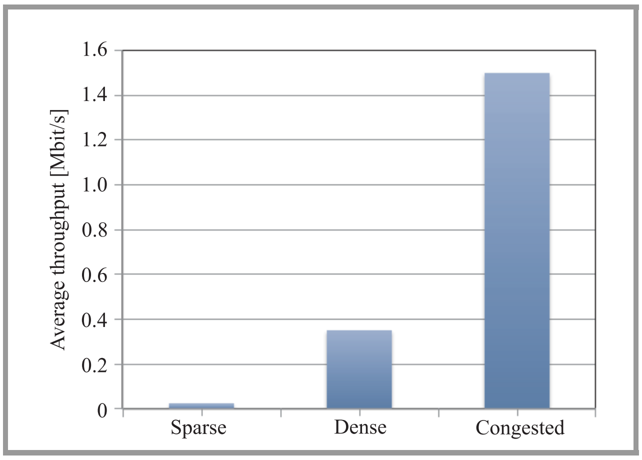


Fig. 5. Comparison of average throughput experienced by vehicles communicating via V2V only, in different traffic scenarios (i.e., sparse, dense and congested).

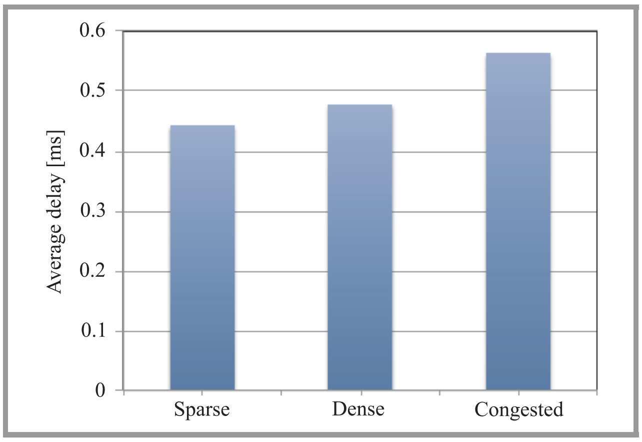


Fig. 6. Comparison of average delay experienced by vehicles communicating via V2V only, in different traffic scenarios, (i.e., sparse, dense and congested).

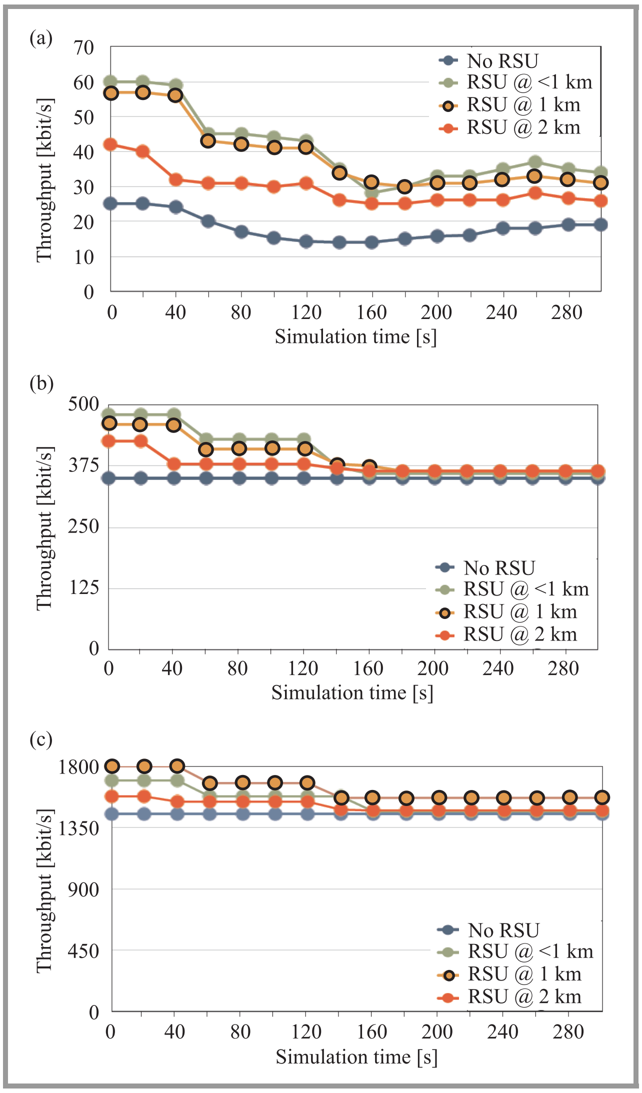


Fig. 7. Throughput [kbit/s] vs. simulation time in different scenarios, i.e., (a) *Smooth-flowing traffic scenario*, (b) *dense traffic scenario*, and (c) *congested traffic scenario*. Notice an increase of performance in all the configurations, mainly due to a higher vehicular traffic density.

In Fig. 5 and Fig. 6, throughput and delay are evaluated in different traffic scenarios, for the case of V2V-oriented simulations. We notice how, on one side, throughput performance increases for higher number of vehicles available to communicate each others and in the other side, the average delay increases due to collisions and congestions. This aspect shows also the influence of market penetration of smart vehicles enabling for inter-vehicle communications.

The most interesting results are when connectivity is supported both by V2V and V2I communications (i.e., V2I-oriented). This represents the most expected scenario configuration for future envision of *smart cities*, where mobility and communications in vehicular environments are supported by vehicles as well as existing network infrastructure. In Fig. 7(a), we show the throughput performance for different values of the number of RSUs; increasing the number of RSUs provides an increasing trend of throughput, reaching up to 60 kbit/s. However, also considering RSU at 1 km of distance each other has a positive impact on throughput. On the other hand, performances get worse when the distance among RSUs increases, as shows the throughput trend for RSUs at 2 km.

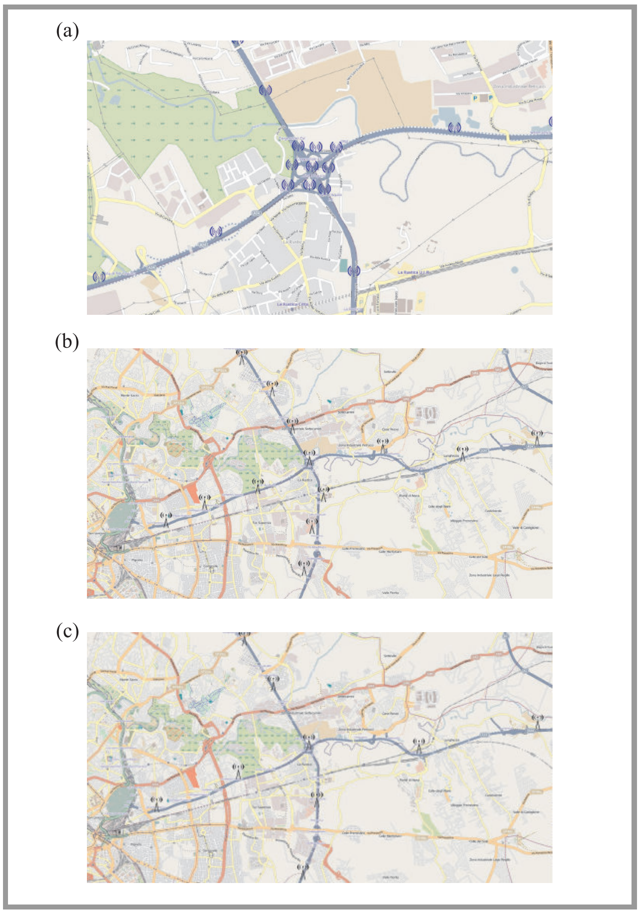


Fig. 8. Variable RSUs deployment in different scenarios. (a) Dense RSUs deployment in the center and near the ramps of the GRA, while low RSU density in highway, (b) 13 RSUs in a dense configuration, deployed at 1 km each other, and (c) 9 RSUs in a sparse configuration, deployed at 2 km each other.

In Fig. 7(b), network performance are evaluated in terms of throughput for the dense traffic case. In this case, the density of vehicles is increased; this leads to an improvement of performances, and all the curves are converging towards the same value. Notice how performances increase for a dense deployment of RSUs in the vehicular network; this also implies that throughput is stable at 350 kbit/s. A variation of performances can be seen at the beginning of the simulation when the vehicles are starting to move before they create dense traffic scenario. Finally, in Fig. 7(c), the obtained throughput when considering a congested network is illustrated. Also for this scenario, an increasing trend of the throughput is obtained when RSUs are at 1 km of distance each other. Notice how performances increase in all the configuration; this can impact on the number of rebroadcast packets, causing network congestions.

The network performances have also been evaluated in terms of end-to-end delay. The end-to-end delay represents an important issue when dealing with vehicular communication systems (e.g., a safety message should be received in a very short time). In this paper, the end-to-end delay is evaluated considering the inter-RSU distance inside the network. In fact, the propagation delay should decrease as the distance of RSUs decreases. For the performance analysis only the smooth-flowing case is taken into account, and the reason lies since for dense or congested traffic scenario there is no lost of connectivity. The end-to-end delay has been evaluated according to previous configurations, i.e., RSUs in a dense deployment (i.e., at < 1 km inter-RSU distance), RSUs laying at 1 km each other, and RSUs laying at 2 km each other. Figure 8 depicts the different scenarios with varying RSUs deployment.

Figure 9 depicts the comparison of end-to-end delays[6] for the three different configurations of RSUs' deployment, in the case of smooth-flowing traffic since this represents the most challenging scenario for packet transmission delay. It is possible to note that, as expected, small increases of delay occur as the inter-RSU distance increases.

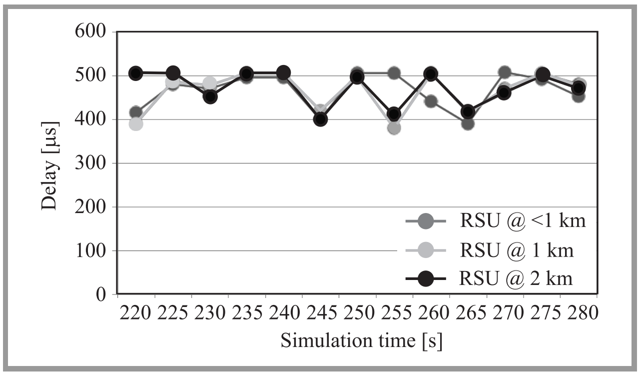


Fig. 9. Packet delay [μs] for a given vehicle vs. simulation time, for different RSUs deployments in the vehicular network. Large gaps between RSUs affect packet delivery delay.

[6] The end-to-end delay is for a reference vehicle only, chosen randomly among all the vehicles in the network. In this particular case, the packet delay exists for $t \geq 220$ s.

Table 2
Performance and cost comparison among different RSUs' deployment configurations within the portion of GRA in Rome

Number of RSUs	Max Throughput in:			Cost [kUSD]
	Smooth-flowing traffic scenario [kbit/s]	Dense traffic scenario [kbit/s]	Congested traffic scenario [Mbit/s]	
0	30	30	1.5	0
18	60	490	1.8	180
13	57	400	1.7	130
9	44	360	1.6	90

Finally, we provide some considerations on the optimal configuration to adopt for the deployment of RSUs, in order to maximize network performance, while keeping low the economic requirements. By considering the installation costs of Savari MobiWAVE OBE, and StreetWAVE RSU, (i.e., 4 and 10 kUSD, respectively), the following Table 2 compares the different configurations of RSUs (i.e., from the absence of RSUs up to 13 RSUs that is the maximum number of RSUs assumed in our simulations). Considering the maximum achievable throughput and the costs, we can conclude that the optimal configuration is for 13 RSUs. Indeed, the differences of maximum achievable throughput in different scenarios are comparable (i.e., only 3 kbit/s low in the smooth-flowing traffic scenario, and around 100 kbit/s in other scenarios). On the other hand, with this configuration we obtain a cost saving of 50 kUSD. This demonstrates that the deployment of a large number of RSUs does not necessary provide a significant enhancement in network performance, but also can have a negative impact on the energy consumption.

5. Conclusion

In this paper, a performance evaluation of broadcast routing protocol in a VANET has been analyzed and discussed, also taking into account the deployment of RSUs inside the vehicular network. The performance analysis has been evaluated in a most realistic scenario (i.e., a selected portion of GRA in Rome) shown that for increasing vehicular densities we obtain high values of throughput.

By including several RSUs along the vehicular network, the end-to-end packet delay – evaluated for a given vehicle – decreases and at the same time, the network throughput shows a better trend. On the other side, an extended number of RSUs has a negative impact on energy and costs consumption in the VANET. A trade-off among an increased throughput and decreased end-to-end delay, and an increased energy/cost consumption is necessary to design an effective vehicular network.

Acknowledgements

This work has been partially supported by RadioLabs Consortium S.p.A. in the framework of GEOSENSE project.

References

[1] S. Olariu and M. C. Weigle (Eds.), *Vehicular Networks: From Theory to Practice*. CRC Press, Taylor & Francis,2009.

[2] A. M. Vegni, C. Vegni, and T. D. C. Little, "Opportunistic vehicular networks by satellite links for safety applications," in *Proc. Fully Netw. Car Worksh.*, Geneva, Switzerland, 2010.

[3] A. M. Vegni, C. Campolo, A. Molinaro, and T. D. C. Little, *Modeling of Intermittent Connectivity in Opportunistic Network: The Case of Vehicular Ad hoc Networks in Routing in Opportunistic Networks*. Springer, 2013 (in press).

[4] H. Moustafa and Y. Zhang (Eds.), *Vehicular Networks: Techniques, Standards and Applications*. Auerbach publishers, Taylor & Francis, 2009.

[5] X. Dong, K. Li, J. Misener, P. Varayia, and W. Zhang, "Expediting vehicle infrastructure integration (EVII)", Tech. rep., California PATH Research Report, UCB-ITS-PRR-2006-20, 2006.

[6] A. M. Vegni and T. D. C. Little, "Hybrid Vehicular Communications based on V2V-V2I Protocol Switching", *Int. J. Veh. Inform. Commun. Sys. (IJVICS)*, vol. 2, no. 3/4, pp. 213–231, 2010.

[7] A. Mostafa, A. M. Vegni, T. Oliveira, T. D. C. Little, and D. P. Agrawal, "QoSHVCP: Hybrid Vehicular Communications Protocol with QoS Priorization for Safety Applications", *ISRN Commun. Netw.*, vol. 2012, p. 14, 2012.

[8] O. Tonguz, N. Wisitpongphan, F. Bai, P. Mudalige, and V. Sadekar, "Broadcasting in VANET", in *Proc. Mob. Netwo. Veh. Environ.*, Anchorage, AK, USA, 2007, pp. 7–12.

[9] S. Yousefi, T. Chahed, M. Moosavi, and K. Zayer, "Comfort Applications in Vehicular Ad Hoc Networks Based on Fountain Coding", in *Proc. 3rd IEEE Int. Symp. Wirel. Veh. Commun. WiVEC 2010*, Taipei, Taiwan, 2010.

[10] S. Ahmed and S. Kanhere, "VANETCODE: Network Coding to Enhance Cooperative Downloading in Vehicular Ad-hoc Networks", in *Proc. Int. Conf. Wirel. Commun. Mob. Comput. IWCMC 2006*, Vancouver, Canada, 2006.

[11] V. Palma, E. Mammi, A. M. Vegni, and A. Neri, "A Fountain Codes-based Data Dissemination technique in Vehicular Ad-hoc Networks", in *Proc. 11th Int. Conf. ITS Telecommun. ITST 2011*, Saint-Petersburg, Russia, 2011, pp. 750–755.

[12] P. Cataldi, A. Tomatis, G. Grilli, and M. Gerla, "A novel data dissemination method for vehicular networks with Rateless Codes", in *Proc. IEEE Wirel. Commun. Netw.*, Budapest, Hungary, 2009.

[13] G. Resta, P. Santi, and J. Simon, "Analysis of multi-hop emergency message propagation in vehicular ad hoc networks", in *Proc. 8th ACM Int. Symp. Mob. Ad Hoc Netw. Comput. MobiHoc'07*, Montreal, Canada, 2007, pp. 140–149.

[14] H. Jiang, H. Guo, and L. Chen, "Reliable and efficient alarm message routing in VANET", in *Proc. 28th Int. Conf. Distrib. Comput. Syst. Worksh.*, Beijing, China, 2008, pp. 186–191.

[15] M. Gerla, B. Zhou, Y.-Z. Lee, F. Soldo, U. Lee, and G. Marfia, "Vehicular grid communications: the role of the internet infrastructure", in *Proc. 2nd Ann. Int. Worksh. Wirel. Internet WICON'06*, Boston, MA, USA, 2006.

[16] G. Marfia, G. Pau, E. De Sena, E. Giordano, and M. Gerla, "Evaluating vehicle network strategies for downtown portland: opportunistic infrastructure and the importance of realistic mobility models", in *Proc. 1st Int. MobiSys Worksh. Mobile Opport. Netw. MobiOpp'07*, Puerto Rico, USA, 2007, pp. 47–51.

[17] H. Hartenstein and K. P. Laberteaux, *VANET: Vehicular Applications and Inter-Networking Technologies*. Wiley, 2010.

[18] A. M. Vegni, A. Stramacci, and E. Natalizio, "SRB: A Selective Reliable Broadcast Protocol for Safety Applications in VANETs", in *Proc. Int. Conf. Selec. Topics Mob. Wirel. Netw. iCOST 2012*, Avignon, France, 2012.

[19] H. Juang, H. Guo, and L. Chen, "Reliable and Efficient alarm message routing in VANET", in *Proc. 28th Int. Conf. Distrib. Comput. Syst. Worksh.*, Beijing, China, 2008.

[20] "Network Simulator 2" [Online]. Available: www.isi.edu/nsnam/ns

[21] "Sumo-simulation of urban mobility" [Online]. Available: http://sumo.sourceforge.net/

[22] "Savari networks" [Online]. Available: http://www.savarinetworks.com/

Veronica Palma received the Dr. Eng. degree in Electronic Engineering from the University of Roma Tre, Rome, Italy, in May 2007, and the Ph.D. degree in Telecommunication Engineering from the Applied Electronics Department at the University of Roma Tre, Rome, Italy, in April 2011. In addition to the research activity for the department, she is now collaborating with RadioLabs, a university-industry consortium participated by large International companies and public Italian universities where she is involved in several European projects. She is a IEEE member and IEEE Communication Society member. Her main research interests focus on routing protocol optimization in mobile networks, network coding and multimedia communications.

E-mail: vpalma1@uniroma3.it
Department of Engineering
University of Roma Tre
Via Vito Volterra, 62
00146 Rome, Italy

Anna Maria Vegni (B.Sc.'04, M.Sc.'06) received the Ph.D. degree in Biomedical Engineering, Electromagnetics and Telecommunications from the Department of Applied Electronics (DEA), University of Roma Tre, Rome, Italy, in March 2010. Since December 2010, she is Assistant Professor at DEA. From May 2009 to October 2009, she was a visiting researcher in the Multimedia Communication Lab, at the Department of Electrical and Computer Engineering, Boston University, Boston, MA, working on VANETs. Her research interests are mobility and interconnectivity in heterogeneous network environments, indoor and outdoor localization techniques, wireless and vehicular communications. She is also involved in several European projects at RadioLabs consortium. Finally, she is a member of GTTI, and a IEEE member, IEEE Computer and Communication Society member.

E-mail: annamaria.vegni@uniroma3.it
Department of Engineering
University of Roma Tre
Via Vito Volterra, 62
00146 Rome, Italy

Paper

Innovative Method of the Evaluation of Multicriterial Multicast Routing Algorithms

Krzysztof Stachowiak and Piotr Zwierzykowski

Faculty of Electronics and Telecommunications, Poznań University of Technology, Poznań, Poland

Abstract—Theoretical considerations of the multicast Quality of Service (QoS) routing have been a rapidly developing and dynamic research area for years. Several algorithms derived from different approaches have been proposed, while the pool of valid solutions to the problem is steadily growing. When new solutions are compared with their predecessors, as much information as possible about their characteristics and differences is needed. Both the graph theory and the optimization theory provide robust and objective means of comparing not only algorithms, but also the results they produce. However, any possible extension to the comparison methods is vital and can bring interesting new information that would eventually lead to innovative conclusions. This article presents a method, derived from practice and experience, that simulates the drainage of resources accumulated by consecutive communication allocations. The nature of this comparison is an extension to the classical measurement of the success ratio and this creates a context of the continuous measure of a success rather than a simple binary value. In this article such a method with regard to algorithms optimizing multicast problems for more than two criteria is used for the first time and leads to an interesting conclusion about the influence of the number of the criteria on the result.

Keywords—evaluation, graph algorithms, multicast, QoS, resource drainage, routing.

1. Introduction

The concept of QoS is the foundation of the process of network convergence. A multitude of services can be provided over the network with the use of a single medium because their requirements are often disjoint. For example, data transfer services may easily coexist with the narrowband real time traffic as the former mainly require large bandwidth, whereas the latter are mostly satisfied with just stable delay guarantees.

One of the more popular techniques in modern networks is the multicast transmission. It enables simultaneous communication of a group of users which, when properly implemented, may offer great resource savings as compared to the basic point-to-point communication based approach. The real time multicast transmission of multimedia content is a widely-used traffic type, which is a challenging research subject as there is a great demand for it in the rapidly developing area of multimedia telecommunications.

The model considered in the article is the *Constrained Minimal Steiner Tree Problem* (CMSTP), [1], [2] that involves connecting a single source with multiple destinations in such way that one of the multiple metrics of the structure is minimal, under the restriction that the others do not violate respective constraints. Therefore, when comparing different algorithms, one has to examine the costs of the multicast tree found in a given graph for given input parameters. The evaluation of the result is a non-trivial task. The metric which is to be minimized should obviously be the lowest, but the constrained metrics may be of greater or lesser importance depending on assumed goals. For example, from the user point of view, any result that satisfies the constraints will be acceptable. It may even be advantageous if the resulting constrained metrics are significantly lower than the proposed constraints. This may, however, lead to an excessive resources drainage which is harmful for the service provider.

From the provider's point of view, the higher the constrained metrics, the better (provided that the constraints are not violated) as it allows providers to save their valuable resources. In this article, the provider's point of view is taken, and so the resources savings process is marked as the main goal. In order to achieve this, an unorthodox comparison technique is to be used. Instead of measuring trees metrics, a special resource drainage scenario has been simulated. In the article, the multicriterial algorithms are compared in this way, and the results of different numbers of criteria are then compared to show how the properties of a given algorithm change with the number of the metrics to be considered.

The article starts with an overview of the available algorithm evaluation techniques and places the one presented by the authors in Section 2. Section 3 introduces a mathematical model used for a description of the algorithms' input and output, which also constitutes the definition of the considered CMSTP problem. In Section 4, the algorithms that have been compared are characterized briefly and the rationale behind the selection of these particular algorithms as the representatives is also provided. Sections 5 and 6 present the experiment description and the presentation and discussion of the obtained results, respectively. Finally, Section 7 concludes the article.

2. Means of Algorithm Comparison

2.1. Evaluation Criteria

The classical purpose of the graph optimization is to find paths, trees or other sub-graphs of the lowest cost. This requirement naturally leads to the cost of the resulting structure as the comparison criterion. In the case of problems with a reasonable complexity, we usually consider algorithms that guarantee finding an optimal solution and, therefore, the running time complexity is the key to evaluate the algorithms' quality [3], [4]. This kind of comparison is one of the fundamental concepts of the optimization theory.

If an $\mathcal{NP}$ problem is considered, such as the CMSTP [1], then optimal solutions are in general non-reachable by any dependable means and thus the computational complexity, while still important, is no longer the only determinant factor. The desired solutions are imminently suboptimal, but the goal is to reach the ones that are possibly closest to the optimum. In order to deal with the algorithm evaluation within limited knowledge, relative values such as the differences between the quality metrics of the results are used. Such an approach is very popular in practice and is presented in [5], among others. What is more, feasible solutions to such problems may not be always readily available and, therefore, often the success rate [3], [6] or the deviation of the actual value from the constraint [3], [7], [8] are to be additionally considered.

As an extension to the aforementioned typical ways of evaluating graph algorithms, another approach is presented in [9]. It simulates the depletion of graph resources under an infinite load of multicast connection requests. The objective is to set up multicast trees for randomly selected node groups one after another, increasing the cost of the occupied edges after each allocation. If a cost of any edge grows beyond a certain limit reflecting the complete depletion of its throughput, the edge is removed from the graph. This is performed until the graph connectivity is broken, after which point the graph is no longer considered valid. The result of the simulation is the number of the trees that were allowed to be set up by the algorithm before the graph became disconnected. Ref. [9] presents the methodology for the optimizations of two criteria only.

One of the advantages this approach gives is the relevance to the real life situations in which dynamical structures are considered and the resources management is important throughout a long period of time. The approach also allows improvement to the success rate measurement. In the case in which two algorithms lead to feasible solutions, the classical approach will judge them equally efficient. However, in our approach further allocations are requested so that we can measure and compare continuous measures of the success instead of a binary value.

As an innovation, in this article multiple criteria are considered and the dependency of the results on the number of the considered criteria are presented in the relevant section.

2.2. Problem Properties

There is a number of important parameters of the experiments that describe the problems solved by the evaluated algorithms.

A very important factor is the *size of test topologies*. Running times directly depend on this parameter, but it may also impact the algorithm procedures indirectly, which is only visible when results for an increasing number of network nodes are presented.

In addition, statistical and topological properties of graphs should be taken into account as there exist a lot of means of obtaining random topologies [10]–[14] and each of them is better suited to reflect different real life networks [15], [16].

This article considers the constrained problems, therefore there is one more important aspect to the graph problems, which is picking constraints so that they are well suited for the comparison. If the constraints are too strict, not many results will be found, if any, and therefore their statistical quality is going to be low unless great amount of computational effort is put into obtaining a sensibly large sample of valid results. On the other hand, if the constraints are too loose, many of the algorithms obtain feasible results early, without any need to perform stronger optimizations, which makes it harder to expose their unique properties. Article [8] presents a technique for picking a single constraint based on a scalar indicating the "toughness" of the problem within the range of $(0,1)$, 0 or less meaning unsolvable problem, and 1 or more meaning a problem that may be solved without any particular optimization with regard to the constrained metrics. In this article, the method has been generalized to include multiple criteria, and this multidimensional variant has been used to generate the problems in the simulations for this article.

Another factor determining how hard the problems are is the *size of the multicast group* to be connected. It not only affects the complexity of the computations, as most of the algorithms' running times depend directly on the number of multicast participants, but also impacts the amount of the resources that is drained from the graph after each tree has been set up.

3. Mathematical Description of the Problem

We model communication network as an undirected graph $G(N,E)$ defined as a finite set of nodes N and a set of edges $E \subseteq \{(u,v) : u,v \in N\}$, each of which reflects a physical point-to-point link. With each of the edges, we associate a set of M metrics modeled with real valued functions: $m_i : E \to \mathbb{R}, i = 0,1,\dots,M-1$. For each of the metrics except the first one we define the constraints C_i, $i = 1,2,\dots,M-1$.

We define a path as a sequence of non-repeated nodes $n_1, n_2, \dots, n_k \in N$ such that for each $1 \le i < k$ an edge

$(n_i, n_{i+1}) \in E$. The cost of the path p with regard to the metric i is defined additively as:

$$m_i(p) = \sum_{e \in p} m_i(e). \tag{1}$$

In this article we evaluate algorithms of the multi-constrained path optimization problem (MCOP), which can be reduced to finding a path p^* such that:

$$\forall_{p \in P(s,t)} m_0(p^*) \leq m_0(p), \tag{2}$$

where $P(s,t)$ is a set of the feasible solutions, i.e., all the paths in the graph G between the nodes s and t that fulfil the following condition:

$$\forall_{i \in (1,2,\ldots,M-1)} m_i(p) \leq C_i. \tag{3}$$

4. Evaluated Algorithms

4.1. HMCMC

The *Heuristic Multi-Constrained MultiCast* (HMCMC) algorithm [3] represents a purely multicriterial multicast algorithm. It is based on a two-pass modified Dijkstra's algorithm in which both the passes utilize a non-linear cost definition. The first of the passes is performed from the destination to all the other nodes in the graph. In this way, a set of labels is defined for each node describing its heuristically defined distance to the destination node. If the tree that is formed this way satisfies all constraints in the paths towards all of the receivers then it is accepted as a final result. Otherwise for each of the destinations that have not been connected to the source via a feasible path, another pass is performed aimed at the optimization of the connection between the particular pair of nodes. The computations for the specific paths are done with use of the information gathered in the initial pass so the results are of better quality than the initial ones at the cost of an additional path finding algorithm run.

4.2. Aggregated MLARAC

In order to demonstrate the discriminating qualities of the presented comparison technique, an algorithm of a very different nature has been selected as the contrasting example. A multicriterial unicast *Multi-dimensional LAgrangian Relaxation based Aggregated Cost* algorithm (MLARAC) [17] has been chosen as its base. In this class of algorithms, the source node is connected with all destinations one by one, resulting in a collection of paths. These paths are then merged into a single subgraph that is, in turn, pruned in order to remove potential cycles from the structure. Such an approach has been earlier demonstrated in [18]–[20], however only two criteria were involved, whereas the MLARAC algorithm handles an arbitrary number of criteria and is used in such an aggregated form for the first time in this article.

4.3. Aggregated HMCOP

In order to provide better exploration of the aggregated unicast algorithms another unicast algorithm is introduced. *Heuristic MultiConstrained Optimal Path* HMCOP [6] is a non-linear Lagrangian relaxation based multicriterial path optimization algorithm. The authors introduce a new, non-linear cost function, which is then used in a two pass Dijkstra's algorithm based search. The first step plays the role of the precomputation providing information for the second pass so that it may efficiently chose good, heuristic result.

5. Experiment Description

The comparison of the multicriterial algorithms is a hard task not only because of the complexity of the algorithms themselves, but also because of the multitude of detail involved in the performance of the simulation, let alone its initiation.
All the parameters that were considered in the experiments were broken into two main categories: the fixed and the variable arguments. The fixed arguments are the assumptions we have chosen experimentally in order to most efficiently expose the searched quantities. The variable arguments are the ones that build up the set of the resulting charts, i.e., the multidimensional results' space.

5.1. Fixed Parameters

Several minor decisions had to be made in order to perform the experiments.

Drainage arguments. The parameters for the drainage simulation were based on the solutions from the OSPF protocol [21] that provided the translation between the edge's cost and the parameters of the underlying physical link:

$$throughput_{ij} = \left\lceil \frac{C_{\min} C_{\max}}{c_{ij}} \right\rceil, \tag{4}$$

where $C_{\min}$ and $C_{\max}$ are the borders of uniform distribution range, and c_{ij} is the cost of the link between node i and j. OSPF uses 10^8 in the numerator, though, based on the actual topologies used in the simulation, we experimentally chose 10^4.
For each stream of data flowing through a link we assumed the drainage of 10 Mbit/s of throughput.

Degree of toughness. A special procedure was used to determine the constraints for the simulated problems. It is presented in [8] and, then, generalized for the multidimensional problems in this article. The coefficient of 0.9 was chosen, which in the scale from 0.0 to 1.0 reflects relatively easy problems. The value was defined arbitrarily in order not to limit the result counts too much so that the differences between the algorithms could be better seen.

Number of graphs. To guarantee the statistical quality, 300 graphs were picked randomly to be considered in each of the major simulation case, which guaranteed the confidence intervals two orders of magnitude less than the obtained average values.

5.2. Experiment Variables

Four of the considered simulation parameters were selected as the variables for the presentation of the results. These are:

- the topology generation algorithm,
- number of the graph nodes,
- size of the multicast group,
- number of the considered criteria.

The first of the above has been chosen in order to reduce the risk of the selected topologies influencing the results too significantly. They are expected to have some impact, so no conclusions should be considered general until confronted with the results for different types of topologies. The following criteria: the number of nodes and the multicast group size are typically used in comparisons [5], [7] and do not require additional explanation. The final variable is one of the improvements of this particular article. As the extension to the previously presented evaluation methodology, the additional dimension of the constraints count is added to the results' space. Choosing it as one of the variables presents an interesting context of the increasing complexity of satisfying an increasing number of QoS requirements.
Two methods of the topology generation have been selected for the experiment. The Waxman's [22] and the Barabasi-Albert's [23] techniques. The numbers: 50, 100 and 150 were selected as the graph sizes. The numbers 2, 3 and 4 were selected for the number of criteria parameters, which reflects a gradual departure from the typical two-criterial comparison. The size of the group was chosen as the main variable and therefore we considered multiple cases of it: 4, 8, 12, 16, 20, 24 and 28.

6. Experiment Results

The experiment results support the claim that the resource drainage evaluation may reveal interesting properties of algorithms. Figures 1–3 present the comparisons of the three algorithms in the Waxman's graphs of 50, 100 and 150 nodes, respectively. Analogically Figs. 4–6 depict the results for the computations in the Barabasi-Albert's topologies. In each of the charts, three sets of plots may be seen. One for the Aggregated MLARAC algorithm, one for the Aggregated HMCOP algorithm and one for the HMCMC algorithm. For each of the three, a set of plots is presented for 2, 3 and 4 criteria.
Each of the charts provides evidence that the HMCMC algorithm produces results that are in general the best in most of the cases. However, further details may be observed as well.

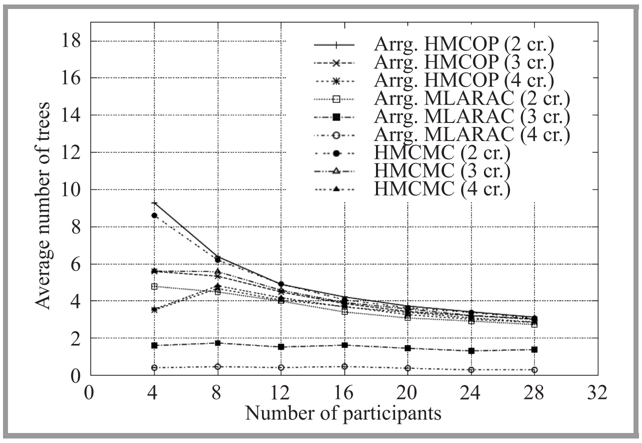


Fig. 1. The comparison results for 50 nodes and Waxman's topology.

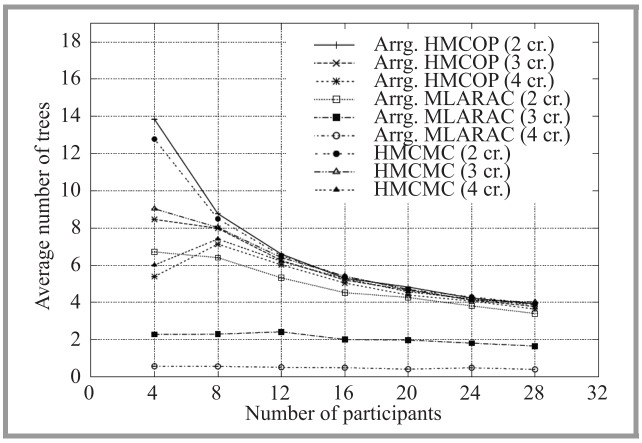


Fig. 2. The comparison results for 100 nodes and Waxman's topology.

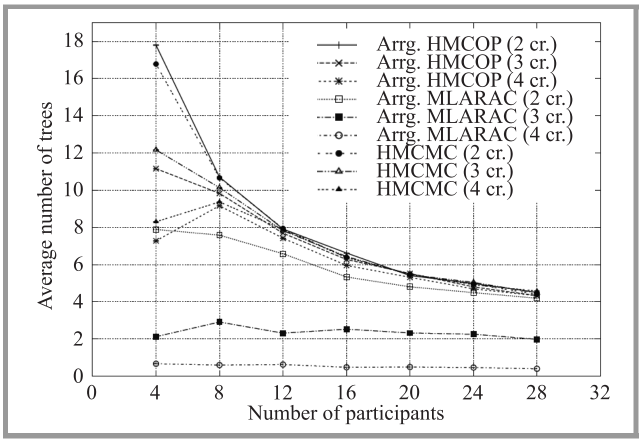


Fig. 3. The comparison results for 150 nodes and Waxman's topology

First of all, a non-linear characteristics of the HMCMC results in the function of the multicast group size may be observed. Also, the curves present different shapes for a different number of the considered criteria, which shows that the experiment presented in this article revealed previously unknown information. For small multicast groups, the al-

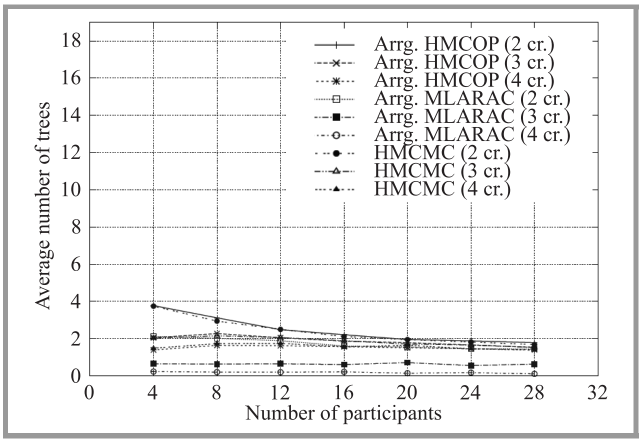


Fig. 4. The comparison results for 50 nodes and Barabasi-Albert's topology.

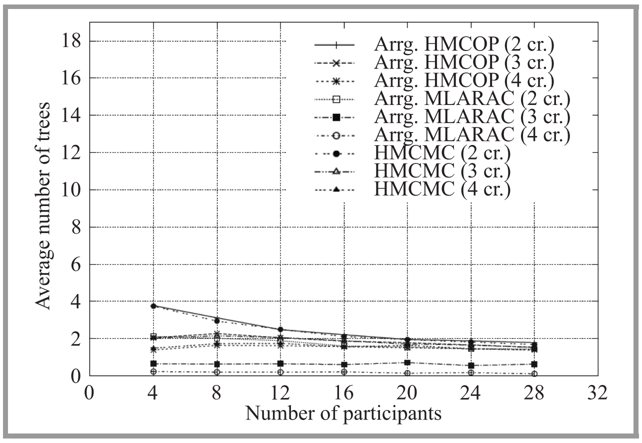


Fig. 5. The comparison results for 100 nodes and Barabasi-Albert's topology.

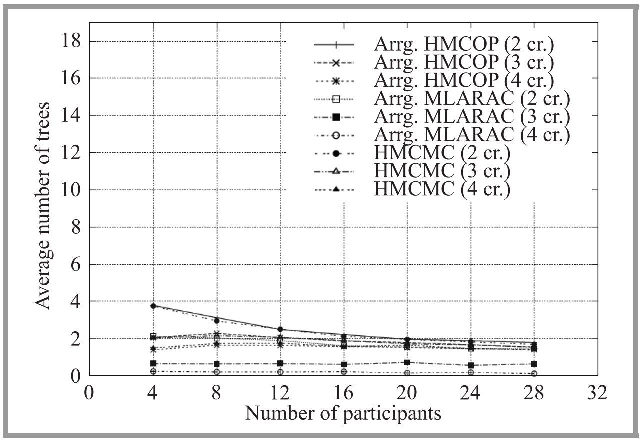


Fig. 6. The comparison results for 150 nodes and Barabasi-Albert's topology.

gorithm tends to produce worse results with the increasing number of the considered criteria, which shows its vulnerability with regard to this parameter. At the same time, the same value is very high in the case with only two metrics being considered.

One of the aggregation based algorithms, the *aggregated HMCOP*, presents comparable performance which may be explained by the fact that it is in principle very similar to the HMCMC at the level of the path finding process. Because the HMCMC approach is optimized in comparison to the aggregation of the HMCOP, and because the final results are similar it may be stated that the HMCMC algorithm turns out better than the aggregated HMCOP with the regard to the assumed comparison criteria.

Different conclusions may be drawn for the *Aggregated MLARAC* algorithm. Firstly, the results tend to be of low quality for greater numbers of the considered criteria, though certain results are still obtained that could potentially present a very good result in the case of the classical success ratio approach. On the other hand, the curve emerging for the low number of the criteria is close to those of the HMCMC.

It is clearly visible that the relationships between different results are very similar in case of both the Waxman's and Barabasi-Albert's topologies. They are however different in scale. It can be noticed that some of the phenomena described above are a lot better visible in case of the Waxman's graphs, especially for the greater amounts of nodes. A minor conclusion may be therefore made that using different topologies, even if does not change the general comparison result, may contribute significantly to the results readability.

In general, a conclusion may be drawn that for a small number of the criteria and large number of participants, all algorithms present comparable performance, though the HMCMC algorithm is still superior. HMCMC and the aggregated HMCOP results present a non-linear asymptotically decreasing trend, whereas those for the Aggregated MLARAC, though being relatively poor, remain constant. In addition, the HMCMC and the HMCOP algorithms present an interesting instability in relation to the number of considered metrics in the case of small multicast participant groups.

7. Conclusion

The class of the multicriterial constrained multicast routing problems presents a non-trivial level of complexity. Following this concept, a need for a broad analysis techniques spectrum arises. In this article, several of the techniques are described, including a presentation of an innovative technique. The resource drainage comparison presents an interesting extension to the concept of the algorithm success rate analysis, which is supported by the provided interesting and valuable results of the experiments. It has been shown that exploring not only the space of the algorithms, but also the space of their comparison is worth an increased amount of effort as the conclusions may render different algorithms useful in different situations. In addition, the stability of the algorithms against changes in different conditions can be shown with the use of the innovative and non-standard analysis.

References

[1] S. Chen and K. Nahrstedt, "An overview of quality of service routing for next-generation high-speed networks: problems and solutions", *IEEE Netw.*, vol. 12, pp. 64–79, 1998.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows: theory, algorithms, and applications*. Upper Saddle River: Prentice-Hall, 1993.

[3] F. Gang, "A multi-constrained multicast QoS routing algorithm", *Comp. Commun.*, vol. 29, no. 10, pp. 1811–1822, 2006.

[4] H. Jiang, P. Yan, J. Zhou, L. Chen, and M. Wu, "Multi-constrained least cost QoS routing algorithm", in *Int. Conf. Telecom. ICT 2004*, J. de Souza, P. Dini, and P. Lorenz, Eds. LNCS, Berlin-Heidelberg: Springer, 2004, vol. 3124, pp. 704–710.

[5] Q. Zhu, M. Parsa, and J. J. Garcia-Luna-Aceves, "A source-based algorithm for delay-constrained minimum-cost multicasting", in *Proc. 14th Ann. Joint Conf. IEEE Comp. Commun. Soc. INFOCOM'95*, Washington, DC, USA, 1995, pp. 377–386 [Online]. Available: http://dl.acm.org/citation.cfm?id=850956.854068

[6] T. Korkmaz and M. Krunz, "Multi-constrained optimal path selection", in *Proc. IEEE INFOCOM 2001*, Anchorage, Alaska, USA, 2001, pp. 834–843.

[7] G. Feng, K. Makki, and N. Pissinou, "Efficient implementations of a delay-constrained least-cost multicast algorithm", *J. Commun. Netw.*, vol. 4, no. 3, pp. 246–255, 2002.

[8] F. Gang, "The revisit of QoS routing based on non-linear lagrange relaxation: research articles", *Int. J. Commun. Syst.*, vol. 20, pp. 9–22, January 2007 [Online]. Available: http://dl.acm.org/citation.cfm?id=1189026.1189028

[9] K. Stachowiak, M. Piechowiak, and P. Zwierzykowski, "The effectiveness of multicast trees with delay constraints", in *Proc. 31th Int. Conf. Inform. Syst. Architec. Technol. ISAT 2010*, Szklarska Poręba, Poland, 2010, pp. 199–209.

[10] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: An approach to universal topology generation", in *Proc. 9th Int. Worksh. Model. Analysis. Simul. Comp. Telecomm. Syst. MASCOTS 2001*, Cincinnati, OH, USA, 2001, pp. 346–356.

[11] C. Jin, Q. Chen, and S. Jamin, "Inet: Internet topology generator", University of Michigan at Ann Arbor, Tech. Research Rep. CSE-TR-433-00, 2000.

[12] E. W. Zegura, K. L. Calvert, and S. Bhattacharjee, "How to Model an Internetwork", *Proc. IEEE INFOCOM 1996*, San Francisco, CA, USA, 1996, pp. 594–602.

[13] E. W. Zegura, M. J. Donahoo, and K. L. Calvert, "A quantitative comparison of graph-based models for internet topology", *IEEE/ACM Trans. Netw.*, vol. 5, pp. 770–783, 1997.

[14] M. Piechowiak and P. Zwierzykowski, "Efficiency analysis of multicast routing algorithms in large networks", in *Proc. 3rd Int. Conf. Netw. Serv. ICNS 2007*, Athens, Greece, 2007, pp. 101–106.

[15] M. Newman, A.-L. Barabasi, and D. J. Watts, *The Structure and Dynamics of Networks (Princeton Studies in Complexity)*. Princeton: Princeton University Press, 2006.

[16] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York: Cambridge University Press, 2004.

[17] K. Stachowiak, J. Weissenberg, and P. Zwierzykowski, "Lagrangian relaxation in the multicriterial routing", in *Proc. IEEE AFRICON 2011*, Livingstone, Zambia, 2011, pp. 1–6.

[18] M. Piechowiak, P. Zwierzykowski, and M. Stasiak, "Multicast routing algorithm for packet networks with the application of the lagrange relaxation", in *Proc. 14th Int. Telecom. Netw. Strat. Planning Symp. NETWORKS 2010*, Warsaw, Poland, 2010, pp. 197–202.

[19] M. Piechowiak, M. Stasiak, and P. Zwierzykowski, "The application of *k*–shortest path algorithm in multicast routing", *Theoret. Appl. Informat.*, vol. 21, no. 2, pp. 69–82, 2009.

[20] M. Piechowiak and P. Zwierzykowski, "A new delay-constrained multicast routing algorithm for packet networks", in *Proc. IEEE AFRICON 2009*, Nairobi, Kenya, 2009, pp. 1–5.

[21] J. Moy, "OSPF Version 2", RFC 2328 (Standard), Internet Engineering Task Force, Apr. 1998, updated by RFC 5709 [Online]. Available: http://www.ietf.org/rfc/rfc2328.txt

[22] B. Waxman, "Routing of multipoint connections", *IEEE J. Selec. Areas Commun.,* , vol. 6, no. 9, pp. 1617–1622, 1988.

[23] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks", *Science*, vol. 286, pp. 509–912, 1999 [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/9910332

[24] A. Juttner, B. Szviatovski, I. Mecs, and Z. Rajko, "Lagrange relaxation based method for the QoS routing problem", in *Proc. IEEE INFOCOM 2001*, Anchorage, Alaska, USA, 2001, vol. 2, pp. 859–868.

[25] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems", *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960 [Online]. Available: http://jmvidal.cse.sc.edu/library/land60a.pdf

Krzysztof Stachowiak received his M.Sc. degree in Telecommunications from Poznań University of Technology, Poland in 2009. Since 2009 he has been pursuing a Ph.D. degree at the Poznań University of Technology in the faculty of Electronics and Telecommunications. He is involved in the research regarding multicast routing in the packet switching networks which mainly consists in the analysis of the existing QoS routing algorithms as well as inventing new proposals. The main subject of his research is multicriterial optimization which has resulted in a series of papers on the subject of the linear and non-linear Lagrangian relaxation. He has taken part in the organization of two international scientific conferences, and has been a participant of several others. Besides the scientific research he is also in charge of the professional training center, coordinating and conducting courses on the usage of the Linux kernel based operating systems. The courses are preparing the participants for the official Linux Professional Institute certification exams.

E-mail: krzysiek.stachowiak@gmail.com
Chair of Communication and Computer Networks
Faculty of Electronics and Telecommunications
Poznań University of Technology
Pl. Marii Skłodowskiej-Curie 5
60-965 Poznań, Poland

Piotr Zwierzykowski received the M.Sc. and Ph.D. degrees in Telecommunications from Poznań University of Technology, Poland in 1995 and 2002, respectively. Since 1995 he has been working in the Faculty of Electronics and Telecommunications, Poznań University of Technology. He received the Stipend of the Foundation for Polish Science for Young Scientists and the Award of the Polish Ministry of National Education and Sport in recognition of his Ph.D. dissertation. He is an Assistant Professor in the Chair of Communications and Computer Networks. He is the author or co-author of over 190 articles and 3 books and served as the reviewer in many international conferences, journals and magazines. He is engaged in research and teaching in the area of performance analysis and modeling of multiservice networks and switching systems. In particular, his scientific interests spread over the following topics: multicast switching, multi-rate systems, traffic modeles for 3G networks and heuristic algorithms for multicast connection. He was supervised final M.Sc. and B.Sc. thesis of more than 80 graduate students.

E-mail: piotr.zwierzykowski@put.poznan.pl
www.zwierzykowski.pl
Chair of Communication and Computer Networks
Faculty of Electronics and Telecommunications
Poznań University of Technology
Pl. Marii Skłodowskiej-Curie 5
60-965 Poznań, Poland

Paper

Recurrent Method for Blocking Probability Calculation in Switching Networks with Overflow Links

Mariusz Głąbowski and Michał Dominik Stasiak

Chair of Communications and Computer Networks, Poznań University of Technology, Poznań, Poland

Abstract—This article presents a new recurrent method for modelling multi-service switching networks with overflow links. In the proposed method, the blocking probability for a given stage of the switching network is determined on the basis of the characteristics of the preceding stage. A particular attention is given to a possibility of a considerable reduction of the internal blocking probability of the switching network that would result from an application of additional overflow links between neighbouring switches of the first stage of the network. The results of the analytical modelling of selected multi-service switching networks with overflow links in the first stage are compared with the results of the simulation experiments. The study confirms the accuracy of all the adopted theoretical assumptions in the proposed analytical model of the multi-service switching network.

Keywords—inter-section links, switching networks.

1. Introduction

Switching networks are essential for the efficient operation of many devices used in network nodes. The parameters of the switching network directly influence the operation of the network in terms of average connection speed and have an effect on the availability of resources for the network's users. Hence, a choice of the optimal switching network with regard to its efficiency and economy is absolutely vital for the effective operation of the network.

Switching networks can be generally divided into non-blocking and blocking networks [1]. Non-blocking networks are the most effective in terms of traffic efficiency as they eliminate the phenomenon of the internal blocking. The phenomenon involves the impossibility of setting up a connection in the switching network that has free available input and output. The structures of non-blocking networks, however, are very expensive since they require a great number of switching elements and advanced controlling algorithms. In this case, the increase in the traffic effectiveness of the network is disproportionally burdened economically as compared to costs of its construction. Blocking networks allow for a loss in part of offered traffic due to the occurrence of the internal blocking phenomenon. Even though this limitation is a hindrance, solutions based on blocking networks structures are viable and economically reasonable (on account of a lower number of switching elements) and are commonly applied in practice.

There are a number of methods for limiting the influence of the phenomenon of the internal blocking in blocking networks [2]. The three most important include: dynamic routing, call repacking and changes in the structure. The first two methods are based on the implementation of dedicated controlling algorithms that decide on the way connections are set up. However, these solutions do not require changes in the structure, they increase considerably the load of the controlling device and, in consequence, slow down the operation of the switching network. The third approach does not involve any interference into the controlling algorithms, but it requires a change in the structure of the switching network. One of the options that is characterized by a relatively small change in the structure is the application of overflow links. This option requires switches with increased number of inputs and outputs to be applied in a given stage of the network. Switching networks with overflow were implemented for the first time in production practice in the Pentaconta cross-bar exchange as early as the 1970s and the 1980s [3]. In Pentaconta switching networks, overflow links were used in the first stage. This solution eventually led to a decrease in the internal blocking probability by several per cent [4]. The possibility of the application of overflow links in switching networks of electronic automatic telephone exchanges were later analysed in many interesting works, including [5]–[7].

Present-day networks service multi-service traffic [8], [9]. The possibility of the application of overflow links in multi-stage switching networks is examined by simulation methods in [10], [11]. These studies have confirmed significant increase in the effectiveness of the switching network with overflow links measured by a decrease in, and in some cases a virtual elimination of, the phenomenon of the internal blocking.

Switching networks with single-service traffic and overflow links are analysed in [3]–[7]. In [4], to model these networks, overflow models were used [12], while to evaluate changes in the blocking probability depending on the capacity of overflow links the exponential function was used. Ref. [7] applies a modification to the effective availability method [13], [14] that was also used in many variants to model single-service blocking switching networks, e.g., in [15]–[18].

The effective availability method is based on a reduction of the blocking probability in a multi-stage network to the calculation of this probability in the single-stage system, i.e., in a non-full-availability group [19]. The accompanying assumption is that the non-full availability group has the same capacity as the output group of the switching network, and that the availability (the so-called effective availability) is determined on the basis of the structure of the switching network and offered traffic. Article [20] proposes an interesting variant of the effective availability method, the so-called recurrent method in which the blocking probability in stage x of a multi-stage switching network is determined on the basis of the blocking probability in stage $x-1$.

In [21]–[23], to analyse multi-service switching networks the effective availability method is used. The method involves an exchange of the multi-service switching network into the equivalent model of a single-service network. The effective availability for calls of a considered class is determined in the equivalent network. After determining effective availabilities for individual classes of calls, on the basis of appropriate models of groups with multi-service traffic [24], the internal and the external blocking probabilities in the switching network are determined. For the analysis of multi-stage multi-service switching networks, [25], [26] propose a recurrent method, the so-called SN-BPPRec method (Switching Networks – BPP – Recurrently). Effective availability methods are characterised by great accuracy, are of universal nature, and can be applied in modelling multi-service switching networks with any structure and any mixture of offered traffic.

In [27], to analyse multi-service switching networks with overflow links and the point-to-group selection, the approach proposed in one of the effective availability methods is used, i.e., the so-called PGBMT method (Point-to-Group Blocking with Multi-rate Traffic) [21]. Article [28] discusses a method for modelling multi-service switching networks with overflow links and the point-to-point selection. Ref. [29] proposes a method for modelling networks with overflow links that are offered Engset traffic. The present article aims at a modification of the recurrent method [25], [26], which will make it possible to model multi-service switching networks with overflow traffic and the point-to-group selection in a much easier way than in the approach proposed in [27], [29].

The structure of the article is as follows. Section 2 describes the structure and operation of a three-stage Clos network [30] with overflow links. Section 3 discusses the concept of the equivalent network and the assumptions for the recurrent method [26]. Section 4 proposes a modification to the recurrent methods and algorithms for modelling multi-service switching networks with overflow links. In Section 5, the results of the calculations are compared with the results of the simulations for selected switching networks. Section 6 sums up the article.

2. The Structure and Operation of the Switching Network with Overflow Links

Figure 1 shows the structure of a multi-service, three-stage Clos switching network. This network is composed of $k \times k$ symmetrical switches in each stage. All input, output and interstage links in the network have the same capacity of f allocation units, the so-called Basic Bandwidth Units (BBU) [31]. Output links of the switching network are grouped according to directions in such a way that each i-th output link of each switch of the last stage belongs to the i-th output direction. The network, composed of $k \times k$ symmetrical switches, has thus k output directions.

In the switching network presented in Fig. 1 overflow links in the first stage are used. When this is the case, the overflow links connect the additional output of a given switch with the additional input of a neighbouring switch of the first stage (Fig. 1). The output of the last switch is connected with the additional input of the first switch of the first stage. A relevant simulation study was performed in [11], [10] in regard to the introduction of overflow links to different stages of the network. The study proves that the introduction of a system of overflow links to the first stage of the network, as shown in Fig. 1, is the most effective. It was also verified in the course of the study that the application of overflow links with a twofold capacity (i.e., $2f$ BBU) is followed by a virtual elimination of the blocking phenomenon in inter-stage links and a stabilization of the internal blocking probability at the level of very low values as compared to the external blocking probability. In such circumstances, the switching network can be considered to be a quasi-non-blocking network. In the model proposed in the article it is assumed that the capacity of the overflow links is at least twice as high as the capacity of the inter-stage link.

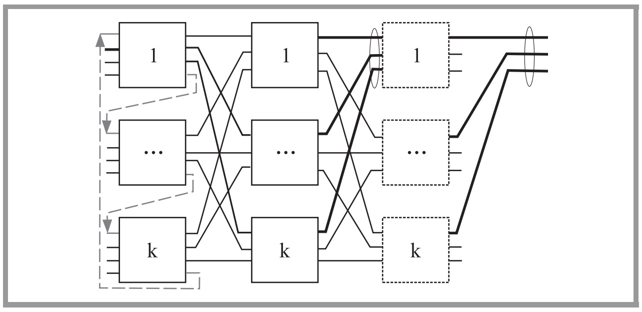


Fig. 1. The structure of a three-stage Clos switching network with overflow links.

The point-to-group selection has been used in the switching network. The algorithm for setting up a connection for this type of selection in the network with overflow links operates in the following way. After a new call appears in a given input link of the first stage switch, the controlling algorithm chooses a switch of the last stage that has an output link in the demanded direction that has free resources to set up the connection. In the next step, the controlling algorithm attempts to set up a connection between selected switches of external stages. If the connection cannot be set up, the controlling algorithm attempts to set up a connection between a selected switch of the last stage and such a switch of the first stage to which the considered switch of the first stage has access to via an overflow link (the assumption is that a given call can make use of only one overflow link that connects neighbouring switches of the first stage). If the attempt at setting up a connection between these switches fails to succeed, the controlling algorithm will choose another switch of the last stage that has a free output link in the demanded direction and will try to set up a connection with this switch directly or via an overflow link. This operation is repeated until the connection is successfully set up or all of the switches that have output links with free resources in a given direction are checked. If, after checking all switches, a connection still cannot be set up, the controlling algorithm rejects the call due to the internal blocking in the switching network. If all the output links of a given direction are busy, i.e., have no free resources, the controlling algorithm rejects the call due to the external blocking in the switching network.

3. Equivalent Network

In this section the concept of the equivalent network will be defined. The equivalent network is a single-service equivalent of a multi-service switching network for calls of a given traffic class. Then, the method for a determination of the effective availability parameter in the equivalent switching network will be discussed. This parameter forms the basis for the evaluation of the internal blocking probability in multi-service switching networks.

3.1. The Structure of the Equivalent Network

Assume that the switching network is offered multi-service traffic that is a mixture of a number of different traffic streams. In order to set up a connection of a given class, an appropriate number of BBUs is required, called the demand of a given class. Demands of different classes are different. The call stream of each of the class is a Poisson stream generated by an infinite number of traffic sources. The service time of calls of each of the class is described by the exponential distribution. It follows that all traffic streams are Erlang streams [8], [9].

The basis for most of multi-service models of switching networks is the notion of the equivalent switching network for the traffic stream of a given class [21]. This means that, for traffic of a given class, a fictitious model of the single-service network servicing fictitious traffic of that class is constructed. The equivalent network is thus a single-service network with the same structure as the multi-service network under consideration. In this network, each link has a capacity of 1 BBU, while each call demands 1 BBU to set up a connection. In networks carrying single-service traffic (with the demand of 1 BBU), the average load of a link (with the capacity of 1 BBU) determines at the same time the value of the blocking probability of the link. In multi-service models of switching networks, for a traffic stream of a given class i, the multi-service switching network is reduced to the equivalent network. Each link of this equivalent network is allocated the load $e(i)$, equal to the blocking probability for a stream of class i in an inter-stage link of the real multi-stage network. This probability can be determined on the basis of the recursive Kaufman-Roberts equations [32], [33] that determine the occupancy distribution in the multi-service full-availability group with the capacity equal to f BBUs:

$$n[P_n]_f = \sum_{i=1}^{M} A_i t_i [P_{n-t_i}]_f \,, \tag{1}$$

$$e(i) = \sum_{i=f-t_i+1}^{f} P[n]_f \,, \tag{2}$$

where:

- $[P_n]_f$ – the occupancy distribution (the probability of the occupancy of n BBUs) in the multi-service full-availability group with the capacity of f BBUs,
- A_i – the average traffic intensity of traffic of class i,
- t_i – call demand of class i, expressed in the number of BBUs,
- M – the number of traffic classes offered to the system,
- $e(i)$ – the blocking probability of calls of class i in the multi-service full-availability group.

3.2. Effective Availability in the Equivalent Network

Having determined the parameter $e(i)$, it is possible to evaluate the effective availability in the switching network for calls of class i. The parameter can be determined on the basis of the modified formula presented in [21] that was derived for the recurrent method [26]:

$$d_{e,z}(i) = d(i, V_z, \pi_z(i)) = [1 - \pi_z(i)]V_z + \pi_z(i)\eta Y_1(i) \,, \tag{3}$$

where:

- $d_{e,z}(i, Vz, \pi_z(i))$ – the notation of the effective availability in the form the function dependent on the parameters i, V_z, $\pi_z(i)$, respectively,
- $d_{e,z}(i)$ – effective availability for the stream of class i in a z-stage equivalent network,
- $\pi_z(i)$ – direct unavailability probability of the last stage switch for calls of class i, i.e., the probability that a connection between the selected first stage switch and the selected last z-stage switch cannot be executed. This probability is determined in many methods for the evaluation of the effective availability, such as in [20], [18]–[23], on the basis of the probability graph method [34],
- V_z – capacity of the output direction in the equivalent network. For the network shown in Fig. 1, this parameter is equal to the number of switches of the third stage:

$$V_z = k_z = k, \tag{4}$$

$Y_1(i)$ – the average fictitious traffic of class i, carried by the first-stage switch; in the considered case of the three-stage switching network, this parameter is equal to:

$$Y_1(i) = ke(i), \tag{5}$$

η – part of fictitious traffic of the first-stage switch that is carried by the considered direction. If we assume that traffic is offered to each of the directions with the same probability, then:

$$\eta = 1/k. \tag{6}$$

The parameter $d_{e,z}(i)$, which forms the basis for the determination of the internal blocking probability, is used in effective availability methods. The parameter determines the average number of switches of stage z that is available from an input of a single first-stage switch for calls of class i. The notion of availability determines those switches with which a connection via free inter-stage links can be set up (the first element of Eq. (3)), as well as those switches with which the first-stage switch has a connection in a given direction (the second element of Eq. (3)).

4. Recurrent Method for Modelling Switching Networks with Overflow Links

This section proposes a method for a determination of the internal, external and the total blocking probability in multi-service switching networks with overflow links. The basis for the proposed method is the SN-BPPRec method, proposed in [26], for multi-service switching networks without overflow links. The latter method will be subsequently modified and then used to analyse switching networks with overflow links. The modification of the method involves an introduction of a number of changes in the determination of the effective availability for subsequent stages of the switching network.

4.1. Internal Blocking Probability

The internal blocking phenomenon in the multi-service switching network occurs when a connection between given switches of the first and the last stage (i.e., between a given input and output link in a given direction) cannot be set up due to the lack of free resources in inter-stage links. In the SN-BPPRec method, the internal blocking probability $E_{\text{in}}(i)$ for calls of class i is determined in a recurrent way.

In the recurrent model, the switching network is considered as a set of sub-systems. The assumption is that each stage of the network is composed of one or more subsystems (Fig. 2). The unavailability probability for s-stage switch, for calls of class i, i.e., the probability $\pi_s(i)$ in a sub-system composed of s first stages of the switching network, can be interpreted as the probability of the occurrence of the blocking phenomenon $E_{s-1}(i)$ in a group of V_{s-1} inter-stage links that lead to one s-stage switch of the equivalent network:

$$\pi_s(i) = E_{s-1}(i). \tag{7}$$

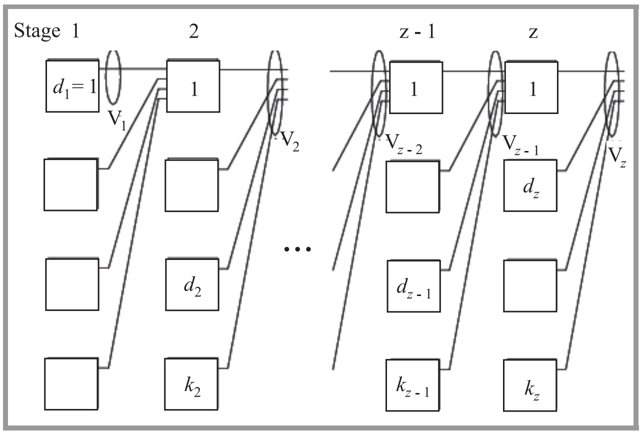


Fig. 2. Multi-stage switching network with indicated parameters used in the recurrent method.

Notice that, according to [16], the blocking probability in a group of inter-stage links that lead to one switch of the s stage, can be treated as the point-to-group blocking probability in a $(s-1)$-stage subsystem of a z-stage switching network. Therefore, the effective availability of the s-stage sub-system is the function of the point-to-group blocking probability of the $(s-1)$-stage sub-system of the switching network:

$$d_{e,s}(i) = d(i, V_s, \pi_s(i)) = d(i, V_s, E_{s-1}(i)), \tag{8}$$

where V_s is the capacity of the output group of the s-stage sub-system of the switching network, i.e., the link group that leads to one switch of the next stage.

In the SN-BPPRec method, the point-to-group blocking probability in each equivalent sub-system of the switching network is approximated by the blocking probability in the Erlang's Ideal Grading (EIG) [19]:

$$E_s(i) = \text{EIF}(A_s(i), V_s, d_{e,s}(i)), \tag{9}$$

where $\text{EIF}(A,V,d)$ determines the blocking probability in the Erlang's Ideal Grading with the capacity V, availability d and the intensity of offered traffic A. This probability is determined on the basis of the so-called EIF Formula (Erlangs Interconnection loss Formula) [19]:

$$\text{EIF}(A,V,d) = \frac{\sum\limits_{i=d}^{V} \binom{i}{d} \Big/ \binom{V}{d} (A^i/i!) \prod_{l=d}^{i-1} \left[1 - \binom{l}{d} \Big/ \binom{V}{d}\right]}{\sum\limits_{j=0}^{V} (A^j/j!) \prod_{l=d}^{j-1} \left[1 - \binom{j}{d} \Big/ \binom{V}{d}\right]} \tag{10}$$

Equations (7)–(9) determine the iterative algorithm for the calculation of the blocking probability in the equivalent switching network. The algorithm proceeds as long as the internal point-to-group blocking probability $E_{\text{in}}(i)$ in z-stage network is determined:

$$E_{\text{in}}(i) = E_z(i) = \text{EIF}(A_z(i), V_z, d_{e,z}(i)). \quad (11)$$

In Eqs. (9)–(11), the parameter $A_z(i)$ is the average intensity of traffic of class i offered to one output direction in the equivalent network with the capacity of V_z of output links.

The operation of the algorithm commences with a determination of the blocking probability $E_1(i)$, i.e., the blocking probability of a group of links that lead from one second-stage switch to one sub-system of the first stage (Fig. 2). The first-stage switch is the sub-system of the first stage of the equivalent switching network. This sub-system is a non-blocking system. Therefore, the following parameters can be adopted: $\pi_1(i) = 1$, $V_1 = 1$ i $d_{e,1} = 1$, because only one output of the first-stage switch leads to a given switch of the second stage. Hence, the parameter $E_1(i)$ can be determined in the following way:

$$E_1(i) = \text{EIF}(A_1(i), 1, 1), \quad (12)$$

where $A_1(i)$ is the average intensity of fictitious traffic of class i offered to one switch of the second stage from one sub-system of the first stage. One sub-system of the first stage (i.e., one switch of the first stage) is connected to only one switch of the second stage via one inter-stage link.

On the basis of Equations (1)–(2), we can determine fictitious traffic $e(i)$, carried by one inter-stage link of this type. Since the link of the equivalent network has the capacity of one BBU, then the relation between offered and carried traffic can be expressed as follows:

$$e(i) = a(i)[1 - E_1(a(i))], \quad (13)$$

where $E_1(a(i))$ denotes Erlang B Formula [8], [9] that determines the blocking probability in the group with the capacity 1 BBU which is offered traffic with the intensity $a(i)$. After elementary transformations in Eq. (13), we obtain:

$$A_1(i) = a(i) = \frac{e(i)}{1 - e(i)}. \quad (14)$$

The SN-BPPRec method adopts the assumption that traffic offered to the output group of stage $s\,(s > 1)$, composed of V_s links, is equal to:

$$A_s(i) = k_s\,a(i). \quad (15)$$

In the switching network presented in Fig. 1, the capacity of the output group in each of the stages is equal to k. The internal blocking probability in the switching network shown in Fig. 1 (without overflow links taken into consideration) can thus be determined in the three consecutive steps.

Step 1:

$$d_{e,1}(i) = 1, \quad (16)$$

$$E_1(i) = \text{EIF}(a(i), 1, 1), \quad (17)$$

Step 2:

$$d_{e,2}(i) = [1 - E_1(i)]k + E_1(i)\eta Y_1(i), \quad (18)$$

$$E_2(i) = \text{EIF}(ka(i), k, d_{e,2}(i)), \quad (19)$$

Step 3:

$$d_{e,3}(i) = [1 - E_2(i)]k + E_2(i)\eta Y_1(i), \quad (20)$$

$$E_3(i) = E_{\text{in}}(i) = \text{EIF}(ka(i), k, d_{e,3}(i)), \quad (21)$$

4.2. The External and the Total Blocking Probability

The phenomenon of the external blocking in the multi-service switching network occurs when all output links in a given direction have no free resources, i.e., the appropriate number of BBUs, to service a call of a given class. The external blocking probability $E_{\text{ex}}(i)$ in the multi-service switching network in the SN-BPPRec method is approximated by the blocking probability in the Limited Availability Group (LAG) [35]. The limited-availability group is a model of k identical links, each with the capacity of f BBU. The group can service a call of a given class only when there is a possibility of a service of this call in one (any) link of this group. The above definition of LAG reflects the operation of the output group of the multi-service switching network. The occupancy distribution and the blocking probability in LAG are described by the Eqs. (22)–(23):

$$n[P_n]_{kf} = \sum_{i=1}^{M} A_i t_i \sigma_i(n - t_i)[P_{n-t_i}]_{kf}, \quad (22)$$

$$E_{\text{ex}}(i) = \sum_{n=k(f-t_i+1)}^{kf} [P_n]_{kf}[1 - \sigma_i(n)], \quad (23)$$

where:

- $[P_n]_{kf}$ – occupancy distribution (probability of the occupancy of n BBU) in LAG with the capacity of kf BBU,
- A_i – the average intensity of traffic of class i offered to LAG. This is traffic offered to a given direction, i.e., to k output links of the considered switching network,
- $\sigma_i(n)$ – the conditional transition probability for the call stream of class i.

The conditional transition probability (state-passage probability) in the LAG model is the parameter that defines the probability of favourable combinations of occupancy in state n, i.e., such combinations that make it possible to service a call of a given class in at least one link that belongs to LAG. In the LAG model [35] the conditional transition probability is approximated on the basis of the following combinatorial formula:

$$\sigma_i(n) = \frac{F(kf-n,k,f,0) - F(kf-n,k,t_i-1,0)}{F(kf-n,k,f,0)}, \qquad (24)$$

where $F(x,k,f,t)$ is the number of arrangements of x free BBUs in k links, each with the capacity of f BBU, with the assumption that initially each link was assigned t free BBUs:

$$F(x,k,f,t) = \sum_{r=0}^{\left\lfloor \frac{x-kt}{f-t+1} \right\rfloor} (-1)^r \times \binom{k}{r} \binom{x-k(t-1)-1-r(f-t+1)}{k-1}. \qquad (25)$$

After determining the internal blocking probability $E_{\text{in}}(i)$ and the external blocking probability $E_{\text{ex}}(i)$, we are in position to determine the total blocking probability $E_{\text{tot}}(i)$ for calls of class i in the multi-service switching network. The probability $E_{\text{tot}}(i)$ is the sum of the probabilities $E_{\text{in}}(i)$ and $E_{\text{ex}}(i)$ that exclude simultaneity of the occurrence of the internal and the total blocking events:

$$E_{\text{tot}}(i) = E_{\text{ex}}(i) + E_{\text{in}}(i)[1 - E_{\text{ex}}(i)]. \qquad (26)$$

The SN-BPPRec method – presented above – is characterized by high accuracy in the evaluation of the total blocking probability [26] and is particularly useful in modelling switching networks with a high number of stages. The simulation study performed by the authors also indicated, however, significant errors in the evaluation of the internal blocking. Due to relatively low values of the internal blocking probability, as compared to values of the external blocking probability, this error cannot be easily discerned. The next section proposes a modification to the recurrent method for the calculations of the internal blocking probability that results in better accuracy in the evaluation of the internal blocking probability.

4.3. Modified Internal Blocking Evaluation Method

The modified version of the recurrent method adopts that the group of output links of a given stage is equal to the sum of availabilities of sub-systems of the preceding stage. If sub-systems of a given stage are identical, then the following can be written:

$$V_s = L_{s-1} d_{e,s-1}(i), \qquad (27)$$

where L_s is the number of sub-systems in stage s. Note that in the Clos switching network (Fig. 3), only the first stage is composed of $L_1 = k_1$ sub-systems. The remaining stages include only one sub-system. Therefore, for the three-stage Clos network we have:

$$L_1 = k_1,\ L_2 = L_3 = 1. \qquad (28)$$

The structure of the equivalent switching network that corresponds to the adopted assumptions is presented in Fig. 3. The effective availability of s-stage sub-system in the modified structure of the equivalent network is then the function of availability and the point-to-group blocking probability in $(s-1)$-stage sub-system of the equivalent switching network. In this case, for each $s > 1$, Eqs. (8)–(9) will be written as follows:

$$d_{e,s}(i) = d_{e,s}(i, V_s, E_{s-1}(i)), \qquad (29)$$

$$E_s(i) = \text{EIF}(A_s(i), V_s, d_{e,s}(i)), \qquad (30)$$

where:

$$V_s = L_{s-1} d_{e,s-1}(i), \qquad (31)$$

$$A_s(i) = L_{s-1} d_{e,s-1}(i) a(i). \qquad (32)$$

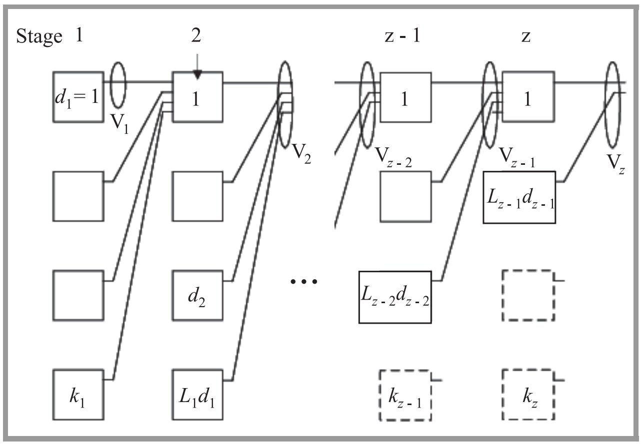


Fig. 3. Multi-stage switching network with indicated parameters used in the modified recurrent method.

For $s = 1$, the parameters V_1 and $d_{e,1}$ are determined exactly as in the recurrent method. The number of sub-systems in a given stage of the considered Clos network is determined by Eq. (28).

To sum up, the internal blocking probability in the switching network presented in Fig. 1 and determined by Eqs. (16)–(21) can be, in the case of the modified recurrent method, written in the following way:

Step 1:

$$d_{e,1}(i) = 1, \qquad (33)$$

$$E_1(i) = \text{EIF}(a(i), 1, 1), \qquad (34)$$

Step 2:

$$d_{e,2}(i) = [1 - E_1(i)]k + E_1(i)\eta Y_1(i), \qquad (35)$$

$$E_2(i) = \text{EIF}(ka(i), k, d_{e,2}(i)), \qquad (36)$$

Step 3:

$$d_{e,3}(i) = [1 - E_2(i)]d_{e,2} + E_2(i)\eta Y_1(i), \tag{37}$$

$$E_3(i) = E_{\text{in}}(i)\text{EIF}(d_{e,2}a(i), d_{e,2}, d_{e,3}(i))\,. \tag{38}$$

4.4. The Internal Blocking Probability in Networks with Overflow Links

In order to decrease the internal blocking phenomenon, overflow links were introduced to the first stage of a three-stage Clos switching network. In this way, each neighbouring switches of the first stage are connected with each other by an overflow link (Fig. 4). After the introduction of overflow links we can thus assume that the availability to one switch of the second stage is equal to two links. In Fig. 4, these two links are marked with bold line. Access to the second link results from the introduction of the overflow link indicated in Fig. 4 by dotted line. One sub-system of the first stage can thus be treated as a system with the capacity and availability equal to two links. Therefore, for a switching network with overflow traffic, Step 1 (Eqs. (16)–(17)) for the recurrent method, and Eqs. (33)–(34) for the modified recurrent method, can be rewritten in the following way:

Step 1:

$$d_{e,1}(i) = 2, \tag{39}$$

$$E_1(i) = \text{EIF}(a(i), 2, 2), \tag{40}$$

The remaining steps of the algorithm, just as the equation determining the external and the total blocking probability, remain without any changes.

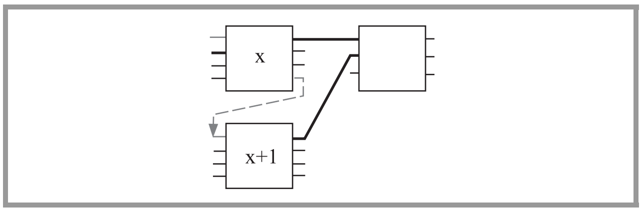


Fig. 4. Structure of links between switches of the first and second sections.

5. Numerical Results

The simulation study and the analytical calculations were performed for a three-stage Clos network. The network was composed of four 4×4 symmetrical switches in each stage. The capacity of the links in the network is 30 BBUs. The overflow links were introduced to the first stage of the network. The assumption was that the capacity of the overflow link was equal to 60 BBUs. The network was offered multi-rate traffic composed of 3 traffic classes that required 1 BBU, 2 BBUs and 6 BBUs, respectively. Traffic of all classes was offered in the following proportions: $A_1t_1 : A_2t_2 : A_3t_3 = 1 : 1 : 1$. To set up connections in the network, the point-to-group selection was used.

The simulation study was performed with a dedicated digital simulator based on the event scheduling method [36]. In the simulation experiments, the 95% confidence interval was determined, evaluated on the basis of the t-Student distribution for 10 series with 100,000 calls of the oldest class in each of the series. The analytical study was per-

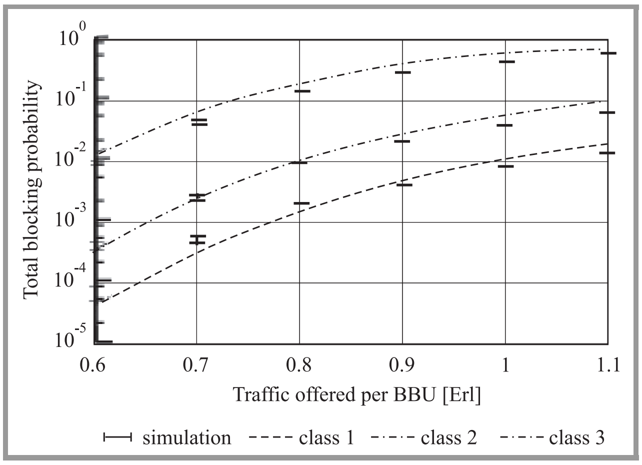


Fig. 5. Total blocking without overflow.

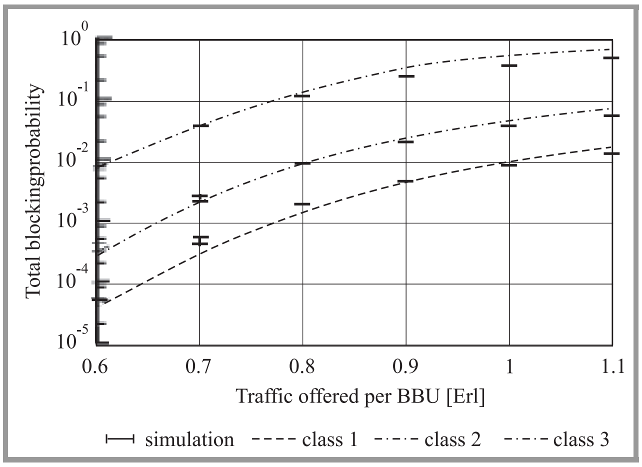


Fig. 6. Total blocking with overflow.

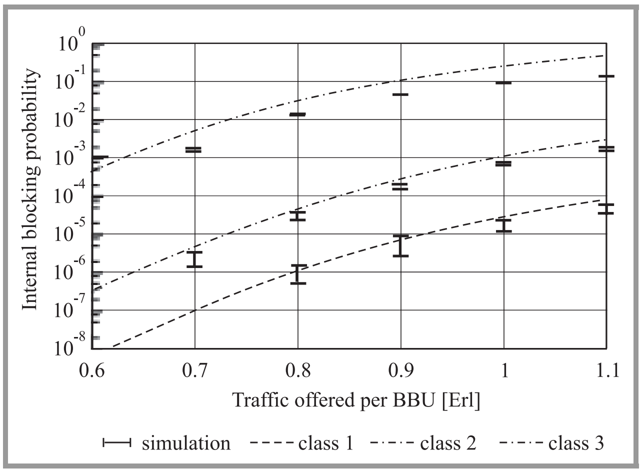


Fig. 7. Internal blocking with overflow links.

formed on the basis of the modified version of the recurrent method proposed in the article.

Figures 5 and 6 show a comparison of the results of the analytical calculations of the total blocking probability with the data obtained in the simulation experiments in a switching network without overflow links (Fig. 5) and in a network with overflow links (Fig. 6). The results confirm high accuracy of the proposed analytical method that is independent from the structure of the network, i.e., whether the system of overflow links was introduced or not. Figure 7 shows, in turn, the results of the analytical and the simulation modelling of the internal blocking probability in the switching network with overflow links. These results also confirm high accuracy of the proposed analytical method.

Figure 8 presents a percentage decrease in the value of the internal blocking probability after the introduction of the overflow links to the switching network for the analytical calculations (a) and in the case of the performed simulation experiments (b). On the basis of these graphs

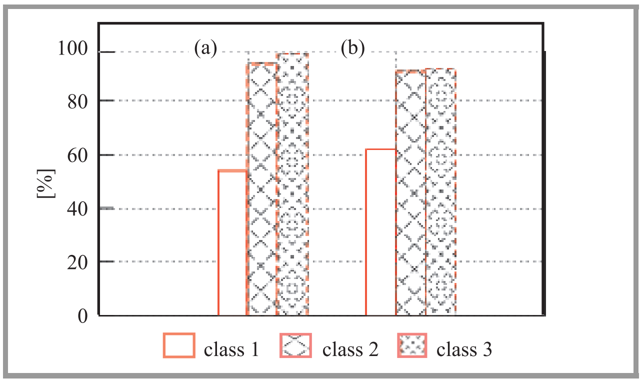


Fig. 8. Percentage decrease in the internal blocking probability.

it is possible to state that the introduction of overflow links results in a considerable percentage decrease in the internal blocking probability in the multi-service switching network. This fact indicates a need for further research into networks with overflow links as regards the possibility of the application of such systems in practice.

6. Conclusions

The article proposes a modified version of the recurrent SN-BPPRec method for analytical evaluation of the internal, external and the total blocking probability in multi-service switching networks with overflow links. The proposed method is characterised by high accuracy, both for modelling switching networks without overflow links and networks with overflow links. The conducted research study also indicates a possibility of the evaluation of the effectiveness of the system of overflow links in multi-service Clos switching networks. The network, in which each two neighbouring switches of the first stage are connected by overflow links, is characterised by a significant decrease in the value of the internal blocking probability. The obtained results indicate thus a need for further research on multi-service switching networks with overflow links that would be conducted within the context of a search for even more effective overflow systems, as well as, as regards of their future practical applications.

References

[1] W. Kabaciński, *Nonblocking Electronic and Photonic Switching Fabrics*. Berlin: Springer, 2005.

[2] A. Jajszczyk, *Wstęp do telekomutacji*. Warszawa: Wydawnictwa Naukowo-Techniczne, 1998 (in Polish).

[3] R. Fortet, *Systeme Pentaconta Calcul d'orange*. Paris: LMT, 1961.

[4] M. Stasiak, "Computation of the probability of losses in switching systems with mutual aid selectors", *Rozprawy Elekrotechniczne, J. Polish Academy of Science*, vol. XXXII, no. 3, pp. 961–977, 1986 (in Polish).

[5] H. Inose, T. Saito, and M. Kato, "Three-stage time-division switching junctor as alternate route", *Electron. Lett.*, vol. 2, no. 5, pp. 78–84, 1966.

[6] L. Katzschner, W. Lorcher, and H. Weisschuh, "On an experimental local PCM switching network", in *Proc. Int. Seminar Integr. System Speech, Video Data Communi.*, Zurich, Switzerland, 1972, pp. 61–68.

[7] V. Ershov, *Switching systems in integrated digital networks*. Moskov: Radio and Swiaz, 1978 (in Russian).

[8] M. Stasiak, M. Głąbowski, A. Wiśniewski, and P. Zwierzykowski, *Modeling and Dimensioning of Mobile Networks*. New York, USA: Wiley, 2011.

[9] V. Iversen, Ed., *Teletraffic Engineering Handbook*. Geneva: ITU-D, Study Group 2, Question 16/2, Jan. 2005.

[10] M. Stasiak and P. Zwierzykowski, "Performance study in multi-rate switching networks with additional inter-stage links", in *Proc. Seventh Adv. Int. Conf. Telecommun. AICT 2011*, M. Głąbowski and D. K. Mynbaev, Eds. St. Maarten, The Netherlands Antilles, 2011, pp. 77–82.

[11] M. Stasiak and P. Zwierzykowski, "Multi-service switching networks witch overflow links", *Image Process. Commun.*, vol. 15, no. 2, pp. 61–71, 2010.

[12] R. Wilkinson, "Theories of toll traffic engineering in the USA", *Bell System Tech. J.*, vol. 40, pp. 421–514, 1956.

[13] N. Binida and W. Wend, "Die Effektive Erreichbarkeit für Abnehmerbundel hinter Zwischenleitungsanungen", *Nachrichtentechnische Zeitung (NTZ)*, vol. 11, no. 12, pp. 579–585, 1959 (in German).

[14] A. Kharkevich, "An approximate method for calculating the number of junctions in a crossbar system exchange", *Elektrosvyaz*, no. 2, pp. 55–63, 1959.

[15] D. Bazlen, G. Kampe, and A. Lotze, "On the influence of hunting mode and link wiring on the loss of link systems", in *Proc. 7th Int. Teletraffic Congr.*, Stockholm, Sweden, 1973, pp. 232/1–232/12.

[16] A. Lotze, A. Roder, and G. Thierer, "PPL – a reliable method for the calculation of point-to-point loss in link systems", in *Proc. 8th Int. Teletraffic Congr.*, Melbourne, Australia, 1976, pp. 547/1–44.

[17] K. Rothmaier and R. Scheller, "Design of economic PCM arrays with a prescribed grade of service", *IEEE Trans. Commun.*, vol. 29, no. 7, pp. 925–935, 1981.

[18] M. Stasiak, "Blocage interne point a point dans les reseaux de connexion", *Annales des Télécommunications*, vol. 43, no. 9–10, pp. 561–575, 1988 (in French).

[19] A. Lotze, "History and development of grading theory", in *Proc. 5th Int. Teletraffic Congr.*, New York, NY, USA, 1967, pp. 148–161.

[20] E. Ershova and V. Ershov, *Digital Systems for Information Distribution*. Moscow: Radio and Communications, 1983 (in Russian).

[21] M. Stasiak, "Combinatorial considerations for switching systems carrying multi-channel traffic streams", *Annales des Télécommunications*, vol. 51, no. 11–12, pp. 611–625, 1996.

[22] M. Stasiak and P. Zwierzykowski, "Point-to-group blocking in the switching networks with unicast and multicast switching", *J. Perform. Eval.*, vol. 48, no. 1–4, pp. 249–267, 2002.

[23] M. Głąbowski and M. Stasiak, "Point-to-point blocking probability in switching networks with reservation", *Annales des Télécommunications*, vol. 57, no. 7–8, pp. 798–831, 2002.

[24] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Modelling product-form state-dependent systems with BPP traffic", *Perform. Eval.*, vol. 67, pp. 174–197, 2010.

[25] M. Głąbowski, "Recurrent calculation of blocking probability in multiservice switching networks", in *Proc. Asia-Pacific Conf. Commun.*, Busan, South Korea, 2006, pp. 1–5.

[26] M. Głąbowski, "Recurrent method for blocking probability calculation in multi-service switching networks with BPP traffic", in *Proc. 5th Eur. Perform. Engin. Worksh. Comp. Perform. Engin. CEPEW'08*, *LNCS*, N. Thomas and C. Juiz, Eds., vol. 5261. Springer, 2008, pp. 152–167.

[27] M. Głąbowski and M. Stasiak, "Internal blocking probability calculation in switching networks with additional inter-stage links", in *Information Systems Architecture and Technology*, A. Grzech, L. Borzemski, J. Świątek, and Z. Wilimowska, Eds. Oficyna Wydawnicza Politechniki Wrocławskiej, 2011, vol. Service Oriented Networked Systems, pp. 279–288.

[28] M. Głąbowski and M. D. Stasiak, "Switching networks with overflow links and point-to-point selection", in *Information Systems Architecture and Technology*, A. Grzech, L. Borzemski, J. Świątek, and Z. Wilimowska, Eds. Oficyna Wydawnicza Politechniki Wrocławskiej, 2012, vol. Networks Design and Analysis, pp. 149–159.

[29] M. Głąbowski and M. D. Stasiak, "Internal blocking probability calculation in switching networks with additional inter-stage links and Engset traffic", in *Proc. 8th IEEE, IET Int. Symp. Commun. Syst., Netw. Digit. Sig. Process. CSNDSP 2012*, Poznań, Poland, 2012.

[30] C. Clos, "A study of non-blocking switching networks", *Bell Syst. Techn. J.*, pp. 406–424, 1953.

[31] J. Roberts, V. Mocci, and I. Virtamo, Eds., *Broadband Network Teletraffic, Final Report of Action COST 242*. Berlin: Commission of the European Communities, Springer, 1996.

[32] J. Kaufman, "Blocking in a shared resource environment", *IEEE Trans. Commun.*, vol. 29, no. 10, pp. 1474–1481, 1981.

[33] J. Roberts, "A service system with heterogeneous user requirements – application to multi-service telecommunications systems", in *Proc. Perform. Data Commun. Syst. their Appl.*, G. Pujolle, Ed. Amsterdam: North Holland, 1981, pp. 423–431.

[34] C. Lee, "Analysis of switching networks", *Bell Syst. Techn. J.*, vol. 34, no. 6, 1955.

[35] M. Stasiak, "Blocking probability in a limited-availability group carrying mixture of different multichannel traffic streams", *Annales des Télécommunications*, vol. 48, no. 1-2, pp. 71–76, 1993.

[36] J. Tyszer, *Object-Oriented Computer Simulation Of Discrete-Event Systems*. Kluwer, 1999.

Mariusz Głąbowski received the M.Sc., Ph.D. and D.Sc. (Habilitation) degrees in Telecommunication from the Poznań University of Technology, Poland, in 1997, 2001, and 2010, respectively. Since 1997 he has been working in the Department of Electronics and Telecommunications, Poznań University of Technology. He is engaged in research and teaching in the area of performance analysis and modelling of multiservice networks and switching systems. Prof. Głąbowski is the author or co-author of 4 books, 7 book chapters and of over 100 papers which have been published in communication journals and presented at national and international conferences. He has refereed articles for many international conferences and magazines, including: IEEE Globecom, IEEE ICC, IEEE HPRS, IEEE Transactions on Communications, IEEE Communications Magazine, Computer Networks, IEEE Communications Letters, IEEE Transactions on Wireless Communications, Performance Evaluation, European Transactions on Telecommunications.

E-mail: mariusz.glabowski@put.poznan.pl
Chair of Communications and Computer Networks
Poznań University of Technology
Polanka st 3
60-965 Poznań, Poland

Michał Dominik Stasiak received his M.Sc. degree in Electronics and Telecommunications from Poznań University of Technology, Poland, in 2010. Since 2010 he is a Ph.D. student at Chair of Communications and Computer Networks at Poznań University of Technology. He is the co-author of a dozen scientific papers. He is engaged in research in the area of modeling of multi-service systems and networks particularly multi-service switching networks.

E-mail: michal.m.stasiak@doctorate.put.poznan.pl
Chair of Communications and Computer Networks
Poznań University of Technology
Polanka st 3
60-965 Poznań, Poland

Paper

Blind Estimation of Linear and Nonlinear Sparse Channels

Kristina Georgoulakis

Department of Science and Technology of Telecommunications, University of Peloponnese, Tripolis, Greece

Abstract—This paper presents a Clustering Based Blind Channel Estimator for a special case of sparse channels – the zero pad channels. The proposed algorithm uses an unsupervised clustering technique for the estimation of data clusters. Clusters labelling is performed by a Hidden Markov Model of the observation sequence appropriately modified to exploit channel sparsity. The algorithm achieves a substantial complexity reduction compared to the fully evaluated technique. The proposed algorithm is used in conjunction with a Parallel Trellis Viterbi Algorithm for data detection and simulation results show that the overall scheme exhibits the reduced complexity benefits without performance reduction.

Keywords—blind estimation and equalization, clustering techniques, sparse zero pad channels.

1. Introduction

The last years an intense research effort has risen in communications problems involving the estimation and equalization of sparse channels, i.e., channels with a large delay spread but with a small non zero support. Estimation and equalization of sparse channels is a challenging problem and sparsity aware estimators and equalizers should be used in order to improve system performance and reduce complexity. Sparse channels are encountered, among others, in High Definition Television (HDTV) [1], in broadband wireless communications [2] and in underwater acoustic channels [3].

Various training based estimators have been developed for the estimation of sparse channels [4]–[6]. Recently a blind algorithm based on the Expectation-Maximization (EM) algorithm for sparse channel estimation has been proposed [7]. The main drawback of the algorithm is its computational burden growing exponentially with the channel length. In this paper the Cluster Based Blind Channel Estimation algorithm (CBBCE) [8] is evaluated for a special class of sparse channels, the zero pad channels. The proposed algorithm, in order to exploit the structured sparsity of zero pad channels, uses a modification of the cluster based blind channel estimation procedure leading to a much lower complexity.

The CBBCE algorithm consists of two steps. Data clusters are first estimated via an unsupervised learning technique and next labelling of the estimated clusters is achieved by unravelling the information hidden in the sequence of received data. Labelling is performed using a Hidden Markov Model (HMM) of the estimation process and by relating data clusters to HMM states. The probability of each cluster to correspond to a specific label is treated as the unknown parameter of the HMM learning task implemented by the EM algorithm [8], [9].

Assuming an *M-ary* alphabet for the data symbols and a channel, with length $L+1$, the HMM is typically evaluated with M^L states [7]. However, in the sparse channel case where only a small fraction of the channel taps is active (i.e., $s+1 \ll L+1$), the full evaluation of the HMM is computationally inefficient [10]. In the zero pad channels all the non zero taps are placed on a regular grid [11], [12]. In this case, the memory of the channel concerns only the transmitted data that correspond to the non-zero taps. Thus, by involving in the HMM states only the $(s+1)$ data corresponding to the non zero taps, the number of states is reduced to M^s. Then, the reduced states HMM results in the appropriate labelling of the data clusters.

The outline of the paper is as follows. In Section 2 system description is given. The proposed CBBCE algorithm for sparse channels is presented in Section 3. In Section 4 the various channel cases where the algorithm can be applicable are referred (i.e., linear and non linear zero pad sparse channels, and a special case of group sparse channels). The tremendous complexity reduction achieved by the algorithm, compared to the full evaluated HMM, is also discussed in the same Section. In Section 5 the use of a Parallel Trellis Viterbi Algorithm (PTVA) [11] employing the channel estimates of the proposed algorithm is evaluated. The performance of the entire scheme in terms of the achieved Bit Error Rate (BER) is illustrated. Finally, conclusions are drawn in Section 6.

2. System Description

Consider the discrete time system described by:

$$g(t) = c(t) + w(t), \tag{1}$$

where

$$c(t) = F(I(t), I(t-1), \ldots, I(t-L)) \tag{2}$$

is the noiseless channel output sequence, $F(.)$ is the function representing the channel action, $I(t)$ is an equiprobable sequence of independent and identically distributed (i.i.d.) transmitted data taken from an *M-ary* alphabet, and $w(t)$ is Additive White Gaussian Noise (AWGN). The channel length is assumed to be $L+1$, however with only $s+1$ taps being non zero. The received data form $Q = M^{L+1}$

clusters in the one dimensional space [8]. Each cluster is represented by a suitably chosen representative which corresponds to the noiseless channel response, i.e.:

$$c(t) \in (c_k,\ k = 1, 2, \ldots, Q).$$

Here, due to channel sparsity, the actual number of clusters formed is:

$$Q = M^{s+1},$$

since the zero valued taps do not contribute to the formation of clusters.

Zero pad channels are sparse channels of a specific form, whose channel impulse response is described by [11]–[14]:

$$H = [h_1 \underbrace{0\ldots0}_{f\ zeros} h_2 \underbrace{0\ldots0}_{f\ zeros} \ldots h_s \underbrace{0\ldots0}_{f\ zeros} h_{s+1}]^T, \qquad (3)$$

where f is the number of zeros between the non zero valued taps and $L = s(f+1)$. In the case of zero pad channels, the received data depend on alternated data symbols [11], [13]. In this case, the noiseless channel output sequence (2) takes the form:

$$c(t) = F(I(t), I(t-(f+1)), ..., I(t-s(f+1))). \qquad (4)$$

3. Clustering Based Blind Channel Estimator for Zero Pad Sparse Channels

Channel estimation can be performed either using a known training sequence of data or identifying the channel based only on the received data (blind mode). Blind channel estimation based on data clustering techniques has been developed for general (non-sparse) channels [8], [9], [15]. The data clustering technique will be adopted in this paper to evaluate a blind estimator for structured sparse channels. Clustering based blind channel estimation is performed in two steps as it is detailed in [16] where initially the clusters representatives c_k are estimated via an unsupervised clustering technique following by clusters labelling, where each cluster is mapped to a specific sequence of transmitted data. When this technique is applied in the case of channels exhibiting a zero pad sparsity profile, the first step remains unaltered. In the second step, the structured sparsity of channels under investigation is taken into account resulting in a novel, reduced complexity labelling procedure. These two tasks are detailed in the sequel.

3.1. Unsupervised Clustering

An unsupervised learning technique is adopted for the estimation of the clusters representatives such as the Isodata algorithm, the Neural Gaz network, etc. [17], [18]. The clusters formed is the contribution of the non-zero taps of the channel only and the number of clusters estimated by the unsupervised clustering technique equals M^{s+1}.

3.2. Clusters Labelling through a Structured Sparsity Aware HMM

The transmitted data input vector:

$$\mathbf{I}(t) = [I(t)\ I(t-1)\ldots I(t-L)]^T,$$

can be described as a first order Markov chain having M^L states denoted by $S(t)$. Since the received data $g(t)$ are a probabilistic function of the state vector $\mathbf{I}(t)$, the channel estimation problem can be formulated as a HMM parameter estimation problem. Thus, a standard HMM parameter estimation algorithm, referred to hereafter as fully evaluated HMM algorithm (FE-HMM), considering M^L states can be applied, being however impractical from the computational point of view, apart from the case when the channel memory L is sufficiently small, which is not the typical case of sparse channels. Since the actual number of the clusters formed is $Q = M^{s+1}$ only, and for reasons of complexity reduction, the proposed algorithm considers M^s states in the HMM, resulting to a novel scheme referred to hereafter as the reduced evaluated HMM algorithm (RE-HMM). This task can be achieved considering instead the cluster model (4), where the actual channel memory pattern is taken into account. In this way, the labelling of the states of the HMM considers transmitted data that are $f+1$ time units apart. Based on the above remarks, the discrete observations RE-HMM for the sparse zero pad channel is characterized by the following elements:

- The states of the model, which according to Eq. (4) are formed as:

  $$S(t) \rightarrow (I(t-(f+1)), \ldots, I(t-s(f+1))). \qquad (5)$$

  The number of states in this case equals to:

  $$N = M^s,$$

  as opposed to the number of states $N' = M^L$ required by the FE-HMM approach.

- The state transition probabilities a_{ij}, which are defined as

  $$a_{ij} = P[S(t+(f+1)) = j \mid S(t) = i], \qquad (6)$$
  $$1 \leq i, j \leq N.$$

  Notice that for each allowable transition ($a_{ij} = 1/M$) a specific noiseless channel output occurs. In other words, each state transition specifies uniquely a cluster label and a cluster transition arises every $f+1$ samples. The cluster labels are specified by:

  $$X(t) \rightarrow (I(t), I(t-(f+1)), \ldots, I(t-s(f+1))) \qquad (7)$$

  and each cluster label, $X(t)\ \varepsilon\ (n_k,\ k = 1, 2, \ldots, Q)$, corresponds to a specific cluster c_k.

- The distinct observation symbols per transition, which in this case are the clusters, c_k.

- The probabilities for each symbol to occur and for each state transition i to j, which denote the probability of a specific cluster to correspond to a specific label, i.e.:

$$\begin{aligned} b_{n_k}(c_k) &= P[c_k \mid S(t)=i, S(t+(f+1))=j] \\ &= P[c_k \mid X(t)=n_k], \end{aligned} \tag{8}$$

$$1 \le k \le Q, \quad 1 \le i,j \le N.$$

- The initial state distribution:

$$\pi_i = P[S(1)=i], \tag{9}$$

$$1 \le i \le N.$$

In the cluster based blind channel estimation procedure clusters' labelling is treated as a HMM learning problem. The EM algorithm is a commonly used numerical iterative scheme to obtain Maximum Likelihood (ML) estimates of a HMM. The resulting ML estimate is given by:

$$\hat{\theta} = argmax_{\theta} P(G \mid \theta), \tag{10}$$

where $P(G \mid \theta)$ denotes the probability of the observation sequence G of length T:

$$G = (g(1), g(2), \dots g(T))^T,$$

given the model parameters (θ) with:

$$\theta = [b_{n_k}(c_k)],\ k = 1, ..., Q.$$

Thus, θ is the $Q \times Q$ probability matrix that maps labels to clusters and it is expected to converge to a matrix whose elements converge either to one, for the case when a specific symbol corresponds to a specific label, or to zero, otherwise. Convergence of the algorithm is achieved when:

$$P(G \mid \theta) > p, \tag{11}$$

with p a predetermined threshold [19].

Clusters' labelling using the RE-HMM approach described by Eqs. (5)–(9) requires the knowledge of the structure of the comb type channel response, which in turn requires the estimation of the distance or the number of unit time delays between all successive non zero elements of the model. In the case of zero pad sparse channels treated in this paper, and due to the specific form of the sparsity structure only a single time delay parameter has to be determined. The required time delay parameter d is estimated using an exhaustive search procedure, starting from $d = 1$, where for each candidate value d, a RE-HMM estimate $\hat{\theta}$ is obtained using Eqs. (5)–(9). When the algorithm converges (11), then the correct value of d is reached ($d = f + 1$) and the correct channel structure is obtained. Then, θ provides the correct labelling.

The proposed CBBCE algorithm is summarized in Table 1. This procedure is further illustrated by a simple example using a channel with impulse response:

$$H = [h_1 \quad h_2 \quad h_3]^T = [1 \quad 0 \quad 0.5]^T,$$

Table 1
The proposed Clustering Based Blind Channel Estimation algorithm

CBBCE algorithm

1. Unsupervised clustering:

- Estimation of the clusters representatives by an unsupervised learning technique.
- The number of the estimated clusters reveals the number of non-zero taps $(s+1)$.

2. Labelling through a RE-HMM

Initialization

Set: Number of states, $N = M^s$

Time delay parameter, $d = 1$

Main

Repeat until convergence (11)

HMM formulation ((5)–(9)) with states:

$S(t) = (I(t-d)I(t-2d)\dots I(t-sd)).$

$d = d + 1$

End

- The correct value of d is reached $(f+1)$,
- The ML estimate, (θ), reveals the labels – clusters correspondence.

where the input data are assumed to be bipolar (i.e., $I(t) = \pm 1$) and the Signal to Noise Ratio (SNR) is set equal to 17 dB. In this particular case, the number of non-zero taps is $s + 1 = 2$, while the channel length is $L + 1 = 3$. Since the data alphabet consists of two symbols, the number of clusters assumed is $Q = 4$. Following the first step of the proposed algorithm, the clusters representatives are estimated using an unsupervised clustering technique. Specifically, Isodata is used for clusters estimation, using $T = 30$ received data, obtaining the estimates:

$$\hat{c}_1 = 1.512,\ \hat{c}_2 = 0.49,\ \hat{c}_3 = -0.507,\ \hat{c}_4 = -1.49.$$

Following the second step of the proposed algorithm, the RE-HMM is formed, with $N = 2$. Initially, d is set equal to 1, thus, a cluster transition is assumed to arise every single sample. Since the assumed channel model is not the correct $((d = 1) \neq (f + 1 = 2))$, the algorithm does not converges according to Eq. (11). The probability matrix θ, after 15 iterations, is shown in Table 2. Obviously, the identification procedure does not converge and no labels – clusters correspondence can be derived. Then, the time

Table 2
Probabilities matrix for channel H = [1 0 0.5]^T after 15 iterations and $d = 1$. Clusters labels cannot be unravelled

Label		Cluster representative			
I(t)	I(t–1)	c_1	c_2	c_3	c_4
–1	–1	0.1235	0.2393	0.4301	0.2071
–1	1	0.2703	0.0632	0.2423	0.4242
1	–1	0.1841	0.5817	0.1136	0.1206
1	1	0.3477	0.1588	0.1175	0.3759

delay parameter, d, is set equal to 2 and a new RE-HMM is formed. This time the algorithm converges. The probability matrix θ, after 10 iterations, converges as it appears in Table 3. In this case the labels – clusters mapping is easily achieved.

Table 3
Probabilities matrix after convergence, for the channel with impulse response H = [1 0 0.5]^T and time delay parameter $d = 2$

Label		Cluster representative			
I(t)	I(t–1)	c_1	c_2	c_3	c_4
–1	–1	0	0	0	1
–1	1	0	0	1	0
1	–1	0	1	0	0
1	1	1	0	0	0

Once the channel estimation process is accomplished, signal detection can be performed by employing a PTVA with reduced complexity [10]. The PTVA is a computationally improved reformulation of the Viterbi Algorithm (VA) which operates into a set of independent trellises for the zero pad sparse channels. The PTVA is optimum for zero pad sparse channels and results in complexity reduction compared to the ordinary VA which uses a single trellis. The evaluation of a channel equalizer employing the CBBCE algorithm followed by a PTVA is described in Section 5.

4. Case Studies

We proceed further our developments on case studies, where a variety of channels amenable to the application of the proposed channel identification method is considered. Complexity issues are also discussed. Notice that, for the sake of simplicity, the symbol values are assumed to be drawn from a binary alphabet set (i.e., $M = 2$).

4.1. Linear Zero Pad Sparse Channels

Linear zero pad sparse channels are successfully identified using the proposed method. Note that, the presence of an arbitrary time delay (number of zeros) at the edges of the non zero taps of the channel does not affect the algorithm, and channels with impulse response of the form:

$$H = [\underbrace{0\dots0}_{x\ zeros} h_1 \underbrace{0\dots0}_{f\ zeros} h_2 \underbrace{0\dots0}_{f\ zeros} \dots h_{s+1} \underbrace{0\dots0}_{y\ zeros}]^T \,, \qquad (12)$$

can be tackled by the method, including, for $f = 0$, a special case of group sparse channels where the non zero taps are located in a single cluster [6]. Consider for example a channel with impulse response given by:

$$H = [0.2\ 0\ 0\ 0\ 0\ 0\ 0.5\ 0\ 0\ 0\ 0\ 0\ 0.9]^T. \qquad (13)$$

In this case we get $L+1 = 11$, $s+1 = 3$ and $f = 4$. Here, the clusters representatives $\hat{c}_k$, $k = 1,..,Q$ are estimated using the Isodata algorithm [18] from a received data sequence of length $T = 300$ [20]. Once clusters identification is completed, the task of clusters labelling is subsequently addressed. A RE-HMM with $N = 2^2$ states is formulated. Application of the proposed algorithm, as it is summarized in Table 1, results in $d = f + 1 = 5$ and the states of the RE-HMM are formed by the (non-successive) data $S(t) = (I(t-5), I(t-10))$. The probabilities matrix θ resulting from the proposed algorithm, after convergence, is tabulated in Table 4.

Table 4
Probabilities matrix after convergence, for the channel with impulse response H = [0.2 0 0 0 0 0.5 0 0 0 0 0.9]^T, $d = 5$

Label			Cluster representative							
I(t)	I(t–5)	I(t–10)	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
–1	–1	–1	0	0	0	1	0	0	0	0
–1	–1	1	0	0	1	0	0	0	0	0
–1	1	–1	0	0	0	0	0	0	1	0
–1	1	1	0	1	0	0	0	0	0	0
1	–1	–1	0	0	0	0	1	0	0	0
1	–1	1	0	0	0	0	0	0	0	1
1	1	–1	0	0	0	0	0	1	0	0
1	1	1	1	0	0	0	0	0	0	0

The Average Squared Error (ASE) is adopted as a metric of the accuracy of the estimated clusters representatives:

$$m = \frac{1}{Q} \sum_{k=1}^{Q} (c_k - \hat{c}_k)^2 \,, \qquad (14)$$

where $\hat{c}_k$ are the estimated clusters representatives and c_k are the noiseless clusters representatives that correspond to Eq. (13).

The ASE versus SNR is illustrated in Fig. 1. Figure 2 shows the impact of the received data sequence length, T, to the accuracy of the estimated clusters values, for SNR = 20 dB.

For the sake of comparison, a supervised Least - Absolute Shrinkage and Selection Operator (LASSO) [21] estimator is used as benchmark. Since LASSO is not capable

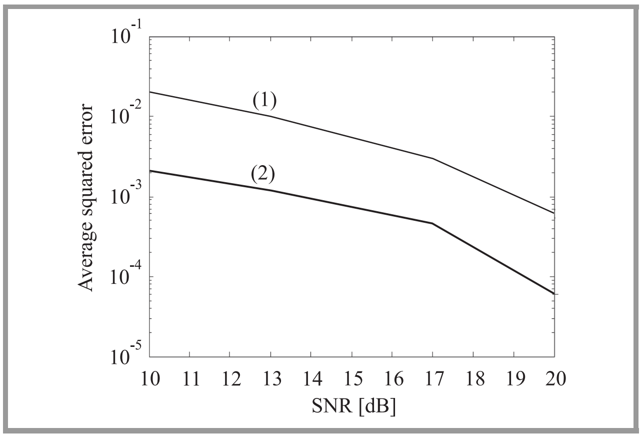


Fig. 1. Average squared error for a channel with impulse response H = [0.2 0 0 0 0 0.5 0 0 0 0 0.9]^T, using a sequence of $T = 300$ data. (1) – proposed CBBCE algorithm, (2) – supervised LASSO channel estimation.

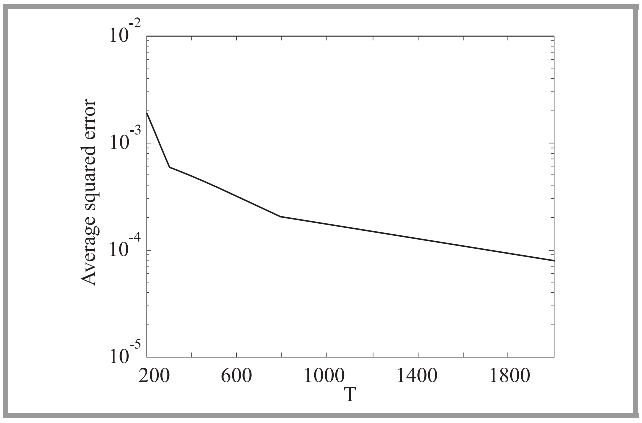


Fig. 2. Average squared error versus T, (number of received data used by the proposed CBBCE), for a channel with impulse response H = [0.2 0 0 0 0 0.5 0 0 0 0 0.9]^T, SNR = 20 dB.

of estimating the clusters representatives directly, the estimates of channel taps are used to calculate the respective clusters representatives. The ASE versus SNR in the case where LASSO is used for channel identification, is also shown in Fig. 1. From a first glance it is evidence that the proposed estimator lacks behind its supervised counterpart, a result which is somehow expected since LASSO is a supervised learning algorithm, while the proposed scheme is a blind identification algorithm. However, as it is shown in the following Section the BER performance of an equalizer using the estimates of the proposed algorithm is very close to that using the supervised LASSO as a channel estimator.

4.2. Nonlinear Zero Pad Sparse Channels

Clustering based estimation algorithms do not adopt any assumption for the impulse response of the channel under consideration, thus, they can efficiently be employed in the case of nonlinear channels [8]. For example a sparse linear channel with impulse response given by $H = [0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0.5\ 0\ 0\ 0\ ]^T$ followed by the nonlinear action described by $g(t) + 0.1g(t)^2 + 0.3g(t)^3$ is considered. In this case we get, $L = 10$, $s = 1$ and $f = 4$. The proposed CBBCE algorithm uses $T = 80$ received data and performs clusters estimation with only $N = 2^1$ states. The resulting ASE versus SNR is shown in Fig. 3.

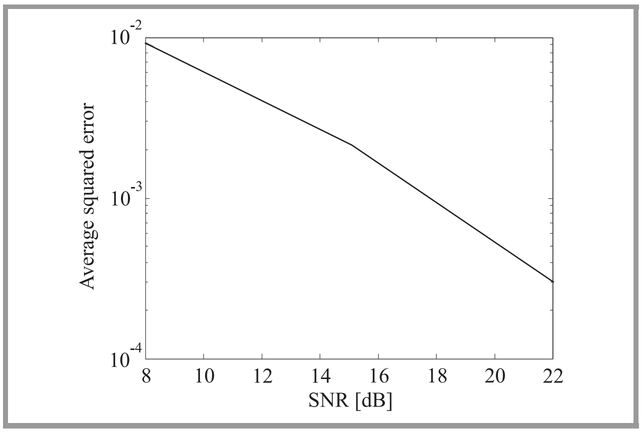


Fig. 3. Average squared error achieved by the proposed blind estimator, for a channel with impulse response H = [0 0 1 0 0 0 0 0.5 0 0 0]^T followed by the nonlinear action described by $g(t) + 0.1g(t)^2 + 0.3g(t)^3$.

4.3. Complexity Assessment

The proposed algorithm reduces the complexity of the HMM scheme from $O(M^L \times T)$ required by the FE-HMM to $O(M^s \times (f+1) \times T)$ operations, required by the proposed RE-HMM, which is a tremendous reduction in the sparse channels case where $s \ll L$. For example, in the experiment described in Section 4.1, the proposed RE-HMM algorithm evaluates the HMM scheme using only $N = 2^2$ states. In this particular case, the HMM procedure is repeated $f + 1 = 5$ times resulting in a complexity of $O(2^2 \times 5 \times T)$ operations which is a major improvement over the FE-HMM algorithm [7], [8] which requires $N' = 2^{10}$ states leading to a complexity of $O(2^{10} \times T)$ operations. In the experiment described in Section 4.2 the complexity of the RE-HMM reaches the $O(2 \times 5 \times T)$ operations while the FE-HMM involves $O(2^{10} \times T)$ operations.

5. Blind Clustering Based Equalizer for Zero Pad Channels

The proposed method can also be applied in the case when instead of channel estimation, channel equalization is under consideration. The proposed CBBCE algorithm combined with a PTVA performs blind clustering based sequence equalization, for the special case of zero pad channels, in an efficient way. We refer to the entire proposed scheme as Reduced Evaluated Blind Equalizer (REBE). Consider for example a channel with impulse response given by $H = [0.2\ 0\ 0\ 0\ 0\ 0.5\ 0\ 0\ 0\ 0\ 0.9]^T$ (13). The proposed CBBCE algorithm has already been described, for this channel, in Section 4.1. Since channel estimation is completed the PTVA algorithm is used for signal detection. The PTVA algorithm uses $f+1=5$ parallel trellises with $M^s=4$ states each [11]. The decision delay for the PTVA is 15. The resulting BER for the proposed REBE appears in Fig. 4.

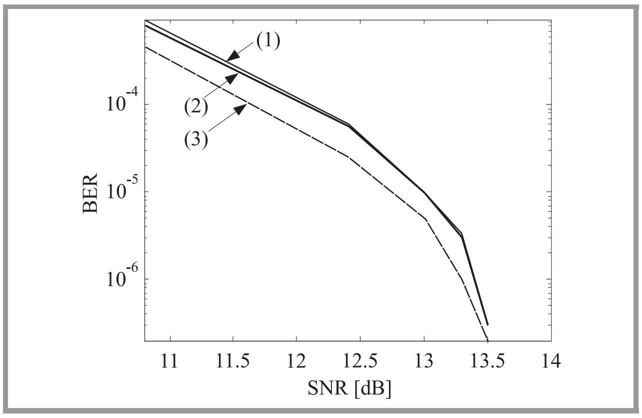


Fig. 4. BER versus SNR for a channel with impulse response H = $[0.2\ 0\ 0\ 0\ 0\ 0.5\ 0\ 0\ 0\ 0\ 0.9]^T$. (1) – proposed REBE (proposed reduced blind estimation algorithm and PTVA algorithm), (2) – FEBE (blind full evaluated clusters estimation algorithm and full evaluated VA), (3) – supervised LASSO channel estimation and full evaluated VA.

The performance of the conventional Full Evaluated Blind clustering based Equalization scheme [8] (FEBE) is also investigated. The FEBE performs clusters identification by an unsupervised clustering technique and clusters labelling using the FE-HMM and data detection through a conventional VA using $M^L = 2^{10}$ states. The decision delay for the VA is 30.

As seen from the Fig. 4 the performance of the FEBE is the same with the REBE.

Moreover, for the sake of comparison, an equalization scheme formed by a supervised estimator and a full – evaluated VA is realized and used as a benchmark. Channel estimation is achieved by the supervised LASSO algorithm [21]. Then the channel estimator is followed by a conventional VA, with 2^{10} states. The number of training data used for the estimator is 300. The decision delay for the VA is 30. As seen from Fig. 4 the resulting BERs of the three schemes are very close, however, the proposed REBE works at a substantially reduced complexity.

6. Conclusions

In this paper a novel reduced complexity blind estimator for zero pad channels is presented. The proposed scheme uses a Clustering Based Blind Channel Estimation algorithm extended to account for the structured sparsity of zero pad channels and exhibits a tremendous complexity reduction compared to the full evaluated counterpart. The algorithm is suitable both for linear and nonlinear channels. The proposed algorithm combined with a Parallel Trellis Viterbi Algorithm is used for signal detection and the proposed sequence equalization scheme exhibits, at a reduced complexity, a performance similar to that compared to other competitive schemes such as a full evaluated blind clustering based sequence equalizer and a supervised LASSO estimator accompanied by a Viterbi Algorithm. Modification of the algorithm for the expansion of its use to tackle general sparse channels is under investigation.

References

[1] W. F. Schreiber, “Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions”, *Proc. IEEE*, vol. 83, pp. 958–981, 1995.

[2] S. Ariyavisitakul, N. R. Sollenberger, and L. J. Greenstein, “Tap selectable decision feedback equalization”, *IEEE Trans. Commun.*, vol. 45, pp. 1497–1500, 1997.

[3] M. Kocic, D. Brady, and M. Stojanovic, “Sparse equalization for real time underwater acoustic communications”, in *Proc. MTS/IEEE Conf. “Challenges of our changing global environment” OCEANS’95*, San Diego, CA, USA, 1995, pp. 1417–1422.

[4] J. A. Tropp and S. J. Wright, “Computational methods for sparse solution of linear inverse problems”, *IEEE Proc.*, vol. 98, pp. 948–958, 2010.

[5] B. Babadi, N. Kalouptsidis, and V. Trokh, “SPARLS: The sparse RLS algorithm”, *IEEE Trans. Sig. Proces.*, vol. 58, pp. 4013–4025, 2012.

[6] Y. Chen and A. O. Hero, “Recursive $\ell_{1,\infty}$ Group lasso”, *IEEE Trans. Sig. Proces.,* vol. 60, pp. 3978–3987, 2012.

[7] G. Mileounis, N. Kalouptsidis, B. Babadi, and V. Tarokh, “Blind identification of sparse channels and symbol detection via the EM algorithm”, in *Proc. IEEE DSP Conf.*, Corfu, Greece, 2011, pp. 1–5.

[8] K. Georgoulakis and S. Theodoridis, “Blind and semi-blind equalization using hidden Markov models and clustering techniques”, *Sig. Proces.*, vol. 80, pp. 1795–1805, 2000.

[9] P. Wang and W. Ser, “On cluster-based channel identification”, *Procedia Engin.*, vol. 29, pp. 2699–2704, 2012.

[10] R. Niazadeh, S. H. Ghalehjegh, M. B. Zadeh, and C. Jutten, “ISI sparse channel estimation based on SLO and its application in ML sequence-by-sequence equalization”, *Sig. Proces.*, vol. 92, pp. 1875–1885, 2012.

[11] N. C. McCinty, R. Kennedy, and P. Hoeher, “Parallel trellis Viterbi algorithm for sparse channels”, *IEEE Communi. Lett.*, vol. 2, pp. 143–145, 1998.

[12] B. Cristea, D. Roviras, and B. Escrig, “Maximum likelihood sequence estimation based on periodic time-varying trellis for LPTVMA systems”, in *Proc. EUSIPCO 2006*, Florence, Italy, 2006.

[13] J. Mietzner, S. Hoeher, S. Land, and P. Hoeher, “Trellis equalization for sparse ISI channels revisited”, in *Proc. IEEE Int. Symp. Inform. Theory ISIT 2005*, Adelaide, Australia, 2005, pp. 229–233.

[14] F. K. H. Lee and P.J. McLane, “Iterative parallel-trellis MAP equalizers with nonuniformly-spaced prefilters for sparse multipath channels”, in *Proc. 56th VTC 2002*, Vancouver, Canada, 2002, vol. 4, pp. 2201–2205.

[15] Y. L. Jeng and C. C. Yeh, “Cluster based blind nonlinear channel estimation”, *IEEE Trans. Sig. Proces.*, vol. 45, pp. 1161–1172, 1997.

[16] K. Georgoulakis and S. Theodoridis, "Efficient clustering techniques for channel equalization in hostile environments", *Sig. Proces.*, vol. 58, pp. 153–164, 1997.

[17] T. M. Martinetz, S. G. Bekovich, and K. J. Schulten, "Neural-gas network for vector quantization and its application in time-series prediction", *IEEE Trans. Neural Netw.*, vol. 4, pp. 558–569, 1993.

[18] S. Theodoridis and K. Koutroubas, *Pattern Recognition*. New York: Academic Press, 1998.

[19] G. Kawas and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels", *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, 1994.

[20] F. Balazs and S. Imre, "Modified radial basis network based blind channel estimation", in *Proc. SoftCOM 2004*, Dubrovnik, Croatia, 2004, pp. 449–452.

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. Royal Stat. Soc.*, vol. 58, pp. 267-288, 1996.

Kristina Georgoulakis received the B.S. degree in Informatics from the Department of Informatics and Telecommunications, University of Athens, Greece, in 1990 and her M.Sc. and Ph.D. degree in 1993 and 1999, respectively, both from the same department. Her research interests are in signal processing for communications systems.

E-mail: kristina@uop.gr
Department of Science and Technology
of Telecommunications
University of Peloponnese
Tripolis, 22100 Greece

Paper

Hidden Context Influence on Pattern Recognition

Juliusz L. Kulikowski

Nałęcz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland

Abstract—The influence of hidden additional information concerning the circumstances of input data acquisition on the quality of decisions based on the data is considered. An analogue to the intuition influencing natural decision making is indicated. The problem of contextual information in decision making based on Bayes rule, on reference data sets in various applications as well as of scene analysis by numerous examples is illustrated.

Keywords—decision making, hidden information, intuition, pattern recognition, scene analysis.

1. Introduction

Since the very beginning of the pattern recognition development as a scientific discipline (i.e., since the middle of 1950's) many attempts were made to make the artificial pattern recognition methods as effective as the natural one were made. Till now, the early expectations of a possible quick reaching this goal by using the approaches to pattern recognition based on artificial neural networks, geometrical, statistical, formal linguistic, algebraic or any other advanced types of models and based on them algorithms only partially came to be true. In many application areas (e.g., in medical diagnosis, biology, criminology, geophysics, etc.) a verification of the computer-aided pattern recognition results by those of human pattern recognition takes place rather than the reverse. Such a situation exists not only because of a much lower number of artificial neural networks' elements than this of the neurons in a natural human brain or because of too low computer performance rates, but also because of an imperfection of the actually available artificial recognition models and methods. One of the human mind's properties which in the to-date artificial intelligence methods is neglected is the **intuition**. The diagnoses made by the experienced medical specialists, even if based on apparently similar or the same input diagnostic data, are usually more accurate than those made by the low-experienced physicians as well as those made by the computer-aided diagnostic systems. This fact is usually explained by the intuition supporting the experienced specialist's thinking and by a lack of the intuition in the two other cases.

Many attempts to explain the phenomenon of intuition were made by the philosophers (e.g., by Aristotle, Plato, R. Descartes, I. Kant, J. Locke, H. Bergson, etc.) [1]. They assumed a primacy of intuitive cognitive process over the one based on the observations. On the other hand, L. Wittgenstein assumed that such metaphysical concepts as intuition in scientific researches should not be taken into considerations [2]. This point of view on the intuition was shared by some naturalists in the past century (e.g., by J. B. Watson and other authors representing a behavioural approach in psychology [3]). Nevertheless, many examples of subconscious processes existence in human thinking and of their influence on human behaviour can be shown [4]. S. Freud tried to explain many aspects of human behavior by subconscious mental processes [5]. The role of the intuition in creative thinking and in discoveries was analyzed by many authors, like R. S. Siegler and E. Stern [6], A. Motycka [7], M. Polanyi [8] and, in particular, in mathematics by B. L. J. Brouwer [9], J. Hadamard [10]. A simple definition of intuition as "…a direct, by a preliminary analysis and/or reasoning not preceded knowledge about something…" is given in [11]. Similarly, D. G. Myers [4] defines intuition as our "ability to getting access to immediate knowledge, prompt inspection into it without observations or reasoning". A. P. Wierzbicki [12] formulates the most general, intuition concerning question: "…what is intuition: is it a supernatural ability of human mind distinguishing it from animals and producing infallible truths or is it rather a very powerful but natural ability, common with animals, producing new ideas but such that require justification and evaluation?".

The paper presented below is an extended version of a non-published plenary speech delivered at the 7th International Conference on Computer Recognition Systems CORES held in Wrocław, Poland, in May of 2011. The following problem is here considered: is intuition a natural mental property worthy and possible to be simulated by artificial decision making systems? If so, then neglecting any its metaphysical aspects, what could it mean from an information processing point of view? In the context of the main problem the following particular items are considered below: a working definition of intuition based on informational approach is proposed in Section 2. Influence of additional information on the decisions based on Bayes rule are described in Section 3. Similar problem concerning pattern recognition based on reference sets is considered in Section 4. The role of contextual information in scene analysis is shortly illustrated in Section 5. Conclusions are formulated in Section 6.

2. Intuition – a Working Concept

Before answering the first of the above formulated questions two remarks should be made:

1) The hidden intuitive mechanisms governing creative (e.g., mathematical, artistic, musical, etc.) thinking are different from those influencing animal or human behavior in suddenly arising situations. In the first case, a sort of unconscious evoking and associating of possible solutions of a problem, while in the second case rather evoking some remembered in the past similar situations and their effects take place. In addition, the results in both cases are different. In the first case, it is an "illumination", i.e., a sudden thought about a possible problem solution, while in the second case it may be an impression of an irrational fear, disapproval, distrust, etc., or in an opposite case it may take the form of a hope, belief, confidence, etc. The last type of intuition has been created, probably by a natural evolution, as a form of adaptation of living beings to a necessity of effectively react to unexpected events of high living importance: a danger of an attack, an occurrence of a potential sexual partner, a possibility of food reaching, etc.

2) Not all intuitively suggested decisions are correct or useful. D. G. Myers [4] describes numerous examples of wrong decisions inspired by intuition. In particular, he remarks that mathematically optimal decisions differ essentially from intuitive decisions based on the most impressive, similar to the present one cases. A noise heard in darkness not always is caused by a wild animal; a smile not always means that somebody is well predisposed to us. Our first impression after arrival to an unknown town is not always correct. The examples show that our intuitive reaction to sudden event is not always justified in a given situation, sometimes it may be evidently unfavourable or wrong.

The simulation of intuition in artificial decision making systems, if possible, is thus desirable only if it helps in system's action improving or in its better decisions making. For this (and in particular, for pattern recognition) purpose an informational model of intuition seems to be suitable. Intuition can thus be defined as an ability of a subject to unintended admission, in certain attending side-circumstances, affecting his (its) behavior or decisions by remembered impact of analogous circumstances on experienced in the past situations.

The informational model of intuition is schematically presented in Fig. 1. It shows that the behavior of a subject and following from it effects evoked by the same stimulus may be different, because of neglecting or unconscious taking into account additional information about the circumstances attending the stimuli. This corresponds to the well known situations of some human decisions that are assessed as "irrational" by other people, because they subconsciously have been affected by information that is unavailable to other people, kept in the mind of the decision making subject.

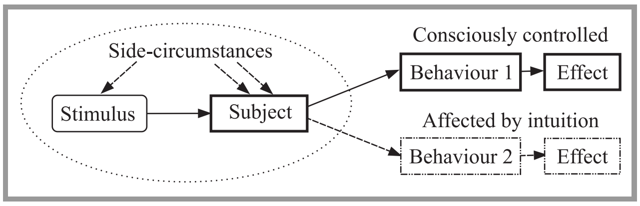


Fig. 1. Informational model of subject's behaviour affected by intuition.

Side-circumstances in Fig. 1 mean in this context that information about them has not been directly included into the decision rules. Otherwise speaking, they constitute a sort of informational context of decisions made under incompleteness of information. Intuition reduces the information incompleteness level by taking into account additional information about the circumstances which may modify the decision. Therefore, an additional (hidden) information may affect the recognition and increase its quality. In the nature, past information influence on human behavior is connected with sub cortical informational processes in the mind. In computer systems such influence may consist in automatic (i.e., hidden for the user) registration of data concerning circumstances attending current decision-making, and in using them in future pattern recognition acts. The concept of including programs that automatically start operations into the computer system is not new; each operational system contains many hidden operations protecting the user against errors, and application programs contain operations protecting against data misusing. The question is whether the above-mentioned cases example an "intuition" of the computer systems. It seems not as they are limited only to an automatic formal analysis of input data or to a hidden control of the calculating operations.

As a more advanced case, it can be mentioned the M. Minsky's concept of "daemons", special procedures included into the knowledge representing systems based on "frames". The role of "daemons" consists in automatic completing the contents of frames by lacking elements, explicitly to the system not provided but existing in the input data in a hidden form [13]. However, even in this case additional information does not concern the circumstances attending the past decisions of the system. Close to the "artificial intuition" concept is this of information retrieval systems equipped by the mechanisms of the replies modification by taking into account the currently modified informational profiles of the users [14]. From the user's point of view, the system able to guess the probable user's requirements for information is equipped with a sort of "intuition". However, this "intuition" is limited to user's preferences only. Below, a more general case of "intuition" consisting in affecting the current pattern recognition by the information about the circumstances of past pattern recognition acts will be considered. Taking into account that additional information affecting current decisions are based on collected in the past experiences, the "artificial intuition"-based information systems can be considered as a sort of self-learning systems. However, like in natural intuition, this additional information may be charged with a dose of uncertainty.

3. Bayes-Rule-Based Pattern Recognition

The Bayes-rule based pattern recognition methods belong to the most early elaborated ones [15]. The simplest case of two recognized classes of objects (patterns), C_1 and C_2 will be here reminded. Let U denote a multidimensional vector space whose elements (vectors) $\mathbf{u}$ represent the objects of a given physical nature. It is assumed that the object occurs randomly. Hence, the classes C_1, C_2 can be described, respectively, by the conditional probability densities $v(\mathbf{u}|C_1)$, $v(\mathbf{u}|C_2)$ and by their a priori probabilities p_1, p_2 where $p_1 + p_2 \equiv 1$. We define a decision function as a function assigning to any vector $\mathbf{u} \in U$ of, as one of the two above-mentioned classes:

$$\chi : U \to \{C_1, C_2\}. \tag{1}$$

Let us also denote by $P_{1|2}$ a conditional probability of recognition of an object $\mathbf{u}$ as belonging to C_1 when, in fact, it belongs to C_2, and by $P_{2|1}$ a conditional probability of recognition $\mathbf{u}$ as belonging to C_2 when, in fact, it belongs to C_1. The optimal Bayes sense decision function $\chi(\mathbf{u})$ is the one that minimizes the mean decision risk:

$$R = P_{1|2} + P_{2|1}\,. \tag{2}$$

It is known [15] that such function should have the form:

$$\chi(\mathbf{u}) = \begin{cases} C_1 & \text{if} \quad \lambda(\mathbf{u}) > \Lambda, \\ C_2 & \text{if} \quad \lambda((\mathbf{u}) < \Lambda, \end{cases} \tag{3}$$

(if $\lambda(\mathbf{u}) = \Lambda$ any of two decisions is admissible), where $\lambda(\mathbf{u})$ is a pre-decision function:

$$\lambda(\mathbf{u}) = \frac{v(\mathbf{u}|C_1)}{v(\mathbf{u}|C_2)} \tag{4}$$

and

$$\Lambda_0 = \frac{P_2}{P_1} \tag{5}$$

is a threshold level [15]. In the simplest case, if $P_1 = P_2 = P$ it is $\Lambda = 1$ and the rule (3) takes a simpler form:

$$\chi(u) = \begin{cases} C_1 & \text{if} \quad v(\mathbf{u}|C_1) > v(\mathbf{u}|C_2)\,, \\ C_2 & \text{if} \quad v(\mathbf{u}|C_1) < v(\mathbf{u}|C_2)\,, \end{cases} \tag{6}$$

as it (in a one-dimensional case) is illustrated in Fig. 2.

For $\Lambda_0 = 1$ a threshold point u_0 between the C_1 and C_2 decision areas, according to Eq. (6) is given by the equation $v(\mathbf{u}|C_1) = v(\mathbf{u}|C_2)$. In such case, the decision risk R is equal to the sum of the areas under the curve $v(\mathbf{u}|C_1)$ for $u > u_0$ and under the $v(\mathbf{u}|C_2)$ for $u < u_0$. Till now, no "intuition" affects the decision. However, let us assume that

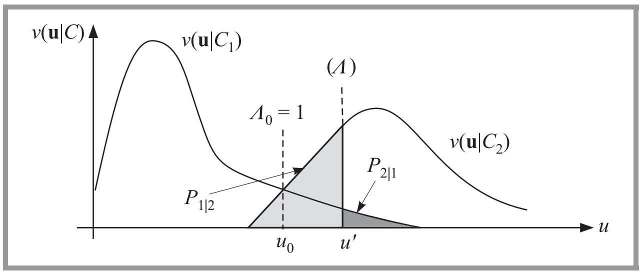


Fig. 2. Bayesian pattern (signal) recognition rule.

for certain reasons, the errors consisting the recognition of C_2 instead of C_1 lead to much larger "costs" than recognition of C_1 instead of C_2. In such case, a user intuitively may believe that it is better to shift the threshold point from u_0 to u', $u_0 < u'$, as shown in Fig. 2. Evidently, the risk R given by Eq. (2) is in this case increased (as being equal to the sum of the gray areas in Fig. 2). It can be shown that it corresponds now to a new threshold value Λ in the rule (3), for $P_1 = P_2 = P$ given by:

$$\Lambda = \frac{P_2 r_{1|2}}{P_1 r_{2|1}} = \frac{r_{1|2}}{r_{2|1}}\,, \tag{7}$$

where $r_{1|2}/r_{2|1}$ denotes a relative cost assigned to the errors of the "C_1 instead of C_2" with respect to this of "C_2 instead of C_1" type. This relative cost can be assessed due to remembered effects (e.g., for the user) of wrong decisions made in the past, i.e., of connected with them circumstances.

The above-presented example of intuitive cautiousness influence on pattern recognition can easily be extended on more realistic, parametric Bayesian decision rules. For this purpose, it will be assumed that the conditional probability density functions of input signal have the form $v(\mathbf{u}|C_1;\gamma_1)$, $v(\mathbf{u}|C_2;\gamma_2)$, where a binary recognition problem once again is considered and $\gamma_1, \gamma_2 \in \Gamma, \Gamma$ being a set of *passive parameters*, i.e., unknown parameters specifying the form of probability distribution. Despite the fact that passive parameters directly don't carry any useful information, their knowledge improves the quality of the decision-making rule [16]. The pre-decision function under passive parameter takes the form:

$$\lambda(\mathbf{u}) = \frac{v(\mathbf{u}|C_1;\,\gamma_1)}{v(\mathbf{u}|C_2;\,\gamma_2)}\,, \tag{8}$$

where the values $\gamma_1, \gamma_2 \in \Gamma$ in a classical Bayes decision rule are exactly known, but in a more general case it may be only approximately given. However, let us assume that a decision maker is, due to his experience, convinced that a perceptible increment (decrement) $\Delta\beta$ of external parameters (circumstances) may cause the corresponding changes of the passive parameters γ_1, γ_2. He doesn't know any exact functional dependence between the increments $\Delta\beta$ and those of the passive parameters. However, for any given $\Delta\beta$ one expects that $\Delta\gamma$ may be *positive* or *negative*, its value may be *small* (S), *moderate* (M) or *large* (L) or it may be *negligible* (N). One interprets the last notions as some *fuzzy variables* characterized by their *membership functions* in the L. Zadeh sense [17], as illustrated in Fig. 3. This justifies a correction of the pre-decision function according to the formula:

$$\lambda'(\mathbf{u}) = \frac{v(\mathbf{u}|C_1;\gamma_1+\Delta\gamma_1)}{v(\mathbf{u}|C_2;\gamma_2+\Delta\gamma_2)}, \tag{9}$$

the values $\Delta\gamma_1$, $\Delta\gamma_2$ being chosen as those maximizing the membership functions $\mu(\Delta\gamma)$ of the fuzzy variables (S, L, M or N).

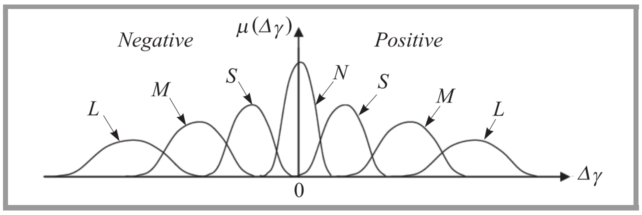


Fig. 3. Membership functions of fuzzy sets describing assumed relative external parameter's influence on passive parameter increments: *N – negligible*, *S – small*, *M – moderate*, *L – large*.

An intuitive correction of probability distribution parameters describing similarity classes can be illustrated by the following example.

Let's assume that in order to detect the most dangerous points on the roads in a town several points of road traffic monitoring have been selected. The vehicles' velocity was measured and the corresponding histograms have been calculated. Two classes of points: C_0 – non-dangerous and C_1 – dangerous have been assumed to exist and to be recognized according to the following rule: a point belongs to C_0 if no more than 50% of passing vehicles exceed the speed limits, otherwise it belongs to C_1. According to the classification rule, the medians m of the velocity histograms $h(V)$ for each monitoring point P_i, $i = 1, 2, 3, \ldots$, were calculated and the classification rule took the form:

$$P_i \in C_0 \quad \text{if} \quad m_i < V_{0i},$$
$$P_i \in C_1 \quad \text{if} \quad m_i \geq V_{0i},$$

where m_i is a calculated median and V_{0i} is a previously established speed limit for the given monitoring point i. However, several histograms of the same median may be of different form, as illustrated in Fig. 4. The histogram A

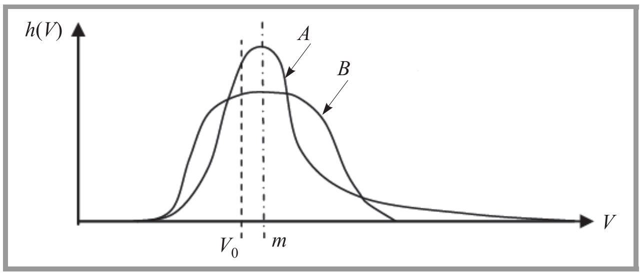


Fig. 4. Approximated histograms of car velocities in the roads; admissible speed limit = 100%.

is characterized by high asymmetry (*skewness*) and low flatness (*kurtosis*), while histogram B is rather symmetrical and more compact. An experience tells us that the type A histogram corresponds to the more dangerous situation than B because it shows that a relatively great rate of vehicles exceeding the speed limits on an extremely high level. As a consequence, a point P_i satisfying the condition $m_i < V_{0i}$ and thus classified as $P_i \in C_0$ in fact, should be reclassified as $Pi \in C_1$ if it is found that its skewness is high-positive.

4. Pattern Recognition Based on Reference Sets

The reference (learning) sets are widely used in pattern recognition in methods based on the assumption that a functional description of patterns (*similarity classes*) is not possible a priori. Reference sets can thus be defined as sets of correctly classified objects representing the similarity classes. It is assumed that the reference sets $S_1, S_2, \ldots, S_N$, where N is the number of recognized patterns, are mutually disjoint finite subsets of the observation space U unambiguously assigned to the similarity classes. Recognition of an object $\mathbf{u}$ consists in indication of the closest element to $\mathbf{u}$ in a certain set reference set S_n, $n \in [1, 2, \ldots, N]$. However, the reference sets can be collected in various ways. It is illustrated by Fig. 5 where composition of reference sets S_r, S_c is shown, representing, respectively, two classes of objects denoted symbolically by "rhombs" (⧫) and "circles" (•).

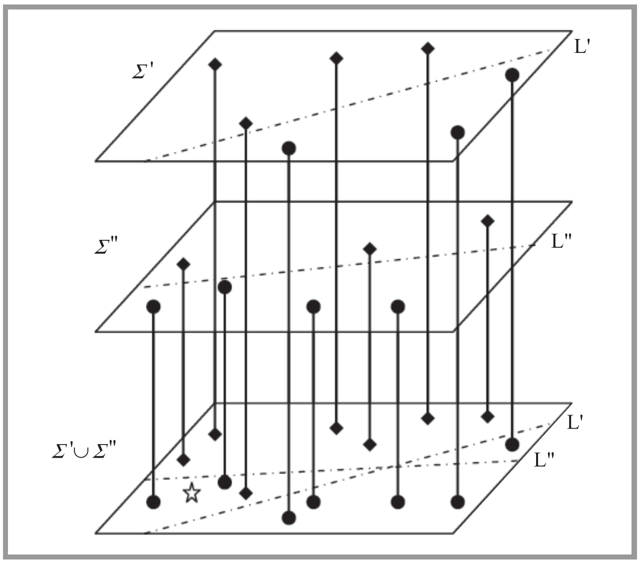


Fig. 5. Collection of reference sets as sums of subsets provided from two independent sources.

In this example it is assumed that the reference sets have been obtained by summing the corresponding subsets: $S'_r \cup S''_r$ and $S'_c \cup S''_c$ collected independently in two different external circumstances attending the observations shown on the planes Σ' and Σ''. Straight lines L' and L'' separate the subsets S_r' and S_c' collected on Σ' as well as S_r'' and S_c'' collected on Σ'. No straight line separates exactly the joined subsets $S_r' \cup S_r''$ from $S_c' \cup s_c''$ (both, L' and L'' admit two misclassifications).

Let's assume that the next object, denoted by a "star" $\star$ has been observed and should be classified. Evidently, it will be classified as a "rhomb" if L' and as a "circle" if L'', as a separating line is used. However, let us assume that it is known that external circumstances of the "star" observation were closer to those attending the Σ' than the Σ'' circumstances. In such case it is reasonable to use the L' separating line and to recognize the "star" as a "rhomb".

The above-presented situation can also be differently interpreted. Different credibility levels can be assigned to the observations collected on Σ' and Σ''. For instance, we believe more that the data on Σ' are correct than those on Σ''. The next "star" object can be recognized by using the "*k most similar objects*" (k-*MSO*) approach. In such a case a similarity measure $\sigma(\omega_i, \omega_j)$ is defined among the pairs of objects in the observation space satisfying the standard conditions [18]:

$$\left.\begin{array}{rl} \text{I.} & \sigma(\omega_i, l\,\omega_i) \equiv 1, \\ \text{II.} & \sigma(\omega_i\,\omega_j) \equiv \sigma(\omega_j, \omega_i), \\ \text{III.} & \sigma(\omega_i, \omega_j)\sigma(\omega_j, \omega_k) \leq \sigma(\omega_i, \omega_j). \end{array}\right\} \quad (10)$$

According to the k-*MSO* recognition rule, an observation ω_i is recognized as belonging to a similarity class C_n if among k most similar to ω_i objects in the reference set $\Sigma' \cup \Sigma''$ most belong to $S_n' \cup S_n''$. However, if the elements of Σ'' are less credible than those of Σ', the similarity measures between ω_i and the elements of Σ' should be taken into account with a weight reducing their influence on the recognition result. This can be done by replacing the value $\sigma(\omega_i, \omega_j)$ by $\sigma^{\beta}(\omega_i, \omega_j)$ where $\beta >$ is a reducing the similarity measure parameter.

The idea of taking a "hidden" inner structure of the reference sets into account has a particular importance in computer-aided medical diagnosis.

As an example, the case of a reference set of total cholesterol rates construction is presented. Data have been extracted from the records gathered in several dozens of analytical laboratories in Poland. The measured cholesterol levels primarily have been grouped according to the geographical regions, sexuality and age of patients. Part of the analysis results are shown in Fig. 6. The range of the recorded cholesterol rates depends on both the sexuality and the age of the patients. An averaged group of cases can be obtained as an algebraic sum of the sets corresponding to the particular cases. An interval $D* = [150 \div 225]$ mg/dcl is a common part of all particular intervals while $D* = [115 \div 285]$ mg/dcl is the minimal one covering all recorded cases. Both intervals describe a rough set (in the Z. Pawlak sense [19]) of total cholesterol levels occurring in all categories of human patients. However, for medical diagnosis a set of "normal" cholesterol levels is desirable. For this purpose, one can select from $D*$ a "core" consisting of its middle 3rd part, i.e, a sub-interval $\underline{D} = [175 \div 200]$ mg/dcl. Then a pair of intervals

$$S = \{\underline{D}, D*\} = \big\{[175 \div 200], [115 \div 285]\big\}$$

describes a rough reference set of cholesterol rates corresponding to the patients "in norm". A computer system may decide that a patient is:

"certainly in norm" if his cholesterol rate belongs to $\underline{D}$,

"possibly in norm" if it belongs to the difference $D* \ \underline{D}$,

"certainly out of norm" if it does not belong to $D*$.

However, the cases recognized as "probably in norm" can be verified if the age and sexuality is taken into account and specific rough reference sets are available instead of the averaged one. For example, let us assume that the patient is 80-year-old woman primarily as "probably in norm" diagnosed. In this case a specified rough reference set is given by a pair of intervals:

$$S_{F,80} = \{[183 \div 227], [140 \div 270]\}.$$

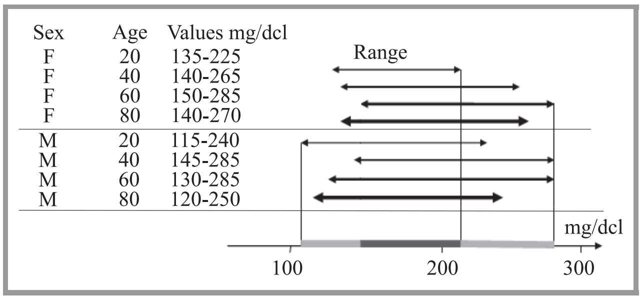


Fig. 6. Construction of a rough set of total cholesterol rate in patients of different sex (*F* – female, *M* – male) and age. The rough set indicated in the bottom-scale: black thick – surely belonging values, grey thick – possibly belonging values, black thin – surely not belonging values (data unpublished, provided by prof. Jerzy Janecki).

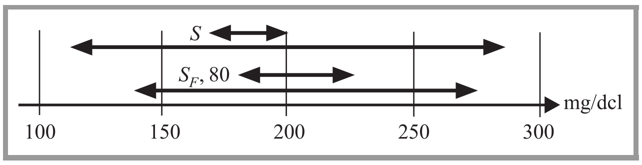


Fig. 7. Comparison of an averaged S and a specific SF, 80 rough reference sets.

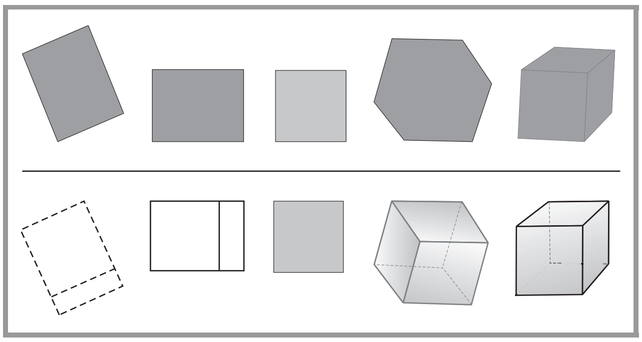


Fig. 8. Recognition of a "square" or a "cube projection" depending on the context.

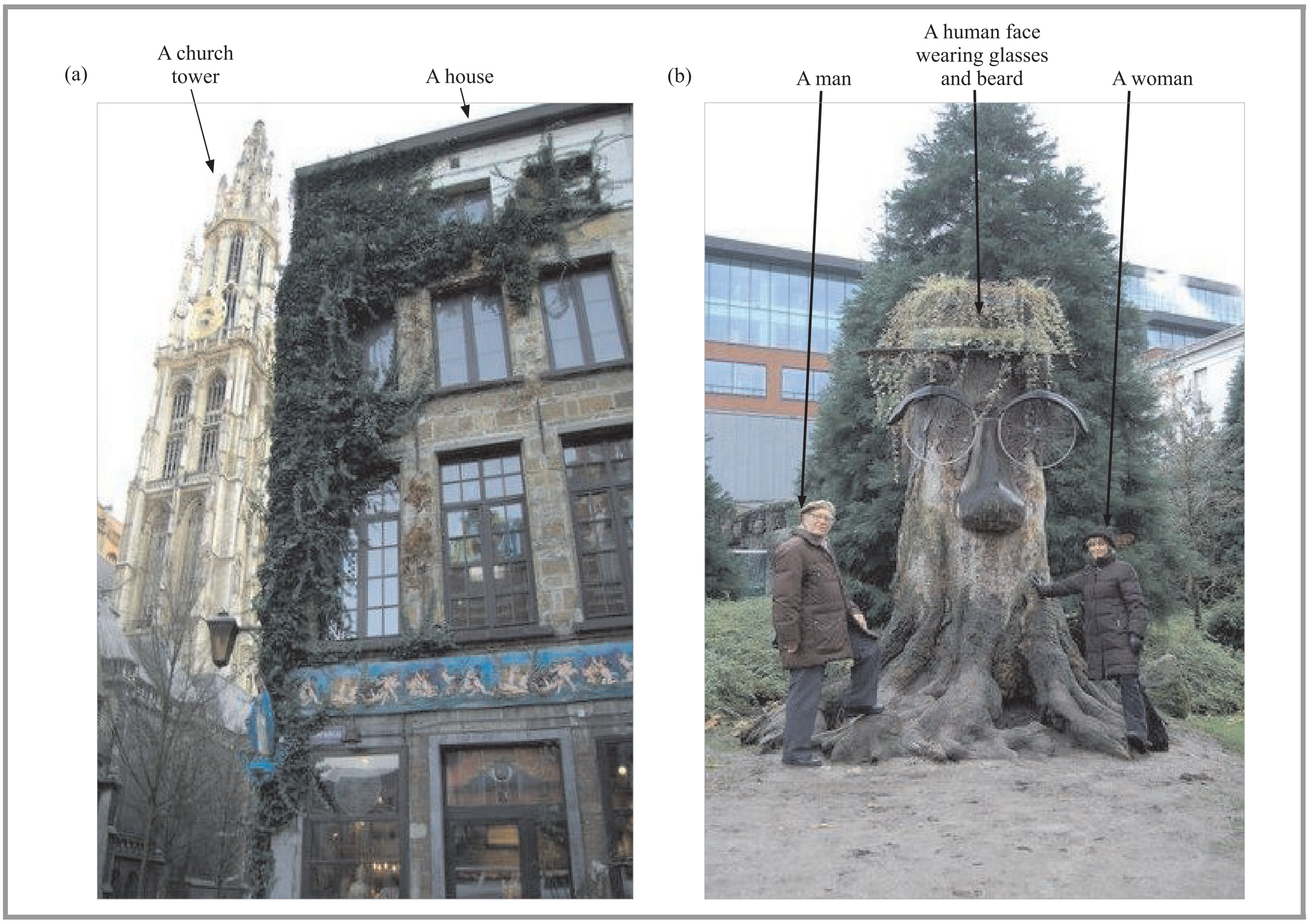


Fig. 9. Examples of scenes: (a) city view, (b) a garden view whose computer interpretation may lead to incorrect results.

In Fig. 7 the averaged and specific rough reference sets are compared. In the case of using the averaged reference set to medical diagnosis the "certainly in norm" interval is relatively small. Any based on it medical diagnosis can be corrected if the decision maker is able to use the specific reference sets instead of the averaged one, and additional information about the patient's age and sexuality is taken into consideration. It can be observed that in such a case a correction may consist of replacing:

– some "certainly in norm" recognitions by "possibly in norm",

– some "possibly in norm" recognitions by "certainly in norm" or by "certainly out of norm".

The influence of contextual information on decision-making can also be illustrated by an example of geometrical figures recognition. In the upper row of figures shown in Fig. 8, a *square* among the rectangles and the hexagons can easily be recognized. However, in the lower-row context the same figure rather as a *projection of a cube* among other cube projections will be recognized.

It follows from the above-given examples that additional information about the reference sets may substantially influence the results of pattern recognition.

5. Scene Analysis

Scene analysis is a domain of pattern recognition where contextual information plays a particularly important role. A *scene* can be defined as a collection of distinguishable objects in a physical 3D space satisfying some geometrical and/or topological relations. A correct image contents understanding and description needs without additional contextual knowledge (experience, intuition, etc.) is practically impossible.

This can be illustrated by the examples presented in Fig. 9. The image 9a can be recognized and described by an "intelligent" computer system as "A church tower and a house of equal heights in a city environment", while the image 9b as "An old man's face wearing glasses and two small human (dwarf?) figures on both its sides". Only additional information about 3D city perspective in the case of 9a and possible design of a garden in the case of 9b makes the correct interpretation of the images.

Hence, besides the basic input data, a correct decision making system should thus, besides the basic input data, take into account some additional data concerning a widely defined environment of the data source and of data acquisition circumstances, as shown in Fig. 10. The problem is how to control the process of additional data acquisition, storage and selection for a given decision process. In natural

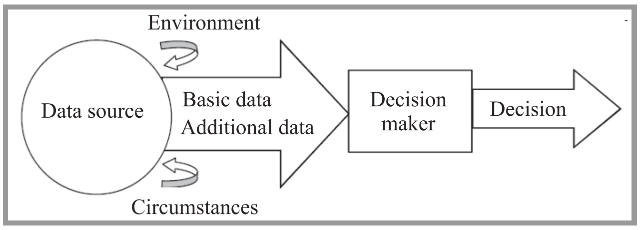


Fig. 10. Recognition of a "square" or a "cube projection" depending on the context.

thinking this is mostly a subconscious process. In artificial decision-making systems it may be hidden to the user. However, it should be a property programmed by the system designers.

6. Conclusions

The role of intuition in human behavior has been known since ages and a lot of concepts explaining the nature of intuition has been proposed in the literature. From an informational point of view intuition can be considered as a sort of additional information influencing decision making besides the basic data taken into account by a decision-maker. The numerous examples show that such additional information may be substantial in the decision making quality. In some cases a correct decision making without taking into account additional information concerning the observed objects environment or data acquisition circumstances seems impossible. On the other hand, inclusion into the decision process additional information stored by the system can be considered as a substitution of intuition in computer-aided decision system, analogous to the natural intuition influencing the human decision making. The investigation of the problem seems thus to be an open and important research problem in computer applications.

Acknowledgements

The author would like to express his deep thanks to Prof. Jerzy Janecki (Nałęcz Institute of Biocybernetics and Biomedical Engineering, PAS in Warsaw) for providing medical data which have been used in Section 3 of this paper and to Prof. Andrzej Wierzbicki for making available his publications concerning the role of intuition in human thinking.

References

[1] A. Kenny (Ed.), *The Oxford Illustrated History of Western Philosophy*. New York: The Oxford University Press, 1994.

[2] L. Wittgenstein, "On Certainty", in *On Certainty*, G. E. M. Anscombe and G. H. von Wright, Eds. Oxford: Blackwell, 1969.

[3] J. B. Watson, *Behaviorism*. New York: Norton Press, 1925.

[4] D. G. Myers, *Intuition. Its Powers and Perils*. New Haven: Yale University Press, 2002.

[5] S. Freud, *Vorlesungen zur Einführung in die Psychoanalyse* (Polish translation). Warsaw: Książka i Wiedza, 1957.

[6] R. S. Siegler and W. Stern, "A Microgenetic Analysis of Conscious and Unconscious Strategy Discoveries", *J. Exper. Psych.: General*, no. 127, pp. 377–397, 1998.

[7] A. Motycka, *Science and Unconsciousness*. Wrocław: Leopoldinum, 1998 (in Polish).

[8] M. Polanyi, *The Tacit Dimension*. London: Routledge and Kegan, 1966.

[9] L. E. I. Brouwer, "Intuitionism and Formalism", *Bull. Amer. Math. Soc.*, vol. 20, pp. 81–96, 1913.

[10] J. Hadamard, *An Essay on the Psychology of Invention in the Mathematical Field*. New York: Dover Publications Inc., 1954.

[11] A. P. Sperling, *Psychology. Made Simple Books*. London: Butterworth-Heinemann Ltd, 1982.

[12] A. P. Wierzbicki, "Knowledge Creation Theories and Rational Theory of Intuition", *Int. J. Knowl. Sys. Sci.*, no. 1, pp. 17–25, 2004.

[13] M. A. Minsky, *Framework for Representing Knowledge in the Psychology of Computer Vision*. New York: Mc Graw Hill, 1975.

[14] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Harlow: Addison Wesley, 1999.

[15] J. Kulikowski, *Cybernetyczne Systemy Rozpoznające* (*Cybernetic Recognizing Systems*). Warsaw: PWN, 1972 (in Polish).

[16] J. Seidler, *Statystyczna Teoria Odbioru Sygnałów* (*Statistical Theory of Signals Receiving*). Warsaw: PWN, 1963 (in Polish).

[17] L. A. Zadeh, "Fuzzy Sets", *Infor. and Contr.*, vol. 8, 1965.

[18] J. L. Kulikowski, "Pattern Recognition Based on Ambiguous Indications of Experts", in *Computer Recognition Systems*, *KOSYR'2001*, Kurzynski *et al.*, Eds. Wrocław: OW Politechniki Wrocławskiej, 2001, pp. 15–22.

[19] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*. Boston: Kluwer, 1991.

Juliusz L. Kulikowski received the M.Sc. degree in Electronic Engineering from the Warsaw Technical University in 1955, Cand. Sc. degree from the Moscow Higher School of Technology in 1959, D.Sc. degree from the Warsaw Technical University in 1966. Since 1966 he was a scientific worker in several Institutes of the Polish Academy of Sciences. Nominated professor in 1973. Since 1981 he is employed in the Institute of Biocybernetics and Biomedical Engineering PAS in Warsaw. He published more than 250 papers in information science, signals detection in noise, image processing methods, artificial intelligence, application of computers in medicine as well as 8 books and monographs in these domains. In 1971 he and his collaborators were rewarded by the Secretary of the Polish Academy of Sciences for construction of the first Polish computer-based image processing system CPO1/ODRA1204. He is an ordinary member of the Warsaw Scientific Society, the Editor in Chief of a scientific quarterly "Computer Graphics & Vision", a member of IFIP TC13 on 11Human-Computer Interaction", for 10 years he was a Chairman of the Polish National Committee for cooperation with the Committee of Data for Science and Technology CODATA.

E-mail: jlkulikowski@ibib.waw.pl
Nałęcz Institute of Biocybernetics
and Biomedical Engineering
Polish Academy of Sciences
Ks. Trojdena st 4
02-109 Warsaw, Poland

Paper

Traffic Analysis in the Network of a Local Voice over Internet Protocol Operator

Jakub Sobczak

Faculty of Electronics and Telecommunications, Poznań University of Technology, Poznań, Poland

Abstract—Voice over Internet Protocol market is growing rapidly. This article is the initial step towards creating a method for modeling systems for Voice over Internet Protocol operators. The purpose of this research is to gather statistics of a live VoIP-operator system, analyze them and determine if it is possible to create a specific model of such system using traditional approach. After evaluation, the data will be used, in future research, to create an analytical model of VoIP-based telecommunications system.

Keywords—VoIP, VoIP modeling.

1. Introduction

As accessing the Internet has recently become easy and affordable, people may be on-line almost everywhere – at home, at work and while commuting. Together with the increasing popularity and accessibility of the Internet, fairly huge number of telecom operators started offering a service of Voice over Internet Protocol (VoIP) calls.

According to the ITU Statistics Newslog [1] the number of VoIP subscribers have grown from just over six million in 2005 to 34.6 million at the end of 2008, which was about 24% of fixed line telephone subscribers in Europe at that time. The number of VoIP subscribers is growing and according to [2] VoIP services revenues are going to reach USD 65 billion in 2012.

What is more, market research made by IBISWorld [3], one of the most trusted independent source of industry and market research in the United States of America, points out that the VoIP industry has the greatest growth factor of all industries (including the search engine industry). The calculated growth percentage is vast 179,035.8%, even though VoIP officially came on the market in 2002. However, the most interesting figure mentioned in that research is the forecast for VoIP which predicts that this technology will still be the best-performing industry in the coming decade (2010–2019) with a growth dynamics 149.6%.

Moreover, according to [4], the VoIP market has grown (in Poland only) by almost 30% (Table 1 and Fig. 1) in 2010 after a vast decrease in growth dynamics between 2007 and 2009 – this clearly shows that the dynamic of the VoIP market has returned to an upward trend. In addition to this, the persisting financial uncertainty has raised consumers' "cash consciousness" and created an additional ground for VoIP market to grow.

It is – however – also worth mentioning, that VoIP's role has recently changed. In the individual segment, Voice over IP is usually an additional part of a larger package (e.g., TV, Internet and telephone) which is offered as a substitute for a more expensive PSTN landline. This may suggest that VoIP's main role is to keep the client and prevent him from migrating to mobile telephone systems [4]. In the business segment, the role of IP telephony is still very strong, mainly due to much higher voice traffic produced by companies.

Table 1
VoIP market in Poland – value and growth dynamics

Year	Value [M PLN]	Growth dynamics [%]
2007	403	90.3
2008	497	23.4
2009	520	4.5
2010	658	26.7

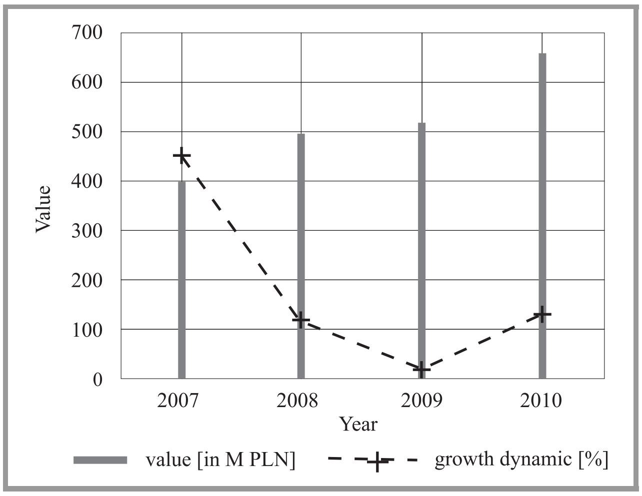


Fig. 1. VoIP market in Poland – value and growth dynamics [4].

According to the Central Statistical Office in Poland cable operators provided VoIP telephone service for about 415,000 customers in 2010, which together with customers of the biggest independent operators (Actio, Freeconet, Tlenofon), adds up to about 0.52 million Voice over Internet Protocol users in Poland. This is merely a 1% of what the whole Europe accounted for at that time – 45 million users [5].

Considering all the aforementioned information, it seems obvious that VoIP market will not stop growing and that is why the author of this article believes it is necessary to create a method of VoIP network analysis and planning. Such methods have already been created and discussed, but mostly for the traditional fixed-line networks. This article is the author's first step to create a method of system modeling for VoIP operators, which would allow them to calculate and predict the type of traffic they may expect and, basing on this calculation, to determine the required resources to serve the estimated traffic.

The purpose of this research is to evaluate and analyze traffic data collected in the network of a live VoIP operator in order to prove if it is possible to model VoIP traffic with traditional telecommunications theory. In order to do this, information about traffic type and structure has to be obtained by observation carried out on a live system. After evaluation, the data will be used in future research, to create an analytical model of VoIP-based telecommunications system.

The article is organized as follows. Section 2 describes Voice over Internet Protocol market as well as the motivation of this study. The next section presents VoIP system architecture and identifies the point of data collection in the VoIP system. The following section shows collected data whereas the results are discussed in Section 3. Finally, Section 4 concludes the article.

2. Voice over Internet Protocol

2.1. VoIP System Architecture

Architecture of a VoIP-operator system differs from a typical Public Switched Telephone Network (PSTN) architecture. First of all, it is usually less complicated and may be based on Internet Protocol (IP) only (Fig. 2).

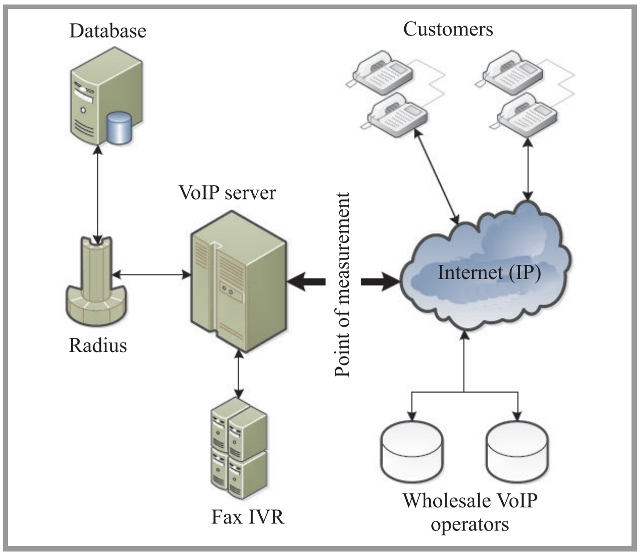


Fig. 2. VoIP system architecture.

Even though different Voice over IP operators usually have their own architectures, the main concept is the same – we can distinguish (Fig. 2) hardware responsible for: authentication and authorization (Radius server connected to the Database), signalling handling (VoIP Server) and providing additional services (IVR, fax). Apart from that, in order to provide its service, each operator has to have the possibility to terminate the outgoing traffic – in this research, the traffic termination is outsourced to other wholesale operators. This kind of VoIP architecture is easily scalable and manageable.

2.2. Point of Measurement

From a traffic engineering of such system point of view, in order to be able to create a model of such a system, it is necessary to gather certain information about the structure of traffic being carried by the system, like: traffic intensity (arrival rate), service time (service intensity), amount of requested resource (type of codec). To collect all the necessary information it is essential to choose a correct point of measurement where it is possible to analyze incoming and outgoing communication. Figure 3 shows the call flow of a typical call. The information which has to be gathered for this research is media format (VoIP codec used – e.g. G.711, G.729, T.38, etc.) and arrival times of the SIP ACK and SIP BYE packages – to determine the service time of a single call.

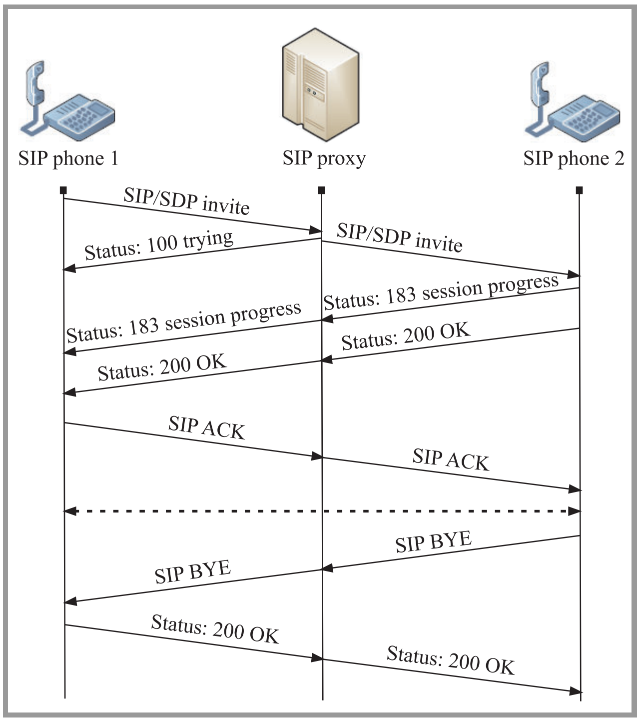


Fig. 3. Connection establishing and terminating procedure (SIP).

For the purpose of this research a live system of a Voice over Internet Protocol operator "BeFree-Mobile" was examined. Apart from providing cheap roaming solutions, BeFree-Mobile also offers VoIP service for a limited number of customers (Section 3). The data was collected on the network interfaces connecting operator's VoIP system to the Internet (Fig. 2, "Point of measurement").

3. Collected Data

The data was collected during a 30-day test period 1 June 2012 – 30 June 2012. The traffic was generated by 298 customers, among which 127 are business customers and 171 are individual customers. Results of this study are listed and analyzed in the following subsections.

In Sections 3.1 and 3.2 Average Number of Calls (ANC) and Average Service Time (AST) distributions are presented and discussed.

3.1. Average Number of Calls (ANC)

ANC parameter is the average number of calls made during a certain hour of a day throughout a 30-day test period. Average number of calls values were calculated for each hour separately. Figure 4 shows the distribution of ANC parameter averaged over all (30) days of observation. According to this chart, it may be noticed that business customers have a major influence on the system – not only does the greatest increase in the ANC parameter occur between 6 and 9 a.m. (when the workday starts), but also it drops significantly after 6 p.m. (when the workday ends). There is also – however – a noticeable increase in the ANC parameter at 8 p.m. which is probably caused by residential customers.

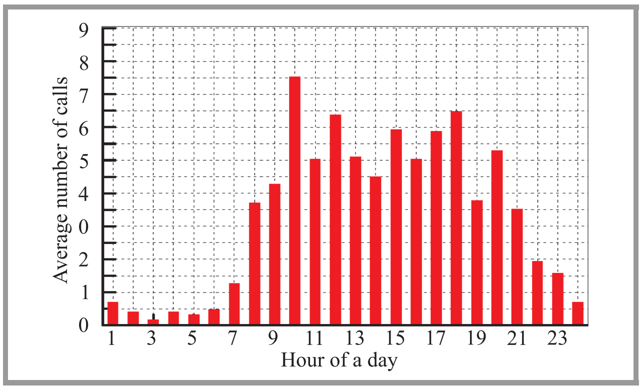


Fig. 4. Average number of calls – 30 day average.

Even though this distribution already depicts the expected outcome, the author decided to check how the distribution changes if separate days were taken into consideration. Figures 5–11 show ANC distribution throughout the whole day – separately for each day of the week. The study of the first five charts (Figures 5–9) brings the same result as in the case of Fig. 4 – the highest value of the ANC is observed around midday.

If weekdays are taken into account, the observation made at the beginning of Section 3.1 – stating that business customers have a major influence on this system – seems to be confirmed. Figures 10 and 11 show that the busy hour is shifted to the late afternoon. What is more, the intensity also seems to have dropped slightly. Also, an interesting fact may be pointed out after analyzing Fig. 11 – people tend to wake up a little later on Sundays.

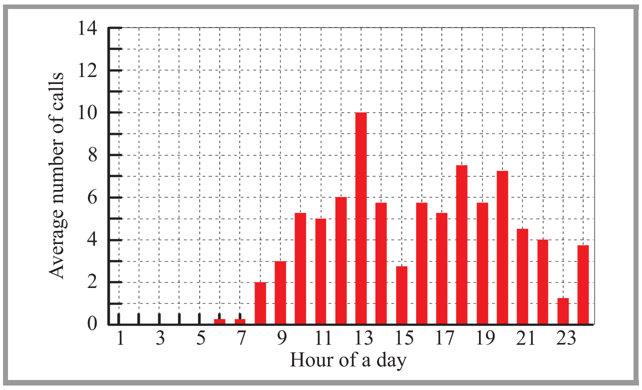


Fig. 5. Average number of calls – Monday.

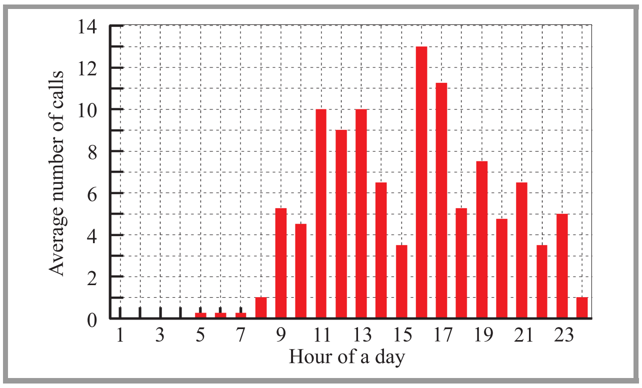


Fig. 6. Average number of calls – Tuesday.

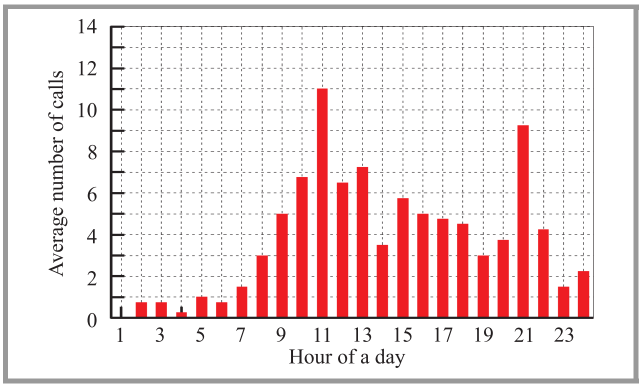


Fig. 7. Average number of calls – Wednesday.

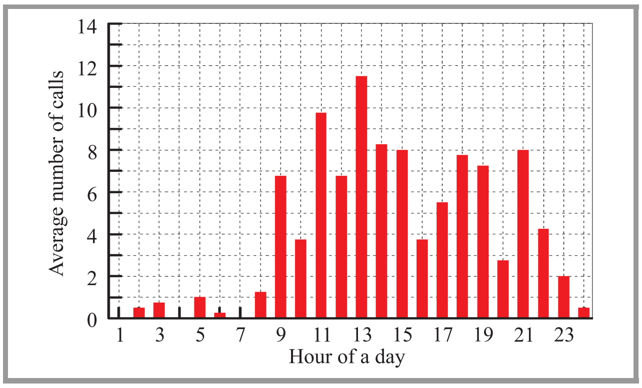


Fig. 8. Average number of calls – Thursday.

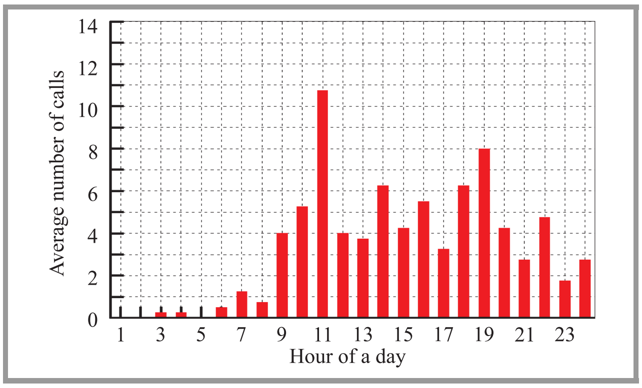


Fig. 9. Average number of calls – Friday.

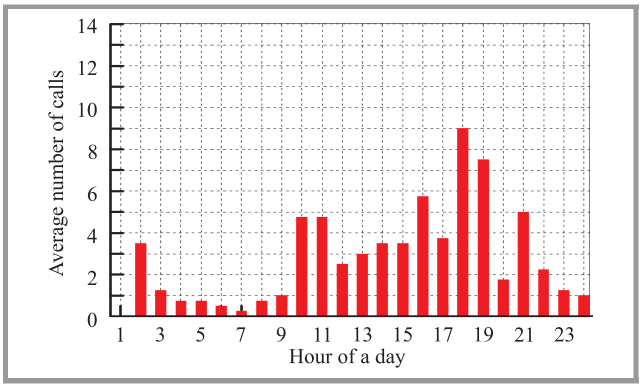


Fig. 10. Average number of calls – Saturday.

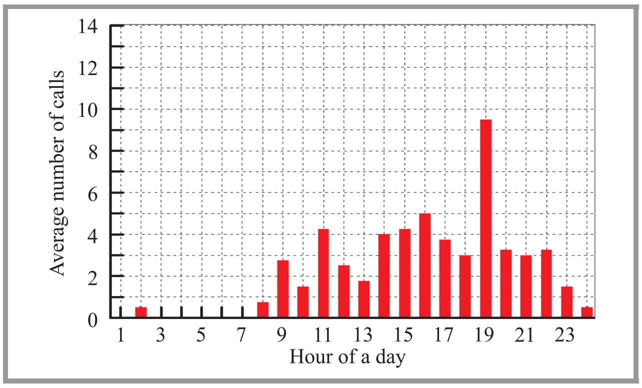


Fig. 11. Average number of calls – Sunday.

3.2. Average Service Time (AST)

AST parameter is the average length of all calls made during a certain hour of a day throughout a 30-day test period. Average Service Time values were calculated for each hour separately. Figure 12 shows the distribution of the AST parameter averaged over all (30) days of observation. The analysis of this chart shows that the longest average service times occurred before midnight – residential customers tend to make less calls, but longer ones. However, it is worth noting that this result is influenced by the fact that the ANC parameter drops significantly near midnight, which – in turn – may be caused by the data set not being big enough.

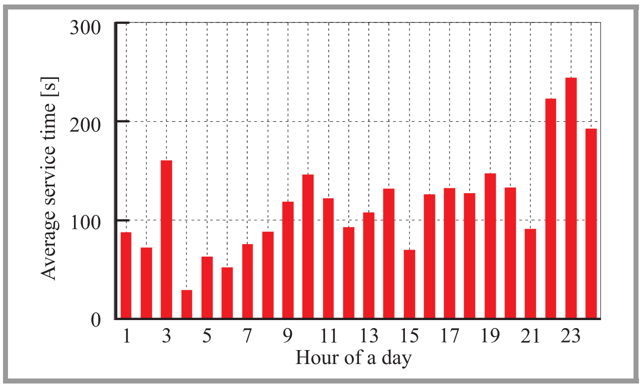


Fig. 12. Average service time – 30 day average.

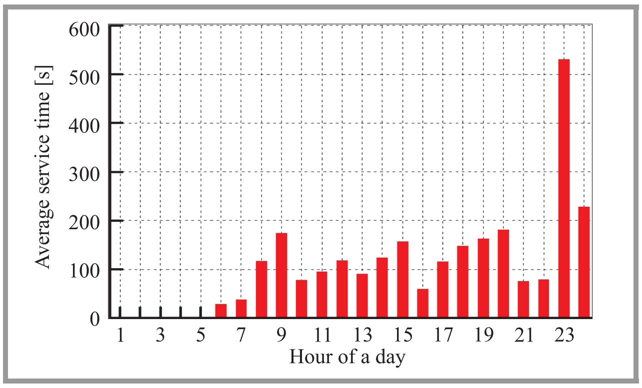


Fig. 13. Average service time – Monday.

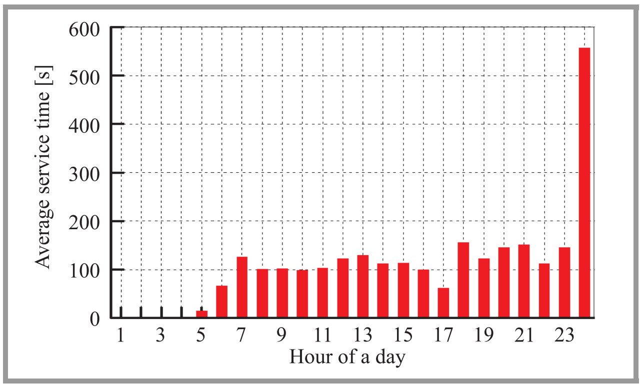


Fig. 14. Average service time – Tuesday.

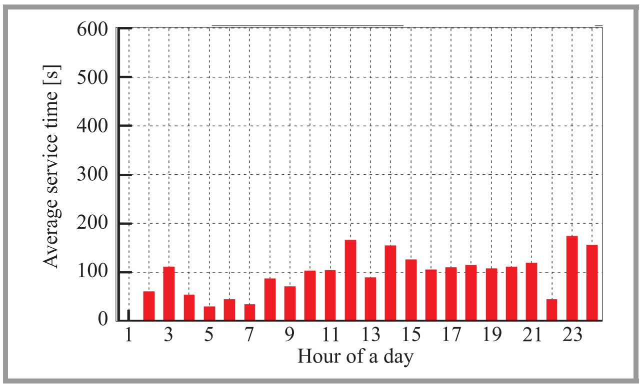


Fig. 15. Average service time – Wednesday.

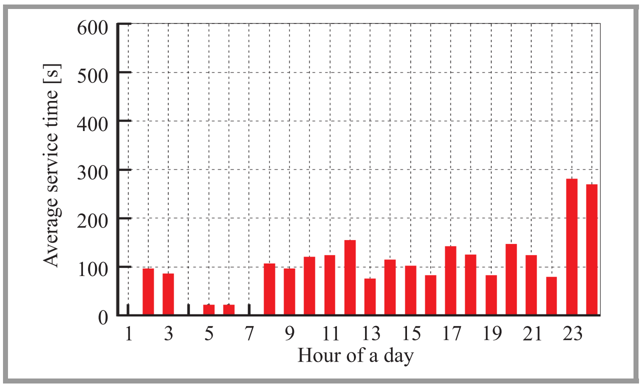


Fig. 16. Average service time – Thursday.

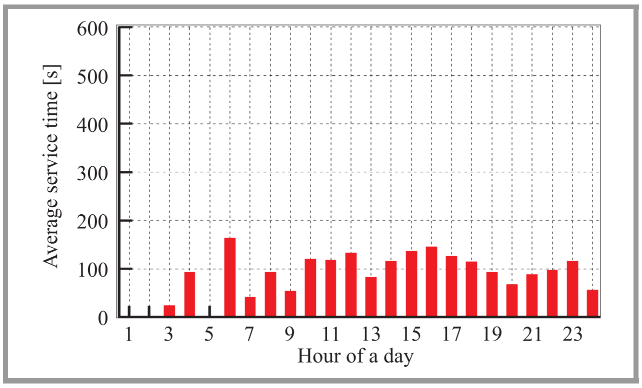


Fig. 17. Average service time – Friday.

As previously, the author decided to check how the distribution changes if separate days were taken into consideration. Figures 13–19 show AST distribution throughout the whole day – separately for each day of the week. The study of the first five charts (Figs. 13–17) shows that in majority of the weekdays there is a drop in AST parameter about 1 p.m. This might suggest that business customers might be having a lunch break at this time. Apart from that, this analysis brings the same result as in the case of Fig. 12 – the peaks in AST parameter near midnight may be caused by not enough data collected (too short observation period).

The analysis of charts depicting AST parameter during weekdays (Figs. 18 and 19) confirms previous observation that people tend to wake up a little later on Sundays. Moreover, it is also worth noting that in the case of the AST parameter, the trend is the same during workdays and weekends – the highest values occur between 9 and 12 p.m.

3.3. Conclusion

Even though a small VoIP operator's system was examined, the outcome of this research is interesting, since it has not only confirmed author's expectations but also provided important observations. Figures 20 and 21 are a summary of the all aforementioned statistics. It is clearly visible that the busy hour is shifted from 11–12 a.m. during workdays to 7–8 p.m. on weekends if we compare ANC parameter for workdays and weekends. What is more, this study also

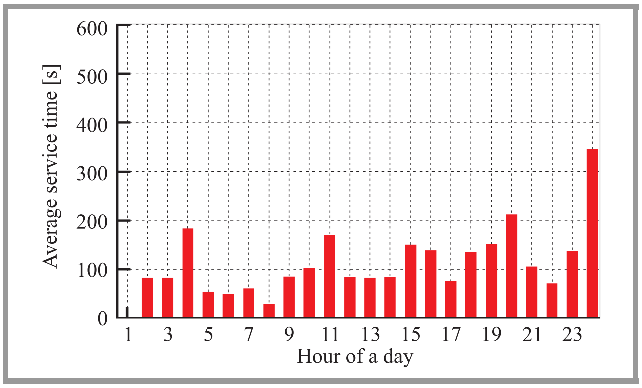


Fig. 18. Average service time – Saturday.

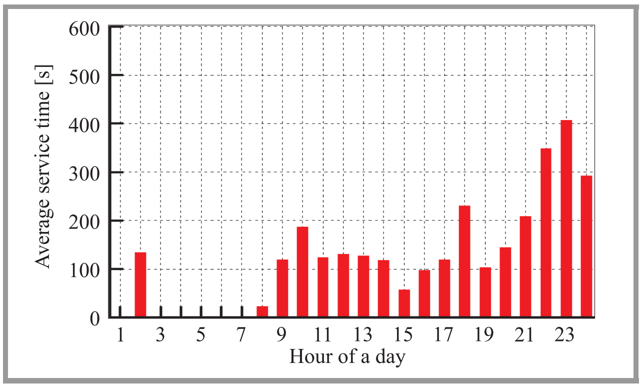


Fig. 19. Average service time – Sunday.

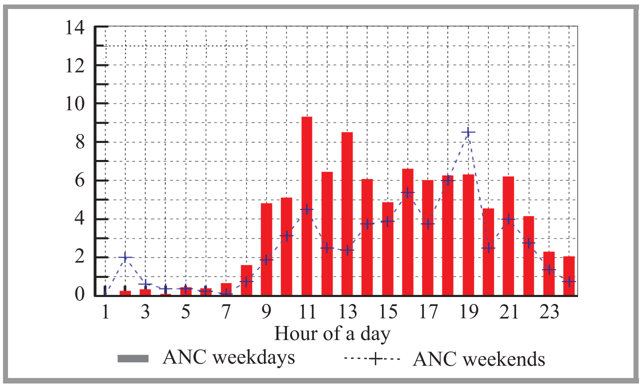


Fig. 20. Average number of calls – weekdays vs. weekends.

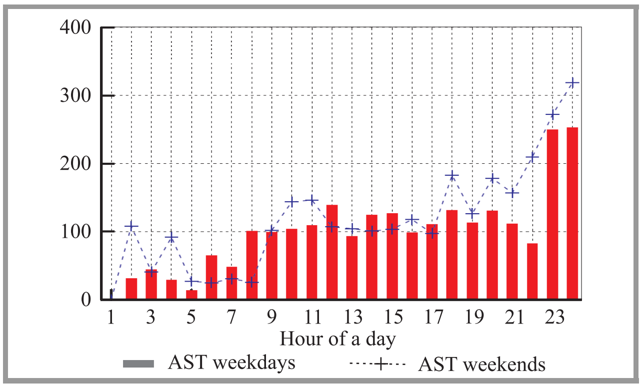


Fig. 21. Average service time – weekdays vs. weekends.

shows that AST parameter achieves the highest values between 9 and 12 p.m. during weekends as well as during workdays.

However, the observed traffic was very homogeneous as all of the customers were using G.711 codec. What is more, in such system the influence of signalling payload is too small to have to be taken into consideration, which would probably change if much bigger network was examined. The single-service nature of the monitored VoIP system clearly shows, that this system may be modeled by the well-known single rate models, based on Erlang-B formula.

4. Conclusions and Future Works

This article presents evaluation of the traffic being carried by a system of a small Voice over Internet Protocol operator.

The most interesting outcome of this research is the conclusion that in small networks with a homogeneous structure of voice streams, VoIP systems may be easily modeled by traditional single-rate traffic models.

In the next stage of this research, networks and systems of much bigger VoIP operators will be taken into consideration. It is worth noting that in the case of bigger VoIP operators due to expected diversity of traffic, it will be necessary to apply multiservice traffic models.

References

[1] ICT Statistics Newslog, *VoIP in Europe: Its bite is worse than its bark*, 2009 [Online]. Available: http://www.itu.int/ITU-D/ict/newslog/VoIP+In+Europe+Its+Bite+Is+Worse+Than+Its+Bark.aspx

[2] Visiongain, *The Voice Over Internet Protocol (VoIP) Market 2012–2017: Prospects for Skype and Other Players*, 2012 [Online]. Available: http://www.visiongain.com/Report/854/The-Voice-Over-Internet-Protocol-%28VoIP%29-Market-2012-2017-Prospects-for-Skype-and-Other-Players

[3] IBISWorld Market Research, *Voice over IP (VoIP) Industry best-performing industry of the past decade (2000–2009)*, 2009 [Online]. Available: http://www.ibisworld.com/

[4] PMR Publications, *VoIP market in Poland 2010. Development forecasts for 2011-2015*, 2010 [Online]. Available: http://www.pmrpublications.com/product/VoIP-market-Poland-2010

[5] TechWatch, *Merger creates Poland's biggest VoIP operator*, 2010 [Online]. Available: http://www.techwatch.co.uk/2010/04/08/merger-creates-polands-biggest-voip-operator/

[6] J. Sobczak and M. Głąbowski, "Top-down approach to Voice over Internet Protocol modeling", in *Proc. 51st Int. Feder. Telecommun. Engin. European Community Congr.*, Poznań, Poland, 2012.

Jakub Sobczak received the M.Sc. degree from Poznań University of Technology in 2010. He currently is a Ph.D. student at Faculty of Electronics and Telecommunications and is engaged in research in the area of multiservice networks and Voice over Internet Protocol modeling. He is also interested in aviation, ham radio operating and learning foreign languages.

E-mail: jakub.sobczak@doctorate.put.poznan.pl
Faculty of Electronics and Telecommunications
Poznań University of Technology
Polanka st 3
60-965 Poznań, Poland

Paper

Diversity of Temporal and Territorial Social Penetration Rates of Information Technology in Europe

Beata Ziewiec

Abstract—This paper presents the diversity of social penetration rates of information and communication technologies (ICT) among selected European countries according to European statistics on diverse ICT indicators. The data considered cover the 2006–2010 time range and was obtained from the Eurostat portal. The scope of the study selected EU countries – Belgium, Germany, France, Finland, Sweden, Bulgaria, Czech Republic, Poland, Slovenia, Spain. The following ICT indicators were analyzed: percentage of households or corporations with broadband access to the Internet (HHBAI), percentage of individuals who are regularly using the Internet (IRUI), percentage of individuals who ordered goods or services over the Internet (IOGSI). These indicators of ICT penetration rate in the countries examined were analyzed in terms of the following aspects: forecasting (estimates until the year 2035), maximum speed of change of these indicators (the pace of social penetration of information technology), delays or advances (in years) as compared to the averages in EU. The results are presented in tables and graphs. General conclusions and directions of future research are indicated at the end of the paper.

Keywords—ICT technology, maximum speed and delays or advances of social penetration rate of information technology, social penetration rate of information technology.

1. Introduction

There is a number of commonly accepted indicators, generally called the ICT indicators, that characterize the scale of adoption of new information technologies in societies of a new civilization – called information society, or network or knowledge civilization [1]. Indicators illustrating progress in the digitalization of the economic and social life of a country are associated with various business areas, such as the overall level of development of the country (e.g., estimated on the basis of their gross domestic product per capita, or some quality of life indicators), Internet access, Internet-based social activities in business, government, education and others.

The development of information society is a continuous and progressive process. However, it is a long term process while the speed and efficiency of this development depend not only on the level of technical infrastructure and efficiency of the telecommunications market. Both these factors generally allow access to the Internet and its various information resources and services. But the role of government policy in this area is undeniable. Government policy influences the digitalization of the economy and this in turn translates into offering new services using new ICT techniques. The development of new services contributes to the development of the economy and in the result – to the development of the information society.

The paper presents an analysis of the dynamics of the development of several selected ICT indicators – the percentage share of households with the access to the broadband Internet, the scale of purchases made by and commercial use of the Internet, and the percentage of individuals accustomed to a regular use of the Internet for diverse purposes (obtaining information, educational activities etc.).

The informational revolution – related to the fast development and slower social acceptance of ICT technologies – resulted in diverse new social phenomena, not all being positive. One of the most important negative phenomena is the so called "digital divide" or digital exclusion, an increasing social exclusion related to inability of using information technologies. This phenomenon has also temporal and territorial diversity and requires a deeper study. However, this paper does not address the problem of analysis and comparison of the digital divide in various European countries. An optimistic assumption (perhaps too optimistic) is that the level of saturation of diverse ICT indicators will eventually reach 100%. The possibility of smaller estimates of the saturation of these indicators and related analysis of digital exclusion will be the subject of further work. In the following sections of this article, the dynamics of social rate of penetration of ICT technologies is analyzed, while including in this analysis selected indicators of ICT penetration and a comparison of selected EU countries, together with an assessment of the place of Poland in such comparisons.

2. Area and Scope of Research and Analysis

The analysis was based on Eurostat data[1]. The historical data available concern the dates of 2006–2010 years. This six year period might be deemed as too short for reliable predictions, but it will be shown that the statistical significance of the obtained predictions is high. The results are presented in national terms, in order to demon-

[1] http://epp.eurostat.ec.europa.eu/

strate the diversity of rate of absorption (penetration) of selected social indicators of ICT in selected European Union countries.

As representatives of the Nordic countries, Sweden and Finland were selected. The core of EU is represented by France, Belgium and Germany. These two groups of countries are also counted as a group of the economically most developed EU countries. Less developed are: a group of post-communist countries of Middle-East Europe, with selected representatives: Poland, Slovenia, Czech Republic. Bulgaria was selected as the representative of Southern Europe group of former communist countries. In addition, Spain was selected to represent Southern Europe and the Iberian Peninsula.

The rate of penetration of ICT is examined in terms of the following aspects:

- Forecasting (with estimations of data until 2025).
- Analysis of the maximum speed the rate of social penetration of information technology (counted for estimated data, in order to smooth out statistical divergences).
- Analysis of delays or advances as compared to the averages of European Union.

In summary, a discussion of the place of Poland in terms of the rate of absorption of new information technologies in comparison to selected countries and to the average value calculated for the 27 EU countries.

The following indicators of ICT penetration were analyzed:

– households with broadband access to the Internet (HHBAI),

– individuals who regularly are using the Internet (IRUI),

– individuals who ordered goods or services over the Internet (IOGSI).

There are diverse models that can be used to estimate the development characterized by temporal data, see the analysis of different models in [2], [3]. The classical logistic function method was selected and thus the data were estimated by the formula:

$$v_2 = a/\big(1+b\exp(-cv_1)\big)$$

with v_1 representing the time (in years) and v_2 – a selected ICT indicator, coefficients b and c determined by using the software package "Statistica 8". After estimation, it is possible to compute the maximal speed of change:

$$V_{\max} = ac/4\,.$$

The source data covers the period 2003–2010, the estimations were computed for the period 1991–2025. The coefficient a was optimistically assumed $a = 100\%$, while it is admitted that this assumption requires further detailed analysis, particularly when addressing the problem of digital divide, cf. [3].

3. Estimation of Development Curves

3.1. HHBAI Indicator

Raw data were drawn from the database Eurostat – presenting the percentage of households with broadband access to the Internet, called HHBAI indicator. These data are available for years 2006–2010 (Table 1).

Table 1
Historical data by Eurostat for the HHBAI indicator, [%]

Country/Year	2006	2007	2008	2009	2010
Belgium	0.0	0.0	59.6	62.9	69.3
Bulgaria	9.1	14.0	19.5	25.1	24.9
Czech Republic	0.0	27.8	36.1	49.0	53.6
Finland	52.4	61.9	65.0	72.4	75.4
France	0.0	0.0	53.8	54.1	64.9
Germany	38.0	53.6	51.7	61.5	73.3
Poland	0.0	29.2	38.0	51.2	57.0
Slovenia	0.0	0.0	50.0	56.0	62.0
Spain	29.2	38.5	43.7	49.9	56.8
Sweden	0.0	0.0	70.0	78.3	81.3
Av. EU 27	33.0	41.4	46.8	54.1	58.4

As a result of estimation by the logistic function Eq. (1), the estimated data presented in Table 2 were obtained. The estimation period starts with 1991, in order to illustrate the beginnings of development for which there are no historical data, and ends with forecasted data for 2025. In the analysis shown by next figures we see that the estimated data revolve closely around or even coincide with historical data. This confirms the preliminary estimation accuracy. Moreover, the estimated results have a 95% confidence level (the "Statistica" program has a built-in mechanism to analyze the confidence levels). It should be also noted that the phenomenon studied concerns a long term development, slow but showing clear trends. Of course, a different model, cf. [3], possibly with an independent evaluation of the coefficient a, might give different results, but except for the problem of digital divide the estimations presented in Table 2 are significant.

Estimations of data for the HHBAI indicator, concerning households with broadband Internet access, show significant differences in the rate of penetration between countries. Figure 1 presents the graphs of the logistic function for HHBAI in the countries studied.

We see that while Belgium starts first, it has a slow development, whereas Sweden starts from lower levels but much faster and becomes the best; Finland is slightly slower. Definitely the worst results, according to the logistic function prediction of the HHBAI rate of penetration, were shown for Bulgaria in the post-communist region of Southern Europe. The strongest penetration according to the HHBAI indicator, as shown by graphs, have Scandinavian and core EU countries. The penetration in Poland, compared with

Table 2
Estimated data for the HHBAI indicator, [%]

Year/ Country	Belgium	Bulgaria	Czech Republic	Finland	France	Germany	Poland	Slovenia	Spain	Sweden	Av. EU 27
1991	4	0	0	2	2	1	0	2	1	1	1
1992	5	0	0	3	3	1	0	2	1	1	1
1993	6	0	0	4	4	1	0	2	1	2	2
1994	7	0	0	5	4	1	0	3	2	2	2
1995	9	1	0	6	5	2	0	4	2	3	3
1996	11	1	1	8	7	3	0	5	3	5	4
1997	13	1	1	10	8	3	1	6	4	6	5
1998	15	1	1	13	10	5	1	8	5	9	6
1999	18	2	2	16	12	6	2	10	6	11	8
2000	21	2	3	20	15	9	2	12	8	15	10
2001	25	3	4	24	18	12	4	15	10	20	12
2002	29	4	6	29	22	15	5	19	13	26	16
2003	34	5	8	35	26	20	8	23	16	32	19
2004	39	7	11	41	30	26	11	27	20	40	24
2005	44	9	16	47	35	32	16	32	25	48	29
2006	49	11	21	54	41	40	22	38	31	56	34
2007	54	14	28	60	46	48	30	44	37	63	40
2008	59	18	37	66	52	56	39	50	43	71	47
2009	64	22	46	71	58	64	49	56	50	77	53
2010	69	27	55	76	63	71	59	62	57	82	59
2011	73	33	64	81	68	77	68	68	64	86	66
2012	77	39	73	84	73	82	76	73	70	90	71
2013	80	46	79	87	77	86	83	77	75	92	76
2014	83	53	85	90	81	90	88	81	80	94	81
2015	86	59	89	92	84	92	91	85	84	96	84
2016	88	66	92	94	87	94	94	88	87	97	87
2017	90	71	95	95	89	96	96	90	90	98	90
2018	92	77	96	96	91	97	97	92	92	98	92
2019	93	81	97	97	93	98	98	94	94	99	94
2020	95	85	98	98	94	98	99	95	95	99	95
2021	96	88	99	98	95	99	99	96	96	99	96
2022	96	91	99	99	96	99	99	97	97	100	97
2023	97	93	99	99	97	99	100	98	98	100	98
2024	98	94	100	99	98	100	100	98	98	100	98
2025	98	96	100	99	98	100	100	98	99	100	99

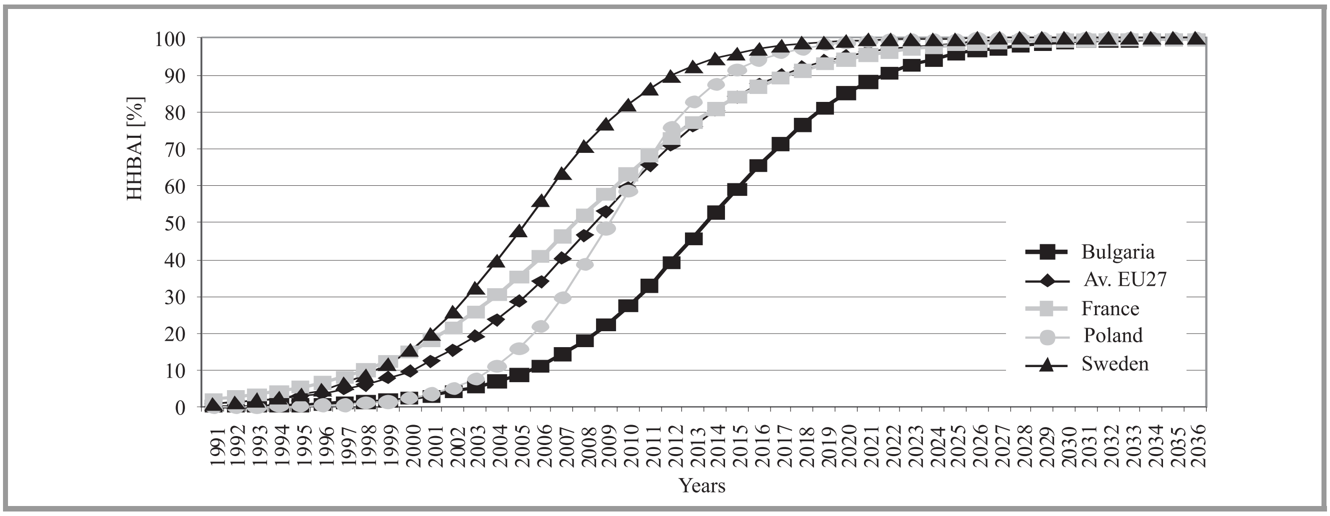


Fig. 1. Estimated data for the HHBAI indicator.

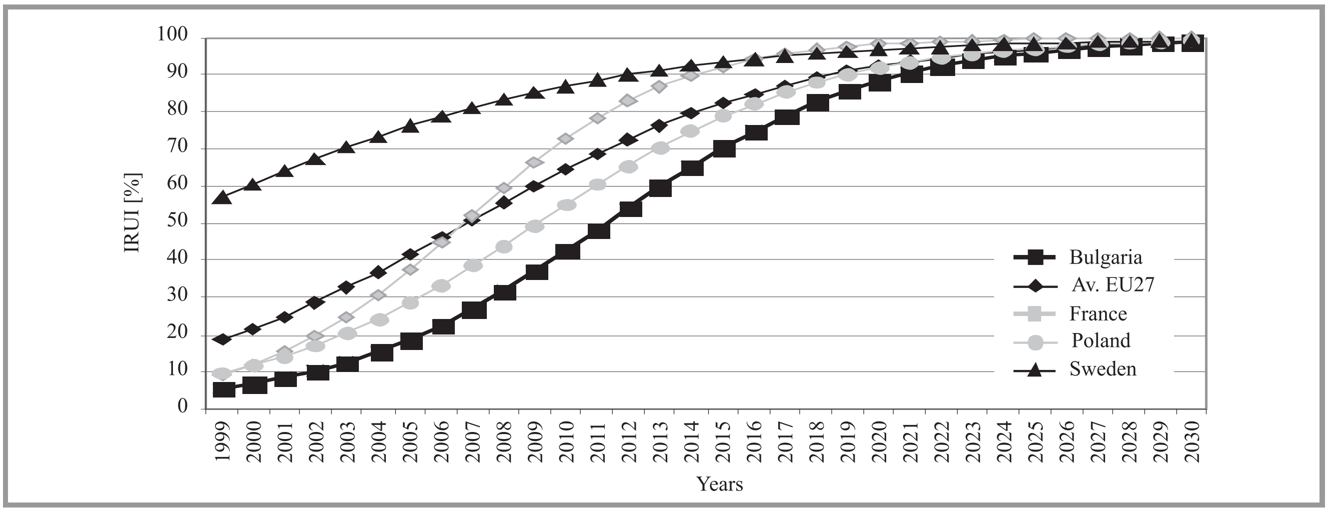


Fig. 2. Estimated data for the indicator IRUI.

Table 3
Historical data by Eurostat for the IRUI indicator, [%]

Country/Year	2003	2004	2005	2006	2007	2008	2009	2010	2011
Belgium	0.0	0.0	53.0	58.0	63.0	66.0	70.0	75.0	78.0
Bulgaria	0.0	13.0	0.0	22.0	28.0	33.0	40.0	42.0	46.0
Czech Republic	20.0	25.0	26.0	36.0	42.0	51.0	54.0	58.0	63.0
Finland	58.0	63.0	62.0	71.0	75.0	78.0	79.0	83.0	86.0
France	0.0	0.0	0.0	39.0	57.0	63.0	65.0	75.0	74.0
Germany	44.0	50.0	54.0	59.0	64.0	68.0	71.0	75.0	77.0
Poland	0.0	22.0	29.0	34.0	39.0	44.0	52.0	55.0	58.0
Slovenia	0.0	33.0	40.0	47.0	49.0	52.0	58.0	65.0	64.0
Spain	29.0	31.0	35.0	39.0	44.0	49.0	54.0	58.0	62.0
Sweden	69.0	75.0	76.0	80.0	75.0	83.0	86.0	88.0	91.0
Av. EU 27	0.0	36.0	43.0	45.0	51.0	56.0	60.0	65.0	68.0

the average measured for the 27 EU countries, occurs in two phases. In the first phase Poland is below the European average, in 2010–2013 it catches up to the average and strongly accelerates to achieve predicted results above the EU-27 average in the second phase of development.

3.2. IRUI Indicator

Primary data for the IRUI indicator were obtained from the Eurostat base and they represent the percentage of all users between the ages of 16–74 years, who have access to the Internet and regularly use it (assuming average – at least 1 time per week; all access methods and every possibility of using from the network, e.g., Internet cafes, were taken into consideration). These data, available for the years 2003–2011 (Table 3), were estimated by the logistic function Eq. (1) for the period 1991–2025 year (even until 2035, but in Table 4 are shown only data until 2025). The estimated data were again (similarly as in the case of HHBAI indicator) very closely oscillating around, or even coinciding with historical data, and the confidence level was over 95%. Thus except for the issue of digital divide, the data in Table 4 are highly significant.

Estimations of data for the IRUI indicator, concerning a regular use of the Internet, show significant differences between the social rate of penetration of information technologies in European Union. The graphs of the logistic function for IRUI and for the countries examined are presented in Fig. 2.

We see that Sweden and Finland are the best in this indicator, although France develops very fast and might overtake them after 2018. Again, the least developed between the examined countries and with a very slow rate of development is Bulgaria. Poland is not much better, with a similar curve to Bulgaria, only shifted in time ahead by two to three years, but Poland is similarly delayed to the European Union average, as Bulgaria is to Poland.

3.3. IOGSI Indicator

Raw data were obtained from Eurostat database for IOGSI indicator that shows the percentage of users purchasing goods and services over the Internet. These data were available for years 2002–2010 (Table 5).

Above data were estimated by the logistic function for the period 1991–2025 (again, even to 2035, but only data until 2025 are shown in Table 6). The estimated data are

Table 4
Estimated data for the HHBAI indicator, [%]

Year/ Country	Belgium	Bulgaria	Czech Republic	Finland	France	Germany	Poland	Romania	Slovenia	Spain	Sweden	Av. EU 27
1991	7	1	1	16	1	9	2	1	5	5	29	5
1992	9	1	2	18	1	11	2	1	6	6	32	6
1993	11	1	2	21	2	13	3	1	7	7	36	7
1994	12	2	3	24	2	15	3	2	8	8	39	8
1995	15	2	3	27	3	17	4	2	9	9	43	10
1996	17	3	4	31	4	20	5	3	11	11	46	11
1997	20	3	6	34	5	23	6	3	13	13	50	14
1998	23	4	7	38	7	26	8	4	15	15	54	16
1999	27	5	9	42	9	30	10	5	18	17	57	19
2000	31	7	11	46	12	34	12	6	21	20	61	22
2001	35	8	14	51	15	38	14	7	24	22	64	25
2002	39	10	17	55	20	42	17	8	27	26	67	29
2003	44	13	20	59	25	46	20	10	31	29	71	33
2004	48	16	25	63	31	50	24	12	35	33	73	37
2005	53	19	30	67	38	55	29	15	40	36	76	41
2006	58	23	35	71	45	59	33	18	44	40	79	46
2007	62	27	41	74	52	63	39	21	49	45	81	51
2008	67	32	47	77	60	67	44	25	53	49	83	55
2009	71	37	53	80	66	71	49	29	58	53	85	60
2010	74	43	59	82	73	74	55	34	62	57	87	65
2011	78	48	65	85	78	78	60	39	67	62	89	69
2012	81	54	71	87	83	81	66	44	71	65	90	73
2013	84	60	75	89	87	83	70	49	74	69	91	76
2014	86	65	80	90	90	86	75	55	78	73	92	79
2015	88	70	83	92	92	88	79	60	81	76	93	82
2016	90	75	87	93	94	89	82	65	83	79	94	85
2017	92	79	89	94	96	91	85	70	86	82	95	87
2018	93	83	91	95	97	92	88	74	88	84	96	89
2019	94	86	93	96	98	93	90	78	90	86	96	91
2020	95	88	95	96	98	94	92	81	91	88	97	92
2021	96	91	96	97	99	95	93	84	93	90	97	94
2022	97	92	97	97	99	96	95	87	94	91	97	95
2023	97	94	97	98	99	97	96	89	95	93	98	95
2024	98	95	98	98	99	97	97	91	96	94	98	96
2025	98	96	98	98	100	98	97	93	96	95	98	97

Table 5
Historical data by Eurostat for the IOGSI indicator, [%]

Country/Year	2002	2003	2004	2005	2006	2007	2008	2009	2010
Belgium	0.0	0.0	0.0	11.0	14.0	15.0	14.0	25.0	27.0
Bulgaria	0.0	0.0	1.0	0.0	2.0	2.0	2.0	3.0	3.0
Czech Republic	0.0	3.0	3.0	3.0	7.0	8.0	13.0	12.0	15.0
Finland	11.0	14.0	24.0	25.0	29.0	33.0	33.0	37.0	41.0
France	0.0	0.0	0.0	0.0	19.0	26.0	28.0	32.0	42.0
Germany	17.0	24.0	29.0	32.0	38.0	41.0	42.0	45.0	48.0
Slovenia	0.0	0.0	4.0	8.0	8.0	9.0	12.0	14.0	17.0
Spain	2.0	5.0	5.0	8.0	10.0	13.0	13.0	16.0	17.0
Sweden	24.0	21.0	30.0	36.0	39.0	39.0	38.0	45.0	50.0
Av. EU 27	0.0	0.0	15.0	18.0	20.0	23.0	24.0	28.0	31.0

Table 6
Estimated data for the IOGSI indicator, [%]

Year/ Country	Belgium	Bulgaria	Czech Republic	Finland	France	Germany	Poland	Slovenia	Spain	Sweden	Av. EU 27
1991	0	0	0	2	1	4	0	0	0	6	2
1992	1	0	0	3	1	5	0	0	1	7	2
1993	1	0	0	3	1	6	0	1	1	8	2
1994	1	0	0	4	1	7	0	1	1	9	2
1995	1	0	0	5	1	8	0	1	1	10	3
1996	1	0	1	6	2	9	0	1	1	12	3
1997	2	0	1	7	2	11	1	1	2	13	4
1998	2	0	1	8	3	13	1	1	2	15	5
1999	3	0	1	9	4	15	1	2	2	17	5
2000	3	0	1	11	5	17	1	2	3	19	6
2001	4	1	2	13	6	19	2	3	3	21	7
2002	5	1	2	15	8	22	3	3	4	24	9
2003	7	1	3	18	10	25	3	4	5	26	10
2004	8	1	4	20	13	28	4	5	6	29	12
2005	10	2	5	23	16	31	6	6	7	32	13
2006	12	2	6	27	19	35	8	8	9	35	15
2007	15	3	8	30	24	38	10	10	11	39	18
2008	19	4	10	34	29	42	13	12	13	42	20
2009	22	6	13	38	34	46	17	14	16	46	23
2010	27	7	16	42	40	50	21	17	18	49	26
2011	32	10	20	47	47	54	26	20	22	53	29
2012	37	13	24	51	53	58	32	24	25	56	33
2013	43	17	29	56	59	62	39	28	29	60	36
2014	48	21	35	60	65	66	46	33	34	63	40
2015	54	27	41	64	71	69	53	38	39	66	44
2016	60	34	47	68	76	73	60	43	44	70	48
2017	66	41	54	72	80	76	67	48	49	72	52
2018	71	48	60	75	84	79	73	54	54	75	56
2019	75	56	66	78	87	81	78	59	59	78	60
2020	79	63	71	81	90	84	83	64	64	80	64
2021	83	70	76	84	92	86	86	69	68	82	68
2022	86	76	81	86	94	88	90	73	73	84	71
2023	89	81	84	88	95	89	92	78	76	86	75
2024	91	85	88	90	96	91	94	81	80	88	78
2025	93	89	90	91	97	92	95	84	83	89	80

again (similarly as for HHBAI and IRUI) very close to historical data and with high confidence level. This fact confirms the estimation accuracy again, except for the issue of digital divide, see Table 6.

Estimations of data for the IOGSI indicator that concern purchases of goods and services made by the Internet, show significant differences in the rate of social penetration of this indicator between examined countries. Figure 3 shows the graphs of the logistic function for the examined countries.

There is a large disparity between EU countries according to IOGSI indicator, even greater than for HHBAI and IRUI indicators. Sweden and Germany are the best, but France might soon overtake them. Bulgaria is again on the weakest position among the tested ten EU countries. However, the development of this index in Bulgaria is fast and it appears that after the year 2020, the percentage of users purchasing goods and services over the Internet in Bulgaria might reach the average level of EU 27, and then – might rise above the average, overtaking even Slovenia and Spain.

Second to last place in the IOGSI index, just before Bulgaria, would be the Czech Republic. However, the Czech Republic accelerates its IOGSI development in recent years, has fast growth of the IOGSI logistic curve, which might result in overtaking the European Union average around 2017. Not much better results than for the Czech Republic can be recorded for Spain – slightly higher today, but much slower in development, thus overtaking the EU average around 2021.

Sweden played the leading role in IOGSI index until 2008. Starting with 2009, Germany is gradually beginning to overtake the Scandinavian countries in the intensity of their purchases of goods and services over the Internet. However, the fastest development in the Internet commerce has

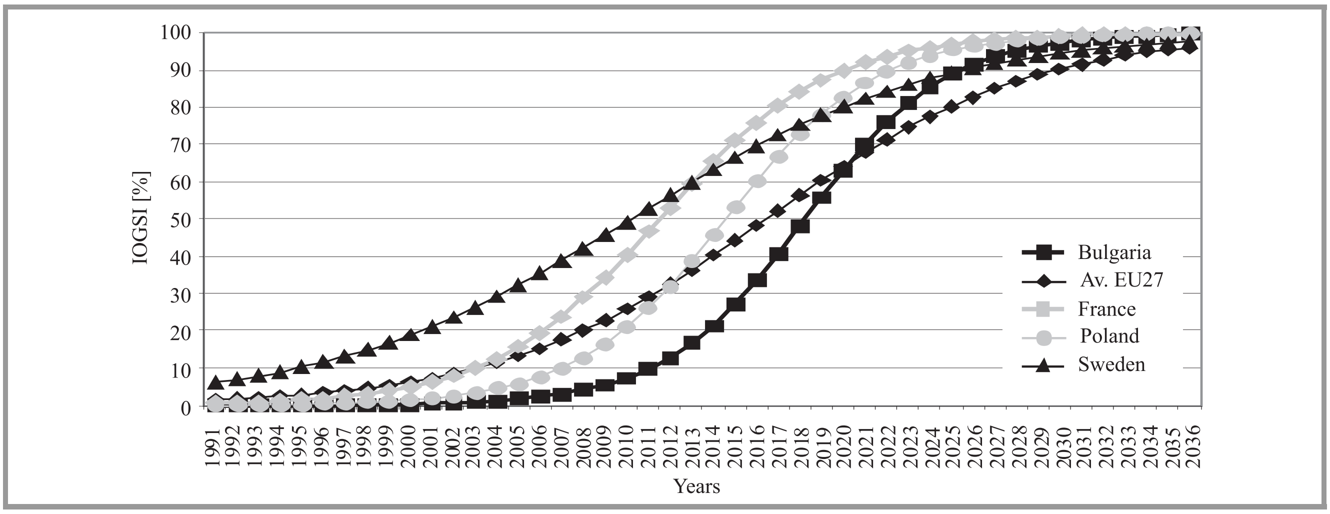


Fig. 3. Estimated data for the indicator IOGSI.

France that might overtake Germany around 2014–2015, to become the undisputed leader in IOGSI indicator.

Poland has also very fast, similar to France, development of IOGSI indicator, although starting at a lower level. Poland might achieve the average EU 27 already currently, in 2012, and in 2015 even overtake a core EU country – Belgium. Even more surprising is the fact that Poland might overtake in the year 2019 Sweden, an undisputed leader in other indicators. Thus, from 2020, according to estimation by the logistic function, leading countries in Europe in the number of users purchasing goods and services over the Internet, might be France – representing the core EU countries and right behind her Poland representing former communist countries of Central and Eastern Europe.

4. The Maximum Speed of Social Penetration of Information Technology

The formula (2) was used to determine the maximal speed of change (smoothed out of statistical perturbations) of the three indicators (HHBAI, IRUI, IOGSI) of social ICT penetration. Since we assume $a = 100\%$, the estimated parameter c determines this speed that is counted in % per year. In [3] it was observed that, for processes of social penetration of new technologies, this speed is strongly limited and rarely exceeds 10% per year. While using the results of estimations from previous section, the following Table 7 is easily computed.

We observe that the maximum speeds, even if they confirm the general conclusion of [3], are very diversified. For HHBAI indicator, fastest development is observed in Poland and Czech Republic; for IRUI, in France and Czech Republic, while the absolute values of the speeds are lower; for IOGSI, fastest development is observed in Bulgaria and Poland. This is illustrated in Figs. 4–6.

Table 7
Maximum speeds of development for indicators, [%]

Country	HHBAI	IRUI	IOGSI
Belgium	5.2	4.7	5.9
Bulgaria	6.8	5.8	7.7
Czech Republic	9.5	6.2	6.5
Finland	6.4	4.2	4.4
France	5.7	7.5	6.4
Germany	8.1	4.4	4.1
Poland	10.1	5.6	7.2
Slovenia	6.1	4.6	5.4
Spain	6.9	4.3	5.1
Sweden	8.1	3.7	3.6
Av. EU 27	6.4	4.7	5.2

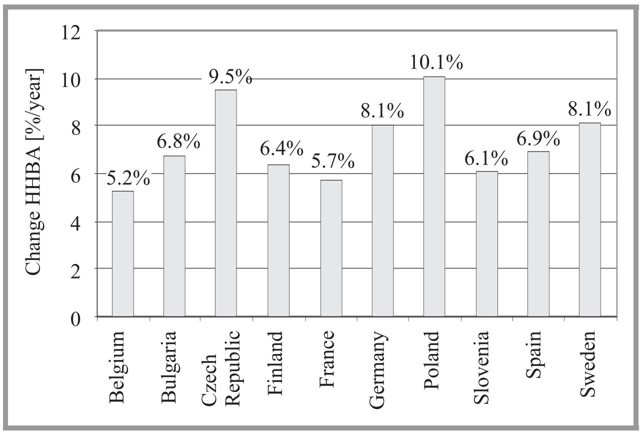


Fig. 4. Maximum annual rate of change for the HHBAI indicator.

Thus the maximal speed of growth of HHBAI indicator is in the range 10.1% per year (Poland) – 5.2% per year (Belgium). The same range for IRUI indicator is 7.5% per year (France) – 3.7% per year (Sweden); for IOGSI indicator 7.7% per year (Bulgaria) – 3.6% per year (Sweden).

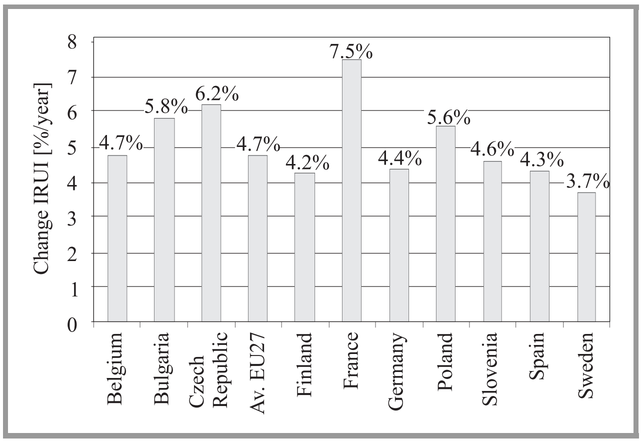


Fig. 5. Maximum annual rate of change for the IRUI indicator.

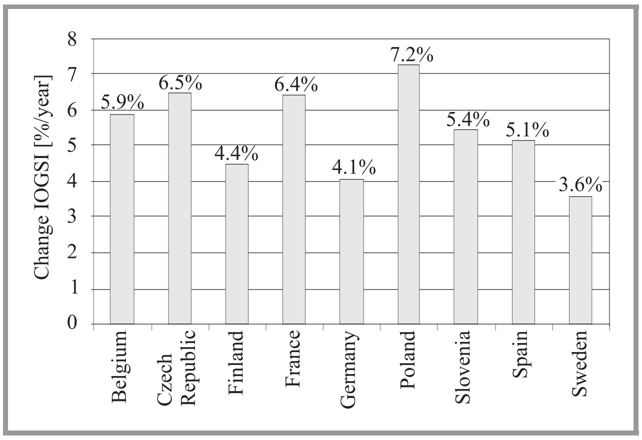


Fig. 6. Maximum annual rate of change for IOGSI indicator.

It can be concluded that even if the range of statistical data is rather limited (six to seven years), the forecasts using logistic curves can give quite interesting information with high confidence. Statistics on the development of ICT indicators were apparently not carried out before 2002 (or even before 2006 for HHBAI indicator), and we can judge upon earlier developments only by backward "forecast" – estimated logistic values for earlier times. For example, the estimation of the HHBAI for the year 2001, see Table 2, indicate, that Belgium and France had the largest percentages of households with broadband internet access – and considerable above the average EU27, while Poland and Bulgaria were well below average.

All the above analysis indicates that while the starting situation was very diversified, countries such as Poland or even Bulgaria do catch up, with smaller or larger delays, which will be analyzed in the next section.

5. Advances and Delays in Comparison to the Current EU Average

To determine the delay or advance of an EU country, we can use diverse approaches, e.g., compare the estimated time of reaching a given threshold. However, the simplest approach is to assume that the threshold is the current average calculated for the 27 EU countries, see [2], [3].

The data for selected European Union countries have been thus analyzed in terms of advances or delays in social penetration of ITC technologies, as compared to the average of the 27 EU countries, in terms of HHBAI, IRUI and IOGSI indicators. The resulting advances or delays for HHBAI indicator are presented in Table 8. The most interesting is the current result – for the year 2011 – when Sweden has 4 year advance over av. EU27, while Bulgaria has 6 year delay; Poland, due to recent fast development, has caught up with European average, better than Spain which has 1 year delay. Forecast for 2015 gives Poland 2 years advance over European average, while the advance of Sweden grows to 5 years, Spain maintains 1 year delay and Bulgaria reduces it's delay slightly to 5.5 years. Further forecasts – to 2020 – might be less reliable, but show an increasing advance of Poland and delay for such core European Union country as Belgium. These results are illustrated in Fig. 7.

For IRUI indicator, the resulting delays or advances of examined countries as compared to the EU27 average are shown in Table 9. We see that currently (2011) Sweden has the largest advance of 6 years, Poland and Spain have delays of 2.5 years, Bulgaria a delay of 5.5 years. Predicted for 2015 is a slight reduction of the delay of Poland to 2 years, while Spain maintains delay of 2.5 years. The situation will not change qualitatively until 2020, with a slight reduction of delay for Poland, a somewhat stronger reduction of delay for Bulgaria. It can be seen that IRUI indicator characterizes a weak point of Poland.

These results are illustrated in Fig. 8. It can be seen that France has the fastest development of IRUI indicator and will advance over EU27 average to over 10 years.

For IOGSI indicator, the resulting delays or advances of examined countries as compared to the EU27 average are shown in Table 10. We see that currently (2011) Sweden has 4.5 year of advance, while Bulgaria 8 years of delay, Czech Republic and Spain 3.5 years of delay, Poland only 1.5 years of delay. According to IOGSI indicator, Poland has a fast growth and in 2015 is predicted to have 1 year of advance, with growing advance until 2025.

These results are illustrated in Fig. 9. We can see the increasing forecasted advances of France and Poland.

6. General Conclusions

The examples of Belgium and Poland will be discussed here in more detail to stress the comparison of a core EU country and a post-communist EU country. The synthetic information for Belgium is summarized in Figs. 10, 11, and 12. We can see that Belgium is good on IRUI (social attitude to Internet) and HHBAI (broadband infrastructure), while it was delayed on IOGSI (broad social commercial use of Internet), but accelerates on IOGSI considerably.

Table 8
Advances and delays of the growth of HHBAI indicator as compared to the av. EU27 average, [year]
(negative entry denotes delay)

Year/ Country	Belgium	Bulgaria	Czech Republic	Finland	France	Germany	Poland	Slovenia	Spain	Sweden
1995	2.50	–5.00	–5.00	1.00	0.00	–4.50	–5.00	–2.00	–4.00	–2.50
1996	2.50	–6.00	–6.00	1.00	0.00	–5.00	–6.00	–1.50	–5.00	–2.00
1997	2.50	–6.50	–6.50	1.50	0.00	–4.50	–7.00	–1.00	–4.50	–1.00
1998	2.50	–7.50	–7.50	1.50	0.50	–4.00	–7.50	–1.00	–4.00	–0.50
1999	2.50	–8.50	–8.50	2.00	0.50	–3.00	–8.50	–1.00	–3.50	0.00
2000	2.50	–9.00	–9.00	2.00	0.50	–2.50	–9.00	–0.50	–3.00	0.50
2001	2.50	–9.00	–8.00	2.00	0.50	–2.00	–8.00	–0.50	–2.50	1.00
2002	2.00	–8.50	–7.00	2.00	0.50	–1.50	–7.00	0.00	–2.50	1.50
2003	2.00	–8.00	–6.00	2.50	0.50	–1.00	–6.00	0.00	–2.00	2.00
2004	2.00	–8.00	–5.00	2.50	0.50	–0.50	–5.00	0.00	–2.00	2.00
2005	2.00	–7.50	–4.00	2.50	0.50	0.00	–4.00	0.00	–1.50	2.50
2006	1.50	–7.00	–3.50	2.50	0.50	0.00	–3.50	0.00	–1.50	3.00
2007	1.50	–7.00	–3.00	2.50	0.50	0.50	–2.50	0.00	–1.00	3.00
2008	1.50	–6.50	–2.00	2.50	0.00	1.00	–2.00	0.00	–1.00	3.00
2009	1.00	–6.50	–2.00	2.50	0.00	1.00	–1.50	0.00	–1.00	3.50
2010	1.00	–6.00	–1.00	2.00	0.00	1.00	–1.00	0.00	–1.00	3.50
2011	0.50	–6.00	–1.00	2.00	0.00	1.50	0.00	–0.50	–1.00	4.00
2012	0.50	–6.00	–0.50	2.00	–0.50	1.50	0.00	–0.50	–1.00	4.00
2013	0.00	–6.00	0.00	2.00	–0.50	2.00	1.00	–0.50	–1.00	4.50
2014	0.00	–6.00	0.50	2.00	–1.00	2.00	1.00	–0.50	–1.00	5.00
2015	0.00	–5.50	1.00	2.00	–1.00	2.50	2.00	–1.00	–1.00	5.00
2016	–0.50	–5.50	1.50	2.00	–1.00	3.00	2.50	–1.00	–1.00	5.50
2017	–0.50	–5.50	2.00	2.50	–1.00	3.00	3.00	–1.00	–1.00	6.00
2018	–1.00	–5.50	2.50	2.50	–1.00	3.50	4.00	–1.00	–0.50	6.50
2019	–1.00	–5.50	3.00	2.50	–1.00	4.00	5.00	–1.00	–0.50	7.00
2020	–1.00	–5.50	4.00	3.00	–1.50	4.50	6.00	–1.00	–0.50	8.00
2021	–1.00	–5.50	5.00	3.00	–1.50	5.00	7.00	–1.00	0.00	8.50
2022	–1.00	–5.50	6.00	3.00	–1.50	6.00	8.00	–0.50	0.00	9.00
2023	–1.00	–5.50	6.50	3.50	–1.50	6.50	9.00	–0.50	0.50	10.00
2024	–1.50	–5.00	7.50	4.00	–1.50	7.00	10.00	–0.50	1.00	10.00
2025	–1.50	–5.00	8.50	4.00	–1.00	8.00	11.00	0.00	1.00	11.00

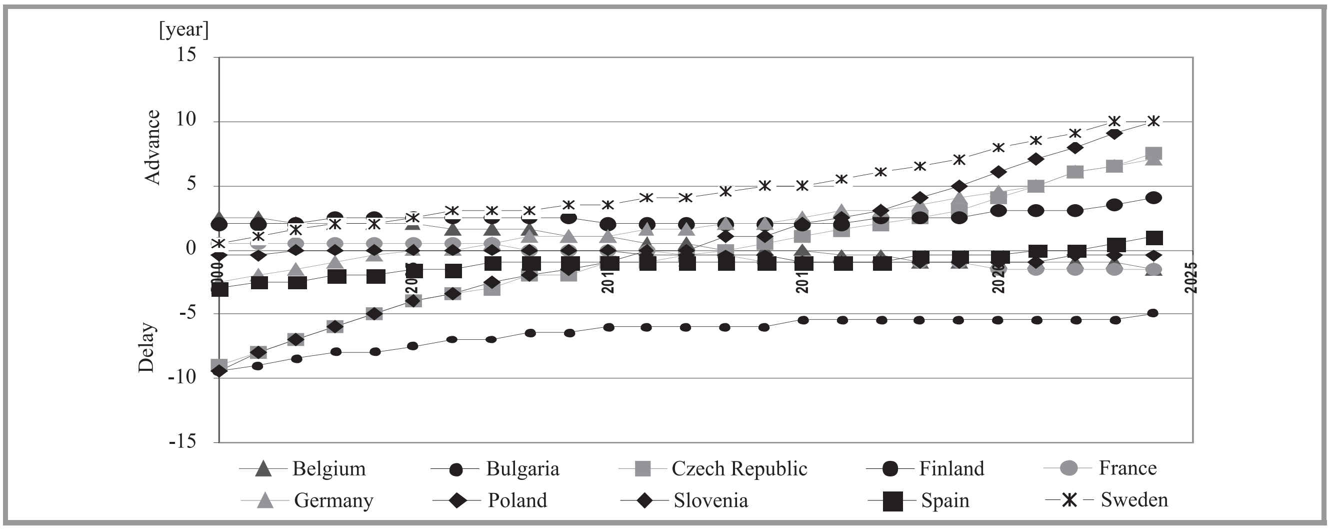


Fig. 7. Penetration of ICT in selected EU countries according to HHBAI indicator.

Table 9
Advances and delays of the growth of IRUI indicator as compared to the av. EU27, [year]
(negative entry denotes delay)

Year/ Country	Belgium	Bulgaria	Czech Republic	Finland	France	Germany	Poland	Slovenia	Spain	Sweden
1995	0.5	–9.5	5	–10.5	1.5	–8	–12.0	–3	–3	9
1996	0.5	–9	5	–9.5	2	–8	–11.5	–2.5	–2.5	9
1997	1	–8	5	–8.5	2	–7.5	–11.0	–2.5	–2.5	9
1998	1	–8	5	–8	2	–7	–11.0	–2	–2.5	9
1999	1	–7	5	–7	2	–6.5	–10.5	–2	–2.5	8.5
2000	1	–7	5	–6	2	–6	–10.0	–2	–2.5	8.5
2001	1.5	–6	5	–5	2	–6	–9.5	–2	–2.5	8
2002	1.5	–5.5	5	–4.5	2	–5.5	–9.0	–2	–2.5	8
2003	1.5	–5	5	–3.5	2	–5	–8.5	–1.5	–2	8
2004	1.5	–4.5	5	–3	2	–5	–8.0	–1.5	–2	7.5
2005	1.5	–4	5	–2	2	–4.5	–8.0	–1.5	–2	7.5
2006	2	–3.5	5	–1	2	–4	–7.5	–1.5	–2	7
2007	2	–3	5	–0.5	2	–4	–7.0	–1.5	–2	7
2008	2	–3	4.5	0	2	–3.5	–6.5	–1	–2	7
2009	2	–2.5	4.5	1	2	–3	–6.0	–1	–2.5	6.5
2010	2	–2	4.5	1.5	2	–3	–6.0	–1	–2.5	6.5
2011	2	–1.5	4.5	2	2	–2.5	–5.5	–1	–2.5	6
2012	2	–1	4.5	2.5	2	–2.5	–5.0	–1	–2.5	6
2013	2	–1	4	3	2	–2	–5.0	–1	–2.5	6
2014	2	–0.5	4	4	2	–2	–4.5	–1	–2.5	5.5
2015	2	0	4	4.5	1.5	–2	–4.0	–1	–2.5	5.5
2016	2	0	4	5	1.5	–1.5	–4.0	–1	–3	5
2017	2	0.5	4	6	1.5	–1.5	–4.0	–1	–3	5
2018	2	1	4	6.5	1.5	–1	–3.5	–1	–3	5
2019	2	1	4	7	1.5	–1	–3.0	–1	–3	4.5
2020	2	1.5	4	8	1.5	–1	–3.0	–1	–3	4.5
2021	2	2	4	8.5	1.5	–0.5	–2.5	–1	–3	4
2022	2.5	2.5	4	9.5	1.5	0	–2.5	–1	–3	4
2023	2.5	3	3.5	10	1.5	0	–2.0	–1	–3	4
2024	2.5	3	3.5	11	1.5	0	–2.0	–1	–3	4
2025	2.5	3.5	3.5	11.5	1.5	0.5	–1.5	–1	–3	3.5

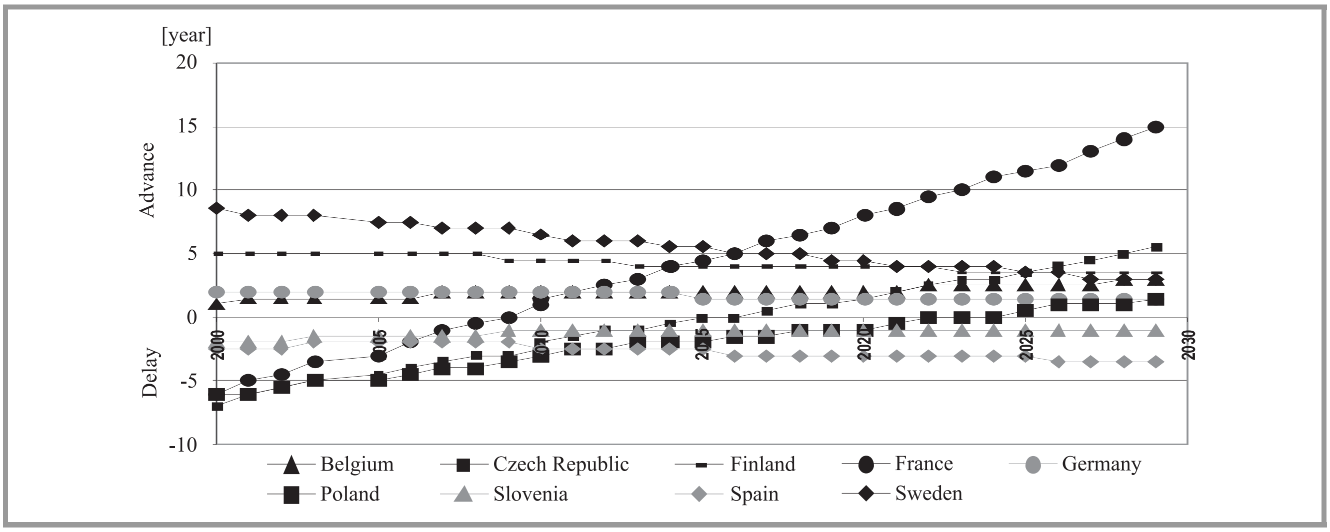


Fig. 8. Penetration of ICT in selected EU countries according to IRUI indicator.

Table 10
Advances and delays of the growth of IOGSI indicator as compared to the av. EU27, [year]
(negative entry denotes delay)

Year/ Country	Belgium	Bulgaria	Czech Republic	Finland	France	Germany	Poland	Slovenia	Spain	Sweden
1995	–8			3.5	–5.5	7		–11	–8	8
1996	–7			3.5	–4.5	6.5		–10	–7.5	8
1997	–6		–15	3.5	–4	6.5	–16	–9	–7	8
1998	–5.5		–13.5	3.5	–3	6.5	–14	–8	–6	7.5
1999	–5		–12	3.5	–2	6.5	–12	–7.5	–6	7.5
2000	–4		–11	3.5	–1.5	6.5	–10.5	–7	–5.5	7
2001	–3.5		–9.5	4	–1	6	–9	–6.5	–5	7
2002	–3	–1	9.–9	4	–0.5	6	–8	–6	–4.5	7
2003	–3	–17.5	–8	4	0	6	–7	–5.5	–4.5	6.5
2004	–2	–15.5	–7	4	0.5	6	–6	–5	–4	6.5
2005	–2	–14	–6.5	3.5	1	6	–5.5	–5	–4	6
2006	–1.5	–13	–6	3.5	1.5	6	–4.5	–4.5	–3.5	6
2007	–1	–12	–5.5	3.5	2	5.5	–4	–4	–3.5	5.5
2008	–1	–11	–5	3.5	2	5.5	–3.5	–4	–3	5.5
2009	–0.5	–10	–4.5	3.5	2.5	5	–2.5	–4	–3	5
2010	0	–9	–4	3.5	3	5	–2	–3.5	–3	5
2011	0	–8	–3.5	3.5	3.5	5	–1.5	–3.5	–3	4.5
2012	0	–7.5	–3	3.5	4	5	–1	–3	–3	4.5
2013	0.5	–6.5	–2.5	3	4	5	–0.5	–3	–2.5	4
2014	1	–6	–2	3	4.5	4.5	0	–3	–2.5	4
2015	1	–5	–2	3	5	4.5	1	–2.5	–2.5	3.5
2016	1	–4.5	–1.5	3	5	4	1	–2.5	–2	3.5
2017	1.5	–4	–1	3	5.5	4	2	–2	–2	3
2018	1.5	–3	–1	3	6	4	2	–2	–2	3
2019	2	–2.5	–0.5	3	6.5	4	3	–2	–2	2.5
2020	2	–2	0	3	7	4	3.5	–2	–2	2.5
2021	2.5	–1.5	0	3	7.5	3.5	4	–2	–2	2
2022	3	–1	0.5	3	8	3.5	4.5	–1.5	–2	2
2023	3	0	1	3	9	3.5	5.5	–1.5	–2	2
2024	3.5	0.5	1.5	3	10	3.5	6	–1	–1.5	1.5
2025	4	1	2	3	11	3.5	7	–1	–1.5	1.5

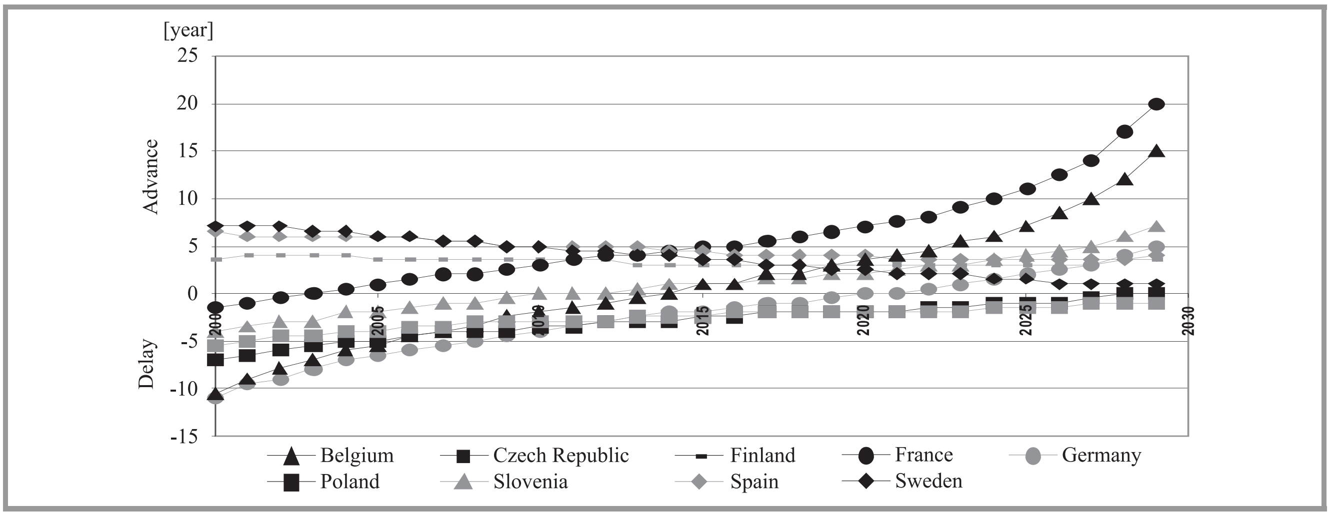


Fig. 9. Penetration of ICT in selected EU countries according to IOGSI indicator.

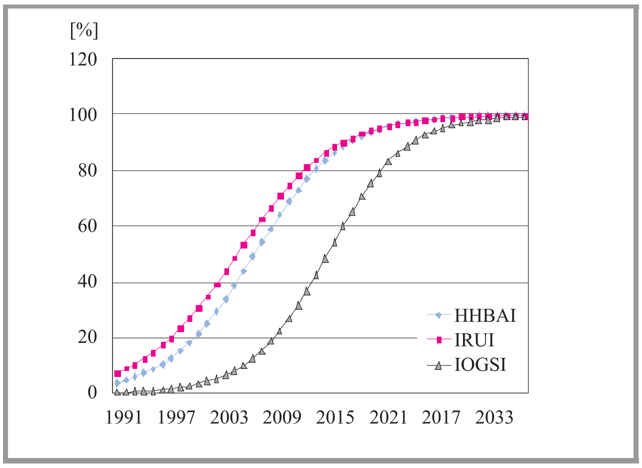


Fig. 10. Estimation for Belgium for the HHBAI, IOGSI, IRUI indicators.

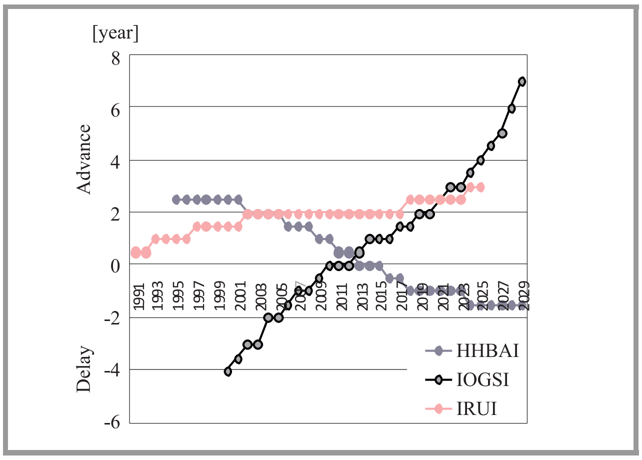


Fig. 11. Penetration of ICT for Belgium for the HHBAI, IOGSI, IRUI indicators.

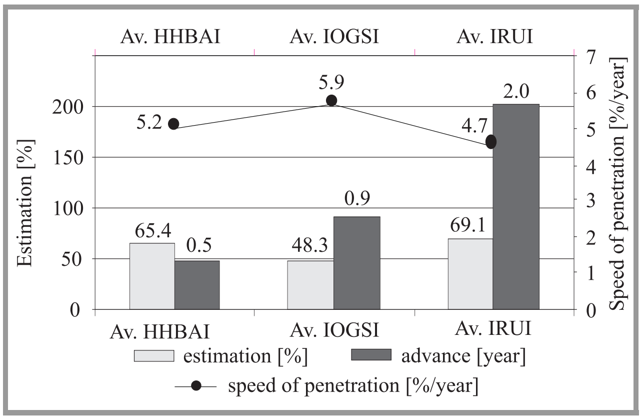


Fig. 12. Development of ICT for Belgium in relation to the av. EU27. Graphs for all indicators based on average values of each indcators over several years.

In Fig. 12, we see again that Belgium is the best in IRUI rate of the general use of the Internet. The same IRUI rate achieved the best result in terms of advance or delay – two years advance ahead of the EU average. However, in the category of the maximum rate of change (annual growth) – the highest rate of change has been an indicator of the development of the commercial use of the Internet – IOGSI, though, of course, percentages in this category are small and much smaller than in other countries. Generally, the pace of social adaptation of new ICT in Belgium is slow.

In Poland, the situation is quite different, as illustrated in Figs. 13, 14, and 15. In terms of the delay or advance, indicators and IRUI and IOGSI look poorly – IOGSI has the delay of 0.7 years to the EU average and IRUI the delay of even 2.8 years to the average. However, HHBAI and IOGSI have large speeds of development, as commented before. In general, Poland has the chance to catch up to the core EU countries in ICT development.

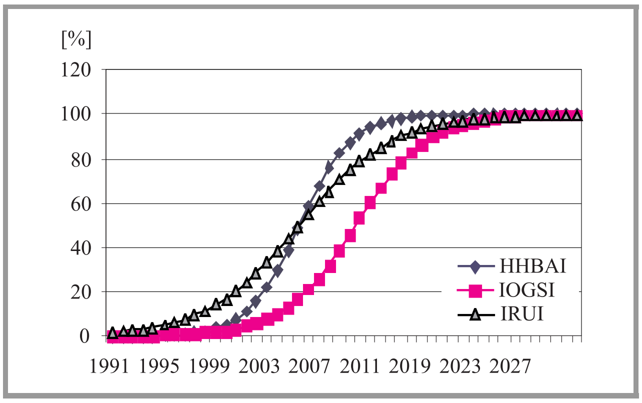


Fig. 13. Estimation for Poland for the HHBAI, IOGSI, IRUI indicators.

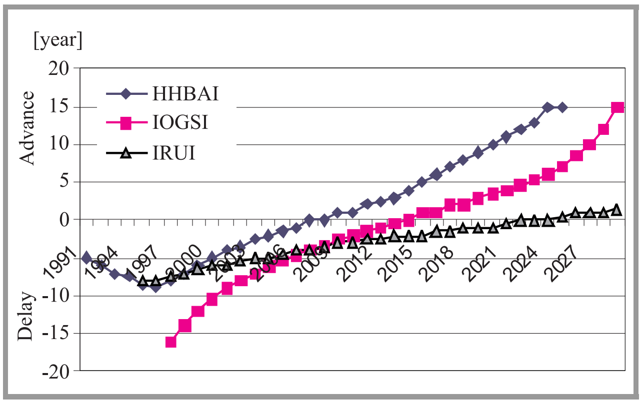


Fig. 14. Penetration of ICT for Poland for the HHBAI, IOGSI, IRUI indicators.

Generally, a graphical presentation of the development of several ICT indicators for a given country, such as in Figs. 10 or 14, gives a convincing kind of "digital signature" of this country. There are many further issues of research that could not be addressed in this paper because of volume limitations. To such issues belong the problem of correlation or causal link of the growth of gross domestic product and the use of ICT. Another already mentioned problem is the issue of statistical estimation and fore-

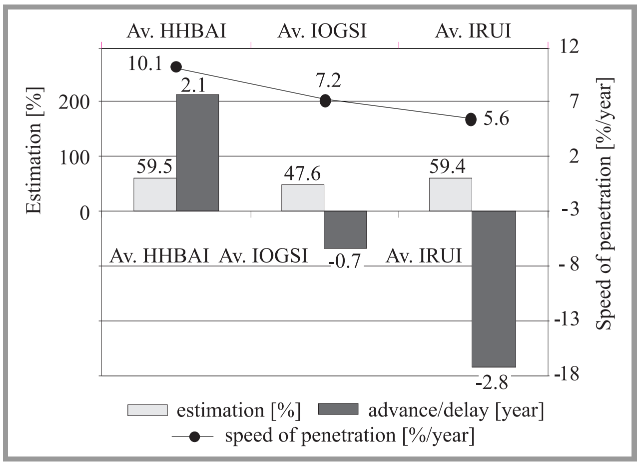


Fig. 15. Development of ICT for Poland in relation to the av. EU27 Graphs for all indicators based on average values of each indicators over several years.

casting of digital divide or exclusion. The richness of these subjects justifies separate articles in this respect.

References

[1] M. Castells, *Information Society*. Warszawa: PWN, 2009.

[2] J. Grzegorek and A. P. Wierzbicki, "New Statistical Approaches in the Systemic Analysis of Regional, Intra-regional and Cross-Regional Factors of Information Society and Economic Development: The Case of Mazovia", *Mazowsze: Studia Regionalne*, no. 3, pp. 117–128, Warszawa, 2009.

[3] J. Grzegorek, "Zastosowanie zaawansowanych metod statystycznych do badania rozwoju społeczeństwa informacyjnego Polsce traktowanej jako system regionów", Ph.D. thesis, Instytut Łączności – Państwowy Instytut Badawczy, Warsaw, 2012 (in Polish).

Beata Ziewiec studied at Faculty of Electrical Engineering and Automation of Kielce University of Technology and received the M.Sc. degree in 1988. She completed post-graduate studies Management, Marketing, Information and Communication Technology – at the Faculty of Information Systems Institute in Warsaw under the auspices of the Polish Academy of Sciences (PAN). She studied at postgraduate Projects Management at the Organization and Management Institute under the auspices of the Polish Association of IT (PTI) (2001–2002). She took part in doctoral seminars including, among others lectures on the development of new technologies – ICT, the characteristics of the information society, research methodology, research including mathematical models and programming tools (2007–2009)

E-mail: beata.maria0962@gmail.com

Paper

Performance and Limitations of VDSL2-based Next Generation Access Networks

George Heliotis, Lowell-Panayotis Dimos, Ioannis Kordoulis, and George Agapiou

Research & Development Department, Hellenic Telecommunications Organization (OTE) S.A., Athens, Greece

Abstract—In this paper, we first briefly review the main operational aspects of FTTC (xPON)/VDSL2 access networks. Then, we present performance measurements with respect to network sub loop operating length and VDSL2 operational profiles (8b, 12a, 17a), studying the rates that can be achieved at different scenarios. We then provide results of the effect of stationary crosstalk and non-stationary impulse noise on the provisioned service in an operating VDSL2 access network.

Keywords—DSL networks, impulse noise, VDSL2 next generation access networks, crosstalk.

1. Introduction

VDSL2-based Next Generation Access Networks based on Fiber-To-The-Cabinet (FFTC) architectures are currently deployed on worldwide scale, and offer a favourable intermediate step towards Fiber-To-The-Home access implementations [1]. FTTC/VDSL2 networks offer the prospect of delivering very high bandwidth (up to 200 Mbit/s) to end-users which should be more than sufficient for the near future, combined with much lower deployment costs than fully optical access networks. Indeed, the combination of optical xPON (e.g. GPON, XGPON) FTTC networks with the legacy copper last mile boosted by VDSL2 technology offers a very attractive balance between cost/performance.

However, to hold the promise of delivering the high bandwidth the VDSL2 technology is capable for, noise-induced performance degradation issues must be studied and addressed. Indeed, the Quality of Service (QoS) in VDSL2 implementations may suffer from severe degradation, with the two principal sources of degradation being stationary (self-crosstalk) and non-stationary (impulsive) noise [2]. While there is a large number of papers in the literature that theoretically consider these two noise sources, experimental studies of the effects of noise on real VDSL2 networks are scarce. In this paper, we first present performance measurements with respect to network sub loop operating length and VDSL2 operational profiles (8b, 12a, 17a), studying the upper limit of the rates that can be achieved at different scenarios and without the effect of noise. We then provide results of the effect of stationary crosstalk and non-stationary impulse noise on the provisioned VDSL2 service.

2. VDSL2 Single-Line Performance

2.1. VDSL2 Operational Profiles

The realization of the first phase of an NGA network is achieved by combining further penetration of the optical technology into the access network with the VDSL2 access technology [3].

The FTTC topology enables the provisioning of extremely high data rates on longer loops, while extending the reach of the fiber network assisting the future transition towards a fully optical access network. Additionally, VDSL2 is an access technology capable of a total rate of up to 200 Mbit/s, while offering a wide range of alternative profiles providing Telcos a very flexible field of options to choose from, depending on their business plan.

The available profiles mainly differ in terms of total output power and available spectrum. Table 1 lists the main characteristics of all the available profiles of the 998 ADE VDSL2 band plan, according to the ITU G.993.2 standard [4]. The inherent spectral and power characteristics of each profile determine whether it should be deployed either from the Central Office (CO), or from a remote DSLAM located in the remote cabinet (or RDF – Remote Distribution Frame). Generally, higher frequencies are more susceptible to attenuation, degrading much faster than lower frequencies as the length increases. VDSL2 8b is characterized by roughly four times the spectrum of ADSL2+ and the same maximum aggregate downstream power. As a result, it provides a higher bit rate than ADSL2+, while maintaining the same effective reach. VDSL2 12a transmits at a lower power level and employs a wider upstream frequency range, thus offering higher data rates but on shorter loops. Similarly, the trade-off between maximum bit rate and effective reach is even more evident on VDSL2 17a and VDSL2 30a.

2.2. Rate Measurements

Rate measurements were performed to assess the performance of the most representative VDSL2 operational profiles for deployment in a FTTC topology (i.e., injection of VDSL2 signal at a remote cabinet with a relatively shot sub-loop distance). A VDSL2 DSLAM (DSL Access Multiplexer) was used that was connected to a CPE (Customer Premises Equipment) via a 0.4 mm diameter UTP copper

Table 1
VDSL2 ADE 998 profile characteristics

Profile	Maximum aggregate downstream transmit power [dBm]	Maximum aggregate upstream transmit power [dBm]	Highest supported data-bearing downstream frequency [MHz]	Highest supported data-bearing upstream frequency [MHz]	Minimum bidirectional net data rate capability [Mbit/s]
8a	17.5	14.5	8.5	5.2	50
8b	20.5	14.5	8.5	5.2	50
8c	11.5	14.5	8.5	5.2	50
8d	14.5	14.5	8.5	5.2	50
12a	14.5	14.5	8.5	12	68
12b	14.5	14.5	8.5	12	68
17a	14.5	14.5	17.660	12	100
30a	14.5	14.5	24.890	30	200

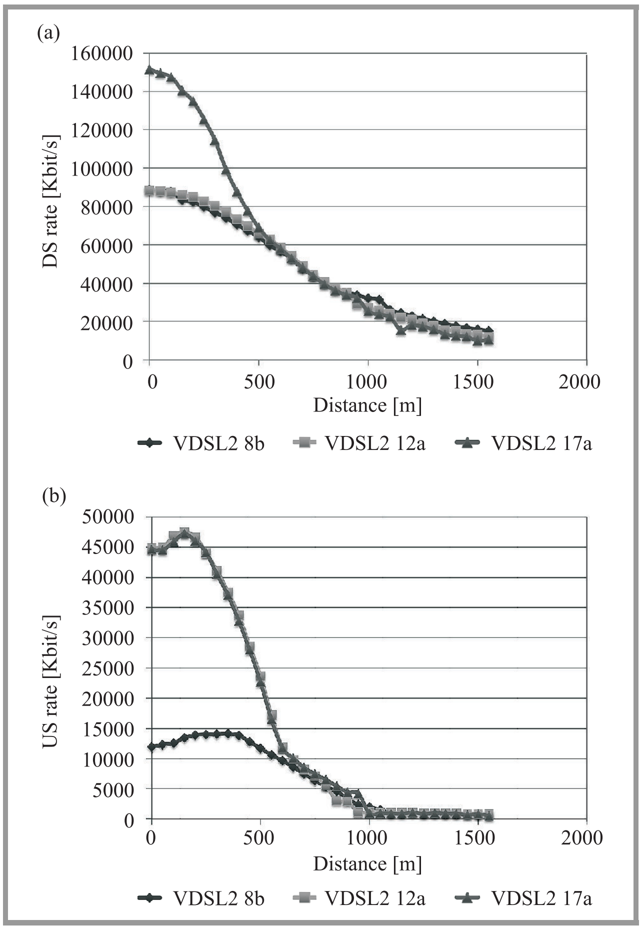


Fig. 1. Attainable downstream rate as a function of the distance between the remote cabinet and the CPE (a); Attainable upstream rate as a function of the distance between the remote cabinet and the CPE (b).

line simulator. The length of the simulated line was varied in a broad range in order to evaluate the performance under different Remote Cabinet CPE distances that may occur in the field. Figures 1(a) and 1(b) show the attainable downstream and upstream rates respectively as a function of the distance between the remote cabinet and the CPE for the VDSL2 profiles 8b, 12a and 17a, the most representative deployment profiles.

3. VDSL2 Performance with Crosstalk

The co-transmission of signals from two distinct points inside the same local loop presents the challenge of crosstalk (electromagnetic interference). The propagation of an electromagnetic signal inside a copper conductor results in the emission of electromagnetic fields, which in turn interfere with adjacent copper pairs and electromagnetic coupling phenomena emerge. As an outcome, the signal affected by crosstalk can sustain severe degradation, shortening its effective reach and/or limiting its data-bearing capability. As Figure 2 shows, the two eminent types of crosstalk caused by the downstream and upstream signals are FEXT (Far End Crosstalk) and NEXT (Near End Crosstalk). FEXT refers to the interference caused by signals on neighbouring lines propagating in the same direction, while NEXT refers to interference from counter propagating signals. Due to the fact that xDSL technologies employ FDMA, NEXT tends to be considered as having a minor effect as the downstream and upstream frequency bands are spectrally separated.

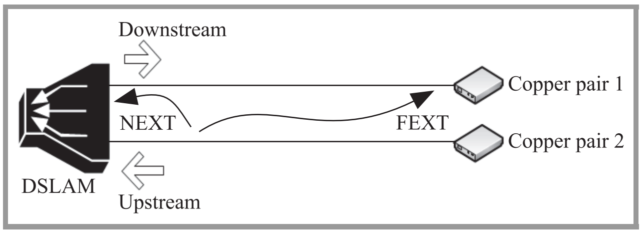


Fig. 2. Illustration of FEXT and NEXT. Downstream directions refers to signals from the DSLAM to the CPEs, while upstream from the CPEs to the DSLAM.

The severity of crosstalk is directly connected to the output power of the disturber line creating the interference, as well as to the power level of the victim line receiving it. Understandably, an attenuated signal is more sensitive to crosstalk than a robust signal. This is the case in access networks including active equipment both at the CO and at the RDF. A strong signal induced at the RDF meets an attenuated signal that has already lost a considerable amount of its power and therefore decreases the available SNR margin for bit loading. This impairment is followed by a considerable decrease of the CO line's effective reach. In that case some sort of downstream power back off (DPBO) protection needs to be applied. Here, we only consider signals that are injected at the remote cabinet, as applying to the case of VDSL2 provisioning in an FTTC topology. Hence, all signals are generated at the same point with equal power. The DPBO technique here cannot be applied, hence our measurements show the rate that can be realistically achieved in such cases.

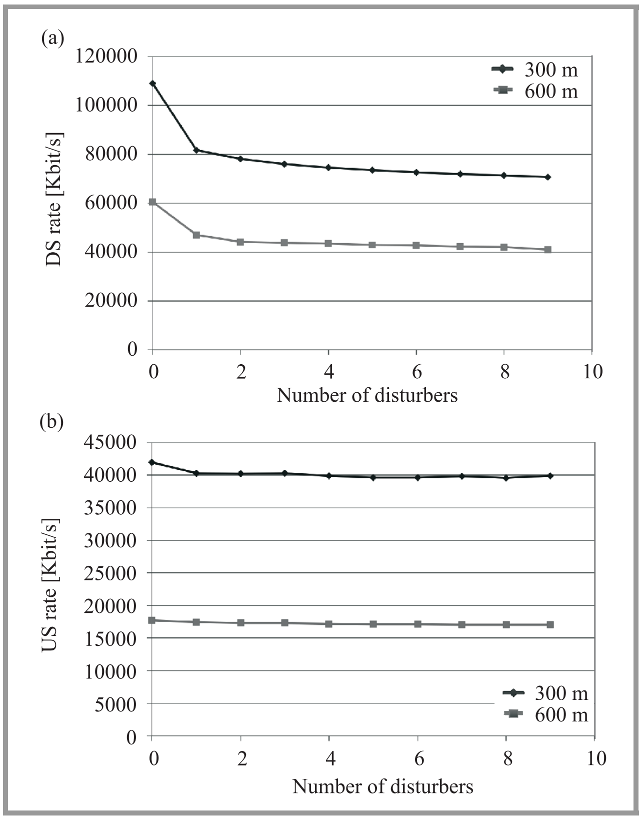


Fig. 3. Downstream (DS) and upstream (US) VDSL2 performance as a function of active disturbers within the same binder.

Figure 3 shows the individual laboratory performance (uplink and downlink rates of transmission) of a line with VDSL2 spectrum allocation plan Annex B: 998ADE17 and how it is affected due to stationary self-crosstalk noise from the gradual activation of identical adjacent lines. The profile used is the 17a, which is the most common for cabinet deployment as it ensures a high bandwidth for the short sub-loops expected in these cases. Ten total lines are used, and this case corresponds to a fully occupied 10-pair distribution cable binder. Data are shown for two representative sub-loop lengths: 300 m and 600 m.

As can be seen, the self-crosstalk phenomena, which are strengthened by activating neighbouring lines have significant adverse effect on the performance can be achieved. Specifically, a decrease in transmission rates as high as 30% can be observed with the increase of the neighboring active lines. The performance drop rate is somewhat saturated when the number of disturbers exceed 4–5, possibly as a result of the pairs geometry within the cable binder.

4. VDSL2 Performance with Impulse Noise

Impulse Noise together with Crosstalk constitute the major performance impairment factors of the VDSL2 copper access network which unlike attenuation their impact cannot be predicted. In contrast to crosstalk, whose nature can be considered as stationary, given that the disturbers' kind and number do not change, impulse noise (which is usually induced by nearby electrical apparatus) is regarded as stochastic. Under the scope of this study, a series of measurements were conducted in order to evaluate the effects of bursts of REIN (Repetitive Electrical Impulse Noise) to the performance of VDSL2 998ADE17 lines. The duration of the impulse noise applied on the lines was predefined to 250 ms, while its power was –110 dBm/Hz. Multiple operational profiles are used, in which the SNR margin is either 9 or 11 dB, either in fast mode or in interleave mode (depth = 10 ms) and with an impulse noise protection of 2 symbols. Figure 4 shows some initial results on the impact of REIN on the performance of VDSL2 lines as a function of sub-loop length and operational profile param-

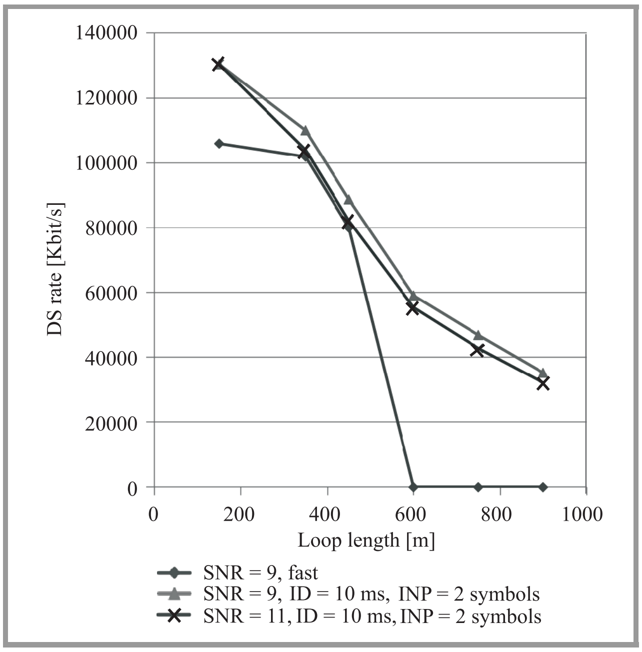


Fig. 4. Downstream rate in a VDSL2 line with REIN for three different service profiles.

eters. It can be seen that REIN can have a severe impact on the sync rates, and varies with length as well as the interleave settings of the service profile. We are currently performing a detailed study to resolve the interplay of these factors with VDSL2 performance and construct optimized service profiles.

5. Conclusions

As was evidenced in the previous paragraphs, crosstalk and impulse noise phenomena can have a very detrimental effect on the performance of VDSL2 NGA networks. Therefore, the further analysis and the addressing (e.g., with dynamic spectral management methods) of such adverse effects of is absolutely necessary in order to improve service quality in such access networks. Here, we presented initial results of our ongoing experimental study.

Acknowledgements

This work was partially supported through the Hellenic General Secretariat of Research and Technology (GSRT) funded collaborative research project PANDA (09SYN-71-839).

References

[1] M. Tsubokawa and K. Kumozaki, "Evolution of Next Generation Access", in *Proc. IEEE Global Telecommun. Conf. GLOBECOM 2008*, New Orleans, LA, USA, 2008.

[2] J. W. Cook, R. H. Kirkby, M. G. Booth, K. T. Foster, D. E. A. Clarke, and G. Young, "The Noise and Crosstalk Environment for ADSL and VDSL Systems", *IEEE Commun. Mag.*, vol. 37, no. 5, pp. 73–78, 1999.

[3] J. M. Cioffi *et al.*, "Very-High-Speed Digital Subscriber Lines", *IEEE Commun. Mag.*, vol. 37, no. 4, pp. 72–79, 1999.

[4] "Very high speed digital subscriber line transceivers 2", ITU-T Rec. G.993.2 (02/2006).

[5] C. Zeng, C. Aldana, A. A. Salvekar and J. M. Cioffi, "Crosstalk Identification in xDSL Systems",*IEEE J. Selec. Areas Commun.*, vol. 19, no. 8, pp. 1488–1496, 2001.

George Heliotis is a telecommunications engineer, holding a Ph.D. from Imperial College London. He currently works at the Access Networks Lab of the R&D department of the Hellenic Telecommunications Organization (OTE S.A.). He has participated in several European Community research projects and has authored and has authored several articles in high-profile scientific journals and conferences. His current work focuses into the transition to next generation access networks, involving the experimental evaluation of fully optical access networks implementations (mainly xPON), hybrid solutions such as Fiber-To-The-Cabinet (FTTC) + VDSL2, as well as value-added broadband services over these networks.
E-mail: gheliotis@oteresearch.gr
Research & Development Department
Hellenic Telecommunications Organization (OTE) S.A.
Pelika and Spartis st 1
15122 Athens, Greece

Lowell-Panayotis Dimos is a graduate student at the Department of Electrical & Computer Engineering of the National Technical University of Athens (NTUA). His Diploma Project Thesis concentrated on the investigation of spectral compatibility issues in NGA networks of varying xDSL technologies, and involved experiments at the R&D Department of OTE S.A. He also held a practical internship at Euroconsultants S.A., where he participated in a consultative study about FTTH deployment in Greece. He currently works at the Access Networks Lab of the R&D department of OTE S.A., focusing on studies of novel optical access network implementations (xPON FTTH, FTTC/VDSL2).
E-mail: pdimos@oteresearch.gr
Research & Development Department
Hellenic Telecommunications Organization (OTE) S.A.
Pelika and Spartis st 1
15122 Athens, Greece

Ioannis Kordoulis received the diploma degree in Electrical Engineering from the Polytechnic School of University of Patras in 2001 and a Master degree in Techno-Economic systems from the National Technical University of Athens (NTUA) in 2007. He worked as Electrical Engineer for Atmel S.A. at VoIP department, focused mainly in VoIP technology. He was experienced in designing and programming of embedded communication systems for Wi-Fi as well as 3G technology. Now he works for OTE S.A. at Voice Switching Lab involved in VoIP, IN, OSA, IP-PBX and IMS projects.
E-mail: ykordoulis@oteresearch.gr
Research & Development Department
Hellenic Telecommunications Organization (OTE) S.A.
Pelika and Spartis st 1
15122 Athens, Greece

George Agapiou is a telecommunications engineer, holding a Ph.D. from the Georgia Institute of Technology. He is the Head of the Measurements and Wireless Technologies Research Laboratory of the Hellenic Telecommunications Organization (OTE S.A.) at the Research & Development Department. He has participated in various IST, STREP, Eurescom and e-ten research projects and has published more than 40 papers in scientific journals and proceedings. His research interests include next generation wired and wireless access networks and novel broadband technologies and services.

E-mail: gagapiou@oteresearch.gr
Research & Development Department
Hellenic Telecommunications Organization (OTE) S.A.
Pelika and Spartis st 1
15122 Athens, Greece

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The author's JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX author's style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, "Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres", *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, "Informational representability: Abstract models versus concrete models", in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314.

Biographies and photographs of authors. A brief professional author's biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.
For copyright form see: http://www.nit.eu/for-authors

A copy of the JTIT is provided to each author of paper published.

(Contents Continued from Front Cover)

INSTYTUT ŁĄCZNOŚCI
PAŃSTWOWY INSTYTUT BADAWCZY

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu