HEMANT SENGAR

IP telephony

BEWARE OF A NEW AND READY-MADE ARMY OF LEGAL BOTS

Hemant Sengar is the co-founder of the vodasec, a voice and data security solution provider to carrier and enterprise networks. His current research interests are in the area of IP telephony and telecommunication network security.

hsengar@vodasec.com

VOICE OVER IP (VOIP), BETTER KNOWN as IP telephony, is aggressively being integrated into the economic and social infrastructure of our lives. But abuse of VoIP will lower people's confidence in this new technology and ultimately hinder its deployment. In this article, I expose a new type of VoIP vulnerability and show how this essential service can be exploited to launch a more potent and stealthy distributed denial-of-service (DDoS) attack affecting both voice and data networks.

IP telephone service providers are moving quickly from low-scale toll bypass deployments to large-scale competitive carrier deployments, thus giving to enterprise networks a choice of supporting a less expensive, single-network solution rather than multiple separate networks. Broadband-based residential customers also switch to IP telephony because of its convenience and cost-effectiveness. In contrast to the traditional telephone system (where the end devices are dumb), the VoIP architecture pushes intelligence toward the end devices (PCs, IP phones, etc.), creating an opportunity for many new services that cannot be envisaged using the traditional telephone system. This flexibility, coupled with the growing number of subscribers, becomes an attractive target to be abused by malicious users.

To break in, attackers may exploit the misconfiguration of devices, the vulnerability of the underlying operating systems, and protocol implementation flaws. Well-known attacks on data networks such as worms, viruses, Trojan horses, and denial-of-service (DoS) attacks can also plague VoIP network devices [1]. Being a time-sensitive service, VoIP is more susceptible to DoS attacks than other regular Internet services. An attacker can disrupt VoIP services by flooding TCP SYN packets, UDP-based RTP packets, or the SIP-based INVITE, REGISTER, etc., messages.

However, if we look at past and current events to identify trends and changes in attacks and targets, then we find that bot-generated DDoS attacks are the most imminent threats to VoIP deployments, as they have been a constant threat to data networks.

The success of bot-generated attacks depends upon two main factors: first, the vastness and diversity of the army of bots, and, second, the bot's distribution over the Internet. Consequently, a bot herder always tries to figure out new ways (such as through worms, Web links, or email attachments) to recruit more hosts into this attacking army. However, there is always a risk of getting caught by law-enforcement agencies for breaking into so many computers. Furthermore, in such digital gang warfare, the rival bot herders may hijack or knock-off these compromised hosts [2], or antivirus scanners may detect and block bot code. But what if, instead of recruiting and compromising new hosts, an attacker finds a ready-made new army of legal bots to launch a more potent and stealthy DDoS attack? This article tries to expose a design error in VoIP services, particularly SIP and the way the INVITE request is used without authentication at the recipient's end of a call. The exploitation of the benign and useful SIP protocol described here deserves our interest for the following reasons: (1) it is new and unexploited; (2) it affects every VoIP telephone, since it is related to the specification rather than the implementation; (3) ironically, many VoIP security devices can also be victimized; (4) it shows a way that a specification can be maliciously exploited.

Background: SIP-based IP Telephony

Session Initiation Protocol (SIP) [3], a standard signaling protocol for VoIP, is appropriately called the "SS7 of future telephony" [4]. It is a text-based application-level protocol to set up, modify, and tear down multimedia sessions with one or more participants. It can also be used to request and deliver presence information as well as instant message sessions. SIP call control uses Session Description Protocol (SDP) for describing multimedia session information. SIP messages can be transmitted over UDP or TCP, but generally UDP is preferred over TCP because of its simplicity and lower transmission delays. However, in some cases, such as the transportation of large SIP messages or the use of TLS, TCP is the only choice.

SIP ARCHITECTURE COMPONENTS

SIP identifies two basic types of components, *user agents* and *SIP servers*. End devices (irrespective of being a softphone or hardphone) are considered *user agents* (UAs). Each UA is a combination of two entities, the *user agent client* (UAC) and the *user agent server* (UAS). The UAC initiates requests, whereas UAS receives requests and sends back responses. Consequently, during a session the UA switches back and forth between a UAC and a UAS. RFC 3261 [3] describes four types of SIP servers, which are implementation-dependent logical entities: *Location Server*, *Redirect Server*, *Registrar Sever,* and *Proxy Server.*

SIP MESSAGES

SIP development is influenced by two widely used Internet protocols: Hypertext Transfer Protocol (HTTP) and Simple Mail Transfer Protocol (SMTP). In SIP, network elements exchange messages as a part of the protocol to set up a call. These messages are classified in two groups: *requests* and *responses*. SIP requests are also called methods and six of them (INVITE, ACK, BYE, CANCEL, REGISTER, and OPTIONS) are described in RFC 3261 [3]. Other methods (in separate RFCs) are also proposed as an extension of the original six methods. Requests are the actions to be taken by UAs or SIP servers. To reply to a request of a UAC, the UAS or SIP server generates a SIP response. Each response message is identified by a numeric status code and, depending upon the range of the numeric status code, there are six different types of responses.

SIP OPERATION

Now we give an example of a typical call setup flow to highlight the usage of SIP request and response messages between user agents UA-A and UA-B. Suppose that the UAs belong to two different domains that each has its own proxy server. UA-A calls UA-B using its SIP phone over the Internet. The outbound proxy server uses the Domain Name System to locate the inbound proxy server of the other domain. After obtaining the IP address of the other proxy server, the outbound proxy server of UA-A sends the INVITE request to the domain of UA-B. The inbound proxy server consults a location service database to find out the current location of UA-B and forwards the INVITE request to UA-B's SIP phone. Exchanging INVITE/200 OK/ACK messages completes the three-way handshake to establish a SIP session [3]. A set of parameters are exchanged via SIP messages (in the message body using SDP) between the two endpoints before an RTP-based voice channel is established. In general, the path of the media packets is independent of that of the SIP signaling messages. At the end of the call, UA-B (or UA-A) hangs up by sending a BYE message. Subsequently, UA-A (or UA-B) terminates the session and sends back a 200 OK response to the BYE message. This example shows the basic functionality of SIP; the detailed description of SIP operation is in RFC 3261 [3].

SIP DEPLOYMENT—PEERING VS. ISLAND-BASED SOLUTIONS

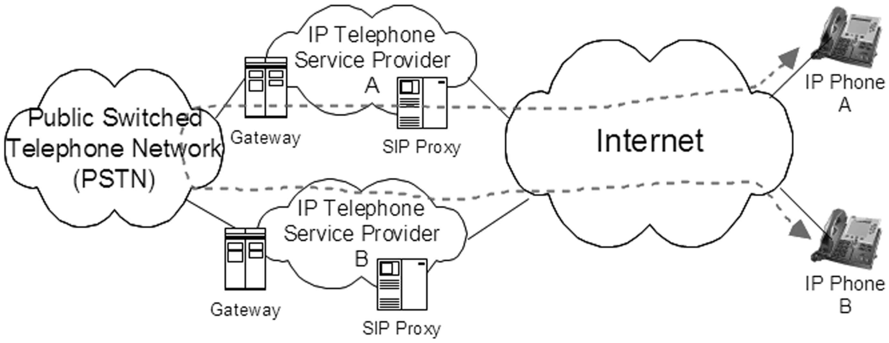

IP Telephone Service Provider A
Public Switched Telephone Network (PSTN)
Gateway
SIP Proxy
Internet
IP Phone A
IP Telephone Service Provider B
Gateway
SIP Proxy
IP Phone B

FIGURE 1: ISLAND-BASED SIP VOIP DEPLOYMENTS

In today's IP telephony world, many of the IP telephone service providers (such as Vonage, AT&T Callvantage, and ViaTalk) operate in a partially closed environment and are connected to each other through the Public Switched Telephone Network (PSTN), as shown in Figure 1. For example, let us assume that user A and user B belong to two different VoIP service providers A and B, respectively. Although the service providers use IP networks to connect with their users, still the calls between users A and B are expected to traverse the PSTN somewhere in the middle. In island-based VoIP deployments the IP traffic is translated into the SS7 traffic [5] (for transportation over the PSTN) and then back into the IP traffic. It is expected that, as VoIP adoption grows, VoIP service providers will interconnect to each other through peering points. Consequently, the calls between any two service providers can be routed through the peering point without traversing the PSTN.

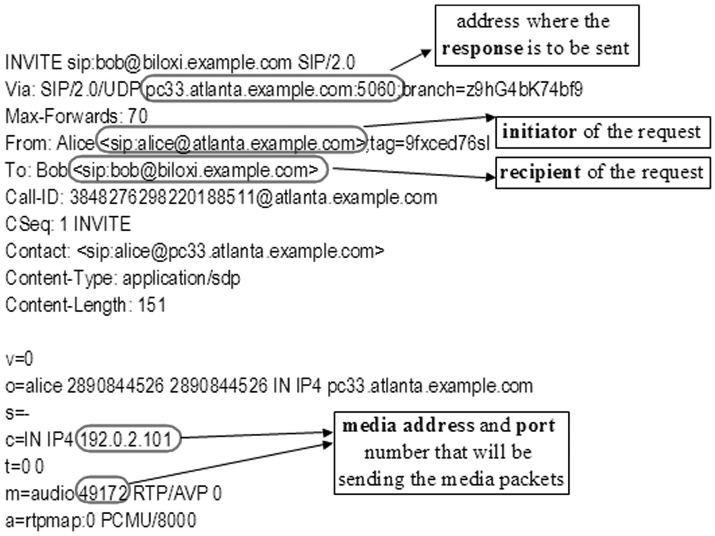
The Threat Model

In a DDoS attack, a number of compromised hosts are used to launch a flooding attack against a particular victim. An attacker installs a daemon on a number of compromised hosts that later on can be requested to start generating spoofed packets directed toward a particular victim target. The enormous number of packets overwhelms the victim's resources, rendering the victim out of service. In the Internet, many network elements such as SIP proxy servers, Web servers, DNS servers, and routers can be defined as *reflectors* because they always respond to some specific type of requests. The attackers can abuse these legitimate and uncompromised reflectors to launch DDoS attacks. The goal of such attacks using reflectors is two-pronged. First, they are used as *stepping-stones,* making such attacks more stealthy so that it is harder to trace back to the actual attacker or real attacking sources. Second, protection becomes difficult, because even if the victimized reflectors are identified, it remains a difficult decision for network administrators to take them out of service, as many legitimate users will also be denied service. However, the use of reflectors is not very lucrative, because a single request generates only one response. Therefore, the number of compromised hosts required to generate spoofed request messages is still large. But what if, instead of a single response, there are a number of response packets for a single request. Such an effect is known as the *amplification* effect. *With the help of reflectors and amplifiers, an attacker can launch a stealthy and more potent DDoS attack using a single machine without possibly compromising any other hosts.*

EXPLOITATION OF THE CALL SETUP REQUEST (INVITE) MESSAGE

Before discussing the exploitation of an INVITE message to achieve both reflection and amplification, we describe its message structure and the purpose of various header fields. As shown in Figure 2, the Via header field contains the address where the caller is expecting to receive the response messages of this request and the From and To header fields contain SIP URIs of the caller and callee, respectively. The Call-ID is a globally unique identifier for this call and the Contact field contains direct route information to reach the caller. The INVITE header fields and message body are separated by a blank line. The session description (media type, codec, sampling rate, etc.) are contained in the INVITE message body. The connection information field (i.e., c=) contains media connection information such as the media's source IP address that will send the media packets. Similarly, the media information field (i.e., m=) contains the media type and the port number.

The exploitation of an INVITE message is based on abusing the *connection information field* contained in the INVITE message body. The SIP proxy server remains at the INVITE header level and routes this message toward the callee without inspecting the message body. The callee's SIP UA parses and interprets the INVITE message and records the media IP address and port number mentioned in the c= and m= fields, respectively. In some cases where the connection information field contains a nonroutable (private) IP address, the SIP UA relies on the *received* parameter of the first Via header field. After a SIP session is established, the callee sends audio packets toward this media IP address and port number. By spoofing the connection information field, an attacker can redirect the media stream toward the spoofed (victim) IP address and port number. In the next section, we give some examples of uncompromised legal bots and the exploitation of the INVITE message.



```
INVITE sip:bob@biloxi.example.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.example.com:5060;branch=z9hG4bK74bf9
Max-Forwards: 70
From: Alice <sip:alice@atlanta.example.com>;tag=9fxced76sl
To: Bob <sip:bob@biloxi.example.com>
Call-ID: 3848276298220188511@atlanta.example.com
CSeq: 1 INVITE
Contact: <sip:alice@pc33.atlanta.example.com>
Content-Type: application/sdp
Content-Length: 151

v=0
o=alice 2890844526 2890844526 IN IP4 pc33.atlanta.example.com
s=-
c=IN IP4 192.0.2.101
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000
```

address where the **response** is to be sent

initiator of the request

recipient of the request

media address and **port** number that will be sending the media packets

FIGURE 2: STRUCTURE OF AN INVITE MESSAGE

Examples of Legal Bots

CASE I: INTERACTIVE VOICE RESPONSE (IVR) SYSTEM

An interactive voice response (IVR) is a phone technology that allows a computer to detect voice and touch tones using a normal phone call. The IVR system can respond with prerecorded or dynamically generated audio to further direct callers on how to proceed [6]. Both IP and traditional (i.e., PSTN) telephone networks are full of IVR systems, in which a user calling a telephone number is briefly interfaced with an automatic call response system. The typical usage of IVR includes call centers, bank and credit card account information systems, air and rail reservation systems, hospital helplines, and college course registration systems.

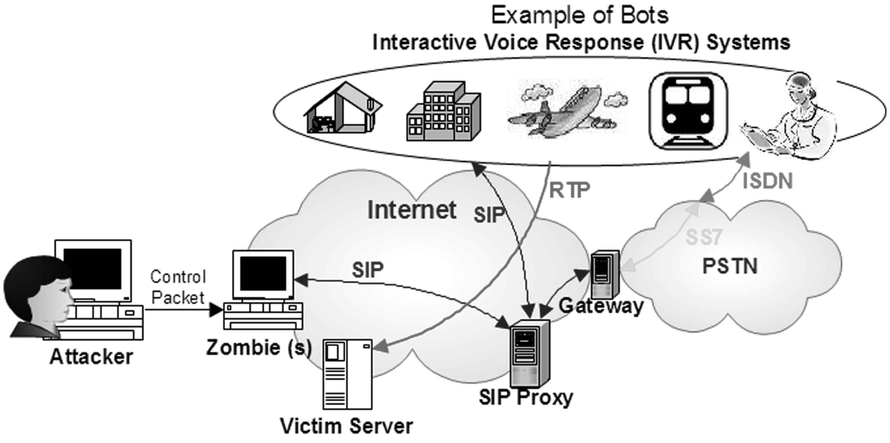


FIGURE 3: IVR SYSTEMS ACTING AS BOTS

As shown in Figure 3, now imagine an attacker, knowing this vulnerability, who sends out a few hundred INVITE messages (while keeping the same media connection address in the message body) to well-known automatic call response systems and establishes fake call sessions with them. In response, the IVR systems flood the victimized connection address with UDP-based RTP packets. In order to establish a call, an exchange of a few call setup SIP messages (i.e., INVITE/200 OK/ACK) can result in a few hundreds to thousands of RTP packets. Such an attack scenario uses both reflection and amplification to make a DDoS attack more potent.

CASE II: USER'S VOICEMAIL SYSTEM

Sometimes when a callee is busy or is not available to answer a phone call, the caller is directed to an answering machine or a voicemail system that plays a greeting message and stores incoming voice messages. An attacker may send fake call requests with the same media connection address to hundreds or thousands of individual telephone subscribers distributed over the Internet. The simultaneous playing of individual greeting messages can overwhelm the link's bandwidth connecting to the victim.

CASE III: USER'S VOICE COMMUNICATION—RTP STREAM

Even if we assume that the callee is not busy and answers a phone call, the callee's voice stream (e.g., Hello, hello . . . or some other initial greeting message) can be directed to a target machine. As in the previous examples, an attacker sends fake call requests with the same media connection address to hundreds or thousands of individual telephone subscribers, and the simultaneous response of subscribers can cause a flooding attack on the victim.

CASE IV: SPIT PREVENTION—THE TURING TEST

In many aspects a voice spam is similar to an email spam. The technical know-how and execution style of email spam can easily be adapted to launch voice spam attacks. For example, first a voice spammer harvests a user's SIP URIs or telephone numbers from the telephone directories or by using spam bots crawling over the Internet. In the second step, a compromised host is used as a SIP client that sends out call setup request messages. Finally, in the third step, the established sessions are played with a prerecorded WAV file. However, voice spam is much more obnoxious and harmful than email spam. The ringing of a telephone at odd times, answering a spam call, phishing attacks, and the inability to filter spam messages from voicemail boxes without listening to each one are time-wasting nuisances.

The Internet Engineering Task Force's informational draft [7] analyzed the problem of voice spam in the SIP environment, examining various possible solutions that have been discussed for solving the email spam problem and considering their applicability to SIP. One such solution is based on the Turing test, which can distinguish computers from humans. In the context of IP telephony, machine-generated automated calls can be blocked by applying an audio Turing test. For example, a call setup request from an unidentified caller is sent to an IVR system where a caller may be asked to answer a few questions or to enter some numbers through the keypad. Successful callers are allowed to go through the SIP proxy server and may also be added to a white list.

VoIP security products such as NEC's VoIP SEAL [8] and Sipera Inc.'s IPCS [9] have implemented audio Turing tests as an important component in their anti-spam product to separate machine-generated automated calls from real individuals. However, an attacker may use these devices as reflectors and amplifiers to launch stealthy and more potent DDoS attacks. For example, to determine the legitimacy of a single spoofed INVITE message, these devices send a few hundred RTP-based audio packets (a 10–20 s audio test) toward the media connection address of an INVITE message. A victimized connection address can be flooded with audio packets if an attacker sends one or two spoofed INVITE messages (with the same media connection address) to several such devices distributed over the Internet.

A Real-World Attack Scenario

To demonstrate a possible DDoS attack, we simulated a real-world attack scenario using IP phones from three different VoIP service providers, namely Vonage, AT&T Callvantage, and ViaTalk. As shown in Figure 4, over the Internet an attacker captures SIP signaling messages exchanged between callers and callees of various VoIP service providers that can later be replayed to launch many different types of DoS attacks toward the subscribers and the SIP proxy server. Most of these attacks are against an individual subscriber, but the INVITE flooding attack can also be launched against a SIP proxy server. However, the media source address spoofing attack discussed in this article is not confined to VoIP systems; rather, it can victimize any voice or data network element.

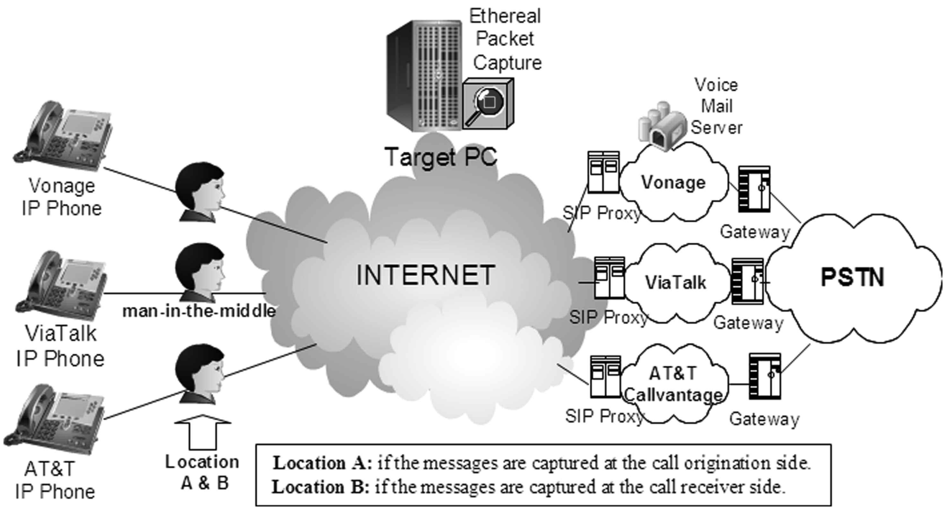


FIGURE 4: REAL-WORLD ATTACK SCENARIO

As shown in Figure 4, the AT&T user talks to both Vonage and ViaTalk customers. The SIP signaling messages exchanged between callers and callees are captured at two locations: *location A* lies between callers and their outbound proxies; similarly, *location B* lies between callees and their inbound proxies. At location A, we observed that in order to prevent *replay* attacks, each of the service providers challenges INVITE messages by sending 401 Unauthorized (in the case of AT&T) or 407 Proxy Authentication required messages that include an MD5 hash of the user's credential and a "nonce" value. This can only be defeated if we have the capability of modifying some header fields (which are not used in MD5 hash computation) and reconstructing the message in real time or by exploiting the implementation of some SIP proxy servers that may accept stale nonce values [10]. However, at location B, there are no such challenge/response messages, leaving the subscribers exposed and vulnerable to abuse. In the sample case study, we exploit the vulnerable and mostly overlooked location B. The captured incoming INVITE messages are reconstructed with a spoofed media address and port number. At a later time, INVITE and ACK signaling messages are replayed while maintaining the same relative order and time. The callee's voice stream (or the playing of the callee's answering machine) is successfully redirected toward the target host.

COULD THIS ATTACK BE PREVENTED?

We now discuss some of the questions that may arise regarding circumvention of the INVITE exploitation attack described in this article. One could argue that the attack can be prevented if the SIP user agent server (UAS) correlates first Via and Contact header fields with the connection address (c=) field of the message body. However, we observe that many services, such as SIP's firewall/NAT traversal and anonymity service, rely on a *media proxy*, thus forbidding the establishment of a correlation between signaling and media destinations.

Conclusion

With the growing acceptance of VoIP and the interconnection between SS7 and IP networks, there is a need to secure both network infrastructures and the protocols used between them. There are many efforts for SIP's implementation vulnerability assessment through syntax testing and test-suite creation. Still, we need to make a thorough revision of the protocol design as well as its intended use. We hope this article will work as a stimulant and bring a concerted effort to prevent any design or implementation flaw that may hinder VoIP deployments or the lowering of IP telephone subscribers' confidence.

REFERENCES

[1] Tipping Point, "Intrusion Prevention: The Future of VoIP Security," white paper, 2005: http://www.tippingpoint.com/solutions_voip.html.

[2] Bob Sullivan, "Virus Gang Warfare Spills onto the Net," April 2007: http://redtape.msnbc.com/2007/04/virus_gang_warf.html.

[3] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, *SIP: Session Initiation Protocol*, RFC 3261, IETF Network Working Group, 2002.

[4] A.B. Johnston, *SIP: Understanding the Session Initiation Protocol,* 2nd edition (Norwood, MA: Artech House, 2004).

[5] H. Sengar, R. Dantu, D. Wijesekera, and S. Jajodia, "SS7 Over IP: Signaling Interworking Vulnerabilities," *IEEE Network Magazine,* 20(6): 32–41, November 2006.

[6] Wikipedia Encyclopedia, "Interactive Voice Response," April 2007: http://en.wikipedia.org/wiki/Interactive_voice_response.

[7] J. Rosenberg and C. Jennings, "The Session Initiation Protocol (SIP) and Spam—Work in Progress," IETF's SIPPING Group, 2007.

[8] NEC Corporation, "NEC Develops World-Leading Technology to Prevent IP Phone SPAM," product news, 2007: http://www.nec.co.jp/press/en/0701/2602.html.

[9] SIPERA Systems, "Products to Address VoIP Vulnerabilities," April 2007: http://www.sipera.com/index.php?action=products,default.

[10] SIPERA Systems, "Some Implementations of SIP Proxy May Honor Replayed Authentication Credentials," May 2007: http://www.sipera.com/index.php?action=resources,threat_advisory& tid=183&.