A SURVEY OF VOIP INTRUSIONS AND INTRUSION DETECTION SYSTEMS

Son Vuong* and Yan Bai**

(*) Department of Computer Science
(**) Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, BC V6T 1Z4 Canada
vuong@cs.ubc.ca and yanb@ece.ubc.ca

ABSTRACT

This paper presents a survey of the security problems in VoIP networks, with an emphasis on both intrusions and intrusion detection methods. It examines the intrusion issues in different components of VoIP systems, points to the strengths and shortcomings in the existing intrusion detection methods and intrusion detection systems and suggests possible future research directions.

Keywords: *VoIP, intrusion, intrusion detection, network security, IPSec, H.323, SIP, H.248.*

1. INTRODUCTION

VoIP applications have grown rapidly and continue to enjoy exponential growth due to largely reduced cost and wider range of advanced services [6], as compared to traditional telephone networks. VoIP applications, however, faces many technical challenges. The two major challenges are the provision of Quality of Service (QoS) and the protection of security. The QoS issues have been extensively studied [13], while the VoIP-related security issues, to our knowledge, have not been adequately investigated. Furthermore, since VoIP is a real-time service, any security threats, such as disruption or Denial of Service (DoS) attack, will compromise the QoS.

This paper aims at providing a critical survey of VoIP security with a focus on intrusions and intrusion detection systems. Section 2 presents an overview of the intrusions in VoIP networks. Section 3 discusses intrusion detection techniques. Finally, in Section 4 we present our concluding remarks and offer suggestions for future research in VoIP security.

2. INTRUSIONS IN VOIP NETWORKS

Intrusion is an action that an intruder breaks into a system or a legitimate user misuses system resources [11]. We classify the intrusions in VoIP networks into three categories: (i) those associated with IP networks, (ii) those inherited from traditional telephone systems, and (iii) those specific to VoIP protocols.

2.1 Intrusions in IP networks

IP-based transmissions are inherently unsecured. Therefore, VoIP applications would face security threats inherited from IP networks. A comprehensive survey of Internet intrusions can be found in [16]. In this paper, the author classifies the Internet infrastructure attacks into four categories: DNS hacking, routing table poisoning, packet mistreatment and DoS, and discusses the impact of these kinds of intrusions on the Internet. Table 1 is an excerpt from this paper.

Table 1. Taxonomy of Internet Intrusions

Type	Target	Intrusion Scenario	Impact
DNS hacking	DNS	- Cache poisoning - Server compromising - Spoofing	- DoS - Masquerading - Information leakage - Domain hijacking
Routing table poisoning	Backbone routers	- Link attacks e.g., interruption - Route attacks e.g., link state router attacks	- Suboptimal routing - Congestion - Overwhelmed host - Looping
Packet mistreatment	Communication links	- Link attacks e.g., interruption - Router attacks	- Congestion - Lowering throughput - DoS
DoS	End systems	- UDP flood - TCP/SYN flood - ICMP/Smurf	Entire groups or whole portions of the Internet affected

2.2 Intruder scenarios in traditional telephone networks

The security risks in traditional telephone networks are typically insignificant [21]. However, a few threats to the switch security still exist and are listed in the Table 2 [22].

Table 2. Basic treats to traditional telephone network

Security Treat	Example
Phone disturbance	Unwanted phone calls
Free phone call	Using other person's phone number by hacking the signaling system
Masquerading of caller or callee	Masquerading of caller or callee
Availability attacks	Access from and to phone services by misusing signaling system or cutting wires

Furthermore, the development of the intelligent network using SS7 (Signaling System No.7) provides greater flexibility to the network through the introduction of new services. It, however, increases its vulnerability to the misuse of those services because certain services allow users access to management information. Free phone service is an example. Mobile technology also impacts telephone security [15]. The above attacks would also affect VoIP users because VoIP networks involve traditional telephone equipment.

2.3 Security Threats in VoIP Protocols

VoIP relies on various protocols to address different aspects of a "call". Typically, these protocols are categorized into *media transport protocols* that are responsible for the digitization, encoding, decoding, packing, reception, and ordering of voice and voice samples, *signaling protocols* that perform session management including locating a user, session establishment and setup negotiation, modifying a session, and tearing down a session, and *other protocols* that are common to any IP-based network such as those involved with QoS provision.

IP telephony-related protocols are not initially designed with security as a prime design goal. Although some of these protocols have added security features in their recent versions, security mechanisms are not secure enough or are still impractical. For example, in a signaling protocol that does not maintain knowledge about changes made to the media path during a call, if one is able to abuse the media path, the signaling path will remain unnotified and clueless about the changes performed to the media path. Another example lies in a signaling protocol that does not have an integrity checking mechanism. This section discusses the security characteristics of the VoIP standards that are currently used in building VoIP systems including SIGTRAN [27], H.323, Session Initiation Protocol (SIP), and Megaco [29]. An overview of all of the above standards can be found in [27,29].

3.3.1. Security issues within SIGTRAN

In the implementation of SIGTRAN, communication security, non-repudiation and system security have been considered [17]. Communication security refers to authentication of peers, integrity of user data transport, confidentiality of user data, and replay protection, while system security refers to avoidance of unauthorized use, inappropriate use and DoS. Among these, the resistance to DoS is provided by Stream Control Transport Protocol (SCTP), a base protocol of SIGTRAN [4]. SIGTRAN also relies on IPSec and TLS for secure communication. IPSec is designed to secure both headers and payload of IP packets by using Authentication Header (AH) and Encapsulating Security Payload (ESP). The IPSec protocols target at addressing the Password Sniffing, IP Spoofing, Session Hijacking, and DoS [20]. Research shows that SIGTRAN with IPSec is more secure than SIGTRAN with TLS [17].

3.3.2 Security issues within H.323

H.323 relies on the H.235 standard to provide security features including authentication, integrity, privacy, and nonrepudiation [29]. The authentication function makes sure that the endpoints are really who they say they are. The integrity function validates that the data is indeed an unchanged representation of the data. The Privacy function hides the data from eavesdroppers. Nonrepudiation protects against someone falsely denying that they participated in a call.

The registration, admission and status (RAS) channel used for gateway-to-gatekeeper signaling is not a secure channel. H.235 allows gateways to include an authentication key in their RAS messages. The gatekeeper can use this authentication key to authenticate the source of the messages. So far, only a few VoIP products can support H.235 features.

3.3.3 Security issues within SIP

SIP covers only signaling aspects. The media stream confidentiality is not treated by the standard. For signaling path, the security mechanisms have been developed to secure both SIP header and SIP message body. The header includes information about communication patterns and contents of individuals, or other confidential information, while the message body contains user information, such as media type, codec, addresses and ports. The mechanisms can be classified into end-to-end and hop-to-hop protection [12]. End-to-end protection are realized by SIP authentication using digest authentication (e.g., HTTP digest), and SIP message body encryption using S/MIME (Secure Multipurpose Internet Mail Extension). Hop-to-hop mechanisms rely on network-level security mechanism, such as Transport Layer Security (TLS) [19]. Recent Internet draft shows call flows demonstrating the use of TLS and S/MIME in SIP [23].

Although the security mechanisms provided with SIP reduces the risk of attack, the SIP communications are susceptible to several types of attack such as Snooping, Modification, DoS and Spoofing [24, 25, 26] (See Table 3). This is due to the limitations in the security mechanisms defined by SIP. In [31], author pointed out two limitations: one is associated with the use of HTTP Digest and another is the assertion and validation of user identity by SIP servers.

Table 3. Some Intrusions in SIP

Type	Description and Example
Snooping	Permit an attacker to gain information on users' identifiers, services, media, and network topology.
Modification	Intercept the signaling path and modify SIP messages in order to change some service characteristics, e.g. force a particular router, change a user registration or modify a service profile
Spoofing	Impersonate the identity of a server or a user to gain some information to modify a session such as termination of a call, or to perform DoS.
DoS	Attempts to "flood" a network, thereby preventing legitimate network traffic

In [23], the author presents some other SIP-related security threats including forking, reflection attack, multi-proxy authentication, encryption limitations, cancel security and NAT and firewall traversal.

3.3.4 Security issues within Megaco

Megaco protocol is a media gateway control protocol. Unlike H.323 or SIP, a peer to peer architecture, Megaco is master/slave architecture for decomposed gateways, in which Media Gateway Controller (MGC) is the master server and Media Gateway (MGs) are the slave clients. Megaco provides support for large-scale end-to-end deployment of VoIP systems. Many industrial companies, such as Cisco, Lucent, Nortel, Microsoft, and Motorola are actively developing related network products.

Security for Megaco includes protection of protocol connections and protection of media connections [28].

1) Protection of protocol connections relies on TLS or IPSec. When Megaco is used with IPSec, AH, ESP and IKE (Internet Key Exchange) are included. The AH header is responsible for data origin authentication, connectionless integrity and optional anti-replay protection of messages passed between the MG and the MGC. The ESP header provides confidentiality of messages. IKE provides a mechanism to negotiate and exchange keys in secrecy. In the protocol specification, AH is mandatory. When the underlying operating system does not support IPSec, an interim AH solution can be employed. The interim AH scheme, however, does not provide protection against eavesdropping and replay attacks [28]. Furthermore, the DoS attacks on MGs or misbehaving MGCs could happen [28]. For example, a DoS attack would occur at a MGC when the attacker send large amount of UDP packets to the protocol's default port 2944 or 2945, which keeps the MGC busy in handle illegal messages, and finally block the normal service. An attacker can keep sending ServiceChange or AuditCapabilities command to a MG and thereby bring down the MG. Other possible security threats in Megaco networks include: call tracking, call redirecting, and toll fraud.

2) Protection of media connections is to prevent eavesdropping or altering of the voice stream between caller and callee. To protect from eavesdropping, the Megaco protocol allows the MGC to provide MGs with session keys that can be used to encrypt the media streams. This solution, however, introduces extra delay for encryption and decryption. The time needed to break the encrypted message depends on the length of the session key. The higher the level of security protection, the larger the delay would be. It is, therefore, not very practical for VoIP applications that tend to be delay and jitter-sensitive.

To combat the problem of "uncontrolled barge-in" in which media packets are directed to the IP address and UDP port used by a connection, Megaco only allows to accept packets from known sources. The source verification can be done by checking both IP source address and UDP source port to see if they match desired values or encrypting and authenticating the packets though the use of a secret key that is conveyed during the call set-up procedure. However, both methods will slow down connection establishment.

Another recent proposal for securing the media part of a Megaco network is the use of Secure Real-time Transport Protocol (SRTP) [9]. The SRTP is designed to provide confidentiality and authentication for RTP as well as RTCP by integrity checks and encryption. However, it could not prevent DoS attack. The nature of DoS atack is the volume of packets it creates towards an unwitting target; whether those packets are signed by the server, or are encrypted with the wrong key, is not relevant for the attack [3].

3. INTRUSION DETECTIONS

To combat security threats, three categories of techniques have been developed based on their functionality: (i) Security Enabling, (ii) Security Protection and (iii) Security Violation Detection techniques. Security Enabling techniques such as Public Key Infrastructure aims at ensuring that messages cannot be intercepted or read by anyone other than the intended person and guaranteeing the authenticity of a message. Security Protection techniques such as firewall focus on protecting from external threats. Security Violation Detection techniques such as Intrusion Detection

concentrate in monitoring the events in a computer system or network and analyzing them for signs of intrusions.

Typically, there are two intrusion detection methods: *anomaly* detection and *misuse* detection. *Misuse detection* methods use information about a known security policy, known vulnerabilities and known attacks on the systems they monitor. This approach compares network activity or system audited data to a database of known attack signatures or other misuse indicators; resulting pattern matches produce alarms of various sorts. When a comprehensive and up-to-date set of attack signatures is used, this approach is fairly reliable but limited for the following reasons [11]:

1) The number of known security vulnerabilities is large and the techniques to exploit those vulnerabilities are also vast. As a result, effective signature databases are difficult to design and maintain, and their execution may become huge and unwieldy.

2) Many applications and systems have unknown and undocumented holes, so new vulnerabilities are constantly being discovered.

On the other hand, *anomaly detection* methods use information about repetitive and usual behavior on the systems. This approach detects events that diverge from expected usage patterns. This approach requires no prior knowledge of invalid behavior, however, it is prone to generating unacceptable numbers of false alarms for the following reasons [11].

1) Normal behavior may also include forbidden behavior, so excluding this activity from a normal data set in a production environment is extremely difficult.

2) Users very frequently do not exhibit consistent behavior; i.e., while undertaking perfectly valid activity they often deviate from a "normal" profile and thus cause many false "positives". Such things as deadline pressure, vacations, or just general user contrariness can cause deviations from normal. In the most extreme case, a completely new behavior profile may be immediately exhibited as a result of a job change or a new assignment.

3) If the system employs a profiling system that adjusts to new user activity over time, knowledgeable, patient, and malicious users can gradually train the system to accept invalid behavior.

Both *misuse* and *anomaly* detection methods have been used in Host-based Intrusion Detection System (H_IDS) and Network-based Intrusion Detection System (N_IDS). A H_IDS is installed on a single computer and checks the integrity of the system files and watches for suspicious processes. H_IDS evaluates information found on host computers by accessing and reading logs and audit-records of interest. The information may include contents from the operating system, file system and software applications. Host computers can be user workstations loaded with specialized applications such as web browsers or servers, peripherals such as printers, or network components such as firewalls, routers, and switches.

Typically, H_IDS should monitor and record login and logoff times, application-processing times, connections with the computer, and time/size changes of critical system files. The H_IDS should be able to read TCP headers in order to detect possible buffer-overflow attacks, and to track specific programs' activities that reveal a possible worm/virus signature activity [7].

Since H_IDS focuses on monitoring the operating system processes during the program execution and alerting on anomalous sequences of system calls [5], the strengths of H_IDS are as follows [7, 8].

1) H_IDS can verify the success or failure of an attack since the reporting is based on examinations of events recorded in the system log.
2) Specific system activities are closely monitored. For example file access, changes to file permissions, user logon/logoff, and administrator functions can be monitored at a level of detail greater than a N_IDS.
3) Attack can be detected that do not cross the point of network entry. It is very useful for protecting from internal users.
4) H_IDS is tightly integrated with the operating system. The network encryption does not affect H_IDS.
5) H_IDS are installed on existing servers without additional requirements. The cost of initial deployment is therefore lower than N_IDS.

However, there are some limitations in H_IDS. First, H_IDS software requires high processing power and memory storage capacity, as well as sufficient system resources to install it [7]. Second, the number and the diversity of computers often make it impossible to protect each computer individually with H_IDS [1]. Third, H_IDS detects an outside intruder only after the intruder has reached the monitored host system, not before, as can N_IDS.

On the other hand, a N_IDS consists of a collection of agent applications strategically placed within a network that monitors WAN or LAN traffic [7]. N_IDS evaluates information captured from network communications. Typically, it analyzes the stream of packets traveling across the network to examine both context and content of defined attacks. N_IDS is installed on dedicated workstations that are placed behind each external firewall, or outside an external firewall, or on major network backbones, or on critical subnets [14]. The strengths of N_IDS are listed below [7, 8, 10].

1) N_IDS can detect unsuccessful attacks when it is deployed outside of a firewall, this property can be used in forensic analysis.
2) N_IDS can detect network-based attacks by checking all the packet headers of any malicious attack, such as IP-based DoS attacks --- TCP SYN attacks, fragmented packet attacks.
3) N_IDS performs a real-time detection. The attacker cannot remove evidence of attack, and those data can be used in evaluation of security policy. Also, real-time detection can get quick response to stop an attack before it compromises the system.
4) N_IDS is easier to deploy since it does not affect existing system or infrastructure.

5) The source data of an N_IDS comes from network packets, which is independent of the operating system.
6) Per-owner cost is low since one N_IDS placed at a critical network entry point can provide security for multiple systems.

However, some concerns about N_IDS is to process all packets for a large or busy network in real-time. When N_IDS is required to keep up with the analysis and storage of information generated by potentially thousands of machines in a network, it might drop packets or fail to recognize an attack launched during busy periods. In the worst case, intruders can launch an attack in an overloaded network by flooding N_IDS with spurious traffic. Other concerns about N_IDS include the following [1, 2]. First, N_IDS cannot handle encrypted data. An N_IDS is designed to monitor network traffic between hosts and deduces behavior based on the content and format of data packets on the network and analyzes overt requests for sensitive information and repeated failed attempts to violate security policy. Second, N_IDS can be subject to an attack where attack packets are broken up into multiple smaller packets, and to insertion and evasion attacks which do not occur in H_IDS.

In sum, both H_IDS and N_IDS have their strengths and weaknesses. They are complementary to each other, and should be used in combination for effective complete intrusion detection.

Several papers compare commercial Intrusion Detection System (IDS) products [7, 8]. There are three main findings in the papers. First, different IDS are designed to handle different type of security treats and use different intrusion detection methods. No IDS can detect all malicious insider activity [11]. Second, different N_IDS are designed to decipher different network and application protocols. The protocols are DNS, HTTP, UDP, TCP and IP, however, no VoIP-specific protocols have been included in existing N_IDS. Finally, H_IDS are targeted to certain users and environments, particularly the operating system [8].

IDS for VoIP Networks

VoIP applications have specific characteristics that require special handling in an IDS. The following points are noteworthy:

(a) VoIP-specific protocols rely on IPSec. IPSec provide a mechanism that can be used to hide both the contents and addresses of network packets between cooperating agents such as firewalls; however, this renders the actual source, destination and the content of packet opaque while they are in transit between agents. If an IDS is positioned along the agent-to-agent path, it will be unable to determine the real origin or destination of the traffic. For this reason alone, H_IDS, which has the ability to view message content even if the message is encrypted in transit, should be considered with a greater emphasis.

VoIP-systems, however, should be designed to work in large-scale environments; so the scalability of the security mechanisms is another issue. Also, one would expect that multimedia terminals will be deployed in proliferation in lightweight devices such as mobile phones and handhelds. These portable computers do not have the computing power as PC's. For these two reasons, it is deemed essential to integrate both N_IDS and H_IDS to capture and analyze both network packets and to analyze information found on host computers. Another benefit of the integrated approach is that the diversity sources of available data can reduce false alarms.

(b) N_IDS for decoding and interpreting VoIP-specific protocols is needed.

(c) VoIP applications are time-sensitive. The data must be processed in fixed time frames to ensure an acceptable result. Security services should be kept as insignificant as possible an impact on this demand.

4. CONCLUSIONS

The VoIP security issues and solutions are increasingly important for the success of VoIP services, especially in the domain of intrusions and intrusion detections. Based on this condensed survey, it is worth noting the following important points:

1. A VoIP system suffers similar security concerns faced by other systems and devices connected to a public network such as the Internet.

2. The increasing frequency and complexity of the intruder make it impossible for any IDS to protect a network from all threats. The new IDS should focus on overcoming the weakness of the current IDS, offering capabilities beyond the current IDS's ability to report intrusions simply by types of signatures and attacks.

3. The characteristics of VoIP applications necessitate VoIP-related IDS. These include time-sensitivity, IPSec dependency, scalability, and VoIP-protocol decoding ability.

In targeting an effective, flexible and holistic approach to VoIP security management, we propose the use of a suitable mobile agent system in an integrated framework which can be applied specifically to VoIP as well as to modern network management in general. Another piece of ongoing work on VoIP security is to develop a taxonomy of vulnerabilities in Megaco, both from a victim and an intruder perspective. This taxonomy will assist in the design and evaluation of an IDS that targets misuse detection.

Acknowledgement

This research is supported in part by the Directorate of Telecom Engineering and Certification of the Department of Industry Canada. We would like to express our gratitude to Peter Chau of Industry Canada for the fruitful discussions leading to this project.

We gratefully acknowledge the contributions of Ken Deeter, Jim Zeng, Sergio Gonzalez, Xiaojuan Cai. Jinmei Yang, Ling Yun, and Ardashir Bahi in the forms of discussions, presentations, inputs and reviews of draft sections. Heartfelt thanks also go to Albert Ip and Sam Lam for the fruitful discussions on practical issues of VoIP networks.

REFERENCES

[1] R. Zhang and D. Qian, "Multi-agent Based Intrusion Detection Architecture", *2001 IEEE International Conference on Computer Networks and Mobile Computing*, Beijing, China, Oct. 2001.

[2] http://www.cs.uct.ac.za/courses/CS400W/NIS/papers2003/tbaving.pdf

[3] J. Rosenberg, "The Real Time Transport Protocol (RTP) Denial of Service (DoS) Attack and its Prevention", *Internet-Draft,* draft-rosenberg-mmusic-rtp-denialofservice-00.

[4] P.T. Conrad and G. J. Heinz, "SCTP in Battlefield Networks", *MILCOM 2001*, Mclean, USA, Oct. 2001

[5] S. Hershkop and et.al. "Host-based Anomaly Detection by Wrapping File System Accesses", *Columbia Tech Report,* Apr. 2003.

[6] M. Hassan, "Internet Telephony: Services, Technical Challenges, and Products", *IEEE Communication Magazine*, April 2000.

[7] B. Mukherjee, "Network Intrusion Detection", *IEEE Network*, Vol.8, No.3, pp.26-41, 1994.

[8] S. Axelsson, "Research in Intrusion-detection Systems: A Survey", *Technical Report 99-15*, Dept. of Computer Engineering, Chalmers University of Technology, Sweden, Mar. 2000.

[9] M. Baugher, "The Secure Real-time Transport Protocol", *Internet Draft,* draft-ietf-avt-srtp-09.txt.

[10] G. B. White and et.al. *Computer System and Network Security*, CRC Press Inc., 1996.

[11] R.Heady and et.al., "The Architecture of a Network Level Intrusion Detection System", *Technical Report CS90-20*, Department of Computer Science, University of New Mexico, Aug. 1990.

[12] J. Rosenberg and et.al. "SIP: Session Initiation Protocol", *RFC 3261.*

[13] X. Chen and et.al. "Survey on QoS management of VoIP", 2003 *IEEE International Conference on Computer Networks and Mobile Computing*, Shanghai, China, Oct. 2003.

[14] NIST Special Publication on Intrusion Detection Systems.

[15] M. Muller, "Intruder Scenarios in Telecom Network", http://www.tml.hut.fi/Opinnot/Tik-110.501/1999/papers/scenarios/scenarios.html

[16] A. Chakrabarti, and et.al "Internet Infrastructure Security: A Taxonomy", *IEEE Network*, Nov/Dec 2002.

[17] J. Loughney and et.al., "Security Considerations for SIGTRAN Protocols", Internet Draft, draft-ietf-sigtran-security-03.txt.

[18] T. Dierks and et.al, "The TLS Protocol", *RFC 2246.*

[19] S. Kent and et.al, "Security Architecture for IP", *RFC 2401.*

[20] R Oppliger, "Security at Internet Layer", *Computer*, vol: 31(9), September 1998.

[21] R. K. Bhattacharyya, "New Challenges for Telephone Companies to Secure Switching Systems", *The 25th IEEE International Carnahan Conference on Security Technology*, Taipei. Taiwan, Oct. 1991.

[22] "Services and Protocols for Advanced Networks: Preliminary Analysis of Broadband Multimedia Services", ETSI Technical Report, ETSI TR 102 99 V1.1.1 (2003-10).

[23] C. Jennings, "Example Call Flows Using SIP Security Mechanisms", *Internet Draft*, draft-jennings-sip-sec-flows-00.txt, October 2003.

[24] http://www.tml.hut.fi/Studies/T-110.551/2003/papers/15.pdf

[25] S. Knuutinen, "Session Initiation Protocol Security Considerations",http://www.tml.hut.fi/Studies/T-110.551/2003/papers/

[26] S. Salsano, "SIP Security Issues: The SIP Authentication Procedure and Its Processing Load", *IEEE Network*, Nov/Dec. 2002.

[27] Signaling Transport (sigtran) Working Group site: http://www.ietf.org/html.charters/sigtran-charter.html

[28] F. Cuervo, and et.al, "Megaco Protocol Version 1.0," *RFC 3015.*

[29] B. Goode, "Voice over Internet Protocol", *Proc. of the IEEE*, Vol.9, No.9, Sept.2002.