Security Challenge and Defense in VoIP Infrastructures

David Butcher, Member, IEEE, Xiangyang Li, Member, IEEE, and Jinhua Guo, Member, IEEE

Abstract—Voice over Internet Protocol (VoIP) has become a popular alternative to traditional public-switched telephone network (PSTN) networks that provides advantages of low cost and flexible advanced “digital” features. The flexibility of the VoIP system and the convergence of voice and data networks brings with it additional security risks. These are in addition to the common security concerns faced by the underlying IP data network facilities that a VoIP system relies on. The result being that the VoIP network further complicates the security assurance mission faced by enterprises employing this technology. It is time to document various security issues that a VoIP infrastructure may face and analyze the challenges and solutions that may guide future research and development efforts. In this paper, we examine and investigate the concerns and requirements of VoIP security. After a thorough review of security issues and defense mechanisms, we focus on attacks and countermeasures unique to VoIP systems that are essential for current and future VoIP implantations. Then, we analyze two popular industry best practices for securing VoIP networks and conclude this paper with further discussion on future research directions. This paper aims to direct future research efforts and to offer helpful guidelines for practitioners.

Index Terms—Computer attacks, network security, voice over Internet Protocol (VoIP).

I. INTRODUCTION OF VOIP TECHNOLOGY

Voice over Internet Protocol (VoIP) is a technology that allows users to make telephone calls using a broadband Internet connection instead of an analog phone line. VoIP holds great promise for lowering the cost of telecommunications and increasing the flexibility for both businesses and individuals. VoIP leverages existing IP-based packet-switched networks to replace the circuit-switched networks used for voice communications since the invention of the telephone.

The VoIP infrastructure consists of endpoints (telephones), control nodes, gateway nodes, and the IP-based network. The IP network can utilize various media including Ethernet, fiber, and wireless. The VoIP system interacts with both local and remote VoIP phones using the intranet and Internet as well as interacting with phones connected to the public-switched telephone network (PSTN) through gateways. Fig. 1(a) illustrates a simple VoIP setup.

The VoIP data processing consists of the following four steps: signaling, encoding, transport, and gateway control [1].

1) Signaling: The purpose of the signaling protocol is to create and manage connections or calls between endpoints. H.323 and the session initiation protocol (SIP) are two widely used signaling standards for call setup and management.
2) Encoding and Transport: Once a connection is setup, voice must be transmitted by converting the voice into digitized form, then segmenting the voice signal into a stream of packets. The first step in this process is converting analog voice signals to digital, using an analog-to-digital converter. Here a compression algorithm can be used to reduce the volume of data to be transmitted. Next, voice samples are inserted into data packets to be carried on the Internet using typically the real-time transport protocol (RTP). RTP packets have header fields that hold data needed to correctly reassemble the packets into a voice signal on the other end. Lastly, the encapsulated voice packets are carried as payload by the user datagram protocol (UDP) for ordinary data transmission. At the other end, the process is reversed: the packets are disassembled and put into the proper order, and then the digitized voice is processed by a digital-to-analog converter to render it into analog signals for the called party’s handset speaker. Fig. 1(b) illustrates the basic flow of voice data in a VoIP system.

3) Gateway Control: The IP network itself must then ensure that the real-time conversation is transmitted across the telephony system to be converted by a gateway to another format—either for interoperation with a different IP-based multimedia scheme or because the call is being placed onto the PSTN.

With the switch to the Internet as a carrier for voice traffic, we see some of the same security issues that are prevalent in the circuit-switched telephone network such as eavesdropping and toll fraud. We are also exposed to new types of attacks that are more prevalent in the data world of the Internet such as the denial-of-service (DoS) attacks.

II. GENERIC SECURITY CONCERNS IN VOIP

A. Simple Security Taxonomy

Computer security has been a field attracting vast interests from researchers and practitioners in information technology disciplines. Several books cover the extensive topics in information security and assurance [2]–[6]. To prepare with a general knowledge map for the discussion in this paper, we can describe the various components in three dimensions. First, security constraints include system assets, vulnerabilities, and threats. The security constraints are the items that make security a problem for any information infrastructure. Second, security requirements are mechanisms, services, and policies to put in place to counteract the security problems. They can either eliminate threats or fix vulnerabilities. Third, security management makes various decisions on tasks and operations regarding security tools, standards, and legal regulations to support specific business functions and processes.

Each of the components in a VoIP system offers potential vulnerabilities for attackers to exploit. Many of the vulnerabilities are similar to those of the public-switched telephone system. For example, in the switched telephone system, eavesdropping can be achieved by physically placing a listening device on the phone line. Physical access to the transmission control protocol (TCP)/IP network can also allow an attacker to place a listening device (packet grabber) on the network and intercept phone conversations. Other vulnerabilities are similar to those of the data network. The control and gateway servers within the VoIP infrastructure are typically built on existing computing platforms (e.g., Linux, Windows). These platforms are constantly under attack, and these attacks would also apply to the VoIP network [7], [8]. Packet networks depend for their successful operation on a large number of configurable parameters: IP and media access control (MAC) (physical) addresses of voice terminals, addresses of routers and firewalls, and VoIP-specific software such as call processing components (call managers) and other programs used to place and route calls. Many of these network parameters are established dynamically every time network components are restarted, or when a VoIP telephone is restarted or added to the network. Because there are so many places in a network with dynamically configurable parameters, intruders have a wide array of potentially vulnerable points to attack.

Security measures are defined and categorized to describe security problems, such as the Central Intelligence Agency (CIA) triplets of confidentiality, integrity, and availability. Each security concern can be classified regarding its compromise on confidentiality, integrity, or availability of the VoIP system. Confidentiality threats generally expose the content of the conversation between two parties, but could also include exposure of call data (telephone numbers dialed, call durations). Integrity threats impact the ability to trust the identity of the caller, the message, the identity of the recipient, or the call record logs. Availability threats jeopardize the ability to make or receive a call. Table I summarizes generic security concerns in a VoIP system and their impact using these three measures.

Attackers are motivated by a variety of factors including free phone service, free long-distance service, impersonation to commit fraud, eavesdropping to commit fraud, and disruption of service to gain notoriety. A number of vulnerabilities in VoIP communications are explored. Many of these vulnerabilities can be exploited using a variety of attack vectors. For many of these attacks, there exist countermeasures that are available or are being developed. Based on the analysis of security constraints in a computer network, we can employ a variety of tools and policies to reduce security risks. Information assurance normally describes operational tasks of threat prevention, detection, and reaction, and evaluation of the security level of the resulting information system.

---

**TABLE I**

<table>
<thead>
<tr>
<th>Security Concern</th>
<th>Confidentiality</th>
<th>Integrity</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denial of Service</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eavesdropping</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Alteration of Voice Stream</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Toll Fraud</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Redirection of Call</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Accounting Data Manipulation</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Caller ID Impersonation</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Unwanted Calls and Messages</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
This paper, therefore, will tackle the VoIP security topic with a focus on security constraints and requirements. More specifically, the previously mentioned above set of security concerns or threats and their impact are discussed in the rest of this section. After this, we will closely examine features and mechanisms of each threat that exploits these concerns and the corresponding countermeasure.

B. DoS—Availability

DOS is an attempt by an attacker to prevent the phone service from operating within the normal operating specifications. This could include the inability to place a call or receive a call. This could be a general attack preventing all calls, or a more selective attack preventing calls to certain addresses (phones numbers) [9]. The motivation for such attacks could be to disrupt business or prevent communication to facilitate another event (e.g., delay an emergency response to a robbery).

These attacks focus on compromising the availability of the system. Availability is one of the cornerstones of the PSTN. Frequently, the switched telephone network is cited as the classic example of availability at five nines (99.999%) or Six-Sigma. Ensuring this level of availability will be important to the widespread adoption of the VoIP [10].

C. Eavesdropping—Confidentiality

Eavesdropping is when a victim’s conversation is secretly monitored by the attacker. This typically involves receiving the voice data from both parties in the call. This data is then used to replay the conversation and use the contents for illicit purposes [11].

Not only are voice conversations subject to eavesdropping but also data type conversations that are carried by the phone system. This includes receiving fax data that is being carried by the VoIP system to obtain a copy of a document and receiving dual-tone multifrequency (DTMF) data to obtain bank and credit card passwords used during interactive voice response sessions.

Reasonable security is expected of the phone system and the conversations that it carries. In the switched telephone network, it is fairly difficult to tap into a call at any place other than the last mile analog circuit due to the use of private networks [10]. This gives most users of the traditional phone system a good level of comfort that their conversation is secure. A similar level of security is mandated for VoIP calls to maintain user’s acceptance of the technology.

D. Alteration of Voice Stream—Confidentiality and Integrity

This is a substitution attack or man-in-the-middle attack. The attacker is able to listen to the conversation between the two victims and also alter the communication. This includes playback of previously captured speech so that the receiver hears a different message than the sender sent.

Due to the unpredictable nature of human conversations, this attack may be difficult to use for wholesale alteration of conversations between two parties. This could be more easily used to change very small portions of a conversation. For example changing “no” to “yes” in response to a question of participation or “sell” to “buy” in a conversation with a financial advisor would severely impact the meaning of a conversation.

This also has value to an attacker when used with interactive voice response phone systems. After capturing a victim’s financial passwords, and depleting funds from an account, the substitution attack could be used when the victim calls the interactive voice response system to check balances. The attacker could playback a previous balance to give the victim the impression that no funds had been removed from an account.

E. Toll Fraud—Integrity

Toll fraud is the attempt by an attacker to receive money by having a large volume of calls placed to a phone number that has a large fee associated with connecting with that number. An example would be to have a phone system open numerous calls to a “900” number that would result in high charges back to the owner of the phone system. These fees would then be split between the owner of the “900” number and the attacker.

Another prevalent attack is the impersonation of a phone to obtain free long distance. The attacker spoofs the system into thinking that their phone is another legitimate phone. The attacker then uses their “cloned” identification to make numerous calls. The charges for the calls are then passed on to the victim [12].

Toll fraud is prevalent within the traditional phone system. Phone “phreaking” and calling card theft have been traditional means for attackers to obtain free long distance within the switched telephone network system. Where holes in security cannot be directly closed, elaborate systems have been constructed to detect patterns of abuse and theft to stop losses early. These systems are similar to intrusion detection systems (IDSs) used in data networks.

F. Redirection of Call—Integrity and Confidentiality

One of the design goals of the VoIP system was flexibility and rich feature sets. The ability to have a single phone number redirected to whoever the owner is present is an advanced feature that gives the caller an easy way to find the person they are looking for by dialing a single phone number.

This rich feature becomes a potential risk if the redirection feature becomes compromised by an attacker. The attacker can then redirect the victim’s phone number to a location of their choice, thus potentially being able to impersonate the victim by having their calls redirected to the attacker’s telephone [13].

Other than the simple call forwarding, this type of feature has not been available in the traditional phone system. Numerous new features available in VoIP telephone systems will provide a rich set of tools for the end users to gain productivity and also for attackers to exploit [14].

G. Accounting Data Manipulation—Integrity and Confidentiality

The accounting database contains entries for each call placed through the system. These call data records (CDR) contain
information about the numbers the call was placed from and to, the time of day, duration, and other information about the call [15].

By gaining access to the CDR database, the attacker can view call patterns. By analyzing call patterns, the attacker could gain insight into confidential matters. An example would be calls being frequently placed from an executive of a company to an executive at another company. The attacker may be able to use that information to determine that a merger or strategic alliance is being formed.

If the attacker gains write access to the CDR database, it is possible to manipulate or delete call records. The motivation for this could be to avoid paying for service (toll fraud) or to cover up other more serious criminal activity. Removing the call records or modifying the phone numbers that the calls were placed from and to would provide a clear mechanism for hiding many activities that are documented through call records.

This type of vulnerability is also found in the switched telephone network. The traditional phone network has the advantage that the voice network is separate from the data network that the accounting database resides upon. The accounting database in the traditional system would likely be on a private network with physical separation from any public networks. VoIP-based telephone networks have the data network as the underlying transport for both the voice communications and the call data. Protecting the call logs within the VoIP system needs extra attention when compared to the traditional system.

H. Caller Identification (ID) Impersonation—Integrity

Each phone has an identity (phone number) that is associated with the device. Having a device impersonate the identity of another can be used as an attack to either receive calls or place calls with the spoofed identity.

An attacker wishing to impersonate someone would setup their VoIP phone device to use the identity of the victim’s device. The attacker’s device then registers with the phone system. Any calls intended for the victim’s phone number would then be directed to the attacker’s phone. The attacker could then answer a call and impersonate the victim. The attacker also has the ability to use the spoofed device to place calls. The caller ID at the phone receiving the call would indicate that the victim is calling, not the attacker. This can also be used by the attacker to impersonate the victim.

Caller ID information has become relied upon for many purposes. Many banking applications now use caller ID to verify the identity of the customer. An attacker spoofing the ID of a victim’s phone could use this to gain access to banking or credit card information.

The traditional phone system is minimally susceptible to this type of attack. The centralized configuration of phone lines in the switched telephone network prevents the spoofing of caller ID information for most lines. This general acceptance that the call ID can be trusted has led to the development of applications that rely on caller ID to establish the identity of the person. This same level of integrity needs to be maintained in the VoIP-based phone system to be accepted by the general public.

### Table II

<table>
<thead>
<tr>
<th>Layer</th>
<th>Attack Mechanism</th>
<th>Confidentiality</th>
<th>Integrity</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>Physical Attack</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Network</td>
<td>ARP Cache Poison</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>MAC Spoofing</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data Link</td>
<td>Device IP Spoofing</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Malformed Packets</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>TCP or UDP Floods</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>TFTP Server Insertion</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHCP Server</td>
<td>DHCP Starvation</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICMP Floods</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buffer Overflow</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Operating System</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Viruses and Malware</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Database Attacks</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SIP Attacks</td>
<td>Registration Hijacking</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Message Modification</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Cancel/Bye Attack</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malformed Command</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redirect</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RTP Attacks</td>
<td>RTP Payload</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RTP Tampering</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

I. Unwanted Calls and Messages (SPIT)—Availability and Integrity

The ability to bulk mail millions of solicitations at little or no cost via email brought the phenomena known as SPAM. Telemarketing could likewise take advantage of the deployment of VoIP phone systems to send out bulk voice solicitations known as SPAM over Internet telephone (SPIT) [16].

The attacker sets up a farm of servers that have lists of VoIP phone addresses (phone numbers). These servers then connect to the addresses and deliver messages at a high volume. These messages are either played through the victim’s phone or fill up the victim’s voice mail box.

Telemarketing has proven to be an annoyance in the switched telephone network. Legislation has been written to option out traditional telephones from telemarketing lists. The combination of lower cost, higher throughput, and lack of legislation has put VoIP on a sure course for exploitation by telemarketing. An explosion in SPIT will surely curtail the growth of VoIP adoption.

III. Attack Vectors in VoIP and Critical Challenges

The threats to the VoIP system can be further broken down into specific attack vectors to disrupt the system, as detailed in Table II and summarized by the system layer where the attack occurs. The “Layer” column in this table tells about the location of the attack occurrence within the system structure of a data communication network. Of course, other classifications of these attack mechanisms can position them to the locations of end nodes (handsets or computers), servers, or other network locations. However, the layer in this table represents a more accurate classification compared to such a location classification. For example, an attack mechanism exploiting the OS
vulnerabilities can happen at the terminal, server, or other network locations. However, this type of attack all happens at the application layer.

We will focus on the attacks that are specific to VoIP applications and critical to VoIP security assurance, shown shaded in the table. We briefly discuss attacks general to IP data networks and PSTN networks at the end of this section.

A. SIP Registration Hijacking

The SIP is an application layer control protocol that can establish, modify, or terminate user sessions. In SIP and other VoIP protocols, a user agent (UA)/IP phone must register itself with an SIP proxy/registrar (control node), which allows the proxy to direct inbound calls to the phone. Registration hijacking occurs when an attacker impersonates a valid UA to a registrar and replaces the legitimate registration with its own address. This attack causes inbound calls intended for the UA to be sent to the rogue UA.

Registration hijacking can result in loss of calls to a targeted UA. This may be an individual user, group of users, or a high-traffic resource, such as a media gateway or voice mail system. By hijacking calls to a media gateway, all outbound calls can be blocked or otherwise manipulated. A rogue UA acting in the middle can also record call contents.

Currently, UDP and TCP are used to carry the registration information between the UA and the control node. Utilizing transport layer security (TLS) to create an authenticated secure connection in place of the open connection will prevent SIP registration hijacking.

B. SIP Message Modification

SIP message have no built-in integrity mechanism. By executing one of the man-in-the-middle attacks (IP spoofing, MAC spoofing, SIP registration), an attacker can intercept and modify an SIP message, changing some or all of the attributes of the message.

This could include the person being called in a session initiation message, giving the victim the impression that he was calling one person while the system connects them to another. By modifying the SIP message, the attacker could impersonate a caller or reroute a call to an unintended party.

By protecting UDP and TCP transport mechanisms with TLS, the contents of the SIP message are protected. This would prevent an attacker from both reading an SIP message and being able to know who was entering into a call as well as modifying the message to create call fraud.

C. SIP Cancel/Bye Attack

The attacker can create an SIP message with the Cancel or Bye command in its payload and send it to an endnode (phone) to terminate an ongoing conversation. If the attacker sends a steady stream of these packets to the phone, the phone will not be able to place or receive calls. This could be expanded to numerous phones creating a systemwide disruption of service.

The lack of strong authentication in SIP makes this attack possible. By adding strong authentication to the communication between the UA and the control node, this type of attack can be prevented. The UA would verify that the incoming Cancel or Bye command was coming from a trusted node using certificate-based credentials.

D. Malformed SIP Command

The SIP protocol relies upon an hypertext markup language (HTML) like body to carry command information. This makes the SIP protocol very flexible and extensible for implementing VoIP features. The downside is that it becomes very difficult to test the SIP parser with every possible input. Attackers can exploit these vulnerabilities as they find them by forming packets with malformed commands and sending them to susceptible nodes. This will either degrade or decommission the node that is attacked making part or all of the VoIP system unavailable [17].

This is difficult to fix from the perspective of error proofing the message parser. The numerous bugs and their resulting exploits seen in Internet browsers show us how difficult it is to test every possible message that may be sent. Testing with both a dictionary of test cases as well as testing with fuzzing should be performed. Fuzzing software provides intelligently malformed requests that attempt to drive out bugs in software that parses HTML-like grammars.

Adding strong authentication to the VoIP SIP command system will also help by preventing an attacker from being able to send malformed SIP commands to a node. This then leaves the authentication protocol as the potential attack point. If there are weaknesses in the parsing logic for the authentication, this attack may still be successful.

E. SIP Redirect

SIP employs a server application that receives requests from a phone or proxy and returns a redirection response indicating where the request should be retried. This allows a person to have a call made to them ring at a different phone depending on where they are located, but the caller only dials a single number to reach the person. By attacking the redirect server and commanding it to redirect the victim’s calls to a number specified by the attacker, the attacker can receive calls intended for the victim. If the attacker wishes to disable the phone network, they could redirect all users’ phone numbers to a nonexistent or null type of device. All calls within the VoIP system would then be routed to this nonexistent extension effectively disabling the phone system from delivering calls.

Once again, this vulnerability is built upon the lack of strong authentication in the SIP protocol. Moving to a more robust authentication system such as TLS with strong passwords will protect against attacks such as the SIP redirect.

F. RTP Payload

The RTP protocol carries the actual encoded voice message between the two callers. It is a simple extension of the UDP protocol adding sequencing information. Using a man in the
middle attack to gain access to the RTP media stream between two nodes, an attacker can inspect or modify the payload of the message.

Inspection in this case becomes eavesdropping on the conversation. If an attacker can modify the payload of messages, they can either inject noise or their own message into the packet. This would either degrade or make impossible conversation between the parties on the call in the case of noise, or potentially alter the meaning of the conversation.

By utilizing the secure RTP (SRTP) protocol this type of attack can be prevented. The RTP packets are encrypted by the sender and travel the entire network encrypted until being decrypted by the receiver. The SRTP approach prevents eavesdropping and modification of packets to contain new messages.

G. RTP Tampering

By manipulation of the sequence number and timestamp fields in the header of the RTP packet, the packets can either be re-sequenced or made unusable. This attack can either make the conversation unintelligible, or in some implementations of the protocol stack, actually crash the node receiving the packets, thus taking the node offline until the software is recycled.

The SRTP protocol will allow the receiving node to determine that the RTP header has been modified. This will prevent unusual behavior from the application software as the packet will be discarded prior to being processed [18].

Maintaining the VoIP traffic on a local area network (LAN)/virtual LAN (VLAN) separate from the non-VoIP traffic will help prevent this type of attack from occurring. Separating the VoIP traffic from the data network makes it increasingly difficult to obtain access to the VoIP traffic, and thus monitoring or modifying the traffic becomes more difficult.

H. Other Attacks General to IP Data Networks

1) Physical Attack: By physically modifying portions of the VoIP network, such as trunk links, handsets, VoIP servers, etc., the attacker can greatly impact availability and confidentiality of the system. The servers, switches, and cabling that make up the VoIP system should not be accessible to anyone except those with administrative duties related to the system by means of locking server rooms, wiring closets, and having physical barriers to areas where backbone areas of the network are present.

2) Address Resolution Protocol (ARP) Cache Poison: By sending forged ARP packets, the attacker can make an association of the attacker’s MAC address and another IP address in the ARP cache of the victim node. This attack can allow the attacker to masquerade as either a control node or an endpoint within the VoIP system [19]. To prevent these attacks, dynamic ARP inspection (DAI) intercepts every ARP packet on the switch, and verifies valid IP-to-MAC bindings before updating the local ARP cache or forwarding them to the appropriate destination.

3) MAC Spoofing: MAC spoofing brings a new node onto the network with a duplicate MAC address. This allows the attacker’s node to appear as another node that has already been configured and authorized into the VoIP system, when the node owning the MAC address is off the network (turned off or disconnected) or unavailable under an attack. One of the best countermeasures is not allowing new nodes without first authenticating with the port that is providing connectivity to the network, such as port authentication detailed in the IEEE 802.1x standard.

4) IP Spoofing: An attacker gains unauthorized access to a computer or a network by making it appear that a malicious message has come from a trusted machine by “spoofing” the IP address of that machine. This technique can be used to impersonate either a control node or an endnode in the VoIP network. To prevent IP spoofing, routers should be configured to not allow incoming packets with source addresses that belong within the local domain. The router should also be configured to disallow packets to be routed out of the local domain that do not contain source addresses that are in the local address range.

5) Malformed Packet: The attacker explicitly sends packets constructed with a flaw in the network protocol to nodes within the system. Processing the packet unleashes the flaw in the protocol stack and either degrades or disables the nodes’ ability to handle further traffic for VoIP calls. Preventing attacks using malformed packets should take a two-pronged approach. All nodes should be kept up-to-date with the latest software and patches. Modern high-capability firewalls also provide a defense against malformed packets with features to deny the passing of malformed packets into the local domain.

6) TCP SYN Flood: The attacker generates a large number of packets with random source addresses and the TCP SYN flag set, requesting allocation of a buffer at the receiving node. After the entire buffer space is exhausted, legitimate users cannot make connection with the victim machine. The VoIP signaling protocols (H.323 and SIP) rely on the TCP transport for communication between nodes. Effective countermeasures include properly configured firewalls, TCP implementations using SYN cookies, or sending a SYN-ACK with a carefully constructed sequence number before buffer allocation.

7) TCP or UDP Replay: From captured packets, the hacker can extract information like device authentication information, voice conversations, or DTMF (touchtone) information. Then, the captured data can be placed back on the network for a new connection. This allows the attacker to register a device with the system using another device’s identification to place and receive calls, or to play back portions of a voice or data call. The best countermeasure method is to encrypt the sessions including a unique sequence number. The receiving node decrypts the packet and checks the sequence number to see if it is the next expected value.

8) Trivial File Transfer Protocol (TFTP) Server Insertion: Endnodes (VoIP telephone handsets) are generally configured to locate a TFTP server for configuration information including updating software on the device. If an attacker
is able to establish a bogus TFTP server on the network, nodes will potentially receive hoax configuration such as the phone number and ID of another handset and cause billing fraud [20]. Utilizing an authenticated connection to the configuration server, such as TLS using secure socket layer (SSL) on top of TFTP will prevent this type of attack. The end node needs to be preconfigured with the certificate information for the server.

9) **Dynamic Host Configuration Protocol (DHCP) Starvation:** By flooding the local network with DHCP requests for randomly generated MAC addresses, an attacker can deplete the available pool of IP addresses in the DHCP server. Then, this attack will prevent a node from obtaining an IP address and subsequently contacting any of the other VoIP servers within the system. The IEEE 802.1x specification provides a mechanism where a node attached to a network port must authenticate its MAC address prior to being able to transmit or receive traffic on the network. This includes the DHCP request, thus preventing the attack.

10) **Internet Control Message Protocol (ICMP) Flood:** Similar to a TCP or UDP flood, by overwhelming a node with a large volume of incoming ICMP packets, the performance of the node is either degraded or taken offline. Degradation is just as effective at bringing down the VoIP system as its time sensitivity dictates that nodes must process packets as soon as they are received. Routers can be configured to block large or unneeded ICMP packets, to restrict ICMP fragments, and to match ICMP requests and responses. The VoIP system should also be on a separate LAN/VLAN guarded by a firewall with the same type of settings, against attacks originating from internal desktop personal computers (PCs).

11) **Buffer Overflow Attack:** These attacks exploit flaws in software that attempts to store more data in a buffer than it was intended to hold. The data then overflows into adjacent buffers or code to hijack the program for the use by the attacker. Buffer overflow attacks could be used against the phone devices and the control nodes in a VoIP network. To mitigate the risk of buffer overflow attacks, the nodes based upon common platforms (Windows, Linux) should be maintained with current patches.

12) **OS Attack:** Many known and unknown vulnerabilities exist in operating system (OS) platforms where the control nodes and gateways in a VoIP network reside with. VoIP phones can be deployed as softphones on desktop or notebook personal computers. To mitigate the risks of these types of attacks, the platforms should be kept current with OS patches issued by the platform vendors and hardening of the platforms should be considered for the control and gateway nodes.

13) **Viruses and Malware:** Just as the vulnerabilities in the OS platforms can be exploited by an attacker, so can susceptibility to viruses and malware. Viruses and malware that infect the nodes of the VoIP network can degrade or eliminate the nodes' ability to handle call traffic [7]. In addition to keeping OS patches current and hardening servers as mentioned earlier, antivirus software should be installed and kept up-to-date on all nodes.

14) **CDR Database Attack:** CDR are stored in commercial databases by the control nodes to log call activity. Known vulnerabilities for commercial database software such as Microsoft’s SQL Server can be used to exploit the call logs. Care should be put into the design of the network to isolate the database from the signaling traffic by placing the database on a separate LAN or VLAN. This separation prevents an attacker from being able to send or receive data from the database server.

### IV. Defense Vectors in VoIP and Key Research Topics

In this section, we examine the most important defense mechanisms. Many of them protect against multiple attacks. A number of defense mechanisms are best practices for networks, in general, even when VoIP is not being deployed. Table III lists the defense mechanisms and the attacks that the mechanism can prevent.

It should be noted that port authentication has a wide sweeping impact on preventing numerous types of attacks [21]. Port authentication utilizes the 802.1x protocol standards to require a device being connected to the network (either wired or wireless) to authenticate against a central authority before being granted access to the network. This prevents rogue devices from being added to the network. This has the potential for preventing many types of attacks as the attacker cannot gain access to the network for their node. The benefits of this become diminished when VoIP traffic leaves the protected confines of the enterprise and enters the Internet. However, experience has shown that many attacks come from within the walls of an organization.

Because of the limit on the available space of this paper, we do not cover intrusion detection in this paper although it is an important line of defense to complement intrusion prevention. Moreover, the survey and analysis of intrusion detection in VoIP can be a separate paper by itself. There have been many studies and IDSs for data networks [2], [22], [23]. Generally all these systems can be used in the VoIP network. Of course, the special attacks on VoIP protocols discussed earlier raise unique requirements on intrusion detection.

We focus to explore the key defense vectors specific to VoIP applications, shaded in Table III. The following mechanisms are beginning to emerge in commercially available VoIP products, and deserve special research efforts in the future.

#### A. Separation of VoIP and Data Traffic

Similar to the port authentication, separating voice and data traffic on to different networks is a key enabler to overall security. Separating the traffic can prevent a number of attacks as PCs and workstations cannot be used by attackers as an easy entry into the VoIP network. Due to the expense of running two separate physical networks, this separation is performed using VLAN technology.

The switches of the network implement VLANs by only allowing routing between devices on the same VLAN as
TABLE III
SECURITY MECHANISMS AGAINST ATTACKS

<table>
<thead>
<tr>
<th>Attack Prevented</th>
<th>Physical Attack</th>
<th>ARP Cache Poison</th>
<th>MAC Spoofing</th>
<th>IP Spoofing</th>
<th>Malformed Packets</th>
<th>TCP/UDP Flood</th>
<th>TCP/UDP Reuse</th>
<th>UDP Port Scan</th>
<th>FTP Server Scan</th>
<th>DNS Cache Poison</th>
<th>ICMP Flood</th>
<th>SP Register Flood</th>
<th>SP Message Modification</th>
<th>SP Cancel Flood</th>
<th>SP Malformed Command</th>
<th>SP Reflect Attack</th>
<th>SP Malformed Header</th>
<th>SP Malformed DNS</th>
<th>SP Reflect DNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Access Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic ARP Inspection (DAI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS Protection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port Authentication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router Configuration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firewall Configuration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separate VoIP and Data traffic (VLANs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configuration Authentication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signaling Authentication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media Encryption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Segmentation of VoIP voice and data using firewalls.

configured by the network administrator. VoIP telephones that implement a second LAN port for a data connection for a PC should implement VLAN technology (802.1q) such that the PC connected to the phone is placed on the data LAN, not the voice LAN.

Some connectivity between the data and voice LAN will be required. Voice mail servers are typically maintained on the data segment of the network. The VoIP call controlling servers will need controlled connectivity to the voice mail servers. To implement this controlled connectivity between the two LAN segments, an SIP-aware stateful firewall should be installed in between the segments. Fig. 2 illustrates the segments with the connecting firewall.

B. Configuration Authentication

The VoIP phones need basic configuration information to get into the VoIP system. Configuring the phones provides a classic bootstrap problem where obtaining configuration information from an untrusted source can build into further problems.

Preconfiguring handsets with the public key of various configuration servers at the time of manufacture provides a possible mechanism for authenticating the configuration server. An alternative would be for the installer of the phone to configure the phone with a public key or shared secret for the configuration server. This could be automated to provide the installer with a hardware key or device that would be attached to the phone and provide a quick and accurate means of copying the public key into the phone.

To obtain the phone’s configuration, the phone would make a DHCP request. In the reply from the DHCP server, the phone obtains both its IP address and the IP address of its configuration server. The phone would then establish a connection with the configuration server using the TLS. During the TLS handshake, the authenticity of the server would be established using the public key the phone device contains and the private key contained in the configuration server. If the two were not a pair, the phone would not load configuration information from the server. If the key pair matches, the configuration information would be loaded into the phone using the FTP over the secure TLS transport (Fig. 3).

C. Signaling Authentication

The connection between the phones in the VoIP network and the servers utilizes the SIP protocol. When a phone registers with the SIP server, the phone provides an identity. This identity is based upon a number of identifiers including MAC and IP addresses of the phone. Safeguards are put into place to help minimize the ability of an attacker to spoof the MAC address or IP address of a phone, but this cannot be totally eliminated.

The IP security (IPSec) protocol provides mechanisms for both authentication and encryption. Utilizing IPSec between the VoIP phone and the call manager server provides a strong authentication mechanism. Key sharing provides the basis for
the trust between the phone and the server. Key sharing has always been a difficult problem to solve for large-scale deployments [24].

IPSec provides three different mechanisms for establishing keys. Manual entry into both hosts is the least desirable. The manual entry would require a technician to visit each phone and configure it with its keys. This could be automated, but would still prove prohibitive in a large-scale deployment. The second mechanism is to utilize a certification authority (CA). The third choice is to utilize the DNS secure (DNSSec) protocol.

If the use of TLS-authenticated and encrypted configuration is adopted for the phones (see previous section), the configuration server can be utilized as the key distribution channel for the phones. When the configuration request for the phone is made, it would receive a set of keys used to setup the IPSec authentication with the call manager server.

IPSec consists of a number of related protocols. The authentication header (AH) protocol is used to provide connectionless integrity and data origin authentication. The protocol also provides an option that prevents replays.

By utilizing the IPSec AH protocol between the phone and the server, the integrity of the call can be maintained. This includes the identity of the caller, the number being called, and the authenticity of the call manager. This provides protection from attackers wishing to impersonate a caller or to receive calls intended for another person.

The AH protocol also guarantees the integrity of the message, allowing the receiver to detect that information in the payload has been tampered with. This would prevent an attacker from modifying a call request by changing the destination phone number to one of the attacker’s choice.

The origin authentication also prevents the phone from responding to commands from a spoofed call manager. The Cancel/Bye attack from a spoofed node would be rejected by the target node as the authentication information would not match the key database.

D. Media Encryption

Protecting the contents of the voice conversation from eavesdropping is a basic concern for many utilizing VoIP to conduct business. An extension to the RTP protocol called the secure real-time protocol (SRTP) has been published by the Internet Engineering Task Force (IETF) as RFC 3711. SRTP provides both authentication and confidentiality services for the payload being carried by the RTP protocol. The SRTP protocol has been designed to add a low overhead to the packet size and to minimize the number of key pairs that must be shared between the two communicating nodes [24].

Even with this minimization of key pairs, a single master key pair must exist between the two nodes wishing to communicate. The RFC relies on emerging key exchange protocols for key exchange such as multimedia Internet keying protocol (MIKEY). MIKEY provides for a lightweight (low bandwidth, low computational needs) protocol capable of exchanging keys in an ad-hoc environment like VoIP phone calls [25].

Preshared secret keys (such as those discussed in the signaling authentication section) provide one mechanism for MIKEY to generate session keys. The session keys are used to encrypt the messages sent via the SRTP protocol and containing voice data. A second method of generating session keys relies upon public key cryptography. The nodes utilize asymmetric cryptography techniques to generate and exchange a session key. The SRTP session then proceeds just as with the preshared secret keys and utilizes the agreed key to encrypt the SRTP messages using secret key cryptography. Secret key cryptographic algorithms are always utilized for the SRTP packets as they have a lower compute cost and thus impose less latency on the real-time delivery of the voice data.

In contrast to the immaturity of SRTP, IPSec has been well established and contains a suite of protocols and key exchange algorithms. The Internet key exchange (IKE) protocol provides a mechanism for two nodes to exchange keys. Using other protocols within the IPSec suite, a secure tunnel can be established between the two nodes entering into a voice conversation [26].

IPSec works well in establishing secure tunnels as a trunk between organizations. This allows two secure facilities to be connected via a secure link that runs through a nonsecure network such as the Internet. This approach is frequently used to connect branch offices with a central corporate office.

V. INDUSTRY CASE STUDIES

There are evolving standards and implementation of standards available from different vendors within the VoIP market. Many VoIP equipment manufacturers have put security at a lower
priority than developing other features within their system. They have tended to rely upon best practices of network operations to keep some level of security. This includes those items listed in the upper rows of Table III.

Here, we explore the recommended practices for VoIP network security from two major vendors, Cisco Systems and Nortel. Cisco has grown from a data network provider into a provider of voice networks. Nortel is an incumbent in the switched circuit equipment market that has moved into voice on IP networks. Both have moved into VoIP through strategic acquisitions.

A. Cisco VoIP Systems

Cisco provides a network security reference model for enterprises of various sizes [27]. This model known as SAFE is a basis for their recommendations for securing a VoIP implementation. SAFE provides in-depth security for the data network, servers, and other network devices while providing connectivity with the Internet, remote offices, and business partners. Cisco’s IP telephony security advocates in-depth segmentation of the voice and data networks as key to securing the VoIP implementation, including the following.

1) They encourage the use of VoIP handsets and not the deployment of PC-based IP phones. PC-based phones or softphones must have access to both the data and voice segments by design. Even with the most flexible firewall, protecting a call manager from attacks based from a PC that has access to the voice segment would be difficult.

2) A stateful firewall is required to control the voice-to-data segment interaction. The firewall should provide DoS protection against connection starvation and fragmentation attacks, spoof mitigation, and general filtering.

3) Cisco strongly advocates using a private address space for all IP telephony devices (as defined by RFC 1918). This address space should be separate from any private address space that is being used for the data segment.

4) Cisco stresses that rouge devices on the network pose serious threats. Locking down switch ports and segments is highly recommended. DHCP servers for the IP phone network should be set up to statically assign IP addresses to known MAC addresses.

5) A phone that has an unknown MAC address cannot obtain configuration information from the configuration manager. To prevent an attacker from plugging in a phone with a spoofed MAC address, Cisco recommends using user authentication. This requires that a user log into the phone using either a password or personal identification number (PIN).

It is recommended that traffic be segmented onto separate VLANs; however, segmentation provides minimal attack mitigation by itself. The true value of segmentation is the ability to set up intrusion detection to monitor the voice segment for intrusion [27]. Network IDS (NIDS) should be deployed in front of the call processing manager. This will detect attacks sourced from the data segment. NIDS should also be deployed between the voice and data segment. This will detect DoS attacks against the voice segment.

B. Nortel VoIP Systems

Nortel provides a number of different recommendations for securing the IP phone system depending on organizational needs. The four levels of security are: minimum, basic, enhanced, and advanced. We will review the advanced level here, which builds upon the three lower levels of security. For this security model, it is assumed that trust does not exist within infrastructure and the WAN [28].

1) Nortel recommends device authentication with the use of MAC address security on all switches. This includes the deployment of separate voice and data VLAN segments. The switches should also include intrusion detection monitoring.

2) The voice segment must be reserved for hard clients (IP phone handsets) only. If softphones must be utilized, the PCs running the softphone should be placed on the data segment and communicate to the voice segment through a VoIP-enabled firewall.

3) Deploying a dedicated set of DHCP servers for the voice network is suggested to protect against DoS attacks due to address depletion. As a second level of authorization, static IP assignment is also suggested; this should be based upon known MAC addresses.

4) Nortel advocates creating a secure voice zone (SVZ) using a stateful firewall. Acting as an application-level gateway (ALG) supporting the SIP protocol, this firewall needs to provide SIP proxy services to protect the IP phone servers from attack.

5) All WAN IP phone traffic streams are encapsulated and encrypted inside IPSec tunnels to provide integrity and confidentiality. This virtual private network (VPN) solution protects all voice traffic between the central office and branches or partners. Encrypting the bearer (voice) stream between two IP phones or between an IP phone and the PSTN gateway is also accomplished using IPSec.

6) Like Cisco, Nortel also requires the proper management of all the servers. The server’s configuration should be hardened and all unnecessary software removed. OS patches and hot fixes should be applied on a regular basis [29].

VI. DISCUSSION AND CONCLUSION

VoIP has become a popular solution for voice communication in enterprises of different sizes. This paper examines security issues unique to VoIP systems as well as issues common to traditional IP data networks and compares some of the issues with the security found in the PSTN networks. We have shown that VoIP deployment still faces great challenges regarding malicious attacks and requires numerous countermeasures to migrate these attacks in existing implementation and future development.

Based on the discussion of the challenges and requirements on VoIP in this paper, in our view, future efforts regarding securing VoIP systems should focus on two essential areas, summarized as follows. First, a great deal of research is still needed to strengthen security mechanisms and services to support VoIP at different system levels. Technologies to handle attacks aiming at specific VoIP protocols such as SIP and RTP and various...
components should be given priority including those handling the connection between traditional and VoIP networks. An integrated approach to securing a VoIP system including security mechanisms to secure both the application layer as well as the underlying IP data network can provide a more robust solution. Second, management of security standards and products specific to VoIP and integrating these tightly with IP data network security standards and products will provide an overall solution to securing enterprise data. Although we have witnessed development of standards for VoIP protocols and services, security management of VoIP systems requires continued evolution of these services and protocols to tighten security. Just like IPSec to VPN networks, protocols such as secure RTP are finding their way into VoIP solutions and will continue to be adopted as more individuals and organizations depend on VoIP for their critical communication needs.

REFERENCES