SECURING REAL-TIME MULTIMEDIA: A BRIEF SURVEY

Bradley Clayton, Barry Irwin, Alfredo Terzoli

Computer Science Department
Rhodes University
Grahamstown

g01c2974@campus.ru.ac.za, b.irwin@ru.ac.za a.terzoli@ru.ac.za

ABSTRACT

Voice over IP (VoIP) enables cheaper and easier communication but can be less secure than the traditional TDM network. This paper is a guide to securing VoIP networks using current technologies and best practices. Physical and logical segregation of data and multimedia traffic is discussed. Current VoIP analysis tools are described with specific reference to their usefulness as a means of evaluating the quality of a secure VoIP system. Protocol enhancements, such as the Secure Real-time Transport Protocol and transport layer protection such as offered by IPSec, are discussed and evaluated.

Finally, various secure VoIP implementation scenarios are discussed, with configurations combining these security solutions presented in the paper.

KEY WORDS

VoIP, SIP, RTP, IPSec, Real-time Multimedia

SECURING REAL-TIME MULTIMEDIA: A BRIEF SURVEY

1 INTRODUCTION

VoIP has the ability to carry voice and video on a network where other devices and operations reside, for example an IP network. Like any other traffic on an IP network, VoIP traffic runs the risk of being intercepted by any network device which exists between two or more communicating parties. The infrastructure used by Time Devision Multiplexing (TDM) networks requires someone to find and tap into telephone wires before communications can be intercepted. In contrast, an eavesdropper of a VoIP conversation is able to intercept voice communication from any node over which s/he has control.

The context of the work undertaken for this paper is the securing of the architecture for the iLanga system developed at the Computer Science Department at Rhodes University [1]. The architecture consists of softswitches running Asterisk (an open source PBX [2]), Sip Express Router (SER), Open Gate Keeper and an innovative user interface [3]. SIP, RTP, H.323, IAX and MGCP end-points are supported but research time has been focused on SIP, IAX and RTP as these are emerging protocols while H.323 and MGCP are loosing popularity. The aim of this paper is to research possible methods to secure the iLanga architecture. To achieve this, current methods of security are investigated with the following priorities in mind:

- The extent to which the security method can provide Confidentiality, Integrity and Authenticity (referred to as CIA).
- The performance cost incurred by the security method.
- The scalability of the security method.

The paper is organised as follows: Section 2 begins with a discussion on basic VoIP security methods, followed by considerations that must be taken into account when adding security to real-time multimedia applications. Section 3 describes the important role of VoIP analysis tools when dealing with quality of service measurements and the interception and monitoring of communication. Internet Protocol Security (IPSec) and its performance when used to secure VoIP communication are discussed in section 4. The Secure Real-time Transfer Protocol (SRTP), used to secure media streams, is discussed in section 5. Methods of securing the Session Initiation Protocol (SIP) are discussed in section 6. Finally, in section 7, combinations of the above mentioned security methods are discussed. Issues surrounding the problem of implementing security into softphones and hardphones, are also discussed in section 7.

2 VOIP CONSTRAINTS ON SECURITY

VoIP nodes are usually connected to an Ethernet network making them as vulnerable to attack as any other computer or network element [4]. The first and most obvious solution to securing VoIP communication is to separate the voice services network from the data network. This follows the classical approach of TDM networks, where the voice infrastructure is completely separate from the data infrastructure. However, this makes for costly installation and reduces the appeal of using VoIP. In comparison to this costly option, virtual LAN technology (VLAN) allows for a logical separation of IP networks within one physical network.

Logical or physical separation of VoIP and data networks within an organisation is possible but separation cannot ensure the confidentiality, integrity and authenticity (CIA) of communication. In addition to this, separation is not possible when communicating via the Internet, unless a Virtual Private Network (VPN) is used. Confidentiality can be achieved by encrypting the media stream and signalling. Integrity could be achieved by hashing messages together with check-sums.

The authentication of a VoIP call is more difficult. Consider a VoIP call between two end-points via a softswitch. The softswitch is able to do minimal authentication of a user or end-point. This can be passed to the other user in the form of a caller ID. While this method of authentication is minimal, it is still better than two end-points communicating with no mediation in between.

The priority for real-time media is performance. Any security addition to a real-time application would be of no value if it induced a delay significant enough to degrade overall quality. When making security additions to VoIP one needs to also be aware of the impact on bandwidth as high quality media streams are already using more than 64Kb/s. Finally, security additions should make economical use of resources such as CPU and memory, should they be implemented into embedded systems or softswitches where many connections are handled concurrently.

3 VOIP ANALYSIS TOOLS

The wide adoption of VoIP technologies has seen an increase in tools that combine network and VoIP analysis. These tools measure aspects of VoIP networks, such as quality, throughput and resource usage. Many are also able to intercept and monitor VoIP content streams. This is an appealing feature should one wish to use these tools in South Africa, where the law requires that any electronic communication system have interception and monitoring capabilities [5, 6]. The use of analysis tools which assess the quality of a VoIP network is key to the study of VoIP security, as they allow one to investigate the trade-off between secure communication and resource usage. Seven current VoIP analysis tools have been benchmarked by Mier and Tarpley [7].

The seven analysis tools use different methods, incorporating the use of both software and hardware, when intercepting VoIP media. The most common method of interception is packet sniffing, using a network interface that is able to promiscuously pass packets up the IP stack. The packet sniffing solutions require that a mirrored port be configured to enable the interception of traffic on a switched network medium. Other solutions make use of a hardware device (referred to as a probe) that is placed inbetween two or more communicating parties. The probe intercepts and forwards data to a central point where the analysis is done. This method is costly, since a separate probe is needed for every physically medium that is to be monitored.

Of the seven analysis tools, none has the facility to decrypt a secured VoIP stream. This is probably due to the fact that their primary objective is the analysis of traffic patterns rather than the monitoring of VoIP media. An ultimate interception tool would be able to decrypt a VoIP media stream without compromising the security of that stream. To accomplish this, one could perform interception at the softswitch by only allowing calls to take place if their media streams pass through the softswitch. Communication between the softswitch and end-points would be encrypted enabling a secure system, with the option of monitoring should one have sufficient credentials on the softswitch.

The sections that follow discuss current security solutions that can be used to secure communication between a softswitch and end-points, and between two softswitches.

4 INTERNET PROTOCOL SECURITY (IPSEC)

IPSec resides below the IP layer of the OSI stack, providing a transparent security solution. IPSec can be implemented on individual hosts or network gateways, hence securing inter-network communication. If a security association is created between two machines, data transmitted between the machines is encrypted and/or authenticated. IPSec follows a user configured policy when deciding whether to transmit or accept encrypted or authenticated data [8]. Securing SIP and RTP with IPSec is simple: with IPSec set up, for example on a softswitch and an end-point, one would need to create a policy on both nodes that requires SIP and RTP packets to be authenticated and encrypted. During transmission, the IPSec layer will encrypt and authenticate SIP and RTP data. Upon reception of SIP and RTP data, the IPSec layer will authenticate and decrypt the packets before passing them to the application layer [8].

IPSec addresses confidentiality, integrity and authentication making it a possible solution to completely securing VoIP communication, whether on a segregated or shared network. Ranganathan and Kilmartin [4] have done simulations using IPSec to secure SIP and RTP protocols using OPNET, a tool for network simulation [9]. The focus of their investigation was to determine the setup time and the delay of real time media when calls are placed through different configurations of an IPSec engine. All of their simulations relate delay to call density, where the call density is the number of calls taking place over a unit of time. It is important to note that their simulations took place between two VoIP gateways and not between a gateway and a number of end-points.

To test the impact of IPSec encryption and authentication on VoIP media, IPSec was configured into tunnel mode with statically assigned security associations. It was found that an exponential increase in call setup and media stream delays occurred when the call density increased. Significant delays would be experienced on a softswitch handling many calls. Ranganathan and Kilmartin attribute the exponential increase in delay to the queue of packets waiting to be processed by the IPSec layer.

Before an IPSec association is created, the Internet Key Exchange (IKE) protocol is used to safely exchange encryption keys. Ranganathan and Kilmartin repeated the simulations incorporating dynamic key exchanges, which are negotiated with IKE, and found a larger delay on call setup and media transmissions. It was found that call setup times decrease when the call density increases provided that the security association lifetime is selected appropriately. This is attributed to the fact that a key exchange only takes place when a call is initiated, in order to create a new security association. Therefore, in the case of high call density, more calls can be set up without a delay: the first call creates a security association and subsequent calls use it.

5 SECURE REAL-TIME TRANSPORT PROTOCOL (SRTP)

The Secure Real-time Transfer Protocol (SRTP) is a profile for the Real-time Transport Protocol (RTP) that offers confidentiality, integrity and authenticity [10]. While SRTP is a profile in its own right, it is essentially an extension of the Audio/Video profile for RTP [11]. The Audio/Video profile has been modified so that the security implementation of SRTP exists between the RTP application and transport layers. RTP packets moving down the stack are intercepted and converted into SRTP packets before being passed to the transport layer. Conversely, SRTP packets moving up the stack are converted to RTP packets and passed to the RTP application layer. Likewise, Real-Time Control Protocol (RTCP) packets are converted into Secure Real-Time Control Protocol (SRTCP) packets when transmitted, and vice versa when received.

The aim of SRTP, as outlined in its RFC [10], is to ensure the confidentiality, integrity and authenticity of RTP and RTCP payloads. This is accomplished through a framework that allows for upgrading to new cryptographic algorithms and a low bandwidth overhead. Currently, a selection of encryption algorithms ensure a low computational penalty and a small footprint, making them ideal for mobile or small embedded devices, such as telephone handsets.

Currently, SRTP's collection of encryption methods consists of the Advanced Encryption Standard (AES) and the null, or no encryption, cipher. SRTP uses stream ciphers, making it vulnerable to statistical attacks. This is due to known formatting bits encrypted in the payload. The length of a stream cipher payload is always known, making it possible to use the known formatting bits, and their position, to derive the corresponding bit of the key-stream. However, it is claimed in RFC 3711 [10] that an attacker will not be able to use these known bits to deduce the rest of the stream if the cipher is *secure*.

SRTP is vulnerable to payload adjustment and source spoofing when message authentication is not used. For this reason SRTP message authentication must always be used. SRTP should also be protected with a strong authentication code. However, the RFC states that the sole use of integrity protection will not protect communications from replay attacks. SRTP counters replay attacks by using a *sliding window* and *replay list*. The replay list contains an index of all packets which have been received and authenticated. Upon receiving a packet, the packets index is compared to a list of recently captured packet indexes. The packet is rejected if its index is smaller then the index of the last received packet, less the size of the sliding window. The sliding window allows the protocol to use a fixed amount of memory for replay protection.

SRTP avoids denial of service attacks by using seekable stream ciphers: a cipher is able to address any position in its key-stream. This feature enables the encryption or decryption of a packet independently from preceding packets.

6 SECURING THE SESSION INITIATION PROTOCOL (SIP)

SIP is becoming common in VoIP implementations, replacing older protocols such as H.323. SIP is responsible for session negotiation between two or more SIP end-points [12]. Once a session has been negotiated, it is up to another protocol, for example RTP, to transport the media. SIP relies on the infrastructure of SIP proxy servers. Proxy servers provide registration and session invitation requests from end-points [12]. End-points are distinguished and addressed by their Uniform Resource Identifier (URI). A SIP URI consists of a user name, domain name and listening port value as seen below.

SIP URI: *sip:bclayton@sip.ru.ac.za:5060*
User: *bclayton*
Domain: *sip.ru.ac.za*
Listening port: *5060*

This SIP URI would identify the user bclayton at the domain sip.ru.ac.za listening on port 5060. When an end-point registers with a SIP proxy server, the server adds a URI to IP address mapping entry to a database. If the user bclayton registered an IP-phone with the IP address 146.231.117.13, the following URI-to-IP address mapping would be created [12].

Registry entry: *146.231.117.13:5060:60:bclayton@sip.ru.ac.za*
IP address: *146.231.117.13*
Listening port: *5060*
Timeout: *60*

If a second end-point were to send an invite request to the proxy server to start a session with bclayton@sip.ru.ac.za, the proxy server would look up the entry in its database based on the SIP URI. The server would use this entry to establish a session with the SIP device associated with the requested URI, in this case 146.231.117.13. Upon the successful creation of a session, real-time media is passed either directly between the end-points or via a softswitch. The path of the media stream depends on session creation rules held by the proxy server.

Unless otherwise referenced, the following discussion of methods of securing SIP are taken from Kuhn et al [13]. SIP is an application layer protocol and is very similar to the Hyper Text Transfer (HTTP) and Simple Mail Transfer (SMTP) protocols. Therefore, we are able to use the following security methods, taken from other application layer protocols and incorporated into SIP:

- HTTP Digest Authentication
- S/MIME
- Transport Layer Security (TLS)

HTTP Digest Authentication simply challenges a remote end-point by requiring a check-sum containing the username, password, HTTP method, requested URI and a nonce value (a random number used to protect against replay or statistical attacks). The password is never sent as clear text. However, HTTP Digest Authentication is susceptible to brute force attacks and is not recommended [12].

SIP already carries a MIME payload making it able to use S/MIME for authentication, integrity protection, confidentiality and key distribution. S/MIME SIP tunnelling is also possible should one wish to also protect the SIP headers, but creates an additional overhead which might be problematic. S/MIME uses certificates, for authentication and integrity, and private keys for confidentiality. Due to the overhead and need for minimal fragmentation, it is recommended that TCP be used instead of UDP when using S/MIME SIP tunneling.

Integrity, confidentiality and anti-replay of SIP packets can be protected by Transport Layer Security (TLS). The SIP RFC [12] requires the use of TLS when passing SIP messages between proxy, redirect and registrar servers. TLS is recommended when passing SIP messages to and from user agents. TLS is referred to as a *hop-by-hop* method of security because SIP messages are only secure while on the wire. Should a SIP message, protected by TLS, be passed from an end-point to a proxy via a second proxy, the second proxy would unwrap the message and re-wrap it with TLS before passing it on. It should be taken into consideration that TLS requires a reliable transport protocol and hence is only supported by TCP, which is unfortunately less suitable than UDP to the needs of a real-time transport protocol.

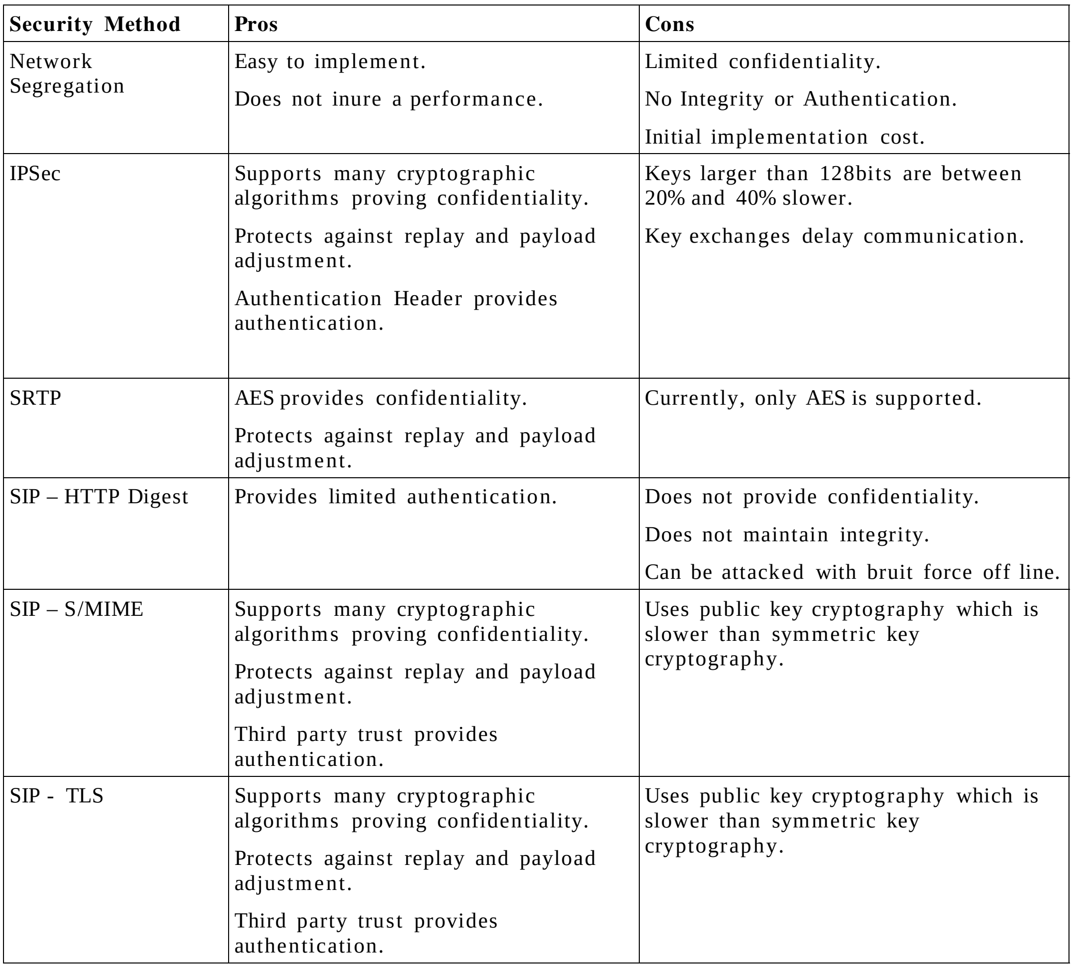
7 CONCLUSION

Table 1 outlines the pros and cons of each security method discussed in this paper. From the table we can see that the security methods:

- Are suitable for protecting different parts of a VoIP system.
- Provide different levels of confidentiality.
- Mostly provide authentication.
- Mostly provide integrity.

It is important to find the best combinations of security methods which do not undermine each other or retard VoIP's real-time performance. For example, one could argue that IPSec be used between SIP proxy servers securing a trunk of many communication channels with a single set of encryption keys. TLS and SRTP could be used between servers and end-points, where encryption on the server side is off-loaded onto hardware, allowing the server to handle a greater number of secure channels. Since TLS and IPSec are *hop-by-hop* encryption protocols they allow interception of communication at the servers. Alternatively, TLS and SRTP could be used throughout the entire VoIP system. This alternative lowers the number of security protocols which minimizes the complexity of the overall solution. IPSec implemented throughout a VoIP system requires security associations to be setup on all nodes. This may make the implementation of a large system complicated, unless an automated method of configuring IPSec, based on options set in the VoIP application on clients, are employed.

The nature of security protocols also governs their ability to be implemented on softphones and hardphones. Securing a softphone with IPSec is trivial as IPSec is already built into most modern operating systems. Implementing IPSec into an embedded hardphone would require substantial modification of its IP stack while RTP, assuming it has been implemented according to specifications, merely requires modification of an already pluggable module to secure the RTP payload. However, SRTP without IPSec would leave the session initiation protocols unencrypted unless TLS or S/MIME were also implemented.



Security Method	Pros	Cons
Network Segregation	Easy to implement. Does not inure a performance.	Limited confidentiality. No Integrity or Authentication. Initial implementation cost.
IPSec	Supports many cryptographic algorithms proving confidentiality. Protects against replay and payload adjustment. Authentication Header provides authentication.	Keys larger than 128bits are between 20% and 40% slower. Key exchanges delay communication.
SRTP	AES provides confidentiality. Protects against replay and payload adjustment.	Currently, only AES is supported.
SIP – HTTP Digest	Provides limited authentication.	Does not provide confidentiality. Does not maintain integrity. Can be attacked with bruit force off line.
SIP – S/MIME	Supports many cryptographic algorithms proving confidentiality. Protects against replay and payload adjustment. Third party trust provides authentication.	Uses public key cryptography which is slower than symmetric key cryptography.
SIP - TLS	Supports many cryptographic algorithms proving confidentiality. Protects against replay and payload adjustment. Third party trust provides authentication.	Uses public key cryptography which is slower than symmetric key cryptography.

Table 1: Pros and cons of each of the researched security methods.

This work was undertaken in the Distributed Multimedia CoE at Rhodes University, with financial support from Telkom SA, Business Connexion, Comverce, Verso Technologies, THRIP, the National Research Foundation and the German Academic Exchange Service.

References

[1] J. Penton and A. Terzoli, "iLanga: a next generation voip-based, TDM-enabled PBX," in *Proceedings of SATNAC 2004 - Next Generation Networks*, September 2004.

[2] J. Penton and A. Terzoli, "Asterisk: A Converged TCM and Packet-based Communications System," in *Proceedings of SATNAC 2003 - Next Generation Networks*, September 2003.

[3] J. Hitchcock, J. Penton, and A. Terzoli, "A multilanguage, open source approach to connecting user interfaces to a next-generation PBX," in *Proceedings of SATNAC 2004 - Next Generation Networks*, September 2004.

[4] M. K. Raganathan and L. Kilmartin, "Performance analysis of secure session initiation protocol-based VoIP networks," *Computer Communications*, vol. 26, pp. 552–565, 2003.

[5] S. A. Government, "Regulation of interception of communications and provision of communication-related information act," *Government Gazette*, vol. 451, January 2003.

[6] S. A. Government, "Electronic Communications and Transactions (ECT) Act," *Government Gazette*, vol. 446, August 2002.

[7] E. Mier and R. Tarpley, "VoIP analysis tools: Picking up VoIP-specific tools for the network management workbench," tech. rep., Network World, March 2003.

[8] C. R. Davis, *IPSec. Securing VPNs*. Berkeley, California: Osborne/McGraw-Hill, 2001.

[9] M. W. Dixon and T. W. Koziniec, "Using OPNET to Enhance Student Learning in a Data Communications Course," in *IS2002 Proceedings of the Informing Science and IT Education Conference*, pp. 349–355, ACM Press/Addison-Wesley Publishing Co., June 2002.

[10] D. McGrew, E. Carrara, M. Baugher, M. Naslund, and K. Norrman, "RFC 3711: The Secure Real-time Transport Protocol (SRTP)," tech. rep., Cisco Systems, Inc and Ericsson Research, March 2004.

[11] S. Casner and H. Schulzrinne, "RFC 3551: RTP profile for Audio and Video Conferences with Minimal Control," tech. rep., Columbia University, Packet Design, July 2003.

[12] H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "RFC 3261: SIP Session Initiation Protocol," tech. rep., Dynamicsoft, Columbia U., Ericsson, WorldCom, Neustar, ICIR, AT&T, June 2002.

[13] D. Juhn, T. Walsh, and S. Fries, "Security Considerations for Voice Over IP Systems," tech. rep., US National Institute of Standards and Technology, January 2005.