JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2015

(*Contents Continued on Back Cover*)

Editorial Board

Editorial Advisory Board

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

Circulation: 300 copies

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

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Quality of Service is becoming one of the most important criteria for the selection of products and services.

If anyone buys a food or a washing machine, almost always pays attention to their quality. It's hard to deny that the quality is also of great importance in telecommunications and services offered by operators. Sooner or later an operator, who does not pay attention to the quality of services, will notice a certain decrease in the number of customers. This is because the functionality of the services offered by different operators is similar or the same. Therefore, one of the most important performance indicator to compete with each other is the quality of service. Poor quality of service is immediately noticed by the user. Signs of deterioration in the quality of services are many. For example, in the case of mobile telephony from time to time there are cases called silent calls. Spectators watching movies or TV do not want the image disappeared, or was frozen. Using the Web you want transmission rate was large enough that you did not notice the delay. Technical reasons for unwanted phenomena can be many. Caring for the quality of services is not easy. This leads many researchers to carry out research in this area. There are also organized numerous conferences on the subject, and also this issue of the *Journal of Telecommunications and Information Technology* presents selected papers on issues of quality of service in telecommunications and mobile systems, to support the topic. Articles describe how to define quality of service, how it is measured and tested. Issues related to quality are becoming more and more frequently a source of ITU-T recommendations.

This issue of the *Journal* presents selected papers on QoS which are the main part of the publication. The content is also supplemented by four other articles. They are devoted to performance comparison of four new ARIMA-ANN prediction models on Internet traffic data, a new implementation of UWB CRLH based antennas for wireless communications, availability of WDM multi ring networks, and assessment of area energy efficiency of LTE macro base stations in different environments.

I hope that readers will find its contents interesting and valuable.

Sławomir Kula
Guest Editor

Paper

New Tool for Investigating QoS in the WWW Service

Tadeus Uhl[1] and Marcus Rompf[2]

[1] *Maritime University of Szczecin, Szczecin, Poland*
[2] *Flensburg University of Applied Sciences, Flensburg, Germany*

Abstract—The subject of this paper is the QoS measurement in the World Wide Web (WWW) service. It will begin with a brief presentation of methods currently used to measure QoS in the WWW service before going on to describe in detail the new tool QoSCalc(WWW). Following that, the potential of the new tool will be put to the test in a series of analysis scenarios. The results and insights obtained will then be presented in several diagrams, and interpreted.

Keywords—Apdex index, communication network, ITU-T G.1030, measurement tool, Power metric, QoS/QoE determination, Triple Play Services, WWW service.

1. Introduction

Quality of Service (QoS) plays a very important role in modern digital networks. The term is becoming a household word and can be found among other things in the definition of Next Generation Networks according to the ITU-T Standard Y.2001 [1]. In 2009 the European Parliament and European Council published directives for the networks standardization and services [2], [3], placing great priority on quality of service.

It cannot have escaped anyone's notice that the WWW service has grown enormously in recent years. Almost everybody accesses the information stored on WWW servers. It is impossible to imagine the modern communications society without this service. The end-user can choose between several browsers with which to access the service, hoping that the desired pages located on a server somewhere on the Internet will be loaded and displayed rapidly. Sadly, this is, however, not always the case. Some WWW servers are often overloaded by requests. The information might often be spread over several servers. The time taken to access the pages becomes stretched. Some network sections might become congested. Congestion causes considerable delays which in turn impair communication within the WWW significantly. All this means that the time taken to open a Web page can become very long. To be sure, the delay may be caused in part by the availability of resources at the end-user. Last-mile transmission rates do play an important role as well. So the question arises: how can the quality of the WWW service be measured quickly and objectively no matter where one happens to conduct the test? At the time of writing there are only very few methods that are capable of doing this. One of them is the Standard ITU-T G.1030 [4], a second is the industry-standard method Apdex [5], and a third is the proprietary metric called "Power" [6]. The first two methods judge the quality of the WWW service from the end-user's point of view, giving a value for the Quality of Experience (QoE). The third method provides not only a way to measure QoE, but also QoS, that is to say the quality of service from the WWW service provider point of view. This method takes not only the time taken to open a Web page into account (which is the case for the first two methods), but also the rate at which the data of a Web page are loaded. That is unique. The metric "Power" has yet to be fully implemented in an efficient tool. This paper aims to do that.

To begin with, the paper in Section 2 provides a short overview of present methods for measuring QoS in the WWW service. Following that, the new tool QoSCalc(WWW) will be described in detail in Section 3. Its capabilities will be put to the test in a variety of study scenarios. The results and insights gained through this study will then be presented in several diagrams, and interpreted in Section 4. The paper will conclude with a summary and an outlook on future areas of work in Section 5.

2. QoE/QoS Measurement Techniques for the WWW Service

2.1. ITU-T G.1030 Standard

The ITU-T G.1030 Standard [4] covers the relationships between the parameters of a service and subjective appraisal of the quality of that service by the end-user. So it is, in fact, a measurement of the Quality of Experience. This method is one of the so-called one-sided QoE/QoS measurement models (see [7] for details). The chief parameter under consideration is the time taken to open a Web page. Various threshold values are used to assess the quality of service subjectively:

- threshold value 0.1 s – maximum value for the reaction time without any impairment to communication,
- threshold value 1 s – maximum value of the reaction time without impairment to the smooth operation of the application,
- threshold value 10 s – maximum value for the time which can elapse during the operation of an application without the user becoming frustrated.

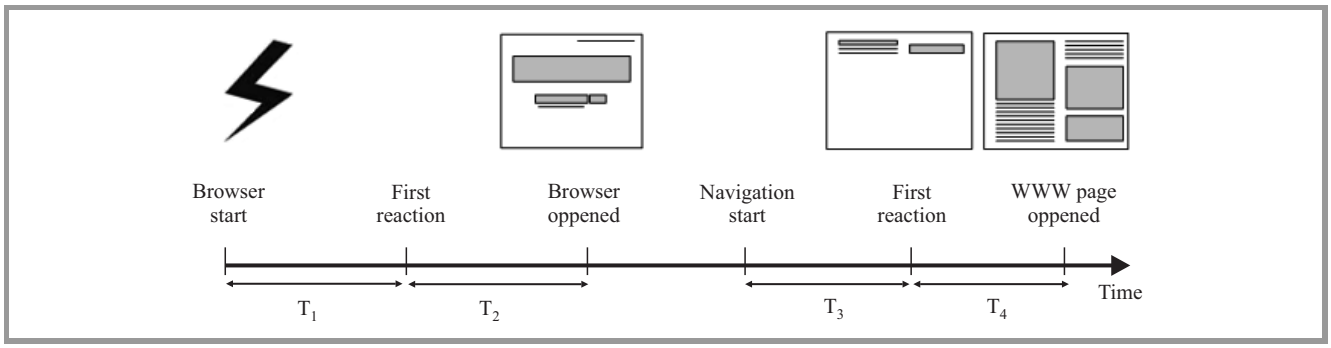


Fig. 1. Time slices in the determination of QoE according to G.1030.

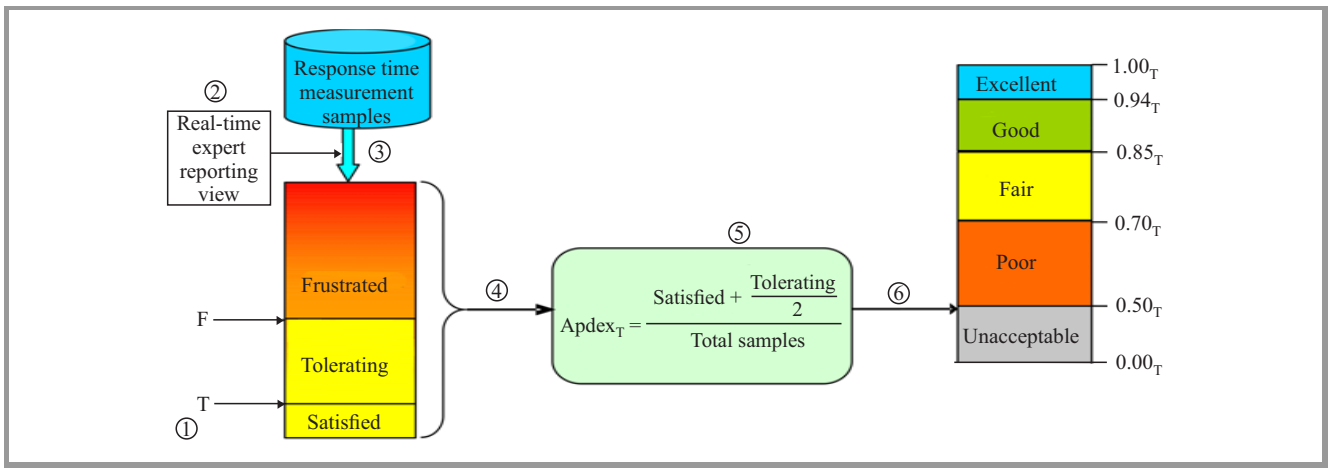


Fig. 2. Apdex process overview.

For the measurement of QoE covered in the Recommendation several time slots are taken into account, as Fig. 1 shows.

The G.1030 Standard includes an experiment description in which a distinction is made between two groups of test persons: (a) test persons with considerable experience and (b) test persons with little or no experience. The information supplied by G.1030 allows one to make the following observations:

- both groups reacted similarly, identifying in their subjective judgements an inversely proportional relationship between service parameters and quality of service;
- persons with little or no experience were more critical than persons familiar with the WWW service;
- as sessions grow longer (in excess of 60 s) the discrepancies between the judgements given by the two groups began to diminish.

The impairment parameter under scrutiny here – whether being measured or appraised – is the time taken for a Web page to open. No account is made of local conditions and parameters. Moreover, this kind of QoE/QoS analysis is extremely time and resource-consuming. Consequently, this measurement technique stands little chance of being used successfully in practice.

2.2. Apdex Index

The Apdex Index Method [5] is a proprietary solution that has yet to be standardized. It, too, is a method for measuring both QoE and QoS in the WWW service. It, too, is a one-sided measuring model. It can be applied without recourse to a group of test persons, but only if specialized tools are then assigned to determine the parameters necessary for this metric. Its chief criterion of evaluation is also the time taken for a Web page to open. The assessment of the quality of service yields a value on a scale from 0 (unacceptable) to 1 (excellent). The assessment depends to a large extent on which threshold time T is used. All relationships (and the metric itself) are clearly shown in Fig. 2.

Steps 1 to 6 have the following meanings:

1. Define threshold T (in seconds) for the application.
2. Real-time expert reporting group – view by application, server, or user.
3. Extract data set from existing measurements for viewing.
4. Count the number of samples in the three performance zones.
5. Calculate the Apdex value (see Apdex_T formula).
6. Display Apdex result.

The T value is the application target time (threshold between satisfied and tolerating users). The F threshold between tolerating and frustrated users is calculated as $F = 4T$. This method can be used in combination with specialized tools in practice.

2.3. Metric Power

The metric Power was first presented in paper [6]. It is a further method of measuring both QoE and QoS that is to be classed as a one-sided measurement model. Unlike the methods described above this method takes a further parameter into account that can be crucial for the WWW service: data download rate. As a result, the method is more versatile and therefore of much more practical use. Equation (1) describes the metric Power.

$$P = \frac{1}{1+\frac{\alpha}{ddr}}, \tag{1}$$

where: α – delay coefficient, ddr – download data rate in [Mb/s], td – total delay in [s], th – threshold of delay in [s],

$$\alpha = \begin{cases} 0 & \text{for } td \le th \\ (td - th) & \text{for } td > th \end{cases}.$$

It is obvious that both the download data rate and the total delay (time lapse between the point in time at which the session begins and the point in time at which the Web page is completely built up) depend on the throughput of the used transmission canal. But they also depend on the locality of the WWW server, its activity level, on the content of the Web pages being accessed, and whether that content is static or dynamic. Total delay can therefore vary within a considerable range. The new metric Power takes that into account.

First results, presented in paper [6], have confirmed that the new metric is of real practical use in principle, the only hitch being that it has not until now been fully implemented in a specialized tool, nor undergone exhaustive tests in a variety of realistic scenarios. And that is where this paper comes in.

3. The New Tool for Analysing Quality of Service in WWW

The tool that was developed in the course of this work to analyse QoS in the WWW service has been called QoSCalc(WWW). It can be used for both objective but also subjective analysis of service quality, i.e. QoE. It was created using the programming language C# / .NET 4.5 [8] and works with the Web browser object Internet Explorer [9]. The tool uses the following, fundamental classes:

- SharpPcap: provides the possibility of recording the network traffic of the hardware,
- PacketDotNet: provides the possibility of converting the recorded network traffic into the corresponding format (TCP/IP packet),
- Webbrowser: facilitates navigation and phase separation,
- WebClient: provides general methods for sending data to and receiving it from a resource labelled with a URI,
- DateTime: manages time stamps,
- TimeSpan: for mathematical operations on date-time objects.

The arrows in Fig. 3 clarify how the new tool works, the single steps being:

1. First navigation phase.
2. First data download phase.
3. Page set-up in Web browser object.
4. Storage in computer cache.
5. Start of a complete page refresh.
6. Refresh data download phase.
7. Refresh page set-up in Web browser object.

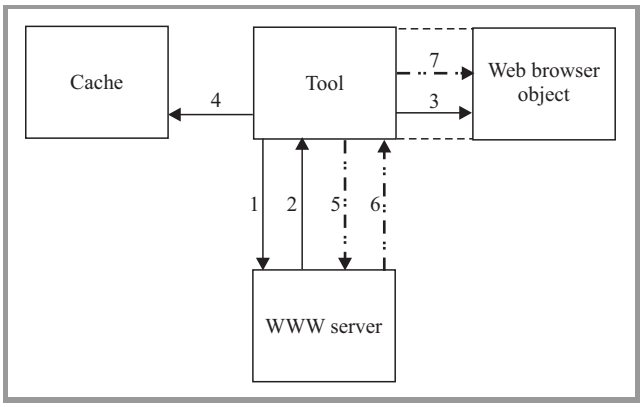


Fig. 3. New tool operation principle.

The first navigation phase and the first data download phase assist in timing the start of the measurement. During these phases the tool checks whether outsourcing is being used for the selected Application Service Provider (ASP), or not. If outsourcing is being used, the tool then looks to see how many WWW servers are being used to store contents elements that are being called up. The fact that the called page must always be built up from scratch in the Web browser object can be used to good effect when it comes to determining QoE. When the measurement has been completed, the tool displays – without any further action on the part of the user – the QoS values calculated in accordance with the metrics Power and Apdex.

Figure 4 shows the finished user tool overlay. The user can choose between three languages: English, German and Polish by using the button EN, DE, PL. The user overlay has three main areas: settings, display and results.

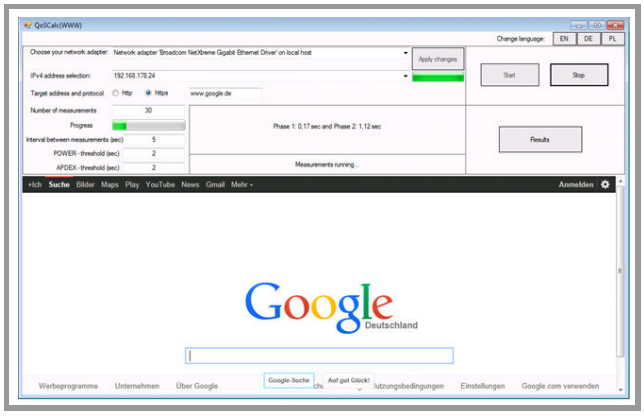


Fig. 4. QoSCalc(WWW) user interface.

In the settings area parameters for the measurement scenarios are set. It contains the following control elements: network adapter (selection box), IPv4 address selection (selection box), target address (text field), protocol (button), number of measurements (text field; default = 30), progress (gas gauge), interval between measurements (text field; default = 2 s), Power threshold (text field; default = 1 s), Apdex threshold (text field; default = 1 s), and accept (button).

In the display area "Preliminary Results" (middle of the user overlay) the confidence interval for the total delay and the results from the metric Power are displayed as soon as the measurement has been successfully completed. The field also outputs the correct target address, along with the instruction to repeat the measurement, should this be necessary.

The results area could be kept relatively simple. It has only three buttons: start, stop and results. The measurement is initiated with the start button. A measurement in progress can be aborted at any time with the stop button. Once a measurement is complete, the results page (shown in Fig. 5) can be called up with the results button.

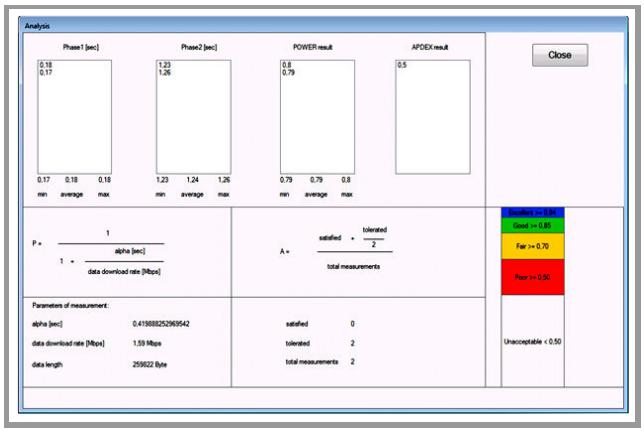


Fig. 5. User overlay of the results page.

The results page comprises 4 scrollable lists. They give the user a clear picture of what the tool has ascertained. As a further help there is a mathematical representation of each metric: P, i.e. metric Power, and A, i.e. Apdex. There is also a colored scale that helps to classify the results. The results page is closed with the close button; the user returns to the main page.

The reliability of the finished tool QoSCalc(WWW) was tested in exhaustive trials. Its operability was then further tested in a real-life environment. This environment and the measurement scenarios in operating within it are described in the next chapter.

4. Measurement Environment, Measurement Scenarios and Measurement Results

Figure 6 shows the measurement environment and how the new tool was implemented. This measurement environment was chosen to provide access both to remote WWW servers on the Internet and a local WWW test server. The authors decided to extend the work scope described in this paper to access external servers in order to be confronted with the problems which real network traffic causes. In addition to the impairment parameters which this causes, a wanulator [10] was included with which network parameters such as packet loss, jitter and delay could be emulated.

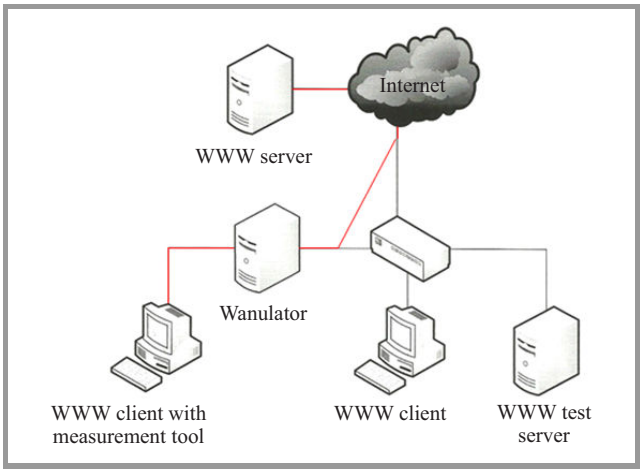


Fig. 6. Arrangement of the measurement environment.

In the following, two measurement scenarios are assumed:

- Scenario 1
  - number of measurements: 30,
  - threshold "Power": 2 s,
  - threshold "Apdex": 2 s,
  - interval between measurements: 3 s,
  - jitter in ms (0, 50, 100, 150, 200, 250, 300).

- Scenario 2
  - number of measurements: 30,
  - threshold "Power": 5 s,
  - threshold "Apdex": 5 s,
  - interval between measurements: 3 s,
  - packet loss in percent (1, 2, 3, 4, 6, 8, 10).

Among others, the following Internet pages were used in the measurement scenarios:

- https://de-de.facebook.com,
- http://www.spiegel.de,
- http://www.chip.de,
- http://www.bundeswehr.de/portal/a/bwde,
- http://www.elektronik-kompendium.de,
- https://www.google.de,
- http://www.amd.com/de/Pages/AMDHomePage.aspx,
- http://www.nvidia.de,
- http://www.fh-flensburg.de/fhfl/index.php.

Representative results from Scenario 1 (with http://www.google.de) are presented graphically in Figs. 7 to 9.

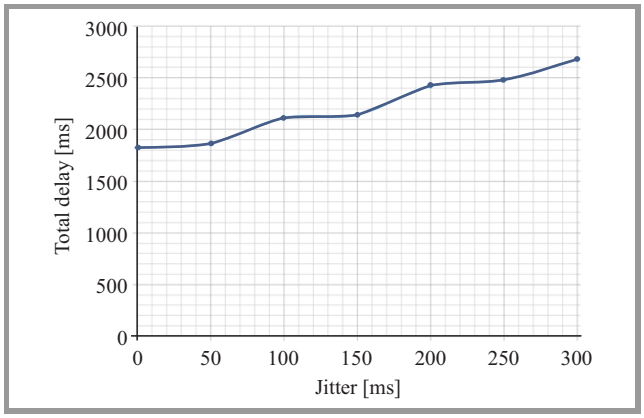


Fig. 7. Total delay as a function of jitter.

It is evident from Figs. 7 to 9 that total delay increases with increases in jitter (which was to be expected). Since the amount of data remains the same whenever the page is called up, the data download rate will decrease accordingly. A comparison between Apdex and Power shows that the two QoS curves develop exponentially but with differing gradients: at a jitter of 50 ms the Apdex curve starts to fall rapidly. The metric Power exhibits much more elasticity, with the subjective appraisal corresponded more to the objective result yielded by Power. While the measurements were being made, it also became evident that the ASP for the Web page www.google.de uses no outsourcing, which

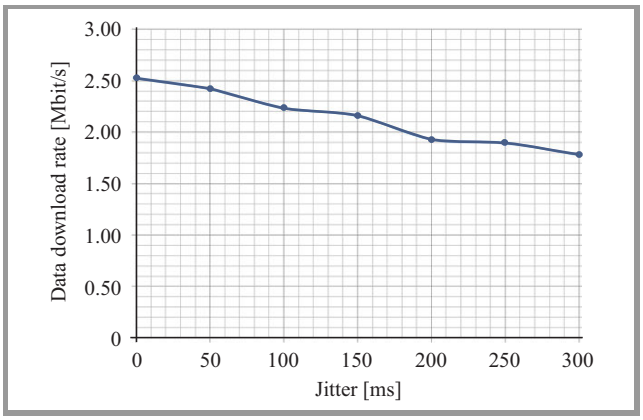


Fig. 8. Data download rate as a function of jitter.

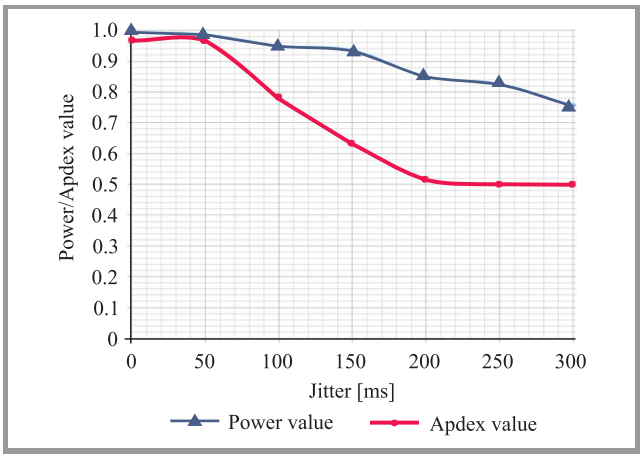


Fig. 9. Power and Apdex values as a function of jitter.

means that the call-up time of approx. 2 s is relatively short.

The representative results from Scenario 2 (with http://www.spiegel.de) are presented graphically in Figs. 10 to 12.

From Figs. 10–12 it is evident that even in a loss-free environment the times taken for a page to open were above the threshold set at 5 s, which resulted in values of less than 1. The time taken is so excessive because the ASP for the Website www.spiegel.de uses outsourcing, which is known to take time. The Apdex values remain almost constant at approx. 0.55. The metric Power demonstrates favorable elasticity too. Whenever packet losses increase, total delay rises rapidly and the data download rate falls noticeably. The small peaks in the download rate at packet losses of 1% and 3% (see Fig. 11) are due to the fact that although statistically the packet loss is higher at 3%, more often than not it was acknowledge packets that were affected when packet loss was measured at approximately 1%, which initiated a resenting of the packets involved. Here again, the subjective evaluation corresponds rather to the objective result of Power.

The results obtained show that the goal of creating an innovative, efficient tool for evaluating the quality of the WWW

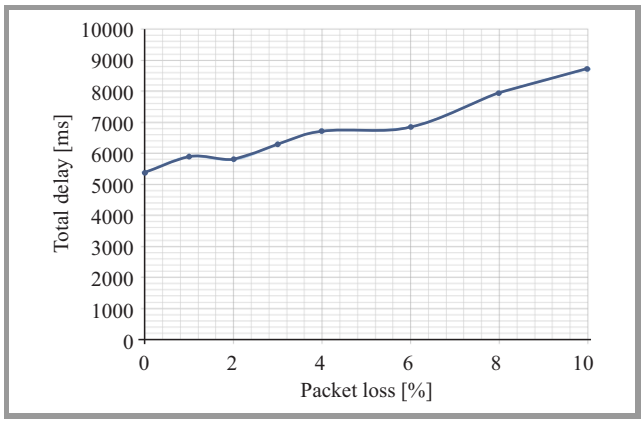


Fig. 10. Total delay as a function of packet loss.

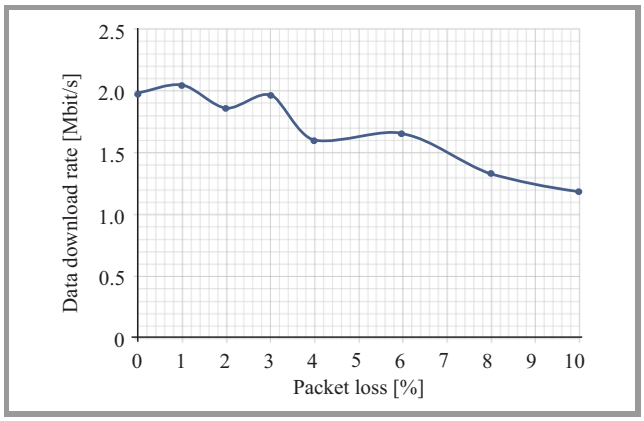


Fig. 11. Data download rate as a function of packet loss.

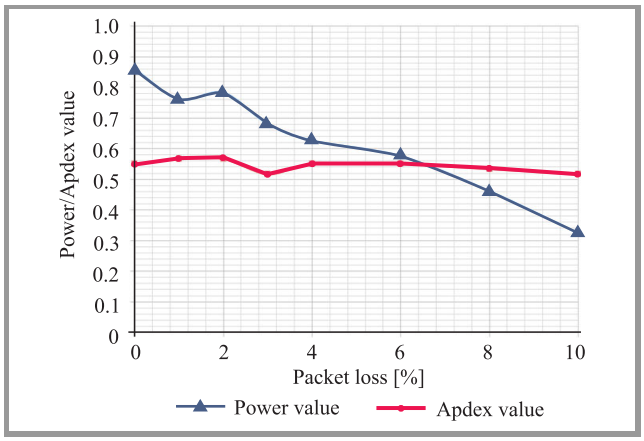


Fig. 12. Power and Apdex values as a function of packet loss.

service has been achieved here. The tool operates dependably and can be implemented practically anywhere. Consequently, the new tool is of immense practical importance for the communications industry.

5. Summary and Outlook

This paper has been devoted to the subject of QoS measurement within the World Wide Web (WWW). It has focussed on the creation of a tool for measuring quality of service in the WWW service. The tool has been given the name QoSCalc(WWW). It takes into account all of the most important parameters that can impair the WWW service, primarily the time taken for a page to open and the data download rate of a page. This two-phase approach to analysis relies on a combination of navigation phase and download phase. The prototype tool was tested in a realistic environment in a range of measurement scenarios using different WWW servers. The measurements have proved that the QoSCalc(WWW) measurement tool yields reliable and duplicable results in an authentic environment. With that, the tool has proved its practical capabilities and is ready to be used effectively in determining the quality of the WWW service.

In future work it would be very worthwhile anchoring the tool in a Testing Cloud Environment to make its use available to a wide sector of the public. Moreover, a range of browser implementations would not be amiss. They would help to identify differences in quality from one browser to the next. Accordingly, it would make sense to use other browser implementations in order to evaluate Web pages that cannot be evaluated at the time of writing due to errors in the internal routines of the browser implementations themselves. Beyond that, it would be worthwhile taking a look at individualized logging-in procedures. It would make sense, for instance, to investigate WWW pages onto which user profiles are loaded. Evaluating the process by which user profiles are loaded would bring us closer to understanding the phenomenon QoE. Another implementation in the end-user spectrum that is sorely needed is one to monitor WLAN adapters. Last but not least, the authors would like to point out that the present version of the tool attempts to resolve DNS entries according to IPv4. This could become a problem when IPv6 becomes the norm. Consequently, an implementation of the tool on IPv6 must follow. Work in this direction is already in the offing.

References

[1] Definition of NGN [Online]. Available: http://www.itu.int/rec/T-REC-Y.2001 (accesed: Nov. 2014).

[2] Directive 2009/140/EC of the European Parliament and of the Council of 25 November 2009 amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and associated facilities, and 2002/20/EC on the authorisation of electronic communications networks and services. Official Journal EU L.337/37.

[3] Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No. 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws. Official Journal EU L 337/11.

[4] "Estimating end-to-end performance in IP networks for data applications", Recommendation G.1030, ITU-T [Online]. Available: http://www.itu.int/rec/T-REC-G.1030/en (accessed: Nov. 2014).

[5] Application Performance Index - Apdex Technical Specification Version 1.1 [Online]. Available: http://www.apdex.org/specs.html (accessed: Nov. 2014).

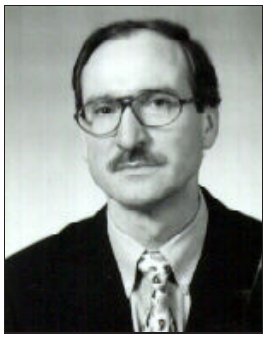
[6] T. Uhl, J. Klink, and P. Bardowski, " New metric for World Wide Web Service Quality", *J. Telecommun. Inform. Technol.*, no. 2, pp. 50–58, 2014.

[7] A. Raake, *Speech Quality of VoIP*. Chichester: Wiley, 2006.

[8] SharpPCap [Online]. Available: http://sharppcap.sourceforge.net/htmldocs/SharpPcap/RawCapture.html (accessed: Nov. 2014).

[9] Web browser object Internet Explorer [Online]. Available: http://msdn.microsoft.com/de-de/library/system.windows.controls.webbrowser%28v=vs.110%29.aspx (accessed: Nov. 2014).

[10] Wanulator [Online]. Available: http://wanulator.de (accessed: Nov. 2014).



Tadeus Uhl received his M.Sc. in Telecommunications from Academy of Technology and Agriculture in Bydgoszcz in 1975, Ph.D. from Gdańsk University of Technology in 1982 and D.Sc. from University at Dortmund (Germany) in 1990. Since 1992 he works as Professor at the Institute of Communications Technology, Flensburg University of Applied Sciences (Germany) and additionally since 2013 as Professor at the Institute of Transport Engineering, Maritime University of Szczecin. His main activities cover the following areas: traffic engineering, performance analysis of communications systems, measurement and evaluation of communication protocols, QoS and QoE by Triple Play Services, Ethernet and IP technology. He is author or co-author of three books and about 130 papers on the subjects LAN, WAN and NGN.
E-mail: t.uhl@am.szczecin.pl
Maritime University of Szczecin
Henryka Pobożnego st 11
70-507 Szczecin, Poland

Marcus Rompf received his B.Sc. in Computer Engineering from the Flensburg University of Applied Sciences (Germany) in 2014. Today he is student for Master's degree in Information Technology at Kiel University of Applied Sciences (Germany). His main activities cover the following areas: quality assurance for Triple Play Services and medical imaging solutions.
E-mail: marcus.rompf@fh-flensburg.de
Flensburg University of Applied Sciences
Kanzlei st 91-93
D 24943 Flensburg, Germany

Paper

Revising and Improving the ITU-T Recommendation P.912

Mikołaj Leszczuk

AGH University of Science and Technology, Department of Telecommunications, Krakow, Poland

Abstract—It was once thought that high Quality of Service (QoS) performance solves recurrent problems of low-quality multimedia services. Since then, solutions have been proposed to ensure a high level of Quality of Experience (QoE). In this paper, the author attempts to outline an understanding of an accurate meaning of multimedia services quality. Starting from QoS and passing through generalized QoE, the author focuses on subjective aspects and objective quality modeling and optimization of visual performance for Target Recognition Video (TRV) applications (such as video surveillance), to outline the ITU-T standardization path in this area. The revising the ITU-T Recommendation P.912 is proposed to reflect improved subjective test techniques developed since this Recommendation was approved. Also at least some existing errors of reasoning are predicted, which are likely to become evident for the industry in the next decade. Finally, the author invites all researchers working on topics related to TRV to join him in the process of improving P.912.

Keywords—CCTV, ITU-T, P.912, QoE, QoS, TRV.

1. Introduction

A decade ago, the telecommunications industry believed that high-performance Quality of Service (QoS) techniques resolve any recurrent problems of low-quality multimedia services. However, within a few years, it became clear that optimization of QoS parameters such as throughput, packet loss, delay, or jitter is not the best way of improving the quality experienced by users. The problem of low bandwidth can be compensated by more efficient codecs. The impact of packet loss is strongly dependent on their distribution, and the use of redundancy coding and transmission. For many applications, buffering multimedia data streams can alleviate major delays and jitter.

Since discovering that QoS is not a sufficient metric of network quality, most proposals have been suggesting that quality should be measured on the user level. This process was named Quality of Experience (QoE) [1], [2]. Such a measurement calls for special structures (frameworks) of quality of video sequences integrated assessment [3]. These structures are increasingly being filled with solutions that attempt to model the overall quality, operating at the intersection of QoS and QoE [4] or only in QoE. However, it has become obvious that such a general approach simply does not work for many visual applications such as target recognition (utility) applications (video surveillance, telemedicine, remote diagnostics, fire safety, backup cameras, games, etc.) [5], [6].

In fact, QoE – the way of perceiving multimedia services quality – depends on a number of objective and subjective contextual parameters [7]. Only a full understanding, usually only possible with strong area limitations of the QoE modeling application, makes it possible to obtain results consistent with the expectations of service users, and, consequently, to optimize quality [8]. Unfortunately, high numbers of contextual parameters mean this research question is still open.

2. Target Recognition Video

In many visual applications, the quality of the motion picture is not as important as the ability of the visual system to perform specific tasks for which it is created, given the processed video sequences. Such sequences are called Target Recognition Video (TRV). Regardless of the different ways in which the concept of TRV quality is understood, its verification is necessary to perform dedicated quality testing. The basic premise of these tests is to find TRV quality limits for which the task can be performed with the desired probability or accuracy.

Such tests are usually subjective psychophysical experiments with a group of subjects. Unfortunately, due to issue complexity and relatively poor understanding of human cognitive mechanisms, satisfactory results of TRV quality computer modeling have not yet been achieved beyond very limited application areas.

Given the use of TRV, qualitative tests do not focus on the subject's satisfaction with the video sequence quality, but instead they measure how the subject uses TRV to accomplish certain tasks. Purposes of this may include:

- video surveillance – recognition of vehicle license plate numbers,
- telemedicine/remote diagnostics – correct diagnosis,
- fire safety – fire detection,
- rear backup cameras – parking the car,
- games – spotting and correctly reacting to a virtual enemy.

The human factor is a significant influence, therefore it is necessary to ask questions on the procedures to be com-

plied with in order to make a subjective assessment of TRV quality. In particular, questions arise on:

- method of selecting the TRV source from which the test TRV (with degraded quality) arises,
- subjective testing methods and the general manner of conducting the psychophysical experiment
- method of selecting a subjects group in the psychophysical experiment, especially identification of any prior task knowledge,
- training subjects before the start of the experiment,
- conditions in which the test will be carried out,
- methods of statistical analysis and presentation of results.

3. Methods for Subjective Evaluation of TRV

Questions formulated in the previous section are addressed by Recommendation ITU-T P.912 "Subjective Video Quality Assessment Methods for Recognition Tasks", published in 2008 [9]. In addition, Recommendation P.912 organizes terminology related to subjective TRV testing, introducing appropriate definitions for the testing methods (psychophysical experiments).
Unfortunately, Recommendation P.912 is only the first step in the standardization of subjective TRV testing methods. In the author's opinion, based on available research results and observations conducted during numerous experiments with TRV, many claims of Recommendation P.912 are formulated at too high generality level. What's more, selected statements are not supported by research results and are significantly disputable. In this situation, a number of steps have been taken to introduce significant modifications (amendments) to the Recommendation. For this purpose, in order to formalize the procedures, the author has established collaboration with the Polish Ministry of Administration and Digitization, and received a formal nomination as a delegate of the Polish government. The procedure for submitting amendments commenced in 2014. The detailed scope of the proposed amendments to Recommendation P.912 is discussed in the following subsections.

3.1. Source Signal

Introduction: in Clause 5, Recommendation P.912 states:

Test sequences should follow the general principles stated in [10] and [11], which specify that scenes should be consistent with the transmission service under test, and should span the full range of spatial and temporal information. It is critical for the nature of these evaluations that the stimuli used actually reflect the true operational parameters of the conditions under which the video material is collected, and cover the entire range of scenarios possible for the application area that one is identifying. Unlike other subjective assessment methods developed for quality evaluations, this method is directed at the usefulness of the video material to complete a task and not the quality of the video itself.

Unfortunately, in certain cases, data availability is very limited. Let us consider the impact of studying the quality of still images on the accuracy of X-ray diagnosis of bone fractures. It is clear that due to the low frequency of certain types of fractures, the availability of a database of corresponding images is very low.
Another example concerns research on the impact of CCTV recordings on the accuracy of license plate recognition [8]. For the purposes of this study, a special video database was created [12]. The recordings have been created using fixed CCTV cameras, recording cars entering the car park at the AGH University of Science and Technology in Krakow, Lesser Poland (Fig. 1). Again, it is clear that due to the abovementioned conditions of acquisition, recordings represent a particular CCTV camera, its specific location and direction, a specific distance from the object, and lighting conditions. What's more, since the recordings were made in Krakow, most of the license plates have the letter "K" (distinguishing the Lesser Poland province) in the first position on the plate and "R" (distinguishing the Krakow county) in the second position.

Fig. 1. Source signal.

As shown, contrary to Recommendation P.912, it is very difficult to ensure complete coverage of the potential applications of the recordings. Any record database expansion is laborious, time-consuming, or even impossible. This does not mean that the cited studies are useless. However, their applicability must be explicitly limited to the scope of the recordings database. Unfortunately, literature frequently includes attempts to extrapolate the applicability of test results (in particular among less experienced researchers), which the author believes may be due to the fact that issues in Recommendation P.912, which frequently include instructions to carry out tests, are not addressed explicitly.

Proposal: the author proposes the introduction of the following amendments to Clause 5 of Recommendation P.912:

Test sequences should follow the general principles stated in [10] and [11], which specify that scenes should be con-

sistent with the transmission service under test, and should span the full range of spatial and temporal information. It is critical for the nature of these evaluations that the stimuli used actually reflect the true operational parameters of the conditions under which the video material is collected. ***If the stimuli used cannot actually cover*** *the entire range of scenarios possible for the application area that one is identifying,* ***the application description needs to be explicitly limited. For example, the results should not be generalized****. Unlike other subjective assessment methods developed for quality evaluations, this method is directed at the usefulness of the video material to complete a task and not the quality of the video itself.*

3.2. Testing Methods and Experimental Design

For videos used to perform a specific task, it may not be appropriate to rate video quality according to a subjective scale such as Absolute Category Rating (ACR) [10]. The goal of test methods for TRV is to assess the viewer ability to recognize the appropriate information in the video, regardless of his perceived quality of the viewing experience. To assess the quality level of TRV, methods that reduce subjective factors and measure the participant ability to perform a task are useful in that they avoid ambiguity and personal preference.

The TRV application is directly related to the user ability to recognize targets at increasing levels of detail. These levels are referred to as discrimination classes (DCs). When determining the DC for particular scenarios, one must consider that for a set distance from the camera to the object of interest, the DC directly correlates to decreasing video resolution of the target, and therefore the object is represented by fewer cycles per resolution degree. Fewer cycles per resolution degree also means that the object subtends less of the information content of the video, making the target identification more difficult.

Experimental methods should consist of responding to questions related to the content in the image or video. The parameter addressed by the question is the target to be recognized.

3.2.1. Multiple Choice Method

Introduction: in Clause 6.1, Recommendation P.912 states:

The number of choices offered to the viewer will depend on the number of alternative scenes being presented. "Unsure" may be one of the listed choices.

It should be noted that subjects tend to abuse the "unsure" response. This problem has been observed when applying a Comparison Category Rating (Table 1), as defined in Recommendation ITU-T P.800 [13], in which subjects tend to abuse the response "0" ("about the same"). A similar trend was observed independently in author's TRV studies.

Unfortunately, Recommendation P.912 is missing a clear warning against the prudent use of the "unsure" response (Recommendation P.912 even encourages its use).

Table 1
Comparison Category Rating (CCR)

3	Much better
2	Better
1	Slightly better
0	About the same
–1	Slightly worse
–2	Worse
–3	Much worse

Proposal: it is proposed that the entry in Recommendation P.912 should be amended as follows:

The number of choices offered to the viewer will depend on the number of alternative scenes being presented. ***The use of*** *"unsure"* ***as*** *one of the listed choices* ***is discouraged but allowed. The experimenter should be aware that individual subjects tend to overuse the "unsure" choice, leading to contamination of results. Consequently, special care must be taken when "unsure" is one of the listed choices****.*

3.2.2. Single Answer Method

Introduction: in Clause 6.2, Recommendation P.912 states:

If there is a non-ambiguous answer to an identification question, the single answer method may be used. This method is appropriate for alphanumeric character recognition scenarios. A viewer is asked what letter(s) or number(s) was present in a specific area of the video, and the answer can be evaluated as either correct or incorrect.

It should be noted that, contrary to Recommendation P.912, it is also possible to apply fuzzy logic [8]. For scenarios where the recognition result is an alphanumeric string, assistance may come from measuring differences between two strings using the Hamming distance (applicable only for strings of the same length) [14], or Hamming distance's generalization – the Levenshtein distance [15]. Using the experiment shown in Fig. 2 as an example, results containing no more than one error may be regarded as correct [8]. This is because even in the event of a plate being recognized incorrectly, by correlating it with a vehicle database containing the make and vehicle colour, the risk of the vehicle being identified incorrectly is substantially reduced.

Proposal: the author proposes that the description of the single choice method be expanded as follows:

If there is a non-ambiguous answer to an identification question, the single answer method may be used. This method is appropriate for alphanumeric character recognition scenarios. A viewer is asked what letter(s) or number(s) was present in a specific area of the video, and the answer can be evaluated as either correct or incorrect. ***Alternatively, fuzzy logic may be used (e.g. Hamming distance or Levenshtein distance), as shown in [8].***

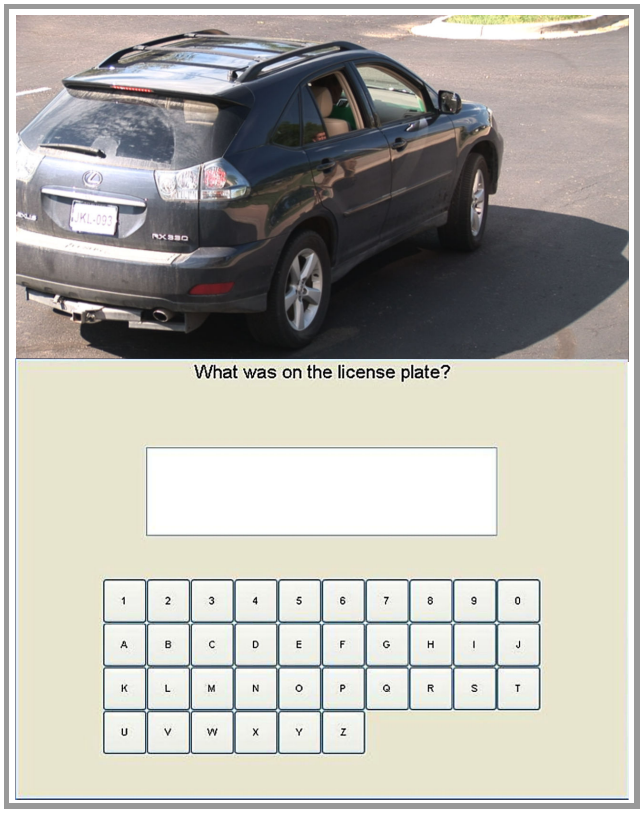


Fig. 2. Single answer method.

3.3. Subjects

Introduction: in Clause 7.3, Recommendation P.912 states:

Subjects who are experts in the application field of the target video recognition should be used. The number of subjects should follow the recommendations of [10].

In order to verify this finding, experiments testing subjects' ability to recognize certain objects (mobile phone, flashlight, gun, mug, radio, aluminum soda can, electric "Taser" stun gun) shown in video sequences were carried out. In the first experiment, the subjects were experts – law enforcement officers [16], [17]. When the experiment was repeated with non-experts, very similar results were obtained, as long as the non-experts were compensated for their time [18].

Proposal: the author proposes an entry introduction which allows the use of non-expert subjects providing they are motivated in an appropriate manner (such as being paid for their time). Naturally, this is only possible for certain areas of testing, since non-experts subjects cannot be used in tests associated with (for example) medical diagnostics.

Subjects who are experts in the application field of the TRV should be used. ***For certain areas of application testing, where neither specific experience nor expertise is required, non-expert subjects may also be used. Such non-experts must be motivated in an appropriate manner (e.g. being paid for their time). The validity of this approach is shown in [18].*** *The number of subjects should follow the recommendations of [10].*

4. Conclusions and Future Work

The discussion of statements contained in ITU-T Recommendation P.912 shows that some of the findings and observations require the certain provisions verification of the Recommendation. The author proposes to revise Recommendation P.912 to reflect improved subjective test techniques developed since this Recommendation was approved. Sufficient justification exists to support a new ITU-T work item, and contributions to this topic have been encouraged by ITU-T.

Ultimately, the amended recommendations should have a broader scope: to expand target testing methods, provide better instruction and training of subjects, improve conditions for testing, statistical analysis and reporting, and extend the applicability of techniques in the field of crowdsourcing for the subjective assessment of the quality of TRV. In cooperation with the US National Telecommunications and Information Administration (NTIA, originator of the Recommendation), there are also plans to expand the Recommendation to include metrics of Video Acuity, created at the NASA Vision Group [19]. The author would like to invite all researchers working on TRV-related topics to join him in the process of improving P.912.

Acknowledgements

This work is funded by the Polish National Centre for Research and Development, contract no. C2013/1-5/MITSU/2/2014 under the EUREKA international programme: Next Generation Multimedia Efficient, Scalable and Robust Delivery.

References

[1] E. Cerqueira, S. Zeadally, M. Leszczuk, M. Curado, and A. Mauthe, "Recent advances in multimedia networking," *Multim. Tools and Appl.*, vol. 54, no. 3, pp. 635–647, 2011.

[2] M. Grega, L. Janowski, M. Leszczuk, P. Romaniak, and Z. Papir, "Quality of experience evaluation for multimedia services", *Telecomm. Rev.*, vol. 81, no. 4, pp. 142–153, 2008.

[3] M. Mu, P. Romaniak, A. Mauthe, M. Leszczuk, L. Janowski, and E. Cerqueira, "Framework for the integrated video quality assessment", *Multim. Tools and Appl.*, vol. 61, no. 3, pp. 787–817, 2012.

[4] M. Leszczuk, L. Janowski, P. Romaniak, and Z. Papir, "Assessing quality of experience for high definition video streaming under diverse packet loss patterns", *Signal Proces.: Image Commun.*, vol. 28, no. 8, pp. 903–916, 2013.

[5] M. Leszczuk, I. Stange, and C. Ford, "Determining image quality requirements for recognition tasks in generalized public safety video applications: Definitions, testing, standardization, and current trends", in *Proc. 2011 IEEE Int. Symp. Broadband Multim. Sys. Broadcast. (BMSB),* , Nürnberg, Germany, 2011, pp. 1–5.

[6] S. Möller and A. Raake, Eds., *Quality of Experience: Advanced Concepts, Applications and Methods*. Cham: Springer, 2014.

[7] K. Brunnström *et al.*, "Qualinet White Paper on Definitions of Quality of Experience", Fifth Qualinet General Meeting, Novi Sad, Rep. of Serbia, Mar. 12, 2013.

[8] M. Leszczuk, "Optimising task-based video quality", *Multim. Tools and Appl.*, vol. 68, no. 1, pp. 41–58, 2014.

[9] "Subjective video quality assessment methods for recognition tasks", ITU-T P.912, 2008.

[10] "Subjective video quality assessment methods for multimedia applications", ITU-T P.910, 1999.

[11] "Digital Transport of Video Teleconferencing/Video Telephony Signals – Video Test Scenes for Subjective and Objective Performance Assessment", ANSI T1.801.01, 1995.

[12] M. Leszczuk and L. Janowski, "Database for video quality assessment in license plate recognition", in *Proc. Sig. Proces.: Algorithms, Architectures, Arrangements, and Applications SPA 2013*, Poznan, Poland, 2013, pp. 51–55.

[13] "Methods for subjective determination of transmission quality", ITU-T P.800, 1996.

[14] R. Hamming, "Error detecting and error correcting codes", *Bell System Tech. J.*, vol. 26, no. 2, pp. 147–160, 1950.

[15] V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals", *Soviet Physics Doklady*, vol. 10, p. 707, 1966.

[16] VQiPS, "Video quality tests for object recognition applications", Public Safety Communications DHS-TR-PSC-10-09, U.S. Department of Homeland Security's Office for Interoperability and Compatibility, June 2010.

[17] VQiPS, "Recorded-video quality tests for object recognition tasks", Public Safety Communications DHS-TR-PSC-11-01, U.S. Department of Homeland Security's Office for Interoperability and Compatibility, June 2011.

[18] M. Leszczuk, A. Kon, J. Dumke, and L. Janowski, "Redefining ITU-T p.912 recommendation requirements for subjects of quality assessments in recognition tasks", in *Multimedia Communications, Services and Security*, A. Dziech and A. Czyzewski, Eds., *Communications in Computer and Information Science*, vol. 287, pp. 188–199. Berlin-Heidelberg: Springer, 2012.

[19] A. Watson, "Video acuity: A metric to quantify the effective performance of video systems", in *Imaging Systems and Applications ISA 2011*, Toronto, Canada, 2011.

Mikołaj Leszczuk started his professional career in 1996 at Comarch Company as manager of the Multimedia Technology Department. Since 1999 has been employed at the AGH Department of Telecommunications. In 2000 he moved to Spain for a scholarship at the Universidad Carlos III de Madrid. After returning to Poland, he was employed at the Department of Telecommunications as a research and teaching assistant. In 2006 he successfully defended his Ph.D. as an assistant professor. His current research interests are focused on multimedia data analysis and processing systems, with particular emphasis on Quality of Experience. He has authored over 100 scientific publications. He has participated more than 20 major research projects. Between 2009 and 2014, he was the administrator of the major international INDECT research project, dealing with solutions for intelligent surveillance and automatic detection of suspicious behavior and violence in urban environments. He is a member of VQEG (Video Quality Experts Group), IEEE, and GAMA (Gateway to Archives of Media Art).
E-mail: leszczuk@agh.edu.pl
AGH University of Science and Technology
Department of Telecommunications
Mickiewicza st 30
30-059 Krakow, Poland

Paper

Quality Expectations of Mobile Subscribers

Przemysław Gilski and Jacek Stefański

Department of Radio Communication Systems and Networks, Gdańsk University of Technology, Gdańsk, Poland

Abstract—Mobile systems, by nature, have finite resources. Radio spectrum is limited, expensive and shared between many users and services. Mobile broadband networks must support multiple applications of voice, video and data on a single IP-based infrastructure. These converged services each have unique traffic holding and quality requirements. A positive user experience must be obtained through efficient partitioning of the available wireless network resources. The 3rd Generation Partnership Project (3GPP) has developed a comprehensive Quality of Service (QoS) parameter to address this problem. The regular control of service quality is critical for operators to ensure user Quality of Experience (QoE), establish new business models and monetize services. It enables operators to employ fair-use resource policies and maintain network performance during peak traffic times. Wireless mobile communication is tending towards an integrated system of Internet and telecommunication technologies, where mobile users move freely anytime and everywhere. They desire to communicate with any device using the best service available. In this paper QoS management issues in mobile communication are described. The authors present an insight into subscriber behavior and related factors that affect the QoE of mobile data services.

Keywords—cellular networks, mobile communication, quality of experience, quality of service, wireless communication.

1. Introduction

Due to the networks evolution from Circuit Switched (CS) to Packet Switched (PS) technologies, telecommunication services have experienced a huge increase in transmission capabilities, e.g., medium, bandwidth, throughput. It also helped new services to emerge, including Voice over IP (VoIP) telephony and multimedia streaming.

Currently, users require only an Internet Protocol (IP) access connection, either via a Wireless Local Area Network (WLAN) hotspot or a cellular connection. This IP communication trend requires an appropriate QoS, in order to fulfill the user expectations.

As the number of users in both telecommunication and Internet networks grows, it becomes clear that real-time services are becoming more difficult to implement due to erratic delay and packet loss.

Nowadays, mobile broadband networks carry multiple services that share radio access and core network resources. In addition to best-effort services, wireless networks must support delay-sensitive real-time services. Each service has different QoS requirements in terms of packet delay tolerance, acceptable packet loss rates and required minimum bit rates.

2. Quality of Service Background

QoS can be defined as a set of predefined technical specifications necessary to achieve the required service functionality. Each user specifies his requirements, so that the network can adjust its bandwidth, making use of different QoS schemes in order to satisfy the request. This can be an important factor when comparing services offered by different vendors or providers. When both price and feature are similar, quality becomes the key differentiator.

The degradation of QoS can be caused by a number of factors, including [1]:

- congestion (caused by traffic overflow – bottleneck effect),
- delays (caused by network equipment),
- distance or retransmission of lost packets,
- shared communication channels (collisions and large delays are common),
- limited bandwidth (poor capacity management).

2.1. QoS in Mobile Telecommunication

As mobile networks evolve to high-speed IP-based infrastructures, the wireless industry is ensuring high-quality services by developing QoS and policy-management techniques in addition to adding network capacity.

Mobile telecommunication is a type of communication used for transmitting voice or data over long distance. It consists of services such as: wireless telephony, satellite communication, WLAN and other 802.1x networks, IP-routed networks including the Internet, etc. However, the current global Internet is a best-effort service.

This service does not guarantee anything, even delivering a packet from one point to another within a single network. The destination node does not know the delivery speed or time. While delivering an e-mail message, delay is not a problem. But when it comes to real-time services like VoIP calls, if the delay becomes too large or too many packets are lost, the service quality becomes unacceptable [2], [3].

2.2. QoS Parameters

In order to keep track whether the contracted QoS are being met, the parameters must be monitored and resources should be reallocated in response to system anomalies. If a change of state happens and the resource management cannot make resource adjustments to compensate it, the application can either adapt to the new level of QoS or degrade to a reduced service level. The measurement of QoS is based on parameters including: delay, jitter, packet loss, throughput and many other, depending on the application and management scheme [4].

- Delay (latency) – a parameter related to communication. Since end points are most often distant, the transfer of information will consume time to reach the other side. Can be measured either one-way (from source to destination node) or round-trip (from source to destination and back to source node). The round-trip delay is used more frequently in the form of the ping command. It only sends a response back when it receives a packet without processing it. The final result is the minimum delay time possible for sending a packet from source to destination in the tested link.
- Jitter – a delay variation introduced by the transmission of multiple packets over the network. Can seriously affect the quality of audio-video streaming. In order to compensate it, all collected packets should be hold until the last (slowest) packet arrives on time and then rearranged to be played in the correct order. Jitter buffers are clearly visible when using audio-video streaming websites.
- Packet loss – occurs when one or more packets transported across the network fail to reach their destination. Some packets may fail to arrive when the buffer is already full. The loss of packets can be caused by other factors, e.g., signal degradation, high network load or defect in network elements. Wireless networks are more vulnerable to packet loss due to interference caused by other systems, multipath fading, multiple obstacles, etc.
- Packet error rate – the number of incorrectly received packets due to corrupted bits, often expressed as a percentage.
- Throughput – the amount of data that can be processed in a fixed time space, usually measured in bits per second. Throughput is a good way of measuring capacity of a communication link, regardless of connection type. However, it may not reflect the real user experience.
- Reliability – the availability of a connection, describes the ability of a system or component to function under stated conditions for a specified time period.

2.3. QoS Class Indicator

The Quality Class Indicator (QCI) specifies the treatment of IP packets received on a specific bearer. The *bearer* is a basic traffic separation element that enables differential treatment for traffic with different QoS requirements. It provides a logical transmission path between the User Equipment (UE) and Packet Data Network Gateway (PDN-GW). Packet forwarding of traffic traversing a bearer is handled by each functional node, e.g. eNodeB in Long Term Evolution (LTE).
The 3GPP has defined a series of standardized QCI types summarized in Table 1 [5].

Table 1
3GPP standardized QCI characteristics

QCI	Packet delay budget [ms]	Packet error loss rate	Exemplary service
1	100	10^{-2}	Conversational voice
2	150	10^{-3}	Conversational video (live streaming)
3	50	10^{-3}	Real-time gaming
4	300	10^{-6}	Non-conventional video (buffered streaming)
5	100	10^{-6}	IMS signaling
6	300	10^{-6}	Video (buffered streaming), TCP-based (e.g. www, e-mail, chat)
7	100	10^{-3}	Voice, video (live streaming), TCP-based (e.g. www, e-mail, chat)
8–9	300	10^{-6}	Video (buffered streaming), TCP-based (e.g. www, e-mail, chat)

where: QCI 1–4 are Guaranted Bit Rate (GBR) and QCI 5–9 Non-GBR resource type; IMS – IP Multimedia Subsystem, TCP – Transmission Control Protocol.

For first deployment, the majority of operators will likely start with three basic service classes: voice, control signaling and best-effort data, whereas in the future premium services such as high-quality video transmission.

3. Quality of Experience

Subscribers expect their mobile devices provide high-quality connectivity and performance at all time. Any interruption in data services is as critical as an interruption in voice. However, while voice services have a standardized measurement of quality called Mean Opinion Score (MOS), there is no equivalent for mobile data.
Mobile data services encompass a wide variety of content types and usage patterns, including e-mail, audio-video

streaming, application downloading or online gaming, all with different characteristics. Depending on the service being used, mobile subscribers have varying quality expectations for mobile data performance and usability.
When subscribers consume content, their QoE is not determined strictly by the speed achieved via wireless technologies. They make subjective assessments based on a combination of factors as: speed, smoothness, latency. Operators know, the better the experience, the longer and more frequently subscribers will consume content.

3.1. Defining Subscriber QoE

Mobile operators do not have unlimited technical resources and capital. The radio spectrum is finite and even if operators increase capacity, bandwidth-hungry applications such as Peer-to-Peer (P2P) services and video streaming will eventually consume any excess capacity. Table 2 demonstrates how subscriber QoE expectation varies by service type [6].

Table 2
Comparison of QoE expectations and performance requirements by service type

Service	QoE expectation	Performance attributes
Internet	Low (best-effort)	• Variable bandwidth consumption, • Latency and loss tolerant
Business services	High (critical data)	• High bandwidth consumption, • Highly sensitive to latency, • High security
P2P	Low (best-effort)	• Very high bandwidth consumption, • Latency and loss tolerant
Voice	High (low latency and jitter)	• Low bandwidth (21–320 kb/s per call), • One-way latency (< 150 ms), • One-way jitter (< 30 ms)
Video	High (low jitter and packet loss)	• Very high bandwidth consumption, • Very sensitive to packet loss
Interactive gaming	High (low packet loss)	• Variable bandwidth consumption, • One-way latency (< 150 ms), • One way jitter (< 30 ms)

There is a significant distinction between real-time services such as video conversation or voice and best-effort services like Internet browsing. Real-time services must reserve a minimum amount of guaranteed bandwidth and are more sensitive to packet loss and latency. Subscriber QoE is based on a number of factors such as:

– mobile application responsiveness,

– time required to download a Web page,

– stalling in a video,

– video content resolution.

Figure 1 describes top Android and iOS applications [7], [8].

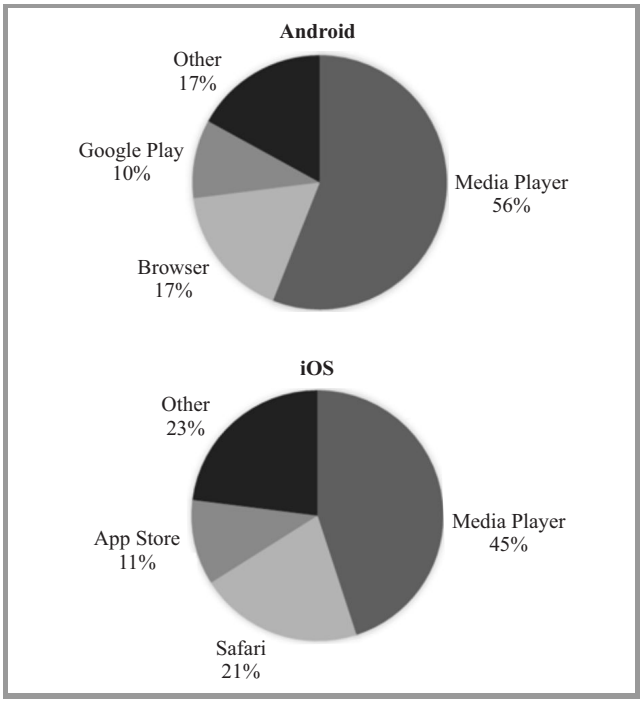


Fig. 1. Top Android and iOS applications.

According to presented data, about a half of mobile data is associated with Media Player software. The dedicated application stores and browsers account more than a quarter, whereas other application including Facebook, eBay and Instagram generate about 20% of mobile data volume.
In case of mobile Web page downloads across multiple mobile operators worldwide results show that over 50% of Web pages take more than 8 s to load and that 20% of Web pages take 20 s or more (Fig. 2) [9].
Depending on network conditions and the time of a day, mobile videos stall between 5–35% of the time. In some cases, stalling can lead some subscribers to abandon their sessions, causing frustration and loss of interest. It is visible, that conventional traffic management solutions do not work well in this case. Video is based on a variable bit rate. Its peak rates can exceed the shaped bandwidth of traditional traffic management solutions, leading to clips, stalling and eventually a poor experience. Figure 3 shows the change in global share of mobile video volume by format between 2010 and 2013 [8].
As shown, in 2010 90% of mobile video data was associated with the FLV format. Currently, the most popular video format is MP4, closely associated with smartphones, representing 67% of the global mobile video volume.

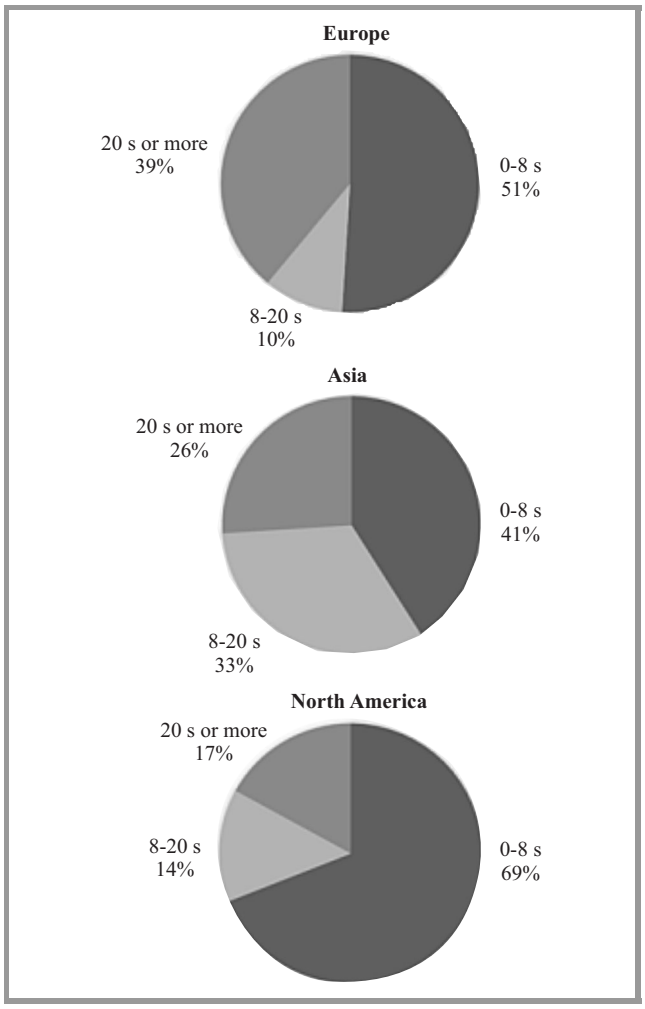


Fig. 2. Web page download times.

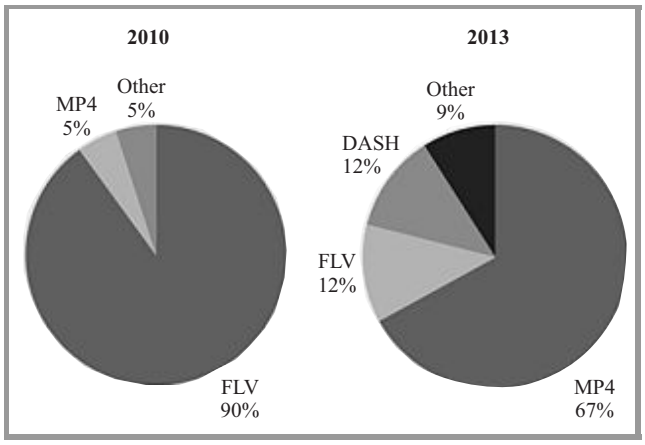


Fig. 3. Global share of mobile video volume by format.

Higher-resolution videos drive a disproportionate percentage of overall wireless network traffic, as shown in Fig. 4 [9].

Wireless networks that support this kind of videos deliver a better visual QoE to their subscribers. However, these multimedia must be effectively optimized to ensure that the overall subscriber QoE is not negative, considering screen size, resolution or connection speed.

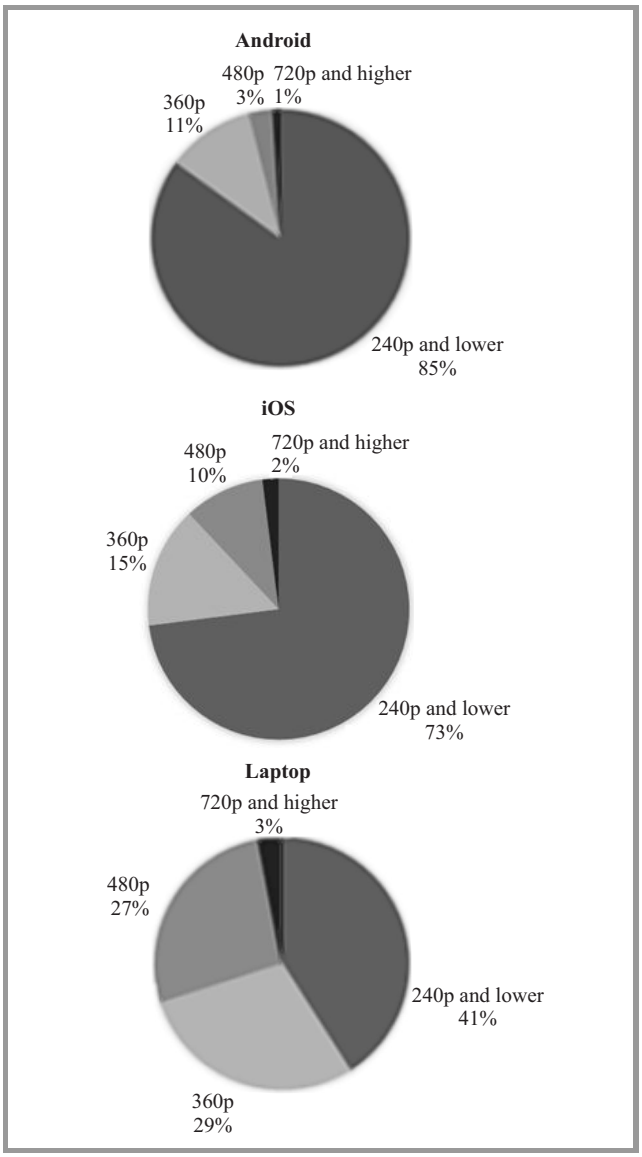


Fig. 4. Video request by resolution.

4. Summary

The widespread and availability of mobile smart devices will fuel the rapid growth in subscribers and sheer data volume. Operators world wide are racing to add new services and more powerful devices. They are making substantial investments to upgrade their networks capacity and performance.

If data continues to grow, operators will be forced to smarter manage the traffic. The economic realities and physical limitations of available spectrum prevent operators from simply adding more and more network capacity. Operators must plan today for future evolution of the network, which means working with vendors that have a solid roadmap for QoS and policy mechanisms in their products.

Quality plays a major role in wireless networks. Further traffic management and optimization technologies could allow network operators as well as service providers and vendors to improve subscriber QoS and QoE. Network efficiency could be optimized through application detection combined with adaptive traffic management in order to dynamically adjust to network conditions in real-time. As a result, it could help to boost mobile data usage, attract new customers, and raise satisfaction.

References

[1] M. Koivisto and A. Urbaczewski, "The relationship between quality of service perceived and delivered in mobile internet communication", *Inform. Syst. and e-Business Manag.*, vol. 2, pp. 309–323, 2004.

[2] J. A. Garcia-Macias, F. Rousseau, G. Berger-Sabbatel, L. Toumi, and A. Duda, "Quality of Service and mobility for the wireless internet", *Wirel. Netw.*, vol. 9, pp. 341–352, 2003.

[3] "Quality of Service (QoS) Concept and Architecture", ETSI TS 123 107 V11.0.0, 2012.

[4] F. Farid, S. Shahrestani, and C. Raun, "Quality of Service concerns in wireless and cellular networks", *Commun. of the IBIMA*, vol. 2013, pp. 1–5, 2013.

[5] 3GPP TS 23.203 V8.9.0, "Policy and Charging Control Architecture", 2010.

[6] "Quality of Service (QoS) and Policy Management in Mobile Data Networks", Ixia, Dec. 2013.

[7] "Mobile Analytics Report", Citrix, June 2013.

[8] "Mobile Analytics Report", Citrix, Nov. 2013.

[9] "Quality of Experience for Mobile Data Networks", Citrix, 2013.

Przemysław Gilski received his B.Sc. and M.Sc. degrees in Telecommunications Engineering from Gdańsk University of Technology (GUT), Poland, in 2012 and 2013 respectively. Currently he is a Ph.D. student at the Department of Radio Communication Systems and Networks (DRCSN), GUT. His research and development interests include digital video and audio broadcasting systems, software-defined radio technology, location services and radio navigation systems, as well as quality measurements in mobile networks.
E-mail: pgilski@eti.pg.gda.pl
Faculty of Electronics, Telecommunications and Informatics
Department of Radio Communication Systems and Networks
Gdańsk University of Technology
Gabriela Narutowicza st11/12
80-233 Gdańsk, Poland

Jacek Stefański received his M.Sc., Ph.D. and D.Sc. degrees in Telecommunications Engineering from Gdańsk University of Technology (GUT), Poland, in 1993, 2000 and 2012, respectively. From 1993 to 2000 he worked as an assistant professor at the Department of Radio Communication Systems and Networks (DRCSN), GUT. Since 2001 he has been working as an associate professor at the DRCSN. His research and development interests include analysis, simulation, design and measurements of cellular, wireless and trunked radio systems, techniques of digital modulation, channel coding, signal spreading, radio signal reception, measurement of radio wave propagation, field strength prediction, software radio design, location services, ad-hoc sensor networks, radio monitoring systems and radio navigation systems. He is the author and co-author of more than 200 papers. He is a member of the Electromagnetic Compatibility Section of the Electronics and Telecommunications Committee, Polish Academy of Science and the Institute of Electrical and Electronics Engineers organization.
E-mail: jstef@eti.pg.gda.pl
Faculty of Electronics, Telecommunications and Informatics
Department of Radio Communication Systems and Networks
Gdańsk University of Technology
Gabriela Narutowicza st11/12
80-233 Gdańsk, Poland

Paper

Silent Calls – Causes and Measurements

Kamil Baran, Paweł Cegłowski, and Sławomir Kula

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

Abstract—The quality of telephone services is very important from either operator or subscriber point of view. One of the negative phenomenon which affects quality of telephone services is lack of speech signal during a call. This situation occurs relatively frequently in mobile telephony, and is called silent call (SC). Lack of speech signal can occur only once or many times during the call, and degrade connection quality. In this paper, an analysis of this phenomenon is presented. The research base are the results of measurements mobile network one of operators in Trójmiasto a large urban area consisting of three cities: Gdańsk, Gdynia, and Sopot. To estimate impact of silent calls on speech quality, mean opinion score index was calculated using POLQA algorithm.

Keywords—key performance indicator, silent calls, speech signal analysis.

1. Introduction

The quality of telephone services depends on many factors such as the type of speech encoder, the type, and parameters of the telephone network, and the transport network performance. They can cause various types of distortions and even break or disconnect call. This paper focuses on the negative phenomenon of temporary or permanent lack of speech signal during a call – so called silent calls (SCs). According to the SwissQual, one of the leading companies involved in research quality phone calls, the silent call occurs when the called party receiving a silence all the time [1]. Here another definition is proposed. SC takes a place when at least one of the called party receiving a silence for a certain time. The problem of SCs is not new, but recently their number is constantly growing. Elimination SCs causes is very important to subscribers and operators because SC lowers the perceived quality of the service. Strong competition in the telecom market and easy change of service provider force operators to eliminate this problem. SCs occur both in mobile and fixed networks. However, in mobile networks due to many reasons SCs more often appear.

This article covers the study the SCs phenomenon only in mobile networks. It is possible to find many sources of the SCs in various network infrastructure parts. SC can be caused at the same time by more than one source. This makes it difficult to find and remove the correct source. The first place where SC can be generated is a subscriber terminal. The second one is radio transmission station. Another possible source is associated with handovers or encryption changes. These processes are taking place in Base Station Subsystem (BSS). Obviously, core networks used to transmission telephone signals between Mobile Switching Center (MSC) can also cause of SC, however the paper only concentrates on the radio interface influence on SCs.

The radio signal transmission conditions have significant impact on the call quality. They have often responsible for short-term signal degradations, but not always causes SC. To prevent the big information loss an interleaving and data redundancy are used. However, sometimes the interruption duration is so long that those methods are insufficient. In such cases, a call is always disconnected by a network, and then user hears the silence in the handset. The time required for dropping the call after detection absence of speech signal depends on the network settings. During a call, voice encoder may be changed because of radio signal transmission conditions degradation, but changing of the encoder is not always fast enough, therefore the handset has no signal while encoder changing. According to the definition adopted in this paper, the silent call is then observed.

Operators constantly upgrades their networks by introducing the new telecommunication systems as in example LTE. Usually areas covered by the new solutions are islands in older networks. Coexistence of different technologies requires multiple protocols to ensure the mobile subscriber an adequate level of service. Unfortunately, a quick introduction of new technical solutions does not always guarantee the calls quality and it can cause SCs.

To investigate sources and rate the negative impact silent calls on quality it was necessary to do many measurements using special procedure and tools.

This paper is organized as follows. In Section 2 the used tools are described. This is followed by the measurements scenario in the Section 3. In Section 4, the results are given and analyzed. Finally, the paper is concluded in Section 5.

2. Used Tools

To investigate SCs detail it was necessary to perform many measurements using so called Drive Test (DT). The DT was prepared by Systemics PAB [2]. It is a procedure for testing the quality of the cellular network and calls from the mobile subscribers perspective. All measured parameters were collected and analyzed using NQDI system [3]. This is the advanced tool that allows for calculation of Key Performance Indicators (KPIs).

To detect silence in the speech samples was used special software tool called Silan [4]. This program reads the sound files and detects silence periods. It is the pro-

gram made for Linux operation system. Its advantage is the ability to work from the command line. Using this tool and special Bash script [5] allow to create tables with information about the duration of speech and silence.

3. Measurements and Samples

The measurements scenario is shown in Fig. 1. The connections are initiated by the voice server and mobile terminal. According to this scheme there are 10 reference speech samples sent in two directions – five times by each of the parties. The first speech sample is always sent by the mobile terminal. In case of successful transmission of all test samples the call is disconnected. Received speech samples were recorded on the both sides.

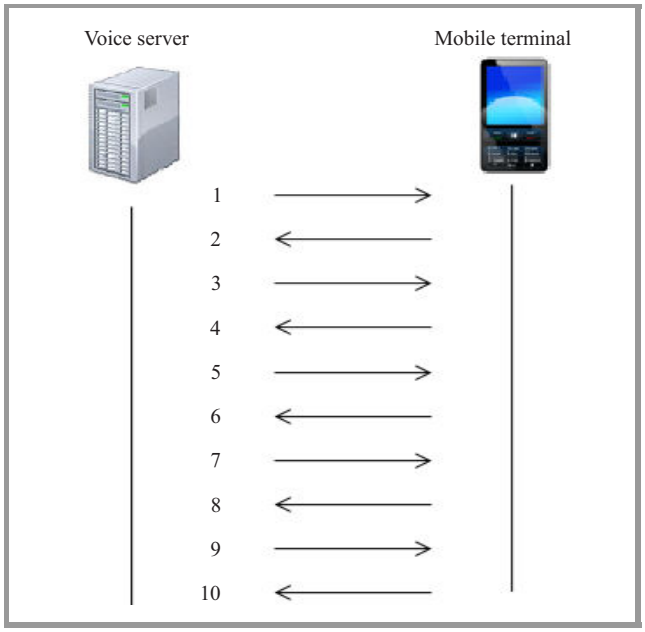


Fig. 1. Measurements scenario.

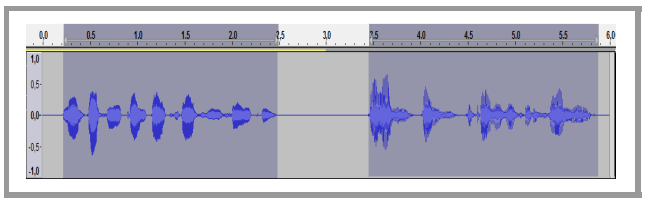


Fig. 2. The reference sample.

The speech sample reference is shown in Fig. 2.
It can be noticed that two fragments of speech signal in the reference sample were detected by Silan. The total duration of the speech signal in the reference sample is 4.697375 s. Duration of the speech signal in the received sample can be shorter if there were silence. In this case the speech signal contents in the each received sample versus the reference sample is then calculated. It is so called Sum Voice Length (SVL) parameter.

4. Results

During the test 51246 reference samples were sent. Among them there were 971 of silent calls. The number of received samples with extra silence versus SVL is presented in Fig. 3.

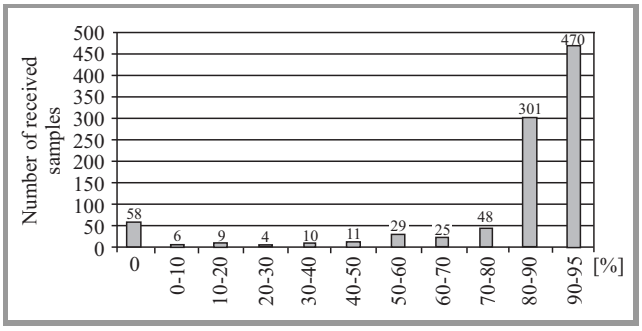


Fig. 3. Number of received samples versus SVL.

Figure 4 shows the dependence of the MOS parameter with respect to the percentage of speech duration in the received sample. MOS was estimated using POLQA algorithm [6]. Each dot in the Fig. 4 corresponds to a single sample. The minimum accepted MOS value is marked by the horizontal dashed line.

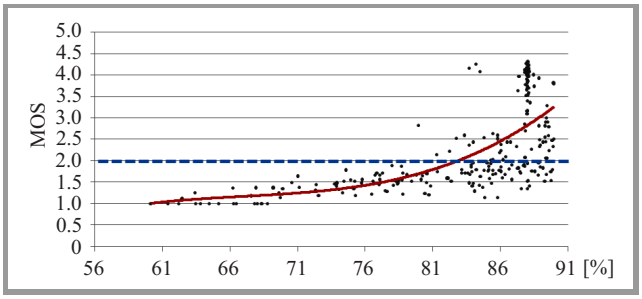


Fig. 4. MOS versus SLV.

Despite the big loss of the speech signal, the estimated speech quality is relatively good. This is probably a result of the mechanism Packet Loss Concealment (PLC) using by application [7].
Only in 58 cases the total voice absence was observed (SVL = 0%). These will be referred to the total SC (TSC). They represent about 0.11% of the sent samples. The

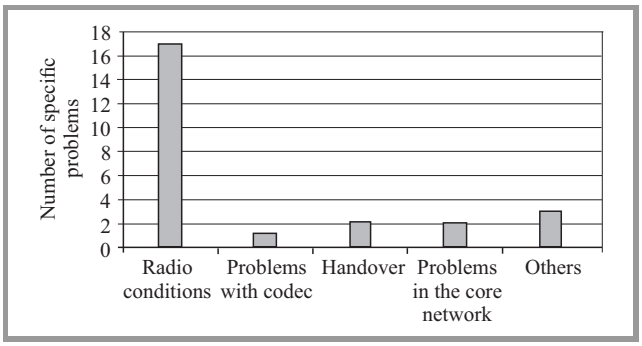


Fig. 5. Reasons of TSC.

33 from these were caused by errors in measuring equipment. The remaining 25 cases were detailed analyzed. The results are shown in Fig. 5. The most common reasons for TSC were wrong radio transmission conditions – not a low signal strength, but too high interferences level. It is obvious in large cities area. In two cases TSC were caused by handover process, and a core network was the reason of another two. In one case, TSC was caused by a problem with an encoder. Unfortunately, in the three cases, the reason was not found.

5. Conclusions

Based on the performed measurements and tests, it was found that:

- 1.89% received samples contained one or more periods of silence,
- only 0.11% of all samples was complete silence,
- the main causes of total silent calls diagnosed in this study were interferences in radio link,
- an unacceptable total duration of silence periods in the received sample was 17% (MOS assumed 2, and estimated using POLQA algorithm).

References

[1] M. Sauter, *From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadband*. London: Wiley, 2010.

[2] Measurement methodology [Online]. Available: http://www.syspab.eu/?en_benchmarking-of-mobile-operators,44

[3] NQDI, SwissQual's post-processing system [Online]. Available: http://www.swissqual.com/index.php/systems/nqdi.html

[4] Silan Software [Online]. Available: https://github.com/x42/silan

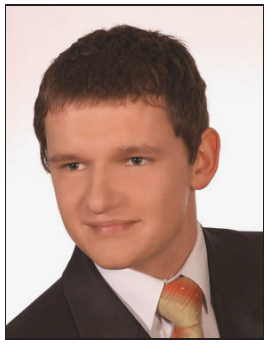
[5] K. Baran and P. Cegłowski, "Research on the phenomenon of silent call in networks mobile operators", Institute of Telecommunications, Warsaw University of Technology, Warsaw 2013.

[6] POLQA algorithm [Online]. Available: http://www.polqa.info/

[7] A. Janicki and B. Książek, "Packet loss concealment algorithm for VoIP transmission in unreliable networks" [Online]. Available: www.tele.pw.edu.pl/ajanicki/docs/aj-missi08.pdf

Kamil Baran received his B.Sc. in Electronics and Telecommunications from the Warsaw University of Technology in 2014. Currently he is a M.Sc. student at the Warsaw University of Technology. Since 2013 he works in Systemics PAB, where he is working at quality of mobile network.

E-mail: kakrzybaran@gmail.com
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland



Paweł Cegłowski received his B.Sc. in Electronics and Telecommunications from the Warsaw University of Technology in 2014. Currently he is a M.Sc. student at the Warsaw University of Technology. Since 2013 he works in Systemics PAB, where he is working at quality of mobile network.

E-mail: pawelc29@gmail.com
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Sławomir Kula received his M.Sc. and Ph.D. from the Faculty of Electronics and Information Technology of Warsaw University of Technology in 1977 and 1982, respectively.Two times he was a vice dean of the faculty, and now he is a deputy director for education of Institute of Telecommunications of Warsaw University of Technology. He is a chairman of Polish Chapter of IEEE Communications Society, and member of scientific committees of many conferences. He is also the guest editor of this issue of JTIT. He is an author of two books, and co-author of two others. One of them being "Transmission systems" (printed in Polish in 2005) which earned him the Ministry of Science and Higher Education Individual Prize. His scientific and research activities include transmission, access systems and networks, and quality of telecom services (QoS and QoE).
E-mail: skula@tele.pw.edu.pl
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Paper

Prospects and QoS Requirements in 5G Networks

Valery Tikhvinskiy and Grigory Bochechka

LLC Icominvest, Moscow, Russia

Abstract—In this article the requirements to some number of KPI that determine the quality of service (QoS) in 5G networks are formulated. The proposed QoS requirements are based on the analysis of functional requirements to 5G networks and traffic parameters for HD video and massive M2M services, which will be highly demanded in 2020. One of the 5G development paradigms is the network function virtualization (NFV) including cloud radio access and cloud core networks. The authors have proposed the concept of function blocks CQMF and CQCF to control and monitor QoS, which are implemented as part of the 5G network cloud infrastructure.

Keywords—5G, M2M, QoS, video services, virtualization.

1. Introduction

5G mobile communication technologies that are expected to appear on the market in 2020 should significantly improve customers' quality of service in the context of snowballing growth of data volume in mobile networks and the growth of wireless devices and variety of services provided [1]–[3]. It is expected that mobile communication networks built on the basis of 5G technologies will provide data transfer speed of more than 10 Gb/s. Previous 4G generation technologies (LTE/LTE Advanced) provide flexible quality of service management based on the division of data transfer characteristics into 9 classes. These classes cover both 4G quality principles – services provision without quality assurance (best effort or non-GBR) and guaranteed quality of service provision (GBR) [4].

Unfortunately, these LTE technological advances in the field of QoS management cover only part of the "end user-end user" (E2E) chain, in particular "5G-5G" and "4G-4G" intranetwork connections. The quality management system does not extend to the part of connections between 5G subscribers and other mobile 2G/3G/4G and fixed networks. Absence of possibility for coordinated and flexible quality management in fixed IP and mobile networks of previous generations will still be for a long period a brake on the new level of subscribers' service quality in 5G networks.

2. Services in 5G Networks

Forecasts of the leading specialists working in international 5G projects [5], [6] show that video services, such as HD and UHD video, with high quality resolution will have a dominant position among the services rendered in 5G networks. According to reports of leading 4G networks operators, video services dominate in the subscribers' traffic and will continue to dominate in 5G networks content.

For instance now the traffic volume of video services is estimated by different operators [5] from 66 to 75% of the total traffic in 4G network, including 33% for YouTube services and 34% for clear video as well as CCTV monitoring (video surveillance) in M2M networks. In addition, by 2020 the volume of mobile M2M connections will grow with CAGR index of 45% [7] up to 2.1 billion connections. Given the growing mass scale of M2M services in all industries, they will dominate over basic services (voice & data) in 4G and 5G networks.

5G European development strategy also aims to enable subscribers by 2025 to choose how to connect to TV broadcast: via 5G modem or antenna with DVB-T, so this will require appropriate quality management mechanisms. Therefore, the efforts of developers to improve the quality management mechanisms will focus on video and M2M services traffic, improvement of quality checking algorithms and creation of new quality assessment methods.

3. 5G Technological Image

Technological development of 5G networks will be aimed at the creation of ultra-dense networks (UDN) of wireless access with heterogeneous cells arrangement and radius of not more than 50 meters. 5G Networks will be based on new methods of modulation and transmission that will significantly increase the spectral efficiency compared with 4G networks and ensure data transfer speed of more than 10 Gb/s.

To provide such data transfer speed in 5G networks the use of broadband channels in the downlink (DL), as well as in the uplink (UL) with a continuous spectrum width of 500 to 1000 MHz will be required. This amount of spectrum is 25–50 times wider than the channels width used in 4G. Allocation of these bands for 5G channels is possible only at the upper boundary of the centimeter and in millimeter wave bands that will significantly reduce base stations coverage up to 50–100 m [8].

The increase in spectral efficiency of 5G networks can be achieved using non-orthogonal multiple access methods (NOMA) in RAN networks and using non-orthogonal signals (e.g., FTN-signals, F-OFDM-signals, etc.) [9]. Requirements to the cell's spectral efficiency in 5G networks

for different transmission channels are shown in Fig. 1 [6]. Comparison of these requirements with the same requirements to 4G networks shows the growth of spectral efficiency by 3–5 times.

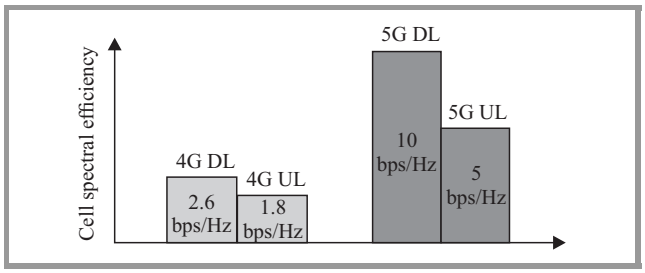


Fig. 1. Cell spectral efficiency in 5G networks.

5G network infrastructure will be based on the use of cloud technologies, both in radio access networks (Cloud RAN) with using Software Defined Radio (SDR) infrastructure and in core network (Cloud CN) with using Software Defined Network (SDN) infrastructure. Full virtualization of NFV network functions implemented in 5G infrastructure will take place. This virtualization of NFV network functions should cover the control and management of QoS, the service policy and prioritization of traffic.

New solutions for 5G networks will be the appearance of moving 5G nodes (base stations) and moving 5G backhaul that are dictated by the need to implement 5G during the construction of intelligent transport network.

Thanks to these solutions, international highways with cars moving at a speed exceeding 200 km/h will introduce moving 5G communication networks constructed on the basis of M2M applications and devices for Vehicle-to-Vehicle (V2V) scenario and ensuring the safety traffic and multimedia data exchange. The role of 5G base stations will carry out the 5G vehicle devices united in mesh network.

4. Traffic in 5G Networks

When forming requirements to QoS in 5G networks two key traffic models should be firstly considered: high-speed video flow "server-subscriber" and massive M2M.

Video transmission services will be an important stimulus to development and a rapidly growing segment of 5G networks traffic. In 2013 the volume of video services in the total traffic of 4G networks subscribers already exceeded 50%, and by 2019 it is forecasted to increase at least by 13 times [3]. Thus we can already observe the first wave of oncoming "tsunami" of subscribers' traffic in 4G networks. Monthly consumption of data transmission traffic in 4G networks has already reached 2.6 GB and monthly consumption of traffic in 5G networks will exceed 500 GB.

The growth of video services traffic volume will be associated with the implementation of various technologies of video services image quality from standard SD TV to UHD TV (8k), which in its turn requires a data transmission speed of up to 10 Gb/s in the network. Technological capabilities of mobile networks of various generations to broadcast video for various video image quality are shown in Fig. 2 [10]–[11]. Capability of video broadcasting depends on data transmission speed in the radio access network.

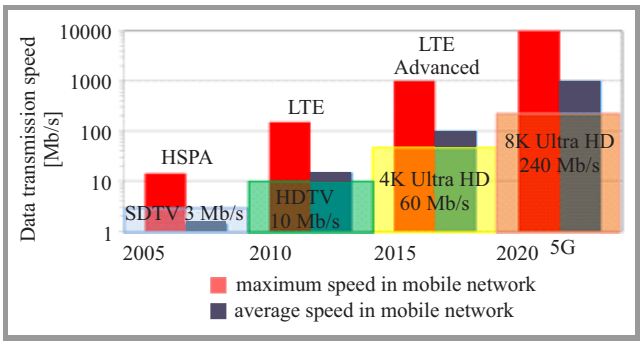


Fig. 2. Technological capabilities of video transfer for mobile networks of various generations.

According to forecasts shown in Fig. 3, in 2018 the number of M2M connections in the networks of mobile operators will exceed 1.5 billion [12], which is 5 times more than the current rate, and in 2022 mobile operators will have more than 2.6 billion M2M connections. At the same time, the share of M2M connections of the total number of connections in the mobile operators' networks will increase from the current 5% to 15% in 2018 and to 22% in 2022.

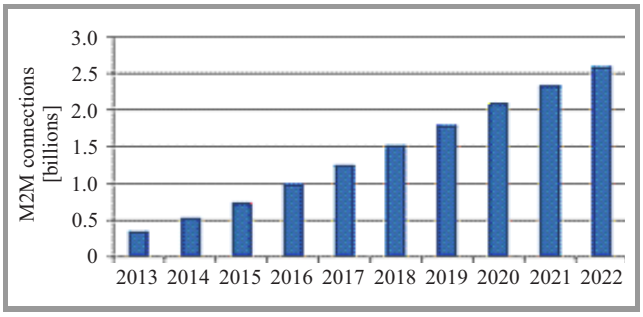


Fig. 3. Number of M2M connections in mobile networks.

Strategies of M2M operators are aimed at creating universal M2M platforms capable of operating in multiple vertical economic sectors. This will lead to the possibility to implement approaches, tools and processing methods for structured and unstructured Big Data derived from M2M networks.

According to ABI Research forecasts, the M2M Big Data and analytics industry will grow a robust 53.1% over the next 5 years from 1.9 billion USD in 2013 to 14.3 billion USD in 2018. This forecast includes revenue segmentation for the five components that together enable analytics to be used in M2M services: data integration, data storage, core analytics, data presentation, and associated professional services.

5. Quality Requirements in 5G Networks

During the evolution of QoS management mechanism in 3GPP (GSM/UMTS/LTE) networks there was a migration from QoS management at the user equipment level to the QoS management at the network level. This approach to QoS management will be maintained in 5G networks as well.

QoS management mechanisms in 5G networks should provide video and VoIP traffic prioritization towards Web-search traffic and other applications tolerant to quality. The service of streaming video transfer without buffering is very sensitive to network delay, so one of the most important parameters that determine QoS requirements is the total packet delay budget (PDB), which is formed on the RAN air interface, and is treated as the maximum packet delay with a confidence level of 98%.

Table 1 lists the requirements for delay in 3G/4G/5G networks formed in 3GPP [4] and METIS project [13]. These data demonstrate that with the increase in mobile network's generation the requirements for the lower boundary of the total data delay across the network decline. Also the analysis of the requirements for the overall 5G network delay revealed that given the accumulation effect the delay in 5G RAN network should be less than 1 ms.

Table 1
Requirements for delay in 3G/4G/5G networks

QoS terms	Packet delay budget [ms]		
	3G	4G	5G
Without quality assurance	Not determined	100–300	Not determined
With guaranteed quality	100–280	50–300	1

Comparison of requirements to delay in control and user planes for signaling traffic and user traffic respectively, presented in Fig. 4, show that requirements to 5G networks will be twice more rigid for traffic in the user plane and 10 times more rigid – in the subscriber traffic plane [6].

Another parameter is the proportion of packets lost due to errors when receiving data packets (IP Packet Error Rate).

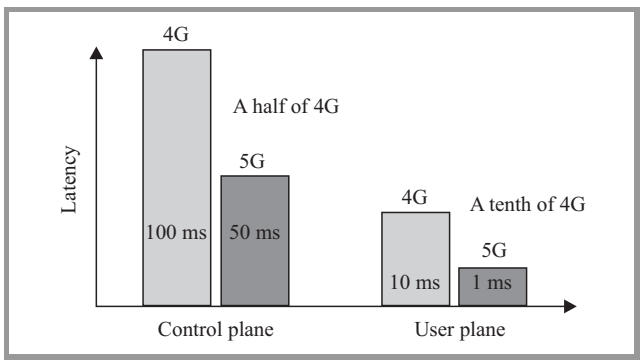


Fig. 4. Requirements to delay in control and user planes for 4G/5G networks.

Values for this parameter, that determines requirements for the largest number of IP packets lost for video broadcasting through 3G/4G/5G mobile networks, are shown in Table 2 [14].

Table 2
Requirements to the Packet Error Loss Rate for video broadcasting

QoS terms	Packet Error Loss Rate			
	SDTV	HDTV	4k UHD	8k UHD
Possibilities of mobile communication generation	3G/4G	4G	4G	5G
Video broadcasting with guaranteed quality	10^{-6}	10^{-7}	10^{-8}	10^{-9}

Table 3
Requirements to the Packet Error Loss Rate for M2M services

QoS terms	Packet Error Loss Rate		
	3G	4G	5G
Without guaranteed quality (non-GBR)	10^{-2}	10^{-3}	10^{-4}
With guaranteed quality (GBR)	10^{-2}	10^{-6}	10^{-7}

For M2M services the quality also will be determined by the proportion of packets lost when receiving in 3G/4G/5G networks. Given that service conditions of M2M subscriber devices will be determined for both cases: with a guaranteed quality of service and without guarantees, requirements to the share of lost packets differ by three orders. Requirements to the Packet Error Loss Rate for M2M services are shown in Table 3.

The development of NFV concept will lead to virtualization of quality management function that could be introduced in the form of two main functions: Cloud QoS management function (CQMF) and Cloud QoS control function (CQCF) shown in Fig. 5.

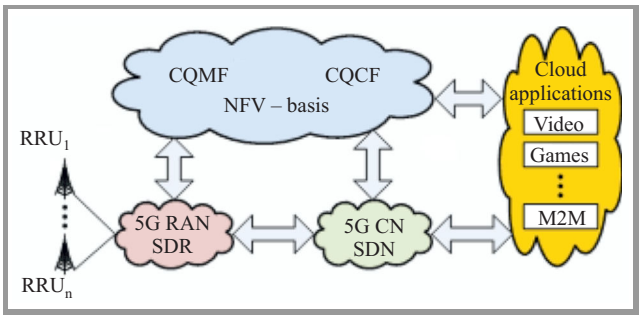


Fig. 5. Virtualization of control and management functions in 5G network.

CQCF function of QoS control provides real-time control of traffic flows in 5G network on the basis of QoS levels established during the connection. Basic QoS control

mechanisms include traffic profiling, planning and management of data flows.
CQMF function of QoS management provides QoS support in 5G network in accordance with SLA service contracts, as well as provides monitoring, maintenance, review and scaling of QoS.
Implementation of algorithms for traffic prioritization in 5G networks will be based on traffic classification procedures with a focus on video traffic priorities and M2M traffic. Traffic classification procedure should be done taking into consideration the adaptation possibility, as the traffic characteristics will dynamically change with the emergence of new applications, both in M2M area and in the field of video services.

6. Conclusion

The emergence of 5G networks in the market in 2020 will be focused on a significant improvement of characteristics of mobile networks, including quality of service. Given that the principles of QoS control will be maintained during the transition from 4G to 5G, main effort of 5G developers should be focused on the virtualization of network functions, responsible for the management and control of QoS in the network.
Another direction for development will be algorithms for traffic classification that will support market's changes including change of demand for services and needs of customers. Future mobile services will be grouped around video services and services based on the massive use of M2M devices in most industries and consumers' households.

References

[1] V. O. Tikhvinskiy and G. S. Bochechka, "Conceptual aspects of 5G construction", *Electrosvyaz*, no. 10, pp. 29–33, 2013.

[2] V. G. Skrinnikov, "Future image of 5G", *Electrosvyaz*, no. 10 pp. 34–37, 2013.

[3] V. O. Tikhvinskiy, "5G World Summit – 2014. Same course from 4G to 5G", - *Electrosvyaz*, no. 7, pp. 2–39, 2014.

[4] V. O. Tikhvinskiy, S. V. Terentiev, and V. P. Visochin, *LTE/LTE Advanced Mobile Communication Networks: 4G Technologies, Applications and Architecture*. Moscow: Media Publisher, 2014.

[5] Y. Weimin, "No-Edge LTE, Now and the Future", 5G World Summit 2014 [Online]. Available http://ws.lteconference.com/

[6] Y. Park, "5G Vision and Requirements", 5G Forum, Korea, Feb. 2014.

[7] V. O. Tikhvinskiy, G. S. Bochechka, and A. V. Minov, "LTE network monetization based on M2M services", *Electrosvyaz*, no. 6, pp. 12–17, 2014.

[8] G. Bochechka and V. Tikhvinskiy, "Spectrum occupation and perspectives millimeter band utilization for 5G networks", in *Proc. of ITU-T Conf. "Kaleydoscope 2014"*, St. Petersburg, Russia, 2014.

[9] G. Wunder, "5th generation non-orthogonal waveforms for asynchronous signalling", in *COST Meeting 2014*, Ferrara Italy, 2014.

[10] "Series H: Audiovisual and multimedia systems. Infrastructure of audiovisual services – Coding of moving video. High efficiency video coding". Recommendation H.265, ITU-T.

[11] E. Puigrefagut, "HDTV and beyond", in *Proc. ITU Regional Seminar on Transition to Digital Terrestrial Television Broadcasting and Digital Dividend*, Budapest, Hungary, 2012.

[12] "The Global M2M Market in 2013", Machina Research, London, Jan. 2013.

[13] Project METIS Deliverable D2.1 Requirements and general design principles for new air interface, 31.08.2013.

[14] ETSI Technical Specification. Digital Video Broadcasting (DVB); Transport of MPEG-2 TS Based DVB Services over IP Based Networks. ETSI TS 102 034 V1.4.1, 08-2009.

Valery O. Tikhvinskiy works as Deputy General Director of LLC Icominvest on innovation technologies – the finance investment company in telecommunication sector, and as Chairman of Information and Telecommunication Technologies branch of Russian Academy of Natural Sciences. He is a Doctor Economical Science (2003), and received the Ph.D. degree in Radio engineering (1988), the Government Prize laureate (2002). He is a Member of State Duma Committee Expert Council (since 2002), Editorial Board Member of Mobile Telecommunications (since 2002), and T-Com Journals (since 2007). He is a Professor of Moscow Technical University of Communications and Informatics (MTUCI, since 2001) and Visit-Professor of Tunisian Telecommunication Institute IsetCom (since 2005).
E-mail: v.tikhvinskiy@icominvest.ru
LLC Icominvest
Ostozhenka st 28
119034 Moscow, Russia

Grigory Bochechka is a Head of Innovation center department of LLC Icominvest and Chairman of WG14 Innovation Management of Telecommunications branch of Russian Academy of Natural Sciences Information and Telecommunication Technologies. He received his Ph.D. degree in specialty Systems, Networks and Telecommunication Devices.
E-mail: g.bochechka@icominvest.ru
LLC Icominvest
Ostozhenka st 28
119034 Moscow, Russia

Paper

Possibilities to Optimize QoS with Next SON Versions

Małgorzata Langer

Institute of Electronics, Lodz University of Technology, Lodz, Poland

Abstract—The paper discusses quality of service in LTE and LTE-A networks seen as a challenge that can be met with Self-Organizing Networks (SON) functionalities. The SON concepts have been included in the LTE (E-UTRAN) standards since the first release of the LTE technology. Self-optimization functionalities will monitor and analyze performance measurements, notifications, and self-test results and will automatically trigger re-configuration actions on the affected network nodes when necessary. The SON specifications have been built over the existing 3GPP network management architecture, the ultimate implementation of SON in 4G networks will bring many advantages. Successive SON procedures are waiting for their time and money to be implemented in 4G, though some essential issues for example of inter Radio Access Technology (RAT) interfaces must be overworked.

Keywords—4G, QoS, SON.

1. Introduction

New and emerging classes of mobile devices are raising large growth of wireless data usage by private and enterprise users. As a result, wireless service providers have to support a growing number of higher bandwidth data applications and services. On the other side their networks are becoming more complex and heterogeneous and the necessity to ensure quality user experience requires more complex Quality of Service (QoS) implementations. For a couple of years the main European body for new radio technology standards, 3GPP Group, has set the goal to support automated procedures in multi-vendor network environments to answer the request of free market with its whole complexity.

Since the first release of the LTE technology (3GPP Release 8) Self-Organizing Networks (SON) concepts have been included in the LTE (E-UTRAN) standards. One can point two categories for SON introduction reasons:

– repetitive processes' automating clearly reduces costs (saves time and effort),

– collecting measurements from many sources must be automated to provide accurate near real time data for algorithms to control fast, granular, i.e. per user, per application processes.

The SON specifications have been built over the existing 3GPP network management architecture, reusing much functionality that existed prior to Release 8. The focus of the Release 8 SON functionality was put on procedures relating to initial equipment installation and integration to support the commercial deployment of the first LTE networks, known as eNB self-configuration.

Release 8 of 3GPP, a basic one for nowadays implementations has several successors (Fig. 1) already and one proves that the SON capability is seen as a key component of the next and emerging networks.

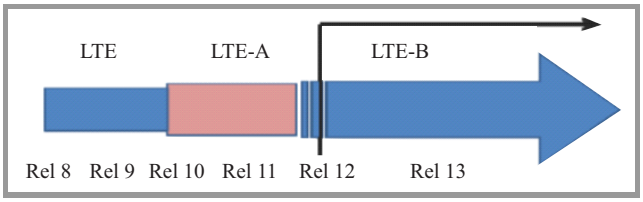


Fig. 1. LTE standardization.

Self-optimization functionalities will monitor and analyze performance measurements, notifications, and self-test results and will automatically trigger re-configuration actions on the affected network node(s) when necessary. This will significantly reduce manual interventions and replace them with automatically triggered re-optimizations or re-configurations thereby helping to reduce operating expenses. This paper aims to discuss the possibilities for QoS features resulting from present and further LTE implementations.

2. Review of Basic SON Procedures Covered by Successive 3GPP Releases

Table 1 [1], [2] shows SON procedures covered by 3GPP Releases since 8 to 12. Release no. 12 completeness was planned for 2014, and first parts of Release 13 (OAM aspects) should be ready soon.

The SON target is to maintain network quality and performance with minimum manual intervention from the operator. More automation allows to manage large networks more efficiently, which consist of thousands of base stations with hundreds of settings each. The role of SON is to enable efficient, and in some cases programmatic means of fine tuning cellular networks.

In 3GPP Release 8 standards, operators have specific SON requirements for simplifying eNodeB deployment and reducing operational cost. Approximately 80–85% of global providers focus on SON only in the 3G portion of their

Table 1
SON procedures in successive 3GPP releases

Release no.	Procedures	Remarks
8	Automatic inventory; automatic software download, automatic neighbor relation, automatic physical cell ID (PCI) assignment	"eNB self-configuration"
9	Handover optimization; RACH optimization; load balancing optimization; inter-cell interference coordination	Network optimization procedures
10	Coverage and capacity optimization; enhanced inter-cell interference coordination; cell outage detection and compensation; self-healing functions; minimization of drive testing; energy savings	SON functions for macro and micro networks overlaid on and interoperating with existing mobile networks
11	SON (UTRAN and LTE) management and coordination between different SON functions; inter-RAT energy saving management	SON Operation, Administration, Maintenance (OAM) aspects
12	High Rate Packet Data inter-RAT SON; enhancements of OAM aspects of distributed SON functions; multi-vendor Plug & Play eNB connection to the network	Next generation SON for UTRA and LTE

networks today [3]. This is because they want to first optimize what is stable and most of the network, while they work out other issues on 4G. Automatic Neighbor Relation (ANR) function, specified in the LTE context, automates the discovery of neighbor relations. It can help the operators to avoid the burden of manual neighbor cell relations management.

The ultimate implementation of SON in 4G networks will bring many advantages. For example, 4G has something called Reserved Quality (talked about on 3G, but not there, yet) as a means of managing QoS and Quality of Experience (QoE). This represents a benefit of SON on LTE in terms of optimizing network to support the QoS/QoE metrics. LTE architecture enables peer-to-peer (eNodeB to eNodeB) connections, which lowers latency and improves round-trip delay times. Successive SON procedures are waiting for their time and money to be implemented in 4G, and some essential issues for example of inter RAT interfaces must be overworked still.

3. Quality of Services in LTE

The QoS must be a particular concern in LTE as multiple applications may be running in a user equipment (UE) (for example at the same time: downloading an FTP file, browsing a Web page, chatting) at any time, each one having different quality of service requirements. Some services itself need better priority handling in the network, e.g., VoIP call, video conference. There are users being willing to pay more for high bandwidth and better network access, wanting to have better user experience on their 4G LTE devices. QoS defines priorities for certain services and customers.

In LTE Network QoS is applied to a set of bearers that are collectively called as EPS bearer. Bearer is a virtual concept and covers some network configuration to provide special treatment to various kinds of traffic. Keeping in mind the LTE architecture, one can point Radio Bearer, S1 Bearer and S5/S8 Bearer that cover QoS configuration in LTE (Fig. 2). As one must consider the end-to-end service, External Bearer is also a key issue here.

There is at-least one default bearer established when UE is attached to LTE network. When there is a need to provide QoS to a specific service, i.e., VoIP, video, etc. a dedicated bearer is established. There are two types of the Dedicated Bearer: Non-GBR and GBR.

Non-GBR Bearer does not provide a guaranteed bit rate and is associated with A-AMBR and UE-AMBR parameters (Application and User Aggregate Maximum Bit Rate). A-AMBR is the maximum allowed total non-GBR throughput to specific APN and is specified interdependently for uplink and downlink. UE-AMBR gives the maximum allowed total non-GBR throughput among all APN (Access Point Name) to a specific UE.

GBR Bearer provides a guaranteed bit rate and is associated with parameters like GBR and MBR; specified independently for uplink and downlink. GBR is the minimum guaranteed bit rate per EPS bearer, MBR – the maximum guaranteed bit rate per EPS bearer.

One should remember that the default bearer could only be non-GBR type.

There are four key parameters that make a QoS set for each bearer, describing: resource type (GBR or Non-GBR), priority (allocation and retention priority), packet delay, packet error or lost packets. Table 2 (GPP TS 23.203 V11.3.0) introduces parameters covered by nine Quality Class Indicators (QCIs) with example services.

Class 8 may be used for a dedicated premium bearer, e.g., associated with premium content for any subscriber/subscriber group or for the default bearer of a UE/PDN for premium subscribers. Class 9 is typically used for the default bearer of a UE/PDN for non-privileged subscribers.

The networks should be able to adapt their quality automatically in response to external factors or to special traffic patterns, i.e. SON procedures are the only solution for this challenge. Decisions are to be selected as the answer to the specified set of performance monitoring counters and indicators for every QCI [1]. They are, among others:

– number of successful sessions,

– number of dropped sessions,

– cell specific customer satisfaction rate,

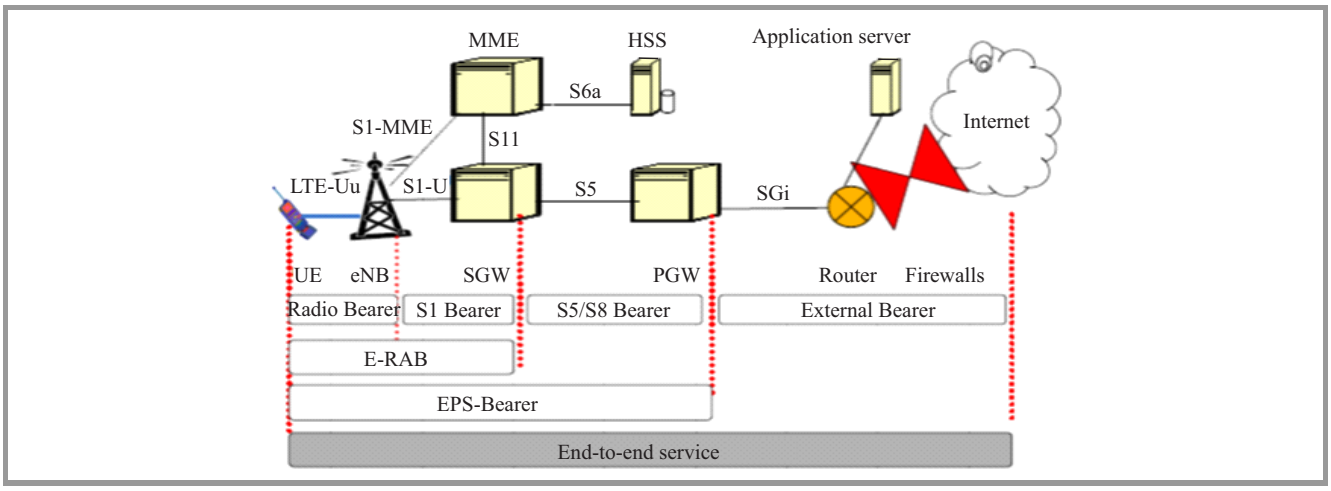


Fig. 2. LTE Bearer types.

Table 2
Standardized QCI with their parameters

QCI	Resource type	Prior-ity	Packet delay [ms]	Packet error/ loss rate	Example application
1	GBR	2	100	10^{-2}	Conversation voice
2	GBR	4	150	10^{-3}	Conversational live video
3	GBR	3	50	10^{-3}	Real-time gaming
4	GBR	5	300	10^{-6}	Non-conversational buffered video
5	NON-GBR	1	100	10^{-6}	IMS signaling
6	NON-GBR	6	300	10^{-6}	Video (buffered streaming), www, email, ftp
7	NON-GBR	7	100	10^{-3}	Interactive gaming voice, live video
8	NON-GBR	8	300	10^{-6}	Video (buffered streaming), www, email, ftp
9	NON-GBR	9	300	10^{-6}	As above

- maximum/average/minimum throughput,
- maximum/average/minimum round trip delay,
- packet loss,
- mean number of Radio Resource Control (RRC) connected users, and users with data to send,
- percentage of users per cell that do not achieve their required GBR, experience higher data rate or delay set as a threshold.

4. Main SON Functionalities

The main functionality of SON includes: self-configuration, self-optimization, self-healing. The SON functionalities' structure is shown in Fig. 3.

According to the location of optimization algorithms, SON can be divided into three classes: centralized SON, distributed SON, hybrid SON.
In centralized SON, all functions are located in OAM systems, so it is easy to deploy them, but since different vendors have their own OAM systems, there is a low support for optimization cases among different vendors. In distributed SON, optimization algorithms are executed in eNB and SON functionality resides in many locations. In hybrid SON part of the optimization algorithms are executed in the OAM system, while others are executed in eNB.

5. After 3GPP Release 8 SON Procedures

Table 1 introduces some SON features covered by after 8 3GPP Releases (up to Release 12). One should remember that SON standards, as all the 3GPP specifications are a work in progress. The functionality of SON is and will continue to expand through the subsequent releases. Some benefits to gain from implementing SON are presented below [2].
3GPP Release 9 provided SON functionality covering operational aspects of already commercial networks, especially: mobility robustness, hand-over optimization, RACH optimization, load balancing coordination, inter-cell interference coordination. Automatic Neighbor Relation (ANR) function, specified in the LTE context, automates the discovery of neighbor relations. ANR can help the operators to avoid the burden of manual neighbor cell relations management. S5 interface (Fig. 2) has a leading role here, as a user may frequently switch on and off the home node and an operator may not be able to access the home node physically.
Minimization of Drive Tests (MDT) for E-UTRAN and UTRAN is an important topic in 3GPP Release 10. With the help of standardized UTRAN MDT solutions, Capacity and Coverage Optimization (CCO) for UTRAN should

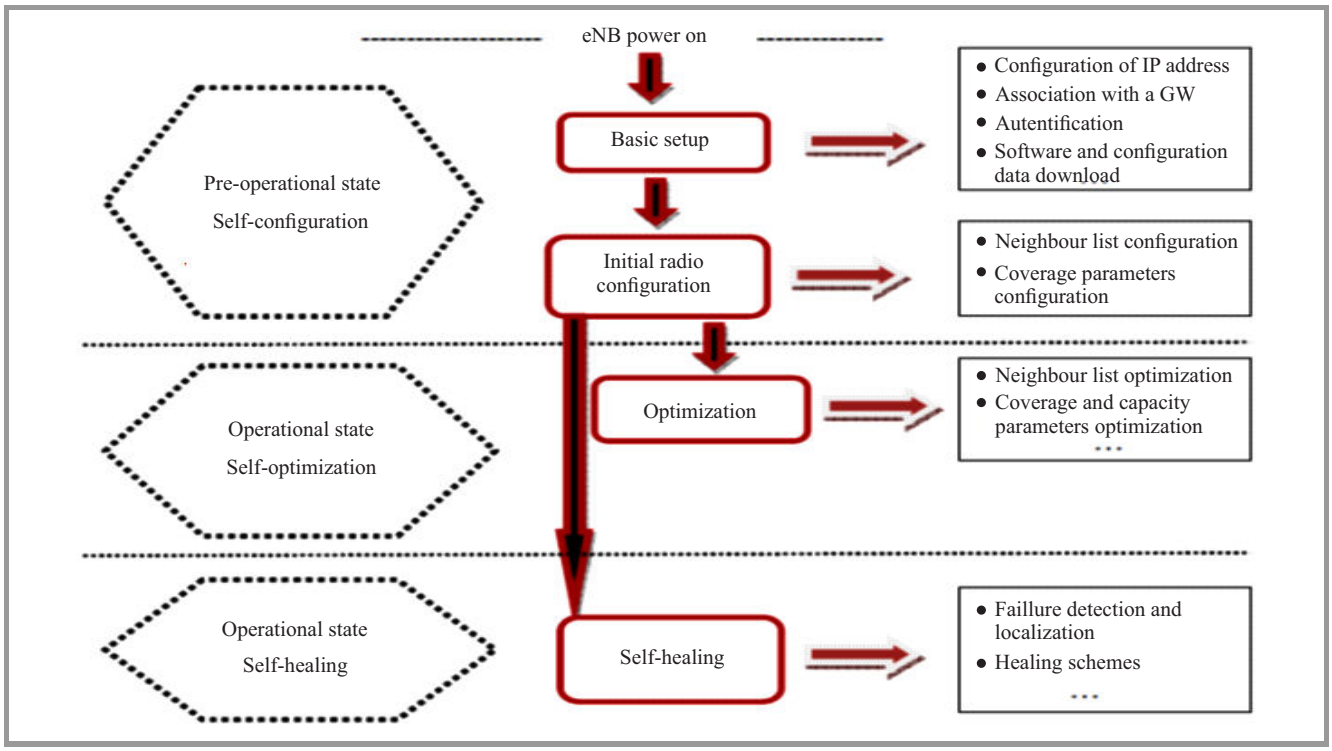


Fig. 3. The SON functionalities' structure.

also be considered in UTRAN SON activities. This Release covers also some self-healing functions and energy saving procedures in heterogeneous networks (macro-pico/femto scenarios). Any node can adjust its transmit output power to avoid interference with other nodes. Network elements can be switched into a stand-by mode and then woken up automatically, without affecting the customer experience such as dropped calls. Energy saving will not only enhance quality service experience but also will reduce operational costs related to power consumption.

In 3GPP Release 11 Mobility Robustness Optimization (MRO) is developed to identify for which UE type the failure has occurred. The release covers key aspects relating to network management, troubleshooting and optimization in heterogeneous networks. Self-testing and self-healing means that a system detects its problems and even solves them avoiding user impact and significantly reducing maintenance costs.

3GPP Release 12 covers SON procedures that can automate the network deployment based on active antennas what enables dynamic cell splitting/merging to handle changing load conditions. There are SON features for small cell functionality in the Release 12, too. Much effort is paid for multi-vendor Plug & Play eNB connection to the network to meet security problems in SON without losing its functionality.

6. Conclusions

SON can fix fundamental problems, such as poor coverage and/or dropped calls in an area and it can also be used for short-term, real-time issues (and then potentially be put back the way the network was in the first place). For example, the network may need optimization locally for a specific event such as a sporting event or live show. In all cases, SON is designed to support wireless carriers desire to provide a multitude of different services with high quality of experience for the end-user. With the SON implementation an operator can gain:

– no need for manual configuration at site, with security certificates without manual intervention,

– providing the node with planning data,

– configuring complex transport patterns with immanent security hidden within them,

– controlling plug and play processes.

All together means substantial savings with higher quality and reliability.

References

[1] M. Sater *et al.*, "Self-Optimizing Networks: The Benefits of SON in LTE", 4G Americas, July 2011.

[2] 3GPP work items on Self-Organizing Networks v0.1.3, June 2014 [Online]. Available: www.3GPP.org

[3] A. Basir, "Quality of Service (QoS) in LTE", Jan. 2013 [Online]. Available: http://4g-lte-world.blogspot.com/2013_01_01_archive.html

[4] M. Mojtahed, "Quality of Service over LTE Networks – Part 1", 2013 [Online]. Available: http://www.wirelessdesignmag.com/articles/2013/09/quality-service-over-lte-networks-part-1-3

Małgorzata Langer graduated in Analogues and Digital Circuits of Automatics, at the Lodz University of Technology and got her Ph.D. degree in electronics. After several years in international marketing she joined the TUL in 1988, first in the Institute of Materials Science and then the Institute of Electronics (since 2004 as Associate Director on Science). She works as Associate Professor in Telecommunications Team and delivers lectures and laboratories for students of various courses in Polish and English. Her main science and research interests are in non-deterministic object simulation, telecommunications, QoS, some aspects of new nonconventional technologies in microelectronics, reliability, theory of experiments, etc. She is the author and the co-author of over 60 publications, also quoted, and has been involved in several international, home, and university research and didactics programs.

E-mail: Malgorzata.langer@p.lodz.pl
Institute of Electronics
Lodz University of Technology
Wolczanska st 223
90-924 Lodz, Poland

Paper

Utilization of the Software-Defined Networking Approach in a Model of a 3DTV Service

Grzegorz Wilczewski

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

Abstract—In article a new concept concerning multimedia service modeling of stereoscopic motion pictures distribution is depicted. Presented conceptual model utilizes functionality approach supported by a Software-defined Networking (SDN) architecture. The core elements composing the proposed 3D television service are stated, depicting internal arrangement of a modern 3DTV service. Moreover, investigated examples of a functional utilization of the SDN approach in a 3DTV service model realization reveal key improvements towards greater flexibility and efficiency in heterogeneous users environment.

Keywords—3DTV service, SDN, Software-defined Networking, streaming architecture.

1. Introduction

Nowadays, intense visual data information exchange is observed. It has to be instantaneous, fast and reliable. In case of contemporary telco market, the focus is put on the telecommunication services delivering high definition quality audiovisual content. Deployed multimedia data processing systems support variety of transferring (distribution) modes. They include real time transmission of live video streams, content on demand features, i.e., Video on Demand (VoD) as far as originated from social networking, forms of prosumer like videos (prosumer, by definition is a service user that not only consumes the content but also produces video materials of its own). In order to keep the service away from an instantaneous market disapproval, from the point of view of the delivered multimedia stream, one has to guarantee appropriate Quality of Service or Experience model (QoS, QoE).

Elaborating on the domain of current telecommunication services, one shall acquire the analysis and forecasts included in the Cisco Visual Networking Index annual report [1]. Therein depicted (Internet Video section) are the researched trends and directions concerning global network congestion schemes. It is delivered, that in the category of the Web-based video applications and video related user activities, the overall, worldwide IP traffic will grow annually in an average of 29% (study was performed for the period of 2012–2017). What is more, extended forecasts provided by the considered report, depict the overall, global IP traffic share. It is indicated that the data transmitted within the multimedia related network activities will contribute, by the year 2017, to nearly 80% of total global IP traffic. Therefore, what follows is the value of the audiovisual data traffic (volume), which by means of Cisco VNI report is evaluated to be in the range of 53 thousands of petabytes (PB) per month. Thus, it is of the upmost importance to consider such a significant value of video related IP traffic, while introducing a new range of functionalities and enablers in refreshed, redesigned network architectures for telecommunication services.

Progressing towards multimedia services of a stereoscopic television (3DTV), investigated projections presented in [2], [3] reveal the extent of intensity of relevant activities in the considered domain of telecommunication services. By means of Compound Annual Growth Rate – CAGR coefficient, the Cisco Visual Networking Index: Forecast and Methodology, 2011–2016 report presents the overall behavior of selected multimedia services. In case of a 3D Video on Demand telecommunication product, the forecasted value of CAGR index reaches 105%. Thus, one can perceive the extent and the behavior of this dedicated service traffic growth, which in the considered case results in a global increase from 2 to 74 PB per month (specific data is included in Table 1).

Table 1
Global consumer managed IP traffic – 3D VoD service

Year	2011	2012	2013	2014	2015	2016	CAGR
Traffic [PB/month]	2	5	11	20	38	47	105%

Expanding the domain of topic related forecasts, as it is presented by DisplaySearch of NPD Group company in [3], the CAGR factor representing the increase of end user devices being 3D playback enabled, reaches 75%. Depicted amongst the Table 2 are the numerical values of projected 3D display sales in the considered time interval, between the years 2010 and 2018. Notable is the fact that nearly 200 million devices, i.e., mobile displays, 3DTV sets, etc. will be ready to present stereoscopic content to their users. Another study delivered by the NPD Group shows that in the 2012 the total market share (saturation) of 3D compatible displays over the total number of shipped screen devices was over 25%. Depicted by means of Table 3 are the saturation levels reflecting LCDs market composition.

Table 2
Worldwide forecast for 3D display sales

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018
3D Display Units [M]	3	11	20	40	71	88	113	143	196

Table 3
3D TV share of global LCD TV panel shipments

Year	2011			2012			
Quarter	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Share [%]	9.4	11.9	14.2	15.2	21.0	23.5	25.7

Aforementioned projections and forecasts deliver significant data, especially those shaping the overall behavior of 3DTV domain, necessary to include while developing and designing of the appropriate mechanisms for the telecommunication service of stereoscopic data transmission. The motivation for an appropriate modeling of a service is supported by the overwhelming numerical values representing forecasted IP traffic volumes required for an efficient service implementation, while the extent of the total number of 3D compatible devices uncovers the scope of widespread service possibilities.

2. 3DTV Service Characteristic

In this section the basics concerning 3DTV service architectural layout are covered, thus one can perceive the general characteristic of the generic service layout. Presented on the schematic diagram within the Fig. 1 are the fundamental elements contributing to the general 3DTV service architecture. Base blocks and relations identified among the considered figure indicate 5 blocks representing separate pillars of the service, listed in a following manner: customer of a service (End User), the Internet Service Provider (ISP), data processing center (Data Center), audiovisual content providers (Content Providers) and an extra block of External Service Provider. The most bottom entity of the Fig. 1 represents so called OTT providers (Over The Top) that expand the basic service offer with a value added functionalities or content.

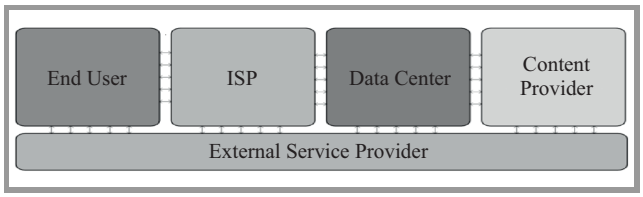


Fig. 1. Generic layout of a 3DTV service.

Nevertheless, the original shape and performance of the service is determined by the joint relation between End User and ISP pillars. Audiovisual content delivery based on that relationship determines the original and attainable levels of QoS or QoE. Moreover, performing monitoring (sensing, probing) over this pair of blocks shall lead towards selection of natural set of parameters describing the quality of the transmitted video content. However, one shall also incorporate such an approach whenever networking capabilities are limited and modification or transcoding of the multimedia data is necessary to balance the provisioned agreements.

Following the interactive path of content, all of the effective multimedia stream processing takes place in the Data Center entity, triggering the Content Providers and External Service Provider bodies to adjust. Furthermore, to keep the overall flexibility of the service, one shall consider supporting heterogeneity over the End User terminals. Not only is the video stream dependent upon the final presentation device but also on a type of a technique utilized to create the 3D perception. Amongst the popular approaches, one can enlist autostereoscopic and filtering methods, where in the latter case active (shutter) and passive (polarization) modes are distinguishable.

In order to complete the investigation of the 3DTV service characteristic, it is necessary to focus on the video stream features, as they are essential in the content driven multimedia service. Depicted by means of Table 4 are the crucial parameters describing stereoscopic 3DTV video streams.

Table 4
Fundamental video stream parameters utilized inside a 3DTV service

Parameter	Description
Image resolution	Depends on a subscribed version of a service (HD/SD) and utilized user's end device. It describes only a single stream resolution, not to confuse with the compound stereoscopic pair.
3D format	Defined as a composition of a pair of frames, might be of a dual settings: video frames positioned one aside another or one over another, respectively for left and right sensor of Human Visual System.
Stream format	Determines the attributes of utilized video stream coder along with specific profiling of selected standardization.
Bitrate	Reflects access network capabilities, may be adaptively selected depending on a subscribed version of a service.
Framerate	Depends on the utilized end device's capabilities. Exemplary values stated for a single video stream.

The common outcome of the analysis of parameters described in previously mentioned juxtaposition leads to the intermediate conclusion that the generic 3DTV video stream shall possess a multilevel flexibility. As it concerns the stage of modeling of a service, this criterion shall be understood as a video stream's property of scalability and adaptivity. What is more, projected model of a 3DTV service implies that the network and environmental conditions behave dynamically and change over time. In order

to maintain such a video stream flexibility one shall approach realization of this feature with separate flows multiplicity, carrying essential visual data sequences. Thus, a logically coherent video content asset is transported over the parallel set of data flows, what in fact enables extended (and also adjustable) QoS control. Moreover, a specific set of primitive parameters describing basic video units may be obtained, improving streaming capabilities with respect to the flexibility criterion. Figure 2 presents multisession transmission mode of a MVC stereoscopic video stream.

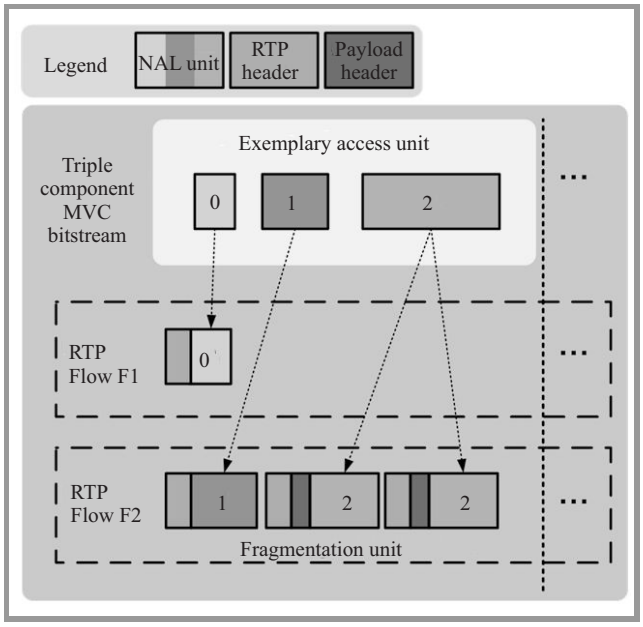


Fig. 2. MVC stereoscopic stream in multisession transmission mode.

Advanced support of a selected QoS mechanism may be realized by means of an initial video stream unit slicing (Access unit module in Fig. 2). A single video content is therefore represented with graded quality streams or may be prioritized accordingly to the proclaimed routine or policy preventing from the severe quality degradation. It also works with an implementation of mechanisms enabling switching between different video streams of a single (logical) multimedia asset with respect to the triggering event, i.e. user's transition between 3DTV service eligible terminals.
Depicted within the Fig. 2 is the scenario of a triple component unit Network Abstraction Layer (NAL) [4] which is orchestrated into a two, separate, manageable network flows. This basic feature is developed over the standard, classical 2D High Definition H.264 codec specification [5]. Worth noticing is the fact of native support of Real-time Transport Protocol (RTP) as an effective transport layer protocol.

3. Software-Defined Networking Architecture

Introducing the concept of a Software-defined Networking architecture [6] it is essential to present the modification that the networking plane has undergone. Conversely to the classical network approach, where multiplicity of layers (i.e., from physical towards application) is vertically merged within a single device, an SDN concept delivers the idea of specific separation of considered stack's planes. The crude significance of a Software-defined Networking architecture is an inherent division into a Control and Data (Forwarding) Planes. Since, control (management) functionalities over data flows and actual data transferring (payload traffic) are logically and physically (in a specific cases) separated. Elaborating on, contemporary network units implement both functionalities (Control and Forwarding) at once, where in case of the SDN approach, the management module of a device is lifted upwards into an unified abstraction plane called Control Plane. Moreover, what follows, is that the network management function may be performed from a single controller unit. Thus, a vast improvement with respect to the classical networking approach is clearly visible, namely efficient resource utilization and simplified reconfiguration of networks.

An exemplary situation depicting the advances in such an approach, may be a case of singular path prioritizing, whenever each device contributing to the considered path had to be personally adjusted, what implied intensive administrative workloads. Herein, by means of a single controller (so called hypervisor unit), at one action call, required task is accomplished. In addition, available solution enables high reliability of control functionalities. Redundancy and hierarchical organization of the Control Plane lays a firm basis for the SDN architecture to support multimedia services, while the data transmission layer may still utilize effective carriage mechanisms. Architectural layout presented on the scheme from Fig. 3 depicts conceptual arrangement

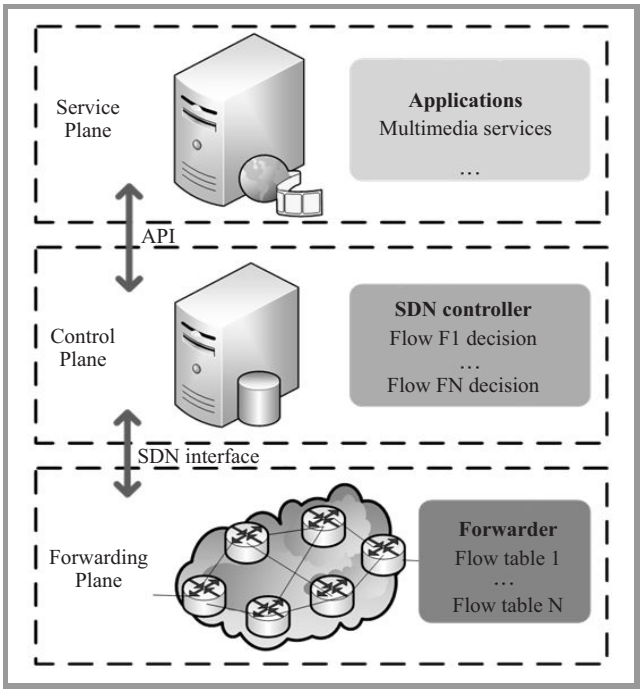


Fig. 3. Layered architecture of an SDN model.

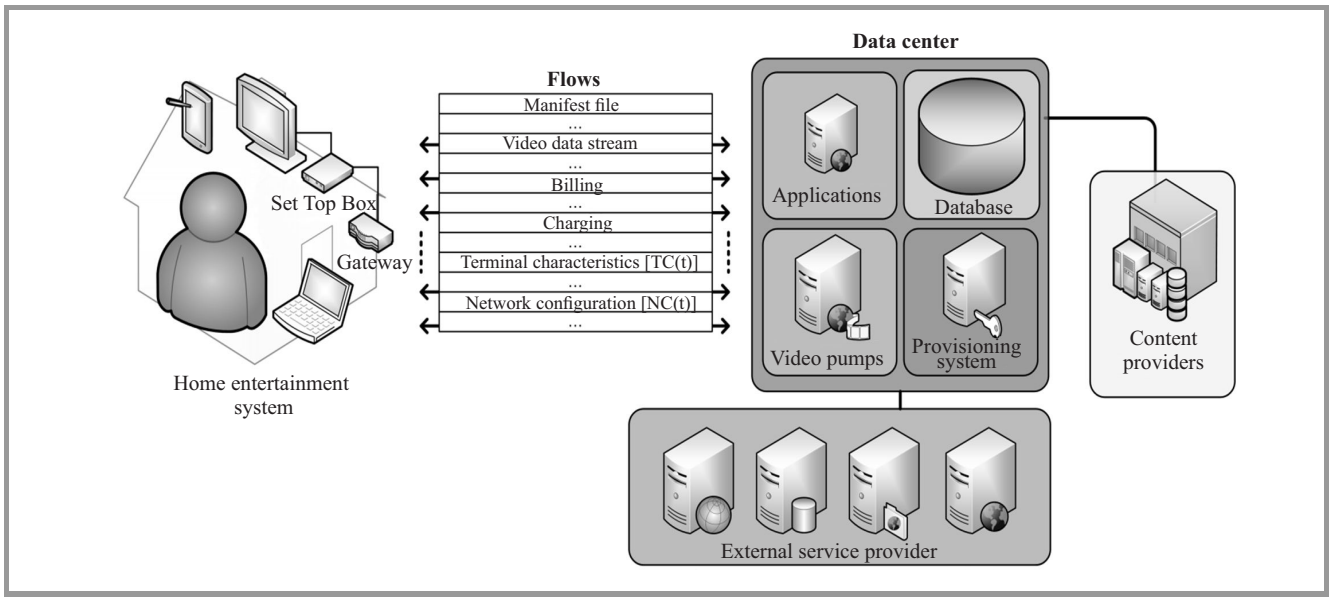


Fig. 4. Compound model of a 3DTV service utilizing an SDN architecture approach.

of layers within the specific Software-defined Networking approach. It is composed of three separate layers, functionally responsible for (from bottom to the top): data transfer (Forwarding Plane), management (Control Plane) and application support (Service Plane).

The interconnection in-between aforementioned planes is performed by two significant interfaces: API and SDN Interface. Service Plane and controlling layer has a joint Application Programming Interface (API) that supports various functionalities, both in a form of an inbound, i.e. management systems, and outbound, i.e. external services scopes. It is a high level programming interface enabling developers to create a new set of services for the deployed system. In case of the SDN Interface connecting Forwarding Plane with the Control Plane, there exist several solutions to the realization of the interface. One can specify open and vendor dependent approaches. In the first case, interface is based on a community contribution model, i.e., OpenFlow [7], whereas in the latter manner, utilization of a commercial (closed) framework is chosen, i.e., Cisco onePK SDN package [8]. The common point of those solutions is the mode of operation, that is based on a network state reporting routines and on on-demand managed flow tables.

4. Model of a 3DTV Service with Use of an SDN Architecture

In the final stage, presenting the model of a 3DTV service based upon an SDN approach, aforementioned requirements, concerning both stereoscopic content as well as the 3DTV service characteristic are combined with the enhancements delivered by the improved network architecture given by the SDN solution. Schematic diagram presenting the compound model of a 3DTV service utilizing described features is illustrated within Fig. 4.

Depicted solution utilizes the model cooperation of five, main entities, revealing their insights and core built. From the point of view of the effective service creation, the part interconnecting Home Entertainment System with the Data Center or else the telecommunication service provider section is the inevitable layer. Appropriate information flow management enables creation of a network aware and optimized service solution. Flow dependent traffic regimes shall improve flexibility feature by creation of a set of policies and other QoS enablers defined upon heterogeneity of end user devices, i.e. singular flow for common set of video stream parametric constraints. Moreover, an administratively simplified and orchestration friendly model of a service becomes while using the SDN architecture key features. Furthermore, in the proposed model of a 3DTV service, one can define rules and actions per single flow setting, thus enabling increased scalability in the domain of stereoscopic video streams. The advantage of assigning different forwarding mechanisms for specific flows may prevent from abundant video quality corruption, as one can exclude essential key frames from being frequently damaged, i.e., by means of real-time resource reallocation, routing paths, etc.

Nonetheless, another improvement may be observed in the domain of Service Level Agreements management. Increased precision and effective resource reservation (utilization) leads to the optimized SLA rules and enables efficient adjustment of those. An additional feature delivered by the utilization of the SDN approach when designing a multimedia streaming service is a layered structure with an open API interfaces (as it was previously indicated), enhancing OTT functionalities and finally improving overall service attractiveness by value added features. Finally in the discussed service layout, the core structure supports di-

rect end-to-end functionality and handling of a customer request from a single contact point, what should simplify the maintenance of the deployed service.
In order to complete the description of the functional model of a 3DTV service, let one perceive the system flows supporting previously discussed features:

- Manifest File (conveys video stream characteristic),
- Video Data Stream (multimedia data of an asset),
- Billing and Charging (basic IT functionalities),
- TC(t) flow (temporal characteristic of a terminal, defining primitives set essential for video stream preparation,
- NC(t) flow (temporal characteristic of a network resources, optimized for an extended QoS support).

5. Conclusions

The following paper covered selected insights out of the compound walkthrough of the process having presented the design of a 3DTV service with use of the SDN architecture approach. Indicated modeling steps followed the essentials and fundamentals of the efficient multimedia service creation.
Described core elements specification of the service, alongside the thorough 3DTV domain characteristic and broad selection of enhancements and advancements delivered by the Software-defined Networking scheme defined the scenario for a successful telecommunication service deployment. Finally, presented model of the 3DTV service supports increased reliability, flexibility of a video content and adaptive network resource allocation.

References

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2012–2017", Cisco Systems, Inc. [Online]. Available: http://bit.ly/P1eXkv (accessed: 14 April 2014).

[2] "Cisco Visual Networking Index: Forecast and Methodology, 2011–2016", Cisco Systems, Inc. [Online]. Available: http://bit.ly/1d4lq97 (accessed: 30 March 2014).

[3] "3D Display Technology and Market Forecast Report", DisplaySearch an NPD Group Company [Online]. Available: http://bit.ly/1eHaqJB (accessed: 24 April 2014).

[4] S. Wenger, M. M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP Payload Format for H.264 Video", IETF, RFC 3984 Internet Engineering Task Force (IETF) Audio Video Transport Group [Online]. Available: http://tools.ietf.org/html/rfc3984 (accessed: 30 March 2014).

[5] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*. Wiley, 2003.

[6] C. DeCusatis *et al.*, "Dynamic, software-defined service provider network infrastructure and cloud drivers for SDN adoption", in *Proc. IEEE Int. Conf. Commun. Worksh. ICC 2013*, Budapest, Hungary, 2013, pp. 235–239.

[7] "ONF Specifications – Open Flow", Open Networking Foundation [Online]. Available: http://bit.ly/1c1Mbrs (accessed: 19 April 2014).

[8] "Cisco Networking Software: One Network Environment Platform Kit, Cisco onePK At-A-Glance Guide", Cisco Systems, Inc. [Online]. Available: http://bit.ly/OvuOqG (accessed: 20 April 2014).

Grzegorz Wilczewski received his B.Sc. and M.Sc. degrees in Electrical and Computer Engineering from Warsaw University of Technology, Poland, in 2009 and 2011, respectively. He is currently a Ph.D. candidate at Warsaw University of Technology. His research interests include 3DTV service quality monitoring, 3D imagery and digital signal processing.
E-mail: g.wilczewski@tele.pw.edu.pl
Institute of Telecommunications
Faculty of Electronics
and Information Technology
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Paper

Subjective Assessment for Standard Television Sequences and Videotoms – H.264/AVC Video Coding Standard

Łukasz Trzcianowski

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

Abstract—This paper presents comparison of videotoms and standard television sequences in terms of image distortions and perceived subjective quality affected by H.264/AVC compression with changed bit rate. Results from initial tests, performed as laboratory exercise can be a reference to show scale of diversity in both level of degradation and Mean Opinion Score (MOS) evaluation. Results and comments included in this paper give overview on the codec influence on videotoms and can suggest approach for further tests and experiments.

Keywords—digital television, IPTV, Quality of Experience, Quality of Service, videotoms.

1. Introduction

The rapid growth of multimedia applications in the recent years has raised needs to provide services to clients in efficient way. In this case, not only network performance should be taken into account but also coding methods, especially those considered as standards. They are still under development in terms of improvements for compression efficiency while preserving same quality. Good example that comes to mind is H.265/HEVC technique created to fully support Ultra High Definition Television (UHDTV) and video resolutions up to 8192×4320 [1]. In fact, this standard is too new to be widely used in commercial systems for providing television services especially in Standard Definition Television (SDTV) and for such a video format it is reasonable to use H.264/AVC method [2], [3] in all the quality assessment tests.

Obviously coding process is only one factor that can affect video quality perceived by viewers. Important is how video signal is prepared, transmitted and received and all those parts should be always covered in video quality approximation. In telecommunication services provided over the IP network, usually parametric models are used in the planning process in order to meet Quality of Service (QoS) requirements, but for video applications it is difficult to create general recommendation that can be applied to all possible implementations and conditions. The most important issue is to find out what is the real relationship between parameters corresponded to coding, transmission, and receiving methods and user satisfaction from provided service. Current works on this topics keep focus on solutions associated with specific conditions considered in experiment process [4]–[7]. Basing on this and even considering existing ITU-T standards for Quality of Experience (QoE) [8] or P.NAMS, P.NBAMS [9], [10] that refer to packet and stream layer, it is difficult to talk about parametric models for IPTV services.

The author's main idea is to use videotoms as simple video sequences in order to create general approach to parametric model creation that could be moved to real TV materials. During tests focused mainly on network conditions [11], [12] and in further experiments it was found that H.264/AVC coding impact is not visible for simple videotoms sequences, whereas in standard television sequences the influence is significant. To verify that and to check what is the scale of degradation between two considered types of video the author decided to do firstly initial subjective tests for coding bit rate without including network parameters usage. There is a lot of papers that cover H.264/AVC coding methods [13]–[15] but not for simple sequences like videotoms.

2. Video Sequences

Video sequences used for this paper purposes can be divided into two groups: standard television video sequences and videotoms.

Difference between both groups is significant. The first group includes various video probes usually extracted from the TV programs source materials, whereas the second one contains sequences created based on human visual perception characteristic. Videotom's definition was introduced in another author's publication [11] and it refers to simple, well-defined and known to users video materials. In order to create such sequences both all image elements and dependencies between them should be taken into consideration. In this case, video pictures should keep constant form and organization. Each irrelevant or abstract information should be removed due to limitations in perceiving process. It is important to take into account also such elements like contrast, brightness, details level, dynamism, and diversity. Videotoms do not fit to television materials

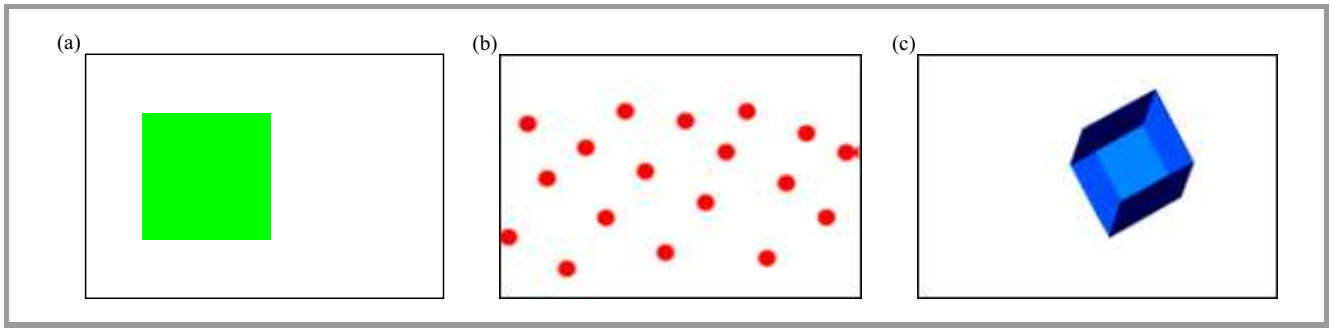


Fig. 1. Videotoms: (a) Green rectangle, (b) Red balls, (c) Blue cube.

Fig. 2. Television sequences: (a) Suzie, (b) Mr. Fins, (c) Cheerleaders.

Table 1
Videotoms description

Name	Duration [s]	FPS [fps]	Resolution [pixels]	Description
Green rectangle	10	25	720×576	Low motion sequence with small number of details (videotom)
Red balls				Average/high motion sequence with average number of details (videotom)
Blue cube				High motion sequence with small number of details (videotom)

Table 2
Television sequences description

Name	Duration [s]	FPS [fps]	Resolution [pixels]	Description
Suzie	10	25	720×576	Low motion sequence with small number of details
Mr. Fins				Average/high motion sequence with average number of details
Cheerleaders				High motion, detailed sequence with background

because of simplification due to very strong adaptation to human seeing process described in many studies [16], [17]. The benefit from their usage is related to easy analysis of the subjective assessment process because of the fact that distortions are easier to notice for users than in the standard television sequences. In studies and tests regarding their diversity, not only transmission should be considered, but also scale and behavior in terms of standard coding methods.

3. Test Conditions

3.1. Test Materials – Reference and Processed Probes

To verify scale of diversity in the degradations caused by H.264/AVC coding between standard television materials and videotoms six sequences were selected. First three were downloaded from "Consumer digital video library" [18] and other were created using Macromedia Flash Professional application in order to be adapted to the nature of human visual perception and according to Young-Helmholtz theory of trichromatic color vision in terms of receptors engagement. All the sequences are presented in Figs. 1 and 2, together with their parameters shown in the Tables 1 and 2.

Presented reference probes were used in the H.264/AVC coding process to create test materials. For this process author decided to use open implementation x.264 installed in the VirtualDub application. Settings setup was in line with ITU-T H.264 standard [2] and with Video Quality Experts Group (VQEG) recommendations – main profile, level 3.0. Coder bit rate was changed in the range of the following values: 3000, 2000, 1000, 500, 250, 100, and 64 kb/s using QP parameter. The subject of this paper focuses on H.264/AVC coding method and influence of network parameters were intentionally omitted.

3.2. Test Method

As a test method, Double Stimulus Impairment Scale (DSIS) was selected – approach and conditions are described in details in ITU-R recommendation BT.500-11 [19]. In this screening process two sequences are shown to assessor in pairs: first one is the reference, second one is impaired (after processing). Important is that viewer is informed about the order and after their playback, he is asked for the quality evaluation using impairment scale: 5 – imperceptible, 4 – perceptible, but not annoying, 3 – slightly annoying, 2 – annoying, 1 – very annoying. To proceed with tests MSU Perceptual Video Quality tool was used [20], where each single projection contains 3 parts: reference sequence, 3s grey area, and impaired sequence.

3.3. Testers

Tests were executed by students (15 persons) in similar age group during laboratory exercise. They were trained on test procedures, used methods, metrics, and tools. The objective for them was not only to assess the quality but also to provide information about observed distortions. In many cases, it is recommended to engage experts in subjective tests but it is expensive what creates difficulties especially in the initial tests. In this paper author assumes that trained non-expert testers can produce similar results if instructions and guidance is provided in a proper way (short training session with examples).

3.4. Results – MOS, Confidence Intervals

After tests execution MOS was computed for each particular test condition as an average of obtained results as:

$$\mathrm{MOS}_k = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathrm{MOS}_{ik}\,, \tag{1}$$

where: N – number of testers after outlier removal, MOS_{ik} – score assigned by tester i to test condition k, MOS_k score for test condition k.

To measure the estimate reliability based on a sample of population (15 persons) confidence intervals of estimated mean were calculated. Results show the relation between estimated mean values and entire mean values of the entire population. Due to small number of students and assuming 95% confidence level, intervals for the mean subjective scores were computed by using Student-T distribution as:

$$\delta = t_{(t-\frac{\alpha}{2})} \cdot \frac{S}{\sqrt{N-1}}\,, \tag{2}$$

where: δ – confidence, $t_{(1-\frac{\alpha}{2})}$ – t value associated with given significance level α for a two-tailed test, N–1 – degrees of freedom, where N is the number of observations in the sample, S – estimate standard deviation of the sample of observations.

4. Results and Conclusions

In this section test results are considered as a comparison two sets of mean MOS scores - one for standard sequences and the second one for videotoms. Figures 3 and 4 show, for each video content, mean MOS values and confidence intervals across changes in the H.264/AVC codec bitrate. Received subjective rates present what is the degradation level and impact on the quality for both types of used materials. MOS results for videotoms are comprised between 2.75 and 5, but for television sequences they span over quality levels entire range. Obviously extremely low coding bit rate negatively affect video quality in all tested sequences, but in case of videotoms even by the smallest bitrate values video pictures are still readable. It is worth to comment on confidence intervals that inform about results reliability. They are definitely more wide for television sequences and it is not possible to identify for which bitrate values the less precise estimates (greater level of variance)can be expected because this depends on particular video sequence. For videotoms, confidence intervals are relatively small, wider

only by extreme values of coding bitrate. It is also significant to indicate that for television sequences MOS values decrease faster, starting by 2000 kb/s. This behavior is in direct relation to TV materials where usually we can expect more details and dynamism and background layers in the video pictures. Standard H.264 mechanisms like motion compensation, entropy coding and inter/intra picture prediction are easier for videotoms and maybe it is reasonable to Basic Line profile instead of Main profile, but the objective for these tests is to have same conditions for both types of sequences.

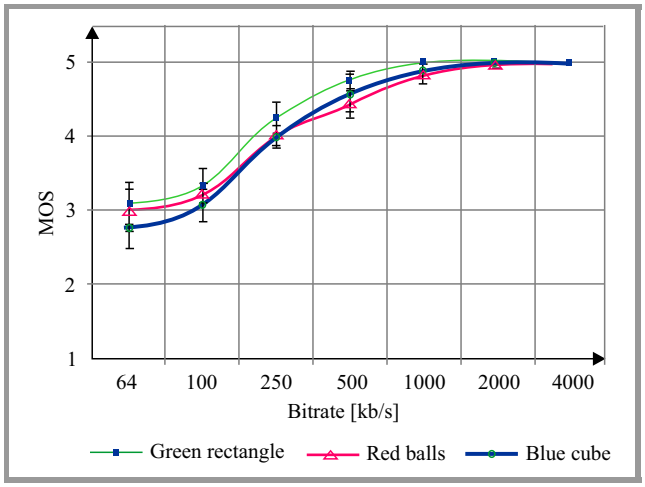


Fig. 3. Mean MOS score for videotoms.

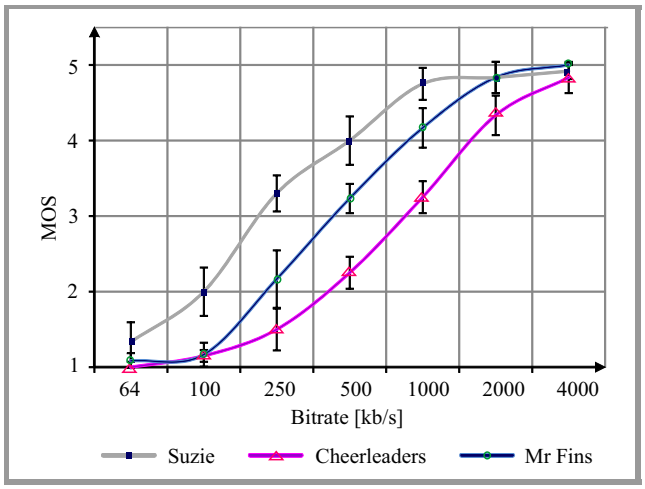


Fig. 4. Mean MOS score for standard sequences.

Regarding to the scale of diversity, by coding bitrate 250 kb/s the MOS value for "Suzie", "Mr. Fins" and "Cheerleaders" is included between 1.5 and 3.5 , whereas for videotoms it is still acceptable (above 4). Figure 5 illustrates what is the scale of difference in degradation level. Observed distortions show that in many cases impact of coding bitrate seems to be negligible for videotoms in comparison to effects in TV sequences. Apparently an error concealment mechanisms works for such simple sequences causing small number of visible degradations and in effect

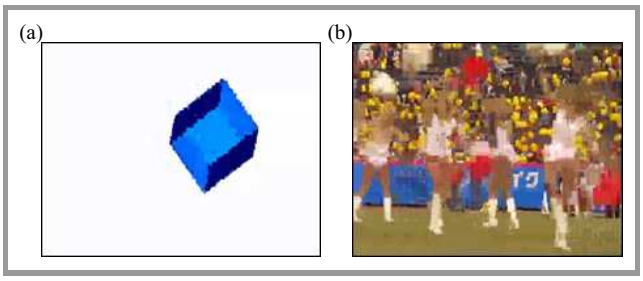


Fig. 5. Blue cube (videotom), 64 kb/s (a), Cheerleaders (TV), 250 kb/s (b).

good quality assessed by viewers. Provided in this article results and comments should be treated as initial experiment, which should be extended in order to draw final conclusions and present mathematic description for the relations. However, results are helpful to indicate difference scale between considered types of video sequences.

5. Summary

This paper is a part of work executed in Institute of Telecommunication at Warsaw University of Technology in terms of multimedia services and quality assessment. Results present how decreased coding bitrate affects video quality perceived by viewers as well as the scale of difference in the level of degradations between videotoms and standard television sequences. The most frequently occurred degradations are same for both types of video material and they are mainly related to blocking and flicker effects and to changes in shapes and colors. Because of the complexity and overall characteristic of TV sequences are more sensitive to negative impact of low bitrate values. For videotoms H.264/AVC coding mechanism seem to work very efficiently and degradations are visible only by lowest bitrate values.

Presented in the scope of this work results have to be extended with additional experiments considering influence of network changes according to earlier author's work [11], [12]. To do that firstly it is required to produce more videotoms sequences grouped into several sets based on their characteristic to compare them in the next step to television sequences with the same settings. It is also needed to engage more viewers in subjective tests in order to allow on more detailed statistic analysis for the results. For objective tests both simple metrics and more complex like Perceptual Video Quality (PVQ), eMOS [21] can be used as point of reference. Final results can allow author to determine functional relationship between considered types of sequences to create proper approach to parametric model creation.

References

[1] Ł. Trzcianowski, "Videotoms in objective and subjective quality tests of video", *J. Telecommun. Inform. Technol.*, no. 1, pp. 27–31, 2014.

[2] Ł. Trzcianowski and S. Kula, "Approach to QoS/QoE monitoring and parametric model for IPTV service", *Scient. J. Telecommun. Electron.*, vol. 16, no. 261, pp. 41–59, 2012.

[3] J. Asghar, "Preserving video quality in IPTV networks", *IEEE Broadcast Technol. Soc.*, vol. 55, 2009.

[4] A. Raak *et al*., "T-V-model: parameter-based prediction of ITV quality", in *Proc. IEEE Int. Conf. Acoust. Speech and Sig. Process. ICASSP 2008*, Las Vegas, NV, USA, 2008, pp. 1149–1152.

[5] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services", in *Proc. IEEE Int. Conf. Commun. ICC 2008*, Beijing, China, 2008, pp. 110–114.

[6] M. Vranjes, S. Rimac-Drlje, and D. Zagar, "Subjective and objective quality evaluation of the H.264/AVC coded video", in *Proc. 15th Int. Conf. Syst. Sig. and Image Process. IWSSIP 2008*, Bratislava, Slovak Republic, 2008, pp. 287–290.

[7] Z. Milicevic and Z. Bojkovic, "Subjective video quality assessment in H.264/AVC video coding standard", in *Proc. 19th Telecommun. Forum TELEFOR 2011*, Belgrade, Serbia, 2011, pp. 1183–1186.

[8] F. De Simone *et al.*, "Subjective quality assessment of H.264/AVC video streaming with packet losses", *EURASIP J. Image Video Proces*., vol. 1, 2011.

[9] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: applications and human-motivated design", *Sig. Process.*, vol. 25, no. 7, pp. 469–481, 2010.

[10] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann, 2013.

[11] L. Zhaoping, *Understanding Vision: Theory, Models, and Data*. OXFORD University Press, 2014.

[12] "Advanced video coding for generic audiovisual services", ITU-T Rec. H.264, 05/2003.

[13] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circ. Syst. for Video Technol*., vol. 13, no. 7, pp. 560–576, 2003.

[14] "High efficiency video coding", ITU-T Rec. H.265, 04/2013.

[15] "Methodology for the subjective assessment of the quality of television pictures", ITU-R Rec. BT.500-11, 06/2002.

[16] "Parametric non-intrusive assessment of audiovisual media streaming quality", ITU-T Rec. P.1201, 10/2012.

[17] "Parametric non-intrusive bitstream assessment of video media streaming quality", ITU-T Rec. P.1202, 10/2012.

[18] "Quality of experience requirements for IPTV services", ITU-T Rec. G.1080, 12/2008.

[19] "Consumer digital Video Library"– CDVL [Online]. Available: http://www.cdvl.org

[20] MSU Perceptual Video Quality Tool [Online]. Available: http://www.compression.ru/video/quality_measure/subjective_metrics_info_en.html

[21] "Perceptual Video Quality (PVQ)" [Online]. Available: http://www.tek.com/products/video-test/sentry-solutions/perceptual-video-quality.html

Łukasz Trzcianowski graduated from the Faculty of Electronics and Information Technology at Warsaw University of Technology (M.Sc., 2012). At present, he continues his work as Ph.D. candidate (Institute of Telecommunications, Warsaw University of Technology). The main focus in his research is given on telecommunication services and parametric models.
Email: ltrzcianowski@tele.pw.edu.pl
Institute of Telecommunication
Warsaw University of Technology
Nowowiejska st. 15/19
00-665 Warsaw, Poland

Paper

A New Implementation of UWB CRLH Based Antennas for Wireless Communications

Mohammad Alibakhshi-Kenari and Mohammad Naser-Mogaddasi

Faculty of Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Abstract—In this article, a novel ultra wideband (UWB) small antennas based on the composite right/left-handed transmission lines structures are proposed. The antennas are presented with best in size, bandwidth and radiation patterns. Their physical size and the operational frequency depend on the unit cell size and the equivalent transmission line model parameters of the CRLH-TL. To realize characteristics of first proposed model, Q-shaped gaps printed into rectangular radiation patches are used. The antenna based on composite right/left-handed transmission lines is composed of two unit cells, each of which occupies only 10.8 × 8.6 mm and covers the 2.7–9.3 GHz bandwidth for VSWR < 2. The peak gain and radiation efficiency, are 5.78 dBi and 42.1% 9.3 GHz, respectively. Moreover, the second designed antenna has the same size and enhancement bandwidth, gain and radiation efficiency than the first proposed antenna with similar design procedure. It is constructed of the printed Q-shaped four unit cells. The length, width and height are 21.6, 8.6 and 1.6 mm, respectively, and it covers 4.1–11.7 GHz bandwidth for VSWR < 2 having highest gain (7.18 dBi) and radiation efficiency (92.69%) at 4.1 GHz.

Keywords—composite right/left-handed transmission lines, metamaterial, modern wireless communication systems, portable devices, printed Q-shaped antennas, small antennas, ultra wideband antennas.

1. Introduction

Since their invention in 1960s, the microstrip patch antennas have found numerous applications for their simplicity in fabrication, compatibility with planar circuitry, low profile and planar structures, and unidirectional radiation capability. Despite of its many electrical and mechanical features, their use at low frequencies has been limited due to limited size and narrow bandwidth.

The conventional approach for miniaturizing the antenna size is to print the radiator on a high-k dielectric substrate. However, because of the capacitive nature of the patch geometry and the existence of high impedance contrast between the antenna substrate and the free space surrounding region, a high energy is trapped inside the dielectric material resulting in a narrow bandwidth and high radiation loss.

The metamaterials (MTMs) are very attractive for the design of small antennas and microwave devices [1], [2]. The composite right/left handed transmission lines (CRLH-TLs) provides a conceptual route for implementing small antennas. CRLH-based antennas can also be made broadband to support today's multi band communication and wireless applications requirements. The commercial uses of frequency band 3 to 10.6 GHz for radar, tracing, and data transmissions were approved by FCC in 2002 [3]. Recently, the research and development of the UWB communication systems including antennas have been widely performed [4]–[6]. One of the main port of the UWB system is an antenna providing low (< 2) VSWR over 3–10.6 GHz band. The designed two CRLH-based antennas support all cellular frequency bands (from 2.7 to 11.7 GHz), using single or multiple feed, which eliminates switches. Significant size reduction is also demanded to achieve the minimization of communication systems or devices. Ideally, the UWB antenna should be small, planar, then low cost and reliable. Also, compatibility and ease of integration with electronics for mobile communications are desirable. Furthermore, in order to satisfy the various demands for communication and wireless services, small antenna with wide bandwidth and good radiation characteristics are needed. Two designed antennas are based on CRLH-TLs, which results of very wideband and good radiation properties.

Developments of wireless communications systems call for more compact and multi band antennas to coexist in a small area, while maintaining their low coupling to support multipath channel decorrelation. Metamaterial structures have the ability to concentrate electromagnetic fields and currents near antenna structures, instead of spreading them along the ground, causing higher coupling. This allows compact antenna arrays to be realized with minimal mutual coupling, to be able to decorrelate multipath channels in MIMO implementations [7]–[8]. In this paper, the authors will focus on transmission lines (TL) based on composite right- and left-hand (CRLH) propagation [9]–[24]. It is nearly impossible to implement a pure left-handed (LH) transmission line due to the right-handed (RH) propagation inherited by using lumped elements [9]. Such transmission lines make possible unprecedented improvements in air-interface integration, over-the-air (OTA) functionality and miniaturization, while simultaneously reducing bill-of-materials costs and specific absorption rate (SAR) values. Metamaterials enable physically small but electrically large

air-interface components, with minimal coupling among closely spaced devices.

Metamaterials (MTM) are manmade composite materials, engineered to produce desired electromagnetic propagation behavior not found in natural media [9]–[10]. The word metamaterial refers to their many variations. Metamaterial antenna structures are copper, printed directly on the dielectric substrate, and can be fabricated by using a conventional Rogers RT Duroid 5880 substrate or a flexible printed circuit (FPC) board. Recently, a novel antennas with these characteristics have been designed by using CRLH-TL metamaterials [11]–[12]. Unlike traditional RH transmission materials, metamaterials based on LH transmission lines have unique features of anti-parallel phase and group velocities ($v_p - \|v_g$) [11]–[13]. Pure LH TLs cannot be implemented due to the existence of RH parasitic effects that occur naturally in practical LH TLs. CRLH-TL structures have been proposed, which also include RH effects. Several metamaterials-based antennas have already been presented, such as backward-to-forward leaky-wave [14]–[15], zeroth-order resonant [16], and so on.

Metamaterials are broadly defined as effectively homogeneous artificial structures exhibiting unusual properties, e.g. theirs index of refraction that may be negative (left handedness), less than one, or modulated in a graded manner. Such materials have spurred considerable interest and led to numerous applications over the past decade [18]–[19].

Metamaterials may be equivalently described in terms of media parameters (electric/magnetic dipole moments, electric/magnetic susceptibilities, permittivity, permeability), or in terms of transmission line parameters (inductance, capacitance, impedance, admittance, propagation constant, characteristic impedance). The latter approach, introduced in [20]–[21], has led to low-loss and broadband metamaterials, due to the non-resonant nature of the structural elements. This has been the foundation for the vast majority of the practical applications reported to date. More particularly, the concept of CRLH transmission line metamaterials, which describes in a simple and insightful manner the fundamentally dual RH/LH nature of metamaterials, has been widely recognized as a powerful paradigm for the understanding of metamaterial phenomena and the design of metamaterial devices [22], [23].

The applications of metamaterials may be classified in three categories:

- guided-wave components: multi-band, enhanced bandwidth, and miniaturized components, tight broadband couplers, compact resonators, uniform power combiners and splitters, UWB filters, agile distributed amplifiers, impulse delay lines and circuits;

- refracted-wave systems: focusing slabs, super-resolution imagers, reflection-less curved refractors, coordinate-transformation-based graded-index structures for electromagnetic manipulations;

- radiated-wave devices: mono/multi band passive/active one dimensional/two dimensional printed planar antennas and reflectors.

This article is concerned with the third category. It presents a selected number of the most practical CRLH metamaterial printed planar antennas based on utilizing CRLH metamaterial transmission lines technology and printed planar methodology, which have enhancement bandwidth and radiation characteristics.

The paper is organized as follows. Section 2 introduced antenna based on composite right/left-handed metamaterial transmission lines. Section 3 recommends a new idea of the design UWB small CRLH MTM antennas. In Section 4 the simulation results and discussions of the proposed printed antennas arrangements are presented. Afterwards in Section 5 provides a brief talk about benefits of the presented CRLH based antennas. Finally, discussion and conclusion are raised in Section 6.

2. Antennas Based on Composite Right/Left-Handed Metamaterial Transmission Lines

2.1. Fundamentals of CRLH Metamaterial TL Structures

Figure 1 shows the equivalent circuit of periodic CRLH metamaterial transmission lines (MTM TLs) in general case (lossy case). It should be noted that periodicity is here a convenience but not a necessity, as long as the largest cell is much smaller than the guided wavelength ($p \ll \lambda_g$) for electromagnetic homogeneity. Another important note is that as long as the effective medium condition, $p \ll \lambda_g$ is satisfied, there is no constraint on the minimum number of unit cells required for metamaterial operation. Even one single cell, when perfectly matched to the external world (i.e., presenting a block impedance equal to that of the external media or ports), behaves in a manner that cannot be distinguished from the behavior of a perfectly continuous medium of the same electrical size for the wave crossing it.

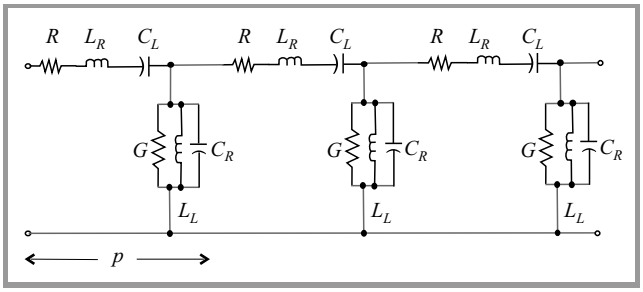


Fig. 1. The equivalent circuit model of periodic homogeneous CRLH metamaterial transmission lines structures composed of P unit cells in general case.

The homogeneous models of RH, LH, and CRLH lossless transmission lines are shown in Fig. 2 [11]. For a RH

lossless TL, its model is developed from conventional infinitesimal circuit model, where L_R is a series inductance and C_R is a shunt capacitance. The LH model is obtained by interchanging the inductance/capacitance and inverting the series/parallel arrangements in the equivalent circuit of the RH-TL, where C_L presents a series capacitance and L_L presents a shunt inductance. In effect, parasitic capacitance C_R due to development of voltage gradients and unavoidable parasitic inductance L_R due to current flow along the metallization will be added to LH TL and result in CRLH TL structure [11].

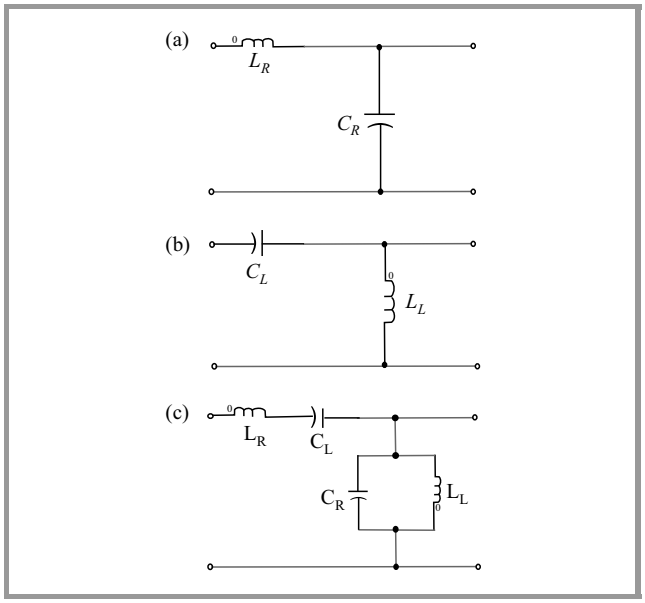


Fig. 2. Equivalent homogenous circuit models: (a) RH TL, (b) LH TL, (c) CRLH TL.

The model values R, L, C, G are known as the primary line parameters, from which the secondary line values are derived. The propagation constant value γ, for a given system is defined by the amplitude at the source of the wave to the amplitude at distance x, and [28] expressed as:

$$\frac{A_0}{A_x} = e^{\gamma x}. \tag{1}$$

Note that γ is a complex form:

$$\gamma = \alpha + j\beta, \tag{2}$$

where α is the real part called the attenuation and β, the imaginary part, is called the phase.
For example in a copper transmission line, the propagation constant can be calculated from the primary line constants by means of the relationship:

$$\gamma = \sqrt{ZY}, \tag{3}$$

where Z and Y are the impedance and admittance of the transmission line. In the special case of the CRLH TL, Z and Y are defined as [11]:

$$Z(\omega) = j\left(\omega L_R - \frac{1}{\omega C_L}\right), \tag{4}$$

$$Y(\omega) = j\left(\omega C_R - \frac{1}{\omega L_L}\right). \tag{5}$$

The dispersion relation for a homogenous CRLH TL is [11]:

$$\beta(\omega) = s(\omega)\sqrt{\omega^2 L_R C_R + \frac{1}{\omega^2 L_L C_L} - \left(\frac{L_R}{L_L} + \frac{C_R}{C_L}\right)}, \tag{6}$$

where

$$s(\omega) = \begin{cases} -1 & \text{if } \omega < \omega_{se} = \min\left(\frac{1}{\sqrt{L_R C_L}}, \frac{1}{\sqrt{L_L C_R}}\right) \\ 0 & \text{if } \omega_{se} < \omega < \omega_{sh} \\ +1 & \text{if } \omega < \omega_{sh} = \max\left(\frac{1}{\sqrt{L_R C_L}}, \frac{1}{\sqrt{L_L C_R}}\right) \end{cases}. \tag{7}$$

Figure 3 shows the ω-β or dispersion diagram of a RH TL, LH TL, and CRLH TL, respectively. The group velocity or $v_g = \frac{\partial\omega}{\partial\beta}$ slope and phase velocity or the line segment slope from origin to curve ($v_p = \frac{\omega}{\beta}$) of these TLs can be inferred from the dispersion diagram. For a purely RH TL, it is shown that v_g and v_p are in parallel ($v_g \| v_p$ and $v_g v_p > 0$).

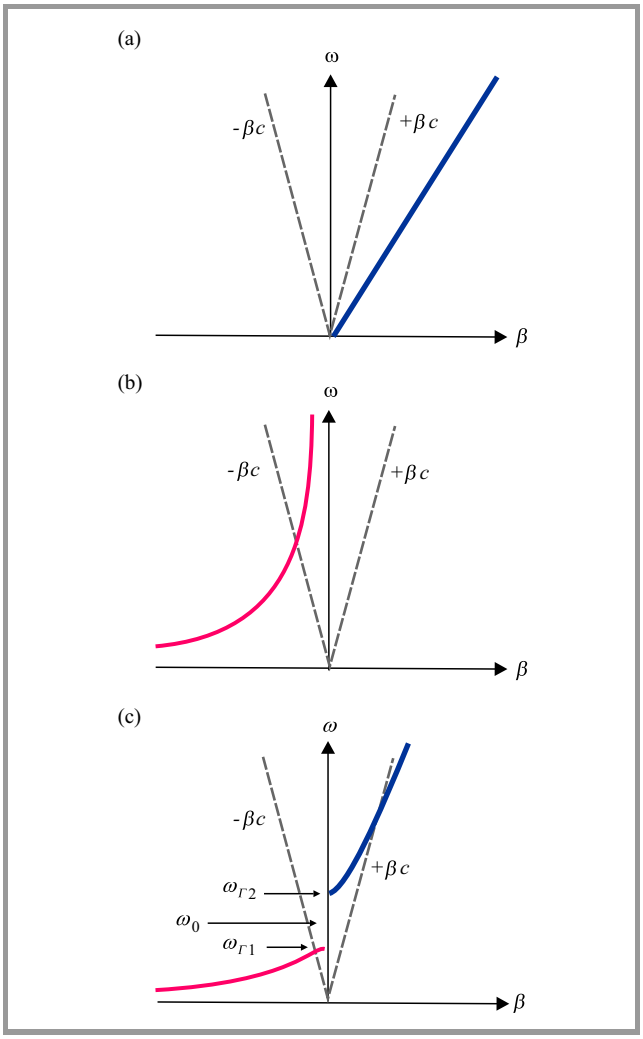


Fig. 3. Dispersion diagrams for homogenous: (a) RH TL, (b) LH TL, (c) CRLH TL (unbalanced case).

However, for a purely LH TL, the negative sign in $\beta(\omega)$ indicates a negative phase velocity and therefore v_p and v_g are anti-parallel ($v_p - \| v_g$ and $v_p v_g < 0$). In addition, Fig. 3 shows that, it has both LH ($v_p v_g < 0$) and RH ($v_p v_g > 0$) region. Also note that the stop band occurs in the frequency range where γ is purely real for a CRLH TL ($\beta = 0$). The group and phase velocities of the transmission line can be defined as:

$$v_g = \left(\frac{\partial \beta}{\partial \omega}\right)^{-1} = s\omega^2 \sqrt{L_L C_L}\,, \tag{8}$$

$$v_p = \frac{\omega}{\beta} = s\omega^2 \sqrt{L_L C_L}\,, \tag{9}$$

where s is a sign function defined as:

$$s = \left\{ \begin{array}{ll} +1 & \text{for RH TL} \\ -1 & \text{for LH TL} \end{array} \right\}. \tag{10}$$

LH-TL ihas high-pass nature, in contrast to that of the RH-TL which is of low-pass nature, in result a CRLH-TL contributes LH property at lower frequencies and RH at higher frequencies with a transition frequency ω_0. When the series and shunt resonances (ω_{se} and ω_{sh}) are equal, i.e.:

$$\omega_{se} = \frac{1}{\sqrt{L_R C_L}} = \omega_{sh} = \frac{1}{\sqrt{L_L C_R}} \tag{11}$$

or

$$L_R C_L = L_L C_R\,, \tag{12}$$

the propagation constant in Eq. (6) reduces to the simpler expression [11]

$$\beta = \beta_R + \beta_L = \omega\sqrt{L_R C_R} - \frac{1}{\omega\sqrt{L_L C_L}}\,, \tag{13}$$

where the phase constants splits up into β_R and β_L. Thus, there is a seamless transition from LH to RH for the balanced case occurring at the transition frequency ω_0 [11]:

$$\omega_0^{\text{unbalanced}} = \frac{1}{\sqrt[4]{L_R C_R L_L C_L}} \tag{14}$$

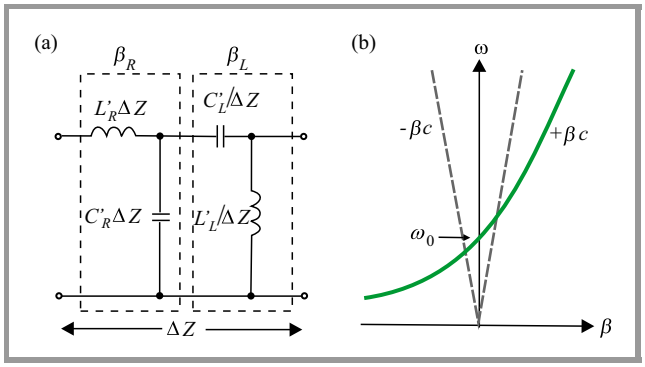


Fig. 4. Balanced form of: (a) simplified equivalent circuit model, (b) dispersion diagram showing seamless transition from LH to RH region.

and in the balanced case, $\omega_0^{\text{balanced}}$ is:

$$\omega_0^{\text{balanced}} = \frac{1}{\sqrt{L_R C_L}} = \frac{1}{L_L C_R}\,. \tag{15}$$

A balanced form of a CRLH TL is shown in Fig. 4. The simplified equivalent circuit model is the series combination of RH and LH TLs. Also, the balanced CRLH TL's dispersion curve does not have a stop band. At ω_0 the phase shift ($\varphi = -\beta d$) for a TL of length d is zero ($\beta = 0$). Phase advance ($\varphi > 0$) occurs in the LH frequency range ($\omega < \omega_0$, $\beta < 0$), and phase delay ($\varphi < 0$) occurs in the RH frequency range ($\omega > \omega_0$, $\beta > 0$) [11].

The TL characteristic impedance is $Z_0 = \sqrt{\frac{Z}{\gamma}}$. In unbalanced case:

$$Z_0 = Z_L \sqrt{\frac{C_L L_R \omega^2 - 1}{C_R L_L \omega^2 - 1}} \tag{16}$$

in the balanced case: $Z_0 = Z_L = Z_R$, and

$$Z_L = \sqrt{\frac{L_L}{C_L}} \tag{17}$$

$$Z_R = \sqrt{\frac{L_R}{C_R}}\,, \tag{18}$$

where Z_L and Z_R are the purely LH and RH impedances, respectively. According to Eq. (16) the characteristic impedance for the unbalanced case is frequency dependent, however, according to Eqs. (17) and (18) for the balanced case is frequency independent and therefore, can be matched over a wide bandwidth.

The TL material permeability and permittivity have been related to the impedance and admittance of its equivalent TL model:

$$\mu = \frac{Z}{j\omega} = L_R - \frac{1}{\omega^2 C_L}\,, \tag{19}$$

$$\varepsilon = \frac{Y}{j\omega} = C_R - \frac{1}{\omega^2 L_L}\,. \tag{20}$$

Equations (19) and (20) prove that for balanced case the permeability and permittivity are negative in LH region ($\omega < \omega_0$).

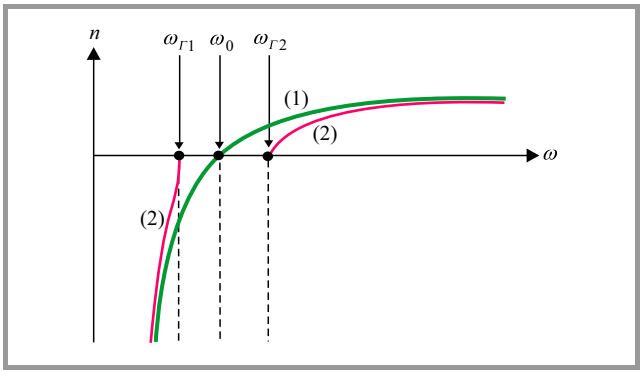


Fig. 5. Typical refraction index plots for the balanced (1) and unbalanced (2) CRLH TL.

The refraction index ($n = \frac{c\beta}{\omega}$) for the balanced and unbalanced CRLH-TL is shown in Fig. 5 [11]. As depicted

the CRLH-TL has a negative refraction index in its LH range and a positive refraction index in its RH range.

2.2. CRLH Metamaterials in Antenna Design

The antenna has become one of the most important component when designing wireless communication systems in portable devices. Due to the limited space available, shrinking conventional antennas may lead to performance degradation and complicated mechanical assembly. Metamaterial technology provides an opportunity to design of a smaller size antenna at lower cost with better radiation performance. Various implementations of metamaterial structures have been reported and demonstrated in [9]–[10]. In this article, a transmission line type of realization CRLH-TL that possesses characteristics of low insertion loss, broad bandwidth, low profile and good radiation performance will be employed.

A metamaterial is usually a periodic structure with N identical unit cells cascading together, where each cell is much smaller than one wavelength at the operational frequency. The composition of one metamaterial unit cell is categorized as a series inductor (L_R) and capacitor (C_L), shunt inductor (L_L) and capacitor (C_R). L_L and C_L determine the left-handed mode propagation properties, while L_R and C_R govern the right-handed mode propagation properties. The behavior of both propagation modes at different frequencies can be easily addressed in a simple dispersion diagram, as shown in Fig. 6. The dispersion curve on the $\beta > 0$ side

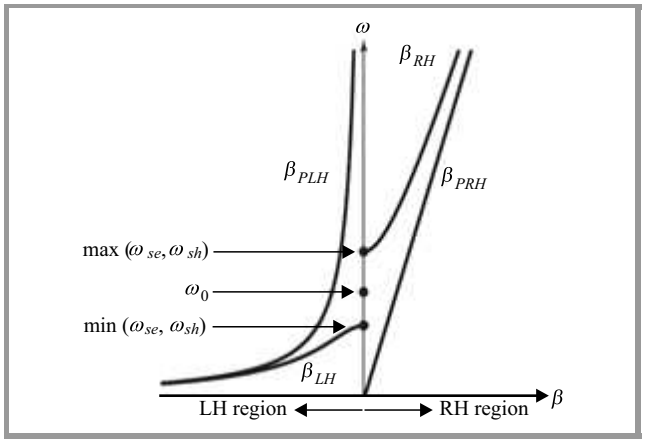


Fig. 6. Dispersion diagrams for a CRLH TL (unbalanced case).

is the right-handed mode, while the dispersion curve on the $\beta < 0$ side is the left-handed mode [9]. The electrical size of a conventional transmission line is strongly related to its physical dimensions and thus reducing device size usually means increasing operational frequency. In contrary, the dispersion curve of a metamaterial is determined by the four CRLH parameters. If these four parameters are realized in a very compact form, the corresponding circuit size will be physically small but electrically large. This concept has been adopted successfully in small antenna designs [7]–[25].

3. UWB and Small CRLH MTM Antenna Design

3.1. Printed Q-Shaped Two Unit Cells Antenna

The equivalent circuit model design of the proposed MTM antenna is based on the CRLH-TL structure shown in Fig. 7. The proposed planar antenna is fabricated on an Rogers RT

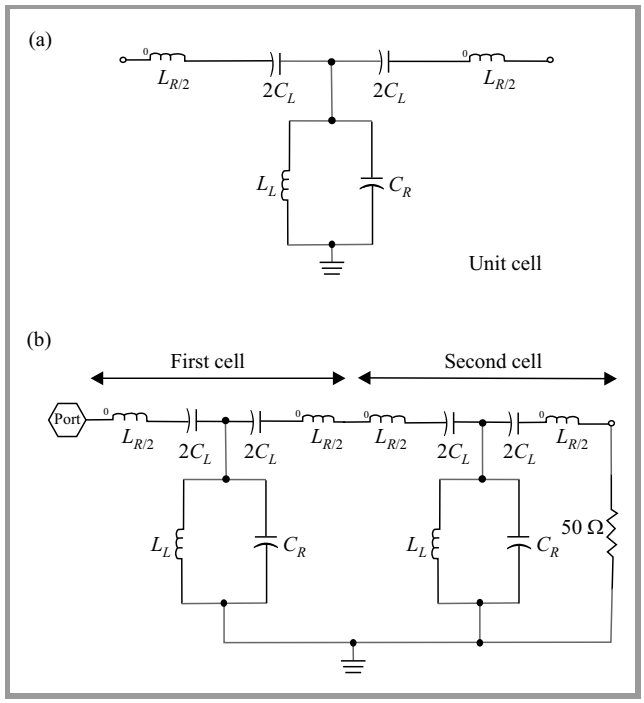


Fig. 7. The equivalent circuit model of the proposed printed Q-shaped antenna: (a) unit cell, (b) whole structure.

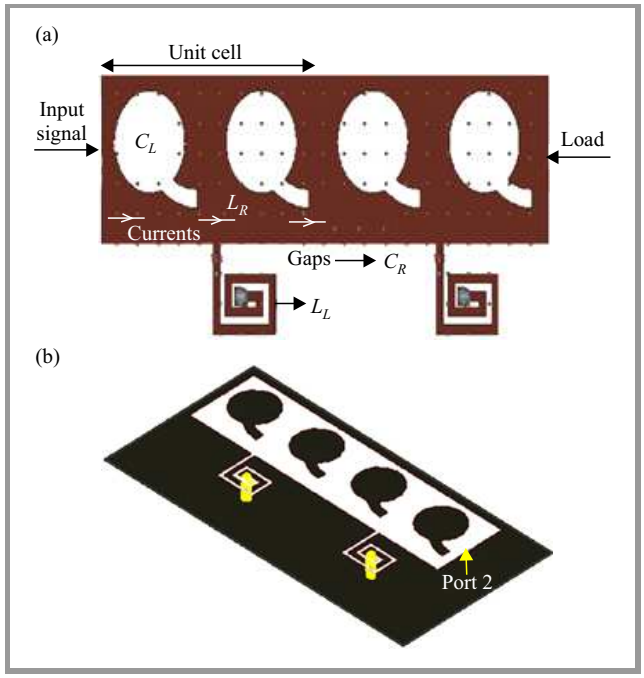


Fig. 8. Proposed printed Q-shaped antenna: (a) top view, (b) 3D view.

Duroid 5880 substrate, with a 2.2 dielectric constant, and a thickness of 1.6 mm. This mushroom type unit cell consisted of a 10.8 × 8.6 mm top patch, printed on top of the substrate and a rectangular inductor attending a metallic via. Each unit cell was coupled to its adjacent unit cell and the vertical via was connected between the rectangular inductor and the ground on the back of the substrate. This antenna was excited by external port as input signal, as shown in Fig. 8. The shape and dimensions were optimized for matching purposes, reducing of the occupy area, enhancement bandwidth and providing good radiation properties.

The antenna is based on two simplified planar mushroom structure unit cells. The unit cell is composed of a host transmission line with two printed Q-shaped gaps into rectangular radiation patches and a rectangular inductor connected to ground plane through a metallic via. The Q-shaped gaps printed within patches operates as series capacitance (C_L) and the rectangular inductor accompanying vertical metallic via hole connected to ground plane performs a shunt inductance (L_L). A purely left-handed transmission line cannot exist physically because, of parasiting effects increasing with frequency. Thus, the CRLH model represents the most general MTM structure possible. This antenna structure is excited by external port (Port 1) as input signal and Port 2 is matched to 50 Ω load impedance of the SMD 1206 size components connected to ground plane through a via. Configuration of the proposed printed Q-shaped two unit cells antenna is shown in Fig. 8.

Presented antenna is formed of the two simplified planar mushroom structure Q-shaped unit cells, each of which occupies only 10.8 × 8.6 mm or $0.18\lambda_0 \times 0.15\lambda_0$ at the resonance frequency $f = 5.2$ GHz, therefore, the physical dimensions are 21.6, 8.6 and 1.6.

The important issue of many conventional metamaterial antennas confront is a lack of bandwidth [26]–[27].

The transmission coefficient is an important frequency domain performance indicator of an UWB antenna [29]. In this paper, the authors proposed several method to extend the bandwidth of the MTM antennas with a fixed antenna size. The points summarize in below, are guidelines for such design:

- traveling wave antennas or antennas having low Q can be very broadband [37];
- antennas incorporating tapers or rounded edges tend to give broad bandwidths because surface currents have a smooth path to follow [30];
- linearly polarized antennas are the simplest to implement in a compact planar package;
- minimizing the thickness of the substrate and using low loss materials maximizes radiation efficiency [40];
- using of the printed planar technology into radiation patches for antenna design with minimizing acceptable distance between gap edges results to extended the bandwidth of the antenna [38]–[39].

In this article, the 2, 4 and 5 approaches for increasing the bandwidth are used. By using a smaller value of the loaded C_L on the CRLH-TL, broadband performance can be obtained. A smaller value of the loaded C_L will be realized by implementation of the Q-shaped gaps with closely space edges printed into rectangular patches. This method is used to increase the bandwidth, as providing UWB antenna with 6.6 GHz bandwidth (from 2.7 GHz to 9.3 GHz) for VSWR < 2. The antenna gain and radiation efficiency at resonance frequency $f_r = 5.2$ GHz are equal to 4.71 dBi and 41.82%, respectively. The simulated reflection coefficient ($S_{11} < -10$ dB) and radiation gain pattern at $f_r = 5.2$ GHz are plotted in Figs. 9 and 10, respectively.

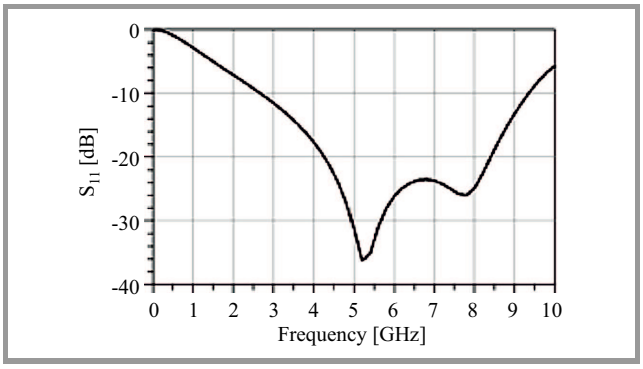


Fig. 9. Simulated reflection coefficient.

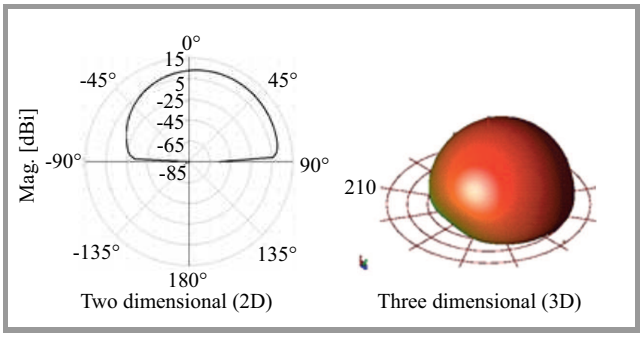


Fig. 10. The radiation gain pattern in elevation plane ($\Phi = 0°$) and at resonance frequency $f_r = 5.2$ GHz.

Next, an useful MTM antenna based on two simplified planar mushroom structure unit cells is designed. Its configuration employing the proposed methods is shown in Fig. 8. The performance of the presented designing methods and antenna structure are verified using Agilent ADS full-wave simulator.

3.2. Improvement Gain Antenna with Printed Q-Shaped Four Unit Cells Structure

In this section, the four unit cells printed Q-shaped antenna structure with enhancement bandwidth, radiation gain and efficiency in comparison structure proposed in previous section. The design procedure is completely same,

but equivalent circuit model and geometry is different. Presented antenna in this section is designed as one rectangular inductor attending a via are considered for each gap area which become one constructing one unit cell, in results this antenna composed of four unit cells as shown in Fig. 12. Equivalent circuit model is shown in Fig. 11. As obvious recent structure composed of four unit cells as each unit cell consist of a series capacitance (C_L) which created by printed Q-shaped gap capacitance, a shunt inductance (L_L) that caused with a rectangular inductor connected to ground plane through a via as these capacitor an inductor plays left-handed roles, also a series inductance (L_R) which established by unavoidable current flow on the patches and a shunt capacitance (C_R) that performed with gap capacitor between patches and ground plane.

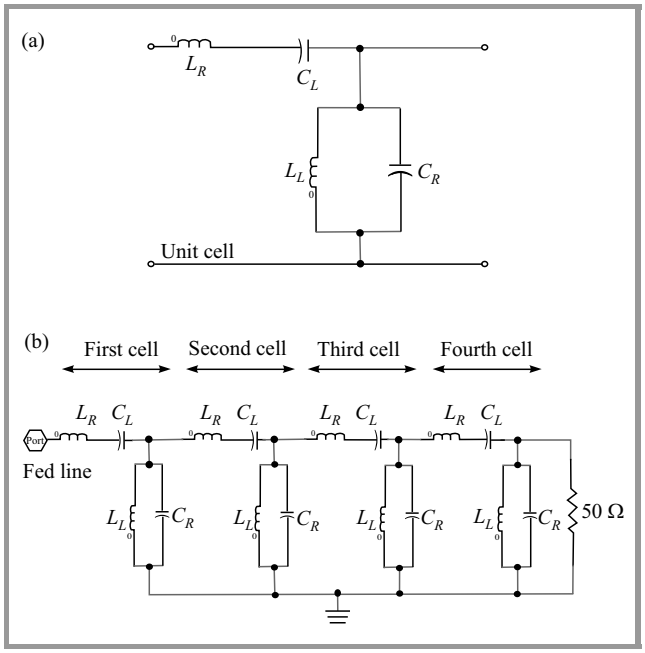


Fig. 11. The equivalent circuit model of four unit cells printed Q-shaped antenna: (a) unit cell, (b) whole structure.

This typical CRLH antenna structure consists of a feed line that is electromagnetically coupled to metallic patches, rectangular inductors, via that connects the rectangular inductor to the ground plane. This feed line through a small gap excites the CRLH unit cells. Typically, the antenna is matched to a port with 50 Ω load impedance. The resonant frequency, matching of multiple right-handed and left-handed modes, and associated efficiencies can be controlled by the size of the cell patch, the via line width, the feed line length, the distance between the antenna elements and the ground, optimizing the rectangular inductor and various other dimensions and layouts [7]–[17]. The gap capacitor and the rectangular inductor accompanying a via could be viewed as C_L and L_L, while the top patch possessed the L_R and C_R to the ground. Therefore, a left-handed resonance could be obtained at the desired frequency by properly designing the gap capacitor and the rectangular inductor that connected to ground plane through a via. Formation of the recommended antenna with enhancement bandwidth and improvement radiation properties is shown in Fig. 12.

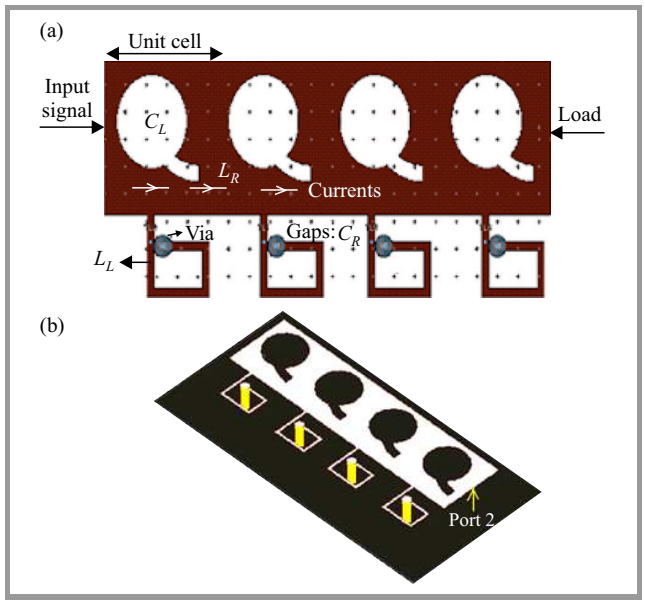


Fig. 12. Formation of the four unit cells printed Q-shaped antenna: (a) top view, (b) isometric view.

Printed Q-shaped antenna was designed on Rogers RT Duroid 5880 substrate with dielectric constant $\varepsilon_r = 2.2$ and thickness $h = 1.6$ mm. It is formed of the four simplified planar mushroom structure Q-shaped unit cells, each of which occupies 5.4×8.6 mm or $0.1\lambda_0 \times 0.16\lambda_0$ at the resonance frequency $f_r = 5.65$ GHz, where λ_0 is free space wavelength, therefore overall size of the Q-shaped antenna is $21.6 \times 8.6 \times 1.6$ mm^3 ($0.4\lambda_0 \times 0.16\lambda_0 \times 0.03\lambda_0$). This antenna has 7.6 GHz applicable bandwidth from 4.1 to 11.7 GHz for VSWR < 2, which corresponds to 96.2% practical bandwidth. In addition at resonance frequency $f_r = 5.65$ GHz the gain and radiation efficiency are 6.52 dBi and 60.47%, respectively. The simulated return loss bandwidth (S_{11}) of the antenna and radiation gain pattern at $f_r = 5.65$ GHz are plotted in Figs. 13 and 14 respec-

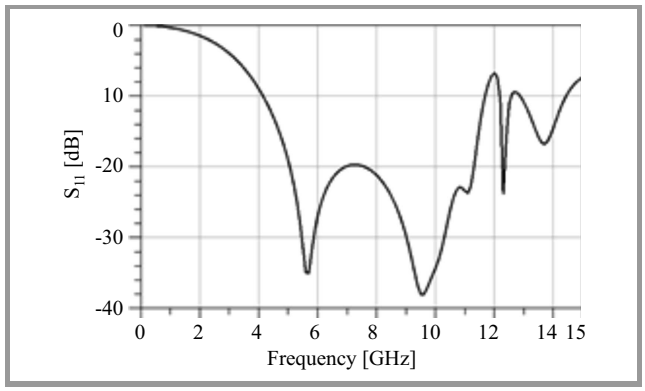


Fig. 13. Simulated return loss (S_{11}).

tively. The results show that with the same physical size, broader bandwidth and superior radiation performances was achieved.

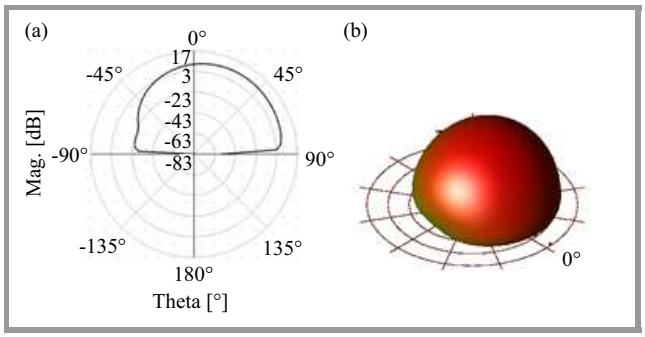


Fig. 14. The radiation gain pattern in elevation plane ($\Phi = 0^\circ$) and at resonance frequency $f_r = 5.65$ GHz: (a) 2D view, (b) 3D view.

Entire equivalent circuit model and configuration of the both Q-shaped antennas are illustrated in Fig. 7. Two designed UWB small antennas based on CRLH MTM-TLs are suitable and useful for microwave and portable devises and wireless communication applications.

4. Simulation Results and Discussion of the Proposed Printed Antennas

Figure 9 and Fig. 13 show the simulated return losses of the two unit and four unit cells obtained by using Agilent ADS full-wave simulator. The simulated return losses bandwidth ($S_{11} < -10$ dB) of the first presented antenna is 6.6 GHz (from 2.7 to 9.3 GHz), this corresponds to 110% practical impedance bandwidth, which is more than conventional. The simulated return losses bandwidth ($S_{11} < -10$ dB) of the second device (Fig. 13) is 7.6 GHz (from 4.1 to 11.7 GHz), which corresponds to 96.2% practical impedance bandwidth, which is more than conventional realizations. The simulated 2D radiation gain patterns of the recommended antennas at different frequency are plotted in Figs. 15 and 16. For first antenna,

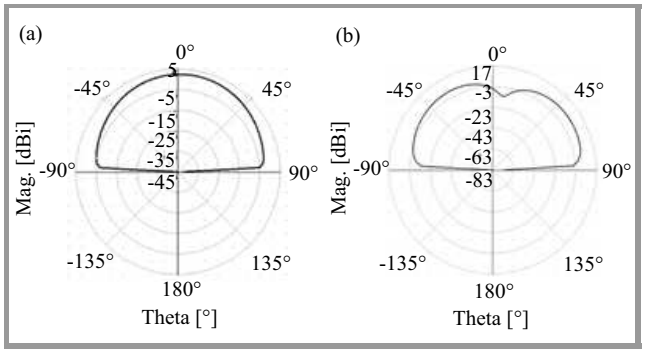


Fig. 15. The simulated radiation gain patterns of the two unit cells antenna in elevation plane ($\Phi = 0^\circ$): (a) at 47 GHz and (b) 9.3 GHz.

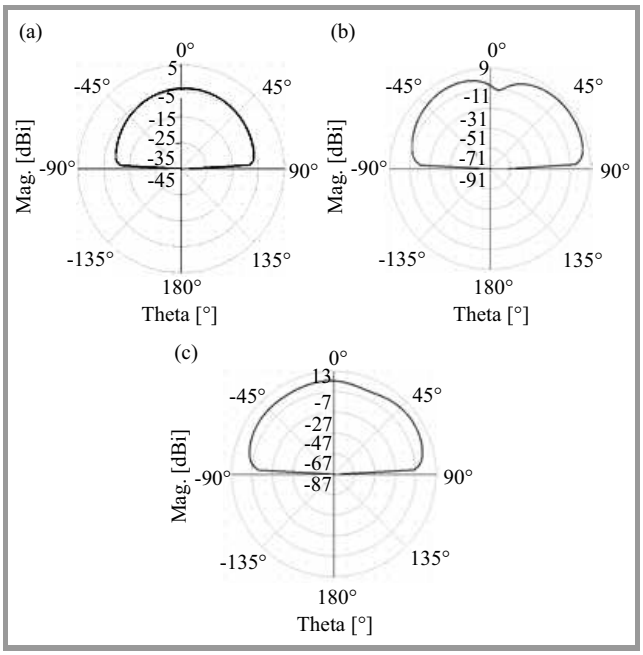


Fig. 16. The simulated radiation gain patterns of the 4-cells antenna in elevation plane ($\Phi = 0^\circ$) at three different frequencies.

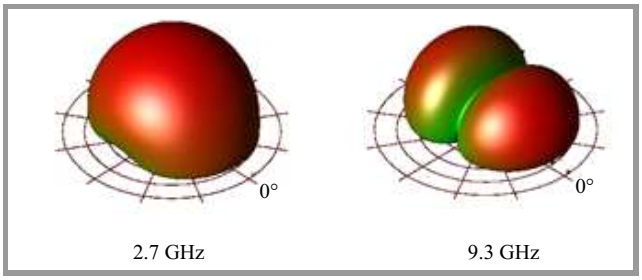


Fig. 17. Simulation of radiation gain patterns of the 2-cells antenna in elevation plane ($\Phi = 0^\circ$) at two frequencies.

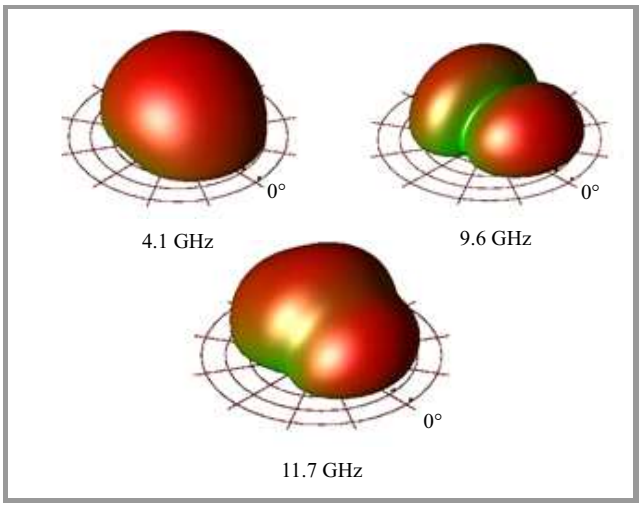


Fig. 18. Simulation of radiation gain patterns of the 4-cells antenna in elevation plane ($\Phi = 0^\circ$) at 3 frequencies.

radiation gains are 2.8 dBi and 5.78 dBi, respectively, at 2.7 and 9.3 GHz. Also, radiation efficiencies are 37.23 and 42.1% at same frequencies, respectively. For second

antenna, at 4.1, 9.6 and 11.7 GHz, gain and efficiency are 7.18 dBi and 92.69%, 5.83 dBi and 34.89%, and 5.42 dBi and 54.19%, respectively.

Table 1
Radiation characteristics of two typical antennas in comparison to the proposed

Parameters	Antenna described in [26]	Antenna described in [31]	Proposed first antenna	Proposed second antenna
Gain [dBi]	0.6	0.45	5.78	7.18
Bandwidth [GHz]	1–2	0.8–2.5	2.7–9.3	4.1–11.7
Efficiency [%]	26	53.6	42.1	92.69

Table 2
Comparison of dimensions of antennas

UWB antennas	Design size
Slotted planar binomial monopole antenna [32]	$30 \times 27.4 \times 1$ mm
Slotted circular monopole antenna [33]	$26 \times 27 \times 1$ mm
Slotted rectangular monopole antenna [34]	$18 \times 20 \times 1$ mm
Fork shaped antenna [35]	$35 \times 30 \times 0.769$ mm
Slotted arc shaped edge rectangular antenna [36]	$24 \times 35 \times 0.8$ mm
Both proposed UWB antennas	$21.6 \times 8.6 \times 1.6$ mm

According to Figs. 15 and 16 radiation patterns are unidirectional. The peak gain and radiation efficiency of the both versions occur at 9.3 and 4.1 GHz and are equal to 5.78 dBi and 42.1%, and 7.18 dBi and 92.69%, respectively. The simulated 3D radiation gain patterns at different frequencies are plotted in Figs. 17 and 18. To validate the design procedures the proposed antennas were compared with well-known compact UWB antennas and their characteristics were summarized in Tables 1 and 2.

5. CRLH-Based Antennas Benefits

The proposed CRLH-based antennas are wideband, low profile and compact in size, smaller than 50 by 10 mm in area on a PCB, and consist of superior radiation performances. In fact, the proposed CRLH-based antennas are typically five times smaller than conventional, while offering equal or better performance. Furthermore, unlike conventional three-dimensional antennas – which must be designed, tooled and fabricated as a complex metal-and-plastic assembly, the proposed shown in Figs. 8 and 12 are a simple fabricate. This offers manufacturers faster time to market and reduced bills-of-materials due to the simplified design. It also offers a greatly reduced need for fabrication and assembly of antenna components.

In addition, the CRLH-based antenna's ability to concentrate electromagnetic fields and currents near their antenna structures results in achieving better performance.

6. Conclusion

CRLH transmission line metamaterials represent a paradigm shift in electromagnetics engineering and, in particular, for antennas. They exhibit exceptional properties, resulting from their rich dispersion and their fundamental left/right-hand duality. They offer simple and deep insight into metamaterial phenomena, and provide efficient tools for the practical design of components and antennas.
In this paper, the authors introduced a new concept of antenna size reduction based on metamaterial design and printed planar technology, and also presented a novel idea of antenna bandwidth enhancement and radiation's properties improvement based on employing appropriate inductive elements accompanying their optimize values and using low loss materials maximizes radiation efficiency. All simulated results show that the proposed CRLH-based antennas have superior performance and smaller size compared to other conventional antennas design. These antennas have the small size, UWB, lightweight, high gain and efficiency, unidirectional radiation patterns, simple implementation and low cost. The simulated results exhibit that the proposed antennas should be potential candidates to use in the modern wireless communication systems and portable devices.

References

[1] C. Caloz and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications*. New York: Wiley, 2006.

[2] N. Engheta and R. W. Ziolkowski, *Metamaterials: Physics and Engineering Explorations*. New York: Wiley, 2006.

[3] *FCC First Report and Order on Ultra Wideband Technology*, Feb. 2002.

[4] Y. J. Wang *et al*., "Novel microstrip-monopole integrated ultra-wideband antenna for mobile UWB devices", in *Proc. Radio and Wirel. Conf. RAWCON '03*, Boston, MA, USA, 2003, pp. 87–90.

[5] T. Taniguchi and T. Kobayashi, "An omnidirectional and low-VSWR antenna for the FCC-approved UWB frequency band", in *Proc. IEEE Int. Symp. Anten. and Propag.*, Columbus, OH, USA, 2003, vo1. 3, pp. 460–463.

[6] W. Sorgel, C. Waldschmidt, and W. Wiesbeck, "Transient response of a Vivaldi antenna and a logarithmic periodic dipole array for ultra wideband communication", in *Proc. IEEE Int. Symp. Anten. and Propag.*, Columbus, OH, USA, 2003, vo1. 3, pp. 592–595.

[7] C. J. Lee, M. Achour, and A. Gummalla, "Compact metamaterial high isolation MIMO antenna subsystem", in *Proc. Asia Pacific Microw. Conf. APMC 2008*, Hong Kong, China, 2008, pp. 1–4.

[8] C. J. Lee, A. Gummalla, and M. Achour, "Compact dualband antenna subsystem for MIMO application", in *Proc. Int. Worksh. Antenna Technol.: Small Antenn. Novel Metamater. iWAT 2009*, Santa Monica, CA, USA, 2009, pp. 1–4.

[9] C. Caloz and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications, The Engineering Approach*. New York: Wiley, 2005.

[10] R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental verification of a negative index of refraction", *Science*, vol. 292, no. 5514, pp. 77–79, 2001.

[11] A. Lai, C. Caloz, and T. Itoh, "Composite right/left-handed transmission line metamaterials", *IEEE Microw. Mag.*, vol. 5, no. 3, pp. 34–50, 2004.

[12] F. J. Herraiz-Martínez, V. González-Posadas, L. E. Garcia-Munoz, and D. Segovia-Vargas, "Multifrequency and dual-mode patch antennas partially filled with left-handed structures", *IEEE Trans. Anten. Propag.*, vol. 56, no. 8, pp. 2527–2539, 2008.

[13] S. G. Mao, S. L. Chen, and C. W. Huang, "Effective electromagnetic parameters of novel distributed left-handed microstrip lines", *IEEE Trans. Microw. Theory Technol.*, vol. 53, no. 4, pp. 1515–1521, 2005.

[14] C. Caloz and T. Itoh, "Novel microwave devices and structures based on the transmission line approach of meta-materials", in *Proc. IEEE MTT-S Int. Microw. Symp. Dig.*, Philadelphia, PA, USA, 2003, pp. 195–198.

[15] Y. M. Pan and S. J. Xu, "A new leaky wave antenna based on channel guide filled with left-hand material for millimeter-wave applications", *Int. J. Infrar. Millim. Waves*, vol. 27, no. 11, pp. 1457–1468, 2006.

[16] A. Sanada, C. Caloz, and T. Itoh, "Zeroth order resonance in the left-handed transmission line", *IEICE Trans. Electron.*, vol. 87-C, no. 1, pp. 1–7, 2004.

[17] W. Huang, N. Xu, V. Pathak, G. Poilasne, and M. Achour, "Composite Right-Left Handed Metamaterial Ultra-Wideband Antenna", in *Proc. Int. Worksh. Anten.a Techno.: Small Anten. Novel Metamater. iWAT 2009*, Santa Monica, CA, USA, 2009, pp. 1–4.

[18] C. Caloz and T. Itoh, *Electromagnetic Metamaterials, Transmission Line Theory and Microwave Applications*. Piscataway: Wiley/IEEE Press, 2005.

[19] N. Engheta and R. W. Ziolkowski, Eds., *Electromagnetic Metamaterials: Physics and Engineering Explorations*. Piscataway: Wiley/IEEE Press, 2006.

[20] C. Caloz and T. Itoh, "Application of the transmission line theory of left-handed (LH) materials to the realization of a microstrip LH transmission line", in *Proc. IEEE In. Symp. Anten. and Propag. Digest*, San Antonio, USA, 2002, pp. 412–415.

[21] A. K. Iyer and G. V. Eleftheriades, "Negative refractive index metamaterials supporting 2-D Waves", in *Proc. IEEE Int. Symp. Microw. Theory Techniq. Digest*, Seattle, USA, 2002, pp. 1067–1070.

[22] C. Caloz and T. Itoh, "Novel microwave devices and structures based on the transmission line approach of metamaterials", in *Proc. IEEE Int. Symp. Microw. Theory Techniq. Digest*, Philadelphia, USA, 2003, pp. 195–198.

[23] R. E. Collin and F. J. Zucker, Eds., *Antenna Theory*, Part 2, Chapters 19 and 20. New York: McGraw Hill, 1969.

[24] G. V. Eleftheriades and K. G. Balmain, *Negative Refraction Metamaterials: Fundamental Principles and Applications*. New York: Wiley, 2005.

[25] C. J. Lee, K. M. H. Leong, and T. Itoh, "Broadband small antenna for portable wireless application", in *Proc. Int. Worksh. Anten. Technol.: Small Anten. Novel Metamater. iWAT 2008*, Chiba, Japan, 2008, pp. 10– 13.

[26] C. J. Lee, K. M. K. H. Leong, and T. Itoh, "Composite right/left-handed transmission line based compact resonant antennas for RF module integration", *IEEE Trans. Anten. Propag.*, vol. 54, no. 8, pp. 2283–2291, 2006.

[27] M. Schussler, J. Freese, and R. Jakoby, "Design of compact planar antenna using LH-transmission lines", in *Proc. IEEE MTT-S Int. Microw. Symp.*, Fort Worth, TX, USA, 2004, pp. 209–212.

[28] E. Weber and F. Nebeker, *The Evolution of Electrical Engineering*, presented at IEEE Press, Piscataway, New Jersey, USA, 1994.

[29] Q. Ye, "Time domain response of ultra-wideband dipole antennas", in *Proc. 2004 Antem/URSI Conf. Anten. Technol. Appl. Electromag.*, Ottawa, Canada, 2004, pp. 661–664.

[30] E. Gazit "Improved design of the Vivaldi antenna", *IEEE Trans. Anten. Propag.*, vol. 135, 1988.

[31] Y. Li, Z. Zhang, J. Zheng and Z. Feng, "Compact heptaband reconfigurable loop antenna for mobile handset", *IEEE Anten. Wirel. Propag. Lett.*, vol. 10, pp. 1162–1165, 2011.

[32] Y. L. Zhao, Y. C. Jiao, G. Zhao, L. Zhang, Y. Song, and Z. B. Wong, "Compact planar monopole UWB antenna with band-notched characteristic", *Microw. Opt. Technol. Lett.*, vol. 50, no. 10, pp. 2656–2658, 2008.

[33] R. Movahedinia and M. N. Azarmanesh, "A novel planar UWB monopole antenna with variable frequency band-notch function based on etched slot-type ELC on the patch", *Microw. Opt. Technol. Lett.*, vol. 52, no. 1, pp. 229–232, 2010.

[34] M. Abdollahvand, G. Dadashzadeh, and D. Mostafa, "Compact dualband-notched printed monopole antenna for UWB application", *IEEE Anten. Wirel. Propag. Lett.*, vol. 9, pp. 1148–1151, 2010.

[35] S. J. Wu, C. H. Kang, K. H. Chen, and J. H. Tarng, "Study of an ultra wideband monopole antenna with a band-notched open-looped resonator", *IEEE Trans. Anten. Propag.*, vol. 58, no. 6, pp. 1890–1897, 2010.

[36] C.-Y. Hong, C.-W. Ling, I.-Y. Tarn, and S.-J. Chung, "Design of a planar ultra wideband antenna with a new band-notch structure", *IEEE Trans. Anten. Propag.*, vol. 55, no. 12, pp. 3391–3397, 2007.

[37] M. Alibakhshi Kenari, "A new compact UWB traveling-wave antenna based on CRLH-TLs for embedded electronic systems", *Int. J. Microw. Wirel. Technol.*, 2014 (in press).

[38] M. Alibakhshi Kenari, "Introducing the new wide band small plate antennas with engraved voids to form new geometries based on CRLH MTM-TLs for wireless applications", *Int. J. Microw. Wirel. Technol.*, 2014 [Online]. Available: http://dx.doi.org/10.1017/S1759078714000099

[39] M. Alibakhshi Kenari, "Printed planar patch antennas based on metamaterial", *Int. J. Elec. Lett.*, 2014 [Online]. Available: http://dx.doi.org/10.1080/21681724.2013.874042

[40] M. Alibakhshi Kenari, "Minimized antenna based on CRLH transmission line for broadband communications", *Int. J. Elec. Lett.*, 2014 (accepted).

Mohammad Alibakhshi-Kenari received his B.S. and M.Sc. degrees in the Electrical Engineering, field of Telecommunication from the Islamic Azad University, Najafabad Branch at Esfahan, Iran on Feb. 2010 and the Islamic Republic of Iran, Shahid Bahonar university of Kerman on Feb. 2013, respectively. His researches interests include microwave and millimeter wave circuits, radars, antennas and wave propagation, composite right/left-handed transmission lines (CRLH-TLs), metamaterial (MTM) applications, integrated RF technologies, embedded systems, electromagnetic waves applications and wireless telecommunication systems. He is now Editor-in-chief in Journal Club for Electronic and Communication Engineering (JCECE) and also works as a reviewer in the several ISI journals such as Elsevier, Taylor & Francis, Wiley, ACES, etc. M. A.-Kenari has served as a Member of the Technical Program Committee (M-TPC) of some of the international conferences such as APACE 2014, APPEIC 2014, MobiWIS 2014, ICEPIT 2014, CICN 2014, ADVCIT 2014, DPNoC 2015, MobiApps 2015, etc. So far, he has been published several papers. His Master Thesis entitled "Designing and Fabricating the Ultra Compact and UWB Antennas based on Metmaterial Transmission Lines with Application in Wireless Radio Transceivers" was approved and granted

by Iran Telecommunication Research Center (ITRC) at Dec. 2012 with grant number of 6987/500/T.
E-mail: makenari@mtu.edu
Faculty of Engineering, Science and Research Branch
Islamic Azad University
Tehran, Iran

Mohammad Naser-Mogaddasi received the B.Sc. degree in Communication Engineering in 1985 from the Leeds Metropolitan University (formerly Leeds Polytechnic), UK. Between 1985 and 1987 he worked as an RF design engineer for the Gigatech company in Newcastle Upon Tyne, UK. From 1987 to 1989, he was awarded a full scholarship by the Leeds educational authority to pursue an M. Phil. He received his Ph.D. in 1993, from the University of Bradford, UK. From 1995, Dr. Naser-Moghadasi joined Islamic Azad University, Science and Research Branch, Iran, where he currently is head of postgraduate studies and also member of Central Commission for Scientific Literacy and Art Societies. His main areas of interest in research are microstrip antenna, microwave passive and active circuits, RF MEMS. Dr. Naser-Moghadasi is member of the Institution of Engineering and Technology, MIET and the Institute of Electronics, Information and Communication Engineers (IEICE). He has so far published over 130 papers in different journals and conferences.

E-mail: mn.moghaddasi@srbiau.ac.ir
Faculty of Engineering, Science and Research Branch
Islamic Azad University
Tehran, Iran

Paper

Availability of WDM Multi Ring Networks

Ivan Rados[1] and Katarina Rados[2]

[1] *HT d.o.o. Mostar, Mostar, Bosnia and Herzegovina*
[2] *Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia*

Abstract—This paper presents a new approach to modeling the availability of the networks composed of multiple interconnected rings in two nodes. For availability modeling method algebraic formulation is used. Using this method, through the availability of multiple ring networks consisting of two and three-rings connected to two nodes, a general expression for the availability between two terminals of multiple ring networks is derived. To perform the expression some real assumptions were taken and the analytical calculation showed that the use of these expressions under these assumptions provides real values for availability between the two terminals of multiple ring networks. Information on the availability of links and nodes is taken from previously published works.

Keywords—algebraic, availability, multiple, WDM.

1. Introduction

Telecommunications services have an important role in the development of modern society. Rapid growth in demand for data transmission and Internet traffic require high-capacity transmission systems. The only transmission media that can meet the needs of such a large capacity is an optical fiber. Unlike other media only the optical fiber has a large unused capacity that can effectively utilize the technologies based on wavelength division multiplexing (WDM).

In such networks, disconnection for any reason, i.e. equipment failure or human error, can cause isolation in terms of telecommunications and profit losses for users and network operators. Therefore, the availability of telecommunication networks is becoming a very important factor for network operators. The basic problem is how to ensure the survival of links between two nodes within large networks in terms of failure and how to ensure high availability. The development of models of reliability and availability of single selfrenewable WDM rings have been dealt with by the authors in [1] using optical add/drop multiplexers in the nodes. There are a lot of papers dealing with the calculation of the availability using the shared-path protection in which it was concluded that the use of these methods guaranteed the achievement of availability and spare capacity utilization, which is especially important for operators as service providers [2]. M. Clouqueur and W. D. Grover [3] investigated how to design a network that is able to restore the connection of all cases of single link failure between any two nodes in the ring (span protection) with less investment. They also concluded that some linked networks are very robust to the simultaneous double link failures. Some authors investigated how the availability of services depends on the total capacity and on the protection resources allocated amount necessary for mastering the protection in the event of failure to the work path [4], [5].

In this paper, the authors will develop models and derive expression for availability between the two terminals for multiple ring networks, starting from a network consisting of two and three-rings connected to two nodes examples.

2. Mathematical Basis for Formulation of Network Models

2.1. General Remarks on Availability

Availability is defined as probability that a system is correct at some point in time T, provided that it properly worked at time $T = 0$ and that defective conditions appear (maintenance or failures), but are always repeared and the system is operative again. Also, the availability of a system at point in time is defined as the ratio of time for which the system is correct in relation to the total time. The Mean Time To Failure (MTTF) is defined as $1/\lambda$, where λ is intensity of the failure. It is usually expressed in Failure In Time (FIT), and 1 FIT means 1 failure in 10^9 hours. Using Mean Time To Repair (MTTR), the availability A can be calculated by [6]:

$$A = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} . \tag{1}$$

Similarly, the unavailability U is defined as probability complementary to the availability, i.e.

$$U = 1 - A , \tag{2}$$

and it is likely that the system is not working properly at some point in time. When reporting about system/network performances, the unavailability is often expressed as Mean Down Time (MDT) in minutes per year, i.e.

$$\text{MDT} = 365 \cdot 24 \cdot 60 \cdot U . \tag{3}$$

2.2. Methods of Algebraic Formulation

Although there are several methods suitable for modeling network availability, i.e. elementary paths enumeration, in this paper the algebraic formulation method is used [6], [7].

Instead of working with probabilities, in this paper the emphasis is on algebraic objects (polynomials) and their appropriate transformation. Let a network $G = (N, E)$ be a directed network with source node s and the destination node t (N – nodes, E – edges). Suppose that the nodes in the network G are ideal and that links $m = |E|$ are subject to failures that occur randomly and are repairable. Let each link $k = 1, 2, \ldots, m$ be corresponds with the variable x_k. Furthermore, let's x_i an event for $1 \leq i \leq m$, which indicates that the link e_i is in working order then its availability is $A_i = p[x_i]$. Events $x_1, x_2, \ldots, x_m$ are independent. If x_1 and x_2 are events then $x_1 + x_2$ are union of mutually exclusive events and $x_1 \cdot x_2$ the case of two independent events. The goal is to calculate the availability of the two terminals s and t as represented by a polynomial function of x elements:

$$A_{st}(x) = A_{st}(x_1, x_2, \ldots, x_m) . \tag{4}$$

which gives the exact availability between the two terminals when the appropriate variables x_k are replaced by the appropriate links availability A_k. To calculate the availability the two new operators is introduced:

- $\otimes$ is used to represent the series combination of two or more elements,
- $\oplus$ is used to represent the parallel combination of two or more elements.

The union of arbitrary events $x_1 \oplus x_2$ and "product" of arbitrary events $x_1 \otimes x_2$ cannot be developed from numerical values, because it depends on the individual statistical correlation or the correlation between two events as opposed to union of two mutually exclusive events, or the product of independent events whose probability is easily calculated [8]. Most of the algebraic approaches for availability exact calculation between the two terminals using the fact that $\otimes$ is commutative and the fact that $x_1 \otimes x_2 = x_i$ and attenuates any repetition of events in $\otimes$ product and authors replace them by ordinary product.
Firstly, the case where links a and b are connected in series is considered. If the link failures are independent, the availability of such a structure is equal to the availability of products A_a and A_b. However, there is a possibility that links contain some common elements. Suppose, for example, that link a contains components $x_1x_3x_7$ and link b components $x_2x_3x_5$. Let A denote the event where components 1, 3 and 7 are correct and if failures of components are mutually independent, availability is $p(A) = A_1 \cdot A_3 \cdot A_7$ and let B denote the event that the components 2, 3 and 5 are correct and the availability of B is equal to $p(B) = A_2 \cdot A_3 \cdot A_5$. Now AB is an event where components 1, 2, 3, 5 and 7 are correct and one link that replaces a and b will be marked as $x_1x_2x_3x_5x_7$:

$$x_1x_3x_7 \otimes x_2x_3x_5 = x_1x_2(x_3 \otimes x_3)x_5x_7 = \\ = x_1x_2x_3x_5x_7 .$$

Replacing the variable by adequate availabilities gives the availability of a replacement link

$$A_s = p(AB) = A_1 \cdot A_2 \cdot A_3 \cdot A_5 \cdot A_7 .$$

Another operator $\oplus$ is applied to links that are connected in parallel, in this case, links a and b. If elements are working independently with an availability A_a and A_b their replacement link will have availability

$$A_a + A_b - A_a \cdot A_b .$$

This formula can be expanded to include possible dependence. For example, let link a include components $x_1x_3x_7$ and link b components $x_2x_3x_5$. Parallel structure will function properly if either link is correct, and the

$$p(A \cup B) = p(A) + p(B) - p(AB) ,$$

so that one link that replaces a and b will have the variable

$$x_1x_3x_7 \oplus x_2x_3x_5 = \\ = x_1x_3x_7 + x_2x_3x_5 - x_1x_2(x_3 \otimes x_3)x_5x_7 .$$

As $x_i \otimes x_i = x_i$ then $x_3 \otimes x_3 = x_3$ so it gets

$$x_1x_3x_7 \oplus x_2x_3x_5 = \\ = x_1x_3x_7 + x_2x_3x_5 - x_1x_2x_3x_5x_7 .$$

Replacing the variables again by the adequate availability gives the availability of alternative link

$$p(A \cup B) = A_1 \cdot A_3 \cdot A_7 + A_2 \cdot A_3 \cdot A_5 - A_1 \cdot A_2 \cdot A_3 \cdot A_5 \cdot A_7 .$$

General algebraic "sum" of two polynomials f and g is defined as:

$$f \oplus g = f + g - (f \otimes g) . \tag{5}$$

Let S represent the set of all polynomials that can arise by a combination of monomer using these two operators, then let $(S, \oplus, \otimes)$ form a distribution lattice with the smallest element 0 (zero polynomial) and the greatest element 1 (unit polynomial). For $f, g, h \in S$ to apply the following axioms [9]:

$$\begin{aligned}
&f \oplus f = f \\
&f \oplus 0 = f \\
&f \oplus g = g \oplus f \\
&f \otimes f = f \\
&f \otimes 1 = f \\
&f \otimes g = g \otimes f \\
&f \otimes (g \otimes h) = (f \otimes g) \otimes h \\
&f \oplus (g \oplus h) = (f \oplus g) \oplus h \\
&f \oplus (g \otimes h) = (f \oplus g) \otimes (f \oplus h) \\
&f \otimes (f \oplus g) = f \\
&f \oplus (f \otimes g) = f
\end{aligned} \tag{6}$$

In order to find a connection between the algebraic structures and network availability P_{st} is defined as a set of all

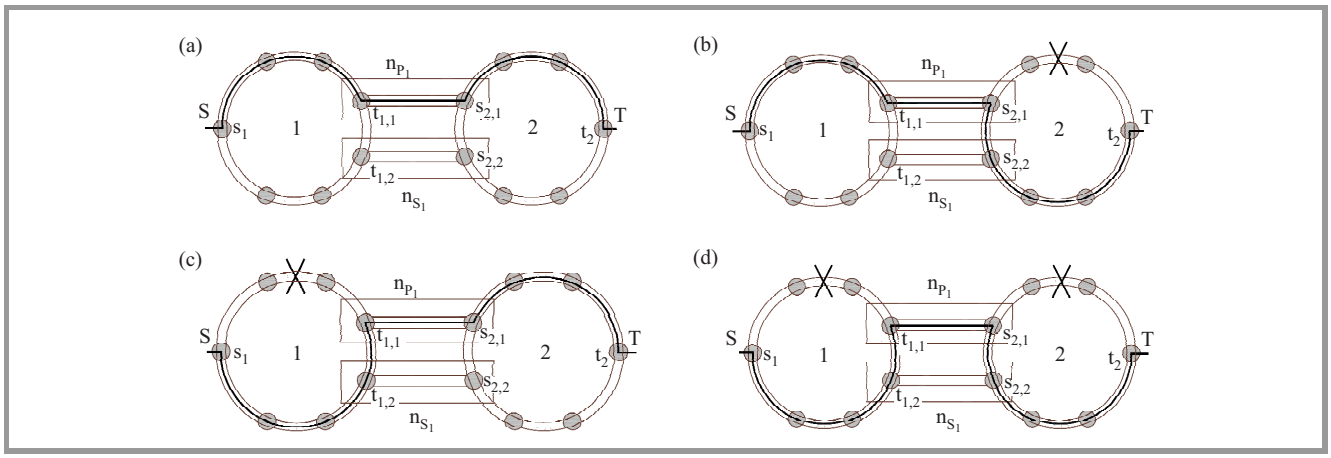


Fig. 1. Working paths and recovery: (a) without fault, (b) fault within the rings 2, (c) fault in side the ring 1, (d) fault with in two rings.

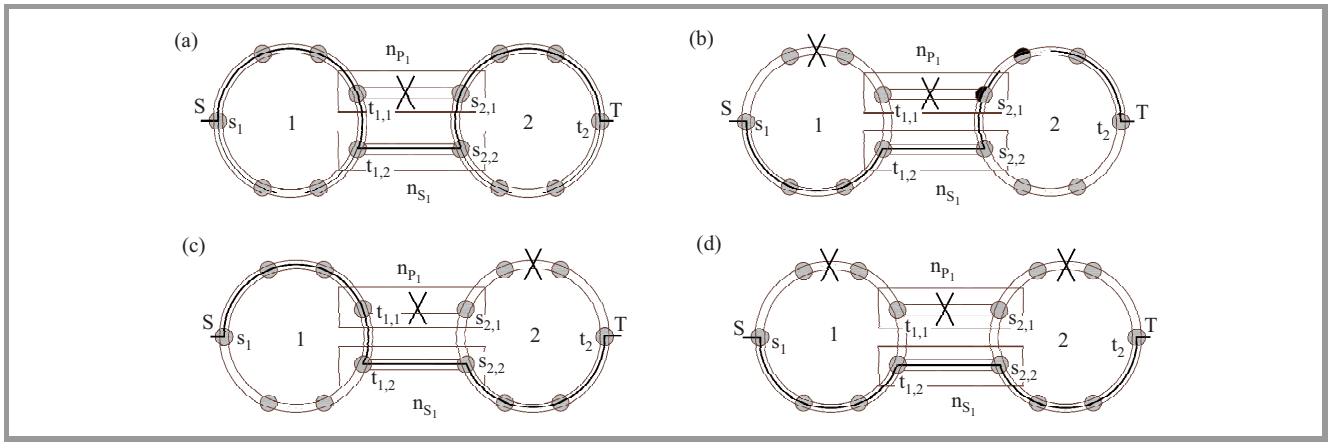


Fig. 2. Path of renewal: (a) without fault, (b) fault within the rings 1, (c) fault in side the ring 2, (d) fault with in two rings.

elementary *s-t* paths P in the network G. It also defines the value of path $v(P)$ as a variables product representing the links along the path P

$$v(P) = \otimes \prod \{x_k : k \in P\} . \tag{7}$$

Availability in this case is the sum of values $v(P)$ over all elementary paths from s to t

$$A_{st}(\mathrm{x}) = \oplus \sum_{P \in P_{st}} v(P) . \tag{8}$$

In the $\oplus$ operation each link was included as many times as the number of occurrences in the set P_{st}, of course, since this operator is $\otimes$, before applying the simple product, repetitions are eliminated.

3. Availability of Multi Ring Networks

Generally, in the multiple ring networks the transmitted signal between two nodes S and T passes N_R rings and interconnecting nodes. The path through, which the signal passes during normal operation is called a work path. When analyzing renewal paths, only those paths that include rings through, which the working path is passing are taken. Although there are more interconnection rings that could provide connectivity between nodes S and T in the case of failure on the working path, those are not analysed because path recovery requires considerable amount of time. In this way only subnets, which are comprised of interconnected rings over the working path are examined, while all other combinations are not analyzed.

In this paper, the models of availability were developed and expressions derived for the availability between two terminals of network consisting of two and three rings that are connected to two nodes. The expression valid for the network of two and three rings is also valid for multiple rings network comprised of N_R as demonstrated by mathematical induction.

3.1. Availability of Network of Two Rings

Another focus is the network comprised of two rings. As possible failure sites ring 1, ring 2 and the interconnecting

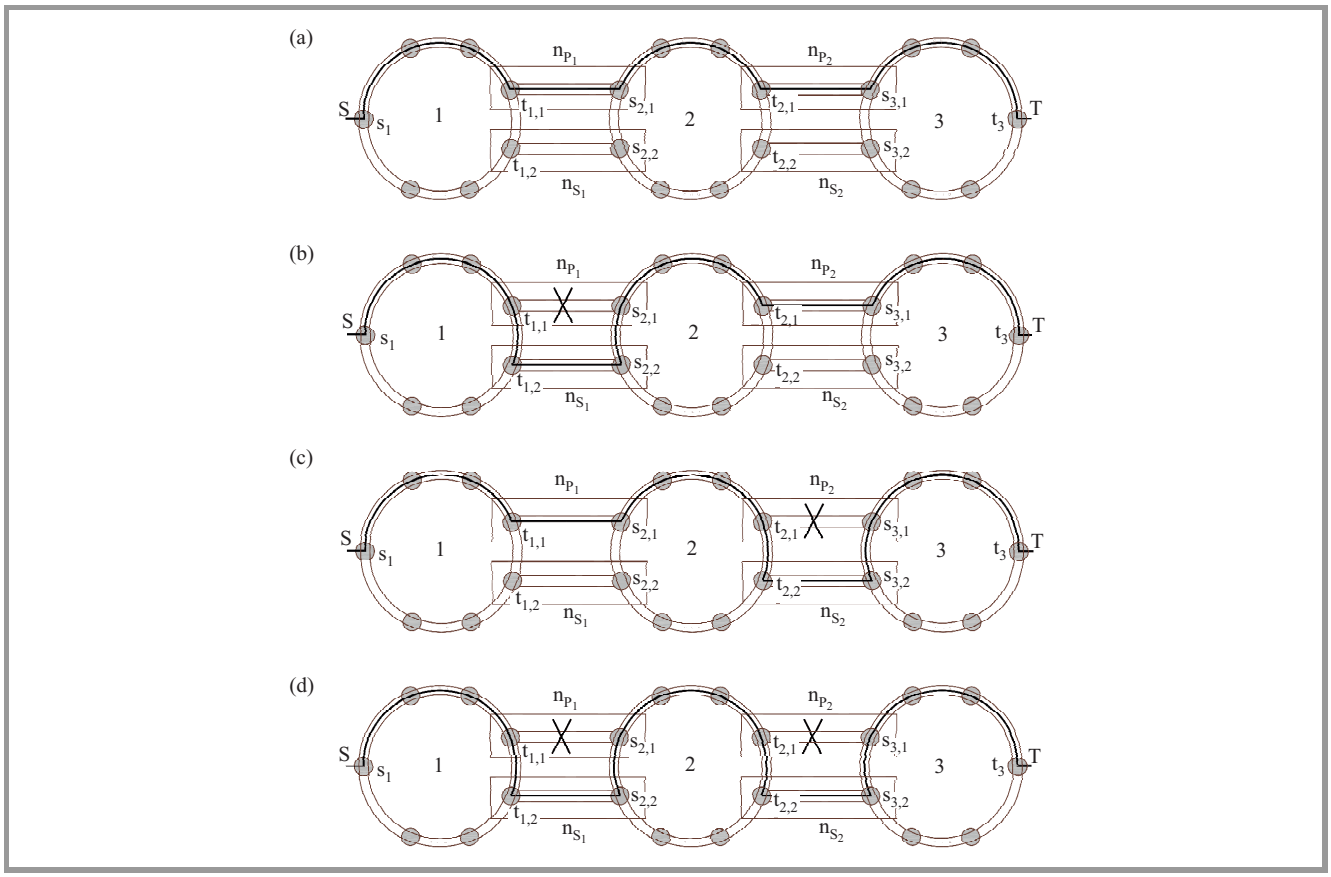


Fig. 3. Three ring network with: (a) no failures, (b) failure of the first interconnect node, (c) failure of another interconnect node, (d) failure of both interconnect nodes.

node are taken so that there is a total of 7 combinations or seven renewal paths:

$$\binom{3}{1}+\binom{2}{2}+\binom{3}{3}=3+3+1=7\,.$$

Renewal paths between nodes S and T need not be mutually exclusive because a parallel sum eliminates identical (duplicate) paths. For a network comprised of two rings there is a work path with no failures and 7 renewal paths which are the result of failures at work paths of rings 1 and 2 and primary interconnecting node. Suppose that each ring is comprised of N links where there are m working and that the connection between $t_{1,1}-t_{1,2}$, and $s_{2,1}-s_{2,2}$ is achieved with one link.

Figure 1 shows working path and recovery paths, but with always correct primary interconnection node.

Figure 2 shows the paths of renewal, but always with a faulty primary interconnect node.

The value of path in failure-free (P_0) case is

$$v_{st}(P_0)=(x_{l_{r1}})^m\otimes x_{P1}\otimes(x_{l_{r2}})^m\,,$$

where $x_{l_{r1}}$ and $x_{l_{r2}}$ denote links that belong to the first (index r_1) and second ring (index r_2).

For each of these 7 paths the renewals are also defined as the values of paths that are marked with $v_{st}(P_1)$ (Fig. 1b) to $v_{st}(P_7)$ (Fig. 2d).

Availability is a "parallel sum" of path values when there are no failures and values of all renewal paths, which are the consequence of failures, and therefore

$$A_{s_1t_2}(\mathrm{x})=\oplus\sum_{P_i\in P_{st}}v_{st}(P_i)\,.$$

Since the true availability between two nodes of an individual ring for the standard size of rings is approximately equal to the minimum availability [10], so in this case availability between the two terminals for the network composed of two rings is

$$A_{ST}(\mathrm{x})=(X_{\min_1}\cdot X_{\min_2})\otimes\left[x_{P1}+x_{S1}-(x_{P1}\otimes x_{S1})\right]\,.$$

Assuming that the rings and interconnecting nodes also have no common elements, sign $\otimes$ could be elided and the common sign of multiplication might be used

$$A_{ST}(\mathrm{x})=\left(\prod_{i=1}^{2}X_{\min_i}\right)\cdot\left[x_{P1}+x_{S1}-x_{P1}\cdot x_{S1}\right]\,.$$

3.2. Availability of Network of Three Rings

A similar analysis can be implemented for a network consisting of three rings as shown in Fig. 3, taking into account all the renewal paths that can occur as a result of

failures of each ring parts, the interconnections of nodes and all their possible combinations. If the number of possible failures is 5, the number possible combinations total is

$$\binom{5}{1}+\binom{5}{2}+\binom{5}{3}+\binom{5}{4}+\binom{5}{5}=5+10+10+5+1=31\,.$$

To avoid examining all possible renewal paths, the authors take into consideration confirmed fact in Subsection 3.1 that all failures of individual rings work path renew within themselves, so that the analysis is brought down to the failures of interconnection nodes, which are only two and therefore the combinations number decreases significantly:

$$\binom{2}{1}+\binom{2}{2}=2+1=3\,.$$

If it is assumed that all the interconnecting nodes of the same type are equal (identical equipment) and assuming that the rings and interconnecting nodes have no common elements, simple multiplication sign can be written instead of sign $\otimes$

$$A_{ST}(\mathrm{x})=\left[\prod_{i=1}^{3}X_{\min_i}\right]\cdot\left[x_P+x_S-x_P\cdot x_S\right]^2.$$

If the analysis is conducted for a network consisting of $4, 5, \ldots, N_R$ rings, there would be a general expression for the availability of two terminals of multiple ring network in which the rings are interconnected in two nodes:

$$A_{ST}(\mathrm{x})=\left[\prod_{i=1}^{N_R}X_{\min_i}(\mathrm{x})\right]\cdot\left[x_P+x_S-x_P\cdot x_S\right]^{N_R-1}, \qquad (9)$$

which means that the availability between two nodes of multiple ring network is equal to the product of minimum availability of individual rings and the interconnections of nodes availability.

4. Numerical Results

Analysis results are presented on a network consisted of two rings to show that the difference is negligible when the actual and minimum availability for individual ring is calculated. Availability data for optical components shown in Table 1 are taken from [10].

Table 2 shows that for the typical size of rings (6–10 nodes), difference between the actual availability and the one obtained by taking a minimum availability for each individual ring are in the worst case of 0.0245 min/year, which is negligible so that the Eqs. (8) and (9) can be used to calculate availability. Note: W is the number of wavelength channel.

Even for larger rings with 14 and 16 nodes, those differences were of 0.1 min/year, which is also negligible. This is especially important for SLA contracts which set a minimum availability threshold and charge penalties for exceeding them.

Table 1
Availability data for optical components

Component	Symbol	Failure rate [FIT]
Booster amplifier	BOA	3200
Line amplifier	LOA	3200
Preamplifier	POA	3200
Multiplexer	MUX	25×W
Demultiplexer	DEMUX	25×W
Optical switch	OSW	1000
Fix transmitter	TRX	186
Tunable transmitter	TX	745
Fix receiver	RX	70
Tunable receiver	RCX	470
Switch	SW	50
Splitter	SPL	50
Cable (per km)	OC	100

5. Conclusion

This paper presents a new approach to modeling the availability of multiple ring network using an algebraic formulation of working with numeric values. It is shown that the paths need not be mutually exclusive because the use of operator $\otimes$ eliminates duplicate routes. The analysis showed that the availability calculation of multiple ring network, in which the rings are connected to two nodes, is actually reduced to a product of minimum availability of individual rings and the availability of the nodes interconnections. This is done with some realistic assumptions and these are: the rings do not have common elements, the interconnection nodes also have no elements in common and neither rings nor interconnecting nodes have also any common elements which are the most common cases in practice, otherwise the protection would not make sense. The analytical results have confirmed the theoretical analysis so that the derived general expression for the availability of the two terminals can be used to analyze and evaluate the multiple ring networks availability.

References

[1] D. Arci, G. Maier, A. Pattavina, D. Petecchi, and M. Tornatore, "Availability models for protection techniques in WDM networks", in *Proc. 4th Int. Worksh. Design of Reliable Commun. Netw. DRCN 2003*, Banff, Alberta, Canada, 2003, pp. 158–166.

[2] L. Song, J. Zhang, and D. Mukherjee, "Dynamic provisioning with availability for differentiated service in survivable mesh networks", *IEEE J. Selec. Areas Commun.*, vol. 25, no. 4, 2007.

[3] M. Clouqueur and W. D. Grover, "Dual failure availability analysis of span-restorable mesh networks", *IEEE J. Selec. Areas Commun.* (Special Issue on Recent Advances on Fundamentals of Network Management), vol. 20, no. 4, pp. 810–821, 2002.

[4] M. C. Arie and A. Koster, "Cost-efficient transparent optical networks with high connection availabilities", in *Proc. 8th Int. Conf. Transparent Optical Netw. ICTON 2006*, Nottingham, UK, 2006, vol. 3, pp. 101–104.

[5] W. Yao and B. Ramamurthy, "Survivable trafficgrooming with differentiated end-to-end availability guarantees in WDM mesh networks", in *Proc. 13th IEEE Worksh. Local and Metropol. Area Netw. IEEE LANMAN 2004*, San Francisco, CA, USA, 2004, pp. 87–90.

Table 2
Comparison of availability between S and T when using the actual availability of the ring and the minimum availability of the ring

Number of ring nodes			The actual availability	Minimum availability	Difference MDT [min/year]
	$N=6$	$N=10$			
No. links of working path	$m=1$	$m=7$	0.9999752896	0.9999752482	0.0217901963
No. links of working path	$m=3$	$m=8$	0.9999752948	0.9999752482	0.0245122162
No. links of working path	$m=5$	$m=3$	0.9999752741	0.9999752482	0.0136182825
	$N=14$	$N=116$			
No. links of working path	$m=2$	$m=13$	0.9999750980	0.9999748389	0.1361543751

[6] L. Schwartz, D. Trstensky, and G. Cepciansky, "Reliability of telecommunications systems", University of Zilina, Slovakia, 2010, p. 193 (in Slovak).

[7] D. R. Shier, *Network Reliability and Algebraic Structures*. New York: Oxford University Press, 1991.

[8] J. Jonczy, "An algebraic approach to network reliability", Institute of Computer Science and Applied Mathematics, University of Berne, Switzerland, Feb. 2008.

[9] D. D. Harms, "A symbolic algebra environment for research in network reliability", Ph.D. thesis, Simon Fraser University, Burnaby, Canada, Sept. 1992.

[10] I. Rados and L. Schwartz, "The worst availability as a parameter for designing and reporting on the network performances", *Communications – Scient. Lett. of the Univ. of Zilina*, vol. 13, no. 1, pp. 101–104, 2011.

Ivan Rados received the B.Sc. degree in Electrical Engineering from the University of Split, in 1983, and M.Sc. degree from the University of Zagreb, in 2000. He received his Ph.D. from the University of Split, Croatia in 2012. In 1985 he joined the PTT (Post and Telecommunication) Tomislavgrad. Since 1992 he works at Department of Transmission Systems of the HT Mostar (Croatian Telecommunication). His research interests include: digital transmission systems, optical systems and networks, availability and reliability of telecommunication systems. He has published 10 papers in international conference proceedings, 4 papers in domestic journals and 7 papers in international journals.
E-mail: ivan.rados@hteronet.ba
HT d.o.o. Mostar
Kneza Branimira bb
88000 Mostar, Bosnia and Herzegovina

Katarina Rados undergraduated Electrical Engineering and Information Technology on Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, in 2014, and she gained the academic title: Bachelor of Science in Electrical Engineering and Information Technology. At the moment, she is studying second level (graduate university study program) Telecommunication and Information Technology on the same faculty. She has published a paper in domestic scientific proceedings.
E-mail: krados@fesb.hr
Kroz Smrdecac 47
21000 Split, Croatia

Paper

Assessment of Area Energy Efficiency of LTE Macro Base Stations in Different Environments

Suhail Najm Shahab[1], Ayad Atiyah Abdulkafi[2], and Ayib Rosdi Zainun[1]

[1] *Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Pahang, Malaysia*
[2] *College of Engineering, Tikrit University, Salahaldin, Iraq*

Abstract—Energy efficiency (EE) of wireless telecommunications has become a new challenge for the research community, governments and industries in order to reduce CO_2 emission and operational costs. EE of base stations (BSs) in cellular networks is a growing concern for cellular operators to not only maintain profitability, but also to reduce the overall negative impact to the environment and economic issues for wireless network operators. In this paper, a framework focuses on the Area Energy Efficiency (AEE) evaluation of LTE BSs is presented. The parameters affect on the AEE and the coverage area of LTE BS in different scenarios are investigated. AEE analysis has been done using a few key performance indicators including transmit power, bandwidth, load factor with the assumption of different scenarios (urban, suburban and rural). The simulation results show that the LTE BSs have better AEE in an urban environment for cell radius less than 750 m compare with the suburban and rural environments. Furthermore, it is obvious that there is a strong influence of traffic load, BW and transmission power on AEE of LTE network. On the other hand, AEE increases significantly as the BW size increases. Finally, it has been shown that the AEE of LTE macro BS decreases with increasing the percentage of traffic load for all scenarios.

Keywords—Area Energy Efficiency, LTE, Macro Base Station.

1. Introduction

Addressing the issue of green communications has benefits to many stakeholders including the industry, academic researchers and government agencies. The cellular industry can realize cost savings and lower their impact to the environment, government agencies realize a fulfillment of administrative goals for energy savings as well as development of standards and metrics, while researchers can push the boundaries of current technologies and theories in material science, distributed computing and system engineering. Telecommunication section and especially cellular networks are parts of Information and Communication Technology (ICT) that is rapidly expanding throughout the globe. With new technologies like Third Generation (3G) and Long Term Evolution (LTE) coming to the market, this section will grow more in a future.

Currently, telecommunication sectors are responsible for about 12% of total energy consumption of the world and generates approximately 1% of CO_2 emissions [1] with percentages expected to rise further. In [2] the authors proposed the deployment of LTE macro base station (BS) to study the impact of modulation and coding schemes (MCS), bandwidth (BW) size and transmitted power on the energy efficiency for urban environment. They showed that the higher transmission power results in lower EE. The difference actually diminishes when cell size increases. At its diameter around 1200 m, it was found that the EE is almost equal for all transmission power considered. On the other hand, EE increases significantly as the BW increases. Similar effect on EE is observed when MCS changes from lower order to the higher-order scheme. In cellular networks, the prime energy users are base stations (BSs), backhaul servers and routers. Around 80% energy is consumed by the BSs [3]. Because of this statistic, most of the energy saving research had been focused on the BS.

This paper investigates the area energy efficiency (AEE) issue on LTE networks and more specifically it is based on simulation for the outdoor environments. The environment's scenarios for the simulation were done with three different environments: urban, suburban and rural. Results have conducted and discussed to show the performance of LTE network from the AEE perspective. A comparison analysis is done in terms of energy saving for a specific macro BS deployment between the three different scenarios.

2. Methodology

2.1. Propagation Model

In general, there are many factors that cause the deterioration of signal quality such as distance dependent path losses, shadowing, outdoor/indoor penetration loss and radiation pattern. The received power P_{rx}, from a BS at a distance of d and angle θ from the main lobe of the antenna can be calculated as [4]:

$$P_{rx}(d,\theta,\psi) = P_{tx} - \left[PL(d) + \kappa + A_h(\theta)\right] + \psi_{dB}, \qquad (1)$$

where P_{rx}, P_{tx}, PL, κ, A_h, ψ and θ denote to receive power and transmit, path loss, penetration loss, antenna radiation pattern, shadow fading and theta, respectively.

Equation (1) assumes that all the signal gains and losses are expressed in decibels. The random variable ψ is used

to model slow fading effects and commonly follows a log normal distribution. The antenna pattern $A_h(\theta)$ depends on the mobile's location relative to the BS which has been adopted from [4]. In addition to path loss and shadowing, another factor which affects the channel quality is penetration loss for users indoors.

In this paper, a 20 dB of attenuation has been assumed to account for outdoor/indoor penetration loss, denoted by κ, which can be found in [5] and [6]. The path loss PL in decibels (dB) for a distance d can be expressed into three different categories, namely urban, suburban and rural areas, which take into account distance, line of sight existence, antenna height, and the average building height with the applicability ranges from 5 to 50 m as proposed in [5] for all environments.

However, the urban scenario usually has a great concentration of BSs due to the demand for capacity. The path loss in urban scenario before the break point d_{BP} can be written in the following form:

$$PL = 22.0\log_{10}(d) + 28.0 + 20\log_{10}(f_c), \tag{2}$$

where d is the distance in meters, and f_c is the carrier frequency in GHz. After d_{BP}, the path loss is founded via:

$$PL = 40.0\log_{10}(d) + 7.8 - 18\log_{10}(h'_{BS}) - \\ - 18\log_{10}(h'_{UT}) = 2\log_{10}(f_c), \tag{3}$$

where h'_{BS} and h'_{UE} are the effective antenna heights at the BS and the User Equipment (UE). The effective antenna heights h'_{BS} and h'_{UE} are computed as follows: $h'_{BS} = h_{BS} - 1.0$ m, $h'_{UE} = h_{UE} - 1.0$ m, where $h_{BS} = 25$ m and $h_{UE} = 1.5$ m are the actual antenna heights proposed in [5] for urban area.

The suburban scenario is modeled to correspond to typical city's periphery with major habitation blocks with several floors. While the remaining territory corresponds to rural low dense populated scenarios that can be crossed by important highways. The path loss in suburban and rural scenarios before the d_{BP} can be written as:

$$PL = 20\log_{10}\left(\frac{40\pi d f_c}{3}\right) + \\ + \min(0.03h^{1.72}, 10)\log_{10}(d)\min(0.044^{1.72}, 14.77) + \\ + 0.002\log_{10}(h)d\,. \tag{4}$$

While after d_{BP}, the path loss for these two scenarios is founded via:

$$PL = 20\log_{10}\left(\frac{40\pi d f_c}{3}\right) + \\ + \min(0.03h^{1.72}, 10)\log_{10}(d)\min(0.044^{1.72}, 14.77) + \\ + 0.002\log_{10}(h)d + 40\log_{10}\left(\frac{d}{d_{BP}}\right). \tag{5}$$

Here h is building height in meters.

2.2. Cell Coverage Area

The cellular system coverage is generally designed for a given minimum received power $P_{\min}$ at the cell boundary. The $P_{\min}$, which is also known as the receiver sensitivity can be written in closed-form for cell coverage area C as [7]:

$$C = Q(a) + \exp\left(\frac{2-2ab}{b^2}\right)Q\left(\frac{2-ab}{b}\right), \tag{6}$$

where:

$$a = \frac{P_{\min} - P_{rx}(R)}{\sigma_{\psi dB}}, \quad b = \frac{1-\alpha\log_{10}(e)}{\sigma_{\psi dB}}, \tag{7}$$

where α denote to path loss exponents and σ_{dB} is the standard deviation of shadow fading [7].

The reference sensitivity $P_{\min}$ level is the minimum mean received signal strength applied to both antenna ports at which there is sufficient SINR for the specified modulation scheme to meet a minimum throughput requirement of the maximum possible. It is measured with the transmitter operating at full power. $P_{\min}$ is a range of values that can be calculated using the Eq. (8) [8]:

$$P_{\min} = \mathrm{kTBW} + \mathrm{NF} + \mathrm{SINR}_{req} + \mathrm{IM} - \mathrm{G_d}\,, \tag{8}$$

where $kTBW$ is the thermal noise level in units of dBm, in the specified bandwidth (BW), NF is the prescribed maximum noise figure for the receiver where LTE defines an NF requirement of 9 dB for the User Equipment (UE), SINR_{req} is Signal to Interference plus Noise Ratio that required for choosing modulation and coding scheme, IM is the implementation margin and $\mathrm{G_d}$ represents the diversity gain which is equal to 3 dB [8]. $P_{\min}$ is a target minimum received power level below which performance becomes unacceptable [7]. Note that $a = 0$, when the target minimum received power equals the average power at the cell boundary, $P_{\min} = P_{rx}(R)$ and $P_{rx}(R)$ is the received power at the cell boundary due to path loss alone. An extra implementation margin is added to reflect the difference in SINR requirement between theory and practicable implementation [8].

2.3. LTE Data Rate Model

Theoretical peak data rates are difficult to achieve in practical situations only in extremely good channel conditions because of limited by the amount of channel impairments noise and interference from own and other cells. The maximum theoretical data rate for single antenna transmission in static channel can be derived through conventional Shannon's formula which is given in Eq. (9). The data rate R_T in unit of bits per second can be expressed in terms of two parameters which are the bandwidth and the signal to noise ratio SNR.

$$R_T = \mathrm{BW} \times \log_2(1 + \mathrm{SNR}). \tag{9}$$

In LTE system, a modified Shannon formula is used to accurately estimate the achieved data rate after taking channel impairments into account.

$$R_T = F \times \mathrm{BW} \times \log_2\left(1 + \frac{\mathrm{SINR}_{req}}{\eta_{\mathrm{SNR}}}\right). \tag{10}$$

where $F = \eta_{BW} \cdot \eta$ in which the η_{BW} accounts for the system bandwidth efficiency of LTE and η_{SNR} accounts for the SNR implementation efficiency of LTE. It should be noted that LTE is performing less than $1.6 \sim 2$ dB from the Shannon capacity bound because it's not constant and changes with the geometry factor (G-factor), the G-factor distribution is defined as the average own cell power to the other cell power plus noise ratio with OFDMA in a wide system bandwidth this corresponds to the average SINR [8]. It was shown that this impact can be accounted for using the fudge factor (η) multiplying by the parameter (i.e. $\eta = 0.9$ and $\eta_{BW} \cdot \eta = 0.75$). η_{SNR} is a parameter for adjusting SNR efficiency which is almost equal to one [9].

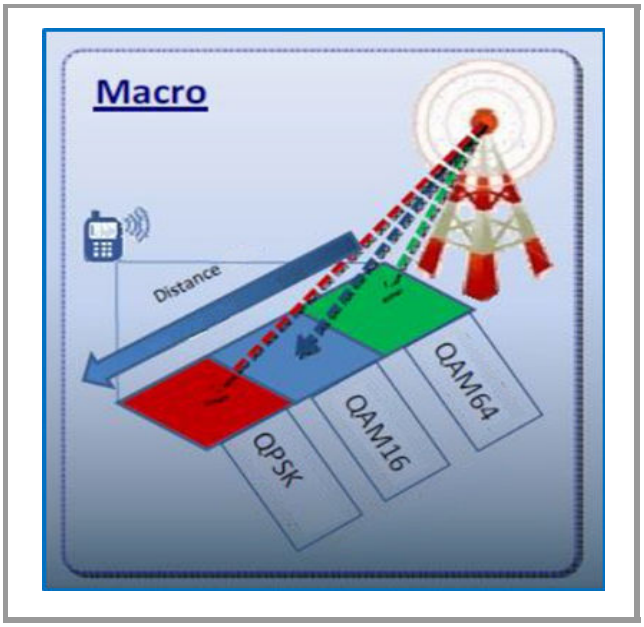


Fig. 1. MCS selected based on user distance.

The MCS selection is depend on the distance between the eNodeB and the UE. The low MCS can be suitable for large distances as the experienced SINR is low while the higher MCS is preferred at short ranges with high data rate demands. Figure 1 shows how the different MCS are selected according to the distance between eNodeB and UE based on the received SINR.

2.4. LTE Power Consumption Model

The main goal of the power consumption model in this paper is to make realistic input parameters available for the simulation. This model also allows fair comparing between different environments and different macrocell BS deployments. The power models have been selected from [10], [11] for different environments cases for LTE deployment.

The power model of macro BS described in [10] has a linear relationship between average radiated power per site and average power consumption. The power consumption calculation is modified to be changed according to the traffic load level and the BS components features. The consumed power P_c by the BS i can be expressed as:

$$P_c = L \cdot N_{sec} N_{ant} (\mathrm{A} P_{tx} + \mathrm{B}), \tag{11}$$

where $L \in [0, 1]$ is the load factor and N_{sec} and N_{ant} denote the BS's number of sectors and the number of antennas per sector, respectively. P_c and P_{tx} denote the total power per BS and the power fed to the antenna, respectively. The coefficient A accounts for the part of the power consumption that is proportional to the transmitted power (e.g., radio frequency amplifier power including losses caused by feeders and cooling of sites), while B denotes the power that is consumed independent of the average transmit power and models the power consumed (e.g., signal processing, site cooling, backhaul, and as well as a battery backup) [10], [12]. Both these coefficients are constant for macro BS. The power model is calculating power consumption with respect to transmit power P_{tx} this assumption is valid because currently deployed macro sites power consumption depends upon the traffic load [10]. The parameter L models the activity level of the BS which describes the portion of resources which are allocated for transmission, where zero and full load correspond to no active user in the cell and providing one or more users with all resources available, respectively.

However, it may be unsuitable to observe only power consumption for comparing the networks with different site densities. This is because they may have different coverage's. In order to assess the power consumption of the network relative to its size, the notion of area power consumption APC measured in [W/km^2] is introduced as the total power consumption in a reference cell divided by the corresponding reference area [10], [13]:

$$APC = \frac{P_c}{A_{macro}}, \tag{12}$$

here A_{macro} is the macro reference area which can be expressed as [10] and [13]:

$$A_{macro} = \frac{3\sqrt{3}}{2} d^2. \tag{13}$$

It was shown that for a hexagonal deployment the area power consumption metric yields an optimal coverage cell size [10].

2.5. Energy Efficiency

The extrapolation of current trends undertaken by many literatures reveals that for a sustainable growth of wireless communications, an improvement of LTE energy efficiency is required. In this study, energy efficiency assessing a framework is studied via network level simulations.

The total network energy efficiency EE_T which is defined as the ratio of total amount data delivered and the total power consumed measured in bits per Joule [14], is represented by:

$$EE_T = \frac{\sum_{i=1}^{N_{BS}} R_i}{\sum_{i=1}^{N_{BS}} P_{c_i}}. \tag{14}$$

where P_c is the power consumption and R_i is the total data rate with a BS i. N_{BS} is the total number of BSs. As know, cell coverage is a primary concern in the design of wireless data communications systems. Increased inter-site distances (ISDs) generate larger coverage areas. With the same transmission power, different cell size can lead to different individual date rate and accordingly various energy efficiency. Therefore, observing the mere energy efficiency per site is not enough for comparing networks with different cell size. Moreover, another important metric is used through this research to evaluate the energy efficiency of the network relative to its size. The Area Energy Efficiency (AEE) metric which is defined as a bit/Joule/unit area is used as a performance indicator metric. The AEE for LTE network can be expressed as [15]:

$$\text{AEE} = \frac{EE_T}{A_T}, \tag{15}$$

where the aforementioned EE_T and A_T are the total energy efficiency and total area of LTE network respectively.

3. Simulation Procedure and Results

The EE performance of the network corresponding to its size and deployment can be more accurately assessed by comparing the AEE performance under different sector radius and scenarios. In the following subsections, the LTE performance in terms of AEE is presented. Furthermore, the effect of environment type on AEE is demonstrated. Later, by considering different traffic load scenarios, the impact of traffic load on AEE has been explained and discussed. The parameters that are affecting the AEE of LTE macro BS are investigated. The impacts of parameters like different traffic load, BW and P_{tx} on AEE.

3.1. Simulation Procedure

In this section, the simulation procedure and system parameters are discussed. There are three different environments are chosen for study campaigns one is an urban type environment. The second is a suburban site like a small city while the third is with a rural environment. Single LTE macro BS covers a hexagonal shaped area as shown in Fig. 2 in which R is the cell radius and A_{macro} is the coverage area.

The cell size is determined according to the minimum received power level constraints. The receiver sensitivity is calculated based on sufficient SINR for the specified

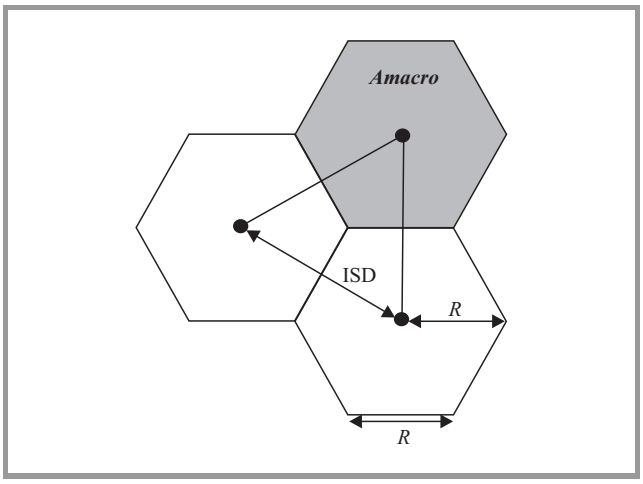


Fig. 2. Corresponding cell geometry.

modulation scheme to achieve a minimum requirement of 95% coverage degree. The received SNR is calculated based on the received power level and white noise which are estimated according to the path loss model described in 3GPP TR 36.814 [5]. Then, the achievable data rate within each BS's coverage area is determined based on the SNR distribution in the cell. The power consumption models consist of dynamic power consumption which is fully depended on the traffic load as expressed in Eq. (11). The simulation parameters are based on 3GPP macrocell model with a system bandwidth of 10 MHz with UE height of 1.5 m. The 2.6 GHz spectrum band is used since this is the band allocated to LTE operators in Malaysia [16]. Effective environment height which is subtracted from the actual antenna height for BS and UE to find their effective antenna heights is assumed to be equal to 1 m. IM of 2.5 dB is assumed for all QPSK modes, while 3 dB and 4 dB are generally expected for 16QAM and 64QAM respectively [17]. However, the typical assumptions for the SINR values for different MCS that are used in the simulation assumptions equal the ones in [8]. The proposed simulation model for evaluating the EE in LTE macro BS in different environments is an extension of the work in [18] as shown in Fig. 3.

Table 1
Simulation parameters

Notation	Description	Default
f_c	Carrier Frequency [GHz]	2.6
BW	Bandwidth [MHz]	10
N_{sec}	Number of sectors	3
N_{ant}	Number of antennas	2
MCS	Modulation Coding Scheme	1/3 QPSK [8]
P_{tx}	Transmit Power [dBm]	46
G_d	Diversity gain [dB]	3 [8]
C	Coverage degree	95%
NF	Noise Figure [dB]	9 [8]
A_i B_i	Power consumption parameters [W]	21.45 354.44 [11]

Various parameters have been used in all simulation scenarios to analyze the EE behavior under specific circumstances. Simulation parameters are listed in Table 1 and simulation procedure flow chart shown in Fig. 3.

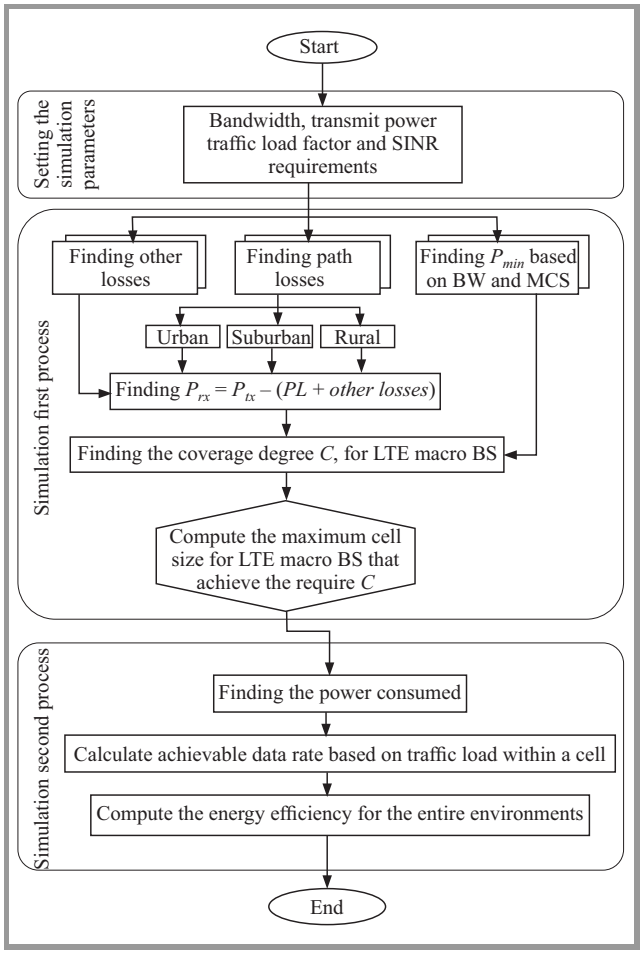


Fig. 3. Simulation model flow chart.

3.2. Simulation Results

AEE for three scenarios at full load. There are different coverage area sizes of LTE BSs due to the deployment environments, there are different data rates for each BS in each environment according to its size and therefore different EE's. Thus, the AEE is used to evaluate the EE of LTE network relative to its size. The AEE has been calculated based on a Eq. (15). In Fig. 4, the AEE versus cell radius for three environments with full load (100%) is plotted. It is obvious that AEE decreases as the macrocell BS's radius increases. Moreover, it can be shown that the LTE BSs have better AEE in urban environment with cell size less than 750 m. For cell radius more than 750 and 1500 m, the LTE performance becomes better in suburban and rural environments respectively. More specifically, at the first 700 m the better AEE is can achieve in urban area but at 710 m the suburban area becomes better than urban and rural, also at 1055 m the rural area became better than urban areas as shown in Fig. 4. This is because the impact of shadowing, path losses as well as the penetration

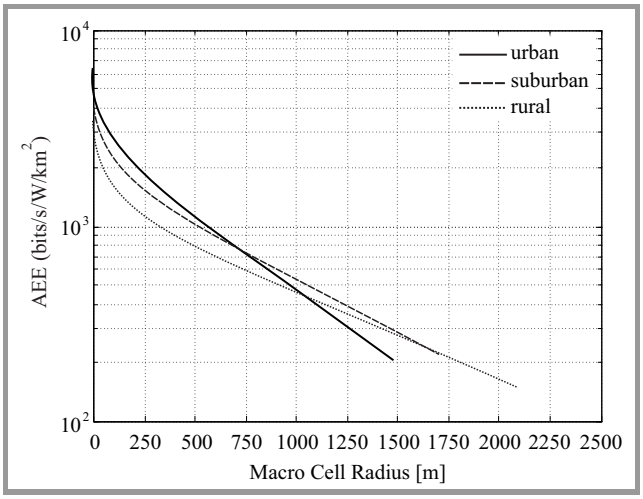


Fig. 4. AEE versus cell radius for three environments.

losses has become more significant in the urban area at long distances as compared with the rest environments.

AEE for three scenarios with different loads. The traffic load is another important factor that affects the network performance. It has a stronger impact on the data rate and the power consumption of LTE network and subsequently on its EE and AEE. The AEE versus cell radius for urban area under different loads shown in Fig. 5. It is clear that

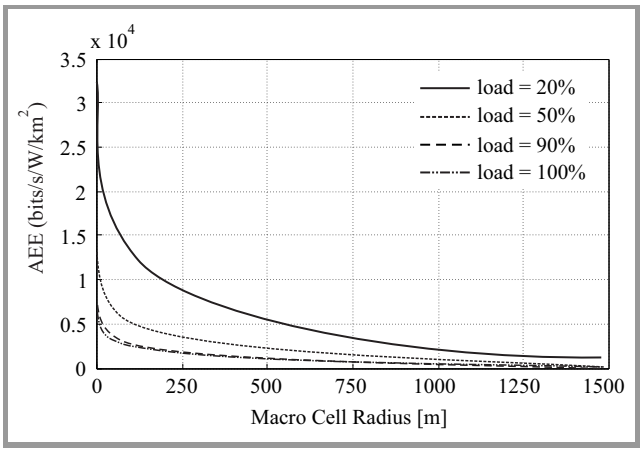


Fig. 5. AEE versus cell radius for urban environment with different loads.

the AEE decreases as the traffic load increases. In fact, the AEE's become almost equals as the traffic loads increased as shown in Fig. 5 the curve with traffic load 90% are very closed to the curve with a full traffic load scenarios. Moreover, it can be shown for all environments that the AEE decreases as the traffic load increases due to increasing in power consumption. The same AEE performance can be concluded for suburban and rural areas when varying the traffic load as shown in Figs. 6 and 7 respectively.

Table 2 summarizes the AEE performance for the three types of environments with different traffic load conditions at 100, 1000 and the cell edge for each environment. As

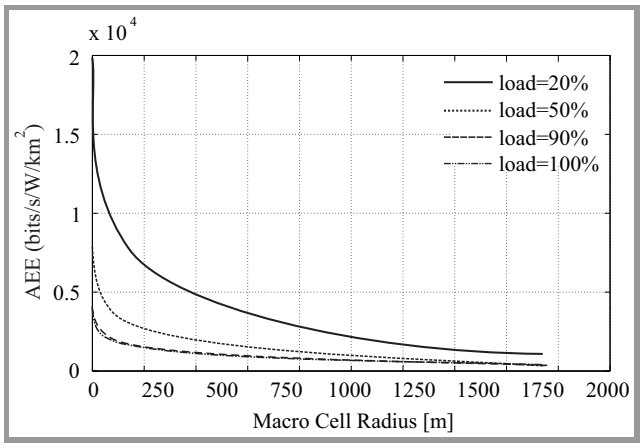


Fig. 6. AEE versus cell radius for suburban environment with different loads.

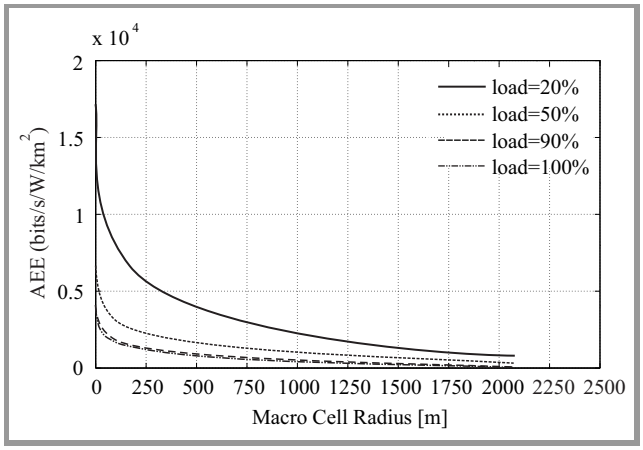


Fig. 7. AEE versus cell radius for rural environment with different loads.

mentioned before, the AEE of LTE macro BS at short distances is better for urban area than suburban and rural for all load conditions. As shown in Table 2 for cell radius more than 1000, the LTE performance becomes better in suburban and rural environments respectively.
Table 3 shows the LTE performance (BW = 10 MHz, 1/3QPSK, full load) in terms of AEE for different transmission powers. However, the AEE decreases as the transmission power increases for the same environment. In addition, it can be concluded that the suburban area achieved better AEE performance due to its suitable cell size compare to urban and rural areas.

Table 2
Training and classification times

Environments	Distance [m]	AEE [bits/s/W/km^2]			
		20% load	50% load	90% load	Full load
Urban	100	13180	5264	2929	2637
	1000	2416	966.5	536.9	483.2
	1475.7	1041	416.4	231.3	208.2
Suburban	100	10780	4311	2395	2155
	1000	2751	1100	611.4	550.2
	1718.1	1107	442.9	246.1	221.4
Rural	100	7841	3136	1742	1568
	1000	2324	929.6	516.5	464.8
	2074.9	758.1	303.2	168.5	151.6

Table 3
AEE at cell edge for different P_{tx}

Environment	P_{tx} [dBm]		
	43	46	49
Urban	453.7886	208.1916	86.3716
Suburban	482.2614	221.4458	92.0362
Rural	330.7610	151.6243	63.1749

Table 4
AEE at cell edge for different bandwidth

Environment	BW [MHz]					
	1.4	3	5	10	15	20
Urban	8.642	34.101	73.337	208.191	382.168	586.726
Suburban	9.192	36.316	78.206	221.445	407.545	626.886
Rural	6.308	24.962	53.715	151.624	279.118	429.570

Increasing the BW for any type of environment will increase the AEE of LTE macro BS. In fact, the better outcomes can be predicted for suburban area while the urban area comes in the next order and finally the rural area as demonstrated in Table 4 ($P_{tx} = 46$ dBm, 1/3QPSK, full load).

4. Conclusion

One of the most important requirements for wireless communication technologies is to be applicable and universally desirable. AEE for LTE macro BS analysis is the main target for this paper. It is considered as the most important process to achieve mobility within wireless networks. Work evaluation has been done by simulating AEE assessing with different scenarios. Three different environments were chosen for this study including urban, suburban and rural. A framework for evaluating the AEE of LTE network in different environments has been proposed. Using few key performance indicators such as coverage size, area power consumption, energy efficiency and area energy efficiency, the network performance from EE perspective for all the three urban, suburban and rural terrains are compared and evaluated. Although, the LTE BSs have large cell size and good coverage degree in rural areas, the simulation results show that they have better AEE in urban environment with small cell sizes while the AEE becomes better in suburban and rural environments for larger cell radius. Also, it can be concluded that there is a strongly impact of traffic load, bandwidth and transmission power on APC and AEE of LTE macrocell networks. For all the three environments, it has been shown that the AEE of LTE macro BS decreases with increasing the traffic load and this effect becomes the same at high loads. Using the proposed framework, the EE of different deployment scenarios can

be evaluated and insights on how to deploy a greener LTE network are provided. The results presented in this work consider only one LTE BS and therefore the impact of the handovers and interference in the LTE network may bring substantial impact on the AEE. These issues have been left for author's future works.

Acknowledgment

This work is supported by University Malaysia Pahang.

References

[1] P. Misar, "Wireless LTE deployment: How it is changing cell site energy and infrastructure design", in *Proc. 32nd IEEE Ann. International Telecommun. Energy Conf. INTELEC 2010*, Orlando, FL, USA, 2010, pp. 510–514.

[2] A. A. Abdulkafi, T. S. Kiong, J. Koh, D. Chieng, A. Ting, and A. M. Ghaleb, "Energy efficiency of LTE macro base station", in *Proc. 1st Int. Symp. Telecommun. Technol. ISTT 2012*, Kuala Lumpur, Malaysia, 2012, pp. 259–264.

[3] M. Pickavet *et al.*, "Worldwide energy needs for ICT: The rise of power-aware networking", in *Proc. 2nd Int. Symp. Adv. Netw. Telecommun. Syst. ANTS 2008*, Bombay, India, 2008, pp. 1–3.

[4] T. T. Tesfay, R. Khalili, J.-Y. L. Boudec, F. Richter, and A. Fehske, "Energy saving and capacity gain of micro sites in regular LTE networks: downlink traffic layer analysis", in *Proc. 6th ACM Workshop on Perform. Monitor. Measur. Heterogen. Wirel. Wired Netw.*, Miami, FL, USA, 2011, pp. 83–92.

[5] 3GPP TR 36.814 V9.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)", 3rd Generation Partnership Project, Tech. Rep., 2010 [Online]. Available: http://www.3gpp.org

[6] A. A. Abdulkafi, T. S. Kiong, D. Chieng, A. Ting, and J. Koh, "Energy efficiency improvements in heterogeneous network through traffic load balancing and sleep mode mechanisms", *Wirel. Personal Commun.*, vol. 75, no.4, pp. 2151–2164, 2014.

[7] A. Goldsmith, *Wireless Communications*. New York: Cambridge University Press, 2005.

[8] S. Sesia, I. Toufik, and M. Baker, *LTE – The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. United Kingdom: Wiley, 2011.

[9] P. Mogensen *et al.*, "LTE capacity compared to the shannon bound", in *Proc. IEEE 65th Veh. Technol. Conf. VTC2007-Spring*, Dublin, Ireland, 2007, pp. 1234–1238.

[10] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks", in *Proc. IEEE 70th Veh. Technol. Conf. Fall VTC-Fall 2009*, Anchorage, AK, USA, 2009, pp. 1–5.

[11] S. Tombaz, M. Usman, and J. Zander, "Energy efficiency improvements through heterogeneous networks in diverse traffic distribution scenarios", in *Proc. 6th Int. ICST Conf. Commun. Netw. in China CHINACOM 2011*, Harbin, China, 2011, pp. 708–713.

[12] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks", in *Future Network and Mobile Summit*, Florence, Italy, 2010, pp. 1–8.

[13] A. J. Fehske, F. Richter, and G. P. Fettweis, "Energy efficiency improvements through micro sites in cellular mobile radio networks", in *Proc. 2nd Int. Worksh. Green Commun.*, in conjunction with *GLOBECOM Workshops 2009*, Honolulu, HI, USA, 2009, pp. 1–5.

[14] A. Chockalingam and M. Zorzi, "Energy efficiency of media access protocols for mobile data networks", *IEEE Trans. Commun.*, vol. 46, pp. 1418–1421, 1998.

[15] W. Wang and G. Shen, "Energy efficiency of heterogeneous cellular network", in *Proc. IEEE 72nd Veh. Technol. Conf. Fall VTC-Fall 2010*, Ottawa, Ontariao, Canada, 2010, pp. 1–5.

[16] Malaysian Communications and Multimedia Commission Annual Report, SKMM-MCMC, 2011 [Online]. Available: http://www.skmm.gov.my/skmmgovmy/media/General/pdf/SKMM_2011.pdf

[17] A. A. Abdulkafi *et al.*, "Energy-aware load adaptive framework for LTE heterogeneous network", *Trans Emerging Tel Tech*, vol. 25, no. 9, pp. 943–953, 2014.

[18] S. N. Shahab, T. S. Kiong, and A. A. Abdulkafi, "A framework for energy efficiency evaluation of LTE network in urban, suburban and rural areas", *Australian J. Basic Appl. Sci.*, vol. 7, no. 7, pp. 404–413, 2013.

Suhail Najm Shahab was graduated in 2010 with B.Eng. in Computer Technology Engineering from Al-Hadba'a University College, Mosel, Iraq. In 2013 he received his M.Sc. in Electrical Engineering with specialization in Wireless Communication from University Tenaga Nasional, Malaysia. Currently, he is enrolled as a Ph.D. candidate in Faculty of Electrical & Electronics Engineering, University Malaysia Pahang, Malaysia. He has authored and co-authored numerous publications in international conferences and journals. His research interest includes LTE, energy-efficient cellular networks, adaptive beamforming, smart antenna.
E-mail: 68suhel@gmail.com
Faculty of Electrical and Electronics Engineering
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia

Ayad Atiyah Abdulkafi received the B.Sc. and M.Sc. degrees in Electrical Engineering (major in telecommunications) from Al-Mustansiriya University, Baghdad, Iraq, in 2001 and 2004, respectively. He received his Ph.D. in Wireless Communication Engineering from University Tenaga Nasional (UNITEN), Malaysia in 2014. He is a staff member in College of Engineering, Tikrit University, Iraq. He is currently a postdoc fellow at Multimedia University, Malaysia. His research interests are in wireless communications, including, LTE, radio resource management and optimization, heterogeneous networks, energy-efficient wireless network design and Green Cellular Networks, OFDM, optical wireless communications, Visible Light Communication.
E-mail: al.ayad@yahoo.com
College of Engineering
Tikrit University
Salahaldin, Iraq

Ayib Rosdi Bin Zainun obtained his B.Eng. in Electrical and Electronics from University Technology Malaysia (UTM), Skudai, Johor, Malaysia in 2000. He received his M.Sc. in Engineering (Adaptive Array Antenna) from Nagoya Institute of Technology, Nagoya, Japan (2005). He has completed Ph.D. (Dye-Sensitize Solar Cell) from University Technology MARA (UiTM), Shah Alam, Selangor, Malaysia (2012). He is currently working as a lecturer at Faculty of Electrical and Electronics Engineering, University Malaysia Pahang. He has been a member of various committees for projects of national interest in Malaysia, and he is referee of various scientific journals. His research is mainly centered on the field of adaptive array antenna and applied science (materials for solar cells applications).

Email: ayib@ump.edu.my
Faculty of Electrical and Electronics Engineering
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia

Paper

Performance Comparison of Four New ARIMA-ANN Prediction Models on Internet Traffic Data

C. Narendra Babu and B. Eswara Reddy

Department of Computer Science and Engineering, JNT University College of Engineering, Anantapuramu, India

Abstract—Prediction of Internet traffic time series data (TSD) is a challenging research problem, owing to the complicated nature of TSD. In literature, many hybrids of auto-regressive integrated moving average (ARIMA) and artificial neural networks (ANN) models are devised for the TSD prediction. These hybrid models consider such TSD as a combination of linear and non-linear components, apply combination of ARIMA and ANN in some manner, to obtain the predictions. Out of the many available hybrid ARIMA-ANN models, this paper investigates as to which of them suits better for Internet traffic data. This suitability of hybrid ARIMA-ANN models is studied for both one-step ahead and multi-step ahead prediction cases. For the purpose of the study, Internet traffic data is sampled at every 30 and 60 minutes. Model performances are evaluated using the mean absolute error and mean square error measurement. For one-step ahead prediction, with a forecast horizon of 10 points and for three-step prediction, with a forecast horizon of 12 points, the moving average filter based hybrid ARIMA-ANN model gave better forecast accuracy than the other compared models.

Keywords—ANN, ANN training, ARIMA, Box-Jenkins methodology, hybrid ARIMA-ANN model, Internet traffic forecasting.

1. Introduction

Time series data (TSD) forecasting has its applications in various domains like agricultural, climatic, econometric, financial and communication. With the growing telecommunication sector, the service providers should be able to effectively distribute their resources for continued services. Internet traffic data forecasting helps service providers manage available bandwidth and resources properly. Consider a situation, where a large part of the bandwidth is being used by a network. Within the next half an hour, if it is a priori known that this network will not consume more than 30% of the available bandwidth, the service provider can reduce the network bandwidth and in-turn divert the rest of the available bandwidth to some other network. This way the resources can be used optimally. Hence, prediction of Internet traffic TSD is drawing more attention in the present days.

2. Related Work

Autoregressive integrated moving average (ARIMA) linear models are popularized by Box and Jenkins in 1970 for time series prediction. These models are applied on various TSD such as electricity prices [1], [2], sugar prices [3], stock market data [4], and wind speed data [5], for the prediction of future values. Next, the pre-processing based ARIMA models were introduced. In [6], a wavelet transformation based ARIMA forecasting is done on global temperature data. In [7], classification and feature extraction techniques were proposed for electrocardiography data. These preprocessing steps help to obtain more accurate predictions.

Later the era has been conquered by the ANN (non-linear) models. ANN was able to model a wide range of TSD compared to ARIMA, as they are capable of modeling non linear variations. ANNs have been applied to electricity demand data [8], financial data [9] river flow data [10], and network data [11], for prediction. Compared to ARIMA, these TSD were accurately predicted with ANN. In [12], neural networks were used to predict earthquakes in Chile.

Instead of individual ARIMA or ANN, research progressed in the direction of combining the benefits of both ARIMA and ANN models to devise hybrid ARIMA-ANN models. Next, a hybrid ARIMA-ANN versions was proposed by Zhang [13], which gave good prediction accuracy compared to individual models, when applied to Wolf's sunspot data, Canadian lynx data, and exchange rate data for one-step ahead prediction. Next, a new hybrid ARIMA-ANN method was proposed by Khashei and Bijari [14], which gave better performance. The hybrid model devised by Zhang was used for electricity price forecasting in [15] and water quality TSD prediction in [16]. In [17], a multiplicative model was proposed (Li Wang *et al.*), which is in contrast to the additive model of Zhang. The results showed that it is no less in comparison to the application of additive Zhang model. In [18], a moving average filter based hybrid ARIMA-ANN model is proposed which first decomposes the data and then applies the apt model on each decomposition. This model is shown to

outperform both Zhang, Khashei and Bijari models, when applied to sunspot data, electricity price data and stock price data, in both one-step ahead and multi-step ahead forecasting.

Many other prediction models are available in the literature. Some of them are based on support vector machines (SVM) [19], and some others on fuzzy logic [20]. Spectral techniques based on SVD were proposed in [15] and the references therein. Most of the forecasting problems showed that hybrid models are a better solution. However, if the hybrid model involves large number of decompositions, the prediction accuracy suffers. Hence, a hybrid model should contain limited number of individual models to retain the model simplicity and prediction accuracy.

In this paper, the Internet traffic TSD predictions for both one-step ahead and multi-step ahead cases are obtained using individual ARIMA, ANN models, hybrid ARIMA-ANN models of Zhang [13], Khashei and Bijari [14], Wang *et al*. [17], Babu and Reddy [18]. From obtained results the best model was identified.

The rest of the paper is organized as follows. In Sections 3–4, the ARIMA, ANN, and some existing hybrid ARIMA-ANN models are described. In Section 5, the results are discussed in four subsections, along with tables of performance measures and graphs of predicted values. Section 6 ends the paper with a conclusion.

3. ARIMA and ANN Prediction Models

Some of the hybrid ARIMA-ANN models available in the literature are outlined, with a brief description.

3.1. ARIMA

To model a TSD using ARIMA, a training data is provided. ARIMA modeling fits a linear equation to this data if it is stationary. If the training data is non-stationary, differencing is performed till it becomes stationary. The corresponding order of differencing is notated as d. The moving average (MA) model order q and auto-regressive (AR) model order p are determined from the decaying nature of auto-correlation function (ACF) plot and the partial ACF (PACF) plot respectively. Detailed correlation analysis for order determination is given in [21]. According to the modeling procedure, the present value of TSD, y_t is considered as a weighted sum of past data points $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ and error values $e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-q}$. The model is shown in Eq. (1), where y_{t-k} is the TSD value at a delay of k time points. The model assumes that the error series e_t has a gaussian distribution.

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots a_p y_{t-p} + e_t + b_1 e_{t-1} + \ldots b_q e_{t-q} \tag{1}$$

The model coefficients $a_1, \ldots, a_p, b_1, \ldots, b_q$ are estimated using Box-Jenkins method [21]. As non-linear likelihood estimation is complex, Gaussian maximum likelihood estimation (GMLE) approaches [22] are used in the estimation of the model coefficients. The model is then validated using diagnostic checks like Akakine Information Criterion (AIC) and Bayesian Information Criterion (BIC). Also normality test like Jarque-Bera test, check on residual autocorrelation plots to meet the confidence limits are also performed. Once the model is identified best according to these diagnostic tests, it is selected for application on the TSD. The selected model is used to predict future TSD values over the prediction horizon.

3.2. ANN

Unlike ARIMA, ANN is a non-linear modeling technique. The neural network model architecture comprises of neurons, similar to the brain's architecture. For example a three layer ANN, with the three layers called as input, hidden and output layers is shown in Fig. 1 [18]. Each layer comprises of one or more nodes. For TSD prediction problem,

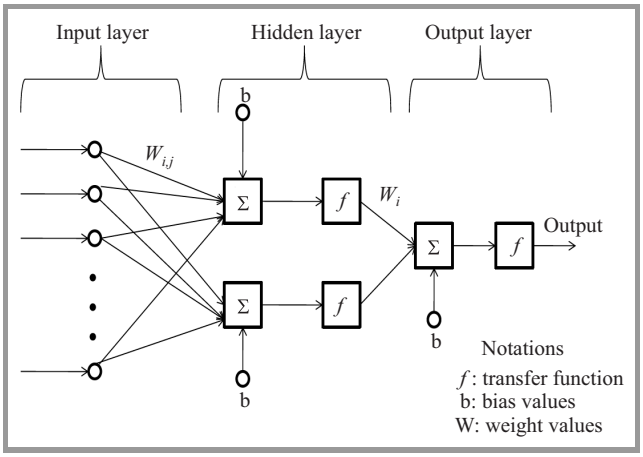


Fig. 1. Three-layer ANN architecture.

the output layer has one node. The hidden layer can have any number of nodes, whose outputs are linked to the output node. The input layer can have one or more nodes depending on the number of TSD points involved in the prediction. There can be more than one hidden layer. The neurons are acyclically linked processing units. Three-layer ANNs are widely used for TSD forecasting. To model TSD using ANN, y_t is expressed as a non-linear function f of $y_{t-1}, \ldots, y_{t-A}$, where A is the lag till which the TSD points are involved in prediction. The model equation is:

$$y_t = g(y_{t-1}, y_{t-2}, \ldots, y_{t-A}) + v_t\,, \tag{2}$$

where v_t is the noise or error term. The transfer function of the hidden layer can be a linear, sigmoid, tan-sigmoid or log-sigmoid in nature. A sigmoid function is:

$$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}}\,. \tag{3}$$

The model coefficients in ANN are weights of each link and the corresponding bias values. To determine these values, a training data is provided to ANN. Many training al-

gorithms are available [23], out of which Levenberg-Marquardt (LM) training algorithm is used in [14], [18]. In [13], a reduced gradient algorithm and in [16], a scaled conjugate gradient algorithm are used. Here LM training is incorporated. The model is diagnosed using validation and testing phase, where the mean square error convergence is verified. If the error is converging, the model is valid, else it is invalid. After the testing phase, the model is used in the prediction of future values.

4. Hybrid ARIMA-ANN Models

Often, the given data may have both linear and nonlinear characteristics. So, hybrid models using both ANN and ARIMA methods are better than individual models for obtaining accurate predictions. Four existing ARIMA-ANN hybrid models considered for discussion in this paper are illustrated as follows.

4.1. Zhang's Hybrid ARIMA-ANN Model

In 2003, Zhang proposed a hybrid ARIMA-ANN model. It is based on the assumption that the given TSD is a sum of two components, linear and non-linear, given in:

$$y_t = L_t + N_t \,. \tag{4}$$

On the given TSD series y_t, ARIMA is fit and the linear predictions are obtained, $\hat{L}_t$, as:

$$\hat{L}_t = a_1 y_{t-1} + \ldots + a_p y_{t-p} + b_1 e_{t-1} + \ldots + b_q e_{t-q} + e_t \,. \tag{5}$$

The difference series is obtained by Eq. (6) on which ANN is fit and the predictions $\hat{N}_t$ are obtained using Eq. (7):

$$n_t = y_t - \hat{L}_t \,, \tag{6}$$

$$\hat{N}_t = f(n_t, n_{t-1}, \ldots, n_{t-A}) + v_t \,. \tag{7}$$

The hybrid model predictions are now obtained by summing the ARIMA and ANN predictions:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \,. \tag{8}$$

This model is suitable for both one-step ahead and multi-step ahead prediction. It is shown to be better than individual models in terms of prediction accuracy. The model is block diagram presented in Fig. 2.

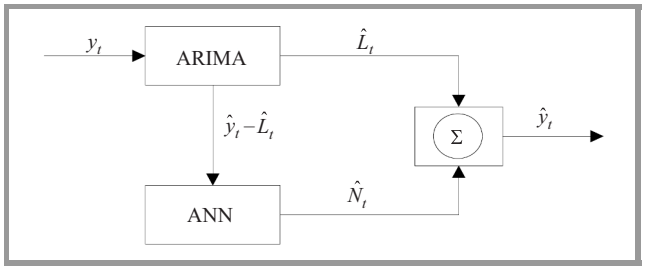


Fig. 2. Zhang's hybrid ARIMA-ANN model.

4.2. Khashei and Bijari's Hybrid ARIMA-ANN Model

In 2010, Khashei and Bijari proposed a new hybrid ARIMA-ANN model for TSD forecasting. Similar to Zhang's model, it also assumes that any TSD has linear and non-linear components, see Eq. (4). But the methodology adopted in prediction is different. An ARIMA model is fit on given TSD to obtain one forecast on the TSD using Eq. (5). Past original values, present prediction, and past error data are all input to the ANN. The ANN gets trained and once the model is validated, the one-step forecast of the given TSD is directly obtained from:

$$\hat{y}_t = f(\hat{L}_t, L_{t-1}, L_{t-2}, \ldots, L_{t-A}) + v_t \,. \tag{9}$$

It is shown to perform better than the Zhang's model in a variety of applications. It is suited for one-step forecasts, but for multi-step forecasting, the model is not suitable. If the past predictions are used as inputs instead of past original values, the model accuracy degrades. The model diagram is illustrated in Fig. 3.

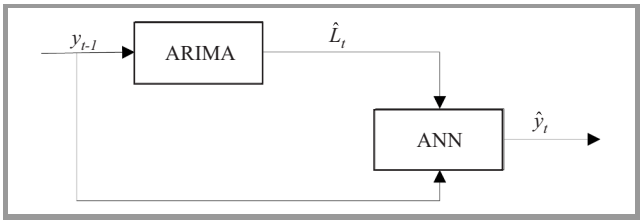


Fig. 3. Khashei and Bijari's hybrid ARIMA-ANN model.

4.3. Multiplicative Hybrid ARIMA-ANN Model

In 2013, Li Wang *et al*. proposed a multiplicative model for forecasting TSD, in contrast to the additive model proposed by Zhang. The model assumes that a given TSD is the product of a linear and a non-linear time series as:

$$y_t = L_t N_t \,. \tag{10}$$

The given TSD y_t is modeled using ARIMA as shown in Eq. (5), similar to the same step in Zhang model. The predictions $\hat{L}_t$ obtained divide the original TSD to obtain the non-linear TSD series as:

$$n_t = \frac{y_t}{\hat{L}_t} \,. \tag{11}$$

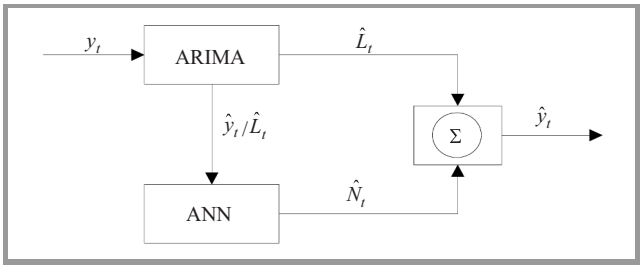


Fig. 4. Multiplicative hybrid ARIMA-ANN model.

The series n_t is modeled and predicted using ANN. The obtained non-linear predictions $\hat{N}_t$ in Eq. (6) and linear predictions $\hat{L}_t$ are multiplied to obtain the final model forecasts as given by Eq. (12). The block diagram of this model is as shown in Fig. 4.

$$\hat{y}_t = \hat{L}_t \hat{N}_t \,. \tag{12}$$

4.4. MA Filter Based Hybrid ARIMA-ANN Model

In [18], a hybrid ARIMA-ANN model is devised using a decomposition step and then applying ARIMA and ANN suitably on each decomposition. The model framework assumes that any TSD is addition of a linear and a non-linear component given in Eq. (4) as in Zhang model. It also assumes that linear processes have less volatility compared to non-linear models, characterized by highly volatile nature. The steps of the model are:

1. An MA filter given by Eq. (13) is used to decompose the given TSD into a low volatile and a highly volatile component. The low volatile component is a smoothened TSD y_{tr}, and the highly volatile component is given by Eq. (14). The length of MA filter m is adjusted such that one of the two decomposed time series is obtained with a kurtosis of 3, which is termed as low volatile decomposition l_t. The difference $h_t = y_t - l_t$ is considered highly volatile. The decomposition is indicated in Eq. (15).

$$y_{tr} = \frac{1}{m} \sum_{i=t-m+1}^{t} y_i \tag{13}$$

$$y_{res} = y_t - y_{tr} \tag{14}$$

$$y_t = l_t + h_t \tag{15}$$

2. The l_t series is modeled and predicted using ARIMA model as in Eq. (16) to obtain $\hat{l}_t$. Note that this modeling using $l_{t-1}, l_{t-2}, \ldots, l_{t-p}$ unlike the ARIMA modeling step of Zhang (5), which uses $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$.

$$\hat{l}_t = f(l_{t-1}, l_{t-2}, \ldots, l_{t-p}, e_t, e_{t-1}, \ldots, e_{t-q}) \tag{16}$$

3. The h_t series is modeled and predicted using ANN model as shown in Eq. (17) to obtain $\hat{h}_t$.

$$\hat{h}_t = g(h_{t-1}, h_{t-2}, \ldots, h_{t-N}) + \varepsilon_t \tag{17}$$

4. The final model predictions are obtained by adding the predictions from steps 2 and 3:

$$\hat{y}_t = \hat{l}_t + \hat{h}_t \,. \tag{18}$$

The steps of this hybrid model are can be represented as a flow chart as shown in Fig. 5 [18].

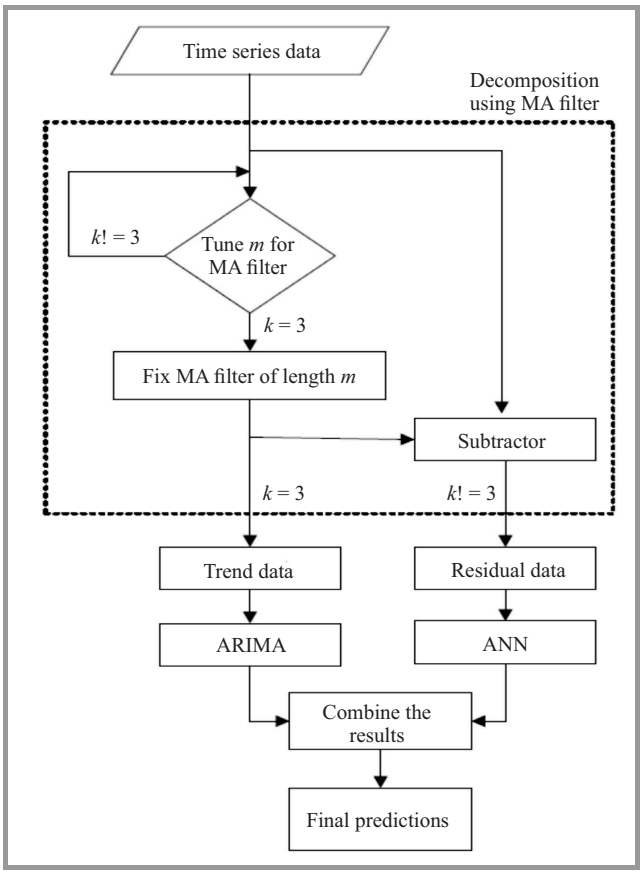


Fig. 5. MA filter based hybrid ARIMA-ANN model.

5. Results and Discussion

The prediction models ARIMA, ANN, Zhang's hybrid ARIMA-ANN, Khashei and Bijari's hybrid ARIMA-ANN, multiplicative hybrid ARIMA-ANN and MA-filter based hybrid ARIMA-ANN are extensively studied for their usage on Internet traffic data. The Internet traffic data obtained from [24] is used in this study. The raw data is

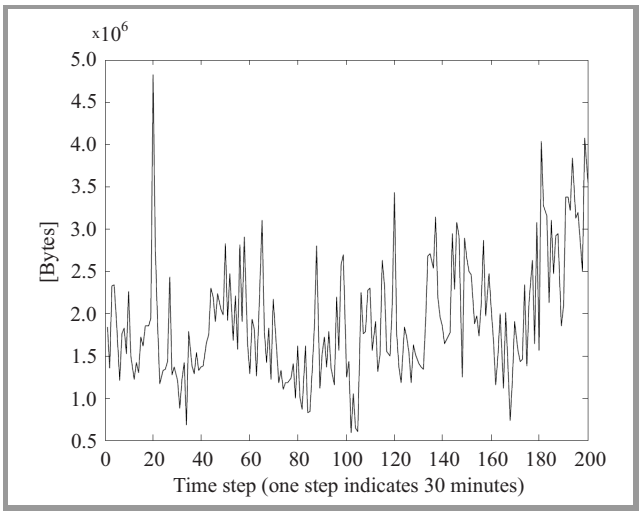


Fig. 6. Actual internet traffic TSD sampled at 30 min steps.

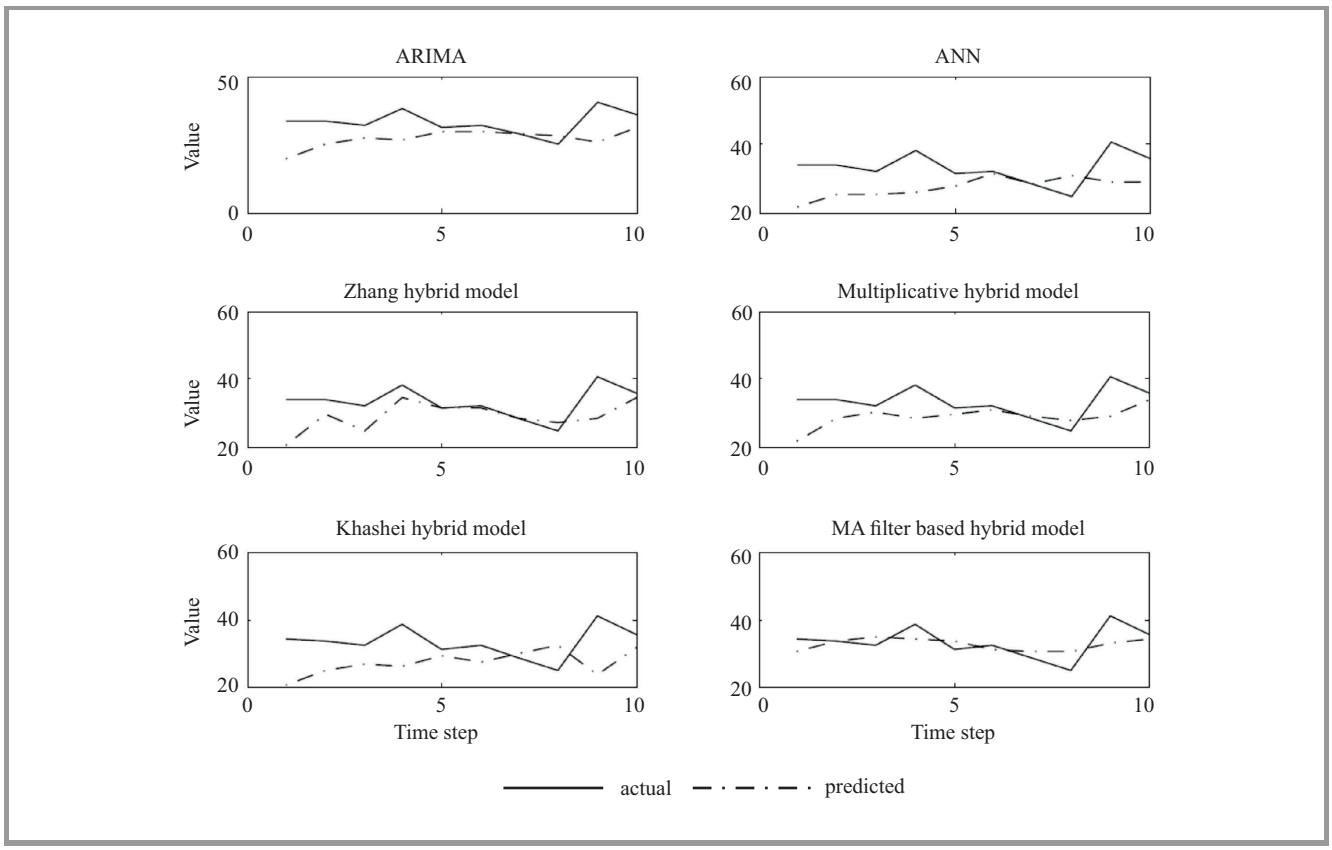


Fig. 7. One-step ahead predictions for TSD1.

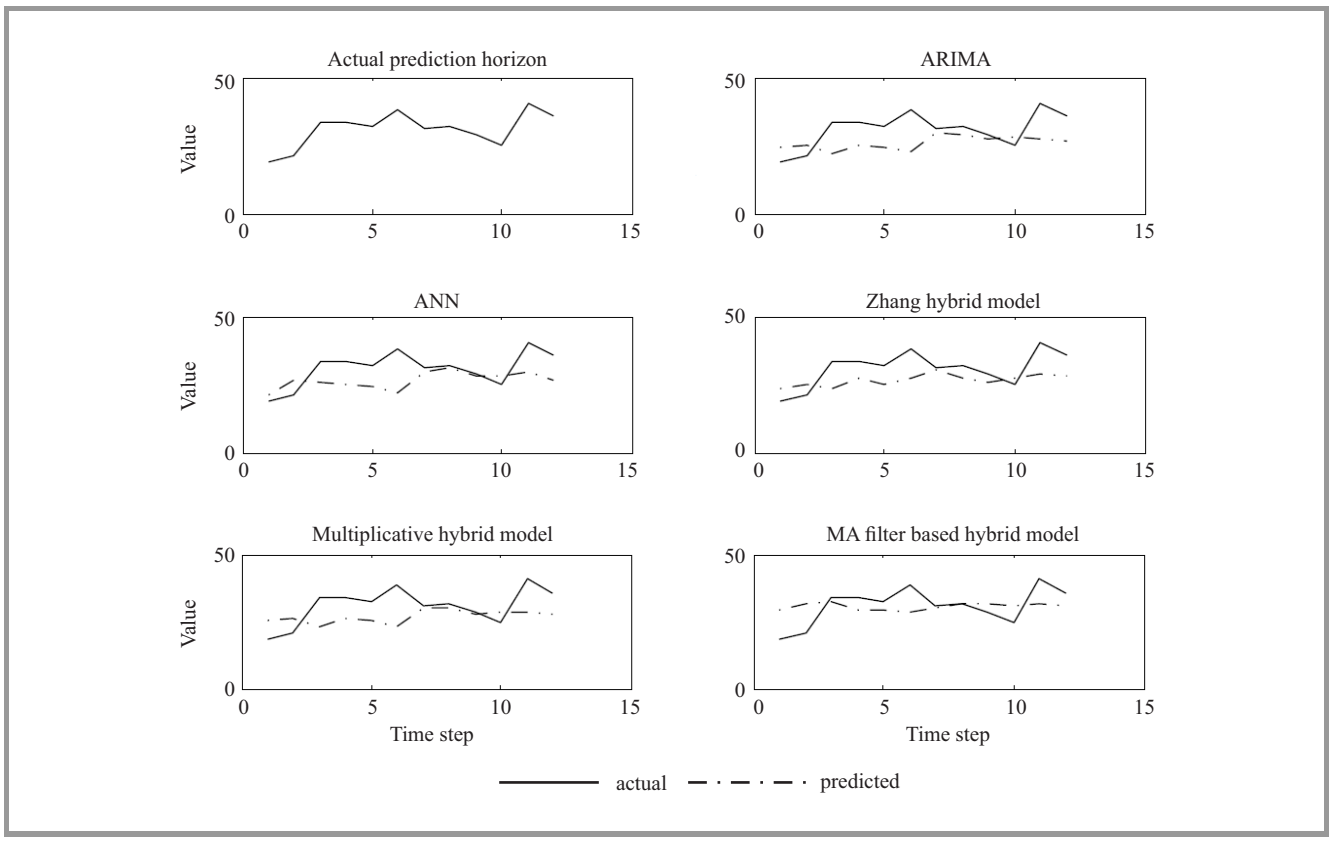


Fig. 8. Three-step ahead predictions for TSD1.

available at every one second for a period of 100 hr. This data is re-sampled to form two different data sets. The first TSD, named as TSD1 is obtained for every 30 min with a total number of 200 points. The second TSD, named as TSD2 is obtained for every 60 min with a total of 100 data points. The processed Internet data is in megabytes. To avoid big numbers, in this study, the authors divided this data by 10^5 and then used it. However the raw data is plotted in Figs. 6 and 9 for 30 and 60 minutes sampling respectively. On both these Internet traffic datasets, the six models are applied and their performances are compared for both one-step ahead and three-step ahead forecasting. The performance measures considered in the study are described.

5.1. Performance Measurement

The two performance measures for accuracy comparison used in this paper are mean absolute error (MAE) and mean squared error (MSE), given by Eqs. (19) and (20) respectively. The smaller these values, the better is the model. In both formulas $E\{.\}$ is the expectation operation, ni and nf indicate start and end points of the prediction horizon, $y_{k,act}$ is the actual value of the time series, and $y_{k,pred}$ is the forecasted time series value at the instant k.

$$MAE = E\left\{\left|y_{act} - y_{pred}\right|\right\} = \\ = \frac{1}{nf - ni + 1}\left(\sum_{k=ni}^{nf}\left|y_{k,act} - y_{k,pred}\right|\right) \tag{19}$$

$$MSE = E\left\{\left|y_{act} - y_{pred}\right|^2\right\} = \\ \frac{1}{nf - ni + 1}\left(\sum_{k=ni}^{nf}\left|y_{k,act} - y_{k,pred}\right|^2\right) \tag{20}$$

5.2. Results for TSD1

TSD1 comprises of 200 points, each indicating the number of packets transmitted. The forecast horizon is taken as 10 data points (which is 5%), and corresponding one-step ahead predictions are obtained. By using a forecast horizon of 12 data points (implying 5%), a three-step ahead prediction is carried out. The MAE and MSE performance results for all the models in both these cases are presented in Table 1. The original TS is shown in Fig. 6. The predictions for the one-step ahead forecast and three-step ahead forecast are shown in Figs. 7 and 8 respectively. The MA-filter based hybrid ARIMA-ANN model outperformed the others in terms of both MAE and MSE.

Table 1
Performance comparison for TSD1

Model	One-step-ahead		Three-step-ahead	
	MAE	MSE	MAE	MSE ($\cdot 10^3$)
ARIMA	6.9352	70.6029	7.1707	72.9144
ANN	6.5810	64.8243	6.3713	61.7111
Zhang	4.6518	44.8343	6.1732	50.2933
Multiplicative	4.9226	45.2739	6.6600	63.3805
Khashei and Bijari	7.7572	85.4724	NA	NA
MA-filter based	**2.9870**	**13.4466**	**5.3093**	**42.2978**

5.3. Results for TSD2

TSD2 comprises of 100 points, each indicating the number of packets transmitted. The forecast horizon is taken as 10%, which implies 10 data points, and one-step ahead predictions are obtained for these points. A three-step ahead prediction is carried out by using a forecast horizon of 12 which is again nearly 10%. The MAE and MSE performance results for all the models in both these cases are presented in Table 2. The original TS is shown in Fig. 9. The predictions for the one-step ahead forecast and three-step ahead forecast are shown in Figs. 10 and 11 respectively. It is noticed that the MA-filter based hybrid ARIMA-ANN model outperformed the others in terms of both MAE and MSE.

Table 2
Performance comparison for TSD2

Model	One-step-ahead		Three-step-ahead	
	MAE	MSE	MAE	MSE ($\cdot 10^3$)
ARIMA	14.3987	313.6585	14.4933	332.1901
ANN	9.2766	129.5465	11.0087	190.3712
Zhang	9.7403	175.3520	10.2397	169.2432
Multiplicative	9.4355	161.6842	13.8069	303.3857
Khashei and Bijari	16.5214	386.2702	NA	NA
MA-filter based	**6.2872**	**57.6054**	**8.4071**	**102.2310**

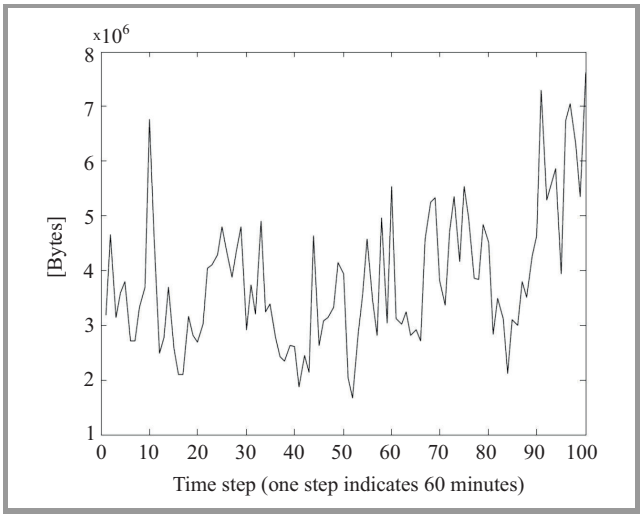


Fig. 9. Actual Internet traffic TSD sampled at 60 min.

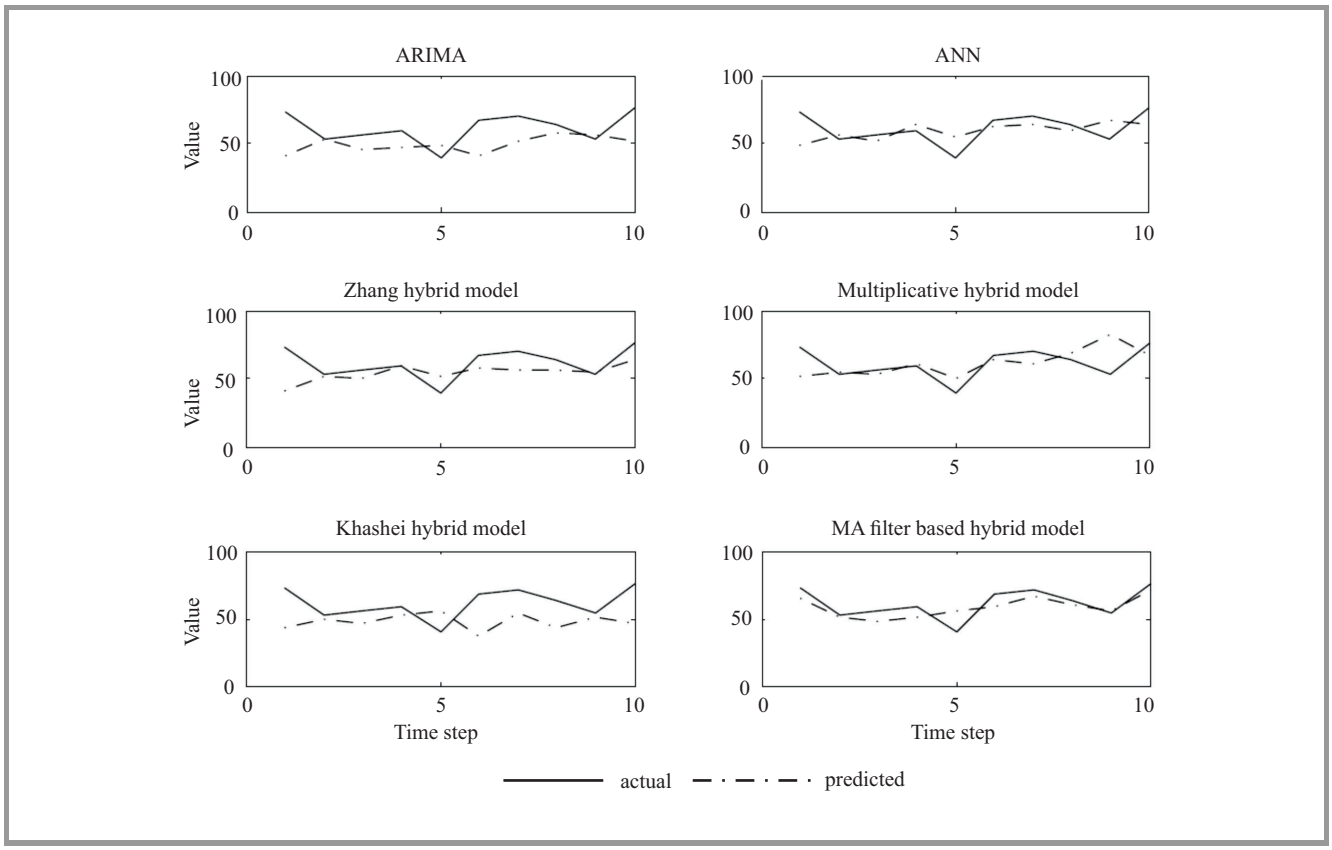


Fig. 10. One-step ahead predictions for TSD2.

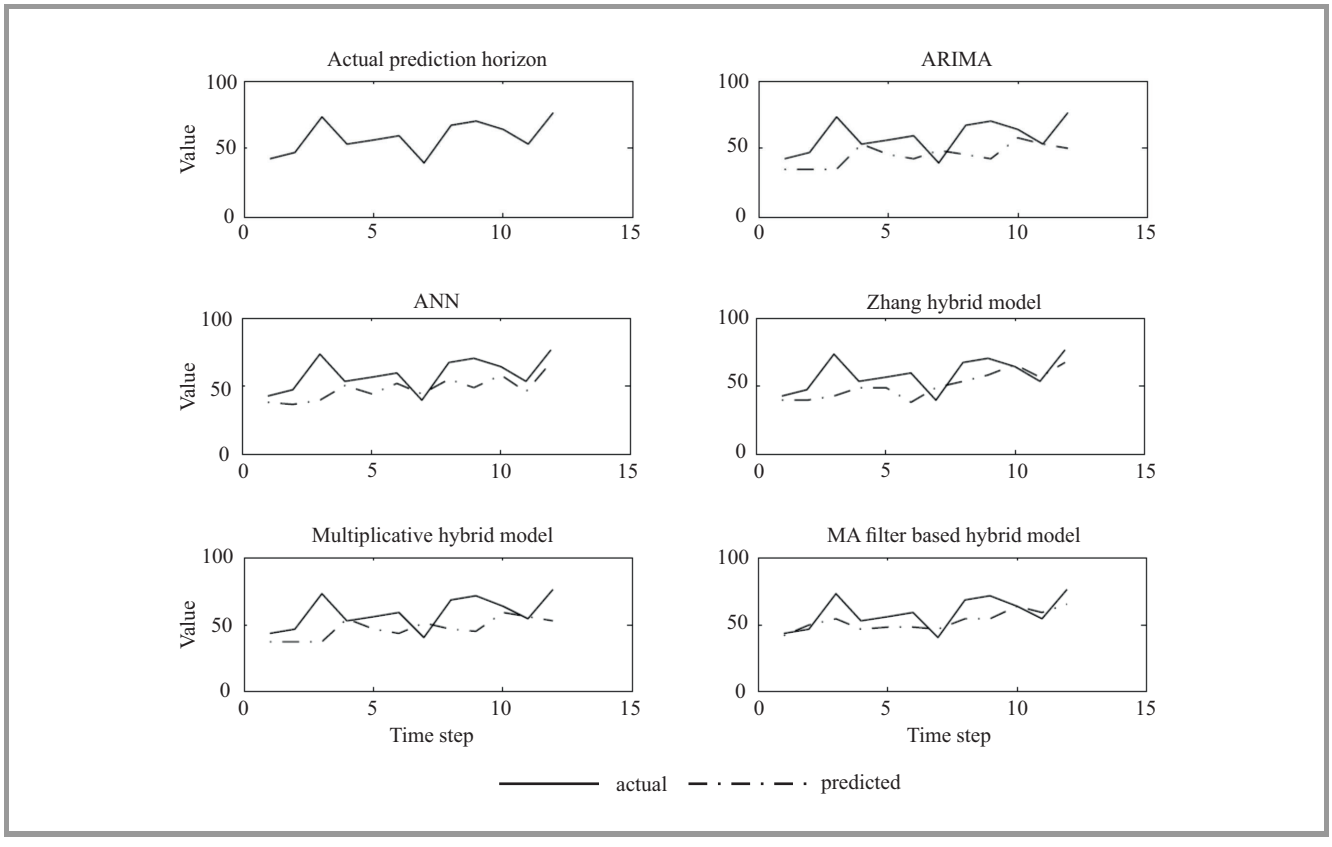


Fig. 11. Three-step ahead predictions for TSD2.

6. Conclusion

In this paper, for the prediction of Internet traffic TSD which is highly volatile in nature, the applicability of various prediction models is explored. The models considered in the study are ARIMA, ANN, Zhang's hybrid ARIMA-ANN, Khashei and Bijari's hybrid ARIMA-ANN, multiplicative ARIMA-ANN, MA-filter based hybrid ARIMA-ANN. Both one-step ahead and multi-step ahead predictions are carried out. The error performance measures, MAE and MSE are used to evaluate the model accuracy.

Two traffic TSD series, one with 30 min sampling and 200 data points, other with 60 min sampling and 100 data points are used in the investigation. The prediction results in all the cases showed that the MA filter based hybrid ARIMA-ANN model outperformed all the other models discussed in this paper, in terms of both MAE and MSE and hence is suitable for predicting Internet traffic data more accurately.

References

[1] J. Contreras, R. Espinola, F. Nogales, and A. Conejo, "ARIMA models to predict next-day electricity prices", *IEEE Trans. on Power Syst.*, vol. 18, no. 3, pp. 1014–1020, 2003.

[2] E. Gonzalez-Romera, M. Jaramillo-Moran, and D. Carmona-Fernandez, "Monthly electric energy demand forecasting based on trend extraction", *IEEE Trans. on Power Syst.*, vol. 21, no. 4, pp. 1946–1953, 2006.

[3] K. Suresh and S. Krishna Priya, "English Forecasting sugarcane yield of tamilnadu using ARIMA models", *English Sugar Tech*, vol. 13, no. 1, pp. 23–26, 2011.

[4] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model", *Omega*, vol. 40, no. 6, pp. 758–766, 2012.

[5] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model", *Renewable Energy*, vol. 35, no. 12, pp. 2732–2738, 2010.

[6] C. Babu and B. Reddy, "Predictive data mining on average global temperature using variants of ARIMA models", in *Proc. 2012 Int. Conf. Adv. Engin. Sci. and Manag. ICAESM 2012*, Tamil Nadu, India, 2012, pp. 256–260.

[7] U. Orhan, "Real-time CHF detection from ECG signals using a novel discretization method", *Comp. in Biology and Medicine*, vol. 43, no. 10, pp. 1556–1562, 2013.

[8] D. Singhal and K. Swarup, "Electricity price forecasting using artificial neural networks", *Int. J. Elec. Power & Energy Syst.*, vol. 33, no. 3, pp. 550–555, 2011.

[9] W.-S. Chen and Y.-K. Du, "Using Neural Networks and data mining techniques for the financial distress prediction model", *Expert Syst. with Appl.*, vol. 36, no. 2, part 2, pp. 4075–4086, 2009.

[10] C. H. F. Toro, S. G. Meire, J. F. Gálvez, and F. Fdez-Riverola, "A hybrid artificial intelligence model for river flow forecasting", *Applied Soft Computing*, vol. 13, no. 8, pp. 3449–3458, 2013.

[11] B. R. Chang and H. F. Tsai, "Novel hybrid approach to data-packet-flow prediction for improving network traffic analysis", *Appl. Soft Comput.*, vol. 9, no. 3, pp. 1177–1183, 2009.

[12] J. Reyes, A. Morales-Esteban, and F. Martínez-Álvarez, "Neural networks to predict earthquakes in Chile", *Appl. Soft Comput.*, vol. 13, no. 2, pp. 1314–1328, 2013.

[13] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomput.*, vol. 50, no. 0, pp. 159–175, 2003.

[14] M. Khashei and M. Bijari, "A novel hybridization of Artificial Neural Networks and ARIMA models for time series forecasting", *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2664–2675, 2011.

[15] H. H. Arash Miranian and M. Abdollahzade, "Day-ahead electricity price analysis and forecasting by singular spectrum analysis", *IET Gener. Transmiss. & Distrib.*, vol. 7, no. 4, pp. 337–346, 2013.

[16] D. Ömer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction", *Engin. Appl. of Artif. Intell.*, vol. 23, no. 4, pp. 586–594, 2010.

[17] L. Wang, H. Zou, J. Su, L. Li, and S. Chaudhry, "An ARIMA-ANN hybrid model for time series forecasting", *Wiely-Syst. Res. and Behav. Sci.*, vol. 30, pp. 244–259, 2013.

[18] C. N. Babu and B. E. Reddy, "A moving-average-filter-based hybrid arima-ann model for forecasting time series data", *Applied Soft Computing*, vol. 23, pp. 27–38, 2014 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494614002555

[19] Y. Sai, Z. Yuan, and K. Gao, "Mining stock market tendency by RS-based support vector machines", in *Proc. IEEE Int. Conf. Granular Comput. GrC 2007*, Fremont, CA, USA, 2007, pp. 659–664 [Online]. Available: http://dx.doi.org/10.1109/GRC.2007.99

[20] M. R. Hassan, "A combination of Hidden Markov Model and fuzzy model for stock market forecasting", *Neurocomput.*, vol. 72, no. 16–18, pp. 3439–3446, 2009.

[21] G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc., 1990.

[22] Q. Yao and P. J. Brockwell, "Gaussian maximum likelihood estimation for ARMA models I: time series", *J. Time Series Anal.*, vol. 27, no. 6, pp. 857–875, 2006.

[23] S. O. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, 2008.

[24] CAIDA, 2014 [Online]. Available: http://www.caida.org/data/

C. Narendra Babu graduated in BE (CSE) from Adichunchanagiri Institute of Technology, Chikmagalur in 2000. He received his M.Sc. in M.Tech. (CSE) from M. S. Ramaiah Institute of Technology, Bangalore in 2004. He submitted his Ph.D. thesis to JNT University, Anantapuramu. He is an Associate Professor in the Department of ISE, Reva Institute of Technology and Management, Bangalore. His research interests include time series data analysis and mining, soft computing. He has published four papers in reputed international journals and two international conferences, also received best author award from IEEE-ICAESM international conference in 2012 held at Nagapattinam. He is a member of IEEE from 2009.

E-mail: narendrababu.c@gmail.com
Department of Computer Science and Engineering
JNT University College of Engineering
Anantapuramu, India

B. Eswara Reddy graduated in B.Tech. (CSE) from Sri Krishna Devaraya University in 1995 and obtained M.Sc. degree from JNT University, Hyderabad, in 1999. He received Ph.D. in Computer Science and Engineering from JNT University, Hyderabad, in 2008. He is as Professor of CSE Dept. and Coordinator for Master of Science in Information Technology (MSIT) program, JNTUA in collaboration with Consortium of Institutions of Higher Learning (CIHL). He has written more than 50 publications in various journals and conferences, and two books. He has received University Grants Commission-Major Research Project (UGC-MRP) titled "Cloud computing framework for rural health care in Indian scenario". His research interests include cloud computing, pattern recognition and image analysis, data warehousing and mining, and software engineering. He is a life member of ISTE, IE, ISCA, IAENG and member of CSI, IEEE.

E-mail: eswar.cse@jntua.ac.in
Department of Computer Science and Engineering
JNT University College of Engineering
Anantapuramu, India

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.

A copy of the JTIT is provided to each author of paper published.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishing's databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(*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

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY 1/2015