

**Simulation Platform for the Planning and Design of Networks
Carrying VoIP Traffic**

by

Abdel Hernandez Rabassa

B.Eng. (2000)

A Thesis submitted to the

Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Masters of Applied Science in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada, K1S 5B6

May, 2010

The undersigned recommend to the Faculty of Graduate Studies and

Research acceptance of the thesis

**Simulation Platform for the Planning and Design of Networks
Carrying VoIP Traffic**

Submitted by

**Abdel Hernandez Rabassa
B.Eng.**

In partial fulfillment of the requirements for the degree of Masters of
Applied Science

Chair, Dr. H.M. Schwartz, Department of Systems and Computer
Engineering

Thesis Co-supervisor, Dr. Marc St-Hilaire

Thesis Co-supervisor, Dr. Chung-Horng Lung

Carleton University

May 2010

Abstract

In order to overcome the known challenges (i.e., latency, jitter, packet loss, and etc.) of transmitting multimedia traffic over a packet switched network, careful network planning needs to take place. Existing simulation platforms, particularly for Voice over IP (VoIP) simulations, have available a limited selections of speech encoding algorithms. The primary objective of this thesis is the creation of a tool aimed at supporting the planning and design phases of packet switched networks carrying voice traffic while considering realistic and current network conditions and simulation features. More specifically, this thesis focuses on the creation of a speech background traffic generation models with the purpose of generating traffic that follows the behaviour of a number of speech encoding algorithms. Also, a model to integrate real speech to the VoIP simulation is offered. Lastly, objective and subjective speech quality assessment methodologies are implemented.

Acknowledgements

To my wife Rachel Jane for giving me the initial impulse and courage to face this challenge, and for the support throughout, thank you very much.

I am very grateful to my lovely kids, Caleb B. and Gaby Lou for having decided to arrive to this world within the frame of this degree. They have made this journey interesting and lively and diverse.

I am grateful to my thesis supervisors, Professors Marc St-Hilaire, Chung-Horng Lung for their support and advice. Also, I am grateful to Professor Ioannis Lambadaris for his invaluable expert advice.

To my parents, for having always being there, because you are now here, thank you very much.

I am grateful to my friend Victor, for being the Toronto office of my simulations and for listening. To Frank, for all the moments together, I am grateful.

Finally, I acknowledge the financial support from Mathematics of Information Technology and Complex Systems (MITACS) and Cistel Technology Inc. towards the completion of this degree.

Table of Contents

Abstract.....	iii
Acknowledgements	iv
Table of Contents	v
List of Figures.....	viii
List of Tables	x
List of Acronyms	xi
Chapter 1 Introduction.....	1
1.1 Problem Statement	3
1.2 Research Objectives	6
1.3 Main Contributions	7
1.4 Thesis outline	7
Chapter 2 Related Works.....	9
2.1 Study of speech encoding algorithms and feasibility of integration to a simulation platform	10
2.1.1 General characteristics of audio compression codecs	11
2.1.2 Fixed data rate speech encoder algorithms	12
2.1.3 Variable data rate speech encoder algorithms.....	14
2.1.4 Study of the feasibility of integrating encoding algorithms to a simulation platform ..	19
2.2 Study of Speech Quality Assessment Models.....	22
2.2.1 Understanding VoIP Quality Assessment Models.....	23
2.2.2 Description of relevant speech quality assessment models for VoIP systems	27

2.2.3 Selection of the speech quality assessment models to integrate to the simulation platform.....	37
Chapter 3 VoIP Simulation Models and OPNET Implementations	40
3.1 Speech background traffic generation.....	40
3.1.1 Available encoding algorithms and parameter calculations.....	41
3.1.2 Analysis of the number of speech frames per packet attribute	51
3.1.3 Computation of the voice payload size	54
3.1.4 Background traffic scalability method	56
3.1.5 General operation of the simulation of speech background traffic generation in OPNET.....	57
3.2 Simulation of real speech traffic in OPNET	61
3.2.1 Encoding algorithms used in real speech simulation and licensing details.....	62
3.2.2 Playout buffer (De-jitter buffer).....	63
3.2.3 General operation of the simulation of real speech traffic generation in OPNET	64
3.2.4 Objective speech quality assessment	70
3.2.5 Subjective speech quality assessment	74
Chapter 4 Model Validation and Simulation Results Analysis	78
4.1 Validation of simulation models	78
4.1.1 Validation of traffic generation models for fixed data rate encoding algorithms	79
4.1.2 Validation of traffic generation model for AMR-NB encoding algorithm	82
4.1.3 Validation of traffic generation model for Speex encoding algorithm	85
4.2 Codec packet loss concealment evaluation and playout buffer optimization	87
4.3 Simulation of a real life scenario.....	92
4.3.1 Deploying VoIP within a single office	92

4.3.2 Deploying VoIP across several branch offices	100
4.3.3 Deploying VoIP across several branch offices. Solutions to low MOS and high network load problems.....	103
4.3.4 Deploying VoIP across several branch offices. Performance analysis using real Internet traffic trace.....	110
Chapter 5 Conclusions and Future Work	115
5.1 Summary of thesis.....	115
5.2 Summary of results.....	115
5.3 Suggestions for future research	116
References	117
Appendix A New Nodes Added to OPNET.....	128
Appendix B Using the Background Traffic Generator node, the Real Speech Traffic Generator Node and the Client Node in an OPNET Simulation.....	136

List of Figures

Figure 2.1 Speech quality assessment models	24
Figure 3.1 Probability Density Function for the random variable Speex frame size	49
Figure 3.2 Autocorrelation function of the Speex frame size random variable	50
Figure 3.3 Obtaining the Innovation Function in TEstool for Speex Q2 frame size.....	52
Figure 3.4 Effects of the increase of the number of frames per packet.....	54
Figure 3.5 Structure of the application layer packet	55
Figure 3.6 Operation of Voice background traffic generator.....	58
Figure 3.7 Traffic generation chart for each of the encoding algorithm groups	59
Figure 3.8 Block diagram of the real voice simulation in OPNET for fix data rate codecs.....	66
Figure 3.9 Block diagram of the real voice simulation in OPNET for AMR-NB codec	68
Figure 3.10 ITU-T G.107 Delay Impairment model.....	73
Figure 3.11 MOS calculations based on E-model implemented in the client node	74
Figure 4.1 Structure of a voice packet considering the contribution of all protocols involved	79
Figure 4.2 Simulated bandwidth usage in an Ethernet link.....	81
Figure 4.3 Variations of the simulated AMR-NB frame size with the M2E delay	83
Figure 4.4 Detailed view of the frame size adjustment procedure for AMR-NB	84
Figure 4.5 TEstool screen shot. Empirical data and OPNET simulated Speex Q2.....	86
Figure 4.6 TEstool screen shot Speex Q2.	87
Figure 4.7 Topology used in the codec performance evaluation	88
Figure 4.8 Evaluation of codec performance at constant network conditions	89
Figure 4.9 MOS as a function of the playout buffer size and the average jitter.....	92
Figure 4.10 Network topology of a small office Ottawa. Data traffic	93
Figure 4.11 Network topology of a small office Ottawa. Data and voice traffic.....	95

Figure 4.12 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	97
Figure 4.13 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	98
Figure 4.14 Network topology. Toronto office and Ottawa office	100
Figure 4.15 Office network topology	100
Figure 4.16 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	102
Figure 4.17 Network topology	105
Figure 4.18 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	106
Figure 4.19 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	109
Figure 4.20 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	110
Figure 4.21 Network topology	111
Figure 4.22 PDF of the E2E delay measured from a Toronto location to Carleton University ..	112
Figure 4.23 MOS and Link utilization for G.711, G.729, AMR and iLBC-30ms codecs	113

List of Tables

Table 2.1 Bit rate modes for iLBC.....	14
Table 2.2 Available modes for ITU-T G.727.....	15
Table 2.3 Bit rate modes and their associated bandwidth for AMR-NB codec	16
Table 2.4 Operation modes for EVCR codec	17
Table 2.5 Operation modes for Speex narrowband codec.....	18
Table 2.6 Absolute Category Rating (ACR) scales.....	28
Table 2.7 Degradation Category Rating (DCR) scale.....	28
Table 2.8 Comparison Category Rating (CCR) scale	29
Table 2.9 R-factor to MOS mapping.....	35
Table 3.1 Frame size and frame duration for fix data rate codecs.....	41
Table 3.2 Threshold M2E values for AMR codec and the resulting frame sizes.....	46
Table 3.3 Properties of the encoding algorithms used in the real voice generator node	62
Table 3.4 Curve fitting parameters for calculation of I_e	72
Table 4.1 Predicted bandwidth requirements in an Ethernet link	81
Table 4.2 Predicted AMR-NB frame size based on the E2E delay.....	84
Table 4.3 Optimum ployout buffer size as a function of jitter for G.729 codec	90
Table 4.4 Characteristics of data traffic inside small office in Ottawa	94

List of Acronyms

3GPP	3rd Generation Partnership Project
3GPP2	3rd Generation Partnership Project 2
ACR	Absolute Category Rating
ADPCM	Adaptive Differential Pulse Code Modulation
AMR-NB	Adaptive Multi-Rate-Narrow Band
CCR	Comparison Category Rating
CNG	Comfort Noise Generation
CS-ACELP	Conjugate-Structure Algebraic-Code-Excited Linear-Prediction
DAM	Diagnostic Acceptability Method
DCR	Degradation Category Rating
DS1	Digital Signal 1
DS3	Digital Signal 3
DTMF	Dual-tone Multi-frequency
DTX	Discontinuous Transmission
E2E	End-to-End delay
EVRC	Enhanced Variable Rate Codec
FTP	File Transfer Protocol
GIPS	Global IP Solutions
GSM	Global System for Mobile Communications
HTTP	Hypertext Transfer Protocol
iLBC	Internet Low Bitrate Codec
IP	Internet Protocol
IT	Information Technology
ITU	International Telecommunication Union
IVR	Interactive Voice Response
LAN	Local Area Network
LPC	linear-predictive coding
M2E	Mouth-to-Ear delay
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
MOS-CQ	MOS conversational quality

MOS-LQ	MOS Listener's quality
MTU	Maximum Transmission Unit
OH	Overhead
P.VTQ	Passive Voice Transmission Quality
PAMS	Perceptual Analysis Measurement System
PAQM	Perceptual Audio Quality Measure
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
PLP	Perceptual Linear Prediction
PPP	Point-to-Point Protocol
PSQM	Perceptual Speech Quality Measure
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RTCP	Real-Time Control Protocol
RTP	Real-Time Protocol
SID	silence descriptor
TCP	Transmission Control Protocol
TES	Transform-Expand-Sample
UMTS	Universal Mobile Telecommunications System
VAD	Voice Activity Detection
VAR	Variable Rate
VoIP	Voice over Internet Protocol

Chapter 1

Introduction

The data network has become a critical resource in most modern business models. Network design and planning stages require precise observation to ensure business continuity by achieving high levels of network and application flexibility. The Information Era that we are all immersed in has brought new challenges to the network design problem. With hundreds of coexisting protocols and technologies, running from the application level to the physical layer, the task of network and application optimization has grown to be far from trivial.

In recent years, as data networks evolve, several of the traditional telecommunications services have adapted to become part of a digital convergence phenomenon. Radio broadcasting, television and telephone services are among the most obvious examples of such services. Converged networks integrate voice, video, and data traffic into a single network [1]. Numerous advantages can be construed from the centralized nature of this model; some of the most relevant to the integration of multimedia applications to a converged network are mentioned below.

- Very efficient use of the data bandwidth, equipment and transmission lines. The ability to transmit more than one telephone call and data over the same broadband connection.
- As a common infrastructure is used for media and data traffic, the transmission costs are considerably lower when compared, for instance, with the traditional Public Switched Telephone Network (PSTN).
- Having one network hosting several services makes possible to consolidate all network expenses.
- The support and maintenance of the provided services simplifies considerably by having fewer systems to interconnect and manage.
- Implementing value added features to the basic telecommunication services becomes a simpler task when supported by a unified network structure. Service providers and

corporate entities will observe an increased in revenues from implementation of new services.

Although the numerous and significant advantages offered by the integration of services in a unified converged network, some disadvantages can also be spotted. The first effect of adding services to the existing data network is that the volume of data traffic will increase. The result of this boost in the traffic volume will affect the data traffic in the form of additional delay and this effect should be carefully analyzed and corrected when possible. Most data transmission services will operate over Transmission Control Protocol (TCP) ensuring that all packets arrive at their destinations, even if retransmission is required, examples are email and File Transfer Protocol (FTP).

Multimedia applications, more specifically Voice over IP (VoIP) applications, require by nature a strict synchronism and stable traffic flow between communication endpoints. Traditionally, audio and video have been transmitted in a connection-oriented environment, ensuring very high levels of quality and reliability. For instance, conventional PSTN has operated with a standard of offering 99.999% uptime, which is sometimes referred as “five nines reliability” [2]. The biggest challenge of using a packet switched network as the support for media services can be summarized as transmitting connection-oriented applications over a connectionless network while maintaining reasonable levels of quality and reliability.

Unlike the transmission of data where control protocols like TCP can be implemented, the transmission of real-time service over a packet switched network presents more complex challenges than just increase of transmission delays. Retransmission of packets is not a suitable flow control mechanism when latency and out-of-order packet arrival are critical network impairments that affect media applications. Packet loss becomes then an extra burden for the multimedia applications to bear. Furthermore, the variation of the total delay in time, commonly referred to as jitter, is also comprised among the most harmful network impairments to affect quality of media transmission.

The rest of this chapter is organized as follow. In Section 1.1, the limitations of existing VoIP simulation platforms and research will be discussed. Also, the problem statement for this thesis is discussed. In Section 1.2, the objectives of the research are presented. In Section 1.3, the

main contributions of this research are outlined while Section 1.4 discusses the structure of the rest of the thesis.

1.1 Problem Statement

Countless efforts and research have been invested in adjusting the packet switched networks to efficiently host VoIP traffic. Hardware and software solutions are now part of a myriad of improvements scattered all over the network and its different layers with the sole purpose to accommodate speech traffic. Network nodes have been completely redesigned, more powerful and comprehensive audio and video encoding algorithms exist [3, 4], Quality of Service (QoS) protocols are now implemented as part of the firmware in nodes and endpoint hardware, various optimization algorithms developed to deal with buffer size and routing schemes. In addition, different standards have been created to effectively measure and predict the perceived quality of audio and video.

In order to overcome the known challenges of transmitting voice traffic over a switched packet network, careful network planning needs to take pace. The prediction of the intricacies of the existing interrelations between the abovementioned factors has become an extremely difficult task; hence, the necessity of powerful simulation platforms capable of encompassing as many realistic factors as possible.

Commercial network planning platforms

VoIP is currently an integral block of the majority of commercial network planning models and platforms. The main ambition of such tools is to create a link between predicted QoS level and network planning/designing decisions. Results should enhance the ability of organizations to predict, avoid, and lessen the effects of network and application functional interruption and the capacity to adapt the network to rapidly changing business conditions.

One of the most powerful tools in the network R&D environment is indeed OPNET Modeler by OPNET Technologies, Inc [5]. In spite of its versatility, the standard libraries of OPNET Modeler provide little support to VoIP traffic analysis. The selection of speech algorithms provided is incomplete and restricted to constant rate algorithms.

Furthermore, two other commercial tools were surveyed. The first one is CISCO Network Planning Solution 2.1 [6] by CISCO Systems Inc. [7]. This tool is extremely powerful and it is based on OPNET technologies. However, for the purpose of general research, it can be found that its performance is limited. The main focus of this platform is to provide traffic growth, consolidation, and migration support to existing networks, with emphasis on encouraging the utilization of CISCO hardware solutions. In addition, using the available documentation for CISCO Network Planning Solution 2.1 [6], no improvement to the OPNET standard VoIP libraries could be spotted.

The second tool surveyed is Vivinet Assessor [8] by NetIQ [9]. This tool, in conjunction with Vivinet Diagnostics 2.3 [10], has the main purpose of assessing the readiness of existing networks in supporting VoIP traffic. Through live analysis of the network, reports are generated showing a predicted behaviour of VoIP calls and hypothetical scenarios. When tested on a real network, it was found that it only supports Avaya [11], CISCO and Nortel [12] Internet Protocol (IP) phones and more importantly, the platform only supports constant rate speech encoding algorithms; namely, ITU-T G.711 [13], ITU-T G.723.1 (5.3 kbps) [14], ITU-T G.723.1 (6.3 kbps) [14], ITU-T G.726 (only 32 kbps) [15] and ITU-T G.729 [16].

Research in the field of speech quality prediction oriented to network planning and design

Beyond the commercially available tools, abundant research has been performed in the field of assessment and prediction of speech quality over packet switched networks. From all literature reviewed on the subject, the two most complete works will be briefly discussed. Showing a truly wide range of points of view to the speech quality assessment “Speech Quality Prediction for Voice over Internet Protocol” [17] covers novel algorithms to assess quality of speech, play-out buffer optimization techniques while using constant rate and variable rate speech codecs. Limitations of this work were found to be related in the first place to the poor analysis involved in the selection of the conditions in which the test beds were implemented. Also, from the very important group of variable rate speech algorithms, only Adaptive Multi-Rate algorithm (AMR) [18] has been analysed with some bandwidth adjustment restrictions.

The work presented in [19] “Implementation of a network simulator supporting VoIP”, shows the design of a network simulation platform done in Java programming language. While the design of a platform from scratch provides the advantage of complete flexibility in the

implementation of traffic, link and node models; to provide a comprehensive set of networking hardware and network technologies is a very difficult task to achieve. A tangible limitation can be observed when comparing the specifications of the simulation platform described in [19] with OPNET (comprising hardware from more than 20 manufacturers and over 15 different network protocols). Finally but more importantly, no analysis is provided as to the effects of the utilization of different speech algorithms and the selection of the correct background traffic for the experiments.

For the particular case of audio and video traffic, many attempts have been made to create models that objectively assess the quality of a media received at the end of a communication channel. More precise and adaptive objective speech quality assessment models are developed [20, 21]; however, all of them use the subjective measurement of speech quality as reference. While every commercial and research test bed surveyed offers some form of objective speech quality assessment, all of them fail to provide the possibility of subjectively assessing the quality of speech in a simulation environment.

Summarizing, it was noticed that in every commercial product and research work surveyed at least two of the following deficiencies were present.

- Lack of support for variable rate speech algorithms. Variable rate speech encoding algorithms are currently in wide use in all voice networks. Variable rate codecs adapt to the impairments of the transmission channel optimizing bandwidth utilization while minimizing packet loss and in some cases jitter. A platform that intends to comprise the present-day state of VoIP communication has to include analysis of variable rate speech codecs.
- Limited selection of the conditions in which the experiments are performed, mainly background traffic relevant to voice networks. The comprehensive study of the environment in which an experiment is performed is vital to its success. Multiple factors determine the quality of speech transmitted over a packet switched network. Furthermore, the interaction between these factors is in most cases dynamic, complex and of contradictory nature. For a speech quality study, the correct selection of the parameters in which the experimentation is conducted can drastically affect the results obtained, possibly leading to wrong conclusions. To pay attention to the soundness and

validity of the conditions in which the measurements are performed is of paramount importance.

- Limited to objective speech quality assessment models. The possibility of injecting real world traffic to the simulation can be used to determine the correctness of the assumptions, parameters and general structure of the simulation model. Besides, it provides the modeller with more reliable quality information than any other objective assessment model could.

1.2 Research Objectives

The main objective of this thesis is the creation of a simulation platform aimed to support the planning and design phase of packet switched networks carrying voice traffic while considering realistic and current network conditions and simulation features. More specifically, the thesis aims to accomplish the following:

- Survey of existing network planning and simulation platforms oriented to support voice traffic over packet switched networks. Study of speech codecs and speech quality assessment models.
- Creation of simulation models relevant to the simulation of VoIP traffic.
 - Creation of a speech background traffic generation models. Background traffic should include constant and variable bit rate speech encoding algorithms as well as being scalable.
 - Creation of simulation models that allow the integration (injection, collection and playout) of real speech data to the simulation. Encoding of real audio should allow for constant and variable bit rate algorithms. This feature enables the utilization of subjective speech quality assessment models.
- Implementation of the simulation models in OPNET Modeler.
 - Implementation in OPNET Modeler of an objective and non-intrusive speech quality assessment methodology.
 - Validation of all new added features to OPNET platform through the design of simulation experiments and results analysis. Study of the impact of various design variables on the objective and subjective quality of speech.

1.3 Main Contributions

The main contributions of the thesis are:

- First, the thesis presents a generic OPNET model which can generate traffic that follows statistical behaviour of a number of speech encoders. The purpose of such a model is to provide relevant background traffic shape to VoIP simulations. Several fixed and variable data rate encoding algorithms (G.729, G.711, iLBC, Speex and AMR) are included in the codec choice for the model. Based on this, the following refereed conference paper is currently under review.

Rabassa, A., St-Hilaire, M., Lung, C.-H., Lambadaris, I., Goel, N. and Zaman, M. "New Speech Traffic Background Simulation Models for Realistic VoIP Network Planning"

- Second, an OPNET model that introduces real speech data to the simulation, allowing the subjective assessment of such data, is presented. An audio file containing real speech is encoded according to one of the available codec/mode (G.729, G.711 (A-law, μ -law, PLC, no-PLC), iLBC(20ms, 30ms), and AMR) and then recovered on the receiving end. The received and decoded file is played out to permit a subjective assessment of the level of degradation introduced by the network.
- Third, an objective non-intrusive speech quality assessment methodology is integrated to OPNET models. The E-model [22] (see Sections 2.2.2 and 3.2.4) is used to offer a Mean Opinion Score (MOS) value to the simulated speech data; permitting the modeller to establish a relationship between perceived quality and objectively measured speech quality.

1.4 Thesis outline

The rest of the thesis is organized as follows:

Chapter 2: Presents results of literature review and study of speech encoding algorithms. Based on the characteristics of the algorithms and the goals of the thesis, a set of codecs is selected. Similarly, a study is performed on relevant speech quality assessment models. An analysis is offered to select the models to be included as part of the present thesis.

Chapter 3: Introduces the implementations of the background traffic generation models, the real speech traffic generation models and the client node. Detailed explanation of the utilization of each model is provided. The OPNET Modeler implementation is emphasized.

Chapter 4: Presents simulations to validate the traffic generation models created. Real-world topologies with emphasis in the extraction of design ideas from the simulation results are discussed.

Chapter 5: A summary of the work performed is discussed. An overview of the simulation results is presented. Possible ways to enhance the research presented in this thesis are suggested.

Chapter 2

Related Works

VoIP refers to the integration of telephone services with the growing number of other IP-based applications; a digital telephone service that uses the public Internet as well as private networks instead of the traditional telephone network. A VoIP system simply converts analog signals such as telephone calls into digital IP packets and distributes these packets across Internet or any other packet-switched network.

VoIP is currently considered one of the most important technologies for telecommunications. In addition to the cost reduction achieved by enabling telecommunications resources to be used for multiple purposes, simplification of infrastructure through network convergence, and the opportunity to provide new and programmable services, VoIP is expected to accelerate the development of rich multimedia services. Some of the most obvious characteristics that make of VoIP a very popular telephony alternative are listed below.

- Very efficient use of the data bandwidth, equipment and transmission lines. VoIP technology allows the transmission of more than one telephone call and data over the same broadband connection. Additionally, speech can be encoded using different algorithms in accordance with channel capacity and quality requirements, permitting a more controlled and efficient use of the channel.
- Generally, the cost of VoIP services is low when compared to traditional telephony and features offered.
- Location independence. Only an Internet connection is needed to get a connection to a VoIP provider or server.
- Integration with other services available over the Internet, including video conversation, message or data file exchange in parallel with the conversation, audio conferencing, managing address books, and notification of availability.
- Advanced telephony features such as call routing, screen pops, and Interactive Voice Response (IVR) implementations are easier and cheaper to implement and integrate.

The rest of Chapter 2 will present two studies relevant to the development of a simulation platform aimed to assist with the planning and design stage of networks carrying VoIP traffic. Section 2.1 offers a comprehensive study of fixed and variable data rate speech codecs and explores the feasibility of integrating each algorithm to the proposed simulation tool. Section 2.2 presents a survey on speech quality assessment models. Also, an assessment of the viability of integration to the simulation platform proposed is discussed for each model.

2.1 Study of speech encoding algorithms and feasibility of integration to a simulation platform

In digital communications, the transmission of audio, more specifically voice, has challenged engineers and researchers for many years. Initially, motivated by military research for secure military radios, where very low data rates were required to allow effective operation in a hostile radio environment. Later, these techniques were available through the open research literature to be used for civilian applications, allowing the creation of digital mobile phone networks with substantially higher channel capacities than the analog systems that preceded them. More currently, with the development and expansion of the Internet and computer networks in general, the need for efficient algorithms offering high voice quality at low data rates has increased dramatically.

The word codec is a portmanteau of the words compressor-decompressor or, most commonly, coder-decoder. An audio codec is a device or software that compresses/decompresses digital audio data according to a defined algorithm. Plenty of research has been done with the objective to improve the existing encoding/decoding algorithms. Criteria as the quality of audio as perceived by humans, data rate and adaptability to channel bandwidth have been the main motivations in this field. Examples of recent research in this area are introduced in [23] and [24], both studies presented in 2008.

In the current section, a survey of the principles and technical specifications of the most popular codecs currently used will be presented. The surveyed encoder algorithms will be assessed based on popularity and feasibility of being integrated to the simulation tool proposed in this thesis.

2.1.1 General characteristics of audio compression codecs

In general, digital data can be compressed in a lossless manner, where the original information can be recovered completely from the encoded version. Also, lossy compression can be implemented where in the codification process part of the original information is discarded according to a certain algorithm.

Lossless coding offers little compression for digital audio and speech data. For a lossless codec applied to digital audio, the size of the output file is approximately 60% to 80% of the size of the original data where using lossy transformations output size can be 5% to 20% of the original one, maintaining in most cases an acceptable perceived quality. More importantly, we have that the compressed stream generated by a lossless audio encoder has a variable non-predictable data rate. Each second of audio can potentially occupy a different number of bits in the stream depending on the statistical nature of the sound [25]. Considering that bandwidth is one of the most important criteria to assess the functionality of an arbitrary encoding algorithm, the unpredictability of lossless encoders is not a desired effect for voice transmission; leaving lossy codecs as the best options. From this point on, all encoders mentioned are lossy algorithms.

A codec can be described through a set of parameters which defines its behaviour. The main codec attributes are bit rate, speech quality, quality degradation due to network impairments, algorithmic and processing delays and computational error. Generally, good performance for one of the attributes leads to poor performance of the others; the interrelation is mainly governed by fundamental laws of information theory and shaped by the algorithm used. The present section will focus on the properties of the encoder algorithms related to output bit rate and its control.

The data rate of a speech codec is generally measured as the average number of bit transmitted or generated in one second. Fix rate codecs use the same number of bits to encode a block of speech while variable bit rate codecs vary over time. Based on the concepts and ideas established by standard telephony communications, where a fix amount of resources is allocated to each conversation during its complete duration (circuit switching), there are a number of standardized fix rate codecs. Examples of such algorithms are the commonly used

ITU-T G.729 [26], ITU-T G.711 [27] and Internet Low Bitrate Codec (iLBC) [28]. In more modern circuit switched networks, like cell phone networks, speech codecs may have a number of modes, each mode with a different and fixed rate. Some authors do not consider this last type of codec as true variable bit rate codecs [25], examples will be analyzed in the thesis. Ultimately, in packet switched communication networks, both bit rate and packet size can be variable which leads to a more elaborated approach to variable rate codecs. For variable bit rate codecs, the number of bits used to encode a speech block is variable and can be adjusted on a frame by frame basis.

2.1.2 Fixed data rate speech encoder algorithms

Fixed output data rate speech encoding algorithms are still very common in voice communications. The simplicity of their implementation, low processing requirements and finally historical reasons account for their popularity.

Tens of fixed data rate encoding algorithms could be analyzed; however, for the purpose of the present investigation and considering the similarities among different algorithms, only three representative ones will be presented: G.711, G.729 and iLBC.

ITU-T G.711

Pulse Code Modulation (PCM) of voice frequencies, International Telecommunication Union (ITU) [29] Recommendation G.711 [13] was first released for usage in 1972 by the ITU. G.711 algorithm was profusely used in telephony, nowadays; it is included in almost every voice communication service and product.

G.711 uses a sampling rate of 8000 samples per second, with the tolerance on that rate 50 parts per million (ppm) [13]. Non-uniform quantization (logarithmic) with 8 bits is used to represent each sample, resulting in a 64 kbps constant bit rate. There are two slightly different versions; μ -law, which is used primarily in North America and Japan, and A-law, which is in use in most other countries outside North America.

G.711 encoding algorithm does not introduce look-ahead delay (this parameter will be discussed in Chapter 3). ITU-T G.711 Appendix I [30] defines a Packet Loss Concealment (PLC) algorithm and Appendix II [31] a Discontinuous Transmission (DTX) algorithm which

uses Voice Activity Detection (VAD) and Comfort Noise Generation (CNG) models to reduce bandwidth during silence periods of the conversation.

ITU-T G.729

ITU-T G.729 [26] became a standard in 1996, it is an algorithm for the coding of speech signals at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP). The encoder is designed to operate with a digital signal obtained by first performing telephone bandwidth filtering of the analogue input signal, then sampling it at 8000 Hz, followed by a conversion to 16-bit linear PCM. The output of the decoder should be converted back to an analogue signal by similar means [26]. Detailed operation of the encoder and decoder process can be found in [26] in clauses 3 and 4.

The output frame size of the encoder is 10 bytes and the frame duration is 10 milliseconds. The output bandwidth of the encoder is 8 kbps. In G.729, a 5 millisecond look-ahead is implemented (this parameter will be discussed in detail in Chapter 3) and the complexity of the algorithm is rated at 15, using a relative scale where G.711 [13] is 1 and G.723.1 [14] is 25. More than twelve annexes have been added to the original ITU-T G.729 [26] standard. In some of these annexes, new functionalities are described including but limited to variable bit rate and DTX.

Table 2.1 Bit rate modes for iLBC [28]

Codec mode	Frame size [bits]	Output bit rate [kbps]
20 ms	304	15.2
30 ms	400	13.3

iLBC

Internet Low Bitrate Codec [28] is a free speech narrowband encoding algorithm. Developed by Global IP Sound, it is suitable for IP voice communications, audio streaming and file archiving. The algorithm uses a block-independent linear-predictive coding (LPC) algorithm and supports two basic frame lengths: 20 milliseconds and 30 milliseconds and a sampling frequency of 8000 Hz. Table 2.1 shows frame size and bandwidth for each mode.

iLBC implements a packet loss concealment algorithm that exhibits a controlled response to packet losses similar to that defined in Appendix I of G.711 [30].

2.1.3 Variable data rate speech encoder algorithms

Variable rate speech codecs seems to go a long way in achieving flexibility to the channel conditions and better utilization of the bandwidth. For variable data rate algorithms, the quantization indexes used to define the output data rate are obtained by means of a lookup table or computation. Having to compute or lookup the quantization index in real time implies a considerable increase in the complexity of these codecs compared to fixed rate codecs, where acquiring the quantization index is a trivial procedure. When it is considered that for current technology and within certain boundaries, the efficient use of the bandwidth is a more relevant problem than the increase of computational complexity, it is easy to conclude that variable rate codecs present some advantage over fixed rate codecs.

The present section will focus on discussing the main attributes for several variable bit rate codecs and assessing the feasibility of integrating some of these codecs to a simulation platform.

ITU-T G.727

ITU-T G.727 [32] contains the specification of an embedded Adaptive Differential Pulse Code Modulation (ADPCM) algorithms with 5-, 4-, 3- and 2-bits per sample; at rates of 40, 32, 24 and 16 kbps, respectively. As specified in [32], ITU-T G.727 codec is recommended for the conversion of a PCM signal sampled at 8 KHz and sample size of 16 bits. A 64 kbps signal obtained, such as those specified by ITU-T G.711 [27] should be converted into 16 bit linear PCM before encoding.

ITU-T G.727 is an embedded algorithm; embedded ADPCM algorithms are variable bit rate coding algorithms with the capability of bit dropping outside the encoder and decoder blocks. This allows bit reductions at any point in the network without the need of coordination between the transmitter and the receiver.

Embedded ADPCM algorithms are characterized by (x, y) pairs where x refers to the enhancement bits and y refers to the core bits. In more detail, the difference between the input

and the estimated signal is quantized into codewords consisting of enhancement bits and core bits. The core bits are used in the prediction process, both in the encoder and the decoder, while the enhancement bits reduce the quantization noise in the reconstructed signal. Thus, while the core bits must reach the decoder to avoid mistracking, the enhancement bits can be dropped to alleviate congestion [33].

For a non-embedded codec, when packets arriving to an intermediate node in the route encounter that the traffic exceeds the available transport capacity, packets are dropped causing a considerable degradation in the speech quality. In the case of embedded codecs, the node has the ability to drop some or all of the enhancement bits as a way to alleviate congestion without dropping packets. The behaviour just described summarizes the main advantage of the embedded algorithms.

Table 2.2 Available modes for ITU-T G.727. Each pair (x,y) represents the (enhancement bits, core bits) [32]

Minimum data rate	Available modes	Maximum data rate
16 kbps	(5,2)	40 kbps
	(4,2)	32 kbps
	(3,2)	24 kbps
	(2,2)	16 kbps
24 kbps	(5,3)	40 kbps
	(4,3)	32 kbps
	(3,3)	24 kbps
32 kbps	(5,4)	40 kbps
	(4,4)	32 kbps

For ITU-T G.727 the four embedded ADPCM rates are 40, 32, 24 and 16 kbps, where the decision levels for the 32, 24 and 16 kbps quantizers are sub-sets of those for the 40 kbps quantizer. Table 2.2 shows the different operation modes for ITU-T G.727. Frame duration for G.727 encoding algorithm is 10 milliseconds.

Adaptive Multi-rate Compression (AMR)

Adaptive Multi-Rate (AMR) is a patented audio data compression algorithm intended for speech coding. AMR was adopted as the standard speech codec by 3rd Generation Partnership

Project (3GPP) in October 1998 [34] and is now widely used in Global System for Mobile Communications (GSM), Universal Mobile Telecommunications System (UMTS) and VoIP networks [35]. AMR was adopted by the 3GPP as the mandatory codec for 2.5G and 3G wireless systems based on the evolved GSM network. The AMR speech coder consists of the multi-rate speech coder, a source controlled rate scheme including a voice activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets.

The speech encoder takes its input as a 13-bit uniform PCM signal either from the audio part of the user equipment or on the network side, from the PSTN via an 8-bit A-law or μ -law to 13-bit uniform PCM conversion. The encoded speech at the output of the speech encoder is packetized and delivered to the network interface. In the receive direction, the inverse operations take place.

AMR Narrow Band (NB) operates at eight bit rates and was specifically designed to improve link robustness. AMR supports dynamic adaptation to network conditions, using lower bit rates during network congestion or degradation while preserving audio quality. By trading off the speech bit rate to channel coding, AMR maximizes the likelihood of receiving the signal at the far end.

Table 2.3 Bit rate modes and their associated bandwidth for AMR-NB codec [34]

Codec mode	Output bit rate [kbps]
AMR_12.20	12.20
AMR_10.20	10.20
AMR_7.95	7.95
AMR_7.40	7.40
AMR_6.70	6.70
AMR_5.90	5.90
AMR_5.15	5.15
AMR_4.75	4.75
AMR_SID	1.8

The AMR-NB codec operates at eight bit rates, Table 2.3 shows all eight modes plus the AMR_SID mode, which refers to silence descriptor (SID) frames. Frame duration for AMR-NB encoding algorithm is 20 milliseconds.

Enhanced Variable Rate Codec (EVRC)

Enhanced Variable Rate Codec Speech Service Option 3 [36] is a 3rd Generation Partnership Project 2 (3GPP2) standard. It was standardized in 2004. EVRC is a variable rate codec designed to serve as voice support for CDMA wireless mobile networks. More specifically, the service option 3 provides two-way voice communications between the base station and the mobile station using the dynamically variable data rate speech codec algorithm described in standard [36].

The transmitting speech codec takes voice samples and generates an encoded speech packet for every Traffic Channel frame. The receiving station generates a speech packet from every Traffic Channel frame and supplies it to the speech codec for decoding into voice samples. EVRC uses frames of 20 milliseconds duration of 8000 Hz, 16-bit sampled speech.

Table 2.4 Operation modes for EVCR codec [36]

Codec mode	Output bit rate [kbps]
Full Rate (Rate 1)	8.55
Half Rate (Rate ½)	4.00
Eighth Rate (Rate ⅛)	0.80

The rate determination algorithm (RDA) is used to select one of three encoding rates: Rate 1, Rate ½, and Rate ⅛. Active speech is encoded at Rate 1 or Rate 1/2, and background noise is encoded at Rate 1/8 [36]. The process of determining the data rate is based on the relation between the energy of the last received frame and the noise energy. This ratio, which is very similar in concept to the SNR of the signal is compared with two thresholds; if the energy ratio for the last frame is greater than the two thresholds then Rate 1 is selected, if it is only greater than one threshold Rate ½ is selected, in case where the energy ratio is lower than the lowest energy threshold Rate ⅛ is selected. In EVRC the output frames are of one of three different sizes as shown in Table 2.4.

Speex

Speex [37] is free software as well as open source. It is developed by Jean-Marc Valin and Xiph.org Foundation. Also, Speex is in constant development and improvement by a community of users and enthusiastic group of programmers. Currently, Speex is integrated into over 15 different platforms and applications and it is widely used in the VoIP environment since it is a codec specifically designed for the integration of speech to IP networks [37, 38]. Speex is utilized in Asterisk, Microsoft Netmeeting, Microsoft Xbox live and Google voice.

The Speex codec is designed to be very flexible and support a wide range of speech quality and bit-rate. Support for very good quality speech also means that Speex can encode wideband speech; i.e. 16 kHz sampling rate in addition to narrowband speech at 8 kHz sampling rate. Speex is capable of varying the complexity allowed for the encoder. An integer ranging from 1 to 10 controls how thoroughly the encoding search process is performed (1 being the mode that yields the least quality and the lower bandwidth and 10 where the best quality is achieved and the most bandwidth consumed). In practice, the best trade-off is between complexity 2 and 4, though higher settings are useful when encoding non-speech sounds like Dual-tone Multi-frequency (DTMF) tones [37].

Table 2.5 Operation modes for Speex narrowband codec [37]

Codec mode	Source codec bit-rate
0	250 bps
1	2.15 kbps
2	5.95 kbps
3	8 kbps
4	11 kbps
5	15 kbps
6	18.2 kbps
7	24.6 kbps
8	3.95 kbps

Speex integrated technologies like VAD, DTX, variable bit rate, variable complexity, adaptive jitter buffer, echo cancellation and some other features related to wideband and ultra wideband sampling. As defined in Section 2.1.1, Speex is considered a true variable rate codec; this means that the algorithm to determine the encoding rate is based exclusively on the signal properties and the codec configuration parameters.

Table 2.5 shows the different modes for Speex narrowband. As explained in previous section, the encoding algorithm will switch from rates depending on the input signal properties and codec setup. Speex can also work on fixed rate mode when defined in the configuration parameters.

2.1.4 Study of the feasibility of integrating encoding algorithms to a simulation platform

In this section, a brief discussion analysing the feasibility of integrating the above discussed encoding algorithm groups to a single simulation platform will be presented. The assessment will consider, whether the codec is an open source project, if an implementation of the codec can be found for research purposes and codec operation principles. Particularly for the implementation of the real speech generation models (see Section 1.2 and 3.2), modifications were performed to the encoder and decoder applications utilized. For this reason, when applicable, operation licensing terms of the encoder and decoder applications used, will be discussed.

2.1.4.1 Feasibility study for fixed data rate codecs

For all three fixed data rate encoding algorithms analyzed, in terms of codec operation principles, the complexity of the integration to a simulation platform is very low. To generate traffic that follows the behaviour of the codecs, only the frame size and frame duration needs to be obtained. This operation can be performed through the utilization of a simple lookup table.

For more complex simulation features, where the actual encoding and decoding algorithms are used (see Section 3.2), codec implementations were found for all three codecs. The source of the codec implementation employed and the pertinent licensing details are discussed below.

G.711 licensing information

Encoder algorithm G.711, PCM of voice frequencies, has been standardized by the ITU-T [13]. The ITU-T, with the purpose of providing a common set of tools that serves for testing and implementation guide of the standard in question has issued a library of portable and reliable software routines. Recommendation ITU-T G.191 [39], software tools for speech and audio coding standardizations, is a compilation of such software examples.

In [39] (Annex B), the license agreement of the software library is described as “general public license”. More explicitly Sections B.1 and B.2 of Annex B, specify the right to distribute, publish and modify any part of the modules included in the software library.

G.729 licensing information

The implementation of the algorithm is based on standard defined by International Telecommunications Union, ITU-T G.729 [16]. G.729 is licensed by Sipro Lab Telecom [40]. Sipro Lab Telecom is the authorized Intellectual Licensing Administrator for G.729 technology and patent pool. To this patent pool belongs, among others, a Montreal based telecom company, VoiceAge® [41]. As part of an open initiative program with the purpose to encourage the integration of G.729 algorithm to new technologies, VoiceAge allows restricted terms utilization of their G.729C implementation. The version C of the algorithm is very similar to the original standard with the only difference that implements floating point calculations, version C is fully compatible with G.729 and most other fix rate 8kbps versions of G.729 standard.

The terms and conditions of this operation are specified on their End User License Agreement for G.729(c) Implementation [42] which has been exhaustively analyzed in search of the reassurance that all restrictions are strictly observed. In addition, VoiceAge has agreed that their implementation of G.729C codec is integral part of the present research.

iLBC licensing information

iLBC is a narrow band speech codec developed by Global IP Solutions (GIPS) [43] (former Global IP Sound) and released publically in 2002. As described in the official GIPS license agreement [44], ILBC is royalty-free, non-exclusive licensed software. More specifically, in

Section 2.1 of [44], it is made explicitly that unless for commercial use, one may use, reproduce, display, perform and/or modify the original software provided by GIPS.

2.1.4.2 Feasibility study for variable data rate codecs

Continuing with the variable rate speech encoding algorithm we have that for ITU-T G.727 [45], codecs developed under this standard are not free; therefore, not open source implementation was found available. While analyzing the operation mode of ITU-T G.727, it can be noticed that the real advantage of this algorithm is achieved while introduced in a network which nodes are capable of supporting embedded codecs. Some of the nodes in the network are capable of adjusting the packet size by dropping the enhancement bits as required. None of reviewed simulation platforms natively support embedded protocols supporting embedded codecs. Given that the complexity of the implementation of such algorithm and the lack of information from manufacturers regarding existing protocols, the integration of ITU-T G.727 standard to a simulation platform is considered outside the scope of the present work.

AMR is not a free codec; hence, no open source implementation is available. However, for research purposes, an implementation have been made available by VoiceAge®, (see below for licensing information). They offer both the binary files for Win32 platform and the C source code for the algorithm. Attending at its operation principle, AMR adjusts its output bit rate based on network congestion information. To provide the implementation of a completely functional variable rate AMR codec, a means for the source node to acquire network load information needs to be implemented. To this end, Real-Time Control Protocol [46] (RTCP) will be included as part of the simulation implementation, making AMR-NB a viable option as variable rate codec for the present project.

AMR-NB licensing details

AMR joint patent pool includes the right of Ericsson, Nokia, Nippon Telegraph and Telephone and VoiceAge®. As for the case of G.729, VoiceAge has created an open initiative program with the purpose to encourage the integration of AMR-NB algorithm to new technologies. The terms and conditions of this project are specified on the End User License Agreement for AMR-NB implementation [47] which has been exhaustively analyzed in search of the reassurance that all restrictions are strictly observed. In addition, VoiceAge has agreed that their implementation of AMR-NB codec is integral part of the present research.

After studying the EVCR variable rate speech codec, the analysis for the potential integration with a simulation platform is simple. Regardless of whether EVCR is free, which is not, the definitive fact is that EVCR is not designed for VoIP networks. EVCR is a codec with the specific purpose to operate over wireless networks, where the traffic performance and network impairments have a completely different behaviour than for packet switched networks. For instance, VoIP codecs implement solid algorithms for packet loss concealment and expend very little effort in compensating for possible corrupt packets; in the case of speech codecs for wireless environment is the opposite. This is more evident in a CDMA environment where the transmission of the minimum possible power is a crucial premise for the operation of the whole network. Simply, EVCR is not used in VoIP networks. Therefore, and according to the objectives of the present work, the integration of EVCR is considered out of scope.

Speex, among all other studied codecs, offers the best possibility of being integrated to a simulation platform. The fact that is open source and it is a true variable rate codec facilitates the process integration to a simulation platform.

Speex licensing details

In Section D and E of [37] the licensing terms for Speex usage and distribution is presented. Distribution of the codec is allowed under the condition that the copyright notices are transferred. Modifications are allowed with no restriction if no distribution is to be performed.

As a summary, attending to the codec structure and design purpose and licensing terms of the presented codecs, the following algorithms are going to be integrated as part of the simulation tool outlined in Section 1.2 of the present report: G.711, G.729, iLBC, AMR-NB and Speex.

2.2 Study of Speech Quality Assessment Models

Digitized voice and data traffic have considerably different requirements when transmitted over a network. Voice traffic is considerably sensitive to network delay and packet loss while data is significantly more tolerant to these network impairments. Also, the nature of voice traffic shape is up to a certain degree smooth and constant when data traffic is *bursty*.

IP networks present significant new challenges to the delivery of real-time voice traffic. Whereas the circuit-switched PSTN guarantees that sufficient bandwidth is reserved and

available for the duration of the call, IP networks, in general, do not. Delay is not guaranteed to be either minimal or constant in an IP network. In addition, dropped packets and packet delay variation introduce distortions not found in traditional telephony. Low bit rate codecs used to reduce required bandwidth distort the original waveform significantly before it is even transmitted.

For a system with the aforementioned difficulties, assessing the quality of speech is crucial. Speech quality measures are used to optimize the design of speech transmission algorithms and equipment, and to aid in the selection of coding algorithms for standardization. It is crucial that networks and terminals are properly designed to constantly monitor the quality of VoIP services, taking the necessary action to maintain the level of service.

In the introduction for Section 2, a basic concept of VoIP systems is presented as well as the main advantages of IP telephony compared to standard PSTN, marking the importance of the former in present-day telecommunication tendencies. Also, the importance of measuring and predicting voice quality for VoIP systems as a tool in designing and/or improving the network infrastructure is introduced. In this section, some of the most relevant and current voice quality assessment models will be analyzed. Emphasis will be put in categorization of speech quality assessment algorithms and their circumstantial advantages and disadvantages.

2.2.1 Understanding VoIP Quality Assessment Models

In this section, three different categorizations for quality assessment models and the parameters used to categorize them are presented. Differences in the usage of the different models are mentioned.

2.2.1.1 Subjective and Objective Assessment models

There are two broad classes of speech quality metrics: subjective and objective.

Subjective measures involve humans listening to a live or recorded conversation and assigning a rating to the perceived quality, see Figure 2.1. This rating can be either a single overall quality rating or a rating of a particular characteristic (i.e., clarity or listening effort) or a particular distortion (i.e., clipping, hum). To ask humans to listen to speech excerpts and then solicit their opinion is considered the most reliable method for assessing voice quality [48].

The subjective method involves presenting a set of speech samples that characterize different conditions of the system under test to human listeners in a controlled environment. After each presentation, subjects are required to grade the sample on a simple discrete opinion scale. For each sample, votes are averaged. Clearly, a metric obtained as described above can be a good measure of perceived speech quality. However, subjective metrics have disadvantages, too. In particular, they can be time-consuming and expensive. Some researchers or organizations may not have the resources to conduct the tests. Certainly, such metrics cannot be used in any sort of real-time or online application for the purpose of speech quality prediction.

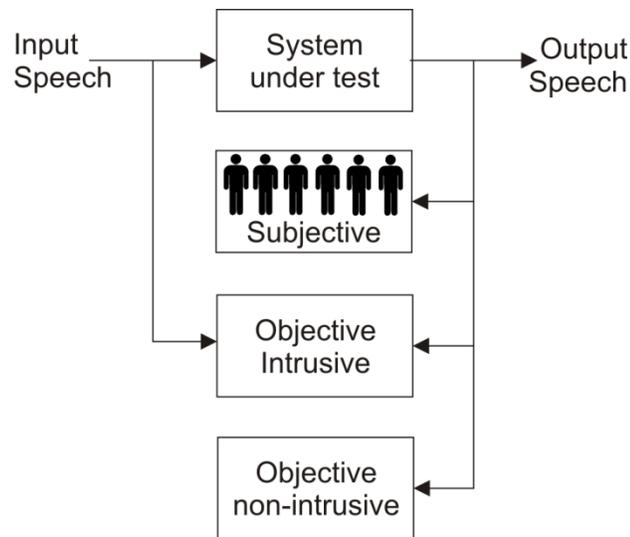


Figure 2.1 Subjective, Objective intrusive and Objective non-intrusive speech quality assessment models

For many years, more automated methods have been investigated with the aim of modeling human behaviour and objectively predicting subjective scores. There was an obvious necessity for finding models that could estimate speech quality from analyzing the physical characteristics of the terminals, networks or directly from the speech signals that they deliver. Such a process is called objective speech quality assessment model.

Objective models were first introduced for evaluating the sound quality of audio systems [49]. An estimation of the distortion introduced by the system under test can be obtained by comparing the signal processed by the system (degraded signal) with the original audio (reference signal). More elaborated models have been developed where only the received (degraded signal) is scrutinized to obtain the final score, see Figure 2.1. Typically, the accuracy

or effectiveness of an objective metric is determined by its correlation, usually the Pearson (linear) correlation¹, with a subjective score for a set of data [50]. If an objective metric has high correlation with a subjective score, then it is deemed to be an effective measure of perceived speech quality, at least for speech data and transmission systems with the same characteristics as those in the experiment.

2.2.1.2 Intrusive and Nonintrusive Assessment models

A common categorization applied to the objective models is related to the signals used to generate or compute the final quality score. Usually, these two classes are referred to as: Intrusive or active and nonintrusive or passive.

Intrusive methods are those that estimate the speech quality by measuring the difference between the input and output speech signals and mapping the distortion values to a predicted quality metric, see Figure 2.1. In contrast to subjective experiments, this model enables extensive testing to be performed over short periods, and is therefore beneficial during codec and equipment development/selection. However, the need for a test signal adds extra load on the network and in some applications requiring objective speech quality evaluation, the input speech signal may not be readily available or may not be compared directly to the output, like in live calls where no clean reference signals are available.

In such cases, an attractive alternative approach is to assess speech quality using only the output signals, i.e., a nonintrusive speech quality evaluation, see Figure 2.1. An effective nonintrusive speech quality measure will be of significant importance to applications where an appropriate input signal is not readily available, e.g., the performance evaluation of hearing aids and nonintrusive performance monitoring of communication systems such as wireless communications and VoIP.

¹ Obtained by dividing the covariance of the two variables by the product of their standard deviations. Pearson's correlation reflects the degree of linear relationship between two variables.

2.2.1.3 Speech-Layer, Packet-Layer and Opinion Models

Objective quality assessment methodologies can be categorized into several groups from the viewpoints of objective, measurement procedure and input information used by the quality assessment model. These categories are listed below.

Speech-Layer Models

Objective models require speech signals as inputs and produce estimates of listening quality [51, 52]. Some different approaches have been developed for this category. From the very simple models where waveform distortion is examined by comparing input and output signal to more elaborated models based on spectral distortion and digital filtering applied to the audio signal(s).

Packet-Layer Models

Objective models that exploit IP packet characteristics only and produce estimates of listening quality are called Packet-layer models [51, 52]. These models have the advantage of being able to monitor the quality of real phone calls as they progress, and are usually implemented in the IP phone or gateway.

Although the speech-layer and packet-layer objective models estimate the same parameter (i.e., the listening quality), they are used in different scenarios. For instance, if it is impossible or difficult to obtain actual speech samples via in-service quality monitoring, we should use packet-layer objective models. Conversely, if it is difficult to capture necessary packet information or we need to obtain quality estimates that are as accurate as possible, we should use speech layer objective models.

Opinion Models

Opinion models are those objective quality assessment methodologies that exploit network and terminal quality parameters to produce estimates of conversational quality [51, 52]. These models are referred to as the most complete ones since they consider quality degradation introduced by speech coding, bit error, distortions introduced by acoustic transducers, as well as impairments specifically related to an IP network, like packet loss, jitter and network delay. Opinion models could be seen as a combination of the two previously mentioned models.

Opinion models have long been studied but not many proposals have reached the category of standard.

2.2.2 Description of relevant speech quality assessment models for VoIP systems

After a study of the most relevant and used speech quality assessment models for VoIP traffic, and based on the categorization presented in Section 2.2.1, this section will discuss the main characteristics, advantages and disadvantages of some of these models.

2.2.2.1 Subjective Quality Assessment Models

The most relevant subjective models are briefly described in this subsection

ITU-T P.800. Methods for subjective determination of transmission quality

The Opinion Rating Method also known as Mean Opinion Score (MOS), which is defined by the ITU-T Recommendation P.800 [53] for VoIP quality assessment, is the most widely used method by which subjective VoIP quality is assessed. MOS is just what the name implies: it's a score derived from users' opinions. A sentence is read aloud over the telephone to a number of listeners. After hearing the sentence, the listeners score the conversation based on their opinion of how it sounded. The scores are averaged to come up with the mean opinion score. After years of testing, the ITU used data from listener opinions to codify a scoring standard. Subjective quality assessment models can be grouped into listening quality and conversational quality. Where for the listening quality tests, subjects listen to a number of distorted recordings and vote their opinion of the quality; for the conversational quality assessment, subjects are involved in a two-way communications.

Table 2.6 Absolute Category Rating (ACR) scales. Listeners are offered a recording from which they have to assess: quality of the speech, effort required to understand the meanings of sentences and loudness preference [53]

MOS	Quality	Listening Effort	Loudness
5	Excellent	No effort required	Much louder than preferred
4	Good	No appreciable effort required	Louder than preferred
3	Fair	Moderate effort required	Preferred
2	Poor	Considerable effort required	Quieter than preferred
1	Bad	No meaning understood	Much quieter than preferred

ITU-T Recommendation P.800 [53] defines various ways to collect information from the listeners. The most obvious and probably more used is the Absolute Category Rating (ACR), defined in Annex B of the ITU-T P.800 recommendation [53]. ACR defines three scales in order to assess three different characteristics or properties of the recording: quality of the speech, effort required to understand the meanings of sentences and loudness preference; see Table 2.6.

Table 2.7 Degradation Category Rating (DCR) scale [53]

MOS	Degradation Rating (DRC)
5	Degradation is inaudible.
4	Degradation is audible but not annoying.
3	Degradation is slightly annoying
2	Degradation is annoying
1	Degradation is very annoying

The ACR method tends to lead to low sensitivity in distinguishing among good quality circuits. A modified version of the ACR procedure, called the Degradation Category Rating (DCR) [53] procedure, affords higher sensitivity. The DCR procedure, which in particular uses an annoyance scale and a quality reference before each configuration to be evaluated, is suitable for evaluating good quality speech. The stimuli are presented to listeners by pairs (*A-B*) where *A* is the quality reference sample and *B* is the same sample that is processed by the system under evaluation. The purpose of the reference sample is to anchor each judgment of the listeners. Table 2.7 shows the MOS scale for DCR tests.

Table 2.8 Comparison Category Rating (CCR) scale [53]

MOS	Comparison Rating (CCR)
3	Much Better
2	Better
1	Slightly Better
0	About the Same
-1	Slightly Worse
-2	Worse
-3	Much Worse

Lastly, the Comparison Category Rating (CCR) [53] is proposed. The CCR method is similar to DCR described before. Listeners are presented with a pair of speech samples on each trial. The only difference is that in the DCR procedure the samples are presented always in the same order: unprocessed and then the processed sample; in the CCR technique the order of the processed and unprocessed samples is chosen at random for each trial. Table 2.8 shows the MOS values for CCR.

The Opinion Equivalent-Q Model

The Opinion Equivalent-Q Model is a variation of the Opinion Rating Model. The MOS of a system obtained using the Opinion Rating Model is dependent on a lot of factors that include such details as the testing date and the mix of quality levels in the experiment [51]. This means that if on a certain date a lot of good-quality condition calls are available on the system, the MOS for certain system condition would be higher than obtained when there are fewer good-quality call conditions existing in the system. The Opinion Equivalent-Q model uses a Modulated Noise Reference Unit (MNRU) defined in ITU-T P. 810 recommendation [53], which is a reference system that outputs a speech signal and speech-amplitude-correlated noise with a flat spectrum. The ratio of signal to speech-correlated noise in dB is called the Q value. The opinion equivalent Q is thus defined as the Q-value of MNRU speech with quality equivalent to that of the speech under evaluation.

The Diagnostic Acceptability Method for Speech Communication Systems

The Diagnostic Acceptability Method (DAM) was proposed in 1976 [55]. It is very similar in principles to ITU-T Recommendation P.800 [51]. The DAM model is not currently used in the industry and the brief description offered below is intended merely to set a chronological reference point in the development and evolution of voice quality assessment models.

Two important features distinguish the DAM approach. First is the fact that it combines a direct (isometric) and an indirect (parametric) approach to acceptability evaluation. The isometric approach requires the listener to provide a direct, subjective assessment of the acceptability of a speech sample, whereas the parametric approach requires the listener to evaluate the sample with respect to various perceived qualities, independently of her/his affective reactions to them.

A second distinguishing feature of DAM is that it solicits separate reactions from the listener with regards to what she/he perceives to be the speech signal itself, the background and the total effect.

2.2.2.2 Objective Quality Assessment Models

Several objective (active) models have been developed over the years, only a few of them had reached the category of standards according to International Telecommunications Union. In this section, the most relevant and current methods are discussed.

Objective Intrusive Quality Assessment Models

This section will review the most relevant objective intrusive models. According to the categorization shown in Section 2.2.1 all intrusive models are considered speech-layer models.

ITU-T P.861. Perceptual Speech Quality Measure (PSQM, PSQM+, PSQM/IP)

Perceptual Speech Quality Measure (PSQM) is a widely used intrusive objective quality assessment model which is derived from Perceptual Audio Quality Measure (PAQM) [56].

PSQM, which is an ITU-T P.861 [57] recommendation, works by injecting a testing source signal for artificial voice, defined in ITU-T P.50 [58] recommendation, with an active speech level signal of -20dBm into the VoIP network; then the distortion experienced when transmitted

through various codecs and transmission media is measured and finally a MOS score is estimated.

The PSQM has the advantage of being able to measure the effects of various impairments and their interactions but has the disadvantage of requiring a call set up each time tests need to be carried out.

At the time PSQM was standardized as ITU-T P.861 [59], the scope of the standard was to assess speech codecs, used primarily for mobile transmission, like GSM. VoIP was not yet a topic at the time. The requirements for measurement equipment have changed dramatically since then. Recommendation ITU-T P.861 standard [57] was reviewed to cope with the new demands arising from next generation networks like VoIP. Within these networks, the measurement algorithm had to deal with much higher distortions than with GSM codecs, but perhaps the most eminent factor is that the delay between the reference and the test signal was no longer constant.

A first approach to overcome such problems was the development of PSQM+ [59]. It handled well the larger distortions caused by burst errors, but still had significant problems with the compensation of the varying delay. An advanced delay tracking feature was added to the standard, the new enhanced procedure was called "PSQM/IP" [59] and it implemented an easy way to solve the varying delay issue in most cases, without losing the option of real-time operation.

ITU-T P.862. Perceptual Evaluation of Speech Quality (PESQ)

With the new ITU standard ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ) [60] the problem of measuring large distortions and delays was greatly alleviated. PESQ combines the excellent psycho-acoustic and cognitive model of PSQM+ [59] with a time alignment algorithm adopted from Perceptual Analysis Measurement System (PAMS) [59], other of the precursors of PESQ, which handles varying delays accurately. PESQ is not designed for streaming applications, which is its only drawback. This is why it cannot fully replace PSQM+. With PSQM and PESQ there are two intrusive standards that cover the entire problem of measuring speech quality.

Using PESQ, a reference signal is played through the system, and the received version is compared to the original reference version. Any degradation that occurred is measured, and a quality score is computed [59]. The resulting score is often mapped to a MOS-LQ value. PESQ has good correlation with subjective listening quality across a very large corpus of tests covering a wide range of narrowband telephony applications.

Because PESQ is an active, or intrusive, monitoring approach, it may not be ideal for networks that are already near operating capacity.

Objective Nonintrusive Quality Assessment Models

Objective nonintrusive quality assessment models are currently the most popular ones. As outlined in Section 2.2.1, objective passive procedures are very attractive when an input signal is not available, allowing real-time assessment. Also, passive methods do not increase the traffic in the network under test.

According to the categorization shown in Section 2.2.1, nonintrusive models can be found in any of the three groups: speech-layer, network-layer and opinion model; depending on their scope and network conditions.

ITU-T P.563 Single-sided method for objective speech quality assessment in-narrow band telephony applications. (Speech-layer)

Up to this point, the solutions provided by subjective models (see Section 2.2.2.1) and objective intrusive models (see Section 2.2.2.2) presented themselves very beneficial during codec and equipment development/selection. However, the need for a test signal adds extra load to the network and live calls cannot be measured because no clean reference signal is available. To provide the industry with a solution for the listening quality assessment of live calls, the ITU-T opened a competition in 2002 with the aim of standardizing a method that does not need the reference signal for estimating the voice quality.

A voice activity detector is used to identify portions of the signal that contain speech and the speech level is calculated. Finally, a speech level adjustment to -26 dB is applied. The speech signal to be assessed will be investigated by several separate analyses, which detect, like a sensor layer, a set of characterizing signal parameters. This analysis will be applied at first to

all signals. Based on a restricted set of key parameters, an assignment to a main distortion class will be made. The key parameters and the assigned distortion class are used for the adjustment of the speech quality model. This provides a perceptual based weighting where several distortions are occurring in one signal but one distortion class is more prominent than the others.

The ITU-T P.563 [61] approach is the first recommended method for single-ended nonintrusive measurement applications that takes into account the full range of distortions occurring in public switched telephone networks (may consider background noise, filtering and variable delay, as well as distortions due to channel errors and speech codecs) and that is able to predict the speech quality on a perception-based scale MOS-LQO according to ITU-T Rec. P.800.1 [62]. This Recommendation is not restricted to end-to-end measurements; it can be used at any arbitrary location or segment in the transmission path.

Vector quantization techniques. (Speech-layer)

Two models will be briefly described using very similar technique to predict speech quality.

The first model is called “Vector Quantization Technique for Output-Based, Objective Speech Quality” [63]. In this model, Perceptual Linear Prediction (PLP) coefficients based on models of hearing perception to extract speaker independent speech parameters are used. The order of the PLP model specifies the amount of detail in the auditory spectrum preserved in the PLP model [63]. 5th order PLP model was found to be effective for speaker independent speech recognition tasks. PLP, PLP cepstrum, and PLP delta-cepstrum parameters are computed for output speech records from an undistorted source speech database and vector quantized. The resulting codebook provides a reference for computing objective distance measures for distorted speech. The obtained objective measures are the transition probability distance, the median minimum distance, and the chi-squared distance.

The second model is called “New Output-Based Perceptual Measure for Predicting Subjective Quality of Speech” [64]. The system is based on computing objective distance measures, such as the median minimum distance, between perceptually-based parameter vectors representing the voiced parts of the speech signal to appropriately matching reference vectors extracted from a pre-formulated codebook. The distance measures are then mapped into equivalent Mean

Opinion scores using regression. The codebook of the system is formed by optimally clustering large number of speech parameter vectors extracted from undistorted source speech database [64]. The required clustering and matching processes are achieved by using an efficient data mining technique known as the Self-Organizing Map. The perceptual-based speech parameters are derived using Perceptual Linear Prediction (PLP) and Bark Spectrum analyses.

Nonintrusive Speech Quality Evaluation Using an Adaptive Neurofuzzy Inference System. (Speech-layer)

This nonintrusive evaluation method is using an adaptive neurofuzzy inference system, called the NISQE-ANFIS method [65], which does not require an artificial reference or codebook and operates just on the output signal. In contrast to other nonintrusive methods, the proposed technique applies a first-order Sugeno-type fuzzy inference system (FIS) to objectively estimating speech quality. The features required for the NISQE-ANFIS method are extracted from the perceptual spectral density distribution of the input speech. The premise and consequent parameters of the FIS, as constructed by the ANFIS, are learned by the back-propagation and least-squares algorithms.

In the proposed NISQE-ANFIS method [65], the input features were first extracted by measuring the distributive characteristics of speech perceptual spectral density. The extracted features were then fed into the first-order Sugeno-type FIS, where each input feature was *fuzzified* by Gaussian membership functions. In the NISQE-ANFIS method, the algebraic product is used as a T-norm operator for the fuzzy AND operator and the overall output are calculated using the weighted average [65].

Real-time monitor for measuring live VoIP call quality. P.VTQ (packet-layer)

The ITU P.VTQ (Voice Transmission Quality) standard defines an endpoint MOS algorithm based on PESQ. A passive type of monitoring approach, P.VTQ looks at the call quality of real phone calls as they progress, and is usually implemented in the IP phone or gateway. The P.VTQ standard defines a quality value called the K-factor. Like the R-value, the K-factor is mapped to the MOS. More precisely, the K-factor is a MOS-LQ value because it includes impairments such as packet loss and jitter discards, but it does not include delay-related impairments [66]. The K-factor is usually calculated over 8-second samples of the audio for a given call.

IP phones and gateways from Cisco Systems implement the P.VTQ standard as a means of calculating a MOS for VoIP phone calls. At the end of a call, the phones and gateways can provide the quality information for that call.

ITU-T G.107. The E-Model, a computational model to be used in transmission planning (opinion model)

The E-Model was originally developed by the European Telecommunication Standard Institute between 1993 and 1996 as a transmission-planning tool. Finally, it was standardized in ITU-T Rec. G.107 [22] as an analytic model for voice quality estimation. The E-model is based on the concept that psychological factors on the psychological scale are additive and that each impairment factor which affects a voice call can be computed separately but remain correlated. The basic result of the E-model is the calculation of the transmission rating R-factor, which is a simple measure of voice quality ranging from 0 (poor) to 100 (excellent). The R-factor is then used to determine the Mean Opinion Score using an R factor to MOS mapping. The R-factor covers such degradation qualities as echo, background noise signal loss, codec impairments, etc. According to [22] the R-factor is related to MOS through the following set of expressions.

$$\begin{aligned}
 MOS &= 1 && \text{if } R < 0 \\
 MOS &= 1 + 0.035R + R(R - 60)(100 - R)7 \times 10^{-6} && \text{if } 0 < R < 100 \\
 MOS &= 4.5 && \text{if } R > 100
 \end{aligned}
 \tag{2.1}$$

Table 2.9 R-factor to MOS mapping [22]

R-Factor	Quality Rating	MOS
90 < R < 100	Excellent	4.34 – 4.5
80 < R < 90	Good	4.03 – 4.34
70 < R < 80	Fair	3.60 – 4.03
60 < R < 70	Poor	3.10 – 3.60
50 < R < 60	Bad	2.58 - 3.10

Based on this R-factor to MOS score mapping is shown in Table 2.9. The MOS for a VoIP network using the E-model for quality estimation reflects the degradation caused by lossy compression, packet loss, voice clipping, jitter in active voice periods, interference signal, noise and excessive attenuation distortion.

Equation 2.2 shows in more detail the calculations of the R-factor as provided in ITU-T G.107 [22]. Where R_0 represents in principle the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise. The factor I_s is a combination of all impairments which occur more or less simultaneously with the voice signal. Factor I_d represents the impairments caused by delay and the effective equipment impairment factor I_e represents impairments caused by low bit-rate codecs. It also includes impairment due to packet-losses of random distribution. The advantage factor A allows for compensation of impairment factors when there are other advantages of access to the user [22].

$$R = R_0 - I_s - I_d - I_e + A \quad [2.2]$$

In Section 3.7 of ITU-T G.107 [22], default values for all factors in equation 2.2 are provided. It is strongly suggested in the same section that the default values are adopted for all factors of equation 2.2 that are not varying during the calculations. Following this principle, equation 2.2 can be written in a simple manner as shown below in equation 2.3.

$$R = 93.2 - I_d - I_e \quad [2.3]$$

From equation 2.3, the R-factor, hence the MOS, can be assumed as a function of the Delay Impairment and the Equipment Impairment factor. In [22], an elaborated analysis is provided for the calculations of these two impairments. Equation 2.4 represents the factors involved in the delay impairment, where The factor I_{dte} gives an estimate for the impairments due to Talker Echo, The factor I_{dle} represents impairments due to Listener Echo and The factor I_{dd} represents the impairment caused by too-long absolute delay T_a , which occurs even with perfect echo cancelling [22].

$$I_d = I_{dte} + I_{dle} + I_{dd} \quad [2.4]$$

In [22], a model is offered to calculate the equipment impairment factor. Equation 2.5 shows all factors involved in the calculation of I_e factor values for codec operation under random packet-loss. Where I_{e0} represents the equipment impairment factor for zero packet loss conditions, the values of the parameter are input independent and recommended values are presented in UIT-T G.103 [67] for different codecs. B_{pl} conveys the packet loss robustness of an arbitrary codec

and the default values are also covered in [67]. The packet loss probability is represented by P_{pl} . $BurstR$ is the packet loss burst ratio.

$$I_e = I_{e0} + (95 - I_{e0}) \frac{P_{pl}}{\frac{P_{pl}}{BurstR} + P_{pl}} \quad [2.5]$$

From equation 2.5, the packet loss burst ratio is defined as:

$$BurstR = \frac{\text{average length of observed bursts in an arrival sequence}}{\text{average length of bursts expected for the network under "random loss"}}$$

2.2.3 Selection of the speech quality assessment models to integrate to the simulation platform

The objective of the proposed investigation is to generate a simulation platform which main function is to serve as a tool or aid in the task of designing and improving IP networks carrying voice traffic. The speech quality assessment model selected has to accommodate and comply with the mentioned research specifications.

As stated in Section 1.2, among the objectives of the present thesis are to invest the simulation platform with the possibility to utilize a subjective speech quality assessment model and an objective non-intrusive model.

Using as starting point the grouping proposed in Section 2.2.1, we have that the two main categories can be defined as subjective and objective speech quality assessment models. The main advantages and disadvantages of subjective assessment models were presented in Section 2.2.1.1. Although expensive and difficult to implement, subjective speech quality assessment models are the only truly accurate tool to evaluate the impact of speech degradation in humans. One of the main objectives of the thesis is to allow the injection of real speech audio into the simulation, the recovery and playout of the degraded audio. With this implementation, any of the subjective quality assessment model presented in Section 2.2.2.1 can be employed. It becomes a preference of the modeller which method to follow.

Focusing on the objective quality assessment models, we have that the two categories with respect to the signals used to generate the final score are intrusive and nonintrusive assessment models. In principle, given that a way to inject real speech audio will be provided, intrusive objective speech quality assessment models are possible. Nonetheless, some properties of these

models make them incompatible with the main objective of the thesis; i.e., to serve as aid tool for the planning/design stage of networks carrying VoIP traffic. First, since intrusive models are focused in a comparison of the original audio signal with the degrade signal, the acquisition of a correlation between speech quality and network conditions is a very difficult task. Secondly, these models only offer a listeners quality MOS, once again, since only the input and output are analyzed, the total delay in the conversation is not considered. The dynamism of the conversation is not evaluated and considered in the final score. Finally, given that the use of the present tool will be focused on the planning of networks that are not built yet, the insertion of real traffic is not possible in this scenario.

For nonintrusive models, we have the categorization suggested in Section 2.2.1.3 where quality assessment algorithms can be defined as speech-layer, packet-layer or opinion models. Starting with the speech-layer models, these models are based on analog signal processing procedures. Models range from time domain analysis to spectral distortion and digital filtering analysis. As explained before, the importance of obtaining correlation between network parameters and speech quality is paramount and that is not possible through speech-layer models.

Ultimately, we are left with objective packet-layer models and opinion models. Packet-layer models are designed to deliver in real time a score for a phone conversation. The results are founded purely on statistical computations based on the IP packets that arrive to a specific node of the network being studied. The opinion model however, is designed to integrate and consider multiple factors related to the network (propagation delay, packet drop, etc.), to the transmission channel (background noise, signal loss, etc.) and to the equipment impairment (echo, codec, etc.).

Considering only the requirements for the two aforementioned algorithms, both are possible in the framework of the proposed research; the input data for both models can be obtained from the simulation software and processed accordingly. Even so, if the project goal is reviewed, it is simple to realize a critical advantage of the opinion model over the packet-layer model. If we focus on the value of the model's output oriented to network design and development, it is obvious that the packet-layer model offers no information regarding the conditions and factors in the network and equipment that generated the results. Consequently, packet-layer models offer no direct input to the decision making process involved in designing and improving a

network to support voice traffic. On the other hand, every time a score is obtained from the opinion model, associated to this result are all the factors that, integrated through a mathematical equation, generated the final score. A whole set of values are obtained from each simulation, allowing the network designer to take steps into much more defined directions and iteratively check the results of the modifications. As demonstrated, the definitive option to be employed in the current investigation is the opinion model group.

Once the group has been defined, the next task would be to find the more appropriate algorithm within that group. For the opinion model group, the most studied and discussed is by far ITU-T G.107 the E-Model [22], described in this section. The E-model offers simple and effective ways to find correlations between the speech quality prediction and the network impairments (see Section 3.2.4 for more details).

Chapter 3

VoIP Simulation Models and OPNET Implementations

The current chapter describes the simulation models created as well as discusses their integration into the OPNET Modeler. More precisely, Section 3.1 will cover the generation of speech background traffic generation models and the integration of such models to OPNET Modeler. In Section 3.2, the models to integrate real speech traffic to the OPNET simulation will be discussed. Also, in Section 3.2, the implementation of an objective speech quality assessment model will be presented. Finally, Appendices A and B complement this chapter. In Appendix A, detailed information regarding the attributes of the new nodes created in OPNET will be offered. Appendix B offers instructions, in a tutorial manner, on how to use the created nodes in an OPNET simulation.

3.1 Speech background traffic generation

The purpose of a background traffic generation is to offer a simulation history, or simulation warm up traffic, that statistically follows real speech traffic. Actual encoding technologies like, packet loss concealment, DTX, VAD, echo cancellation or noise suppression are not implemented. Additionally, a traffic generation node and a traffic sink node, when placed in a simulation, could represent a simple VoIP phone, a computer or even the aggregated traffic coming from a subnet.

OPNET provides effective ways to generate traffic related to most popular protocols and applications, some of them are HTTP, FTP, email traffic, database access traffic, printing applications. Consequently, the addition to the background traffic generation procedures has been reduced to voice traffic exclusively

To facilitate the control of the traffic volume generated, a traffic scalability model has been added to the speech traffic generation node. The methodology used to create the traffic growth is linked to the nature of the traffic generated: telephone conversations.

Summarizing, two features have been added to the generation of speech background traffic in OPNET: (1) new voice encoding algorithms and (2) a traffic scalability methodology that properly adapts to the nature of the voice traffic. Both attributes will be discussed in the next sections.

3.1.1 Available encoding algorithms and parameter calculations

As mentioned, dealing with codecs that describe the current trends of the VoIP world is of vital importance to achieve a reliable simulation. After detailed analysis, the following encoding algorithms were selected according to their popularity and with the purpose to represent the main categories of speech encoding algorithms, see Section 2.1 for more details.

Fixed rate speech encoding algorithms

In the fixed data rate category, four encoder methods have been selected, they are listed below.

- ITU-T G711. Fixed output data rate of 64 kbps
- ITU-T G729. Fixed output data rate of 8 kbps
- iLBC 20ms. Fixed output data rate of 15.2 kbps
- iLBC 30ms. Fixed output data rate of 13.33 kbps

Variable rate speech encoding algorithms

Based on analysis provided in Sections 2.1.3 and 2.1.4, the following variable data rate algorithms will be integrated into the simulation models.

- **AMR**: algorithm designed to adapt the output data rate according to network conditions; hence, it requires information regarding network status to operate in variable rate (VAR) mode; see Section 2.1.3 for more details.
- **Speex**: Free open source codec that adapts the output data rate depending on the statistical properties of the analog audio. Two modes will be implemented, Speex low quality with Quality = 2 and Speex high quality with Quality = 8 [37, 68]. See Section 2.1.3 for more details.

Only a small subset of the speech encoding algorithms selected are already implemented by default in OPNET Modeler, as the research started and carried out. This is the case of G.711

and G.729A. When investigating the possibility of reusing those OPNET modules associated to G.711 and G.729 codec in the proposed background traffic generator models, the following considerations were assessed:

- (a) The simplicity of isolating the OPNET source code corresponding to the traffic generation for both codecs;
- (b) The complexity of implementing the codecs from the start in the new background traffic generator node;
- (c) The integration in the background traffic generator node of these two algorithms (G.711 and G.729) with the rest of the encoding algorithms that are not part of OPNET natively.

It was found that, given the simplicity of the codecs and the nature of the OPNET native implementation, a simpler solution was to implement the algorithms from the start in the background traffic generator node. This decision ensures consistency with the newly implemented codecs that are not part of OPNET by default. Nonetheless, both implementations are completely compliant with the official standard for each codec [13, 16].

Table 3.1 Frame size and frame duration for fixed data rate codecs

Codec name	Frame size [bytes]	Frame duration [ms]
G.729A	10	10
G.711	80	10
iLBC 20ms	38	20
iLBC 30ms	50	30

The aforementioned encoding algorithms can be divided in to three groups considering the methodology used by the encoder to select the output data rate. First, we have the fixed data rate algorithms (G.711 [13], G.729 [16] and iLBC [28]). For these codecs, the output data rate is fixed and defined in the official standards or recommendations. For AMR-NB [18], the output data rate is adjusted according to network conditions. As a result, AMR-NB requires information regarding network status to operate in variable rate mode. Finally, Speex [37], uses exclusively the statistical properties of the input analog audio to adapt its output data rate. The

three groups present remarkable differences in the method used to compute the frame size. The details of such process will be described separately in following subsections.

3.1.1.1 Computation of frame size and frame duration for fix rate algorithms

In the case of the codecs with fixed output data rate, the frame size and frame duration are constant values and they are defined in the official standard for the particular codec. No computation is required to obtain the desire frame size and duration during the simulation; instead, a frame size/duration map is configured into the background traffic generator source code. Table 3.1 shows the value of the nominal frame sizes and durations for the fixed data rate codecs integrated into the background traffic generator.

3.1.1.2 Computation of frame size for AMR-NB algorithm

AMR-NB operates at eight bitrates and was specifically designed to improve link robustness. AMR supports dynamic adaptation to network conditions, using lower bitrates during network congestion or degradation while preserving audio quality. By trading off the speech bitrate to channel coding, AMR maximizes the likelihood of receiving the signal at the far end.

In the AMR standard, the mechanism through which the output data rate is adjusted is not defined [18, 69]; it is up to the modeller or protocol designer to define such a procedure. However, the most common method is to use the network delay or Mouth-to-Ear (M2E) delay as a condition to update the data rate [70, 71, 72, 73]. In voice communications carried over a packet switched network, the delay perceived from the time a sound is generated by the talker until the time this same sound is played out in the listener's end is a random variable and depends on multiple factors. This factor is commonly referred to as M2E delay. Equation 3.1 accounts for all factors involved in the M2E calculation. From equation 3.1, during the simulation the processing and packetizing delay cannot be measured; instead, only the End-to-End (E2E) network delay can be measured. Considerations will be made to estimate the necessary values. In most instances, worst case estimation will be used.

$$\begin{aligned}
 M2E = & \textit{packetization}_{delay} + \textit{compression}_{time/fr} + \textit{algorithmic}_{delay} + \\
 & + \textit{E2E network}_{delay} + \textit{dejitter}_{delay} + \textit{decompression}_{time}
 \end{aligned} \tag{3.1}$$

Where:

$packetization_{delay}$: Refers to the time taken to fill a packet payload with encoded/compressed speech. The packetization delay can be expressed by equation 3.2, where $frame_{duration}$ represents the nominal value of the speech frame duration (20 ms for AMR-NB) and $number_{fr/pkt}$ is the number of speech frames to be encapsulated per packet.

$$packetization_{delay} = frame_{duration} \times number_{fr/pkt} \quad (3.2)$$

As determined in the ETSI standard documentation for AMR [18], the duration of one voice frame is equal to 20 ms. Using equation 3.2, packetization delay can be redefined as equation 3.3.

$$packetization_{delay}[ms] = 20ms \times number_{fr/pkt} \quad (3.3)$$

$compression_{time/fr}$: Time taken by the encoder software to encode one speech frame. The estimation of this parameter is not trivial since it will depend on CPU speed for a specific codec. In some literature, this parameter may be referred to as coder or processing delay. According to research performed in [74], a safe value to assume for this parameter in the case of AMR codec is 5ms.

$algorithmic_{delay}$: Usually referred to as look-ahead delay. In most algorithms, in order to properly encode the current block N , part of the next block $N+1$ needs to be analyzed. This process adds an extra delay to the encoding of one block or frame. Look-ahead is implemented in most codecs to ensure a smooth decoding transition between adjacent encoded blocks and help the operation of packet loss concealment (PLC) procedures when applicable. The use of algorithmic delay term varies according to the author. In some literature, the compression time per voice frame and the look-ahead delay are added together and sometimes referred to as algorithmic delay. For the present document, the nomenclature used will be that adopted in [75] where algorithmic delay refers exclusively to the look-ahead delay.

As established in the ETSI standard [76], algorithmic delay is equals to 5 ms for all modes except for AMR12.20 where no look-ahead is implemented. However, when the codec is

operating in variable rate mode, for AMR12.20, a dummy 5ms delay is introduced to ensure inter-mode synchronism.

dejitter_{delay}: Corresponds to the size of the playout buffer in the receiving end. The function of this buffer is to minimize the effects of variable latency in the network providing a steadier stream of frames to the decoder. As it can be observed from equation 3.1, there exists a compromise between the size of the buffer (possibility of correcting jitter) and the total M2E. For the current experiments, the playout buffer size will be considered as 40 ms (duration of 2 voice frames).

decompression_{time}: The decompression time per frame is usually ten percent of that of compression [77]. However, this value has to be factored by the number of voice frames received per packet. Equation 3.4 shows such a relationship.

$$decompression_{time} = decompression_{time/fr} \times number_{fr/pkt} \quad (3.4)$$

Using the estimated value of *compression_{time/frame}* it can be concluded that, from equation 3.3, decompression time can be expressed as in equation 3.5.

$$decompression_{time} = 0.5 \text{ ms} \times number_{fr/pkt} \quad (3.5)$$

E2E network_{delay}: End-to-End network delay is the time that takes a packet of an arbitrary size to traverse the network from starting point to end point. The factors included in building up to the total delay are: serialization delay, queuing/buffering delay and switching delay. In the case of our work, the E2E network delay is obtained from the simulation. The individual analysis of the different factors that influence the E2E network delay lacks of relevance in our case beyond the point of acknowledging their existence and nature.

Table 3.2 Threshold values of M2E for AMR codec and the resulting frame sizes

M2E [ms]	AMR mode	Bitrate [kbps]	Fr size [bytes]
0 – 150.0	AMR_12.20	12.20	31
150.0 – 164.3	AMR_10.20	10.20	26
164.3 – 178.6	AMR_7.95	7.95	21
178.6 – 192.9	AMR_7.40	7.40	19
192.9 – 207.2	AMR_6.70	6.70	18
207.2 – 221.5	AMR_5.90	5.90	16
221.5 – 235.8	AMR_5.15	5.15	14
235.8 – 250.0	AMR_4.75	4.75	13

Based on the range of M2E that are considered acceptable to achieve an intelligible and dynamic conversation [78], Table 3.2 shows the decision thresholds for the AMR-NB frame size.

Using all calculations and estimations shown above in conjunction with equation 3.1, the M2E delay can be written as in equation 3.6, where M2E is only a function of the number of frames per packet and the E2E network delay.

$$M2E = 20.5 \text{ ms} \times \text{number}_{fr/pkt} + 50\text{ms} + ETE \text{ network}_{delay} \quad (3.6)$$

Considering that the number of frames per packet is a configuration parameter defined by the modeller, the problem of finding the M2E reduces to obtaining the E2E network delay for each packet or group of packets. The E2E network delay is the simulation metric used then to

determine the frame size for subsequent packets. A procedure has been implemented to obtain the E2E network delay that is consistent with the protocols and constrains used in real networks. Such a procedure will be explained next.

Obtaining the E2E network delay

Speech packets are transmitted using the Real-Time Transport Protocol (RTP). According to RFC3550 [46], where RTP technical specifications are discussed, RTP provides E2E network transport functions suitable for applications transmitting real-time data, such as audio and

video. RTP comes accompanied by a control protocol, also presented in [46] and [79]. The main function of RTP Control Protocol (RTCP) is to provide statistics and control information for an RTP flow. Five different RTCP packet types are defined in order to carry a variety of control information. These packet types are listed and explained below:

SR: Sender report, for transmission and reception statistics from participants that are active senders.

RR: Receiver report, for reception statistics from participants that are not active senders and in combination with SR for active senders reporting on more than 31 sources.

SDES: Source description items, including CNAME.

BYE: Indicates end of participation.

APP: Application-specific functions.

For the purpose of the thesis, only RR packets will be sent from the client node back to the background traffic generator node. The RTCP RR packets can carry information regarding time-stamp when a packet was generated (used to calculate round trip time in the sender), packet loss and jitter. Since earlier in this section it was defined that E2E delay would be the metric to be used as frame size control, only the value of the time stamp will be used in our case; however, the total packet size will be simulated. As stated in [46], for a single receiver model, the RTCP packet size is 32 bytes.

RTCP packets are useful in a scenario where the sender has some way to control the flow properties in order to improve the conditions of reception. This is the case of the background traffic generator when operating with the AMR-NB codec. The more frequent RTCP packets are sent by the client node, the more precise and efficient the bandwidth control carried out by the background traffic generator will be. However, the increase of the RTCP generation frequency also implies that the bandwidth used in the control plane of the stream increases as well. To ensure that the primary function of the transport protocol to carry data is not impaired, it is suggested in [46] that the bandwidth to be employed for the control protocol is limited to 5% of the total bandwidth used by the RTP flow. This limit should be observed by the modeller; nevertheless, no boundaries have been added to the implementation of the client node in order to allow freedom to the modeller even beyond the recommended parameters.

3.1.1.3 Computation of frame size for Speex algorithm

Analyzing real-world random phenomena using exclusively first-order statistic, is in many cases a valid approach; in other instances, it is a common oversimplification mistake. For variable data rate audio and video algorithms, second-order statistic dependence is to be expected [80]. *Undermodeling* is the term used in [81] referring to the case where some of dependencies that are relevant to the system in consideration are overlooked. Temporal dependence, when present, needs to be closely observed by the modeller; failure to do so will render the model unrealistic and results misleading.

In order to generate background traffic that follows the statistical behaviour of a real Speex encoded audio, two parameters need to be computed: frame size, frame interarrival time. Frame interarrival time is a constant value equals to 20ms in the particular case of Speex [37]. Frame size is a random variable that, as explained before, is to be expected to present first-order and second-order statistic dependencies.

To investigate the statistical behaviour of the frame size random variable, 30,000 Speex frames were analyzed for each of the Speex quality modes analyzed. The thesis covers only American English language and a mixture of male and female speech sample was chosen to minimize the influence of gender. Audio samples were obtained from the Open Speech repository located in [54].

The probability density function for the Speex frame size (Quality = 2) random variable is depicted in Figure 3.1.

Autocorrelation of the frame size random variable was calculated up to a maximum lag of 10 samples. Figure 3.2 shows the autocorrelation function for the frame size random variable (Quality = 2)². As inferred, a strong autocorrelation is observed. Also, the fact that the function is monotonically decreasing as the sample lag increases corresponds with the principles of the Speex algorithm, where the properties of the analog audio signal determine the compression rate of the encoded stream.

² It was confirmed by performing identical computations that the nature of the frame size for Speex Quality = 8 is similar to that of Quality = 2; i.e., to present an evident temporal dependence.

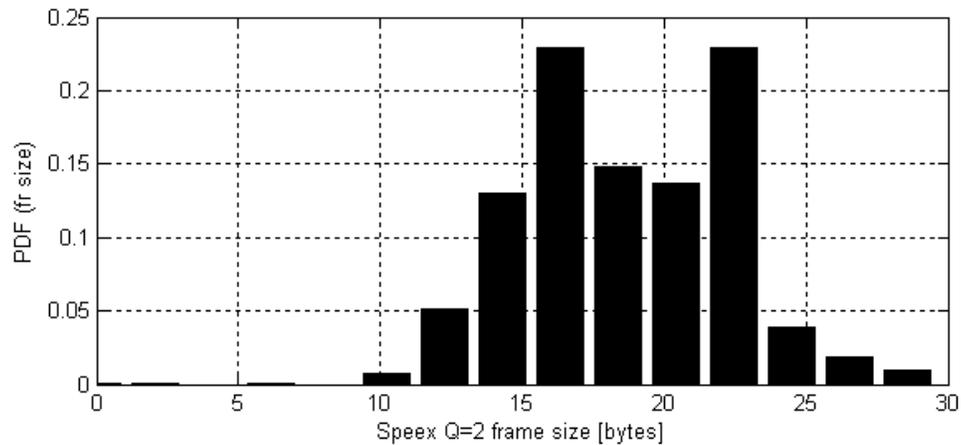


Figure 3.1 Probability Density Function for the random variable Speex frame size [bytes]

From the results derived in Figure 3.1 and Figure 3.2, it becomes obvious that both first-order and second-order statistic need to be considered in the generation of the random variable representing the frame size.

It is a fair generalization to say that from the Probability Density Function the Autocorrelation Function cannot be derived, and the Autocorrelation Function gives no information about the Probability Density Function. Consequently, a random number generation method needs to be utilized that at the same time contemplates first-order as well temporal dependence.

Few methodologies were found in the literature that complies with the aforesaid requirements and fewer that were of relatively simple implementation. In [82], the desired time series is obtained by shuffling the series to minimize a sum of squares criterion between desired and actual autocorrelation functions. In [83] and [84] the wavelets principle is used instead. However, in [81], the author describes the Transform-Expand-Sample (TES) algorithm. The use of TES has rendered useful results in numerous studies [85, 86] and its implementation is uncomplicated and well documented. A comprehensive implementation of the TES algorithm is offered in [87]. Ultimately, a computer software, called TESstool, that is intuitive yet robust [88], facilitates the calculation of the parameters used by TES model to generate data.

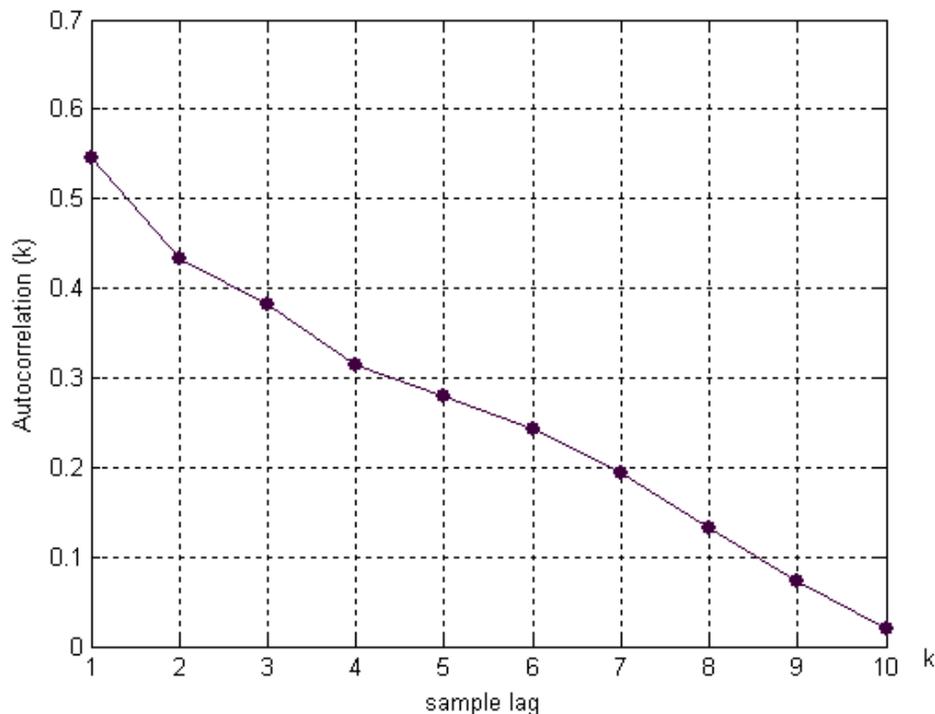


Figure 3.2 Autocorrelation function of the Speex frame size random variable

TES is an approach to modeling stationary time series capable of capturing both the marginal distribution and the autocorrelation function of empirical data. Most importantly, TES aims to fit both marginal and autocorrelation functions simultaneously. The derivation of TES models is performed in two stages. The first phase consists of generating a correlated sequence with uniform marginal distribution $[0,1]$. This is achieved as shown in equation 3.7 where V_n represents a sequence of iid random variables independent of U_0 . V_n is called the innovation sequence [81, 83] and $\langle \rangle$ denotes the modulo-1 operation. The amount of correlation depends on the structure of the density function of the innovation sequence.

$$U_n^+ = \begin{cases} U_0 & \text{for } n=0 \\ \langle U_{n-1}^+ + V_n \rangle & \text{for } n>0 \end{cases} \quad (3.7)$$

In the second stage, a histogram of the original sample is constructed and an inversion technique like that shown in [81] is applied. This inversion technique allows the transformation of any uniform random variable to another variable with a specific distribution.

As described in [87] and [88], TESTool takes the stationary time series to be studied and outputs the Probability Density Function of that time series and an Innovation Function which is related to the computation of the Autocorrelation Function. After deriving the Marginal Distribution and Autocorrelation Function from the empirical data, a heuristic search is performed for fitting the TES model generated data to empirical data. TESTool provides a visual interactive interface that allows the process of finding a good-fitting Innovation Function to flow in an intuitive manner.

Figure 3.3 shows a screen captured from TESTool showing the results of finding the Innovation Function and the validation of the implementation of the TES algorithm in OPNET. Captured screen is only shown for Speex Quality = 2 mode. However, equivalent results were obtained for Quality = 8 mode.

Using the Innovation function graphically represented in Window 4 of Figure 3.3 and the Marginal Distribution shown in Window 3 of the same figure, the implementation of the TES algorithm described in [87] was incorporated into OPNET model for the background traffic generator node.

3.1.2 Analysis of the number of speech frames per packet attribute

The effective bandwidth of a voice stream on a packet switched network is not exclusively defined by the data rate of the speech encoder. The size of each speech packet that traverses the network will have two components: (1) the voice payload and (2) headers corresponding with different protocols associated with various network layers. This last component of the packet structure is usually referred to as overhead. The overhead is a fixed number of bytes per packet and its compression is a difficult task [89].

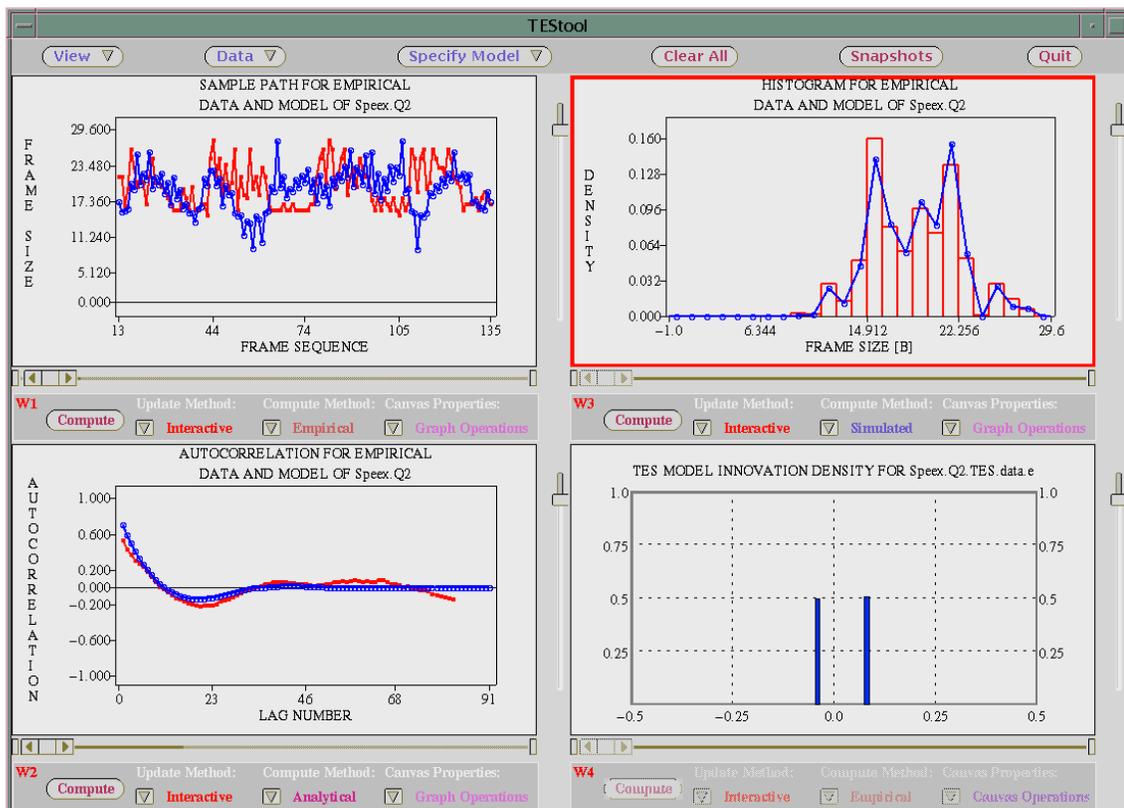


Figure 3.3 Obtaining the Innovation Function in TESstool for Speex Q2 frame size. Window 1 (top-left): Sample paths of empirical data and modeled data based on the current Innovation Function. Window 3 (top-right): PDF of empirical data and modeled data. Window 2 (bottom-left): Autocorrelation Function for empirical data and modeled data. Window 4 (bottom-right): Innovation Function.

In VoIP communications, the most common way to mitigate the effect of the overhead in the bandwidth consumption is to increase the *payload / overhead* ratio. This can be achieved by encapsulating more than one speech frame per packet. Equation 3.8 shows a generic way of calculating bandwidth utilization for any data stream where, BW represents the bandwidth required, $number_{pkt/sec}$ is the packets per second rate and pkt_{size} is the size of each packet expressed in bits.

$$BW = number_{pkt/sec} \times pkt_{size} \quad [bps] \quad (3.8)$$

Both terms on the right side of equation 3.8 can be expressed as a function of the codec parameters and overhead size. Based on this, equation 3.8 can be redefined as:

$$BW (number_{fr/pkt}) = \frac{codec_{fr\ rate}}{number_{fr/pkt}} \times (OH + (codec_{fr\ size} \times number_{fr/pkt})) \quad [bps] \quad (3.9)$$

In equation 3.9, $codec_{fr\ rate}$ and $codec_{frame\ size}$ are the nominal frames per second rate (inverse of the frame duration) and the frame size in bits for a particular codec, respectively. OH represents the fixed size of the overhead for each packet expressed in bits and $number_{fr/pkt}$ is a design parameter that represents the number of speech frames that are encapsulated in each packet.

Figure 3.4 illustrates the effects of increasing the number of speech frames per packet. Figure 3.4 (top) shows how a desired decrease of the bandwidth is observed as the number of frames per packet is increased. However, as depicted in Figure 3.4 (bottom) by encapsulating more than one frame per packet, an additional effect occurs. An increase in the M2E delay can be observed with the increase of the number of frames per packet. If considering the generic case when n frames are encapsulated per packet; in the encoder, a delay is introduced while waiting for the n^{th} frame to be produced before the packet can be sent to the lower layers and eventually transmitted through the network. Similarly, in the decoder, the decompression time of the packet is a function of the number of frames per packet. Figure 3.4 (bottom) shows a linear increase in the mouth-to-ear delay as consequence of the increase in the number of frames per packet. For the calculation of the increase of the M2E delay, only the factors of equation 3.1 that depend on the number of frames per packet have been considered. Additionally, other parameters that strongly influence the quality of speech, like packet loss robustness, may also be a function of the number of speech frames per packet [90].

For the constructions of the graphs in Figure 3.4, the following assumptions were taken: encoding algorithm G.729 ($codec_{frame\ size} = 80$ bits, $codec_{pkt\ rate} = 100$ pkt/sec), traffic on Ethernet link, OH calculated including eth v.2 header, IP header, UDP header and RTP header (OH = 58 bytes).

Based on the versatility of the results that can be achieved by modifying the number of frames per packet, this parameter has been included as a configurable attribute in the background traffic generator node for each of the encoder algorithms implemented.

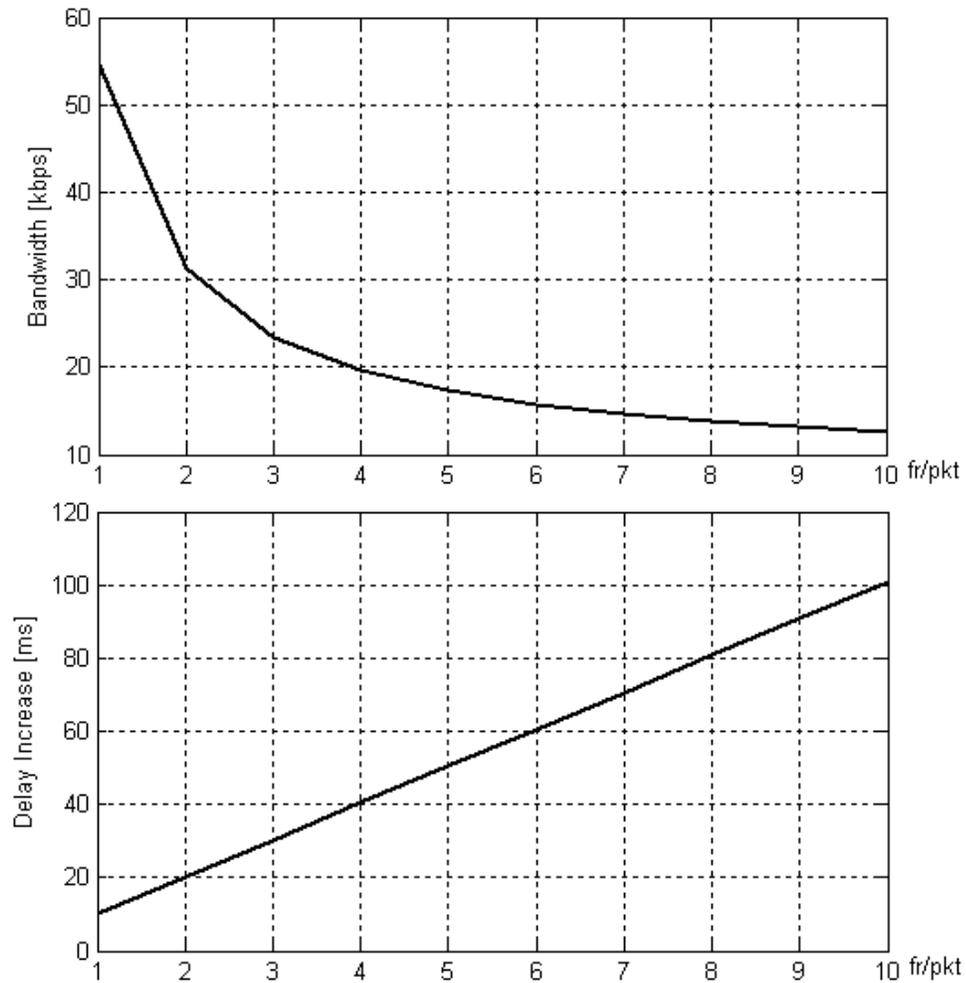


Figure 3.4 Effects of the increase of the number of frames encapsulated per packet in the bandwidth (top) and the ear-to-mouth delay (bottom)

3.1.3 Computation of the voice payload size

In Section 3.1.1, the methods to compute the frame size have been discussed for each of the codecs. This process occurs in the application layer of the node. In the simulation sequence, the next step is to construct the voice payload in order to send the packet to the lower layers.

The application layer packet has two fields; the first one corresponds with the header of the application protocol to be used, RTP in this case, and the second field corresponds with the

encoded voice data itself, the voice payload. Figure 3.5 depicts the structure of the application layer packet carrying n speech frames.

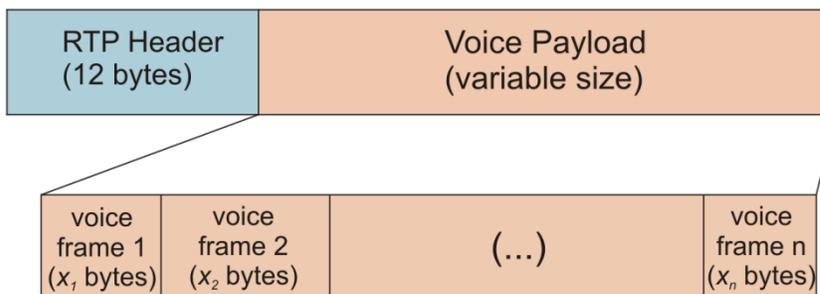


Figure 3.5 Structure of the packet produced in the application layer of the background traffic generator node

Field 1: RTP header

As specified in [46], the RTP header is 12 bytes long. Bit alignment and field description is irrelevant to the background traffic generator, only to model the size of the header is required. Detailed information can be found in Section 4 of [46].

Field 2: Encoded speech data

The encoded speech data field is formed by one or more speech frames according to the value of the attribute “Number of frames per packet” specified by the modeller in the background traffic generator node. In Figure 3.5, the generic case of n voice frames is shown. More formally, the size of the encoded speech data field can be defined as shown in equation 3.10 for each of the encoding algorithm groups.

$$Voice\ payload_{size} = \begin{cases} number_{fr/pkt} \times codec_{fr\ size} & \text{For fix data rate} \\ number_{fr/pkt} \times AMR_current_mode_{fr\ size} & \text{For AMR} \\ \sum_{i=1}^n fr_size_i & \text{For Speex} \end{cases} \quad (3.10)$$

Where $number_{fr/pkt}$ represents the value of the “Number of frames per packet” attribute, $codec_{fr\ size}$ is the nominal frame size for each fixed data rate algorithm; i.e., 80 bytes for G.711, 10 bytes for G.729, 38 bytes for iLBC20ms and 50 bytes for iLBC30ms. $AMR_current_mode_{fr\ size}$ represents the frame size for the AMR mode that was active before

the packet was produced, see Table 3.2, and $f r_{size_i}$ is the size of one Speex frame calculated on frame by frame basis, see Section 3.1.1.

3.1.4 Background traffic scalability method

The methodology proposed to achieve speech background traffic scalability has been designed to reflect the nature of VoIP traffic. One of the simplest and effective ways to define speech traffic volume is by the number of conversations. Therefore, the attribute used to create scalable traffic is “number of conversations”. More formally, it can be defined as number of calls per codec per destination.

To better illustrate, let us take a simple situation where two branch offices *A* and *B* are located in different cities and interconnected somehow. It is of interest to determine the reason for the poor subjective quality of the conversations between the two offices. In order to properly model the problem, the analyst will scrutinize the network statistics for a number of days. Information such as the average number of calls between the two branches will be collected. Obviously, as for any modeling problem, the more detailed the information describing the real-world system, the more reliable the results of the simulation will be. A practical example could be that the simulation modeller knows that between office *A* and *B* occur an average of twelve G.729 calls for a certain period of the day. These values are fed to a background traffic generator node collocated in *A* (and one in *B* if a conversational speech assessment wants to be obtained) in order to generate a realistic traffic background for the simulation. As mentioned, OPNET will natively provide support for other traffic models that could exist between branches *A* and *B*, e.g., HTTP, FTP, email.

Finally, for the case of Speex, it is valid to clarify that if more than one conversation is specified to a particular destination, the frame size random variables generated according to TES model are independent among conversations, as it is in a real life scenario.

3.1.5 General operation of the simulation of speech background traffic generation in OPNET

Figure 3.6 depicts how the background traffic generator node is integrated into a simulation as well as the most relevant inputs. The background traffic generator node will receive three types of input to operate, they are: (1) attributes defined by the modeller in the simulation setup stage, (2) attributes embedded in the simulation and (3) feedback parameters obtained from the simulation. The shape of the traffic generated in this node as well as the destinations where such traffic will be sent are to be defined by these three set of inputs. Appendix A presents a detailed description of the background traffic generation node implemented in OPNET.

In Figure 3.7, a flow diagram describes the different stages involved in generating speech background traffic for the three groups of encoding algorithms. More detailed information regarding the process of calculating the frame size for each group will be discussed below.

Speech background traffic generation for fixed data rate codecs

1. The first step of the simulation is the definition of the simulation attributes by the modeller. The following parameters are defined at this stage:
 - 1.1 In the background traffic generator node, one of the following codec/modes is selected: G.729, G.711, iLBC-20ms or iLBC-30ms. The selection of the codec will univocally define the encoded frame size and frame duration (see Table 3.1).
 - 1.2 In the background traffic generator node, the IP address of the destination node and the number of speech frames encapsulated per packet are defined (see Section 3.1.2 and Appendix A for more details).

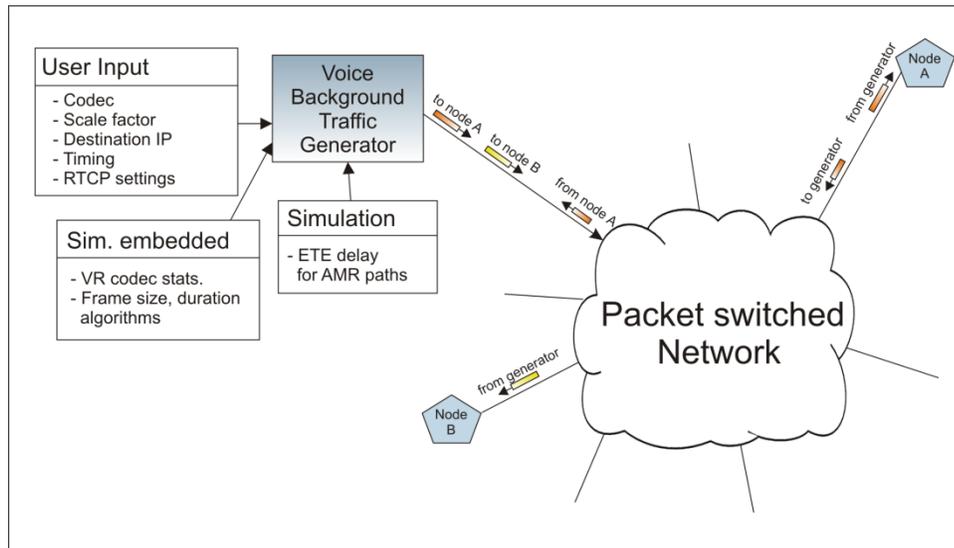


Figure 3.6 Operation of Voice background traffic generator

- 1.3 In the background traffic generator node, the number of conversations per codec per destination is entered. Along with some timing settings, the number of conversations attribute defines the traffic scalability for each codec-destination.
2. Using the codec type attribute defined in the simulation setup stage and information from Table 3.1, the frame size and frame duration for the selected codec are obtained.
3. Using the number of conversations attributes, separate independent streams are created, each one representing a conversation.
4. For each conversation, using the number of voice frames per packet attribute and based on equation 3.10, the voice payload size and application layer packet size are calculated.
5. The application layer packet is sent down to all other layers and finally transmitted through the simulated network. Each packet is potentially exposed to latency, jitter and packet loss creating the simulation speech background traffic.

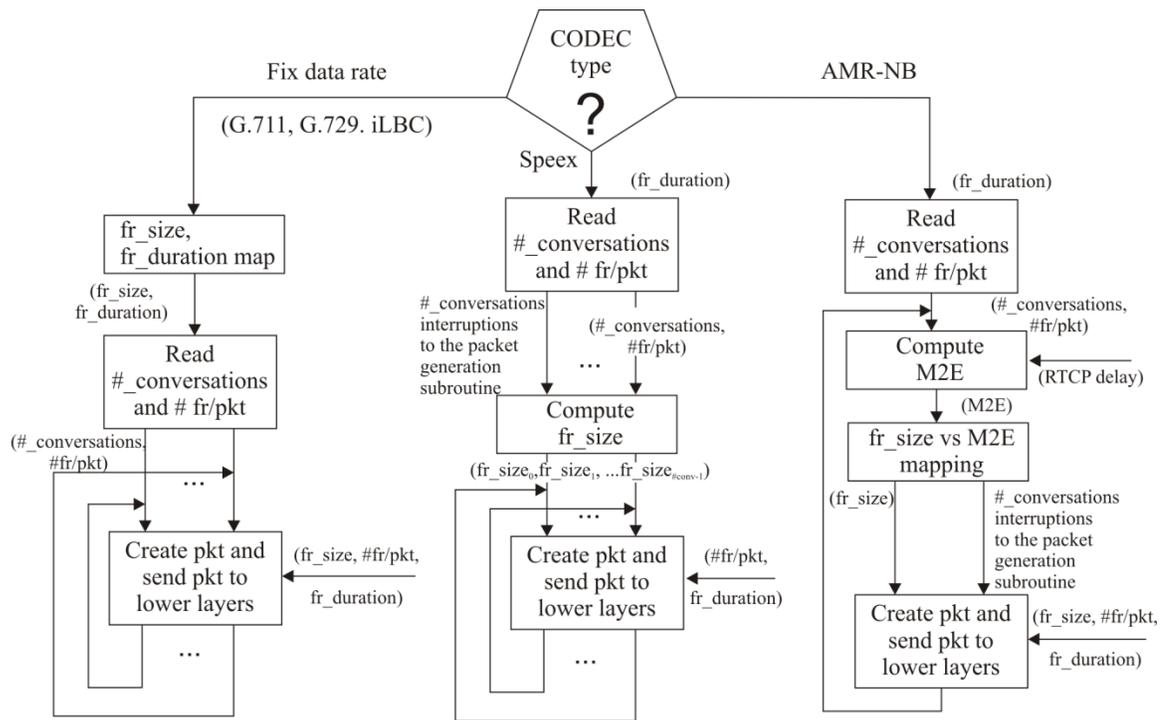


Figure 3.7 Traffic generation chart for each of the encoding algorithm groups encompassing all inputs and attributes

Speech background traffic generation for AMR-NB codec

1. The first step of the simulation is the definition of the simulation attributes by the modeller. The following parameters are defined at this stage:
 - 1.1 In the background traffic generator node, AMR-NB codec is selected. The selection of the codec will define the frame duration of 20ms.
 - 1.2 In the background traffic generator node, the IP address of the destination node and the number of speech frames encapsulated per packet are defined (see Section 3.1.2 and Appendix A for more details).
 - 1.3 In the background traffic generator node, the number of conversations per codec per destination is entered. Along with some timing settings, the number of conversations attribute, defines the traffic scalability for each codec-destination.
 - 1.4 In the client node, the RTCP generation frequency is defined in terms of number of RTP packets received before one RTCP packet is sent back to the background traffic generator node (see Section 3.1.2 and Appendix A for more details).

- 1.5 In the client node, it is defined if the delay information to be sent back to the background traffic generator node is that measured from the last received RTP packet or the average of the last group of RTP packets received since the previous RTCP packet was sent.
2. Initially, since no information regarding the delay associated with the network path has been obtained, AMR-NB will start transmitting in the highest mode (12.2 kbps).
3. The client node, after receiving the number of AMR-NB frames defined in step 1.5, will generate a control packet that will be sent back to the real voice traffic generator node. The RTCP packet will offer the real voice traffic generator node updated information regarding the delay on the path connecting the two nodes.
4. At the reception of the RTCP packet from the Client node, the background traffic generator node updates the M2E parameter based on the E2E delay obtained from the RTCP packet and using equation 3.5.
5. Using Table 3.2 and the M2E value obtained in step 4, the frame size for the subsequent frames is computed. At the reception of the next RTCP packet the process of adjusting the frame size will be repeated. During the interval between RTCP packets reception, the frame size is maintained Using the number of voice frames per packet attribute and based on equation 3.10, the voice payload size and application layer packet size are calculated.
6. The application layer packet is sent down to all other layers and finally transmitted through the simulated network. Each packet is potentially exposed to latency, jitter and packet loss creating the simulation speech background traffic.

Speech background traffic generation for Speex codec

1. The first step of the simulation is the definition of the simulation attributes by the modeller. The following parameters are defined at this stage:
 - 1.1 In the background traffic generator node, Speex codec is selected (one of the two modes available). The selection of the codec will define the frame duration of 20ms.

- 1.2 In the background traffic generator node, the IP address of the destination node and the number of speech frames encapsulated per packet are defined (see Section 3.1.2 and Appendix A for more details).
- 1.3 In the background traffic generator node, the number of conversations per codec per destination is entered. Along with some timing settings, the number of conversations attribute, defines the traffic scalability for each codec-destination.
2. According to the parameter entered in step 1.4, a number of traffic streams are generated.
3. For each conversation independently, a number of frame sizes equal to the parameter defined in step 1.3 are calculated using the methodology discussed in 3.1.1. Within each conversation, the frame size variable maintains its autocorrelation even when frames are encapsulated in different packets. Analogously, the frame size variable is independent between conversations.
4. Using the number of frames per packet attribute defined in the simulation setup stage and equation 3.10, the voice payload size and the application layer packet size are obtained.
6. The application layer packet is sent down to all other layers and finally transmitted through the simulated network. Each packet is potentially exposed to latency, jitter and packet loss creating the simulation speech background traffic.

3.2 Simulation of real speech traffic in OPNET

From the perspective of simulation theory and principles [91], in a typical simulation environment for existing systems the use of real data may play an important role on the verification process of a model. Once the model is built to depict a real-world system, actual data can be fed to the model and the output compared to the output of the real system to the same input. The results of the comparison can be used as information describing the correctness of the assumptions, parameters and general structure of the model.

For systems in design stage, the possibility of injecting real data to the simulation is also appealing, adding a new dimension to the simulation results. Particularly interesting is the case when, like audio or video, the effects of network impairments over a voice stream can be better appreciated in a subjective mode. The ways to objectively quantify speech quality have been

numerous [22, 60, 63, 64] and still remains an area of research; it is difficult to determine in which magnitude certain degradation of the data stream will affect the way humans perceive the speech quality. Note, from Section 2.2.1, that every speech quality assessment algorithm uses as a reference a subjective assessment model, usually ITU-T P.800 [53]. The *what-would-it-sound-like* feature added to OPNET considerably improves the power of speech simulation in this platform.

3.2.1 Encoding algorithms used in real speech simulation and licensing details

In order to find the set of algorithms to be integrated into the real voice traffic generator node, a study was performed to assess the popularity of each codec and the availability of source code with the appropriate licensing that allows modification and therefore the addition to our OPNET model. The results lead us to the implementation of the codecs mentioned below.

- G.729
- G.711 A-law (Packet loss concealment (PLC) and no-PLC)
- G.711 μ -law (PLC and no-PLC)
- iLBC (20 ms and 30 ms)
- AMR-NB

Table 3.3 Properties of the encoding algorithms used in the real voice generator node

Codec	Sampling rate [Hz]	Frame duration [ms]	Input fr. [B]	Output fr. [B]	PLC	Output data rate [kpps]
G.729	8,000	10	160	10	Yes	8
G.711 (PLC)	8,000	10	160	80	Yes	64
G.711 (no-PLC)	8,000	10	160	80	No	64
iLBC-20 ms	8,000	20	320	38	Yes	15.20
iLBC-30 ms	8,000	30	480	50	Yes	13.33
AMR-NB	8,000	20	320	variable	No	4.75-12.20

Table 3.3 refers to the main technical specifications of the encoder-decoder applications used in the real speech generation module (see Section 3.2.5 for more details). Licensing details for all codec applications used can be found in Section 2.1.4.

Note that Speex algorithm is not included in the set of codecs implemented in the real speech traffic generator model. Speex controls the data rate based on the statistical properties of the original analog audio file (see Section 2.1.3 for more details). In order to integrate Speex to the real speech traffic generator model, the input audio file needs to be encoded before the simulation starts and the frame size for each frame extracted from the encoding process. This procedure implies that modifications need to be applied to the Speex encoder source code to ensure that an ASCII file containing frame size for each speech frame is generated along with the encoded audio file. However Speex is a free and open source codec, the source code for the implementations available do not compile successfully in any of the compilers attempted. The inclusion of Speex algorithm to the real speech traffic generator model will be suggested in Section 5.3 as future work to this thesis.

3.2.2 Playout buffer (De-jitter buffer)

The E2E delay of packet transmission is of paramount importance in the implementation of a voice network. Adequate QoS must be maintained. Authors in [92] define jitter as “a measure of time between when a packet is expected to arrive to when it actually does arrive”. The main cause of jitter in packet switched network is determined by the queuing variations and caused by the randomness of network traffic. In addition, the nature of a packet switched network implies that packets of the same stream can traverse the network using different paths, the less cost paths according to the routing algorithms used and the present network loads. The fact that packets arrive out of sequence is a common problem that nodes involved in real-time traffic have to face in packet switched networks.

The most common approach to removing jitter is a playout buffer, sometimes called jitter or de-jitter buffer. The playout buffer is implemented in the node where the media will be played and sometimes in intermediate media gateways. It could be hardware or software (embedded in the decoder) implemented. The principle of de-jitter using a buffer consists in holding the packets long enough as to permit the slowest packet to arrive in time. It can be understood as adding a tolerance to the expected arrival time, solving at the same time the problem of out of sequence packets.

Reports [93, 94, 95] can be found on the subject of playout buffer size optimization. Defining the optimal buffer size deals with the conflicting goals of minimizing overall delay and removing jitter. The bigger the playout buffer the larger the jitter values that can be compensated but this occurs at the cost of increasing the total E2E delay with a direct impact in the conversational interactivity. In the reviewed literature, the universally accepted way to approach buffer size optimization seems to be focused in dynamic playout buffer size. Some of these optimizations algorithms have yield better results depending of the jitter level, i.e., some will work better in low jitter conditions and some others perform well for severe jitter environments.

For this thesis, only constant buffer size is available in the client node and it is defined as a user input parameter. Also, the buffer size is defined as the duration in milliseconds that the audio frames are delayed before they are accepted as received. Depending on the codec, more specifically on the nominal frame duration of each codec, a conversion is made from time in milliseconds to number of frames to be compensated. For instance, if the playout buffer defined is 100 milliseconds, for G.729, 10 frames will be stored in the buffer before they are played out; for AMR-NB, only five frames are compensated and for iLBC-30ms, only three frames are stored before playing them out (see Appendix A for more details) .

3.2.3 General operation of the simulation of real speech traffic generation in OPNET

The present section will discuss the procedure followed in order to integrate real voice to an OPNET simulation. Appendix A presents a detailed description of the real speech traffic generation node implemented in OPNET.

For simplicity of the thesis, it will be assumed as irrelevant to the research the manipulations with headerless (RAW) audio files; e.g., obtaining RAW audio files, playing back a RAW audio file, etc.

As for the background traffic generator, a very important component of the simulation of the real voice traffic generator lies on the procedures to find the frame size and the frame duration

for each codec simulated. Two different methodologies are used to perform such a task. Their difference is associated to the nature of the output data rate.

Real speech traffic simulation for fixed data rate codecs

The simulation sequence and principles for the integration of voice data encoded using one of the fixed data rate algorithms integrated to the real voice traffic generator node will be explained step by step in the current subsection. Interaction among the integral modules is clearly depicted in Figure 3.8.

1. The first step of the simulation is the definition of the simulation attributes by the modeller. The following parameters are defined at this stage:
 - 1.1 It is assumed that one of the following codec/modes is selected: G.729, G.711 A-law (PLC or no-PLC), G.711 μ -law (PLC or no-PLC), iLBC 20ms or iLBC 30ms. The selection of the codec will univocally define the input frame size, encoded frame size and frame duration (see Table 3.3).
 - 1.2 Also, in the pre-simulation settings, the size of the playout buffer in the client node will be defined (see Section 3.2.2).
 - 1.3 The name of the real audio file is specified.
 - 1.4 The IP address of the destination node is entered.
2. Using information from Table 3.3, which matches codec type with input frame size, the number of audio frames to be simulated is obtained from the original RAW audio file (see Appendix A for more detail).
3. As many frames as obtained in step 2 and with size equivalent to the encoded frame size as specified in Table 3.3 are sent through the simulated network to the destination IP address specified in step 1.3. Each packet is potentially exposed to latency, jitter and packet loss. Note that real data contained in the audio file is not used in the simulation, only the size of the frames is simulated and dummy frames are transmitted.

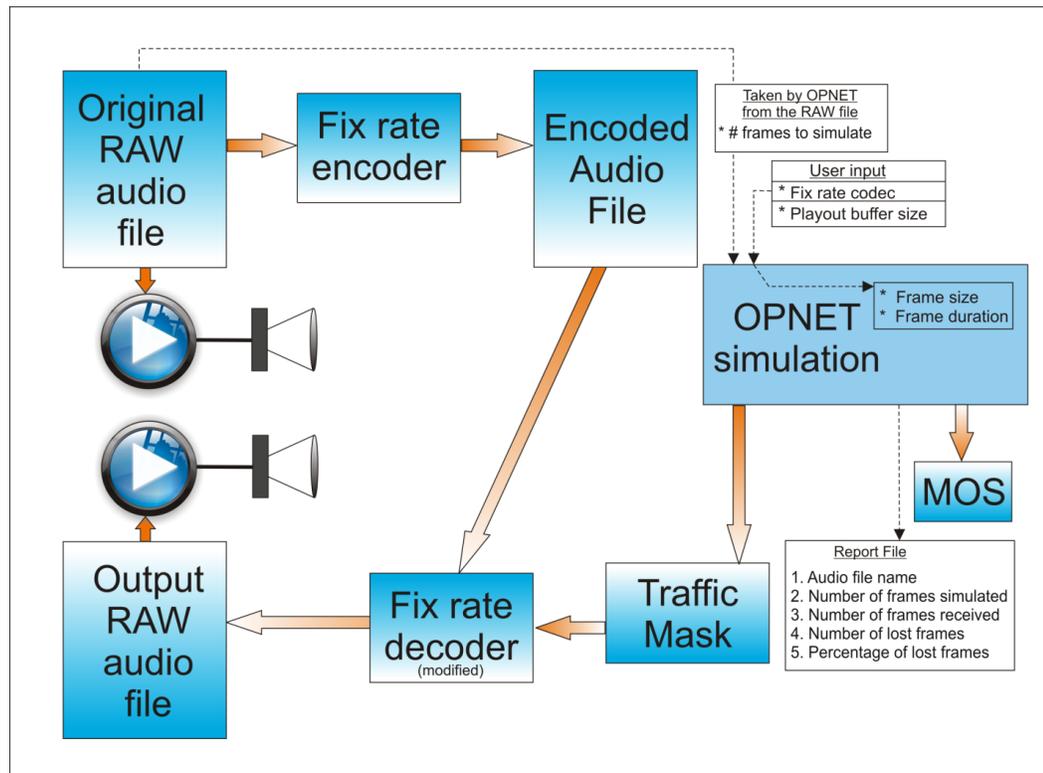


Figure 3.8 Block diagram of the real voice simulation in OPNET for fixed data rate codecs

4. Depending on the simulation configuration all or some frames will arrive to the client node. Furthermore, for the client node, not all the frames that arrive are considered *received*. For a real-time stream, the latency of the packets is critical; if a packet arrives at a later time than the instant when the decoder is supposed to process it, the frame or frames contained in the packet will be dropped. The client node will define if a frame that arrived is considered *received* based on the size of the playout buffer specified in step 1.2, the frame duration and the arrival time of the first frame (see Appendix A for more detail). The client node generates an ASCII file that contains the precise information of whether each frame was *received* or not. This file will be referred to from now on as traffic mask.
5. Also in the client node, calculations of MOS are performed as the frames are received. E2E delay is calculated from the timestamps in the frames and packet loss rate is updated as the simulation progresses (see Appendix A for more information).
6. At the end of the simulation a report ASCII file is generated containing the following information: selected codec, number of frames transmitted by the real voice traffic

generator, number of frames *received* by the client node and total frame loss rate. Also a batch file is generated to help automate steps 7- 9. Furthermore, objective quality assessment (see step 5) can be accessed through the **Discrete Event Simulation** menu in OPNET Modeler (see Appendix A for more information).

7. After simulation ends and using the batch file generated in step 6 the original audio file is encoded.
8. Then, the decoder algorithm, taking as input the encoded audio file and the traffic mask file, decodes the encoded file accordingly. Only the *received* frames are decoded and packet loss concealment is applied whenever available or applicable. The decoded file will possibly exhibit some quality degradation compensated somehow by the playout buffer.
9. Assisted by the batch file created in step 6, the RAW audio file obtained in step 8 is played out allowing the modeller to subjectively perceive the effects of the simulation over a real voice stream. Alternatively, the original audio file can also be played out with the purpose of highlighting the distortion introduced by the simulation.

Real speech traffic simulation for AMR-NB codec (variable data rate)

The simulation sequence and principles for the integration of voice data encoded using one of the AMR-NB algorithms will be explained step by step in the current subsection. Interaction among the integral modules is clearly depicted in Figure 3.9.

1. The first step of the simulation is the definition of the simulation attributes by the modeller. The following parameters are defined at this stage:

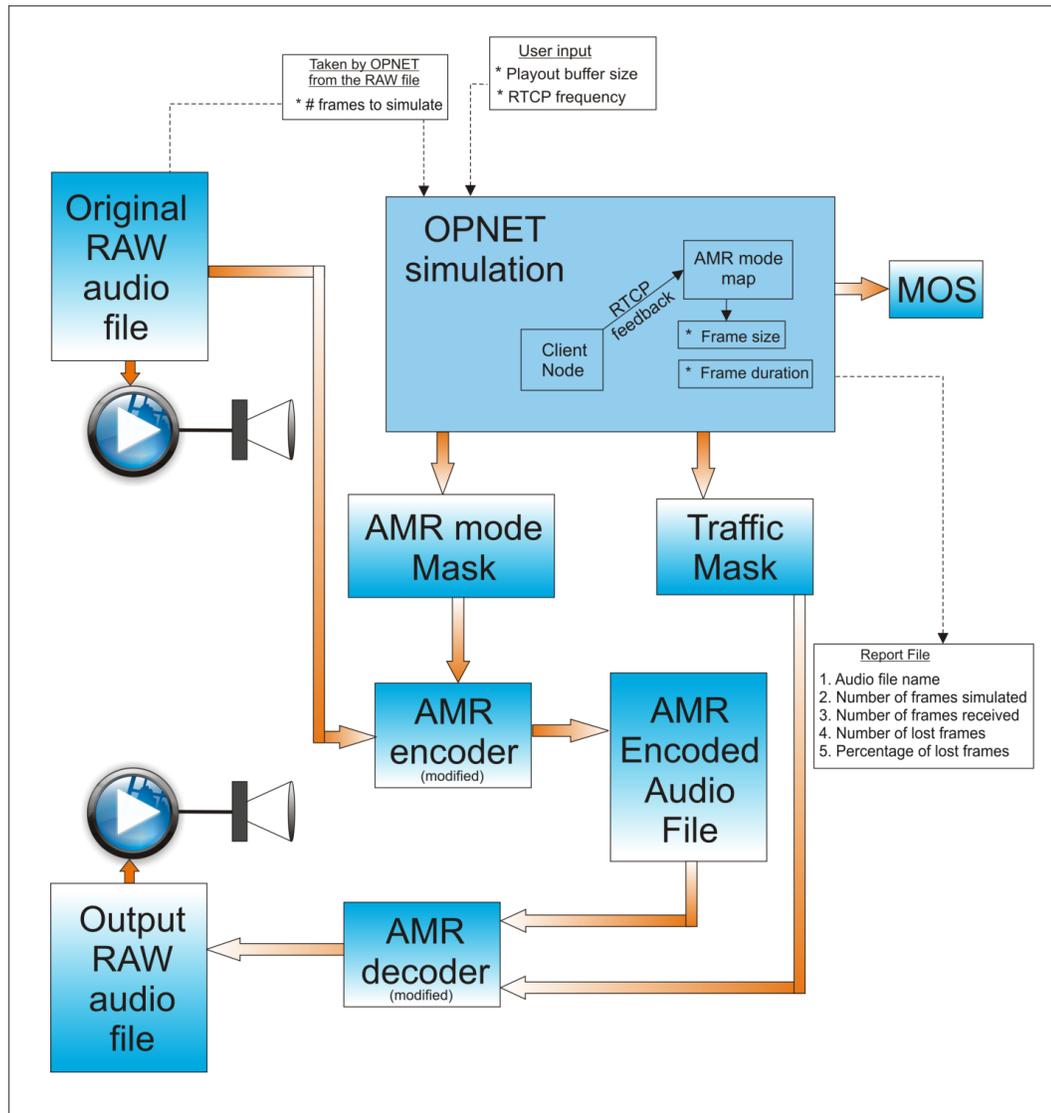


Figure 3.9 Block diagram of the real voice simulation in OPNET for AMR-NB codec

- 1.1 It is assumed that AMR-NB codec is selected. The selection of this codec will define the input frame size and frame duration as 320 bytes and 20 ms respectively.
- 1.2 Also, in the pre-simulation settings, the size of the playout buffer in the client node will be defined (see Section 3.2.2).
- 1.3 The name of the real audio file is specified.
- 1.4 The IP address of the destination node is entered.
- 1.5 In the client node, the frequency that will define how often RTCP packets are sent back to the real voice traffic generator is defined. Besides, whether each control

packet will carry information associated to the delay of the last received packet or the average of the last group of packets is specified.

2. Using information from Table 3.3 for AMR-NB, the number of audio frames to be simulated is obtained from the original RAW audio file (see Appendix A for more detail).
3. In Section 3.1.1 the method to obtain the frame size for AMR-NB codec was explained for the background traffic generator. The exact same mechanism is used in the real voice generator node. The client node, after receiving the number of AMR-NB frames defined in step 1.5, will generate a control packet that will be sent back to the real voice traffic generator node. The RTCP packet will offer the real voice traffic generator node updated information regarding the delay on the path connecting the two nodes. Based on the value of this delay, and using Table 3.2, the frame size for subsequent frames will be computed. At the reception of the next RTCP packet this process will repeat. As the simulation progresses the AMR mode (and consequently the frame size) used in each frame is stored in an ASCII file. This file will be referred as the AMR mode mask.
4. As many frames as obtained in step 2 and with variable sizes according to the AMR mode computed in step 3 are sent through the simulated network to the destination IP address specified in step 1.3. Each packet is potentially exposed to latency, jitter and packet loss. Note that real data contained in the audio file is not used in the simulation, only the size of the frames is simulated and dummy frames are transmitted.
5. Depending on the simulation configuration all or some frames will arrive to the client node. Furthermore, for the client node, not all the frames that arrive are considered *received*. For real time stream, the latency of the packets is critical; if a packet arrives at a later time than the instant when the decoder is supposed to process it, the frame or frames contained in the packet will be dropped. The client node will define if a frame that arrived is considered *received* based on the size of the playout buffer specified in step 1.2, the frame duration and the arrival time of the first frame (see Appendix A for more detail). The client node generates an ASCII file that contains the precise information of whether each frame was *received* or not. This file will be referred to as traffic mask.

6. Also in the client node, calculations of MOS are performed as the frames are received. E2E delay is calculated from the timestamps in the frames and packet loss rate is updated as the simulation progresses (see Appendix A for more information).
7. At the end of the simulation a report ASCII file is generated containing the following information: selected codec, number of frames transmitted by the real voice traffic generator, number of frames *received* by the client node and total frame loss rate. Also a batch file is generated to help automate steps 8 - 10. Furthermore, objective quality assessment (see step 6) can be accessed through the **Discrete Event Simulation** menu in OPNET Modeler.
8. After simulation ends, the original RAW audio file is encoded using the AMR mode mask file generated in step 3 as an input for the encoder. The AMR mode mask file provides the encoder exact information of the frame size used to encode each frame.
9. Then, the AMR decoder, taking as input the encoded audio file and the traffic mask file, decodes the encoded file accordingly. Only the *received* frames are decoded, PLC is not implemented in the VoiceAge open initiative implementation of AMR-MB. The decoded file will possibly exhibit some quality degradation compensated somehow by the playout buffer.
10. Assisted by the batch file created in step 7, the RAW audio file obtained in step 9 is played out allowing the modeller to subjectively perceive the effects of the simulation over a real voice stream. Alternatively, the original audio file can also be played out with the purpose of highlighting the distortion added by the simulation.

3.2.4 Objective speech quality assessment

The current section will discuss in detail the objective quality assessment selected, the E-model [22]. Also, adjustments and enhancements from relevant research will be applied to this model and its integration to the real speech traffic simulation model will be introduced.

As detailed in Section 2.2.3, the speech assessment model that best fits the requirements of the present work is ITU-T G.107 (The E-Model). The primary result of the E-model is the calculation of “Rating Factor” R. The R-Factor, which is a measure of speech quality taking values from 0 to 100, can be transformed to give estimation of customer’s opinion or Mean Opinion Score. In equation 2.1, the conversion from R-factor into MOS was shown.

Taking in consideration assumptions and suggestions from ITU-T G.107 [22] discussed earlier in this thesis, specifically those modifications added in 2000 and compiled in Appendix A of [22], equation 2.2 was reduced to equation 2.3. Where I_e represents the Equipment Impairment, this factor considers codec effect and packet loss for random distributions and I_d is the Delay Impairment, associated with echo and total delay on the transmission path. For clarification purposes, equation 2.3 is redefined below as equation 3.11.

$$R = 93.2 - I_d - I_e \quad (3.11)$$

Calculation of the equipment impairment (I_e)

Equation 2.5, represents the model proposed by ITU-T G.107 [22] to calculate equipment impairment I_e . From this equation, two independent factors can be singled out. First, the equipment impairment factor at zero packet loss contemplates exclusively the degradation of the speech quality as a consequence of low bitrate encoding process. The second factor includes the effect on speech quality as a consequence of packet loss. These two factors are independent between them, although both of them are codec dependant. A simplified yet valid version of equation 2.5 can be written. In equation 3.12, I_{e0} represents the equipment impairment due to codification on zero packet_loss condition and $I_{e_pktloss}$ represents the component of the Equipment Impairment associated with the packet loss ratio.

$$I_e = I_{e0} + I_{e_pktloss} \quad (3.12)$$

The process followed by ITU to find the correct codec dependant parameters that contribute to the calculation of the Equipment Impairment (i.e., I_e at zero packet loss and packet loss robustness) requires that extensive subjective tests be performed. I_e at zero packet loss and packet loss robustness values for some of the ITU standardized speech encoding algorithms can be found in ITU-T G.113 [67]. Subjective tests can be time-consuming and expensive processes. Some researchers or organizations may not have the resources to conduct the tests. To overcome this limitation, a methodology is offered in [96] that allows the generation of the necessary E-model parameters for an arbitrary codec. Also, study in [90] enhances the work described in [96]. The E-model becomes available to a wider variety of codecs including some variable rate algorithms. In [90], parameters to integrate the E-model to AMR-NB and iLBC were obtained.

Table 3.4 Curve fitting parameters for calculation of I_e . From [96] and [90]

Parameters	AMR (H)	AMR (L)	G.729	G.711	iLBC
a	16.68	30.86	21.14	30	12.59
b*100	30.11	4.26	12.73	15	9.45
c	14.96	31.66	22.45	0	20.42

The algorithm introduced in [90], as a first step uses PESQ [60], intrusive objective speech quality assessment model (see Section 2.2.2.2 for more details) to calculate I_{e0} . A MOS value at zero packet loss conditions is obtained from PESQ. Solving equation 2.1 for the R-factor, and considering that the value of MOS calculated by PESQ does not consider any M2E delay ($I_d = 0$), an equipment impairment at zero packet-loss (I_{e0}) can be easily obtained. The second step is to assess how much each encoding algorithm is affected by packet loss ($I_{e_pktloss}$). A similar process is performed, where speech that has been exposed to incremental values of packet loss are injected to the PESQ model [60]. Knowing the I_{e0} value for the codec in used, a graph describing $I_{e_pktloss}$ vs. packet loss can be plotted. Finally, using curve fitting algorithms, an analytic function is obtained. Equation 3.13 shows such a function, where ρ is the packet loss percentage and a , b and c are curve fitting parameters. Table 3.4 presents the values of a , b and c for G.711 PLC, AMR-NB (H=12.2), AMR-NB (L=4.75), iLBC-20ms and G.729 algorithms.

$$I_e = a \ln(1 + b\rho) + c \quad (3.13)$$

Calculation of the delay impairment (I_d)

Equation 2.4 shows all elements involved in the calculation of the delay impairment I_d . Since the objective of the present work is focused on the development of an assessment tool for packet switched networks, the effects observed in circuit switched network interworking are negligible. Furthermore, considering ideal echo cancellation, we have that the listeners and talkers echo factors are considered to make no contribution to the delay impairment. From ITU-T G.107 [22], I_d as a function of the one way M2E delay in milliseconds, T_d , can be written as:

$$I_d = I_{dd} = 25 \left[\sqrt[6]{1 + X^6} - 3 \sqrt[6]{1 + \left(\frac{X}{3}\right)^6 + 2} \right] \quad (3.14)$$

Where $X = \log_2(T_a/100)$

For reference purposes, equation 3.14 is plotted in Figure 3.10 for M2E of up to one second. It can be noticed that for delays under 200 ms, almost no voice degradation due to delay impairment is registered. From M2E between 200 ms to 600 ms, the effect of the M2E delay reflects approximately linearly in the delay impairment. For M2E delays higher than 600 ms, the I_d moves asymptotically to a value of 50 ($\lim_{X \rightarrow \infty} eq. 3.14 = 50$).

Equation 3.1 accounts for the calculation of one-way M2E delay. Note that regardless of this equation being located in the AMR-NB frame calculation section, it equally applies to all other codecs. Reference values of delay related to encoding-decoding algorithms for G.729, G.711 and iLBC can be found in Appendix I of ITU-T G.114 [97] while estimations for AMR-NB codec were offered earlier in Section 3.1.1.2.

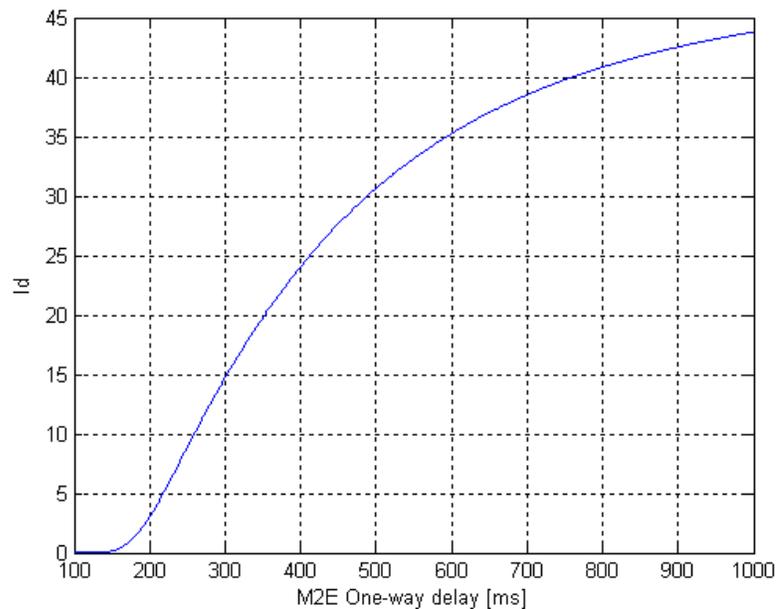


Figure 3.10 ITU-T G.107 Delay Impairment model

Most current researches based on E-model [90, 98] do not consider the time variation nature of the impairments; therefore offering misleading results. As a clarifying example, a hypothetical 100 seconds conversation can be considered. For a 1% of packet loss, conventional models will detect almost unnoticeable quality degradation. However; if the 1% packet loss is concentrated

in three seconds, then during that time, 30% packet loss is actually experienced, yielding considerable and noticeable quality degradation. In the thesis, since the packet loss ratio, M2E delay and playout buffer duration values are obtained during the simulation, the time variance of the network impairments causing speech degradation is accurately accounted for.

Figure 3.11 shows the OPNET implementation of the MOS calculations based on E-model implemented in the client node. For the calculation of the equipment impairment, the parameters *gamma1*, *gamma2* and *gamma3* correspond with the parameters *c*, *a* and *b* from equation 3.13 respectively.

```

//Calculation of Delay Impairment
X = 1.442695*log ((double) delay_total/100); //log(x) is actually ln(x) in C programming
//log2(x) = ln(x)/ln(2); 1/ln(2) ~ 1.443
Id = 25*( pow( 1+pow(X,6), (double)1/6) - 3*( pow( 1+ pow((double)X/3,6), (double)1/6) ) + 2 );
//Calculation of Equipment Impairment
Ief = gamma1 + gamma2*log(1+ (double)gamma3*pkt_loss_ratio/100);

//Calculation of R-Factor
R = 93.2 - Id - Ief;

//Conversion of R-Factor to MOS
if (R<6.5)
    MOS = 1;
else if (R> 100)
    MOS = 4.5;
else // 6.5<R<100
    MOS = 1 + 0.035*R + 7e-6*R*(R-60)*(100-R);

```

Figure 3.11 MOS calculations based on E-model implemented in the client node

3.2.5 Subjective speech quality assessment

The advantages of combining objective and subjective quality assessment models, particularly for speech traffic, were discussed earlier in Section 3.2. The present subsection will discuss the requirements and general procedures complementary to the OPNET simulation associated to the subjective speech quality assessment.

Reading the real audio file

During the simulation setup stage, in the real speech traffic generator node, a file to be simulated is specified. The input audio file has to have the following format:

- Headerless file, also called RAW file. This type of file only contains the audio samples without any header or other file structure.
- 1 audio channel (mono).

- Each sample is 16-bit signed and bit order little endian.
- Sampling rate 8000 samples/second.

Obtaining and reproducing the simulated audio file

The subjective quality assessment comes only after listening to a real speech audio. After the real speech file has been simulated, it is necessary to obtain the resultant simulated file. For this purpose, after each simulation, a batch file is created containing the necessary commands to ensure proper decoding and playout of the speech file. The following procedures will be performed in the order specified below:

1. Encode the original RAW speech audio file. If AMR-NB traffic is the encoding algorithm used, the AMR mode mask file is used to perform the encoding.
2. Decode the encoded file using the traffic mask file.
3. Deletion of temporary files.
4. Playout of the decoded audio file.
5. Presentation of the simulation report file.

The name of the batch file will depend on the encoding algorithm selected during simulation setup stage. The obvious correspondence between encoding algorithm and file name can be appreciated from the filename list below.

- *G729_encode_decode_play.bat*
- *iLBC20_encode_decode_play.bat*
- *iLBC30_encode_decode_play.bat*
- *G711A_encode_decode_play.bat*
- *G711u_encode_decode_play.bat*
- *AMR_encode_decode_play.bat*

To perform the aforementioned tasks, the following software has to be placed in the folder *C:\Voice\tools*. The encoder and decoder files are provided in [99]. The three last files in the list are open source free software and can be freely obtained from the sources specified in the references. The files which filename commences with *AHR* have been modified as part of the present work to utilize the mask files produced by the OPNET simulation.

- *G711_encoder.exe*
- *AHR_G711_decoder.exe*
- *G729_encoder.exe*
- *AHR_G729_decoder.exe*
- *AHR_iLBC_codec.exe*
- *AHR_AMR_encoder.exe*
- *AHR_AMR_decoder.exe*
- *sox.exe* [100]
- *cygwin1.dll* [101]
- *VideoLAN VLC media player* installed [102]

Notice that if consecutive simulations of the same codec are performed, the batch file, report file and applicable mask files will be overwritten without warning.

In regards to the subjective speech quality assessment methodology offered, one important factor needs to be considered. The quality of speech perceived from playing the output audio file does not consider the total M2E delay. This is a known deficiency of subjective and objective intrusive speech quality assessment models; see Section 2.2 for more details. In other words, subjective and intrusive speech quality assessment algorithms only offer a MOS-LQ (Listeners Quality). Incorrect conclusions regarding the conversational quality (MOS-CQ) of the speech sample may be derived based exclusively on the perceived quality of the speech from the output audio file. The two examples below pretend to illustrate the case.

Scenario 1: Consider the case of a network with very high and almost constant E2E delay (very low jitter). Assuming low packet loss, it can be observed that the obtained audio file will be almost identical to the original encoded file.

Scenario 2: Consider a network with a very high jitter and in order to compensate its effects the modeller has set a very large playout buffer. Once again, the result audio file will be very similar to the original encoded file.

For both cases, the resulted speech quality derived from the obtained audio file will be excellent. In actuality, a conversation occurring under either of the scenarios described above will show poor interactivity.

To avoid arriving to misleading conclusions, the associated MOS calculations associated to each voice stream should be added as a variable to the final assessment procedure. Since MOS calculations offered in the thesis as an alternative to objective quality assessment, consider both the packet loss and the E2E delay, the MOS value for both scenarios above will be low.

Chapter 4

Model Validation and Simulation Results Analysis

In this chapter, three sets of simulation results will be discussed. The objective for each set of simulations is as follows:

1. A first group of simulations is performed with the purpose of proving the validity of the proposed models discussed in Chapter 3. Results obtained from the simulation will be compared to theoretical expected values. Statistical comparisons between empirical and simulated traffic will be performed when relevant.
2. The second set of simulations focused on proving the usefulness of the proposed simulation tool in evaluating codec behaviour and properties under arbitrary network conditions. Results obtained through the objective non-intrusive speech quality assessment, i.e., the E-model, will be further corroborated through a subjective evaluation of the resultant speech.
3. Finally, several real-life scenarios are implemented and analyzed. The experiment follows the growth process of a company (i.e. in increasing order of complexity). Multiple possible solutions to the challenges and difficulties rose from the expansion process are offered.

4.1 Validation of simulation models

The current section evaluates the validity of the traffic generation models for all codecs. The validation of the models is equally applicable to both the background traffic generation model and the real speech traffic generation node. The same principles for calculating the frame size and frame duration are applied in both nodes. Given that the codecs and codec modes used in the real speech traffic generator node are a subset of those implemented in the background traffic generator node, the latter is used in the simulations performed in this section.

The validation process has been performed independently for the fixed data rate group of codecs, AMR-NB codec and Speex codec. During the analysis of the simulation, emphasis will be made in comparing the results obtained in the simulation with theoretical values and

boundaries. Also, when relevant, statistical comparison between the simulated data and empirical data will be provided.

4.1.1 Validation of traffic generation models for fixed data rate encoding algorithms

While validating the results for the fixed data rate algorithms, not only the correct creation of speech frame is assessed but the entire process of creating the final physical layer packet. By analyzing the contribution to the packet size of all protocols involved, a bandwidth prediction can be made. The predicted bandwidth, based on the analysis of the standards for each protocol, can then be compared to the results obtained from the simulation.

For the current experiment, the bandwidth usage of the simulated data will be compared with the predicted bandwidth requirements in an Ethernet link. The following protocols are involved in the simulation of speech traffic in an Ethernet link: Ethernet (data link layer) [103], Internet Protocol (network layer) [104], User Datagram Protocol (transport layer) [105] and RTP [46]. Figure 4.1 shows the structure of a packet once it leaves the link layer. Each block only represents the contribution in size to the total size of the packet, the actual allocation of bits in the packet is not accurately depicted; e.g. the Ethernet header has 4 bytes at the end of the packet for cyclic redundancy check, the position of those bits is not considered but their contribution of the packet size is.

Ethernet Header (18 B)	IP Header (20 B)	UDP Header (8 B)	RTP Header (12 B)	voice payload (variable)
---------------------------	---------------------	---------------------	----------------------	-----------------------------

Figure 4.1 Structure of a voice packet on an Ethernet link considering the contribution of all protocols involved

Based on Figure 4.1 and using the simulation parameters and the encoder algorithms nominal parameters, the bandwidth usage on an Ethernet link can be formally calculated as shown in equation 4.1 for each codec. Where $codec_{frsize}$ and $fr_{duration}$ represent the nominal frame size (in bytes) and duration for each fixed data rate encoder respectively. $number_{fr/pkt}$ is the number of speech frames encapsulated in each packet, which is defined by the modeller at the simulation setup stage.

$$BW [bps] = \frac{8 (58 + codec_{frsize} \times number_{fr/packet})}{fr_{duration} \times number_{fr/packet}} \quad (4.1)$$

Table 4.1 shows bandwidth requirement calculations based on equation 4.1 for each of the encoder algorithms in the fixed data rate group, codec nominal parameters can be found in Table 3.1. It can be observed that by increasing the number of speech frames encapsulated in a packet, the bandwidth requirement decreases. This effect is to be expected from equation 4.1 given that the overhead for each packet is constant while the packet size and the packet interarrival time change with the size of the voice payload. Despite the decrease of the bandwidth requirement, the effect on the increase of the M2E delay also needs to be considered (in Section 3.1.2, a study of the effect on the increase of the number of speech frames encapsulated per packet is offered). A balance between M2E and codec bandwidth consumption needs to be found by the modeller or network designer. The equilibrium point may depend on the network load, packet loss ratio, jitter and even on the subjective performance expectations of the users of the system.

Figure 4.2 depicts the results of an OPNET simulation where a speech background traffic generator node and a client node were introduced in a simple bus topology and was set to generate traffic corresponding to the four fixed data rate algorithms discussed at two frames per packet. The simulation was performed with no jitter, constant E2E network delay and zero packet loss.

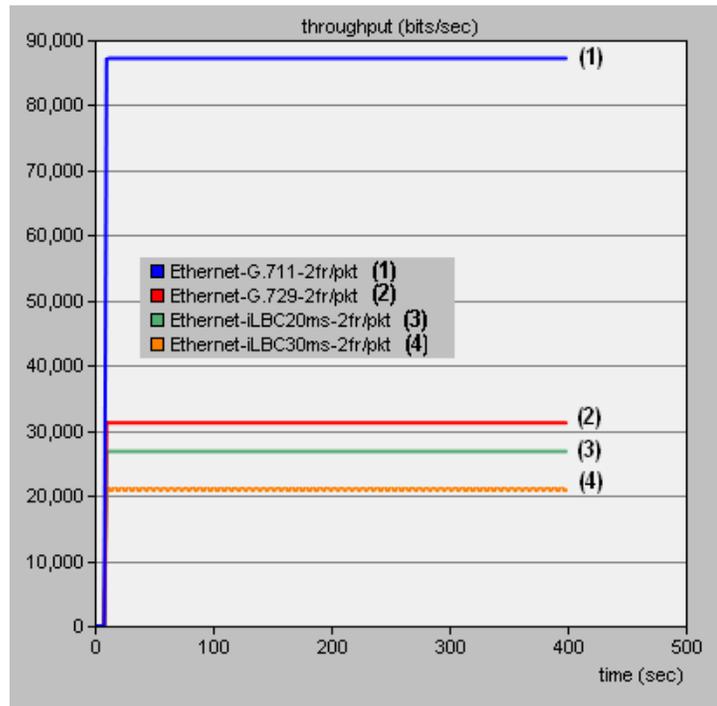


Figure 4.2 Simulated bandwidth usage in an Ethernet link for G.711, G.729, iLBC-20ms and iLBC-30ms at 2 frames per packet

Table 4.1 Predicted bandwidth usages in an Ethernet link for G.711, G.729, iLBC-20ms and iLBC-30ms for 1-4 frames per packet

Codec	BW [bps] 1 fr/pkt	BW [bps] 2 frs/pkt	BW [bps] 3 frs/pkt	BW [bps] 4 frs/pkt
G.711	110400	87200	79466.67	75600
G.729	54400	31200	23466.67	19600
iLBC20ms	38400	26800	22933.33	21000
iLBC30ms	28800	21066.67	18488.89	17200

It can be observed that the results correspond exactly to those predicted in the **2 frs/pkt** column of Table 4.1. Based on the results of the simulation illustrated in Figure 4.2, it can be concluded that the traffic generation algorithms for fixed data rate algorithms have been correctly

modeled. Also, it can be stated that the behaviour of the lower network layers (presentation layer to MAC layer) in the traffic generation node is as expected³.

4.1.2 Validation of traffic generation model for AMR-NB encoding algorithm

To test the validity of the speech background simulation model for AMR-NB traffic, the background traffic node and the client node were placed in a simple topology where the traffic path was forced through an IP cloud network model in OPNET. The IP cloud model allows for the configuration of latency experienced by packets inside the cloud.

The AMR-NB configuration in the traffic generator node was set to 1 frame per packet. In the client node, the RTCP frequency was set to 1/150 RTP packets (approximately 3 seconds between RTCP packets). The delay in the IP cloud was modelled as a Normal distribution, $N(0.175,0.001)[s]$.

For AMR-NB, it is expected that the size of the speech frame will fluctuate according to the M2E delay values. Figure 4.3 shows a plot of the simulated AMR-NB frame size and the M2E delay over time. It can be observed that the frame size (top) and the M2E (bottom) are highly correlated, when M2E increases the frame size decreases and vice versa, common sense and Table 3.2 indicate likewise. Also, the frame size is bounded by 13 bytes and 31 bytes, Table 3.2 confirms these values.

³ All modifications applied to the three nodes described in Chapter 3 and Appendix A were implemented in the application layer of the nodes. The implementation of the lower network layers remains as originally provided by OPNET.

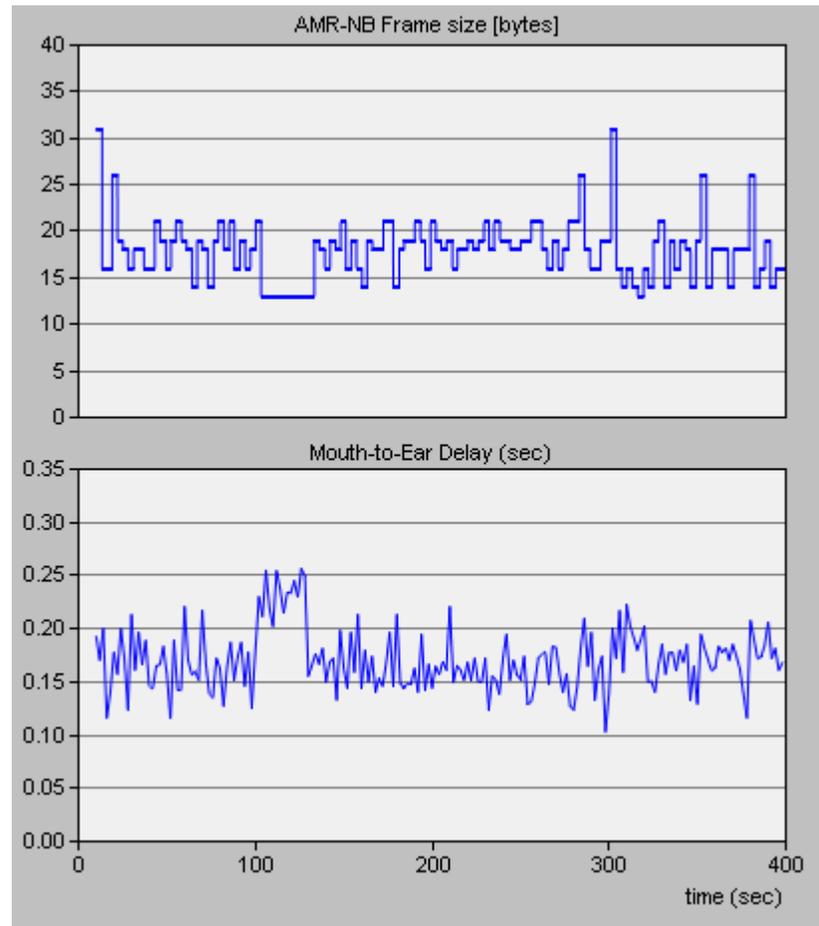


Figure 4.3 Variations of the simulated AMR-NB frame size with the M2E delay

Furthermore, the scale in the graph in Figure 4.3 does not allow for accurate measurements of the frame size adjustment method as a function of the M2E delay. Figure 4.4 shows in more details this process. In Figure 4.4, the top graph represents the E2E delay read from the RTCP packets at the traffic generator node, each transition corresponds with the reception of one RTCP packet. The top graph offers information on the time instant when the RTCP packet was received (transitions) as well as the E2E delay extracted from such packet (value of the curve on vertical axis). From the top graph can also be appreciated that a RTCP packet is received approximately every 3 seconds, this is consistent with the settings of the simulation where 1 RTCP packet is generated by the client node for every 150 RTP packets received. The bottom graph in Figure 4.4 shows the variations on the AMR-NB frame size generated in the background traffic generator.

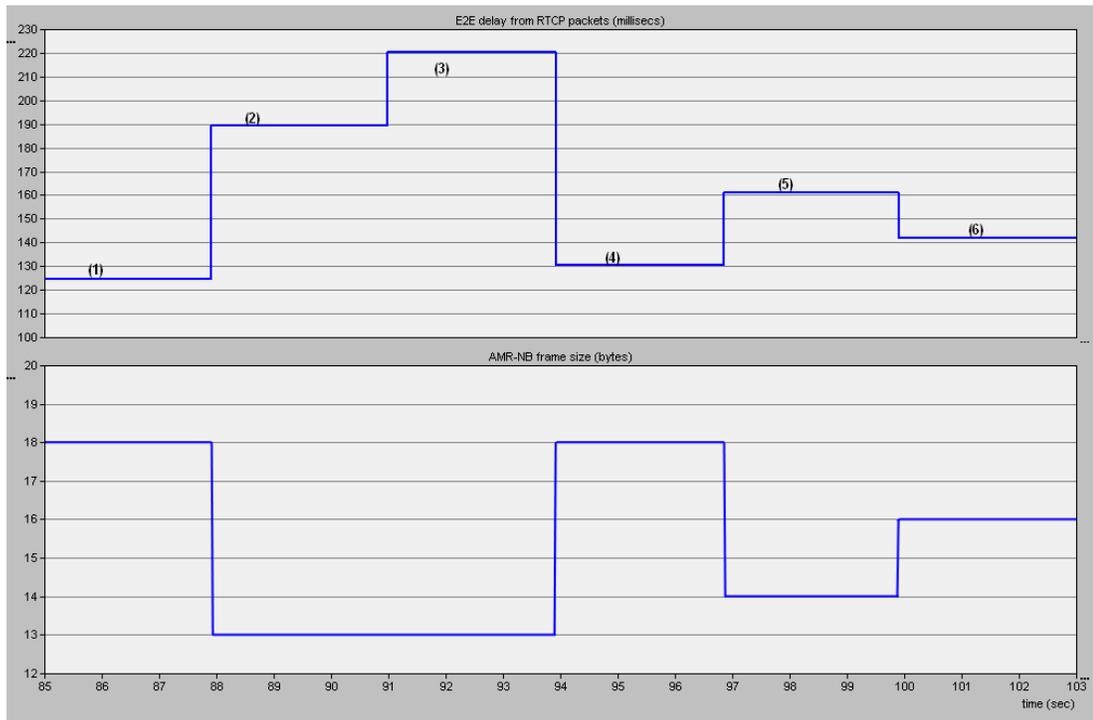


Figure 4.4 Detailed view of the frame size adjustment procedure for AMR-NB encoder algorithm

Due to the scale of both graphs, it seems that one AMR-NB packet is generated right after a RTCP packet is received (transitions in both graphs are very close in the time axis). These two events are not correlated; however, it should be noticed that one AMR-NB packet is generated every 20ms, in the scale of the graph, a 20 ms interval is indistinguishable.

Table 4.2 Predicted AMR-NB frame size based on the E2E delay values read in the background traffic generator node

Interval	E2E [ms]	M2E [ms]	Predicted frame size [B]
1	125.0	195.5	18
2	190.0	265.5	13
3	221.0	291.5	13
4	131.0	201.5	18
5	162.0	232.5	14
6	142.0	212.5	16

Table 4.2 shows the E2E delays for every time interval marked in the top graphic of Figure 4.4 with numbers from 1 to 6, the M2E delay corresponding with each of these intervals (see equation 3.6 and Section 3.1.1.2 for more details) and the frame size obtained from Table 3.2.

From close inspection of the bottom graph in Figure 4.4, an exact match between the predicted frame size values and the values obtained in the simulation can be appreciated.

4.1.3 Validation of traffic generation model for Speex encoding algorithm

Finally, the approach used to proof the validity of the traffic generation model for Speex traffic was the statistical comparison between the empirical data obtained from encoding real speech and the OPNET simulated data.

A simulation was configured where 360,000 Speex quality 2 frames were generated in the traffic generator node. One Speex frame was encapsulated per packet. Since Speex frame size does not depend on the network delay or jitter, the experiment was performed under constant network delay and zero packet loss conditions.

The simulated Speex frames were collected in the client node and injected to TESTool⁴ along with the original empirical data set (see Section 3.1.1.3 for more information on TESTool and Speex frame size generation model).

From Figure 4.5, it becomes clear the good visual match in terms of marginal distribution and autocorrelation function between the simulated data and the empirical data. Also, a resemblance can be noticed between the two sample paths. An exact model validity analysis was performed on the generation of Speex Quality 8 traffic obtaining similar conclusions.

Additionally, a visual notion of how traffic generated from Quality 2 and Quality 8 mode could impact network utilization can be derived from Figure 4.6. When closely analyzed, Figure 4.6 reveals that for Speex Quality 8, the frame size has its higher probability density between 40-55

⁴ For Figures 4.5 and 4.6, only the functionality of TESTool of computing and plotting marginal distribution and autocorrelation function are exploited, some other tool like MATLAB could have been used.

bytes (16 kbps -22 kbps) and for Speex Quality 2, the frame size is heavily concentrated between 10-25 bytes (4 kbps - 10 kbps). The obtained results coincide with the properties of Speex codec discussed in Section 3.1.1.

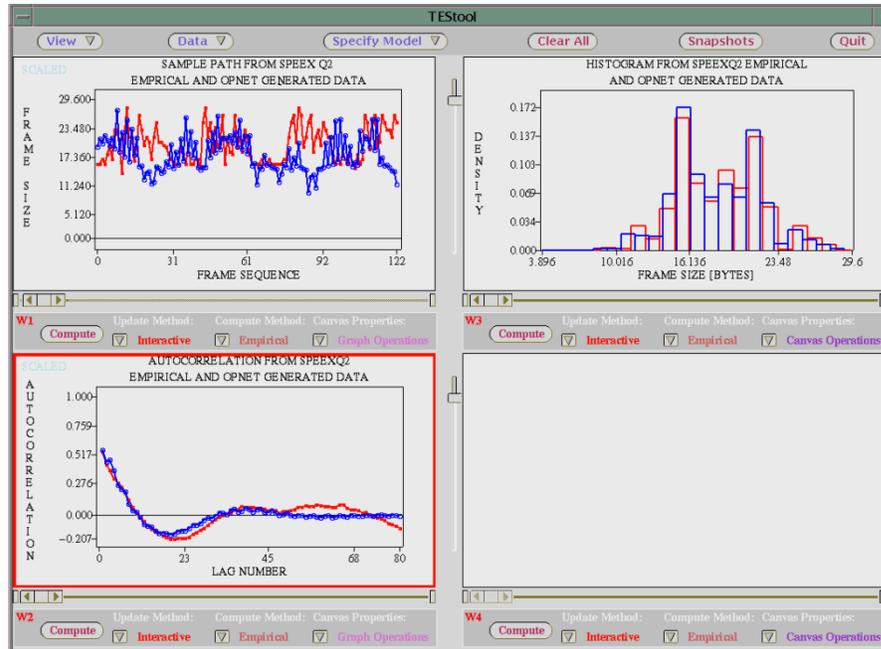


Figure 4.5 TEStool screen shot. Empirical Speex Q2 data (red), OPNET simulated Speex Q2 data (blue). W1 (top-left), sample path. W2 (bottom-left) autocorrelation function. W3 (top-right) marginal distribution

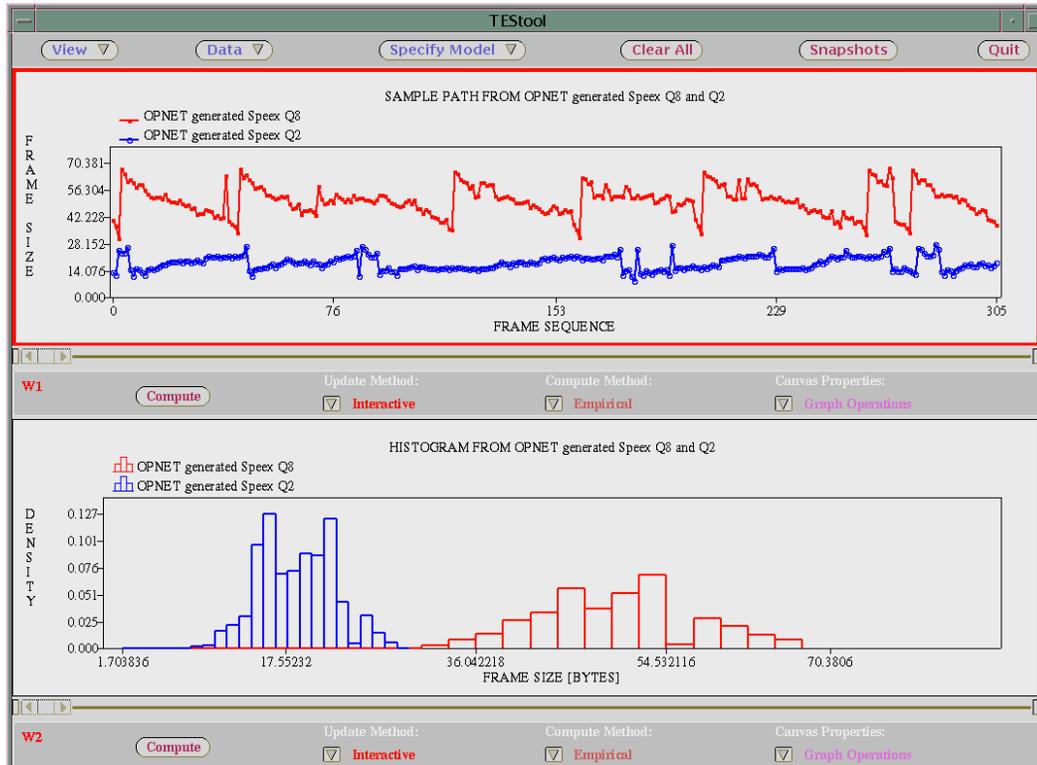


Figure 4.6 Window 1 (top): Sample path and PDF for Speex data generated in OPNET using TES model for Speex Quality = 2 and Speex Q = 8. Window 2(bottom): PDF of Speex frame size [bytes] for Speex Quality = 2 and Speex Q = 8

4.2 Codec packet loss concealment evaluation and playout buffer optimization

The simulations and analysis discussed in this section are intended to illustrate the possible use of the simulation platform as an aid in the codec selection process and the design of playout buffer size optimization algorithms.

The objective of the current simulation is to evaluate the performance in terms of packet loss robustness of four different speech encoding algorithms under constant network condition. This experiment is not meant to determine superiority among codecs. The selection of the correct codec for each situation is a complex analysis that comprehends variable and more realistic network conditions.

Each codec is affected by the effects of packet loss and constant delay (see Sections 2.2.2 and 3.2.4 for more details). On a constant jitter and packet loss environment, the increase of the

playout buffer has two consequences. The first one is to improve the packet loss ratio by permitting the encoder to process packets that have arrived delayed. When referring back to the E-Model [22], it can be noticed that improvement in the packet loss ratio brings a decrease in the equipment impairment I_e , therefore, an increase in the MOS for the call. Also, by increasing the playout buffer size, the total delay experienced by the call increases; this increment affects the delay impairment I_d according to the E-Model, causing degradation in the MOS for the call. The two opposite effects of increasing the playout buffer size manifest differently for each codec depending on how robust is the algorithm to packet loss.

In the simulation, a conversation of five minutes duration is put through the topology depicted in Figure 4.7. Four different codecs (G.729, iLBC-30ms, G.711 and AMR-NB) are tested independently and their performance evaluated. One speech frame was encapsulated per packet. The E2E delay measured in the *Office 2-Callee* node was randomly distributed according to a Normal distribution $N(195 \text{ ms}, 78 \text{ ms})$. The playout buffer size was modeled from 10 ms to 610 ms with 20 ms interval.

Ten replications using different seeds were performed for each codec. For the results shown in Figure 4.8, the MOS is bounded by $\text{MOS} \pm 0.17$ for AMR-NB with 90% confidence. For all other codecs, the error margin is smaller.

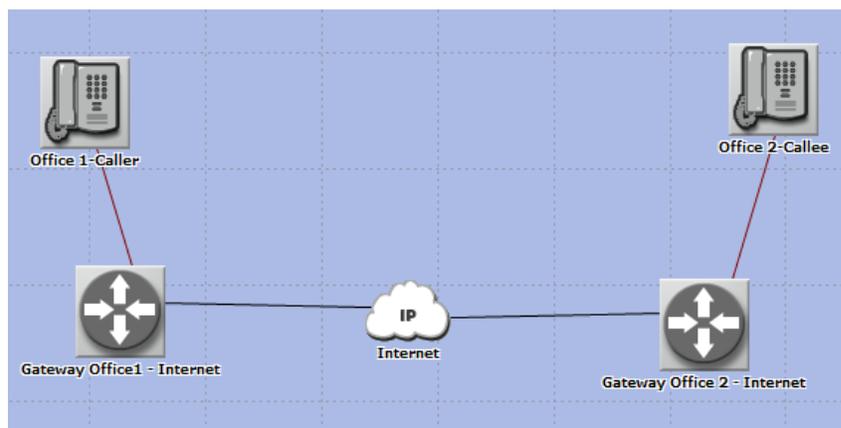


Figure 4.7 Topology used in the codec performance evaluation as a function of playout buffer size

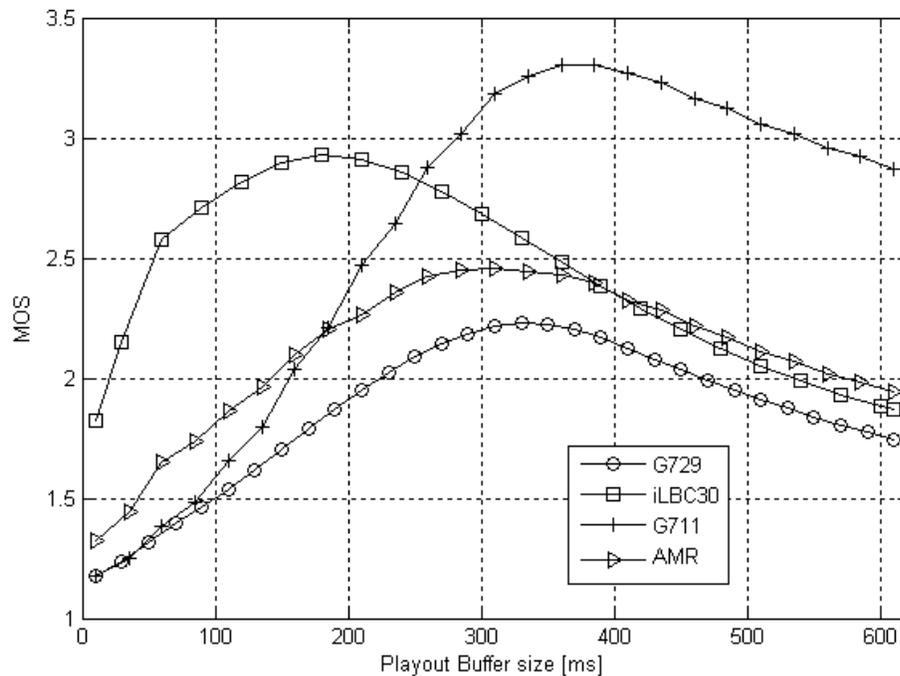


Figure 4.8 Evaluation of codec performance at constant network conditions with the increase of the playout buffer size

From the curves obtained from the simulations and shown in Figure 4.8 some generalizations can be made in regards to the principles driving the shape of the graphs. As a general interpretation, the playout buffer size value yielding the maximum MOS value is clearly the optimum buffer size for the network conditions of the experiment. For the segment of the curve to the left of the optimum buffer size point, the effect of the packet loss is predominant in causing a low MOS value. For the right side of the curve, the predominant effect becomes the delay introduced by increasing the buffer size.

In Figure 4.8, it can be noticed that the codec that offers the highest MOS value is G.711. This is not a surprising result considering that G.711 has a codec bandwidth of 64kbps, eight times higher than G.729 and approximately five times higher than iLBC 30ms and AMR-NB (when in 12.2 kbps mode). The algorithmic complexity of G.711, usually directly associated to the compression rate and inversely related to the output speech quality, is very low compared to all other speech codecs. It can be noticed that iLBC achieves its optimum MOS value at a lower buffer size than all other codecs. This fact indicates that the packet loss concealment

implemented in iLBC performs very well compared to the other codecs (similar conclusions can be found in [90]). iLBC reaches the optimum MOS value at a buffer size approximately equals to 180 ms. AMR-NB and G.729 seems to be affected very similarly by the effect of packet loss.

In both cases, the optimum MOS value is achieved for a buffer size approximately equals to 300 ms. However, G.711 yielded the maximum value of MOS, it needed a buffer size of approximately 360 ms to achieve its maximum MOS value. By examining the slopes of the left and right sides of the curves, it can be concluded that G.729 and AMR-NB seem to be affected in similar magnitude by the positive effect of the packet loss decrease (left side of the curve) and the negative effect of the total E2E delay increase (right side of the curve). For iLBC and G.711, they benefit quickly from the decrease in the packet loss ratio (steeper slope of the left side of the curve) while the decrease in MOS value as consequence of the increase on E2E delay is slower (flatter slope of the right side of the curve).

Table 4.3 Optimum plyout buffer size as a function of jitter for G.729 codec

Average Jitter [ms]	Buffer size [ms]	Maximum MOS
6.71	30	3.58
8.29	50	3.52
10.77	70	3.44
11.91	70	3.35
13.28	110	3.22
14.49	130	2.98
15.11	150	2.80
15.57	190	2.68
15.86	210	2.58
16.71	290	2.20
17.60	490	1.79

From the perspective of the design of buffer size optimization algorithms, if the experiment is repeated for different values of jitter, three-dimensional curves can be obtained for each codec. Using these curves, a set of optimal buffer sizes for all jitter values can be obtained for each codec. It is to be expected the optimal buffer to be dependant on the jitter value. A very low-complexity buffer size optimization algorithm can be implemented based on lookup tables. Such algorithms would be easily scalable to new speech codecs.

Figure 4.9 shows a graph of MOS as a function of average jitter and playout buffer size for G.729 codec. It can be observed that as expected, when the jitter increases, the maximum MOS value is achieved at higher playout buffer values. Larger buffers are needed to compensate for the wider spread delay of the packets. It also becomes clear, when comparing equi-jitter curves, that with the increase of the jitter, the maximum achievable MOS value for each curve decreases. This reflects the negative impact of the increase of the M2E delay (when buffer increases, M2E delay increases) on the speech quality. Table 4.3 offers the optimum playout buffer sizes for all values of average jitter; also, the maximum MOS value for each curve is offered. This experiment can be repeated for as many codecs as desired.

Finally, and with the purpose to corroborate the validity of the results offered in Figure 4.8, a subjective test has been performed. Two real speech files were put through the simulation and encoded according to G.729 and iLBC 30ms. Both conversations were simulated with playout buffer size = 150 ms. From Figure 4.8, it can be confirmed that at this point, the average MOS for G.729 simulations is approximately 1.75 and for iLBC-30 ms is approximately 2.8. The two conversations obtained can be found in [99]. By listening to files (*experiment1_G729_buffer=150ms_MOS=1.55.wav* and *experiment1_iLBC30_buffer=150ms_MOS=2.92.wav*), it can clearly be appreciated that the difference in speech quality between the two encoding algorithms are as expected from the simulation results.

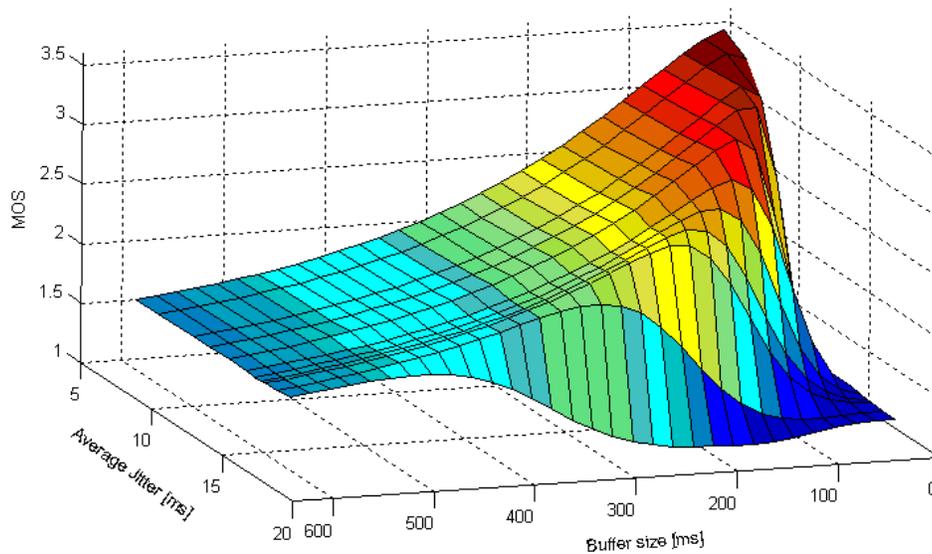


Figure 4.9 MOS as a function of the playout buffer size and the average jitter for G.729 codec

4.3 Simulation of a real life scenario

The experiment described in this section intends to prove the usefulness of the proposed simulation platform in dealing with real life scenarios. For that purpose, a potential real life scenario will be simulated and the results analyzed. The experiment will focus first on a company with a small office located in Ottawa and subsequently, the effects of the addition of a bigger office in Toronto. The challenges of expanding the network infrastructure and traffic will be considered and possible solutions offered.

4.3.1 Deploying VoIP within a single office

In this section, the traffic in a small office in Ottawa will be simulated and results analyzed.

This scenario focuses in the decision making process associated to the transition from standard telephony to VoIP telephony for a small company. The motivations and advantages of the utilization of centralized data and voice network models were presented in Chapter 1.

The network topology of the company is presented in Figure 4.10. At the moment of the transition, the following network services were registered.

- HTTP
- FTP
- Email
- Database access

The network of the office is divided into three sections, subnet_0 (connected to *Switch_0*), subnet 1 (connected to *Switch_1*) and the servers' subnet (connected to *Swith_servers*). Each subnet 0 and subnet 1 hosts half of the workstations from the network. All traffic associated to the network services abovementioned is exchanged between the *Ottawa LAN_0*, *Ottawa LAN_1* nodes and the servers. The *app_config* and *Profiles* nodes allow the configuration of the data traffic.

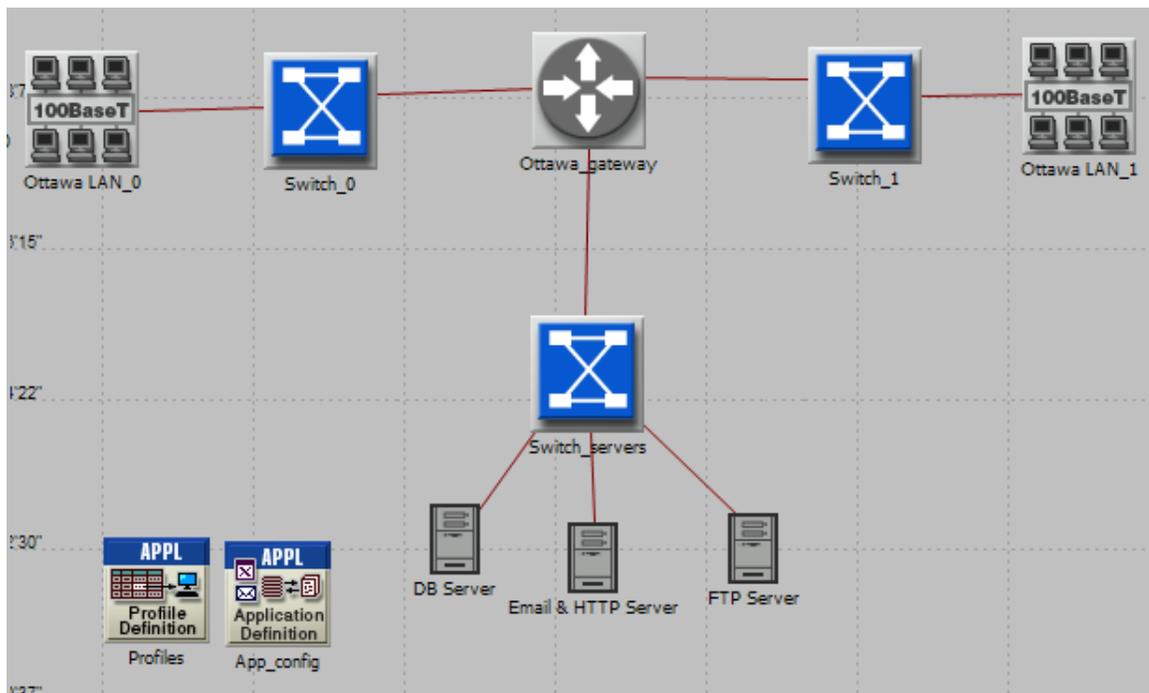


Figure 4.10 Topology of the Ottawa office network carrying data traffic only

Table 4.4 Characteristics of data traffic inside small office in Ottawa

Application	Interarrival time [sec]	Packet size [bytes]
HTTP	exp (60)	Page size Text: constant (1000) Image: uniform (50000,150000)
e-mail	Send: exp (60) Receive: exp(60)	constant (2000) constant (2000)
FTP	exp(10)	constant (100000)
Database	exp (3)	exp (32768)

Data Traffic

The simulation of data traffic is performed using the procedures and libraries offered in OPNET Modeler by default. The *app_config* and *Profiles* nodes help define the interarrival time, packet size, duration and repeatability of the profile for each application traffic. Table 4.4 shows the characteristics of the traffic for each application during 10AM-12PM, which has been found the busiest time of the day in terms of data traffic. Most values shown in Table 4.4 correspond with the OPNET default values for the heavy traffic profile for each application (see OPNET Modeler documentation for more information).

The data traffic flows from the computers in the subnet 0 to the servers and identically for subnet 1. Each LAN node in each subnet is formed by three computers. It is important to note that the number of computers does not reflect a real numbers of workstations, instead it reflects the average number of systems sending and receiving data traffic according to the profiles shown in Table 4.4 at the same time at all times during the length of the experiment. The same concept will be applied to all simulation scenarios discussed in this chapter.

After rigorous analysis of the utilization of the telephone system in the office, the average number of calls occurring between subnet 0 and subnet 1 was obtained. Figure 4.11 represents the new network topology hosting both data traffic and the equivalent VoIP traffic to the current standard voice traffic.

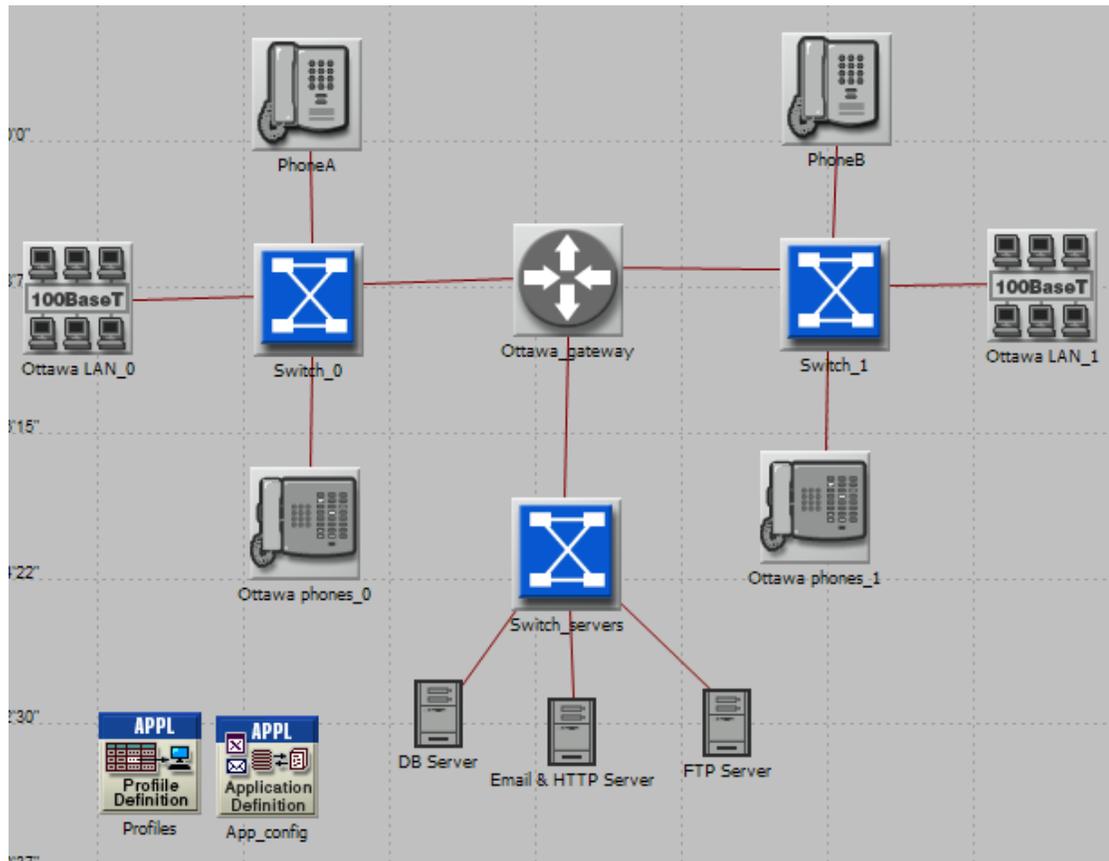


Figure 4.11 Topology of the Ottawa office network carrying data and voice traffic

Voice Traffic

The simulation of voice traffic has been discussed in Sections 3.1 and 3.2. There are two components to the voice simulation: the speech background traffic and the insertion of a real speech file (see Chapter 3 and Appendix B for more details).

In the network topology illustrated in Figure 4.11, the nodes involved in the generation of the speech background traffic are the *Ottawa phones_0* and *Ottawa phones_1* nodes. These two nodes correspond with the background traffic generator node and the client node respectively; these nodes have been described in Chapter 3 and Appendix A. The *Ottawa phones_0* node (speech traffic generator node) generates the traffic equivalent to the average number of phones minus one located in subnet 0 that are engaged in a call with phones in the subnet 1 simultaneously and during the total duration of the experiment. This number does not necessarily reflect the physical number of phones located in each subnet. Analogously, *Ottawa*

phones_1 node encompasses the average number minus one of phones answering a call from phones in the subnet 0. The same concept will apply to all scenarios simulated in this chapter. It is assumed that for this experiment, all VoIP traffic occurs is confined to the office.

The nodes involved in the generation and processing of the real speech audio traffic are the *PhoneA* and *PhoneB* nodes. *PhoneA* and *PhoneB* correspond with a real speech traffic generator node and a client node respectively; these nodes have also been described in Chapter 3 and Appendix A.

The objective of the simulation is to determine if the current office network infrastructure is suited for supporting the additional VoIP traffic. Furthermore, it is also a goal to find out which speech codec offers the most advantages given the abovementioned network conditions.

During the busiest period in the office (10AM-12PM), twelve conversations were registered to take place simultaneously between the phones located in subnet 0 and those located in subnet 1.

Twelve conversations were simulated using G.711, G.729, AMR-NB and iLBC-30ms speech codecs. Eleven of these conversations were routed between the *Ottawa phones_0* and *Ottawa phones_1* nodes. The last conversation was routed between the *PhoneA* and *PhoneB* nodes. In the *PhoneB* node, statistics for speech quality were collected. This conversation is a real audio file lasting five minutes; the conversation was injected after the simulation had been running for 5,000 seconds.

From Figure 4.12, the MOS for the simulation of twelve conversations for each codec is depicted (left vertical axis). During the simulation, the link utilization statistics were collected for each link of the network topology. The busiest links with impact on the speech quality are *Switch_0* ↔ *Ottawa_gateway* and *Switch_1* ↔ *Ottawa_gateway*. The link utilizations for these links are almost identical when the simulation is run for a long time. Note that the network topology is symmetric around the *Ottawa_gateway* node. The link utilization for the link connecting *Switch_0* ↔ *Ottawa_gateway* is also shown in Figure 4.12 for each of the codecs simulated (right vertical axis).

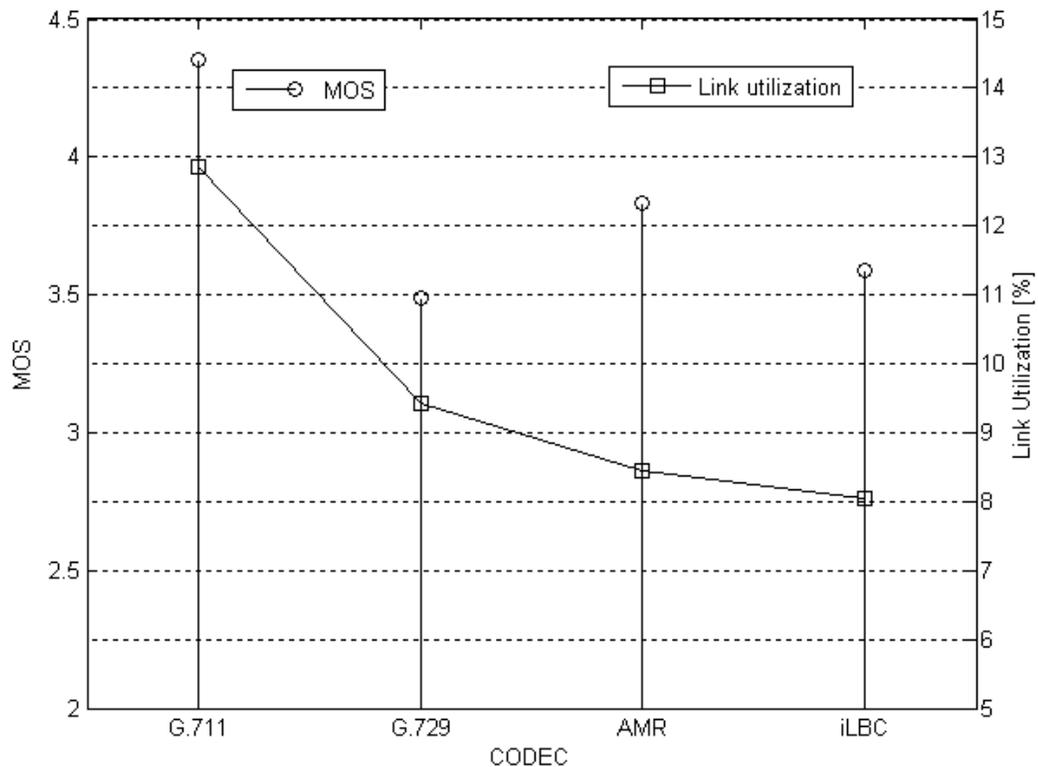


Figure 4.12 MOS (left vertical axis) and Link utilization (right vertical axis) for G.711, G.729, AMR and iLBC-30ms codecs.

From Figure 4.12, the codec offering the best performance is G.711 with a value of 4.35. According to Table 2.6, this MOS value is considered between **Good** and **Excellent**. The rest of the codecs, given the network load conditions, also perform considerably well, with values ranging from 3.49 (G.729) to 3.83 (AMR). Analyzing the link utilization, it can be noticed that the highest value is also reached for the case of G.711. This peak value on the link utilization however, is only of 12.86 %. Considering that 12.86% link utilization includes data traffic and voice traffic in that link, it can be concluded that the use of G.711 does not make a negative impact in the network performance. After analyzing the speech quality provided by the G.711 codec and corroborating that the effect in the link utilization is acceptable, it can be concluded that for the present network conditions, the preferred codec is G.711. Also, it is important to consider that G.711 presents the lowest algorithmic complexity of all speech codecs; perhaps allowing the employment of simpler VoIP equipment.

Despite all the advantages offered by G.711, the main characteristic that always needs to be observed closely is its bandwidth requirement. Based on the bandwidth requirement values for an Ethernet link shown in Table 4.1, it can be noticed that G.711, for the case of two frames per packet, requires between 2.8 to 4.1 times more bandwidth than the rest of the codecs analyzed. To ensure that the selection of G.711 codec will not hinder the possible expansion of the network, two more simulations were run. The new simulations consider an increase in the speech and data traffic in the network of two and three times the current traffic volume. For the first simulation, 23 conversations were run between the *Ottawa phones_0* and *Ottawa phones_1* nodes and one conversation was routed between the *PhoneA* and *PhoneB* nodes. The LAN in each subnet was formed in this scenario by 6 computers. For the next scenario, 35+1 conversations were simulated and nine computers were included in each LAN node. The results of these simulations are depicted in Figure 4.13.

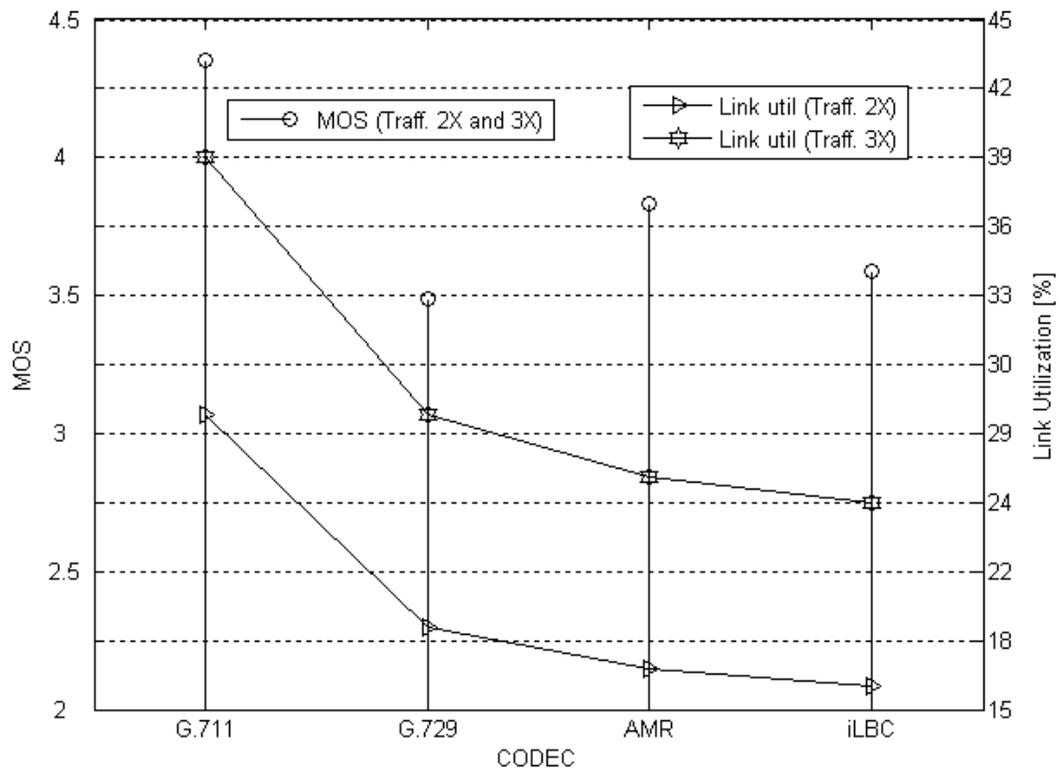


Figure 4.13 MOS (left vertical axis) and Link utilization (right vertical axis) for two times and three times traffic for G.711, G.729, AMR and iLBC-30ms codecs.

In Figure 4.13, the maximum link occupancy for both two-time and three-time traffic curves is achieved, as expected, while using G.711 codec. In the two-time traffic scenario, the link utilization reaches 27.82 % and it climbs to 39.05 % in the three-time traffic scenario. The values of MOS for all codecs remain the same as shown in Figure 4.12.

As a conclusion, even after considering that the network size increases twofold and threefold, the MOS values obtained remain the same as in the original scenario. Besides, the link utilization remains within acceptable values, allowing for good performance of the network services as well as good voice quality. Then, the current infrastructure of the network was found to be sufficient to carry data and the new voice service. In addition, the selection of G.711 as speech codec, does not pose a threat to future network expansion allowing for the best speech performance and low network load.

Ten replications were performed in all three scenarios for each codec. The MOS values did not manifest any random behaviour. Any randomness observed in the MOS values is driven by the non-deterministic nature of the data traffic and the variable rate codecs; causing variable delays and queue lengths in the switching nodes. For small values of link utilization, and given that the relationship between link delay and link utilization is not linear [106], the delay variations are negligible and do not affect the MOS values. All small delays are perfectly compensated by the playout buffer in the client node and no packets are lost. For the link utilization statistic, a similar effect is observed. This time the reason of the deterministic nature of the statistic is linked to the fact that all codecs operate at constant bit rate. AMR-NB codec, which potentially can operate in variable rate mode, modulate its output rate based on the delay on the path. In the scenarios analyzed, the delays always remained under the smallest threshold, allowing AMR to operate at all times in its highest mode (see Table 3.2).

4.3.2 Deploying VoIP across several branch offices

As a natural growing process of the company used as example in Section 4.3.1, it is assumed in this section that a new branch office has been opened in Toronto. Figure 4.14 shows the topology of the company network, the two offices are shown as subnets interconnected by a DS1 (1.54 Mbps) dedicated link. Figure 4.15 shows the topology of the intranet for the Toronto office (left) and Ottawa office (right).



Figure 4.14 Network topology. Toronto office and Ottawa office interconnected by a DS1 link.

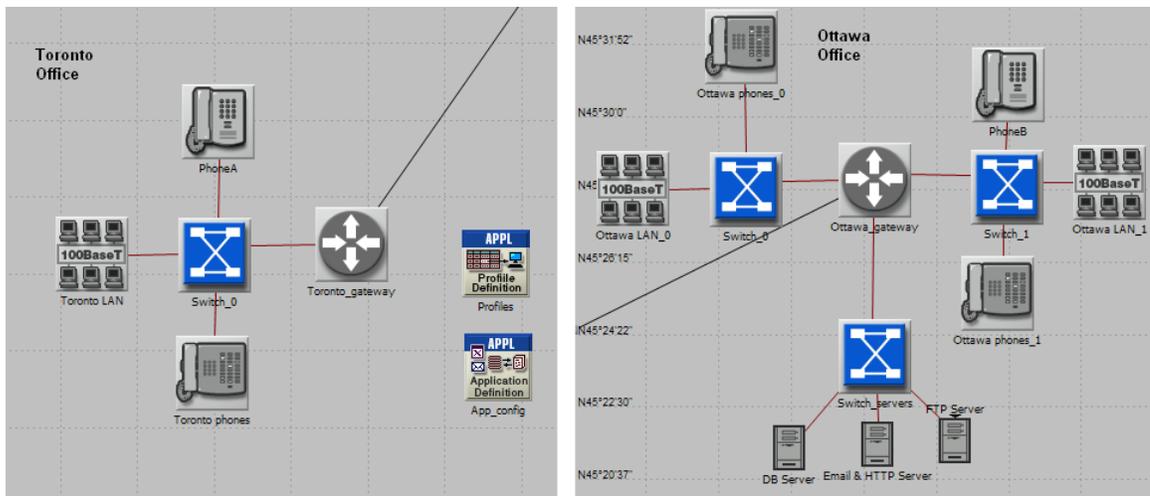


Figure 4.15 Office network topology. Left: Toronto office's network topology. Right: Ottawa office's network topology

Data Traffic

The specifications and the simulation principles of the data traffic is identical to the definitions discussed in Section 4.3.1 (see Table 4.4). The *Toronto LAN* node is composed by 10 computers, while the *Ottawa LAN* nodes remains the same size as in the experiment in Section

4.3.1; i.e., three computers in each Ottawa subnet. Note that the servers that handle the data traffic remain in the head office in Ottawa.

Voice traffic

The same structure discussed in Section 4.3.1 for the simulation of the speech traffic applies to this scenario. It was determined, that in the busiest time of the day (10AM-12PM), 12 conversations are exchanged between the Toronto office and the Ottawa office. To this end, in the Toronto office, the *Toronto phones* node generates 11 conversations that are routed to the Ottawa office. Of these conversations, six are routed to *Ottawa_phones_0* node and five are routed to the *Ottawa_phones_1* node. In the Ottawa office, the local voice traffic is generated as defined in Section 4.3.1. Finally, after the simulation is running for 5,000 seconds, a five minute real audio file is inserted in *PhoneA* (Toronto) and routed to *PhoneB* (Ottawa), which is located in the same subnet that *Ottawa_phones_1* node. The experiment is repeated for G.711, G.729, AMR-NB and iLBC-30ms codecs. Statistics are collected in the *PhoneB* node at the arrival of the speech traffic.

The objective of the simulation is to determine if the company network's infrastructure is suited for the proposed expansion. It is also a goal to find out which speech codec offers the most advantages given the abovementioned network conditions.

Figure 4.16 illustrates the MOS measured in the *PhoneB* node for each of the encoding algorithms. Link utilization of all links in the network topology was collected to find that the most critical link is the DS1 link connecting the two offices. The average link utilization for this link is shown in Figure 4.16 for each codec.

Contrary to the scenario analyzed in Section 4.3.1, where G.711 proved to be the indisputable codec option; in this case, G.711 yields the lowest MOS value. When analyzing the link utilization for the G.711 scenario, the reason for the low MOS becomes obvious; link utilization equals to 100% implies considerably high delays and packet loss. The highest MOS value is achieved in this scenario by iLBC-30ms codec; the values measured were MOS = 2.77 and link utilization = 73.37%. In terms of MOS, the next best value is offered by G.729 codec; values measured for G.729 were MOS = 2.67 and link utilization = 76.7%.

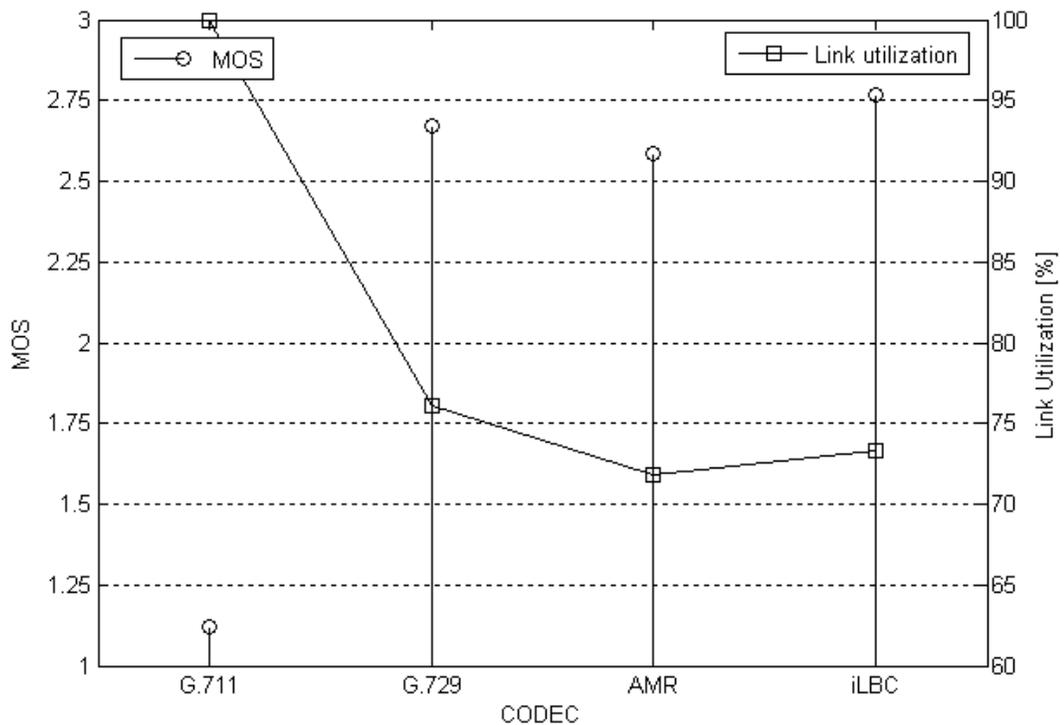


Figure 4.16 MOS (left vertical axis) and Link utilization (right vertical axis) for G.711, G.729, AMR and iLBC-30ms codecs

Focusing on the MOS, the difference between the values measured is 0.1. This difference is barely perceptible; however, the advantage of iLBC-30ms is reinforced when noted that the difference in link utilization is over 3% in favour of iLBC-30ms. Finally, AMR-NB yielded lower MOS value than iLBC-30ms and G.729. Interesting result when, going back to Figure 4.12, AMR-NB was the best codec after G.711 in the scenario presented in Section 4.3.1. Since AMR-NB adjusts its bandwidth based on the M2E delay, which is directly related to link delay, it is to be expected that with link utilization over 70%, the M2E delay is considerable high. Therefore, AMR-NB operates in the lower, if not the lowest, bit rate; yielding low MOS values (see Table 3.2). This assumption is supported by the fact that, although still high, AMR-NB causes the lowest link utilization of all four codecs simulated. Under the operation conditions, the best codec option remains iLBC-30ms.

Although a best-choice codec can be found, the results obtained are far from satisfactory. According to Table 2.6, the quality of conversations with MOS values between 2 and 3 is

considered to be **Poor** [53]. In addition, the link utilization for the trunk connecting the two offices is over 70% for all codecs analyzed, causing high delay for all other network services using this link. Furthermore, the possibility of infrastructure growth is very limited under these circumstances. It can be concluded, that under the present conditions, the company's infrastructure is not sufficient to support the data and voice traffic while offering an acceptable QoS.

For reference purposes, real conversations corresponding to a G.711 simulation and to an iLBC-30ms simulation can be found in [99]. When listening to these conversations, the difference in terms of speech quality is evident.

For the current simulation, ten replications were run. For the MOS computation, it was found that with 90% confidence can be said that, for the case of G.729 codec, the MOS is confined in $MOS \pm 0.12$. All other cases presented narrower error margins. For the link utilization statistics, the G.711 simulations presented zero variance. All values of link utilization obtained were 100% due to link saturation. For all other three codecs, the wider error margin is observed in the case AMR-NB, the link utilization is bound by link utilization $\pm 0.17\%$, with 90% confidence.

In the next section, two modifications with the purpose of improving the speech quality and link congestion will be presented and discussed.

4.3.3 Deploying VoIP across several branch offices: Solutions to low MOS and high network load problems

In the previous section, simulations to determine the performance of the company's network were presented and results analyzed. From the conclusions offered, it became clear that the quality of speech experienced in conversations between the two offices was below the accepted level. In addition, high link occupancy for the DS1 link connecting the two branch offices was detected, causing considerable delays to the data traffic and almost no possibilities of future expansion.

In this section, modifications to the network topology will be proposed and the impact on speech quality and link utilization will be analyzed.

For the network described in Section 4.3.2, the reason for poor speech quality was the high link utilization of the DS1 link. The traffic in this link has two components; first, the data traffic exchanged between the Toronto LAN computers and the servers located in the head office in Ottawa. The second component is the voice traffic between the employees working in Toronto and those based on the Ottawa office. Focusing on the data traffic on the complete network, there are also two components, the traffic exchanged between the Toronto computers and the servers and the traffic exchanged between the Ottawa LANs and the servers. This last traffic component is local to the Ottawa network. Given that the Ottawa office is the head office and the first one that existed, the servers and the Information Technology (IT) infrastructure necessary to support them are located in Ottawa. Considering that the computers in the Ottawa LAN are less in number than those in Toronto, a benefit can be seen in the network if some of the Toronto data traffic can be kept local (at the expenses to send some of the Ottawa data traffic over the inter-office link).

The proposed modification is focused on balancing the load of the data traffic between the two offices. To that end, one server, the database server, will be moved from the Ottawa office to the Toronto office. With that adjustment, the database traffic from the Toronto LAN will remain local to the Toronto subnet and the database traffic from the Ottawa LAN will be transmitted over the Ottawa-Toronto link.

Figure 4.17 shows the new topology, the only difference with the previous network topology shown in Figure 4.15 is the position of the data base server.

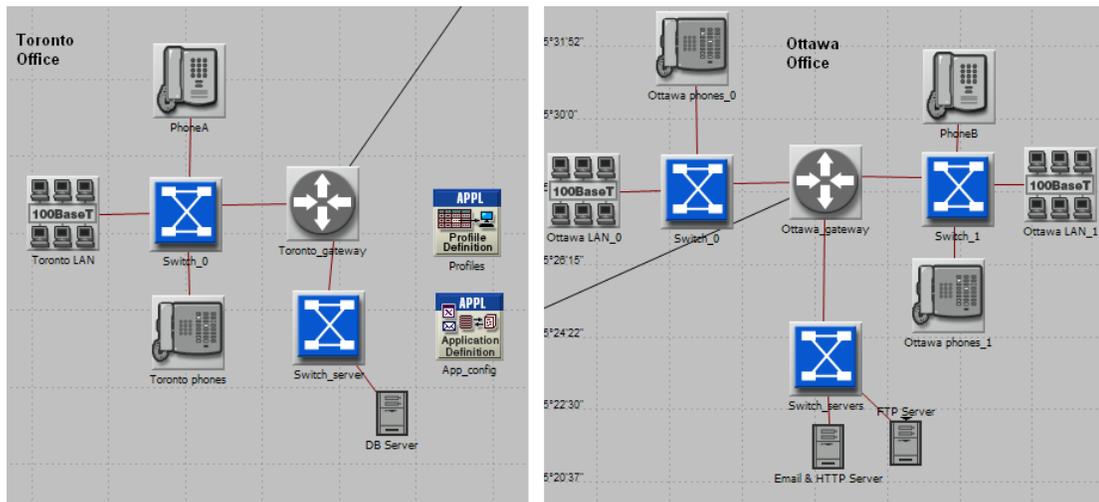


Figure 4.17 Network topology. Left: Toronto office with the DB server. Right: Ottawa office

Data Traffic

Identical to the traffic described in scenario described in Section 4.3.2.

Voice Traffic

Identical to the traffic described in scenario described in Section 4.3.2.

The results of the simulations for G.711, G.729, AMR-NB and iLBC-30ms are illustrated in Figure 4.18. Similarly to scenario discussed in Section 4.3.2, the MOS for G.711 codec is the worst of all codecs simulated. The reason for this behaviour remains the large bandwidth requirement of this codec and the impact on the link occupancy for the link connecting the two offices. The goal of the modification performed in this scenario is to ease the occupancy of the DS1 link, which is not achieved for the case of G.711 codec. In Figure 4.18, two codecs present considerably similar performance, for AMR-NB the statistics measured were MOS = 2.99 and link utilization = 56.95% and for the iLBC-30ms, MOS = 3.03 and link utilization = 56.54%. The difference between the two codecs in terms of MOS and link utilization is minimal, to determine which codec is a better option is near to impossible based on the data collected. Finally, the measured MOS value for G.729 codec was 2.70 and the link utilization was 60.11% in this case.

It can be concluded that overall, an improvement in the link utilization was observed for all codecs but G.711. This improvement has produced better MOS values for all codecs in addition

to increasing the possibilities of future expansion when compared to the previous scenario (see Section 4.3.2). The best-choice codecs for this simulation are indeed iLBC-30ms and AMR-NB, with almost identical MOS and link utilization.

The cost involved in moving a server from the Ottawa head office to the Toronto office needs to be considered. Beyond the initial cost in equipment and configuration, the management cost of the server needs also to be considered.

Furthermore, a similar approach to that discussed in this section can be extended to all servers in Ottawa. Once again, taking into account that probably the IT team would need to follow the servers to their new location and the expenses that this change implies.

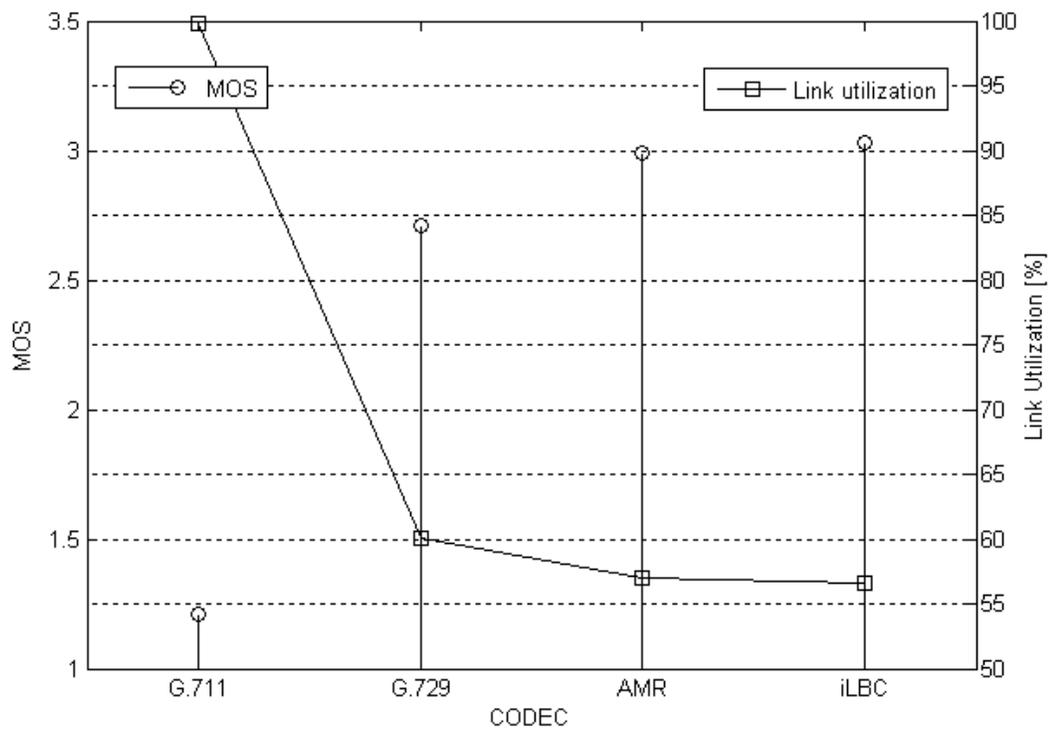


Figure 4.18 MOS (left vertical axis) and Link utilization (right vertical axis) for G.711, G.729, AMR and iLBC-30ms codecs

For reference purposes, real conversations corresponding to an AMR-NB simulation and an iLBC-30ms simulation can be found in [99]. It can be noticed that the quality of the two conversations is very similar, corresponding with the data collected and shown in Figure 4.18.

For the current simulation, ten replications were run. For the MOS computation, it was found that with 90% confidence can be said that, for the case of G.711 codec, the MOS is confined in $MOS \pm 0.11$. All other cases presented narrower error margins. For the link utilization statistic, the wider error margin is observed in the case G.729, the link utilization is bound by link utilization $\pm 0.68\%$, with 90% confidence.

Comparing the result of this scenario with the scenario in Section 4.3.2, an improvement in the link utilization on the DS1 link of 15% has been recorded. Also, in terms of speech quality, the MOS measured in the *PhoneB* node, for the calls originated in Toronto, increased from approximately 2.7 to 3. According to Table 2.6, for a MOS value equals to 3, the quality of the conversation is considered **Fair** [53]. The improvements in speech quality and link utilization are indisputable but modest. The speech quality barely reaches the **Fair** category and the link utilizations recorded over 55%.

In the next scenario, the proposed solution focuses in the increase of the capacity of the link connecting the two offices. The network topology for this scenario was described in Figure 4.14 and Figure 4.15. The only difference is that the capacity of the link connecting the Toronto office with the head office in Ottawa has been changed to two DS1 links (3.088 Mbps).

Data Traffic

Identical to the traffic described in scenario described in Section 4.3.2.

Voice Traffic

Identical to the traffic described in scenario described in Section 4.3.2.

The results of the simulations for G.711, G.729, AMR-NB and iLBC-30ms are illustrated in Figure 4.19. Similar to the results described in Section 4.3.1, the codec offering the best MOS is G.711. The value of MOS measured in the *PhoneB* node in Ottawa was 4.01 and the utilization for the link connecting the two offices was 57.64 %. The next codec to yield better result is AMR-NB with MOS = 3.79 and link utilization = 36.26 %. Finally, G.729 and iLBC-

30ms offer a similar results, with MOS = 3.55 and MOS = 3.47, respectively. Link utilization for G.729 and iLBC-30ms was recorded at 37.52 % and 35.55 %, respectively. Focusing the analysis in the two codecs yielding the highest MOS values, based on Table 2.6, G.711 is considered a **Good** speech quality and AMR-NB is located between **Fair** and **Good** category but almost in the **Good** zone. Based only on the MOS results, G.711 seems the best option of codec for this scenario; however, the MOS difference compared to AMR-NB is of only 0.22 when the link utilization difference is approximately 21 % on favour of AMR-NB.

It can be concluded that a considerable improvement in link utilization was detected for all codecs. As a logic consequence, an improvement in the speech quality for all codecs is also evident. In the conditions of the current scenario, although any of the codecs would offer a high subjective quality perception, the best-choice codec is whether G.711 or AMR-NB. The selection of one of these two codecs would be a judgment of the network designer. The criteria presented below may be relevant to the decision process.

- Based on MOS, G.711 is the best option.
- Based on the overall network performance, AMR-NB offers lower delays for all traffic exchanged between the Toronto office and the Ottawa head office.
- Based on the possibilities for future network expansion, AMR-NB offers better potential since the link utilization is only 36 % when compared to 57 % for G.711.
- Based on operation cost, G.711 is supported by default by all VoIP equipment but AMR-NB requires an operation license.

For the current simulation, ten replications were run. For the MOS computation, it was found that with 90% confidence can be said that, for the case of G.711 codec, the MOS is confined in $MOS \pm 0.085$. All other cases presented narrower error margins. For the link utilization statistic, the wider error margin is observed in the case AMR-NB, the link utilization is bound by link utilization $\pm 0.23\%$, with 90% confidence.

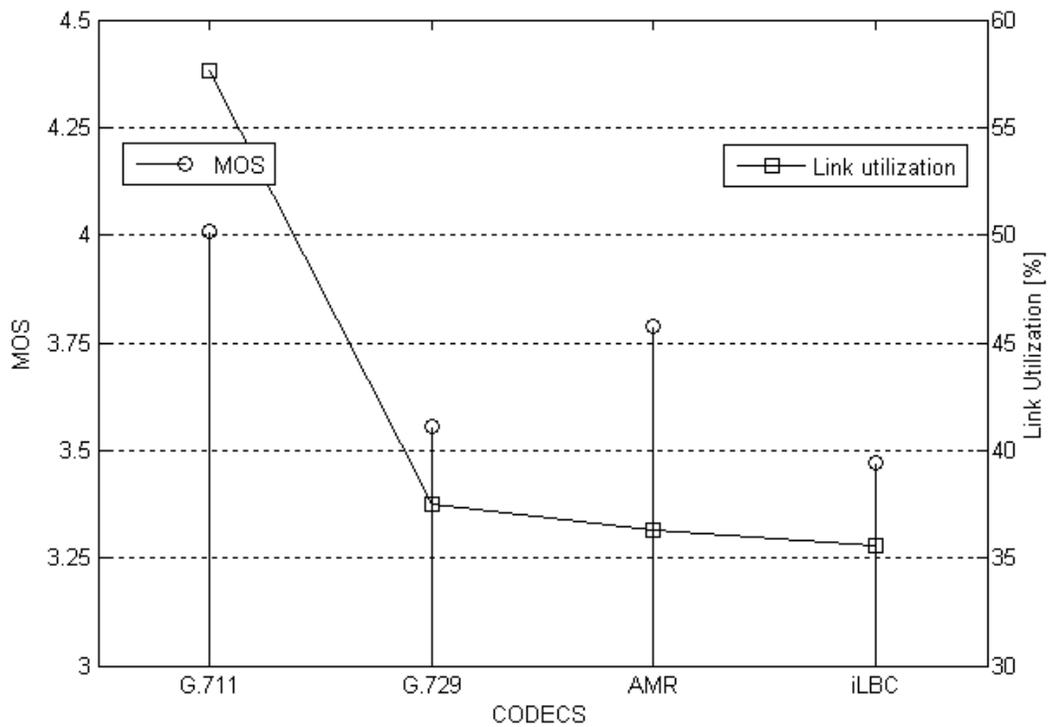


Figure 4.19 MOS (left vertical axis) and Link utilization (right vertical axis) for G.711, G.729, AMR and iLBC-30ms codecs

In Figure 4.20 a compilation of the simulation results from the three scenarios discussed in Sections 4.3.2 and 4.3.3 is presented. The top graph describes the behaviour in terms of MOS for the four codecs simulated and the bottom graph the link utilization for each codec in each scenario. This consolidated view of the simulation results allows for an easier codec comparison under the diverse link utilization conditions. In the legend of Figure 4.20, the label “Original Scenario” describes the scenario presented in Section 4.3.2. Similarly, the two scenarios discussed in this section are described by the “Load Balancing” and “Link 2xCapacity” labels.

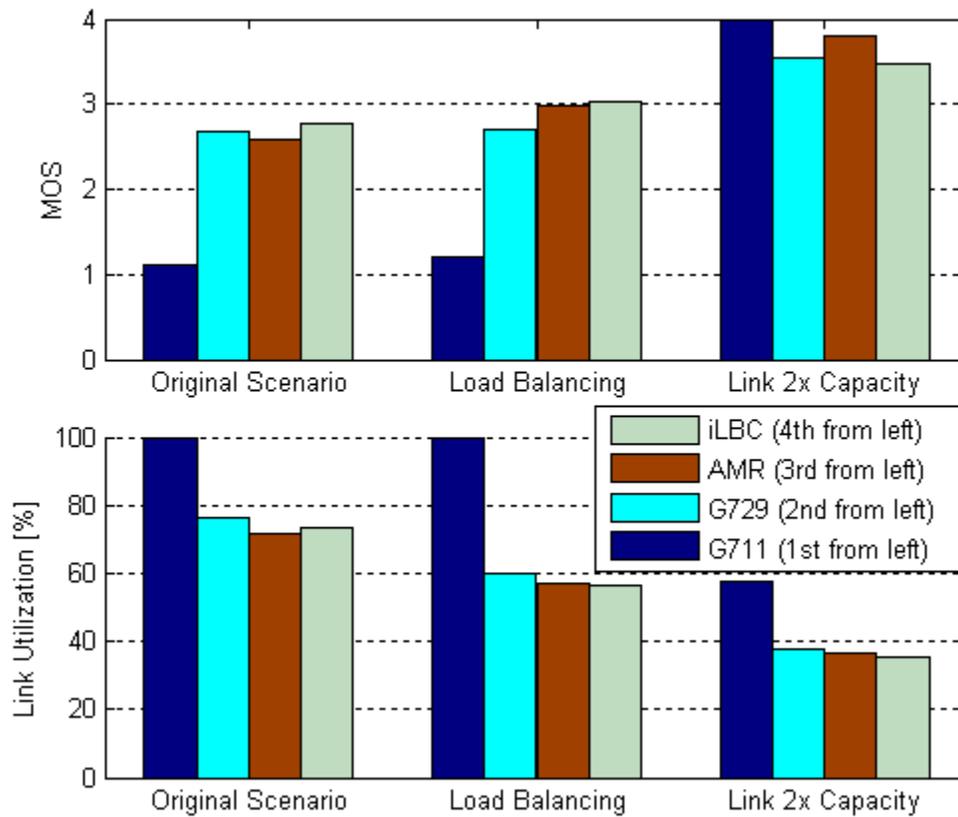


Figure 4.20 MOS (top) and Link utilization (bottom) for G.711, G.729, AMR and iLBC-30ms codecs for three simulation scenarios presented in Sections 4.3.2 and 4.3.3

4.3.4 Deploying VoIP across several branch offices: Performance analysis using real Internet traffic trace

This section presents a variation for the topology of the company's network discussed in Section 4.3.2. For all scenarios discussed in Sections 4.3.2 and 4.3.3, the offices are interconnected using Point-to-Point Protocol (PPP) links. In this section, an Internet connection will be used. The advantages of using the Internet over dedicated PPP links are mainly associated to reducing costs. Allowing interconnecting virtually any two places on the Globe, Internet access is affordable and significantly reliable. For the purpose of this experiment, the principal disadvantage related to the use of Internet over PPP links is associated to the variability and lack of predictability of the network delay and packet loss.

To model packet delay, delay variation and packet loss in an Internet connection is not easy task; current research is performed in that filed [95, 107]. To model the behaviour of the Internet in the scenario discussed in this section, a real Internet traffic trace will be used.

In Figure 4.21, the topology of the company's network is depicted. The intranet topology for both offices remains as illustrated in Figure 4.15.

Data Traffic

Identical to the traffic described in scenario described in Section 4.3.2.

Voice Traffic

Identical to the traffic described in scenario described in Section 4.3.2.

Internet modeling

In order to model the behaviour of the *Internet* node, 282,000 speech frames were sent from an asterisk server in the head office of CubaLlama [108] in Toronto to Carleton University during the time of 11AM-12PM. RTP packets and RTCP packets were captured using [109] and the E2E delay extracted. Figure 4.22 shows the PDF of the one-way E2E delay in the link.

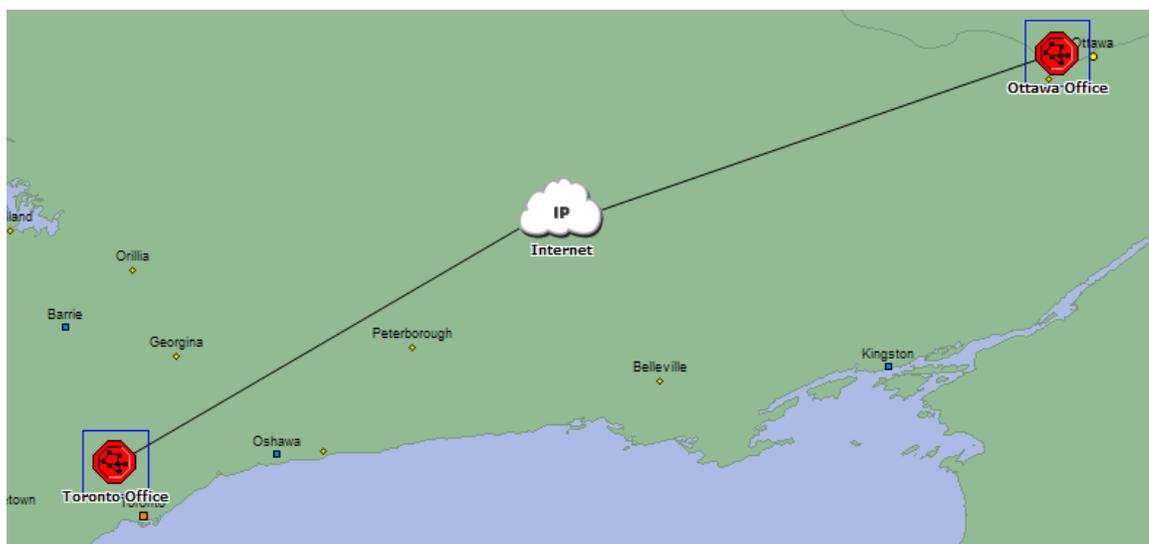


Figure 4.21 Network topology. Toronto office and Ottawa office interconnected through the Internet

The OPNET IP cloud node allows the utilization of a predefined PDF to model latency in the cloud. Also, it allows the employment of a user created PDF. Through the OPNET PDF Model editor, the marginal distribution shown in Figure 4.22 was integrated to OPNET and further utilized to model the delay in the IP cloud node. Since all the delays present in the link; i.e., transmission delay, queuing delay and processing delay, are contained in the values measured, high data rate links were selected to connect the two offices with the purpose to minimize the transmission delay introduced by the links. The links connecting the Toronto office to the *Internet* node and this last node to the Ottawa office are DS3 links (44.736 Mbps).

From Figure 4.22, it can be noticed that (1) the highest probable delays range between 50 ms and 60 ms and (2) that the delay spread is considerably small. From this path, low latency and low jitter will be measured. High values of MOS for all codecs are to be expected.

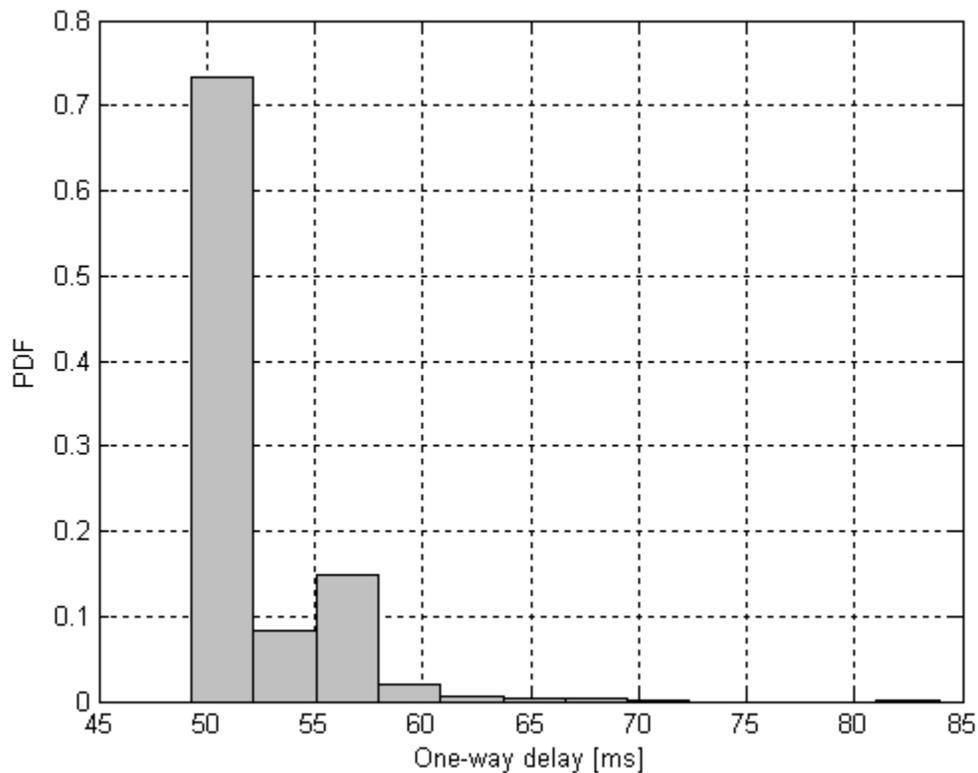


Figure 4.22 Probability Density Function of the E2E delay measured from a Toronto location to Carleton University

Figure 4.23 shows the speech quality values measured in *PhoneB* node. As predicted above, all values of MOS are in the **Good** and **Excellent** zone, according to Table 2.6. Based on the MOS values, the best-choice codec is G.711. The quality measured for G.729 codec (MOS = 3.52) is also in the **Good** area, it needs to be considered, that the bandwidth requirement for G.729 codec when operated in 2 frames per packet mode are 2.8 times lower than for G.711 codec (see Table 4.1).

For reference purposes, real conversations corresponding to all four codecs can be found in [99]. It can be noticed that the quality of all conversations is almost perfect. This result corroborates the data collected and shown in Figure 4.23.

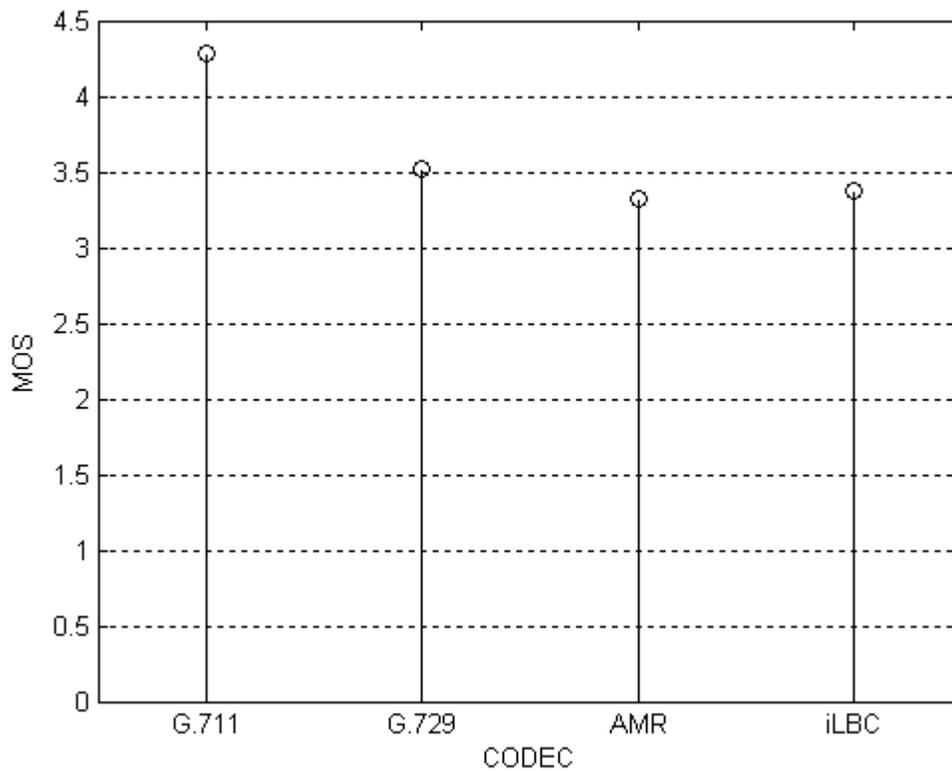


Figure 4.23 MOS for G.711, G.729, AMR-NB and iLBC-30ms codec

It can be concluded that, based on the delay data collected for the link between Toronto and Ottawa, the predicted speech quality is good for all codecs simulated. However, as already mentioned in this section, the factors involved in defining the delay, delay variation and packet loss in an Internet link are unpredictable or at least uncontrollable; making the Internet a transmission mean far from ideal for real-time traffic.

Chapter 5

Conclusions and Future Work

In this chapter the conclusions to the thesis will be presented. Additionally, suggestions for future expansion of the work described in this document will be provided.

5.1 Summary of thesis

In this thesis, a study of the most popular constant and variable data rate speech codec was performed. Also, a comprehensive survey of speech quality assessment models was presented with emphasis in the objective and non-intrusive algorithms. An evaluation of the state of the art in VoIP simulation for commercial products and research was carried out. The main purpose of this evaluation was to determine the needs in the field of VoIP simulation oriented to network planning and design.

It was found that, in regards to the codec availability for VoIP simulations, there are gaps particularly for variable bit rate algorithms. Also, it was determined that, in general, the importance of the generation of the relevant warm up traffic for the VoIP simulations and the time-dependant nature of the network impairments are overlooked. Finally, most products and research focus only in the objective speech assessment, not considering the subjective nature of the speech quality assessment process.

With the purpose of creating a VoIP platform that improves the aforementioned deficiencies, VoIP traffic generation models were created for G.711, G.729, iLBC, Speex and AMR-NB codecs. In addition, a model that allows the injection of real speech data to the simulation was created. Finally, all models were integrated to OPNET Modeler nodes along with the implementation of a non-intrusive objective speech quality assessment model.

5.2 Summary of results

The nodes and models mentioned in Section 5.1 have been included in the simulation of networks carrying VoIP traffic to test their efficacy.

First, all traffic generation models were tested and compared with theoretical bandwidth predictions or statistical models as applicable. The usefulness of the models implemented was tested as a tool to research the effects of playout buffer size in the speech quality. A simple buffer size optimization algorithm was outlined based on the simulation results. Ultimately, simulation of possible real-life scenarios was performed. From the simulation results, ideas to improve the speech quality and network performance were gathered, implemented and tested. Actual traffic behaviour was obtained from Internet traces collected from Toronto and Ottawa and integrated to the simulations.

5.3 Suggestions for future research

In order to complement the present study, the following additional work is suggested:

1. Include support for variable rate embedded codecs in the simulation. This implies that the nodes in the network are able to drop bits of the voice samples before the destination node is reached. Several advantages can be inferred from the use of embedded speech codecs.
2. Allow the playout buffer to be of variable (automatically adjusting) size. Implement one or more buffer size optimization algorithms.
3. Allow for some sort of network capacity planning, including projecting future traffic demands and automatically sizing links.
4. Integrate a network topology optimization algorithm to facilitate project growth.
5. Integration of Speex algorithm to the real speech traffic generator model

References

- [1] Mark A. Miller, Voice over IP Technologies: Building the Converged Network, 2nd Edition. John Wiley & Sons, Inc., 2002.
- [2] Mark A. Miller, White Paper: “Introduction to Converged Networking . A Technical Briefing Series on VoIP and Converged Networks”, DigiNet Corporation . Vol. 1, 2005.
- [3] M. Neuendorf, P. Gournay, M. Multrus, J. Lecomte, B. Bessette, R. Geiger, S. Bayer, G. Fuchs, J. Hilpert, N. Rettelbach, R. Salami, G. Schuller, R. Lefebvre, B.Grill, “Unified Speech and Audio Coding Scheme for High Quality at Low Bitrates”, IEEE International Conference on Acoustics, Speech and Signal Processing, page(s) 1-4 , 2009.
- [4] J. Škorupa, J. Slowack, S. Mys, P. Lambert, R. Van de Walle, C. Grecos, “Stopping Criteria for Turbo Coding in a WYNER-ZIV Video Codec”, IEEE Picture and Coding Symposium, page(s) 1-4, 2009.
- [5] OPNET Technologies Inc., <http://www.opnet.com/>, last accessed on May 20, 2010.
- [6] “Cisco Network Planning Solution 2.1 and Cisco Network Planning Solution - Service Provider 2.1. Data Sheet”, available at https://www.cisco.com/en/US/prod/collateral/netmgts/ps6341/ps6363/ps8794/prod_bulletin0900aecd80712089.html, 2007, last accessed on May 20, 2010.
- [7] CISCO Systems, Inc, <http://www.cisco.com/>, last accessed on May 20, 2010.
- [8] “NetIQ Vivinet Assessor. Data Sheet”, available at <http://www.netiq.com/products/va/>, 2008, last accessed on May 20, 2010.
- [9] NetIQ, <http://www.netiq.com/>, last accessed on May 20, 2010.

- [10] “NetIQ Vivinet Diagnostics. Data Sheet”, available at <http://www.netiq.com/products/va/>, 2008, last accessed on May 20, 2010.
- [11] Avaya, <http://www.avaya.com/>, last accessed on May 20, 2010.
- [12] Nortel, <http://www.nortel.com/>, last accessed on May 20, 2010.
- [13] International Telecommunication Union, “Pulse Code Modulation (PCM) of Voice Frequencies,” ITU-T Recommendation G. 711, 2003.
- [14] International Telecommunication Union, “Dual Rate Speech Coder for Multimedia Communication Transmitting at 5.3 and 6.3 kbit/s,” ITU-T Recommendation G.723.1, 1996.
- [15] International Telecommunication Union, “40, 32, 24, 16 kbps Adaptive Differential Pulse Code Modulation (ADPCM),” ITU-T Recommendation G.726, 1990.
- [16] International Telecommunication Union, “Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP),” ITU-T Recommendation G.729, 1996.
- [17] L. Sun, “Speech Quality Prediction for Voice over Internet Protocol Networks”, PhD Thesis, School of Computing, Communications and Electronics, University of Plymouth, 2004.
- [18] European Telecommunications Standards Institute, “Digital Cellular Telecommunications System (Phase 2+); Adaptive Multi-Rate (AMR) Speech Transcoding,” ETSI-EN- 301-704 V7.2.1, 2000.
- [19] JW. Choi, KH. Lee, “Implementation of a Network Simulator Supporting VoIP,” Proceedings of the International Conference on Information Networking, 2006.
- [20] L. Sun, E. C. Ifeachor, “Voice Quality Prediction Models and their Applications in VoIP Networks”, IEEE Transactions on Multimedia, Vol. 8, Issue: 4, page(s) 809-

820, 2006.

- [21] A. W. Rix, White Paper: “Comparison Between Subjective Listening Quality and P.862 PESQ Score”, Psytechnics Limited, 2003.
- [22] International Telecommunication Union, “The E-model, a Computational Model for Use in Transmission Planning”, ITU-T Recommendation G. 107, 2003.
- [23] G. Hotho, L. F. Villemoes, J. Breebaart, “A Backward-Compatible Multichannel Audio Codec”, IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, Issue: 1, page(s) 83-93, 2008.
- [24] S. Mehrotra, W. Chen, K Koishida, N. Thumpudi, “Hybrid Low Bitrate Audio Coding Using Adaptive Gain Shape Vector Quantization”, Proceedings of the 10th IEEE Workshop on Multimedia and Signal Processing, page(s) 927-932, 2008 .
- [25] J. Benesty, M. M. Sondhi, Y. Huang, Handbook of Speech Processing, Springer, 2008.
- [26] International Telecommunication Union, “Coding of speech at 8 kbps using conjugate structure algebraic-codec-excited linear-prediction”, ITU-T Recommendation G. 729, 1996 .
- [27] International Telecommunication Union, “Pulse Code Modulation (PCM) of Voice Frequencies”, ITU-T Recommendation G. 711, 1993.
- [28] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn, J. Linden, “Internet Low Bit Rate Codec (iLBC)”, 2004
- [29] International Telecommunications Union, <http://www.itu.int/>, last accessed on May 20, 2010.
- [30] International Telecommunication Union, “Pulse Code Modulation (PCM) of Voice Frequencies. Appendix I: A High Quality Low-Complexity Algorithm for Packet

Loss Concealment for G.711” ITU-T Recommendation G.711, 1999.

- [31] International Telecommunication Union, “Pulse Code Modulation (PCM) of Voice Frequencies. Appendix II: A Comfort Noise Payload Definition for ITU-T G.711 Use in Packet-Based Multimedia Communication Systems” ITU-T Recommendation G.711, 2000.
- [32] International Telecommunication Union, “5-, 4-, 3- and 2-bits Sample Embedded Adaptive Differential Pulse Code Modulation (ADPCM)” ITU-T Recommendation G.727, 1990.
- [33] M. Sherif, B. Duane, G. Bertocci, A. Bruce, M. Gonzalo, “Overview and Performance of CCITT/ANSI Embedded ADPCM Algorithms” IEEE Transaction on Communications, Vol.41, Issue: 2, page(s) 391-399, 1993.
- [34] 3GPP, “AMR Speech Codec; General Description”, TS 26.071, 1999.
- [35] I. Johansson, T. Frankkila, “Bandwidth efficient AMR Operation for VoIP” IEEE Proceedings of the Workshop on Speech Coding, page(s) 150-152, 2002.
- [36] 3GPP2, “Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems”, C.S0014-A, 2004.
- [37] J. M. Valin, “The Speex Codec Manual Version 1.2 Beta 3,” unpublished. Available: <http://www.speex.org/>, 2007, last accessed on May 20, 2010.
- [38] The Xiph Open Source Community, <http://www.speex.org/>, last accessed on May 20, 2010.
- [39] International Telecommunication Union, “Software Tools for Speech and Audio Codec Standardization” ITU-T Recommendation G.191, 2005.
- [40] Spiro lab Telecom, <http://www.sipro.com/>, last accessed on May 20, 2010.

- [41] VoiceAge Corporation, <http://www.voiceage.com/>, last accessed on May 20, 2010.
- [42] VoiceAge, “End User License Agreement for G.729(c) Implementation”, available at http://www.voiceage.com/openinit_g729_eula.php, last accessed on May 20, 2010.
- [43] Global IP Solutions, <http://www.gipscorp.com/>, last accessed on May 20, 2010.
- [44] “iLBC License Terms”, available at http://www.ilbcfreeware.org/documentation/gips_iLBClicense.pdf, 2004, last accessed on May 20, 2010.
- [45] International Telecommunication Union, “5-, 4-, 3- and 2-bits Sample Embedded Adaptive Differential Pulse Code Modulation (ADPCM)” ITU-T Recommendation G.727, 1990.
- [46] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson., “RTP: a Transport Protocol for Real-time Applications”, RFC 3550, IETF, 2003.
- [47] VoiceAge, “End User License Agreement for AMR Narrowband Implementation” available at http://www.voiceage.com/openinit_amr_eula.php, last accessed on May 20, 2010.
- [48] S. Keagy, Integrating Voice and Data Networks. Practical Solution for the World of Packetized Voice over Data Networks, 4th printing, CISCO Press, 2000.
- [49] M. Karjalainen, “A new Auditory Model for the Evaluation of Sound Quality of Audio Systems”, IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 10, page(s) 608-611, 1985.
- [50] T. A. Hall, “Objective Speech Quality Measures for Internet Telephony,” Proceedings of SPIE, Vol. 4522, page(s) 128-136, 2001.
- [51] A. Takahashi, H. Yoshino, N. Kitawaki, "Perceptual QoS Assessment Technologies for VoIP", IEEE Communications Magazine, Vol. 42, Issue: 7, page(s) 28- 34,

2004.

- [52] A. Takahashi, H. Yoshino, N. Kitawaki, “Objective Assessment Methodology for Estimating Conversational Quality in VoIP”, IEEE Transaction on Audio, Speech and Language Processing, Vol. 14, Issue: 6, page(s) 1984-1993, 2006.
- [53] International Telecommunication Union, “Methods for Subjective Determination of Transmission Quality” ITU-T Recommendation P.800, 1996.
- [54] International Telecommunication Union, “Modulated Noise Reference Unit (MNRU)” ITU-T Recommendation P.810, 1996.
- [55] W. D. Voiers, “Diagnostic Acceptability Measure for Speech Communication Systems”, IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, page(s) 204-207,1977.
- [56] J. G. Beerends, J.A. Stemerink, “A Perceptual Audio Quality Measure”, AES Convention, 1992.
- [57] International Telecommunication Union, “Objective Quality Measurement of Telephone-Band (300-3400Hz) Speech Codecs”, ITU-T Recommendation P.861, 1998.
- [58] International Telecommunication Union, “Artificial Voices”, ITU-T Recommendation P.50, 1999.
- [59] M. Goudarzi, “Evaluation of Voice Quality in 3G Mobile Networks”, MASC thesis, School of Computing, Communications and Electronics, University of Plymouth, 2008.
- [60] International Telecommunication Union, “Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs”, ITU-T Recommendation

P.862, 2001.

- [61] International Telecommunication Union, “Single-Ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications” ITU-T Recommendation P.563, 2005.
- [62] International Telecommunication Union, “Mean Opinion Score (MOS) Terminology” ITU-T Recommendation P.800.1, 2003.
- [63] Ch. Jin, R. Kubichek, “Vector Quantization Techniques for Output-Based Objective Speech-Quality”, IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, page(s) 491-494, 1996.
- [64] D. Picovici, A.E. Mahdi, “New Output-Based Perceptual Method for Predicting Subjective Quality of Speech”, IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, page(s) 633-636, 2004.
- [65] G. Chen, V. Parsa, “Nonintrusive Speech Quality Evaluation Using an Adaptive Neurofuzzy Inference System”, IEEE Signal Processing Letters, Vol. 12, Issue: 5, page(s) 403-406, 2005.
- [66] NetQoS, Inc., VoIP: Do You See What I Am Saying? Managing VoIP Quality of Experience on Your Network, E-Book, 2008, available at http://www.netqos.com/SEO_promo/ebook/, last accessed on May 20, 2010.
- [67] International Telecommunication Union, “Transmission Impairments Due to Speech Processing” ITU-T Recommendation G.113, 2007.
- [68] J.-M. Valin, “Speex: A Free Codec for Free Speech,” unpublished. Available: <http://www.speex.org>, last accessed on May 20, 2010.
- [69] J. Sjöberg, M. Westerlund, A. Lakaniemi, Q. Xie, “Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs,” RFC 3267, IETF,

2002.

- [70] J. Yang; Sh. Yu; M. Zhao, "The Implementation and Optimization of AMR Speech Codec on DSP," International Symposium on Intelligent Signal Processing and Communications Systems, page(s) 365-367, 2007.
- [71] J. W. Seo, S. J. Woo, K. S. Bae, "Study on the Application of an AMR Speech Codec to VoIP," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol.3 page(s) 1373 – 1376, 2001 .
- [72] W. Van Den Boeck, K. J. H. G. Venken, "Optimizing the Use of Network Resources for Data Transmission over an IP-Network," European Patent Specifications, 2009.
- [73] I. Johansson, T. Frankkila, P. Synnergren, "Bandwidth Efficient AMR Operation for VoIP," IEEE Proceedings of Workshop on Speech Coding, page(s) 150-152, 2002.
- [74] International Telecommunication Union, "Talker Echo and Its Control," ITU-T Recommendation G. 131, 2003.
- [75] R. J. Bates, D. W. Gregory, Voice & Data Communications Handbook, 3rd Edition, McGraw-Hill, 2000.
- [76] 3GPP, "Adaptive Multi-Rate (AMR) Speech Codec; Transcoding Functions", TS 26.090 V8.1.0, 2009.
- [77] CISCO Systems Inc., White paper, "Understanding Delay in Packet Voice Networks", CISCO, 2008.
- [78] J. Janssen, D. De Vleeschauwer, G. H. Petit, "Delay and Distortion Bounds for Packetized Voice Calls of Traditional PSTN Quality", Proceedings of the 1st IP-Telephony Workshop, page(s) 105-110, 2000.
- [79] H. Schulzrinne, S. Casner, "RTP Profile for Audio and Video Conferences with

- Minimal Control”, RFC 3551, IETF, 2003.
- [80] B. Melamed, J. R. Hil, “A Survey of TES Modeling Applications,” Sage, 1995.
- [81] B. Melamed, “An Overview of TES Processes and Modeling Methodologies,” Springer, 1993.
- [82] I. W. Hunter, R. E. Kearney, “Generation of Random Sequences with Jointly Specified Probability Density and Autocorrelation Functions,” *Biological Cybernetics Journal*, page(s) 141-146, 2004.
- [83] M. Bayazit, H. Aksoy, “Using Wavelets for Data Generation,” *Journal of Applied Statistics*, Vol. 28, page(s) 157-166, 2001.
- [84] S. Ma, Ch. Ji, “Modeling Heterogeneous Network Traffic in Wavelet Domain”, *IEEE/ACM Transactions on Networking*, Vol. 9, Issue: 5, page(s) 634-649, 2001.
- [85] A. Matrawy, I. Lambadaris, Ch. Huang, “MPEG4 Traffic Modeling Using the Transform Expand Sample Methodology,” *IEEE 4th International Workshop on Networked Appliances*, page(s) 249-256, 2002.
- [86] M. Ismail, I. Lambadaris, M. Devetsikiotis, A. R. Kaye, “Simulation and Modeling of Variable Bit Rate MPEG Video Transmission over ATM Networks”, *International Journal of Communication Systems*, Vol. 9, page(s) 283-297, 1996.
- [87] P. R. Jelenkovik, B. Melamed, “Algorithmic Modeling of TES Processes”, *IEEE Transactions on Automatic Control*, Vol. 40, Issue: 7, page(s) 1305-1312, 1995.
- [88] D. Geist, B. Melamed, “TEStool : An Environment for Visual Interactive Modeling of Autocorrelated Traffic,” *IEEE International Conference on Communications, SUPERCOMM/ICC*, Vol. 3, page(s) 1285 – 1289, 1992.
- [89] C. Bormann, L. Jonsson, “Robust Header Compression”, *53rd Internet Engineering*

Task Force, 2002.

- [90] L. Sun, E. C. Ifeachor, "Voice Quality Prediction Models and their Applications in VoIP Networks", IEEE Transactions on Multimedia, Vol. 8, Issue: 4, page(s) 809-820, 2006.
- [91] J. Banks, J. S. Carson, B. L. Nelson, D. M. Nicol, Discrete-Event Systems Simulations, 5th Edition, Prentice Hall, 2009.
- [92] J. F. Ransome, J. W. Rittinghouse , VoIP Security, Elsevier Digital Press, 2005.
- [93] C. Wu, K. Chen, C. Huang, C. Lei, "An Empirical Evaluation of VoIP Playout Buffer Dimensioning in Skype, Google Talk and MSN Messenger", Network and Operating System Support for Digital Audio and Video, 2009.
- [94] M. Narbutt, L. Murphy, "VoIP Playout Buffer Adjustment Using Adaptive Estimation of Network Delays" 18th International Teletraffic Congress (ITC-18), September 2003.
- [95] L Sun, E Ifeachor , "New Models for Perceived Voice Quality Prediction and their Applications in Playout Buffer Optimization for VoIP Networks", IEEE Proceedings Communication, Vol.3, page(s) 1478-1483, 2004.
- [96] R. G. Cole, J. Rosenbluth, "Voice over IP Performance Monitoring," ACM Computer Communication Review, Vol. 31, page(s) 9–24, 2001.
- [97] International Telecommunication Union, "One-Way Transmission Time," ITU-T Recommendation G.114, 2003.
- [98] L. Ding, R. A. Goubran, "Speech Quality Prediction in VoIP Using the Extended E-Model", IEEE GLOBECOM, Vol.7, page(s): 3974 – 3978, 2003.
- [99] OPNET models source code, supporting software and reference audio files,

- http://www.csit.carleton.ca/~msthilaire/voip_bkg_traffic/, last accessed on May 20, 2010.
- [100] Sound eXchange, available at <http://sox.sourceforge.net/>, last accessed on May 20, 2010.
- [101] Cygwin libraries, available at <http://www.cygwin.com/>, last accessed on May 20, 2010.
- [102] Video LAN-VLC Media Player, <http://www.videolan.org/>, last accessed on May 20, 2010.
- [103] IEEE Computer society, “Carrier Sense Multi Access with Collision Detect (CSMA/CD) Access Method and Physical Layer Specifications,” IEEE standard 802.3, Revision 2008.
- [104] Information Sciences Institute, “Internet Protocol DARPA Internet Program Protocol Specification,” RFC 791, IETF, 1981.
- [105] J. Postel, “User Datagram Protocol,” RFC 768, IETF, 1980,
- [106] T.C. Piliouras, *Network Design: Management and Technical Perspective*, 2nd Edition, Auerbach, 2005.
- [107] B. Zhang, T. S. Eugene Ng, A. Nandi, R. Riedi, Pe. Druschel, G. Wang, “Measurement Based Analysis, Modeling, and Synthesis of the Internet Delay Space”, *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, page(s): 85 – 98, 2006.
- [108] CubaLlama, <http://www.cuballama.com/>, last accessed on May 20, 2010.
- [109] Wireshark, available at <http://www.wireshark.org/>, last accessed on May 20, 2010.

Appendix A

New Nodes Added to OPNET

In Section 3, enhancements applied to OPNET simulation of VoIP traffic have been discussed. The current Annex will refer to the OPNET nodes where these new features have been implemented. Section A.1 will refer to the Background Traffic Generator node, Section A.2 will discuss the Real Voice node and finally Section A.3 the Client node.

The three nodes possess some attributes in common as well as the structure of the node. Figure A.1 shows the inner structure of each node. The functional differentiation of each node is implemented in the traffic generator model (traf_gen module). Subroutines of packet generation and packet destruction are the basic components of the traffic generator process model. The OPNET source code for the traffic generator models for each of the three nodes is presented in Annex B.

Common attributes of the three nodes

name: Each object in and OPNET simulation (node, link, etc) has to have a unique identifier. The name usually makes reference to the nodes function in the simulation but it is completely open and does not carry any importance in the outcome of the simulation.

IP Host parameters: This set of attributes allows assigning an IP address to the Traffic Generator node. Also, a subnet mask can be assigned as well as a Maximum Transmission Unit (MTU).

A.1 The voice Background Traffic Generator node

The Traffic Generator node can generate traffic that resembles G.729, G.711, iLBC 20ms, iLBC 30ms, AMR-NB codec and Speex codec (2 modes); specifications for these codecs have been discussed in Section 2.1 and earlier on this section. The configurable attributes for this node are illustrated in Figure A.2.

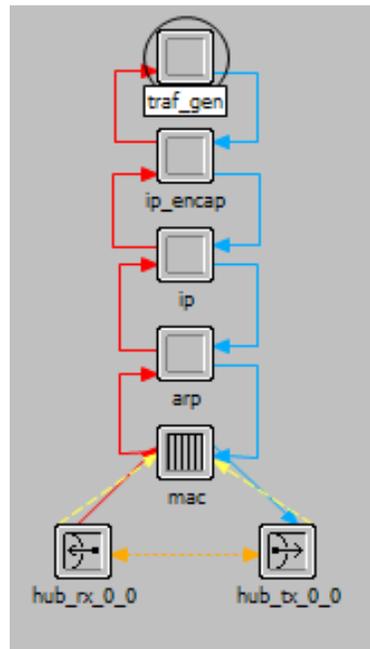


Figure A.1. OPNET structure of Background Traffic Generator node, Real Voice node and Client Node.

Attributes of Background Traffic Generator node

Traffic Generation Parameters: This attribute is a compound attribute, according to OPNET's definitions a compound attribute is formed by several sub-attributes. As it can be appreciated in Figure A.2, several rows can be defined; as many rows as desired can be defined. Each row corresponds with a VoIP stream going to a particular destination and with a specific speech codec. For the speech traffic generator node depicted in Figure A.2, four rows have been defined. Also, each row comprises eight attributes which describe the data stream or voice traffic. Details for each attribute follows.

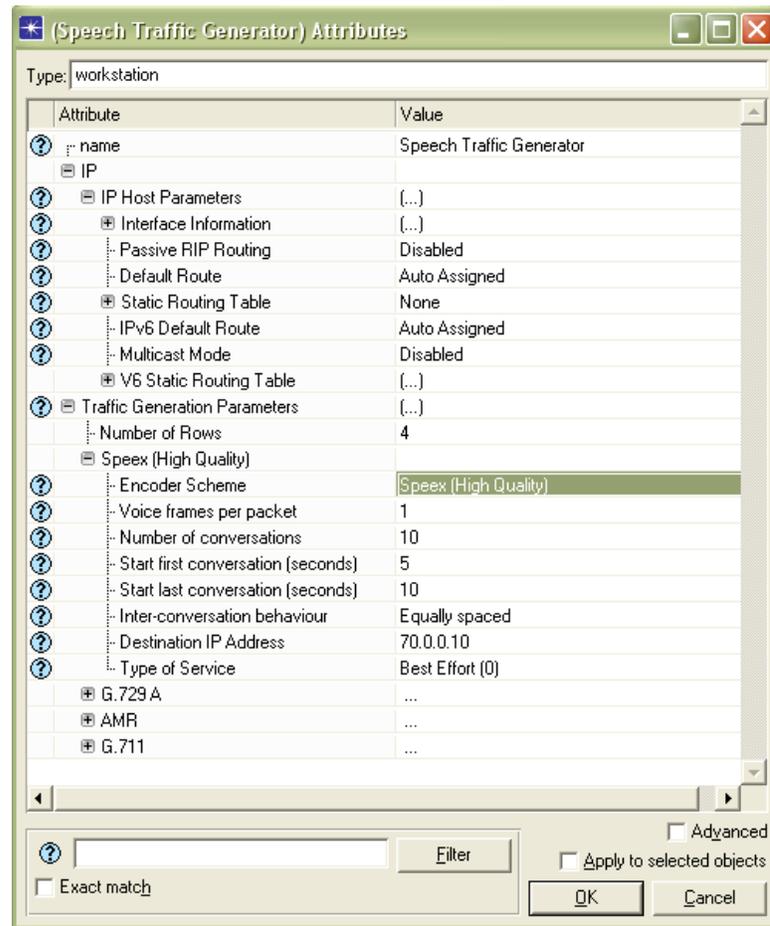


Figure A.2. Configurable attributes of the Traffic Generator node.

Encoder scheme: One of seven available VoIP encoding modes can be selected. The selection of the codec will define how the speech frame size and frame duration are going to be calculated during the simulation (see Section 3.1).

Voice frames per packet: Often in voice over IP communications more than one voice frame is encapsulated in a packet. This is done with the purpose of decreasing overhead of the IP packet. It should be noticed that as a consequence of increasing the number of voice frames per packet the overhead decreases and the total latency for the conversation increases. A common value used is two frames per packet and that is as well the default value of this attribute.

Number of conversations: The Traffic Generator node is designed to be versatile enough as to simulate one phone conversation coming from one IP phone client or more complex voice

traffic coming from a subnet, this subnet can be for example an office containing a number of IP phones from which a certain number of conversations as average are generated. The number of conversations parameter allows generating more than one conversation using the same codec and going to the same destination. Default value is one conversation.

Start first conversation, Start last conversation, Inter-conversation behaviour: The combination of these three parameters will help define when the different conversations will start, in case more than one conversation was defined in the previous attribute. The beginning of the conversations will be equally spaced or exponentially spaced according to the value assigned to Inter-conversation behaviour attribute. Default values are (5 seconds, 10 seconds, equally spaced respectively).

Destination IP address: IP address of the node where the data stream will be directed to. A client node has to be placed in the network and an IP address assign to it.

Type of Service: It represents an attribute which allows packets to be processed faster in IP queues. It is an integer between 0 - 252, 252 being the highest priority. Default value is Best Effort (0).

A.2 The Real speech traffic generator node

The Real Voice node is also a traffic generation node; the shape of the traffic being generated by this node follows G.729 (PLC), AMR-NB, iLBC 20ms, iLBC30ms, G.711 A-law (PLC and no-PLC), G.711 A-law (PLC and no-PLC) with one frame per packet. The configurable attributes for this node are illustrated in Figure A.3.

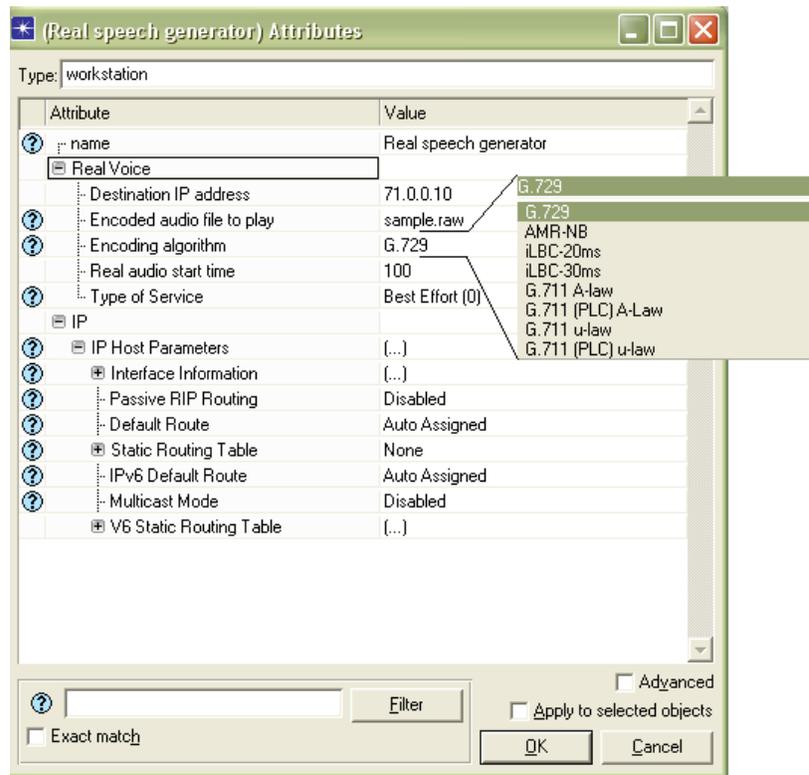


Figure A.3. Real Voice node attributes.

Despite of its name, the VoIP frames generated here do not carry real audio encoded information, dummy frames are sent instead to traverse the network. When studying a physical network is clear that link delay, probabilities of packet loss, position in queues, routing priority and other attributes that could define a packet circulating through a network, do not depend on the content of the packet payload. This said, a packet containing dummy information, for example all logical ones, will statistically behave the same as a packet carrying encoded audio information. From the simulation point of view though, the simulation time (and memory) required to process these two kinds of packets is not the same. To simulate real data takes considerable more resources than simulating dummy packets.

Attributes of Real Voice node

Real Voice: Attribute Group⁵ related to the generation and timing of real encoded audio.

Destination IP address: IP address of the node where the real encoded audio stream will be directed to. A client node has to be placed in the network and an IP address assign to it.

Encoded audio file to play: File name of the G.729 encoded data. The default location where the file should be stored is C:\VOICE\. The filename is not a fix value and the default value is *sample.raw*. Before the simulation starts, the number of encoded audio frames is extracted from the specified file. If the file is not found no data is generated and error is generated through the *Real_Voice_Simulation_report.txt* file also located in C:\VOICE\.

Encoding algorithm: One of eight available VoIP encoding modes can be selected. The selection of the codec will define how the speech frame size and frame duration are going to be calculated during the simulation as well as the number of frames to simulate (see Section 3.1.3).

Real Audio Start time: This is the time in seconds when the real voice stream will start. The relevance of this value was discussed earlier in this section. The duration of the real voice data stream can be expressed as:

$$\begin{aligned} \text{Real audio simulation}_{time} = \\ \text{MIN}(\text{number frames extracted from encoded file} * \\ \text{frame duration}, \text{Simulation end}_{time} - \text{Real audio start}_{time}) \end{aligned}$$

Type of Service: It represents an attribute which allows packets to be processed faster in IP queues. It is an integer between 0 - 252, 252 being the highest priority. Default value is Best Effort (0).

⁵ A Group is different than a Compound attribute. A Compound attributes can contain more than one row (or sets of the same attributes) when the Group is only an encapsulation for attributes with similar purpose.

A.3 The Client node

The Client node is the sink for the traffic generated by the Traffic Generator nodes and by the Real Voice node. Besides destroying the arriving packets, this node has three additional functions. The first one is to provide feedback to generator nodes for AMR-NB codec regarding the ETE delay for the specific path. The feedback is sent back using the RTCP protocol (see Section 3.1.3). The second functionality is to provide jitter compensation through the playout buffer for real voice traffic. The third feature is to collect the relevant statistics for the calculation of MOS based on the E-model as discussed in Section 2.2.2.2. Figure A.4 shows the attributes for the Client node.

Attributes of Client node

Playout Buffer Duration: Duration in milliseconds of the de-jitter buffer. The value of this attribute is not bounded. If the value entered is not a multiple of 10ms (frame duration) the value used in the simulation is adjusted to the next lower multiple of 10 ms of the value entered by the user.

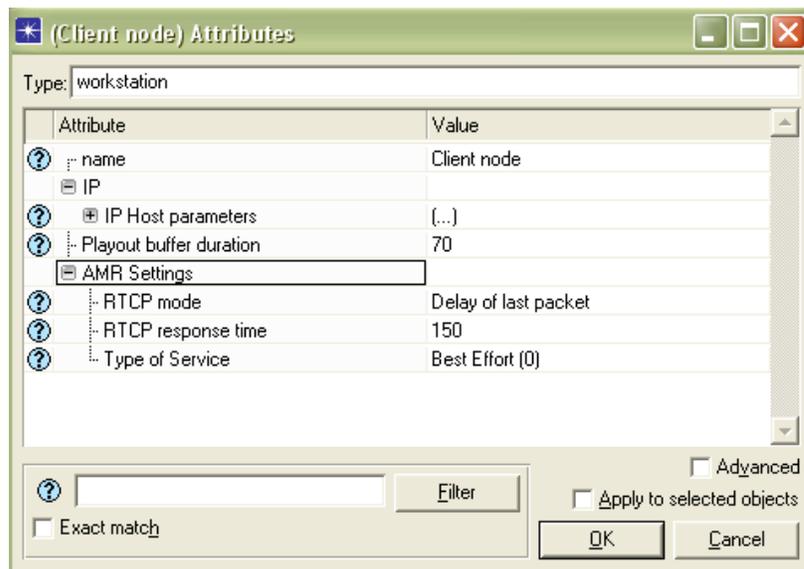


Figure A.4. Attributes of the Client node.

AMR Settings: Attribute Group related to AMR-NB bandwidth control mechanism.

RTCP mode: The two possible values of this attribute are: **Delay of the last packet** or **Average delay of the last group of packets**. When delay of the last packet is selected, the RTCP packet generated and sent back to the generator node will carry the information of the E2E delay of the last AMR-NB packet received before the RTCP packet was generated. When average of the last group of packets is selected, the RTCP packet sent back to the generator will carry the average of all E2E delays of AMR-NB packets received since the previous RTCP packet was sent.

RTCP response time: This parameter will define how many packets flagged as AMR to receive from a particular traffic generator before sending an RTCP packet back. This number is related to how often the bitrate control mechanism in the Generator node receives information about the network status. The smaller this number the more frequent the Client node will send a packet back to the generator causing the bitrate to be adjusted in a more efficient way but also creating an increasing in the network load. Default value is 150 packets; to put it in perspective, for AMR-NB (frame duration 20 ms), one speech frame per packet, a RTCP packet will be sent to the generator node approximately every 3 seconds. Note this attribute represents number of RTP packets, no number of frames.

Type of Service: It represents an attribute which allows packets to be processed faster in IP queues. It is an integer between 0 - 252, 252 being the highest priority. Default value is Best Effort (0).

Appendix B

Using the Background Traffic Generator node, the Real Speech Traffic Generator Node and the Client Node in an OPNET Simulation

The current Annex will illustrate how to construct an OPNET simulation using the three nodes described in Annex A. Instructions will be provided in a tutorial manner enriched by the insertion of images, screen shots and tips.

B.1 Placing the node models and process models in a valid OPNET model folder

The background traffic generator node, the real voice traffic generator node and the client node are new models added to the OPNET libraries. Before we are able to use these three nodes, the node models and the associated process models need to be placed in the correct folder. Once this task is performed, OPNET Modeler will be able to operate with the nodes and they can be added to a scenario.

The folders where OPNET Modeler searches for models are specified in a configuration variable named “Model Directories”, to access the value of this variable follow the simple steps listed below.

1. If Modeler is not already running, start it.
2. Select **Edit > Preferences**.
3. In the search bar on the top of the window type “Model Directories” and click **Search**.
The result of the query will be similar to those shown in Figure.
4. Click on the text shown in the red rectangle in Fig. B.1 (the red rectangle has been added by the author of the present document to highlight the correct option).
5. A list of all folders where models are potentially stored is shown. Take note of one of them or add a new one.

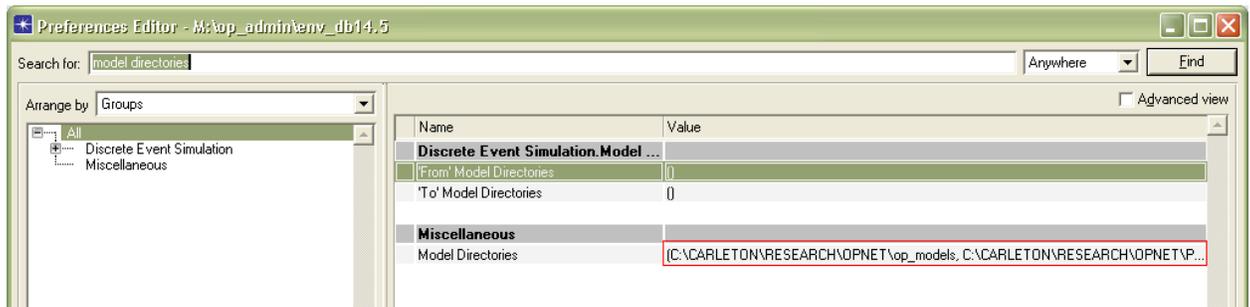


Figure B.1. Preference editor of OPNET Modeler showing the “Mode Directories” variable

Table B.1. Files to be added to one of OPNET Modeler model directories.

Node name	Associated files	File type
Background traffic generator	AHR_background_traffic_gen.nd	Node model
	AHR_application_layer_background_node.pr	Process model
Real voice traffic generator	AHR_Real_Voice_Traffic_gen.nd	Node model
	AHR_Application_layer_real_voice_node.pr	Process model
Client node	AHR_Client_node.nd	Node model
	AHR_application_layer_client_node.pr	Process model

6. Add all 6 files listed in Table B.1 to the folder noted or added in step 5.

B.2 Creating an OPNET project

When creating a new network model, you must first create a new **project** and **scenario**. A project is a group of related scenarios that each explores a different aspect of the network. Projects can contain multiple scenarios.

1. If Modeler is not already running, start it.
2. Select **File > New...**
3. Select **Project** from the pull-down menu and click **OK**.
4. Name the project and scenario, as follows:
 - 4.1 Name the project **Test_voice_sim**.
 - 4.2 Name the scenario **simple_scenario**.

4.3 Click **OK**.

The **Startup Wizard** opens.

Enter the values shown in the Table B.2 in the dialog boxes of the **Startup Wizard**. The workspace of the size you specified is created. The object palette for the project opens in a separate window.

Table B.2. Values to enter in the setup wizard

Dialog Box Name	Value
1. Initial Topology	Select the default value: Create empty scenario .
2. Choose Network Scale	Select Office . Select the Use metric units checkbox.
3. Specify Size	Select the default size: 100 m x 100 m
4. Select Technologies	Leave it as it is.
5. Review	Check values, then click Finish .

B.3 Adding the three nodes to the object palette in OPNET Modeler

The object palette in OPNET Modeler facilitates the selection of nodes, links and simulation objects in general for the creation of the simulation scenarios. In order to place the nodes in a simulation they need to be added first to the object palette. Follow the steps below.

1. When creating a project the object palette is opened automatically; if it is closed click on the icon  on the tool bar.
2. When the object palette is opened, click on the icon  located in the top left corner to change the view to icon view mode.
3. Once in icon mode, click on the **Configure Palette** button on the top right corner of the window.
4. A **Configure Palette** new window is shown, click on the Node Models button and select the three node names from the list by clicking on the “not included” text on the right of each node name. The text will change to show “included”. Click **OK**.
5. When all three nodes are added, save the object palette by clicking on **Save As** button. The object palette should now look similar to Figure B.2. Note the three nodes have been added.

B.4 Adding the background traffic generator node to the OPNET simulation

The next step is adding the three nodes and other complementary entities to the simulation. Let us start with the background traffic generator node. The steps to perform this task are listed below.

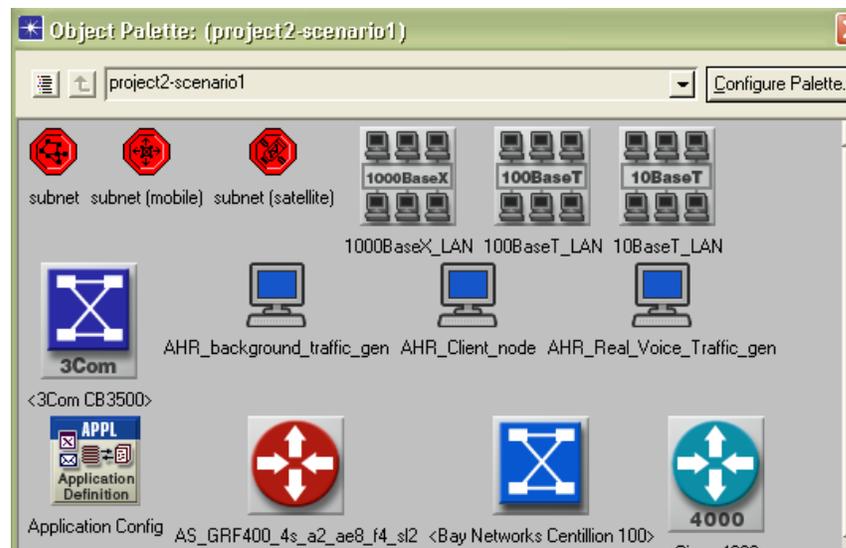


Figure B. 2. OPNET Modeler object palette showing the newly added nodes.

1. Drag the node from the object palette and drop it anywhere in the scenario workspace.
2. Right click on the node once it is placed on the workspace and select **Edit Attributes (Advance)**.
3. Change the field **name** to **Office 1 aggregated traffic gen**. Optionally, click on the **icon name** attribute and select the **svr_enterprise** icon.

B.5 Adding the real voice traffic generator node to the OPNET simulation

Let us add the real voice traffic generator node. The steps to perform this task are listed below.

1. Drag the node from the object palette and drop it anywhere in the scenario workspace.

2. Right click on the node once it is placed on the workspace and select **Edit Attributes (Advance)**.
3. Change the field **name** to **Office 1 Phone 1**. Optionally, click on the **icon name** attribute and select the **phone** icon.

B.6 Adding the client node to the OPNET simulation

Finally, let us add two client nodes to the simulation. The steps to perform this task are listed below.

1. Drag the one node from the object palette and drop it anywhere in the scenario workspace.
2. Right click on the node once it is placed on the workspace and select **Edit Attributes (Advance)**.
3. Change the field **name** to **Office 2 Phone1**.
4. Click on the **Office 2 Phone1** node and press **Ctrl+C** and **Ctrl+V**. Place the new node anywhere in the workspace.
5. Change the field **name** to **Office2_all_other_phs**.

At this point the workspace should look like shown in Figure B.3. If needed, nodes can be relocated in the workspace.

B.7 Adding other network nodes

In order to create a network that, although simple, reflects a relevant and reasonable scenario, other network nodes need to be added.

1. Open the object palette. Go to the very top and open **> Node Models > Fix Node Models > By Name > ethernet 2 ***.
2. Select the **ethernet2_slip8_gtwy** node model and place two nodes in the work space.
3. Change the field **name** of the two newly created nodes to **Gateway Office1 – Internet** and **Office2_all_other_phs**.

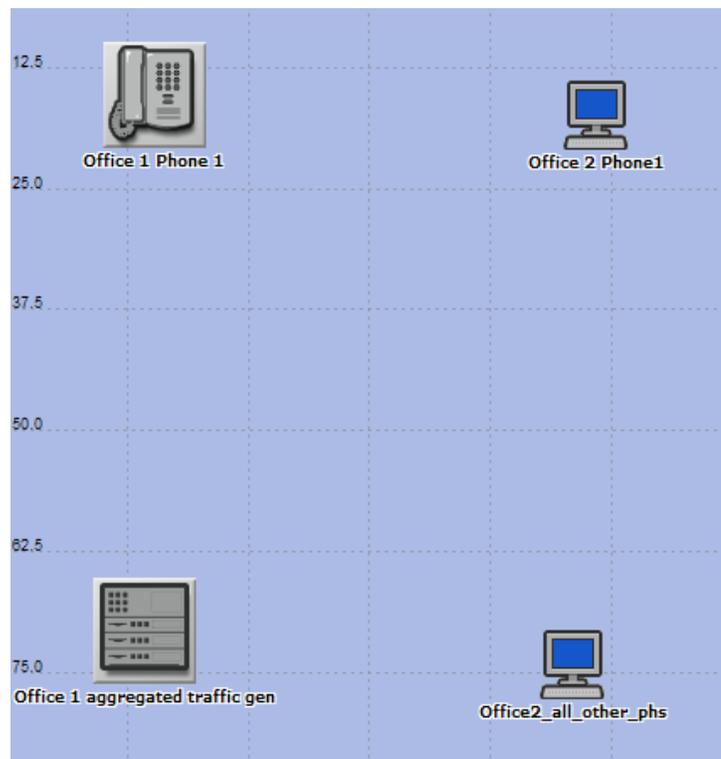


Figure B.3. Nodes placed on OPNET Modeler workspace.

4. Again in the object palette, open > **Node Models** > **Fix Node Models** > **By Name**> **ip8** *.
5. Select and place in the workspace one **ip8_cloud**.
6. Name the **ip8_cloud** node **Internet**.

B.8 Interconnecting all network elements

Finally, all network elements should be interconnected. To achieve this task, follow the steps described below.

1. Open the object palette. Go to the very top and open > **Link Models** > **Duplex Models** > **By Name**> **10BaseT** *.
2. Select the **10BaseT** link and connect nodes using information offered in Table B.3.
3. Again, from the object palette open > **Link Models** > **Duplex Models** > **By Name**> **PPP** *.
4. Select a **PPP_DS1** link and follow Table B.3 for interconnections.

When all the nodes have been interconnected, the network topology created should look similar to Figure B.4.

Table B.3. Node interconnection table describing link type.

Link description	Link type
Gateway Office1-Internet ↔ Office 1 Phone 1	Ethernet 10BaseT
Gateway Office1-Internet ↔ Office 1 aggregated traffic gen	Ethernet 10BaseT
Gateway Office1-Internet ↔ Internet	PPP DS1
Gateway Office 2 - Internet ↔ Office 2 Phone1	Ethernet 10BaseT
Gateway Office 2 - Internet ↔ Office2_all_other_phs	Ethernet 10BaseT
Gateway Office 2 - Internet ↔ Internet	PPP DS1

B.9 Understanding the topology

The network created represents a common setup for modern business models. The topology represents two offices located possibly hundreds of kilometers apart⁶. Between the two offices, data and phone conversations are exchanged on a regular basis. For the current example only voice traffic has been modeled. In this case, Figure B.4 is self explanatory, the three nodes in the left-hand side of the figure; represent Office 1, interconnected to the Internet by the **Gateway Office1 – Internet** node. Similarly, Office 2 is placed in the right-hand side of the figure and it is connected to the Internet through the **Gateway Office 2 – Internet** node.

⁶ Real geographical scale has not been considered in the current scenario for simplicity purposes.

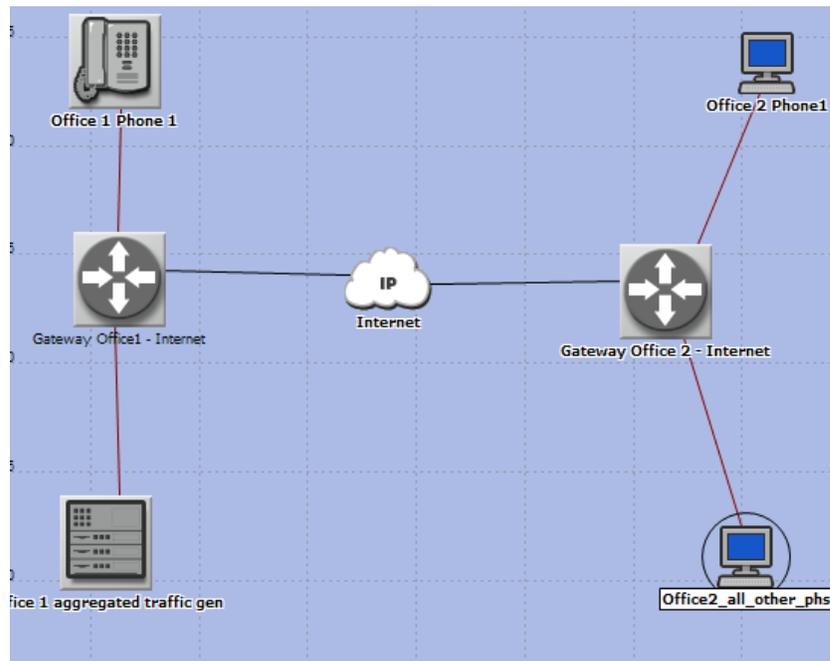


Figure B.4. OPNET Modeler network topology

For Office 1, and with the purpose to model the voice traffic, the office has been split in two nodes. The **Office 1 aggregated traffic gen** node will generate the equivalent average voice traffic generated by all minus one phones existing in Office 1. All this traffic will be sent to the **Office2_all_other_phs** node in

Office 2 that analogously represent the aggregation of all minus one phones in Office 2 receiving calls from phones in Office 1. The stream between these two nodes represents the background or warm up traffic for the simulation.

In order to test the quality of a call between Office 1 and Office 2, separate phones have been created in each office. A call will be placed from **Office 1 Phone 1** to **Office 2 Phone 1**. To make the simulation more realistic, a real audio file will be the content of the simulated call. This audio file will be then recovered by **Office 2 Phone 1** and played back to test the distortions introduced by the network transmission process. Additionally, **Office 2 Phone 1** will collect statistics from the incoming packets and offer a quality assessment for the call in question, this quality parameter will be MOS.

B.10 Setting up the simulation

After we understand the topology and the network model is complete, it's time to setup all the parameters for the simulation to run and statistics to be collected. Please, refer to Annex A for more detail information on each node's attributes.

To access the configuration parameters for each node, right click on the node icon and select **Edit Attributes**. Figures B.5 and B.6 show the settings that need to be modified for each node.

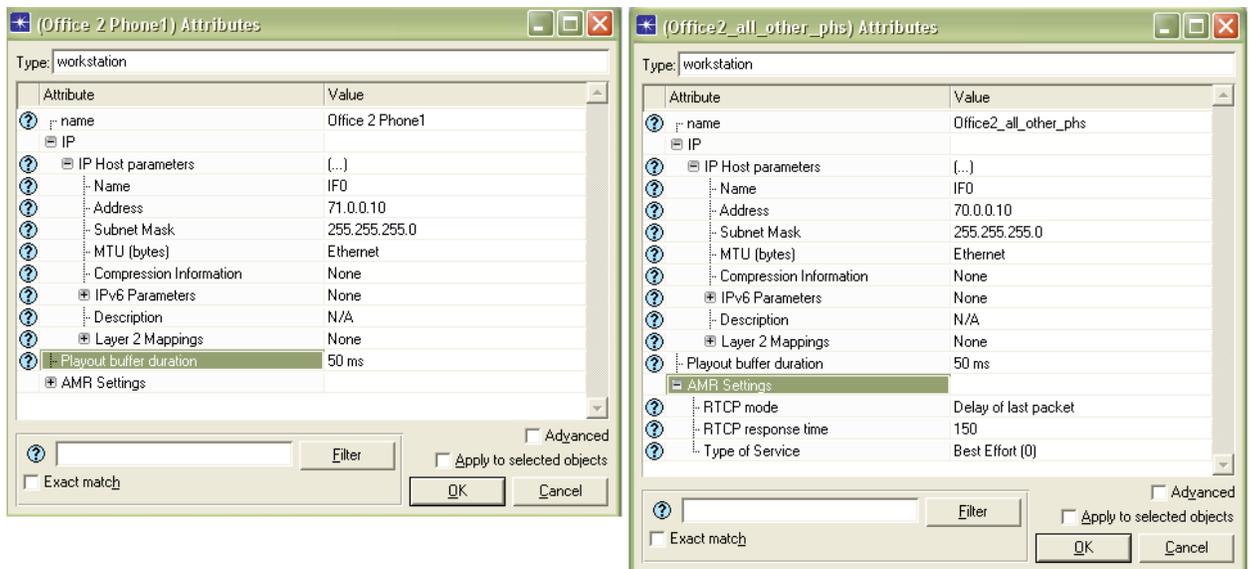


Figure B.5. Simulation settings for both client nodes (Office 2).

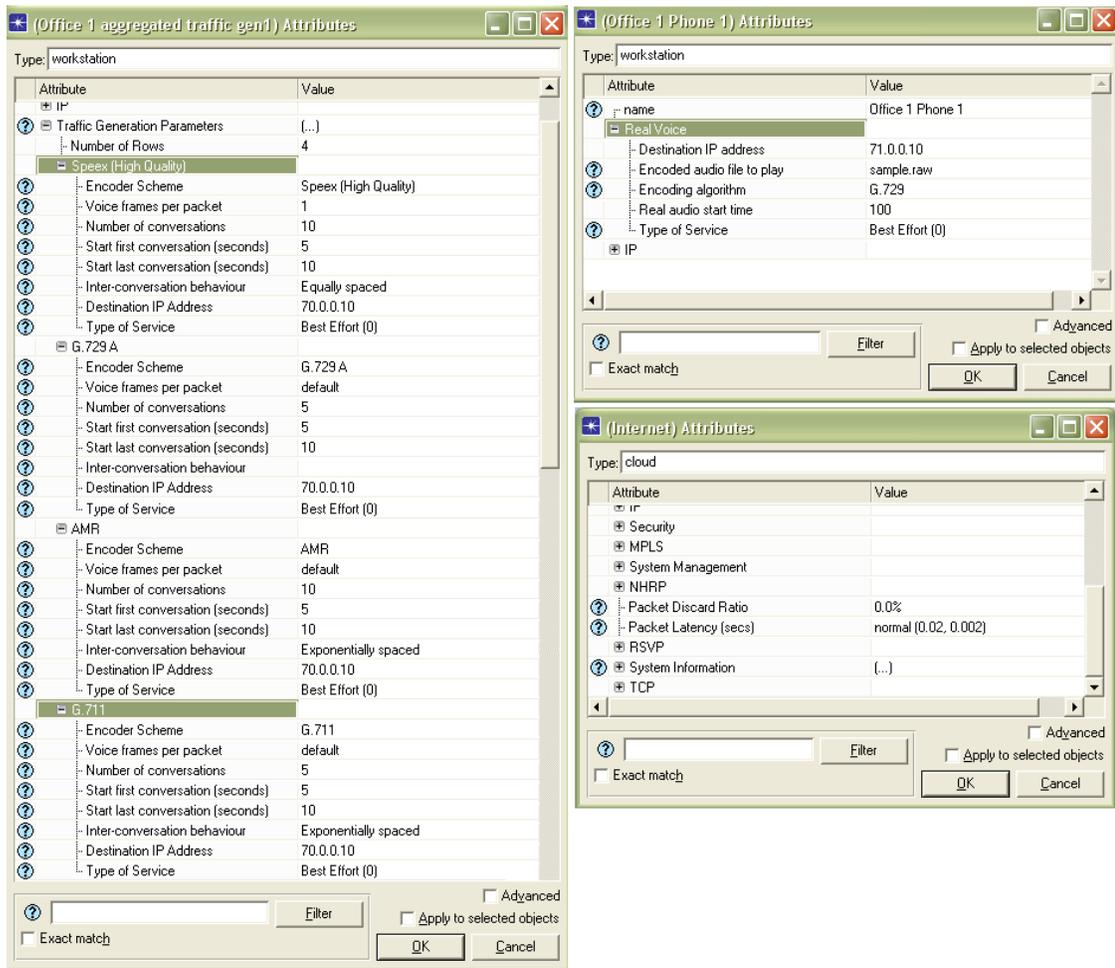


Figure B.6. Simulation settings for the background traffic generator node, the IP cloud node and the real speech traffic generator node (Office 1 and Internet).

A brief description of the simulation is as follows:

- 10 conversations encoded according to Speex quality 8 (1 frame per packet) are originated in the **Office 1 aggregated traffic gen** node with destination **Office2_all_other_phs** node (IP address 70.0.0.10).
- 5 conversations encoded according to G.729 (2 frame per packet) are originated in the **Office 1 aggregated traffic gen** node with destination **Office2_all_other_phs** node (IP address 70.0.0.10).

- 10 conversations encoded according to AMR-NB (2 frame per packet) are originated in the **Office 1 aggregated traffic gen** node with destination **Office2_all_other_phs** node (IP address 70.0.0.10).
- 5 conversations encoded according to G.711 (2 frame per packet) are originated in the **Office 1 aggregated traffic gen** node with destination **Office2_all_other_phs** node (IP address 70.0.0.10).
- One audio file (*sample.raw*) is generated from **Office 1 Phone 1** node with destination node **Office 2 Phone 1** (IP address 71.0.0.10). The file will be encoded according to G.729 algorithm at 1 frame per packet. As specified in 3.2.5, a file **sample.raw** has to exist in the location C:\Voice\.

B.11 Collecting data and running the simulation

According to the simulation objective, during the simulation statistics need to be collected and analyzed. Likely, relevant conclusions will be obtained from such analysis. Before the simulation commences, it is necessary to define the statistics to be collected.

For our simulation, the objective will be to assess the capability of our network to support the voice traffic described above and yield a speech quality between fair (MOS 3) and excellent (MOS 5), see Table 2.5. For that end, the following statistics will be collected.

- Mean Opinion Score collected in the **Office 2 Phone 1** node.
- Packet Loss ratio collected in the **Office 2 Phone 1** node.
- Average Jitter collected in the **Office 2 Phone 1** node.

In order to setup the collection of the statistics, first, right click on the node which statistics wants to be collected from (e.g. **Office 2 Phone 1** node). Select **Choose Individual DES Statistics** and select the three statistics as shown in Figure B.7.

To run the simulation click on the **DES** menu and then select **Run Discrete Event Simulation**. In the simulation setup window set the **Duration** time to 300 seconds and click the **Run** button. The simulation will start. When the simulation finishes, click the **Close** button.

After the simulation runs successfully, the statistics selected need to be visualized. Usually collected statistics are presented as a graph. To see the results of the simulation click on **DES> Results>View Results**. In the **Results Browser** window check all three statistics selected before (MOS, Average jitter and packet loss). On the right-hand side panel of the **Results Browser** window, the graphs of the three statistics will show. The graphs should look similar to Figure B.8.

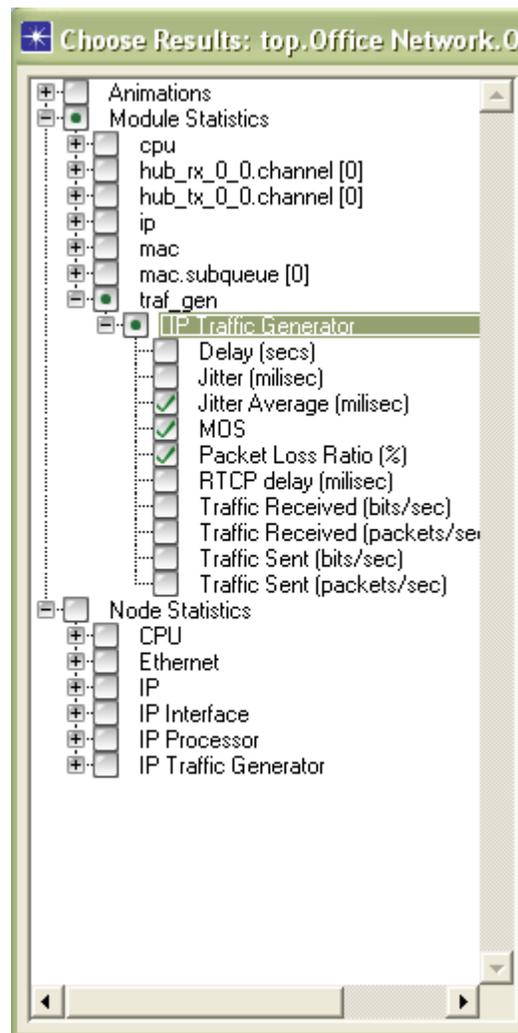


Figure B.7. Selecting statistics to be collected in the Office 2 Phone 1 node.

To help with the subjective speech assessment, a file with the name **G729_encode_decode_play.bat** will be created in C:\Voice\. By executing this file the following procedures are performed: (1) the obtained file is decoded using G.729 decoder and a traffic mask file (see Section 3.2.5), (2) the decoded audio file is played using VLC player (see Section 3.2.5) and (3) a report file is shown describing information linked to the simulation of the real speech file (i.e. simulated frames, received frames, packet loss ratio, destination node, etc).

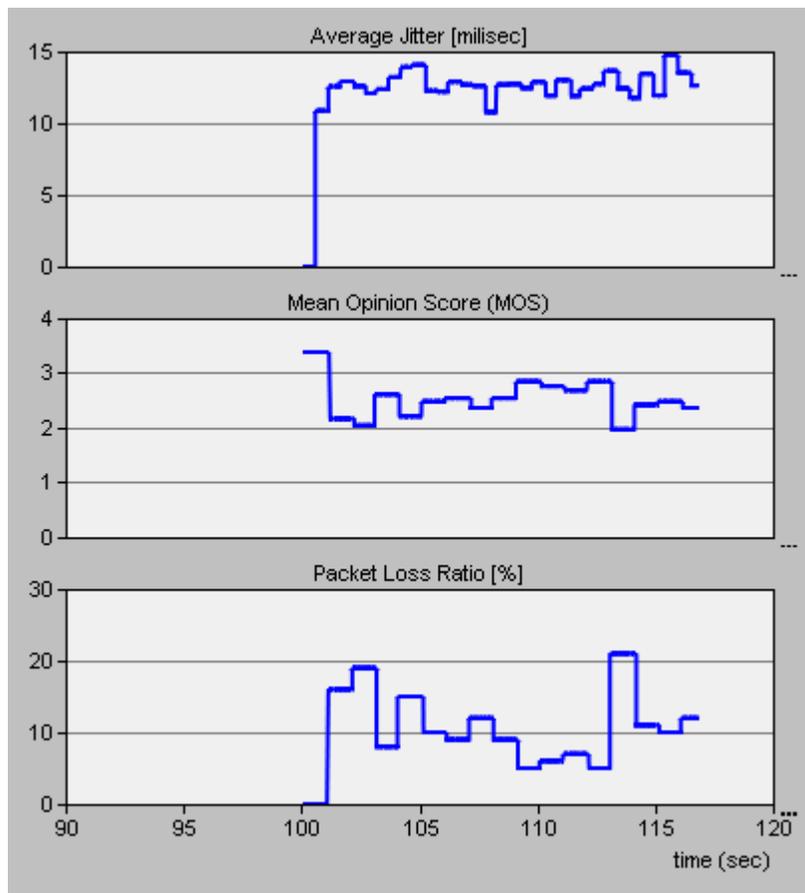


Figure B.8. Mean Opinion Score, Average Jitter and Packet Loss measured at the Office 2 Phone 1 node.

B.12 Analyzing the results of the simulation

However the objective of this Annex is not focused on the interpretation of the simulation results but in the steps and procedures of setting up a simulation instead, a brief analysis of the results can be performed. According to Table 2.5, it can be observed that for the whole duration of the conversation, the call quality oscillates between poor (MOS 2) and fair (MOS 3). It should be concluded then, that the simulated network, under the current load conditions and based on the simulation objective, cannot accommodate all the required voice traffic and yield the desired speech quality (MOS 3- MOS 5).

Furthermore, by inspecting the three collected statistics, a strong correlation can be observed; i.e. the MOS follows the fluctuations in the jitter and packet loss. This fact should help the modeller to define the reasons of the poor speech quality and design a strategy to improve it.