Quality of Service Analysis of VoIP Services

Abdullah Alhayajneh
School of Engineering and Computing Sciences
New York Institute of Technology
New York, USA
alhayajneh@nyit.edu

Alessandro N. Baccarini
Fordham Center for Cybersecurity
Fordham University
New York, USA
abaccarini@fordham.edu

Thaier Hayajneh
Fordham Center for Cybersecurity
Fordham University
New York, USA
thayajneh@fordham.edu

***Abstract*—Voice over Internet Protocol (VoIP) is a method of transmitting audio and video over a network. It is a low-cost and readily available alternative to traditional landline phones. In this paper we discuss the foundational protocols of VoIP and analyze the Quality of Service (QoS) of several applications. Our testing consisted of capturing packets during a conversation to determine packet loss, throughput, delay, and jitter. Several listeners subsequently provided Mean Opinion Scores (MOS) of several audio snippets based on phonetic clarity. We determine that Skype provides the highest quality and most consistent service.**

***Index Terms*—VoIP, Quality of Service, softphones, Mean Opinion Score**

I. Introduction

Voice over Internet Protocol (VoIP) is a transmission technique for audio calls and multimedia over an Internet Protocol (IP) network. It is an attractive and widely used service as a result of its low cost and widespread availability compared to traditional public switched telephone network. Implementation of VoIP can take either the form of a "hardphone" (a physical dedicated communication device such as a Cisco IP phone) or a "softphone" (a software-based program such as Skype). Currently, softphone applications are the most popular due to the ease-of-use and convenience for the user.

Furthermore, VoIP softphones are not exclusive to personal computers; mobile applications are available for smartphones. The diffusion of these applications put their services in front of many challenges in Quality of service (QoS) and security attacks.

In this paper we suggest a tool to analyze the performance of any VoIP applications by studding the expected attacks and vulnerabilities, and evaluate the QoS metrics and Mean opinion score (MOS). We focus our attention on VoIP applications, specifically choosing the most common applications including Skype, Google Talk, and Yahoo Messenger. QoS calculations and comparisons are done for the chosen applications by taking two input text files which contain sniffed UDP packets from a source and destination to evaluate the delay, jitter, and packet loss. Listening tests using ITU-T-recommended conditions are used to evaluate sound quality and returns MOS values for different VoIP applications.

978-1-5386-7693-6/18/$31.00 ©2018 IEEE

Our paper is outlined as follows. In Sec. II we discuss the fundamental protocols VoIP is comprised of. We discuss the currently available security measures for VoIP in Sec. III. The metrics we use to determine QoS of a VoIP service is presented in Sec. IV. Sec. V outlines our evaluation procedure, as well as the parameters for our MOS test (Sec. V-A). Our results and calculations are presented in Sec. VI, and we conclude our paper and propose future areas of work in Sec. VII.

II. Foundational Protocols

We outline the fundamental protocols of VoIP. In this technology, we have two main types of traffic: *signaling traffic* to make the connection, and *media traffic* to carry the voice. The two most common protocols used in signaling are H.323 and Session Initiation Protocol. . The protocols which are responsible for transmitting real-time data (such as audio and video) are Real-Time Transport and Real-Time Control Protocols. Additionally, ZRTP is a VoIP security protocols used to exchange the key, and Secure Real-Time Transport Protocol and Transport Layer Security are responsible for encrypting the media and the signaling stream. Fig. 1 shows the protocol stack of VoIP.

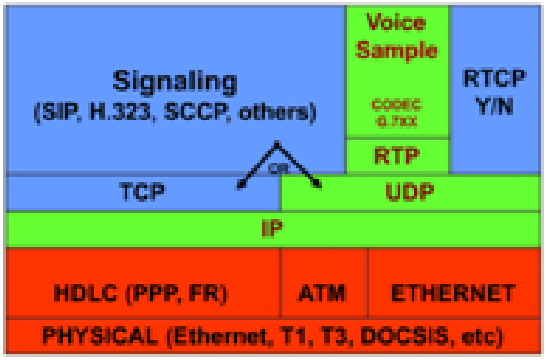

Fig. 1. VoIP protocol stack.

A. H.323

H. 323 is a standard published by the international telecommunication Union Telecommunication sector (ITU-T) to provide multimedia services. It consists of several protocols working together to provide VoIP services, including H.225.0 RAS (Registration, admission, and Status), H.225.0 Call Signaling,

H.245 Control Signaling, and Real-time Transport Protocol (RTP). H.323 is predominantly based on preceding ITU multimedia protocols, including H.320 (ISDN), H.321 (B-ISDN), and H.324 (GSTN) [1].

H.323's architectural components are shown in Fig. 2. The terminals are the LAN endpoints that support multimedia communications. Gateways connect a H.323 network with non-H.323 networks. The Gatekeeper, although optional, provides crucial services by translating between an IP address and a telephone number. The MCU (Multi control Unit) provides connectivity for more than two H.323 networks [2].

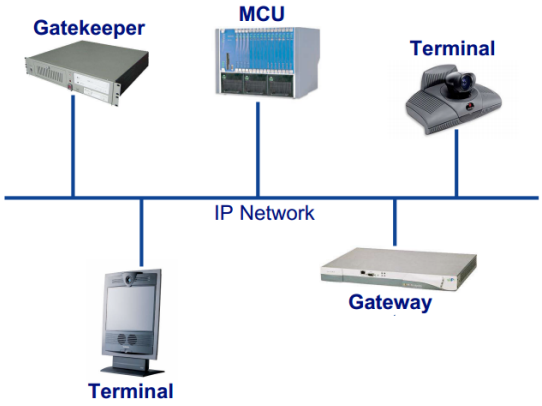

Fig. 2. H.323 architectural components.

B. Session Initiation Protocol

Provided by Internet Engineering Task Force (IEFT), the Session Initiation Protocol (SIP) is an application layer signaling protocol used to manage and control voice sessions over IP networks. When a user wants to access SIP services, authentication is performed by verifying the remote user and the server [3]. SIP uses a request/response method to establish communications between various components in the network. This will eventually establish a call or session between two or more endpoints.

The SIP architecture is composed of User Agents (UA), the Proxy server, the registrar server, and the redirect server. User agents are the users that originate or receive the calls using a softphone (e.g. Skype) or hardphone (e.g a Cisco IP phone) [4]–[6]. UA's can behave as both a user agent client and a user agent server. [4] The proxy server functions as relay server that forwards SIP requests on behalf of the clients. The registrar server tracks a user's position, and it accepts the RIGESTER message requests from the UA. It saves the information received in a location database. Lastly, the redirect server informs the client about the next hop(s) a message should take.

Fig. 3 demonstrates the registration process of user agent. The UA sends a REGISTER message that contains its information. Here, `iptel.org` is the SIP service provider address, `jiri@iptel.org` is the user agent address, and the contact address `195.37.78.173` is the IP address of the user agent. The register server accepts this message and saves the IP address of the user agent on the location database. Fig. 4 shows SIP call flow from a UA with address `caller@sip.com` to another UA with address `jiri@iptel.org` who is already registered on location database [5]. The SIP request method begins with an INVITE request to initiate the call, an ACK confirm a final response for INVITE, and a REGISTER to register with location server. Other requests are possible, such as BYE to terminate a call and CANCEL to cancel searching and ringing.

C. Real-Time Transport Protocol

Provided by IEFT, the Real-Time Transport Protocol (RTP) is used to transmit a real time data stream such as audio across IP network. It is an end-to-end data delivery service with real-time characteristics, and functions above UDP (Fig. 5). It is designed to only transmit data in a timely matter, without considering whether the packet is lost or not. Note, it cannot function above TCP because of TCP's innate retransmission characteristic.

RTP headers have a minimum size of 12 bytes, and contains the following fields: Version, Padding, Extension, Marker, Payload type, Sequence Number, Timestamp, SSRC, CSRC, and Header Extension [7]. RTP does not guarantee quality of service, despite the existence of the sequence number (which is responsible for handling out of sequence packets) and the timestamp (used to avoid jitter) [8].

D. Real Time Control Protocol

The Real-Time Control Protocol (RTCP) works in conjunction with RTP. While RTP is responsible for the delivery of the actual data, RTCP is used to send control packets between the caller and callee. RTCP primarily functions to provide feedback regarding the quality of service being provided by RTP [8], [9].

The specification defines several RTCP packet types to carry various control information, specifically a Sender Report (SR), a Receiver Report (RR), a Source description, a Goodbye (BYE), and an Application-specific message (APP). The SR's and RR's exchange information regarding packet losses, delay and jitter [8].

Additionally, RTCP packets carry a transport-level identifier (the canonical name) for a RTP source [9]. This is used to monitor and track each participant. Source description packets may contain textual information (e.g. username, email address) regarding the source. An application may implement multiple RTP streams regardless of prior identification by the SSRC identification, which can be trivially associated with this textual information. The RTCP packets are sent periodically by each session member via multicast to the other participants [8].

III. Security Protocols

We examine VoIP security protocols that are used to secure systems and maintains confidentiality, integrity and availability of the data. this subsection analyzes all VoIP security protocols and their corresponding algorithms.

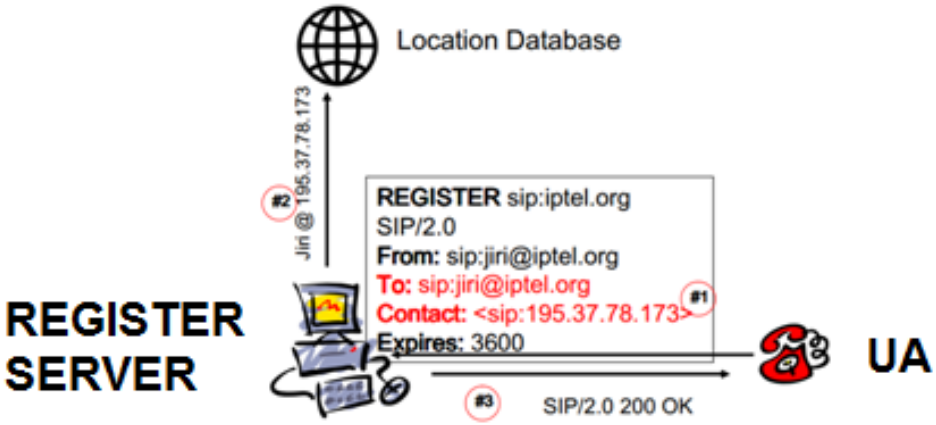

Fig. 3. SIP registration process.

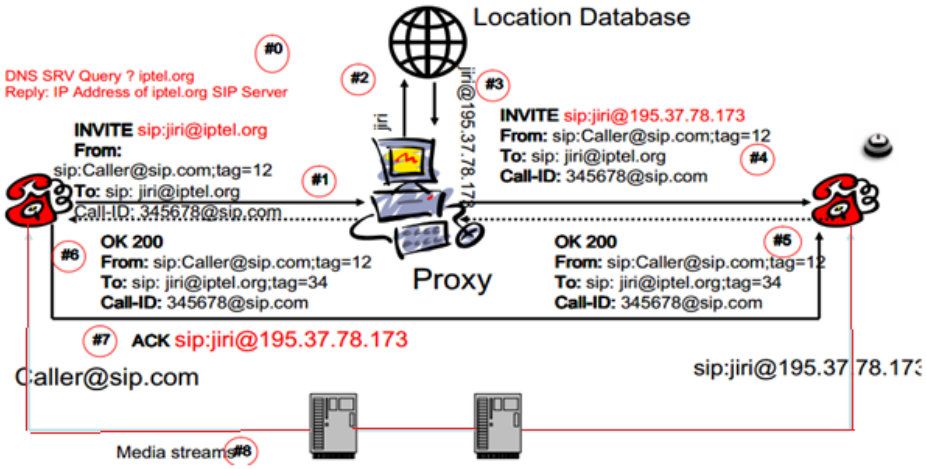

Fig. 4. SIP call flow diagram.

Fig. 5. RTP position among other protocols.

A. Transport Layer Security

Transport Layer Security (TLS) is a security protocol between the application layer and the transport layer as shown in Fig. 6 [10]. Implementing TLS is a means of protecting SIP signaling messages from the attacks that jeopardize the integrity, confidentiality and authentication of VoIP calls [11]. TLS offers mutual authentication, integrated key-management and secure key distribution.

It has a static information including trusted public key from trusted certification authorities and server certificate [12]. Fig. 7 shows the full handshake of TLS protocol. The Hello message is the first message sent during TLS session setup from the client to server.

This initiation message contains the higher version of TLS that the client can support, the acceptable Cipher Suites, and Compression Methods. An example cipher suite is TLS_RSA_WITH_DES_CBC_SHA, where TLS is the protocol version, RSA is the selected algorithm used for the key exchange, DES_CBC is the encryption algorithm (using a 56-bit key in cipher block chaining (CBC) mode), and SHA-1 is the hash function. Table I shows the algorithms are used for key establishment, media encryption and signature, and the type of certificate.

The server responds with the server-Hello message containing the strongest cipher that both client and server can support. The server certificate contains the public key of a server to be used by the client to authenticate the server. A server key exchange is optional and a temporary key created by the server to be used by the client. The client certificate request message that requests the certificate from the client and its optional message only required when the server want to authenticate the client. Server hello done indicates that the server is finished and awaiting a response from the client [14], [15].

Then, the Client responds to the server hello with a client

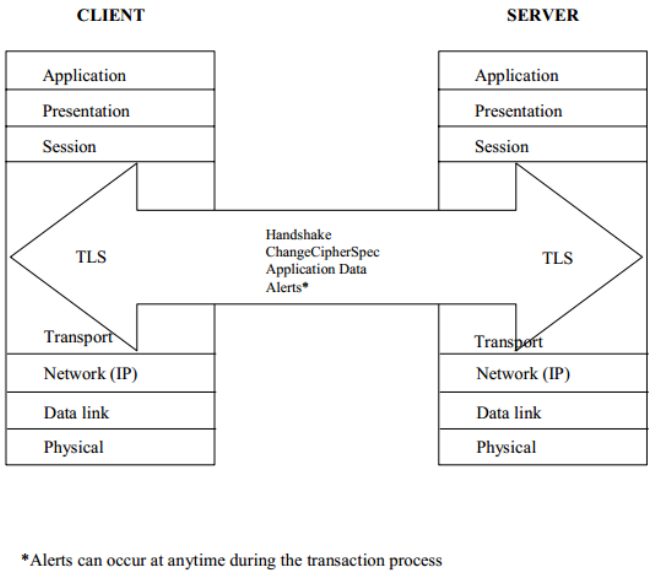

Fig. 6. Location of TLS in client/server stack.

TABLE I
TECHNICAL SPECIFICATIONS OF TLS [13].

Mechanism	TLS
Key establishment	RSA DH-RSA DH-DSS DHE-RSA DHE-DSS DH-Anon
Confidentiality	IDEA-CBC RC4-128 3DEA-EDE-CBS Kerberos AES
Signature	RSA DSA
Hash	MD5 SHA-1
Certificate Format	X.509

certificate message if the certificate requested in server hello message, along with the client key exchange message containing a secret key encrypted using the public key of the server. The Certificate verify message is required if client certificate is sent. Lastly, the Finished message are sent from both sides to verify the previous steps of handshake messages [16].

B. Secure Real-Time Transport Protocol (SRTP)

SRTP (Secure-Real-Time Protocol) and SRTCP (secure real time control protocol) provide confidentiality, message authentication, and replay protection to RTP and RTCP traffic [17].

As shown in Table II, SRTP encrypts RTP traffic using Encryption Algorithm (AES - CM, AES - F8), and uses Master key to derive session keys, and authenticates the message and its source by Authentication Algorithm (HMAC-SHA1 and MD5).

TABLE II
TECHNICAL SPECIFICATIONS OF SRTP [17].

Mechanism	SRTP
Key Method	SDES (Session Description Protocol Security Descriptions)
Encryption	AES-CM AES=F8
IP transport	UDP
Authentication Method	MD5 SHA-1

This protocol will protect the traffic from eavesdropping and modification, even if unauthorized users were able to capture the audio packets [17], [18]. SRTP is efficient for transporting packets, since it only encrypts the payload. A SRTP packet is distinguished from an RTP packet by the addition 4-byte authentication tag. SDES key method is used and exchanged within the SIP signaling. The SRTP-TLS protects media from being heard by unauthorized persons, and provides a high level of security for data. The internet now can stops eavesdroppers, hackers by Using TLS and SRTP to encrypt signaling and media [19].

C. ZRTP

As mentioned above, the SRTP protocol is used for encryption and message authentication. However, we need a method to share the keys are used for encryption between the call parties. The ZRTP protocol is used as a key agreement protocol to share a session key and parameters for establishing SRTP sessions. [20]

Ephemeral Diffie-Hellman keys are used in ZRTP protocol. A fresh key is generated each time a session is established. It does not rely solely on a public key infrastructure or on certification authority [20] ZRTP provides protection from man-in-the-middle attacks by using the short authentication string method where the two parties compare a value by reading it aloud.

If the two values match then no man in the middle attack has occurred [21]. It also provides protection from denial of service attacks by using of a sequence of keyed-Hash Message Authentication Codes (HMAC), that allow each party to discard false message injected by an intruder, This is achieved through the hash images H0, H1, H2 and H3 sent throughout the run of the protocol [21]. Fig. 8 shows the ZRTP call flow.

IV. QUALITY OF SERVICE

Quality of Service for VoIP describes a various quality of service concepts and features that are applicable on VoIP and depends on the traffic flow [22]. We outline a tool for analyzing the QoS for an application. The QoS of a VoIP application depends on the following metrics:

- *End-To-End Delay*: the time that will be lost until the packet reaches the destination. Delay is primarily a

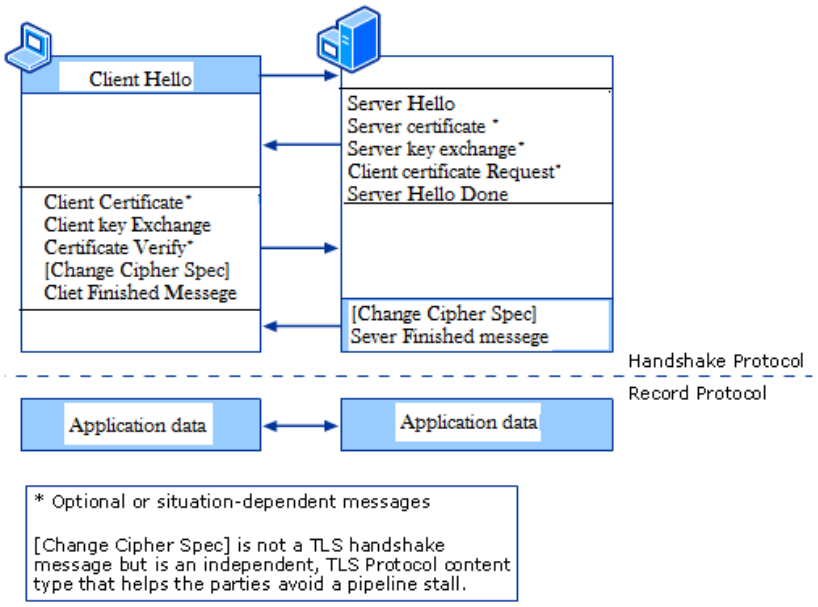

Fig. 7. The full TLS handshake.

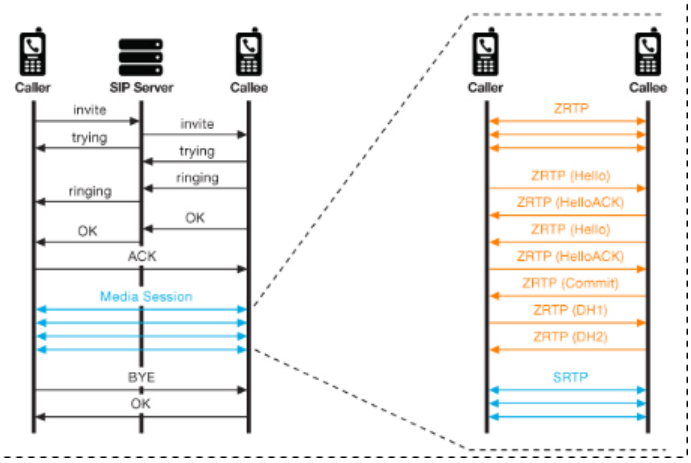

Fig. 8. The ZRTP call flow.

byproduct of processing and transmission, and represents one of the most important challenges to VoIP services.

- *Jitter*: the difference in total end-to-end delay of two consecutive packets in the data flow.
- *Out-of-Order packet delivery*: occurs in the complex topology where more than one path exists between the sender and the receiver.
- *Latency*: the amount of time required to transmit data from the source to the destination; it is an end-to-end delay that occurs during the information exchange between two nodes.
- *Bandwidth*: the total capacity of a transmission medium to transfer data. This capacity decreases if the network is saturated with users. The higher the bandwidth speed allows for more data to be sent over a broadband connection.
- *Packet Loss*: occurs when packets are dropped due to congestion or a limited buffer size on the receiving end. When packets are lost, they cannot be recovered without retransmission by the sender. This impacts the speed and quality of the network [22].

V. Evaluation

Having established our QoS metrics, we outline our evaluation procedure for VoIP applications, specifically Skype, Google Talk, Tango, Yahoo Messenger, Windows Live, and MicroSIP (encrypted and unencrypted). One of the most important functions of our tool is synchronization. We will use NetTime such that the testing computers will have the same time such that the times of the sent and received packets will be equal. The transmitted packets will be captured via Wireshark, where the majority of the packets are UDP.

The captured packets are divided into source and destination files, which are then read into our code. The program determines the average delay and jitter of each packet, the number of packets sent and received, the total packet loss, and the ratio of packet loss to throughput.

A. Mean Opinion Score Test

The Mean Opinion Score (MOS) is one of the tools to measure QoS in VoIP. It provides a numerical quantity to assess the quality of voice that is received after transmission [23]. MOS values range from 1 to 5, with 1 representing excellent speech quality and 5 for poor quality (shown in Table III, adapted from [24]). Since voice quality depends on delay, packet loss, jitter, and the codec used, we can use these factors to evaluate the MOS via E-modeling or listening tests [25].

TABLE III
THE POSSIBLE RANGE OF MOS VALUES [24].

Quality	Rating
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

For our analysis, MOS is evaluated using a *listening test*. We record the voice on the sender's (talker) and on the receiver's (listener) ends, and compare the difference in quality of the two audio files. The steps of the ITU-standard test are outlined as follows:

1) The talker should be seated in quite environment, with no reverberation.
2) The recording system must be high-quality; it can be a computer-controlled digital storage system.
3) Speech material should be meaningful, short sentences that are easy to understand. ITU-T recommends the following sample sentences:
   - *You will have to be very quiet.*
   - *There was nothing to be seen.*
   - *They worshiped wooden idols.*
   - *I want a minute with the inspector.*
   - *Did he need any money?*
4) Speech is recorded using microphone positioned 140-200 mm from the talker's lips.
5) There must be multiple talkers. Five talkers are used in this project to test each softphone.
6) The experiment should take place in different places or in the same location at different times [26].

VI. RESULTS AND CALCULATIONS

We outline our findings for each of the applications we tested. We calculated the average packet loss ratio, throughput, delay, and jitter for five trials of each service. For the MOS test, five listeners were provided five varying speech samples taken from different talkers. They scored the quality of each audio sample according to Table III. The MOS score is derived from the average of all the testers' scores.

The delay indicates how long it takes for a sent packet to be received. A comparison of the delays is presented in Fig. 9. Windows Live, Skype, and Tango produced similar, positive results. Yahoo Messenger had the greatest delay of all services. We observe a nontrivial difference between Encrypted MicroSIP and Unencrypted MicroSIP, indicating that added security will negatively impact performance.

Jitter calculates the variation in arrival time of packets. Our measurements are presented in Fig 10. Less jitter indicates clearer audio quality, since packets will arrive continuously over small time intervals. Skype and Windows Live perform the best of all the services tested, while Tango and Yahoo Messenger experienced the most jitter, and therefore poorest quality.

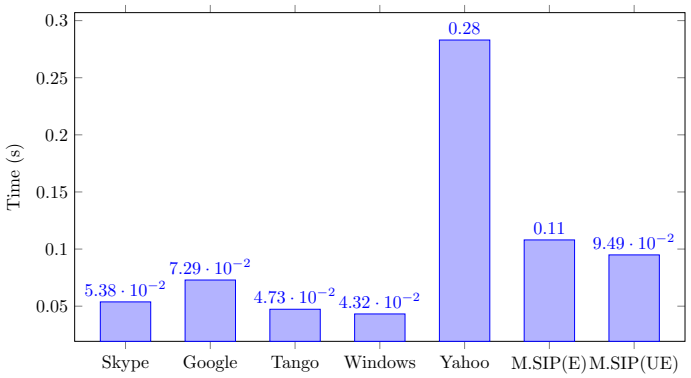

Fig. 9. The delay comparison for our VoIP applications.

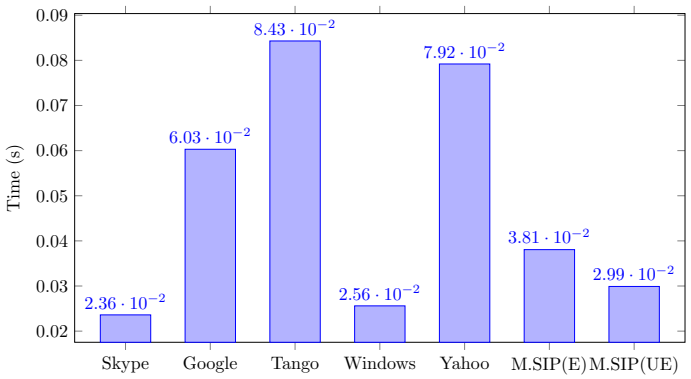

Fig. 10. The jitter comparison for our VoIP applications.

Packet loss impacts audio quality by creating abrupt "cuts" in continuous speech. This notion is crucial for VoIP, since it is a real-time application. Fig. 11 compares the packet loss ratio for our applications. Skype performed considerably better than all of the other applications, with Tango as the second best option being half as good as Skype. Interestingly, Encrypted MicroSIP outperformed the Unencrypted version.

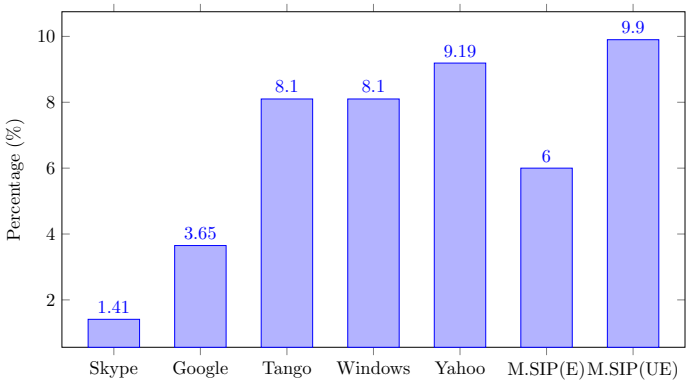

Fig. 11. The packet loss ratio comparison for our VoIP applications.

Throughput indicates the average of packets that are delivered successfully to the destination through communication channel. Our throughput findings are presented in Fig. 12.

Skype and Windows Live exhibited the highest throughput ratio. Tango, Yahoo Messenger, and Google Talk performed equivalently poorly.

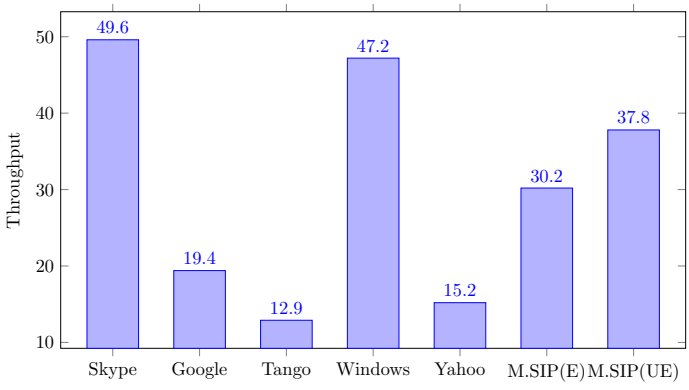

Fig. 12. The throughput comparison for our VoIP applications.

Lastly, we have our MOS test (Fig. 13) that we outlined in Sec. V-A. Once again, Skype performed better than its competitors, being the only one to achieve a "Good" quality rating with a MOS of 4.04. All other applications ranged from "Fair" to "Moderately Fair." MicroSIP yielded the lowest score with 2.52.

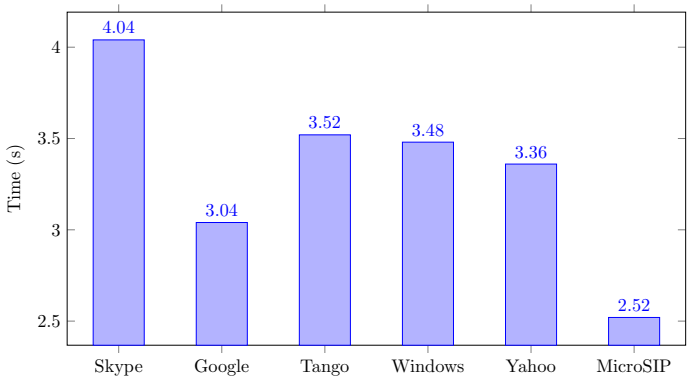

Fig. 13. The MOS comparison for our VoIP applications.

VII. Conclusion

In this paper we have discussed the foundational protocols of VoIP, as well as existing methods to secure communications. We conducted testing to evaluate the QoS for several VoIP applications. We captured transmitted packets to resolve the packet loss, throughput, delay, and jitter. Several testers scored audio samples for each application based upon the MOS test. Our results determined that Skype consistently outperformed all other services we examined. We plan on evaluating the security of these applications by testing common attacks (man-in-the-middle, denial of service, registration hijacking, session tear down, theft of service, and spam flooding) in our future work. We will develop a program to enhance VoIP security using DES-CBC encryption.

References

[1] H. Schulzrinne, J. Rosenberg *et al.*, "A comparison of sip and h. 323 for internet telephony," in *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*. sn, 1998, pp. 83–86.

[2] P. E. Jones, "h. 323 protocol overview," *Read April*, vol. 10, p. 2011, 2007.

[3] S. Kumari, S. A. Chaudhry, F. Wu, X. Li, M. S. Farash, and M. K. Khan, "An improved smart card based authentication scheme for session initiation protocol," *Peer-to-Peer Networking and Applications*, vol. 10, no. 1, pp. 92–105, 2017.

[4] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "Sip: session initiation protocol," Tech. Rep., 2002.

[5] IPTEL, "Session initiation protocol overview," Online, Sep. 2018.

[6] A. Suryawanshi, M. Bhati, and R. Dubey, "Session initiation protocol (sip)," 2013.

[7] J. Mattila, "Real-time transport protocol," *Department of Computer Science, University of Helsinki, Helsinki October*, vol. 19, 2003.

[8] T. Koistinen, "Protocol overview: Rtp and rtcp," *Nokia Telecommunications*, pp. 1–8, 2000.

[9] C. Huitema, "Real time control protocol (rtcp) attribute in session description protocol (sdp)," Tech. Rep., 2003.

[10] C. Shen, E. Nahum, H. Schulzrinne, and C. P. Wright, "The impact of tls on sip server performance: measurement and modeling," *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 4, pp. 1217–1230, 2012.

[11] B. Goode, "Voice over internet protocol (voip)," *Proceedings of the IEEE*, vol. 90, no. 9, pp. 1495–1517, 2002.

[12] S. Santesson, R. Housley, S. Bajaj, and L. Rosenthol, "Internet x. 509 public key infrastructure–certificate image," Tech. Rep., 2011.

[13] C. M. Chernick, C. Edington III, M. J. Fanto, and R. Rosenthal, "Guidelines for the selection and use of transport layer security (tls) implementations," Tech. Rep., 2005.

[14] G. Apostolopoulos, V. Peris, and D. Saha, "Transport layer security: How much does it really cost?" in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2. IEEE, 1999, pp. 717–725.

[15] E.-K. Kwon, Y.-G. Cho, and K.-J. Chae, "Integrated transport layer security: end-to-end security model between wtls and tls," in *Information Networking, 2001. Proceedings. 15th International Conference on*. IEEE, 2001, pp. 65–71.

[16] H. L. McKinley, "Ssl and tls: A beginners guide," *SANS Institute*, 2003.

[17] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman, "The secure real-time transport protocol (srtp)," Tech. Rep., 2004.

[18] D. Butcher, X. Li, and J. Guo, "Security challenge and defense in voip infrastructures," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1152–1162, 2007.

[19] Ingate, "The tls and srtp combination," Online, Apr. 2009.

[20] R. Bresciani, S. Superiore, S. Anna, and I. Pisa, "The zrtp protocol security considerations," *Technical Report, École Normale and Supérieure Cachan*, 2007.

[21] R. Bresciani, "The zrtp protocol analysis on the diffie-hellman mode," *Computer Science Department Technical Report TCD-CS-2009-13, Trinity College Dublin*, 2009.

[22] S. Chandra, "Qos in voip," in *Seminar Report, Department of Computer Science and Engineering*, 2004.

[23] M. Viswanathan and M. Viswanathan, "Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (mos) scale," *Computer Speech & Language*, vol. 19, no. 1, pp. 55–83, 2005.

[24] F. Ribeiro, D. Florêncio, C. Zhang, and M. Seltzer, "Crowdmos: An approach for crowdsourcing mean opinion score studies," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2416–2419.

[25] A. Rämö, "Voice quality evaluation of various codecs," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 4662–4665.

[26] *Mean opinion score (MOS) terminology*, Reccomendation itu-t p.800.2 ed., International Telecommunications Union, Jul.