11TH INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2024
Serbia, Tara, 9-11 October 2024

PERFORMANCE CHARACTERIZATION OF SECURE IP COMMUNICATION SYSTEMS FOR VARIOUS INTERACTIVITY LEVEL APPLICATIONS

LARA KAŠCA
School of Electrical Engineering & Vlatacom Institute, Belgrade, lara.kasca@vlatacom.com

ĐORĐE NEŠKOVIĆ
School of Electrical Engineering & Vlatacom Institute, Belgrade, djordje.neskovic@vlatacom.com

MARKO MARKOVIĆ
School of Electrical Engineering & Vlatacom Institute, Belgrade, marko.markovic@vlatacom.com

STEFAN STANKOVIĆ
School of Electrical Engineering & Vlatacom Institute, Belgrade, stefan.stankovic@vlatacom.com

MIROSLAV PERIĆ
Vlatacom Institute, Belgrade, miroslav.peric@vlatacom.com

Abstract: *This paper describes a measurement methodology for characterizing the performance of secure communication systems, with a particular focus on the third layer of the OSI model (Layer 3). Special emphasis is placed on analyzing the impact of network characteristics such as delay and packet error rate on Quality of Service (QoS). The network environment is simulated using multiple routers and links, incorporating the mentioned network parameters that will be varied. We have presented impact on various types of interactive communications, such as VoIP audio data and video streaming. Special care is also taken on differences when using encryption method.*

Keywords: *Secure communication systems, Quality of Service, VoIP intelligibility, Video streaming, Network simulation*

1. INTRODUCTION

In battlefield scenarios, communication networks face numerous challenges that can impact link reliability. The reliability of the link, aside from the drop of the link itself, can be reflected in, for example, among the others, the packet delay and the number of error packets. These factors are critical for the performance and effectiveness of military operations and should therefore be addressed. Additionaly, encrypted communicationis are crucial in contexts where security and privacy are paramount, such as in military, governmental and corporate environments, for reasons like confidentiality, protection from eavesdropping, data integrity, handling of sensitive information etc.

In military contexts, the exchange of specifically audio and video data serves as a pivotal component for real-time communication, coordination and situational awareness. Encrypted radio transmissions facilitate rapid dissemination of commands and updates, crucial for prompt decision-making and synchronized actions. Meanwhile, live video feeds from aerial platforms such as cameras and drones provide surveillance, target identification and precise mission planning.

This study will focus on evaluating the audio and video quality, specifically voice data and video streaming, by manipulating parameters in a hybrid network using GNS3 software. The network will include multiple routers configured with OSPF protocol inside GNS3 and two Linux PCs on the end sides of the network. By adjusting parameters of interest and varying input signal characteristics, we aim to assess the resulting audio and video quality.

Quality of Service (QoS) in VoIP and video streaming directly impacts user experience by managing key parameters such as latency, jitter and packet loss, which are crucial for maintaining clear voice communication and high-quality video playback. QoS ensures that network resources are efficiently utilized and prioritizes traffic to minimize delays and packet loss, thereby enhancing overall service reliability and user satisfaction.

This paper will also demonstrate how various softwares and applications can be used for networking purposes and for assessing the quality of audio/speech and video. Specifically, this paper utilized GNS3, Wireshark, VLC, FFmpeg, qTox, and the Linux Ubuntu 22.04 operating system. In this paper, an improvement on work [1] is introduced by simulating the desired network environment without the use of certain physical devices.

Chapter two provides an overall network configuration description and outlines the differences for two test scenarios: one used for evaluating audio quality and the

Corresponding author: L.Kašca;
E-mail:lara.kasca@vlatacom.com
DOI: 10.5937/OTEH24071K

other for video streaming. Chapter three outlines the methodology for obtaining objective speech quality metrics, discusses relevant theoretical aspects and presents the resulting quantitative values, all adjusted for the serbian language. Analogously, chapter four details the setup for evaluating video streaming quality, followed by observational outcomes in form of subjective assessments and comments. Chapter five discusses the results and concludes the study.

2. NETWORK CONFIGURATION

As mentioned in introduction, a hybrid setup was used: GNS3 on Linux PC Eva (IP address: 192.168.2.13) on one side, in conjunction with a physical switch and another PC, Bob (IP address: 192.168.1.11), on the other side. The combination of the cloud and eno/eth interface enables 'exiting' the GNS3 environment as well as the host computer itself, providing connectivity with Bob's PC and thus creating a hybrid setup.

Routers R5 and R6 are connected to the networks of interest and the topology between them consists of four routers (R1-R4) connected in a square constellation with a cross-connection, simulating real life network, providing a backup path in case of a link drop. The topology of the entire network is shown in Figure 1 and the addresses of the routers’ interfaces are listed in table 1. Routers are configured with the OSPF protocol.

Eva, although the GNS3 was launched on it, was used as an end device, accessing the Eva’s terminal from the GNS3 environment using a later defined tap0 interface (192.168.2.25).

For each of the setups explained below, the same analysis was performed both with and without using IPsec, which was implemented within the GNS3 environment between routers R5 and R6. IPsec stands for Internet Protocol Security and is a secure network protocol suite that authenticates and encrypts packets of data to provide secure encrypted communication over IP. [2]

Table 1. Routers' interface names and addresses

Router	Interface	Interface address
1	f0/1	172.24.15.1
	s0/2	172.24.12.1
	s0/3	172.24.13.1
2	s0/1	172.24.12.2
	s0/3	172.24.23.1
	s0/4	172.24.24.1
3	s0/1	172.24.13.2
	s0/2	172.24.23.2
	s0/4	172.24.34.1
4	f0/1	172.24.46.1
	s0/2	172.24.24.2
	s0/3	172.24.34.2
5	f0/1	172.24.15.2
	f1/0	192.168.1.12
6	f0/1	172.24.46.2
	f2/0	192.168.2.22

2.1. Audio test setup

In the first part of the testing, when the focus was on audio data, communication was established between Bob and Eva using the qTox software. qTox is a completely free and open source, end-to-end encrypted messenger, focusing on user’s privacy and security. [3] It supports messages, audio and video calls. Therefore, qTox has the ability to facilitate voice over IP (VoIP), enabling users to make voice and video calls using internet protocols for transmission which matches this study focus. Namely, the focus of the QoS testing was on speech intelligibility, as speech, unlike other audio data, is crucial in military applications.

Here and later during testing, the connection establishment was verified using ping. From there on, in qTox Bob's system sound was selected as the input sound source instead of microphone, while on Eva, the output sound was recorded in .wav format using FFmpeg. FFmpeg is a free and open-source software comprising a suite of libraries and programs for handling video, audio and other multimedia, with its core being a command-line tool. [4] Following call establishment, audio files (whose content will be explained in the next chapter) were played in the background on Bob using a preinstalled media player, thus being sent over the network and recorded on the other side of the topology.

Within GNS3, various filters were applied, including packet error rate (abbreviated in later text as PER) and link delay in combination with jitter effect. These filters were applied to the links through which the traffic passed (which can be seen, represented by a funnel icon, in Figure 1). The specific values used will be listed in the corresponding chapters.

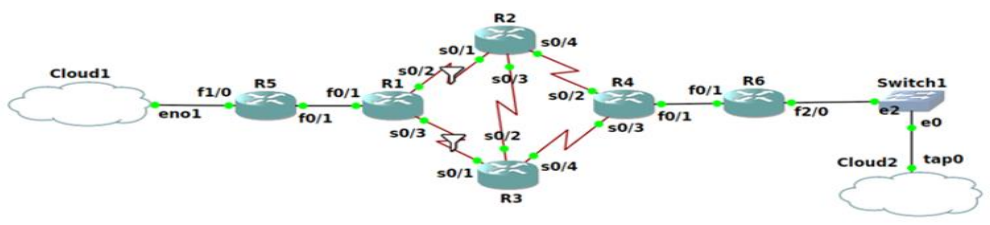

Figure 1. GNS3 network topology with interface labels and filtered links

2.2. Video streaming test setup

During testing of video stream behavior VLC media player ([5]) was used as a streaming platform, running as a server on Bob and as a client on Eva. VLC, widely known for playing movies, music and other media files, is also highly versatile and useful for various other purposes like streaming, allowing you to broadcast a media stream from one device or receive it on another via network or internet. UDP protocol and port 1234 were used. Similar to the VoIP testing, filters were applied to the links, with their specific values to be provided later. In Figure 2, it can be seen how video traffic appears in Wireshark before, during and after the establishment of IPsec. It can be noticed that the source and destination addresses, which initially show the real addresses of the end devices are changed afterward to the addresses of the routers between which the protocol is established.

3. VOIP AUDIO DATA

The quality of the audio signal regarding speech intelligibility in the various rooms/halls like classrooms, offices, concert halls, sports arenas etc. and on various channels (telecommunications systems, VoIP, etc.) can be determined using different objective measures, such as Speech Transmission Index (in further text: STI) factor, RASTI (stands for: rapid STI), STIPA etc. Methods of determining objective parameters are diverse and complex and they evolve over time. [6]

Before objective measures of speech intelligibility were developed, subjective methods were employed, requiring the participation of many individuals and extensive statistical analysis. In one widely cited publication, [7], it is shown that STI is correlated with subjective intelligibility scores from 167 different transmission channels obtained using pseudowords (also abbreviated as PB). Many research studies and papers have repeated and produced similar results. A graph from one of these ([8]) is shown in Figure 3. A literature review uncovers a wide array of subjective tests, primarily concerning the selection of test materials ([6]). Subjective tests are more time-consuming and obtained results can sometimes be difficult to compare with one another, e. g., due to the different participants, which however, in some situations, if chosen properly, can be an advantage because it introduces diversity of listeners and their auditory abilities ([9], [10]).

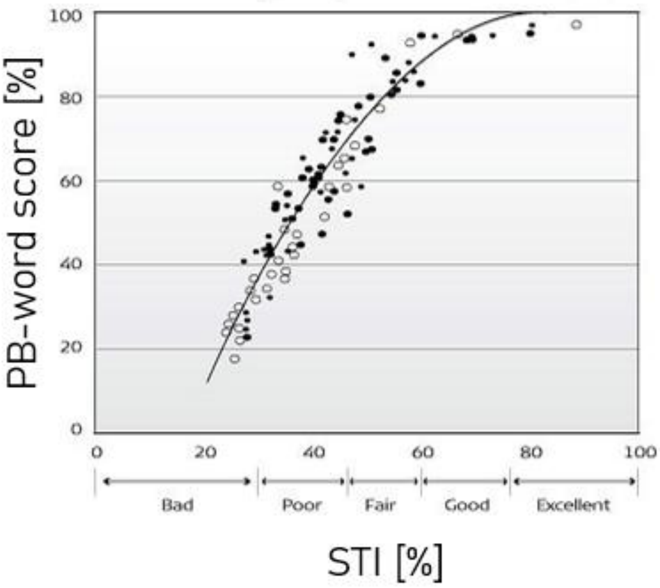

Figure 3. Relation between STI and PB-word score

One popular method (also used in [8], [9], [10], [11]) for assessing speech intelligibility involves using nonsense words or syllables in acoustic experiments known as logatomes or mentioned pseudowords. Logatomes are brief, meaningless words created by altering at least one letter in an existing word of specific language or combining various bigrams/trigrams. Speech contains multiple layers of information/redundancy that provide additional cues for understanding.

No.	Time	Source	Destination	Protocol	Length	Info
3349	22.807192	192.168.1.11	192.168.2.25	UDP	194	33445 → 33445 Len=152
3350	22.816598	192.168.2.25	192.168.1.11	UDP	274	33445 → 33445 Len=232
3351	22.817319	192.168.1.11	192.168.2.25	UDP	194	33445 → 33445 Len=152
3352	22.826844	192.168.2.25	192.168.1.11	UDP	274	33445 → 33445 Len=232

a)

7846	1804.212481	172.24.46.2	172.24.15.2	ISAKMP	150	Identity Protection (Main Mode)
7847	1804.253449	172.24.15.2	172.24.46.2	ISAKMP	346	Identity Protection (Main Mode)
7848	1804.273147	172.24.46.2	172.24.15.2	ISAKMP	346	Identity Protection (Main Mode)
7849	1804.314003	172.24.15.2	172.24.46.2	ISAKMP	150	Identity Protection (Main Mode)

b)

8025	2108.274415	172.24.15.2	172.24.46.2	ESP	166	ESP (SPI=0x673eaf9e)
8026	2108.284202	172.24.46.2	172.24.15.2	ESP	166	ESP (SPI=0xce272467)
8027	2108.284450	172.24.15.2	172.24.46.2	ESP	166	ESP (SPI=0x673eaf9e)
8028	2108.304535	172.24.46.2	172.24.15.2	ESP	166	ESP (SPI=0xce272467)

c)

Figure 2. Wireshark capture of video traffic before (a), during (b) and after (c) IPsec establishment

The redundancies that will be explained here are: acoustic, phonetic and lexical. Acoustic redundancy helps listeners through sound characteristics; phonetic is achieved by reinforcing phonemes and lexical redundancy through contextual clues within the language. These forms of redundancy help ensure effective communication, even when the sound transmission medium, which affects the speech signal itself, is not ideal, whether it is a communication channel like VoIP or room acoustics.

Logatomes are convenient for measuring speech intelligibility because they lack contextual information found in real-life speech, enabling researchers to focus solely on phonetic processing and recognition without the influence of external factors. Logatomes differ across languages as they adhere to each language's phonotactic rules. Some examples of logatomes in Serbian language are: mumi, fase, rare, džuđi, šizu… Due to gender-specific vocal characteristics, two distinct versions of logatomes are recorded—one tailored for female voices and another for male voices—reflecting the nuanced differences in speech patterns and phonetic traits between genders.

Experiments were conducted by playing on Bob’s side five audio signals with different logatome sets consisting of fifty logatomes spoken by a female and fifty spoken by male and recording them using FFmpeg on Eva. Each set is phonetically and structurally balanced.

The network parameters that were varied during different audio playbacks, excluding the original one, were: 1) delay and jitter together, which were 10ms and 2ms, respectively; 2) PER, which was 15%; 3) PER which was 25% and 4) a combination of delay+jitter/PER of 10±2ms/20%. The entire explained procedure was repeated without and with the use of IPsec.

Collected recordings were reproduced offline to listeners, on headphones and in quiet surroundings, who noted what they heard. By comparing with the correct responses, the quantity of correctly interpreted words for each scenario was obtained. Multiplying by 0.02 (or 2%) yields the value of the STI factor, which ranges from 0 to 1 (or 0% - 100%), determined for both female and male voice and graphically shown, expressed as a percentage, on Figure 4.

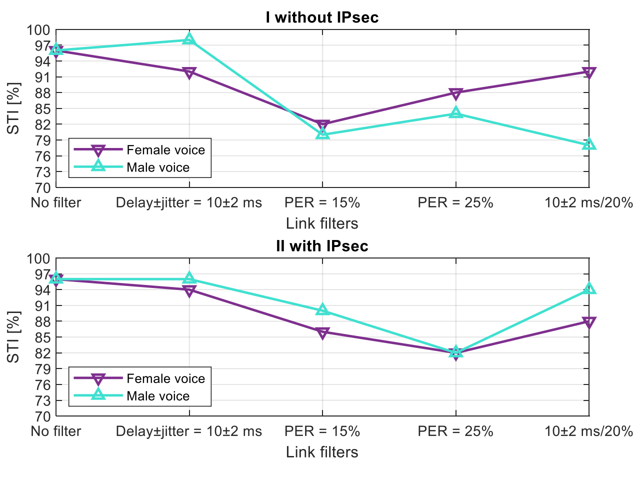

Figure 4. Comparison of female and male voice intelligibility without and with IPsec

For native language speakers, the relationship between STI and speech quality can be expressed using the following scale ([12]) with five bands:

- STI ∈ {0, 0.3} — bad quality,
- STI ∈ {0.3, 0.45} — poor quality,
- STI ∈ {0.45, 0.6} — fair quality,
- STI ∈ {0.6, 0.75} — good quality and
- STI ∈ {0.75, 1} — excellent quality.

Another measure, derived from STI, is CIS (stands for Common Intelligibility Scale) based on a mathematical relationship with STI (CIS = 1 + log(STI)). [13] [14]

4. VIDEO STREAMING

For testing quality of a video stream, Big Buck Bunny short movie was used. It is an animated film, made using Blender by Blender Institute. The video is freely available in multiple resolutions, including high ones like HD and 4K, making it suitable for testing different display qualities and resolutions. It is a widely recognized and standard piece of media. Different resolutions were downloaded from official source ([15]), in order from smallest to largest: 640x360 (360p), 853x480 (480p) and 1280x720 (720p). Tests, that is, streaming, were done, as mentioned earlier, using VLC Media Player.

The network parameters that were varied during different video playbacks, excluding the original one, were: 1) delay and jitter together, which were 10ms and 2ms, respectively; 2) PER, which was 2%; and 3) a combination of delay+jitter/PER of 5±1ms/1%.

By simply observing the video quality on the receiving end, comments were derived from and listed in tables 2 and 3. It should be noted that all comments refer to the frames after the stream establishment and stabilization, during which image freezing is more likely to occur.

Table 2. Comments on video quality across resolutions and applied filters when not using encryption

Without IPsec			
Filter	**360p**	**480p**	**720p**
Original	Watchable, like on Tx side	Watchable, like on Tx side	Unwatchable, much worse, picture freezing
Delay±jitter 10±2 ms	Watchable, but sometimes choppy	Watchable, but sometimes choppy	Unwatchable, block artifacts, smearing, lagging and losing frames
PER = 2%	Unwatchable, rare occurrence of continuous scenes	Similar to 360p, more pronounced effect	Similar to 480p, even more pronounced effect
Delay±jitter 5±1 ms PER = 1%	Similar to PER = 2%	Similar to 360p, but worse	Completely unwatchable and unusable

Table 3. Comments on video quality across resolutions and applied filters when using encryption

With IPsec			
Filter	**360p**	**480p**	**720p**
Original	Watchable, like on Tx side	Watchable, little video stuttering	Unwatchable, more frequent stuttering, block artifacts
Delay±jitter 10±2 ms	Watchable, like on Tx side	Watchable, little video stuttering	Unwatchable, more frequent stuttering
PER = 2%	Unwatchable, rare occurrence of continuous scenes, mostly freezes rather than plays	Similar to 360p, more pronounced effect	Completely unwatchable and unusable
Delay±jitter 5±1 ms PER = 1%	Unwatchable, frequent video stuttering	Similar to 360p, block artifacts and smearing	Completely unwatchable and unusable

5. CONCLUSION

The speech intelligibility results show that when no filters were applied, the number of correctly interpreted words was the same without and with IPsec, for both male and female speech. On average, adding additional latency did not affect intelligibility, but increasing the packet error rate substantially degraded it. The combination of latency and PER highlighted that PER had a more significant impact. Male speech was generally understood as well as or better than female speech and IPsec did not considerably affect intelligibility. Any inconsistencies in conclusions could be due to different sets of logatomes used. Intelligibility did not drop below 75% in any scenario, indicating overall good quality.

For video quality, increasing resolution made the video more susceptible to degradation from filters. Even lower resolutions became unwatchable with more aggressive filters. Effects such as freezing, lagging, choppiness, block artifacts and smearing occurred more frequently with increasing degradation through filters. Also, video content could tolerate much lower PER levels compared to speech. Freezing, lagging and choppiness were unavoidable with higher resolutions. The observations and the data table show that IPsec did not impact video quality degradation.

Overall, this study explored the reliability of communication networks by evaluating audio and video quality under various network conditions. Using GNS3 software, a hybrid network was configured with multiple routers and Linux PCs, with a focus on voice data traffic and video streaming. The results showed that network parameters like delay, jitter and packet error rate noticeably affect speech intelligibility and especially video quality. Implementing IPsec provided secure communication and did not influence much in other ways except maybe introducing additional latency. The research highlighted the importance of Quality of Service (QoS) in managing network resources to maintain clear communication and high-quality video playback. The tools and methodologies demonstrated in this study offer valuable insights for optimizing communication networks.

Acknowledgement

This paper was supported by Vlatacom Institute of High Technologies under the project P176.

References

[1] PETROVIĆ,R., SIMIĆ,D., STANKOVIĆ,S., PERIĆ,M.: *Educational Platform for Examining the Influence of the Simulated Satellite Link on Overall Communication inside of Different IoT Systems*, Proceedings of the 16th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), 2023

[2] https://en.wikipedia.org/wiki/IPsec

[3] https://www.linuxbabe.com/desktop-linux/install-qtox-messenger-on-linux

[4] https://en.wikipedia.org/wiki/FFmpeg

[5] https://www.videolan.org/vlc/download-ubuntu.html

[6] KUROWSKI,A., et al.: *A Novel Method for Intelligibility Assessment of Nonlinearly Processed Speech in Spaces Characterized by Long Reverberation Times*, Sensors, 2022, 22.4: 1641.

[7] STEENEKEN,H.,J.,M., HOUTGAST,T.: *A physical method for measuring speech-transmission quality*, The Journal of the Acoustical Society of America, 1980, 67(1), 318–326. doi:10.1121/1.384464

[8] https://www.troldtekt.com/product-advantages/good-acoustics/advanced-acoustics/speech-intelligibility-and-speech-intelligibility-goals/

[9] KRUGER,K., GOUGH,K., HILL,P.: *A comparison of subjective speech intelligibility tests in reverberant environments*, Canadian Acoustics, 1991, 19.4: 23-24.

[10] HODOSHIMA,N., ARAI,T.: *Effect of talker variability on speech perception by elderly people in reverberation*, Proceedings of the International Symposium on Auditory and Audiological Research. 2007. p. 383-388.

[11] BRACHMAŃSKI, S.: *Automation of the logatom intelligibility measurements in rooms*, Archives of Acoustics, 2014, 32.4 (S): 159-164.

[12] STEENEKEN,H.,J.,M.: *The measurement of speech intelligibility*, TNO Human Factors, Soesterberg, the Netherlands

[13] BARNETT,P.,W., KNIGHT,R.,D.: *The Common Intelligibility Scale*, 1995, Proc. I.O.A. Vol 17, part 7.

[14] BARNETT,P.,W.: *Overview of speech intelligibility*, 1999, Proc. I.O.A Vol 21 Part 5.

[15] https://peach.blender.org/download/