

A Model for Spam Prevention in IP Telephony Networks using Anonymous Verifying Authorities

N.J Croft and M.S Olivier

April 2005

Information and Computer Security Architectures Research Group
Department of Computer Science
University of Pretoria
South Africa
ringtingting@gmail.com

Abstract

Unsolicited emails, more commonly known as spam, have plagued the use and efficiency of email since its inception. With the introduction of ungoverned cheap voice communications, such as IP telephony, spam over telephony (SPIT) and its prevention is set to dominate and drive whether the technology is widely adopted or not. A possible example of voice spam schemes includes the use of Interactive Voice Response (IVR) systems in conjunction with automated telemarketing sales to repeatedly initiate call setups and fill voicemail boxes. The problem faced is to design a system that allows legitimate voice call establishment while anonymously blocking unwelcome ones. An anonymous approach prevents voice spam by prohibiting spammers from establishing call and do not call lists.

In this paper we propose a new model as an extension to a call signalling phase (or call setup phase) as a means to prevent unwanted voice spam. In doing so we define a IP telephony “call me back” scheme using an Anonymous Verifying Authority (AVA) and Mediator in blocking unsolicited voice calls. We illustrate the use of a “call me back” request in allowing the callee to accept or reject the call. Verification may be based on the callee’s personal call policy thus filtering unsolicited voice spam.

Key words

IP telephony, Spam, SPIT, Model, Verifying Authority

1 Introduction

Since the telephone was invented in the late 1800's, telephone communication has not changed substantially. In the 1990's a number of people in the research field took a serious interest in carrying voice over IP networks. Recently, interest has been renewed in IP telephony due to the mass deployment of the Internet and the benefits it brings to businesses in terms of significant communication and infrastructure cost reductions.

With the adoption of new communication technologies comes inherent risks. Vulnerability considerations in communication mediums include amongst others; eavesdropping, identity theft, Denial of Service (DoS) attacks and unwanted communication often referred to as spam. IP telephony is no different and is not immune to such concerns [13].

This paper focuses on unsolicited voice messages in IP telephony environments and provides a means to combat voice spam through a model using Verifying Authorities (VA) in the call setup stage. By encompassing anonymity in the verification process one eliminates a direct path from the caller to the callee and the ability of the spammer in gaining direct access to callee (by simply placing a call).

This paper is structured as follows: Section 2 provides an overview of IP telephony focusing on architectures, protocols and session establishment. In Section 3 we outline Spam over IP Telephony (SPIT), define what constitutes voice spam and provide some possible approaches to dealing with problem. In Section 4 we implement a model for the prevention of SPIT using Anonymous Verifying Authorities (AVA) and coupled with a Mediator providing the abstraction needed in order to remain anonymous from the caller in the call setup phase. Section 5 concludes this paper.

2 Background

Voice communication over IP or IP telephony as it is commonly known, is a real-time event displaying properties of connectedness between the communication parties. We begin with an overview of IP telephony, illustrating common IP telephony architectures and how voice communication sessions are established.

2.1 Overview of IP Telephony

IP telephony is a technology that allows standard telephone voice signals to be compressed into data packets for transmission over the Internet or other IP network. The protocols used in carrying the voice signals over the IP networks are commonly referred to as Voice over IP (VoIP). IP telephony moves away from traditional circuit switched voice networks, such as Public Switched Telephone Networks (PSTN's) to a packet switched one where IP packets containing voice data are sent over the network. The advantages of IP telephony over traditional telephony are amongst

others: lower costs per call (or even free calls [4]) and lower infrastructure costs. IP networks are considered best-effort networks, so unfortunately there is no guarantee of constant bit flow. Therefore the problems facing IP telephony include: Quality of Service (QoS) guarantees, latency and possible data integrity problems.

IP telephony usually comprises, independent of the protocols used, a signalling plane and a multimedia plane. The signalling plane is used for transporting the necessary signalling information, while after call setup, the media transport plane is used to carry voice data packets between IP telephony components.

Before audio can flow between two devices, various protocols must be employed to find the remote device and to negotiate the means by which audio will flow between the two devices.

We now turn our attention and investigate in more detail IP telephony's infrastructure in terms of signalling (session establishment) and transport of real-time audio.

2.1.1 Signalling

As previously discussed, voice is based on a connection oriented technology which implies there must be a call setup phase before any voice traffic is carried across the IP network. Signalling commands in establishing and terminating a call as well as providing some special features such as call forwarding and call waiting. There are currently two common Internet telephony signalling protocols, Internet Engineering Task Force (IETF) Session Initiation Protocol (SIP) [10] and International Telecommunication Union Standardization Sector (ITU-T) H.323 [7]. Both SIP and H.323 can be described as intelligent end-point protocols meaning the establishment of media streams between the local and remote device are integral parts of these protocols. In order to prevent SPIT we must understand how these protocols are used and how a session is established in more detail.

2.1.2 Session Initiation Protocol (SIP)

SIP, published as RFC 3261 [10], is an application-layer protocol that can establish, modify, and terminate multimedia sessions with one or more users. SIP was adopted by 3gpp [12] and has been designed with scalability in mind while remaining simple in its implementation. SIP does not by default provide any mechanism to ensure that data packets are delivered in sequential order.

SIP is a text-based protocol, based on an HTTP/SMTP request-response model where SIP addresses users by and email-like address typically containing a username and a host name. SIP is a peer-to-peer protocol where each peer is referred to as a User Agent (UA) where UA's can either act in client or server mode. SIP identity, a type of Uniform Resource Identifier [16], called a SIP URI is used for initiating interactive communication sessions between users. SIP uses the Session Description

Protocol (SDP) [6] in describing the capabilities and media types supported by terminals.

SIP's architecture comprises of five entities (Figure 1), namely:

1. SIP Terminals
2. Proxy Servers
3. Redirect Servers
4. Location Servers
5. Registrar Servers

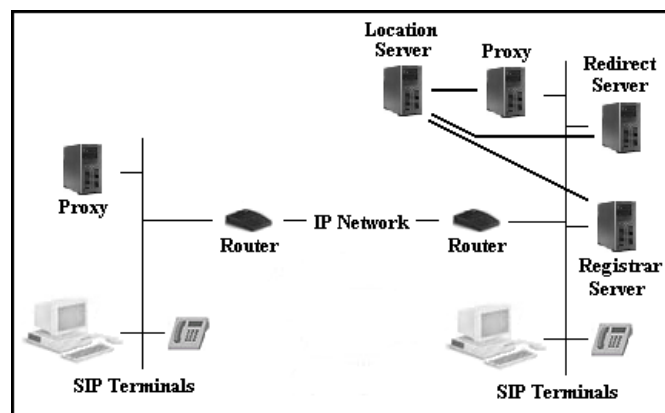


Figure 1: SIP architecture

Terminals initiate 2-way multimedia communication, sending and receiving SIP requests. Proxy servers act as intermediary, checking the SIP terminal destination and forwarding requests. SIP proxies are responsible for maintaining the state of calls. A Redirect server, instead of forwarding requests, advises the caller (if need be) to contact another server directly. Location servers contain information about the location of the callee and is usually integrated in Redirect or Proxy servers. A Registrar server processes requests from terminals wishing to join a SIP network. The Registrar server will in turn update the Location server with the client information.

2.1.3 H.323

H.323 [7] is a binary-based protocol standard approved by the International Telecommunication Union (ITU) which supports real-time point to point multimedia data communications over non-guaranteed bandwidth, packet-based networks, such as the Internet. H.323 is an umbrella specification as it encompasses various other ITU standards where the latest version (v5) was released in 2003.

In general H.323 implementations includes four logical entities (Figure 2), namely:

1. H.323 Terminals

2. Gateways (GW)
3. Gatekeepers (GK)
4. Multipoint Control Units (MCU)

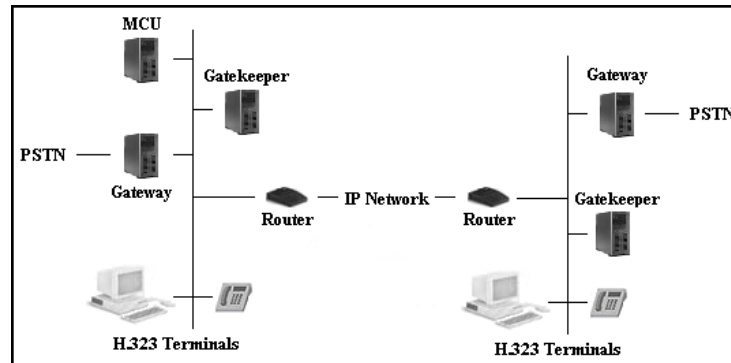


Figure 2: H.323 architecture

A H.323 Terminal provides real-time two way communication with another H.323 terminal, gateway or MCU sending multimedia messages. H.323 terminals support audio codecs for example the G.711 [5] codec and signalling using Q.931, H.245 and Registration, Administration and Status (RAS) protocols. Gateways are optional components and are only required when communicating between different networks for example between an IP-based network and Public Switched Telephone Networks (PSTN's). A Gateway provides data format translation, control-signaling translation, call setup and termination functionality as well as compression and packetization of voice. Gatekeepers are responsible for translating between telephone number and IP address and routing of calls. They also manage bandwidth and provide mechanisms for registration and authentication by terminals. All H.323 endpoints register with a single GK and in the process build an H.323 zone. To support multi-terminal conferences all terminals must establish a direct connection to an MCU.

2.2 Transport

Both SIP and H.323 use the Real-Time Transport Protocol (RTP) for carrying real-time multimedia traffic. RTP was developed by the IETF to transport audio and video across an IP network. RTP, published as RFC 1889 [15], provides end-to-end delivery services but does not provide any mechanism to ensure timely delivery or provide Quality of Service (QoS) guarantees. As voice is tolerable of a small amount of packet loss but intolerable of delay, RTP typically transports voice packets over UDP.

RTP consists of two parts, namely:

- RTP, to carry data that has real-time properties

- RTP control protocol (RTCP), to monitor QoS and to convey information to the participants in an on-going session

Having completed the call setup phase, RTP encapsulate the audio stream. Each terminal samples the input audio stream, encodes it using for example the G.711 [5] codec, and encapsulates the payload into a RTP header. Speech codecs [3] are optimized for compressing voice, which significantly reduces the bandwidth consumption compared to uncompressed audio streams.

2.3 Session Establishment (Call Setup)

Figure 3 illustrates a SIP session setup and a H.323 session setup with two endpoint terminals without the use of a Gateway (GW).

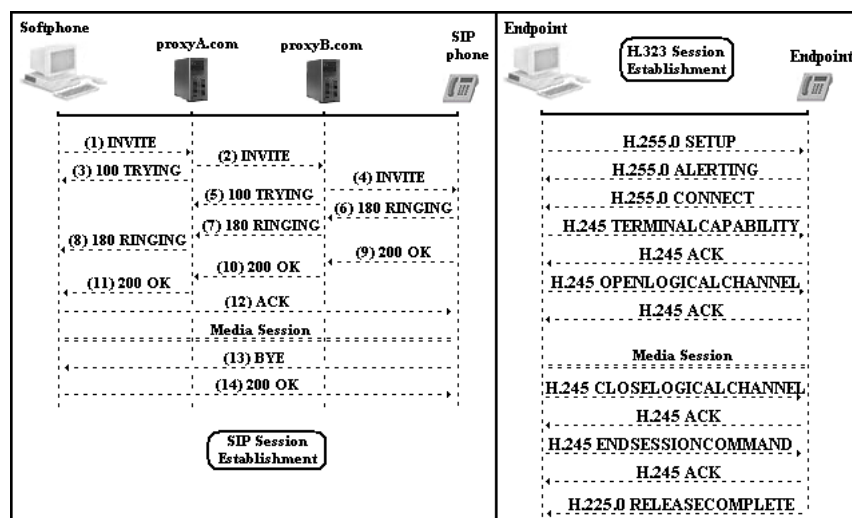


Figure 3: SIP session setup and H.323 session setup without a GW

Both SIP and H.323 uses request/response models in the call setup phase. If the callee endpoint is known to the caller a magnitude of requests may be sent to the callee in a relatively short period of time. We now investigate Spam over IP Telephony, commonly referred to as SPIT and how it may plague the adoption of IP telephony.

3 Spam over IP Telephony (SPIT)

Spam over IP Telephony is best described as a combination of a telemarketing call and an email spam message. SPIT works on the same principle as email spam, allowing for example a 30 second voice-recording to be sent to thousands of voice IP addresses within seconds. However, unlike email messages which you can stop and

analyze for content, one can not add latency (delay) to a IP telephony call as this effects audibility of the voice communication and may result in what is refer to as jitter.

SPIT will undoubtedly become the next pervasive medium for spammers due to its low associated costs. Once a person's IP telephony numbers is published or harvested, it could become a serious threat to security and infringe on a users privacy. An agent could, in parallel, generate and setup a large number of calls. If a call should connect, the agent then generates an acknowledgement and proceeds to play a telemarketing prompt after which it terminates the call. Privacy mechanisms [11] provide guidelines for the creation of messages that do not divulge personal identity information, however in IP telephony, like in traditional communication systems, once a user is identifiable by some means (alias), this identity may be distributed without the consent of the owner [2].

The SIP SPAM Internet draft [8], outlines the following permissible approaches to SIP spam prevention as what are seen as common ways in preventing email spam:

- Content Filtering. Useless as once the call is answered the session is establish and content delivered. It is further noted that it becomes extremely difficult in filtering stored voice content such as voicemail.
- Black Lists. Filters according to a blocked list of addresses (both usernames and entire domains). Useless as email addresses are easily spoofed and email addresses are almost limitless in supply.
- White Lists. Filters according to a list of valid senders the user wishes to receive emails from. Successful defence against spam however its not a completely flexible solution as it does not allow for the receipt of new wanted emails.

It seems that most successful anti-spam measure taken to combat email spam is almost useless for the prevention of IP telephony spam. Probably the most effective prevention measure of voice spam lies in the use of white lists, however as it is not considered a flexible solution, we introduce a Verifying Authority (VA) in our spam prevention model where the VA is controlled by a trusted party.

4 IP Telephony SPAM prevention model

In order to prevent spam from being directly intended for the user's intranet, our proposed model removes the call setup phase into an abstracted layer of verification. Any provider that performs inter-domain messaging must use techniques that depend on strong identity techniques [9]. This verification process communicates a IP telephony "call me back" offer to the callee which is either accepted, rejected or simply ignored. For this reason we introduce two new components to the IP telephony architecture, namely a Mediator and a Verifying Authority (VA) (Figure 4).

The Mediator's role is twofold:

1. To locate and forward call setup request information to a Verifying Authority (VA_n) (anonymous to the caller)
2. Hook-up calls, by joining the caller to the callee with a unique communication token.

The VA may:

- Verify amongst others the identity, validity, location of the caller and callee destination
- Forward a call setup request to another VA for verification
- Apply a generic policy of the intended callee's intranet in determining whether to send a "call me back" request on behalf of the caller
- Fetch and apply any callee policies from the callee Proxy/Gateway server

If the callee policy should comply, then once again, the VA forwards on behalf of the caller a "call me back" request.

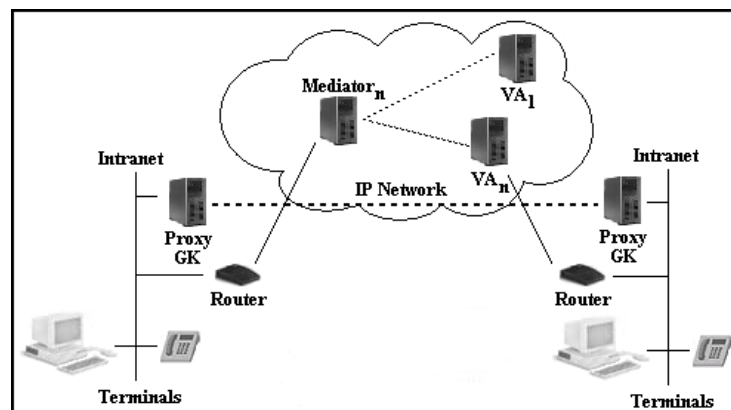


Figure 4: IP Telephony SPAM prevention model

In our model it is important to note that the Proxy can only perform the following actions:

- Can only send a call setup request to a Mediator
- Can only receive "call me back" and policy requests from a VA
- Can only send and receive multimedia streams (using RTP) if a caller in the intranet has a pending call setup request and has a valid token assigned to it by the VA authority through the Mediator

A Mediator can only perform the following actions:

- Can only receive call setup requests passed from a Proxy server
- Can pass this call setup request to any of the n present VA's

A Verifying Authority (VA) can only perform the following actions:

- Verify amongst others: identity, validity, destination and location of the both communicating parties
- Can request any call related setup policies from the callee Proxy server
- Generate a unique valid token

Lets take the following example where Susan wants to call Bob where Susan and Bob are on different intranets and for simplicity sake they both use SIP-enabled phones. Figure 5 illustrates our proposed SPIT prevention session setup.

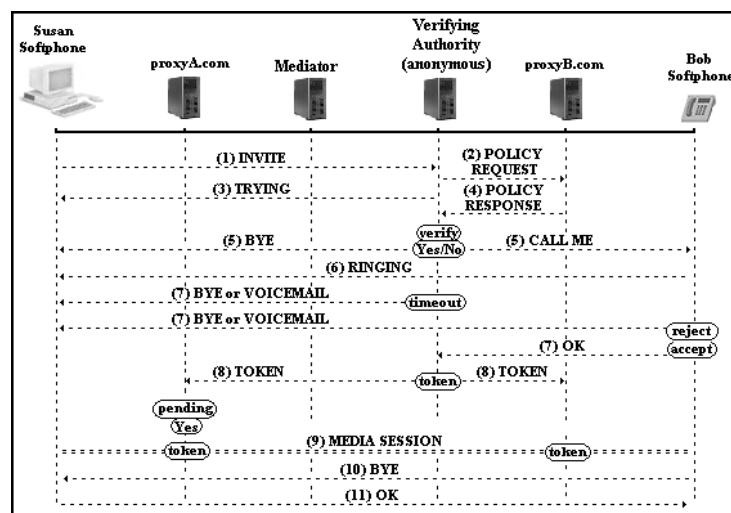


Figure 5: IP Telephony SPAM prevention session establishment

Susan initiates and invites Bob to participate in a voice call. Susans Proxy server forwards this request onto a Mediator which in turn forwards it to a randomly chosen Verifying Authority. Susan has no way of knowing which VA was chosen and is therefore anonymous to Susan. The VA verifies amongst others the identity, validity, destination and location of both Susan and Bob (the intended recipient). The VA may request any related policies regarding Bob’s call setup requirements from Bob’s Proxy server. During this period Susan will be prompted with a “ringing” tone. At this point verification takes place, and if approved the VA sends Bob a “call me back” request, if not the call is terminated. Three different outcomes are possible at this stage, either the call is terminated due to an elapsed expiry time, or Bob accepts or rejects the request. If call setup request is rejected or is simply ignored, caller Susan may not be able to determine even if callee Bob exists. The existence of Bob as a valid IP telephony user may be the single most contributing factor in the prevention of SPIT. Voicemail is an option where call setup requests are timed-out or rejected depending on Bob’s profile and the status of the VA’s verification process. If Bob decides to accept the call from Susan and “OK” response is sent to the VA upon which the VA generates a unique communication token. The VA distributes this token to

both Susan's and Bob's Proxy server. Susan's Proxy will not be able to create the media session unless it has a call invite in pending state for Susan going out to Bob. If so, then the media session is established making use of this unique communication token. The media session may then be terminated and is acknowledged as such.

5 Conclusion

In this paper we designed a system that allows legitimate voice call establishment while anonymously blocking unwelcome ones through the use of Anonymous Verifying Authorities (AVA). Our proposed model for voice spam prevention allows for filtering of call setup request, user consent-based acceptance ("call me back") and a tokenized means to call setup and media session.

Future work includes the investigation of Denial of Service (DoS) or distributed Denial of Service (DDoS) attacks on the Proxy/Gatekeeper servers and the Mediator causing sufficient latency on the IP telephony network. One possible solution may lie in the adoption of a mix network [1] or crowds [14] approach to Proxy/Gatekeeper servers and the Mediator. We may also investigate the role the token may play in the RTP media session.

References

- [1] David Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.
- [2] N.J. Croft and M.S Olivier. Using a Trusted Third Party Proxy in achieving GSM Anonymity. In *South African Telecommunication Network and Applications Conference*. SATNAC, September 2004.
- [3] N.J. Croft and M.S. Olivier. Codec-Hopping: Secure and Private Voice Communication in Bandwidth Constrained Networks. SecPerU 2005, Workshop on Security and Privacy in Pervasive Ubiquitous Computing, Santorini, Greece. Submitted, April 2005.
- [4] Free World Dialup. <http://www.pulver.com/fwd/>. Web Reference, Accessed April 2005.
- [5] ITU-T Recommendation G.711. Pulse Code Modulation (PCM) of voice frequencies, November.
- [6] M. Handley and V. Jacobson. SDP: Session Description Protocol. RFC 2327, April 1998.
- [7] International Telecommunications Union. Recommendation H.323, Packet based multimedia communication system. Technical report, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, February 1998.
- [8] J. Peterson J. Rosenburg, C. Jennings. The Session Initialization Protocol (SIP) and SPAM. IETF Internet Draft, October 2004.

- [9] J. Peterson. Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP). IETF Internet Draft, September 2004.
- [10] E. Schooler J. Rosenberg M. Handley, H. Schulzrinne. SIP: Session Initialization Protocol, RFC 3261. Technical report, Internet Engineering Task Force (IETF), March 1999.
- [11] J. Peterson. A Privacy Mechanism for the Session Initiation Protocol (SIP). RFC 3323, November 2002.
- [12] Third Generation Partnership Project. <http://www.3gpp.org>. Web Reference. Accessed April 2005.
- [13] U. Roedig R. Steinmetz R. Ackermann, M. Schumacher. Vulnerabilities and Security Limitations of Current IP Telephony Systems. Technical report, Darmstadt University of Technology, 2002.
- [14] Michael K. Reiter and Aviel D. Rubin. Anonymous Web transactions with Crowds. *Commun. ACM*, 42(2):32–48, 1999.
- [15] H. Schulzrinne. RTP profile for audio and video conference with minimal control, RFC 1890. Technical report, Internet Engineering Task Force (IETF), January 1996.
- [16] U.C. Irvine L. Masinter T. Berners-Lee, R. Fielding. Uniform Resource Identifiers (URI): Generic Syntax. RFC 2396, August 1998.