Modern Applied Science; Vol. 7, No. 7; 2013
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

A Review of Methods for Preventing Spam in IP Telephony

Saeed Farooq Khan[1], Marius Portmann[1] & Neil W. Bergmann[1]

[1] School of Information Technology & Electrical Engineering, University of Queensland, Brisbane, Australia

Correspondence: Neil W. Bergmann, School of Information Technology & Electrical Engineering, University of Queensland, Brisbane 4072, Australia. Tel: 61-7-3365-1182. E-mail: n.bergmann@itee.uq.edu.au

Received: May 19, 2013 Accepted: June 3, 2013 Online Published: June 22, 2013

doi:10.5539/mas.v7n7p48 URL: http://dx.doi.org/10.5539/mas.v7n7p48

Abstract

Voice over IP telephony reduces communications costs, but it is also subject to spam, i.e. unwanted, unsolicited calls. This paper discusses the problem of spam over IP telephony and reviews published techniques to reject likely nuisance callers. The techniques investigated include content filtering, black lists, white lists, gray lists, call rate monitoring, IP/domain correlation, reputation systems, consent-based communications, address obfuscation, limited-use addresses, computational puzzles, Turing tests, payments at risk, legislation, centralized SIP providers, circles of trust and authenticated identity. The advantages and disadvantages of each are analyzed. It is concluded that no single technique is sufficient and that a framework of multiple techniques is required. One proposed framework is analyzed and it is found that it has limited support for voicemail spam and contact request spam.

Keywords: internet telephony, voice over IP, spam, privacy

1. Introduction

Traditional Public Switched Telephone Network (PSTN) voice communications is under considerable pressure from Internet-based telephony. Voice over Internet Protocol (VoIP) encodes voice signals and then transmits these signals over IP (Internet Protocol) networks (Hung & Martin, 2006). VoIP offers significant advantages over PSTN. Line costs are negligible if an IP Internet connection is already available. Features such as conference calls and video calls are easily accommodated in an IP-based communication system. Since VoIP compresses data packets during transmission, more calls can be handled on one access line. In addition, VoIP can operate over a wide range of underlying network technologies such as ATM, SONET, Ethernet and Wi-Fi, as long as they use the IP protocol at the network layer (Varshney et al., 2002).

As well as these advantages, VoIP has several significant disadvantages. Unlike PSTN, VoIP uses shared network bandwidth. Quality of Service is a significant issue if real-time telephone calls share limited Internet bandwidth with other data types. Distortion or call dropouts are common if bandwidth is inadequate. Voice data have to be compressed and transmitted, then decompressed and delivered. This process can be problematic if computing resources are limited. Another problem is that if the Internet connection is unavailable or very congested, then VoIP is not available. This can be very annoying for users of VoIP services (Varshney et al., 2002). Security is also a very important issue in VoIP. Because VoIP telephones are potentially accessible by any Internet connected device, VoIP is vulnerable to issues like identity and service theft, viruses and malware attacks, as well as denial of service and phishing attacks (Davidson et al., 2006; Dantu et al., 2009).

This paper addresses one particular issue, which is the problem of spam in IP telephony. A continuing problem for the Internet is controlling spam, i.e., undesired, unsolicited emails. The increasing quantity of spam messages circulating through the Internet results in problems such as low availability and network congestion, and most importantly, disruption of users. Unfortunately, VoIP is not immune to spam. VoIP spam refers to unsolicited Internet telephony calls that consume computing and communications resources of the end users and intermediate infrastructure, and which also cause interruptions and inconvenience to the call recipient. VoIP spam is also referred to as SPIT or SPam over IP Telephony. Of particular concern is automated spam, i.e. where computer software generates automated nuisance calls which advertise particular products or services, or else seek to gather private information.

Voice spam is like email spam in that it annoys the recipient, consumes resources and congests the network. For email spam, there are successful filtering systems that are in widespread use. Based on the content, sender address or recipient addresses, email servers can discard messages with no time or computation cost to the end user. However, when it comes to preventing voice spam, traditional email spam filtering techniques are not adequate. In VoIP, identifying spam calls based on contents requires the spam calls to be analyzed in real time which is not a simple task and not practical. Many of the techniques that are used or proposed for email spam detection are based on content analysis (e.g. using Bayesian filters). Content-based filtering is less useful for VoIP spam, since the voice content can only be analyzed once the call is underway, by which time the receiver has already been interrupted by answering the call. Voice spam is much more irritating than email spam. Hence it is a challenge to detect and block it before the user is alerted to a spam call. Therefore, if the advantages of VoIP technology are to be maintained, then specific VoIP mechanisms for spam prevention are needed.

The goal of the paper is to review a range of proposed techniques to detect and mitigate VoIP spam and to discuss the advantages and disadvantages of each technique. The paper will also analyze the resources that are needed to implement such techniques and changes needed to be made to existing protocols.

2. VoIP Protocols

There are several different standards for VoIP. The most widely adopted open standard is the SIP (Session Initiation Protocol) from the Internet Engineering Task Force (IETF) (Rosenberg et al., 2002). H.323 is an alternate open signaling protocol that provides call connection, call management and call termination in a VoIP session. H.323 is recommended by the International Telecommunications Union (ITU) (Schulzrinne & Rosenberg, 1998). SIP and H.323 adopt different approaches to signaling. Typically H.323 is a high overhead, feature-rich protocol used by telecommunications providers. SIP is a low-overhead distributed protocol, more common for open-source unregulated solutions.

The popular VoIP application, Skype, uses a proprietary standard that is based on a peer-to-peer computing paradigm (Baset & Schulzrinne, 2006). It includes facilities for traversing firewalls and NAT (Network Address Translation) servers which are problematic for other protocols. Other commercial providers of VoIP services, such as Cisco, also use more complex gateway protocols such as MGCP (Media Gateway Control Protocol) for systems that need to bridge multiple VoIP and PSTN networks (Khasnabish, 2003). In this review, the SIP protocol will be used as an example signaling protocol, but many of the spam prevention approaches can be easily applied to other protocols as well.

The ideal time to reject spam callers is during the SIP call setup phase, so that users are not bothered at all. Therefore, it is useful to explain SIP in some detail. SIP is an application−layer signaling protocol that manages multimedia sessions. It is used for establishing, modifying and terminating media sessions (Rosenberg et al., 2002). SIP does not restrict the media types that users can use during a session. The session can involve IP phone calls, conferencing and messaging. SIP is a text - based protocol based on the HTTP request−response model. SIP uses addresses like an email address consisting of username and hostname. SIP is a peer−to−peer protocol where each peer is referred to as a User Agent (UA). The UA can either act in client or server mode. In the client mode, the peer is called the User Agent Client (UAC) whereas in server mode, the peer is known as the User Agent Server (UAS). In order to initiate communication sessions between peers, a SIP Uniform Resource Identifier (URI) is used. The SIP architecture comprises five entities namely, SIP User Agent, Proxy Server, Redirect Server, Location Server and Registrar Server (Rosenberg et al., 2002).

2.1 SIP Call Setup

A call is usually set up between two users as a result of a message exchange in the form of a 'trapezoid' as shown in Figure 1 below:

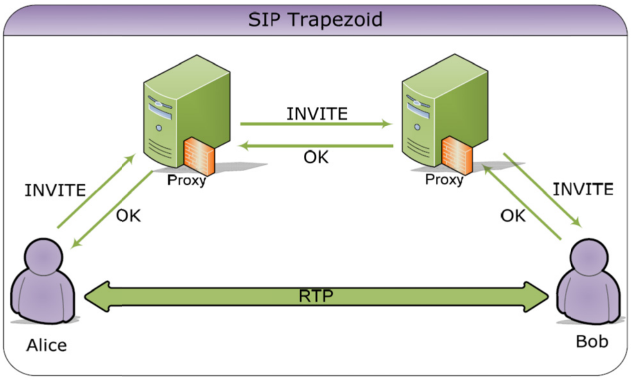

Figure 1. Call setup in SIP

Alice (the caller) uses her SIP phone to ring Bob (the receiver) on his SIP phone over the Internet. There are two proxy servers which act on behalf of the end-users to facilitate session establishment (Rosenberg et al., 2002). The caller sends an INVITE request to the receiver's SIP URI address. The request contains several header fields in the INVITE message, including a unique identifier of the call, the destination address (Bob's address), Alice's address and the type of session that is to be established.

After receiving the INVITE request from the caller, her proxy server acknowledges with a 100 (Trying) response, indicating that the proxy is now forwarding the INVITE to the destination. The proxy server then uses a location service to identify the correct destination proxy to send the data to. The destination proxy forwards the INVITE request to the receiver's SIP phone. The receiver's phone rings and sends a 180 (Ringing) message back through the proxy chain to the caller. When the receiver picks up the phone to answer it, the phone sends a 200 (OK) message back through the chain to the caller. The caller's phone sends an acknowledgment message (ACK) to the receiver's phone to confirm reception of the final response. After this message is received, the session is established between caller and receiver and they can communicate via Real time Transmission Protocol (RTP) messages.

When one of the users hangs up at the end of the session, a BYE message is sent to the other user, who in turn sends an OK message to acknowledge it and this terminates the session.

SIP conversations may include an Instant messaging (IM) session as well as a voice session, and these IM sessions are often used as part of the spam prevention mechanism.

2.2 SIP Spam

SIP spam can be classified into the following three types (Rebahi, Sisalem, & Magedanz, 2006):

(i) *Call Spam:* This refers to bulk unsolicited session initiation requests to establish a voice, video or instant messaging session (Day, Rosenberg, & Sugano, 2000). If the user answers, the spammer software plays its message.

(ii) *IM Spam:* This is similar to email. It is a set of unwanted instant messages with content that the spammer tries to convey. This spam is usually sent using SIP MESSAGE requests. INVITE requests with the content in the subject header or INVITE requests with text or HTML bodies can also be used for IM spamming (Rosenberg & Jennings, 2008).

(iii) *Presence Spam:* This is similar to IM spam. It consists of unsolicited presence requests in an attempt to get on the allowed caller list of a user. Then the spammer can send instant messages or voice spam (Day, Rosenberg, & Sugano, 2000).

PSTN call spam typically consists of telemarketing calls. However, because there is a significant cost (at least a few cents) to make a PSTN call, this limits the number of spam calls. However, these line costs are reduced by VoIP. IM spam is similar to email spam in terms of costs involved. IM is more intrusive than email; messages may automatically popup and be presented to the user. On the other hand, email needs to be deliberately selected and opened. IM systems can operate in two different modes: page mode and session mode. In page mode systems, each message is sent separately, similar to email. In session mode, advance signaling sets up a conversation, and then IM messages are exchanged. Anti-spam techniques for email can also be used for page mode IM but it is difficult to apply the same techniques to session mode because it is more like telephony (Rosenberg & Jennings, 2008). The cost for presence spam is similar to IM spam. Presence spam can be thought of as a type of IM spam with limited, deceptive content conveyed to the user, such as naming the spammer "System Update".

3. Literature Review

Research on VoIP spam has been undertaken for the past decade. Many techniques have been deployed to combat spam but none of these are ideal. Most of the techniques are variations on approaches that have been proposed in the context of email spam. This section surveys proposed techniques and their suitability for VoIP spam prevention or mitigation.

The various approaches can be grouped into three major categories, depending on where the spam prevention occurs. The first category is called system-based approaches which apply to the whole Internet. Second are server-based approaches which require knowledge of the behaviour of users across an administrative domain and are implemented on SIP servers. The third category of approaches is client-based approaches where spam is rejected at the client machine.

3.1 System Based Approaches

3.1.1 Legislation

This refers to making laws and regulations to make spam illegal. Such laws could apply to any kind of spam: IM, email or SIP. It usually takes the form of "do not call" lists. People who do not want to be bothered by advertisements and junk calls put their numbers onto this list. These laws have limited scope of enforcement. Calls from international sources cannot be treated under the laws of the recipient country. Only domestic callers can be penalized for spam calls under this scheme.

In (Park, Kim, & Kang, 2006), the applicability of email spam regulations to counter VoIP spam is investigated. These regulations are called Opt-out, Opt-in and Labeling. However, their effectiveness depends on the ability to sanction spammers who operate outside the law.

3.1.2 Provider Policies

If VoIP systems require users to register with a service provider, then that service agreement will include an injunction against spam. For example, Skype's terms of use state "*In particular, you are responsible for ensuring that you do not submit material that is … (iv) an advertisement or solicitation of business; …*" (Skype, 2013). If the system allows users to report spam abuse, then those users can be sanctioned and their accounts disabled. However, like email spam, it is relatively easy for the spammer to simply open another account if one account is closed.

3.2 Server Based Approaches

If a VoIP system employs a server during call setup, as shown above in Figure 1, then that server can monitor the behaviour of various users of the VoIP system and reject spam calls before these are even notified to the client. In general, the success of such server-based systems depends on the ability of servers or cooperating domains of servers to accumulate statistics about users.

3.2.1 Call Rate

This method is based on the rate at which calls are made by an individual caller within a certain time interval (Schlegel et al., 2006). The call rates of individuals are updated whenever a new call is made. A server will only allow a certain number of calls to be made in a fixed period. A higher call rate suggests a higher probability of spam. However, this method requires that calls are all visible by the receiver's server. Also, the spammer can rotate through a large number of sender addresses so that each sender has a low call rate.

3.2.2 IP/Domain Correlation

The IP/Domain module works by keeping a log of past calls with SIP identity and corresponding IP address. The SIP identity and IP address of new calls are checked against the logged information for matches in either the SIP domain or IP address. Schlegel et al. (2006) explain that the module identifies three potential spam situations:

Calls with a different caller SIP domain are placed from the same IP address

Calls from the same caller are placed from different IP addresses

Calls from callers with the same domain but different usernames are placed from the same IP address

This method aims to detect spammers who avoid detection techniques such as call rate or black lists by rapidly changing their SIP identity. Legitimate and illegitimate uses of these situations need to be distinguished. The second situation is valid for mobile users, so the timing between such calls needs to be considered. A central server for an organization which routes all calls through one server could lead to the third situation.

3.2.3 Circles of Trust

This is a model in which a group of domains collaborate and agree to exchange calls amongst each other. If any one of them is caught spamming, they have to pay a fine (Rosenberg & Jennings, 2008). This technique depends on being able to authenticate between domains. In other words one domain can authenticate that calls are received from a second domain. This can be done by using mutually authenticated transport level security between SIP providers (Dierks & Rescorla, 2008). This kind of technique is suitable for small domains where policies can easily be enforced. It is not clear how well such systems would scale, and how large public domains would ensure their users are not spammers.

3.2.4 Centralized SIP Providers

This is a variation of circles of trust (Rosenberg & Jennings, 2008). It uses a separate class of service providers who operate as gateways between domains. SIP networks would route all inter-domain calls through these providers, and would only accept SIP messages from their chosen inter-domain provider. The inter-domain provider charges the local provider per message for the delivery of message to other local providers. The inter-domain SIP providers later on form bilateral agreements with each other and exchange SIP messages according to the contract. Based on the contract, each inter-domain provider is responsible for a charging fee per message to their customers. This structure is similar to the PSTN. However, these kinds of architecture are unattractive because they complicate the SIP protocol and restore the centralized bottleneck that the protocol aims to eliminate.

3.2.5 Signaling Protocol Analysis

A VoIP call consists of a sequence of signaling messages that are exchanged during call setup, during the session, and at call termination. By looking at the characteristics of these signaling messages, it is possible to identify characteristics that apply to spammers but are less likely to apply to other users. Spam calls are unidirectional – a lot of calls are made by the spammer but few or none are received. The receiver usually hangs up before the end of the message. Based on these characteristics, MacIntosh and Vinokurov (2005) have defined a number of scenarios for termination behavior. This technique does not require user input to identify spammers. However, at least ten calls are needed to make a decision about one caller, and so the technique is easily circumvented by changing caller IDs. Moreover, this technique may also block some legitimate services, such as wakeup calls or emergency notification warnings.

3.2.6 A Biometric Framework for SPIT Prevention

A major problem in countering SPIT is ease and frequency with which spammers can change their SIP identity. Methods such as blacklisting are not useful in the presence of continuously changing identities.

One way is to bind identities to persons by using biometrics such as voice features to identify a caller. However, there are practical challenges. For example, if the servers are to check all calls on a global scale, the certification servers must be powerful and pervasive. Additionally it is not clear how easy it is to identify the voices of thousands or millions of users since voice patterns of legitimate users and spammers may be indistinguishable, and spammers may use background noise to alter the features (Baumann, Cavin, & Schmid, 2006).

3.3 Client Based Approaches

In these approaches, call filtering is done on a per-client basis. This doesn't mean the algorithm always needs to be implemented on a client machine. If a client always uses the same server machine, and client data (such as white lists and black lists) are available to that server, then these algorithms can be implemented on the server.

3.3.1 Black Lists

These are lists that contain addresses of known spammers and are usually maintained by spam filters. The addresses comprise both usernames and the domains. In case of email, black lists are not very effective because addresses can easily be forged and the spammer may come up with a spoofed address which can render the black lists useless. Even if the spammer does not forge addresses, he or she can get a new 'From' address, as email addresses are in abundant supply. The blacklisting cannot be done based on domains, because a domain may also contain legitimate users. Blacklisting must be done on individual identities. In summary, black lists are not effective for SIP spam, mainly due to the fact that SIP uses email-like addresses (Rosenberg & Jennings, 2008).

3.3.2 White Lists

White lists are lists which contain the addresses of valid users, and so they are the complement of black lists. The communication requests from these addresses are not blocked. White lists are also vulnerable to spoofing but that can be prevented with strong authentication mechanisms. One problem with white lists is that if a legitimate caller or sender is not on the white list of the intended recipient, there needs to be a lightweight way for them to be added. This means white lists need to overcome the "introduction problem" which refers to how a legitimate but unknown caller can get onto the white list. White lists can benefit from social media applications which already incorporate friend or contact lists (Rosenberg & Jennings, 2008).

3.3.3 Gray Lists and Gray-Leveling

Gray lists are used in a technique known as Progressive Multi Gray-Leveling (PMG) (Shin, Ahn, & Shim, 2006). PMG calculates the "gray level" of a caller to determine whether the call will be accepted or rejected. The gray level is based on previous caller history, and is a measure of whether the caller is a likely spam source or not (Shin, Ahn, & Shim, 2006). Unlike the white and black lists that always accept or reject calls from a given caller, PMG determines the legitimacy of a caller, based on calling patterns. Spam-like behaviour, such as many calls in a short time raises the gray level, and if a threshold is exceeded, the caller is blocked from further calls. The gray level reduces during periods where calls are not initiated. Individuals can set their incoming gray level threshold to suit their situation. Because PMG needs to be able to track user behaviour, it needs a strong mechanism for identity management. PMG can only rank those calls that the ranking system sees, so it performs best with a more centralized infrastructure. For a single client or proxy server, PMG can only rank users based on calls that are directed to that server.

3.3.4 Consent-Based Communication

Consent-based communication is a technique that is used in combination with black or white lists. If a caller is not on the receiver's black or white lists, then the caller needs to seek consent before a call will be put through. This may happen as part of the call itself, or may be part of an IM contact request sent before calling. This technique can lead to additional presence spam or IM spam in order to try to fool a user into adding a spammer to their white list. As with many other techniques, such a system works best with strong authentication of identities (Rosenberg & Jennings, 2008).

3.3.5 Payments at Risk

In this technique, the caller deposits a small amount of money into the receiver's account. If the receiver decides that the call is not spam, the amount is refunded, otherwise the receiver keeps the money. There are two transactions involved here. First is transfer of money from caller to receiver, and second is the return of funds to the caller. The first transaction occurs before the call is established and the second one occurs after it is proved that a call is not spam. The advantage of this technique is that it can make it expensive for spammers to send bulk spam. The small cost would add up to a large amount making it unattractive for spammers. The problem here is the extra overhead in the form of transaction costs. Also, there is a problem of inequities in the value of currency between caller and receiver. A relatively poor receiver might keep the deposit even if the call is legitimate, and a poor caller might not be able to afford a deposit which is sufficient to discourage a rich spammer.

A charging mechanism has been proposed in (Rebahi, Sisalem, & Magedanz, 2006). It makes users pay for sending SIP requests. This is managed by a server which provides authentication, authorization and accounting for the used services within a SIP community. However, adding a charging mechanism in a relatively unregulated environment is difficult and most likely highly unpopular.

3.3.6 Content Filtering

In email, content filtering is commonly used for spam prevention. Spam content filters analyze the text of emails and watch for particular signatures to see if the email is spam. Published approaches for this include the use of Bayesian filters, or Support Vector Machines (Liu & Cui, 2009).

Content filtering is successful and effective for email spam but when it comes to voice spam, it is of limited use. The reason is that when the receiver answers a call, it is only after the content is delivered that it can be analyzed. If the content is stored in the form of voicemail, the speech recognition engines are currently not sophisticated enough to analyze the content to detect if the call is spam. Even if speech is analyzed, it will incur a processing overhead. Spammers may also come up with varied accents and styles to baffle the speech recognition engine. On the other hand, since IM spam is similar to email spam, it can be detected with content filtering methods (Kolan, Vaithilingam, & Dantu, 2007).

3.3.7 Reputation Systems

This is another technique that tries to categorize users that are not on a black or white list, and is an enhanced version of consent-based communication (Rebahi, Sisalem, & Magedanz, 2006). When a user receives a consent request, a reputation score is associated with the prospective caller.

Reputation is calculated based on previous user assessments of the caller. For example, the VoIP user interface might include a button that terminates the call and marks the caller as a spammer. The input of any single user would not be sufficient to damage one's reputation, but continued negative feedback would increase the negative reputation score. Like many systems, reputation systems depend on strongly secure identities. There is also the possibility for a spammer with many identities to generate positive reputation feedback between these identities. If a legitimate user's identity is compromised and used to send spam then that will impact the user's reputation negatively, and it may be difficult to ever recover a positive reputation.

Rebahi, Sisalem and Magedanz (2006) propose a reputation based technique to identify and deal with spam. This mechanism uses the trust that a receiver has in a caller to distinguish between a spammer and a non-spammer. A reputation based technique has also been proposed by Wang, Mo and Huang (2007). In this method, a call request is judged as spam or otherwise by calculating the reputation of the caller. This is an evaluation based on four components: the user's white/black lists, the user mode (e.g. accepting urgent calls only), a call model (e.g. call density and call length) and users' subjective evaluations. This reputation evaluation is then securely transferred to the network using a hashing mechanism. A trust network propagates personal evaluations between peers, so that users can share their experience with other trusted users in order to avoid spam attacks.

3.3.8 Address Obfuscation

This refers to making addresses difficult or impossible to gather. Spammers usually build their spam lists as a result of collecting email addresses from websites. They look for text in the form of abc@xyz.com and assume that such a string of text is an email address. To avoid the addresses being collected, addresses can be obfuscated in a form like "abc at xyz dot com". Such addresses can be used by humans but cannot easily be identified by a machine as an email address. This technique can be used for SIP spam as well but it is not a particularly effective way of countering spam.

3.3.9 Limited-Use Addresses

This method is also related to address obfuscation. A user has a large number of addresses for incoming messages, and each of these is limited to particular uses, such as being limited to particular time intervals, or groups of users. Contact requests outside these situations can be refused.

This technique is applicable to SIP but the drawback is that it can make it hard for people to be reached. If an address becomes spammed, changing it requires informing every contact, which may be a difficult task to do.

In the extreme case, a different address is given to each correspondent. This way, not each and every correspondent needs to be notified of an address change. This scheme is called single-user address. Again, managing multiple addresses is a cumbersome process (Rosenberg & Jennings, 2008).

3.3.10 Computational Puzzles

This technique involves solving computational puzzles in order to be able to make a call. When a call request is made, the receiver's software requests the caller's software to perform a computational task and pass the result back. The nature of the computation is such that it is expensive enough to prevent spammers from sending bulk messages, but insignificant for legitimate users.

However, there is a problem associated with this technique. There may be significant variance in the computational power of the devices making and receiving calls. Spammers who use machines infected with malware to generate spam might get substantial computational power from those machines. Computational puzzles are under active research and it is expected that the outcome for email spam will likely be applicable to SIP spam (Clayton & Laurie, 2004).

3.3.11 Authenticated Identity

There are two types of authenticated identity techniques that can be used for VoIP spam prevention.

Sender Checks: Transport Layer Security (TLS) can be used to check identities between different domains. The address provided by the sender must match the domain from which it comes. If the domain address is authenticated and that domain is trusted to have authenticated the user, the whole sender address can be trusted. This requires all messages to be routed through trusted proxies for each domain.

Signature-Based Techniques: Domain Keys Identified Mail (DKIM) Signatures allow strong assertions to be made about user identity by signing email messages with digital signatures. It requires trusted authorities for providing digital signatures.

These techniques can also be applied to SIP calls.

The key to detecting spam is identifying the sender of the message. SIP Identity can provide the solution for this problem. First each domain needs to identify its users. Then domains can sign connection requests before sending to a different domain to validate that the SIP identity is correct (Peterson & Jennings, 2006).

The P-Asserted-Identity header field can be used to provide a weaker identity assertion (Jennings, Peterson & Watson, 2002). Mutual trust needs to be maintained amongst all interconnected domains. As the number of inter-connected domains increases, it becomes increasingly difficult to provide identity assertion. The overall strength of identity assertions is only as good as the weakest domain in the group (Rosenberg & Jennings, 2008).

3.3.12 Human Telephony Communication Patterns

This method is based on the study of human communication patterns in legitimate versus spam calls. Quittek et al. (2007) have analyzed certain patterns such as mutual silence, double talk and call start pattern. For example, if a call is answered by first sending an automated greeting to the caller, then a polite human caller would wait for that greeting to finish, but a spam call would immediately promote its product. So such a call would be classified as spam.

3.3.13 Turing Test Challenges

Another spam prevention technique is based on a Turing test (Browne, 1991) or Human Interaction Proof (HIP) which is a kind of test in which a human user interacts with a machine (computer) to prove that he or she is an actual human being and not some other machine or a bot. When used in the context of email, the sender of the message is given a challenge which is not easy for a software program to answer. If the challenge is answered correctly, the sender is authenticated (i.e. confirmed to be human) for that call, and depending on user preferences, is placed on the white list of the receiver. One example of such a challenge is CAPTCHA (Completely Automated Public Turing test to Tell Computers and Humans Apart) (Von-Ahn, Blum, & Langford, 2004). It asks users to read and type characters from a distorted image.

Humans can identify the distorted words relatively easily, but a software algorithm cannot. The strength of this technique depends on the ability of spammers to come up with artificial intelligence algorithms to crack the challenge. The cost of generating and verifying such tests is small, but to do the same computationally is quite difficult for current artificial intelligence software. This makes it difficult for the spammers to use automated techniques to send spam. Just as a Human Interaction Proof can be used to prevent IM spam, a Human Interaction Proof can also be used to prevent voice spam (Lindqvist & Komu, 2007).

Turing test techniques mainly rely on the complexity of test challenges. However, in the future, better pattern recognition techniques may be developed to circumvent Turing tests (Chellapilla & Simard, 2004). This may require the CAPTCHA engine to generate more complex sets of characters which may, in return, be difficult even for human beings to recognize correctly. Another consideration could be performance optimization. The process needs to be fast enough so as not to lock down the resources and allow the human caller swift access to the receiver. Also, there is a scalability issue to take into account. SIP clients who do not have the modified authentication process will not be able to communicate with each other.

Moreover, visually impaired people or calls made on devices without a graphical display would require an audio version of a CAPTCHA (Markkola & Lindqvist, 2008). This kind of voice-based Turing test can be used by using the SIP application interaction framework (Rosenberg, 2009). This framework provides information about the capabilities of the devices involved in the call so that an appropriate challenge can be given.

The key problem with a Turing test is that it only distinguishes between humans and machines. Low paid human workers can be used to quickly solve the challenges. Due to this problem, Turing tests may never completely counter spam.

4. Frameworks for SPIT Prevention

Each of the methods described above has its advantages and its disadvantages. As for email spam, a single technique is unlikely to be effective in all cases. Realistically, many of the techniques will need to be combined to effectively counter VoIP spam. So rather than a single technique, VoIP spam is best prevented by a layered framework.

Such a framework is presented in (Quittek et al., 2008). The idea is that at each layer of processing, a call can be classified into one of three classes-a valid call which is passed onto the receiver, a spam call which is rejected, or an undecided call which is passed onto the next stages of processing. The framework of Quittek et al suggests the following stages:

Stage 1: Modules at this stage use techniques such as black lists, white lists, reputation systems, circles of trust and anomaly detection. Stage 1 happens without caller interaction.

Stage 2: This stage requires caller interaction (but not the receiver). Modules here include gray lists, computational puzzles, sender checks, communication patterns and Turing tests.

Stage 3: Receiver interacts with Caller. This stage involves consent-based communications allowing users to be added to a white list.

Stage 4: Receiver receives the call-this is where content filtering techniques would be applied, but in general such techniques are not useful since the receiver is already disturbed.

Stage 5: After the call-reputation systems require the user to give positive or negative feedback on a caller.

The framework proposed by Quittek et al. (2008) consists of three layers. Stage 3 above is not considered in the framework, since it is really an ancillary part of white lists and does not involve a voice call. Stage 4 is not generally considered useful, so is not in the framework.

The framework consists of three layers.

Layer 1 consists of stage 1 modules which each give a score in [-1, 1] (higher means more likely to be spam). The weighted sum plus low and high thresholds are used to classify the message as valid, spam or undecided. Note that strong identity authentication would be needed for high scores on white lists.

Layer 2 deals with undecided messages from layer 1 and sequentially uses techniques from Stage 2 modules. If all modules are passed, then the call is forwarded to the receiver.

Layer 3 incorporates Stage 5 user feedback from the call into stage 1 (such as the user hanging up after identifying spam) modules.

This framework is a promising approach, although it has some drawbacks. It does not include support for preventing spam connection requests in consent-based communications and it doesn't deal with one place where content-based filtering is most useful, viz., in rejecting spam voicemail.

5. Implementation Costs

Key costs involved with any spam prevention system are the implementation cost in terms of the memory, storage and computation costs of the networked computing equipment, the time cost in terms of the additional workload imposed on valid users of the system, the system cost involved in requiring a particular network configuration (such as mutually trusted domain servers) and the inconvenience cost if the system either designates a valid user as a spammer, or fails to significantly reduce spam calls.

Algorithms which depend on a central server to identify spam calls (the server-based systems above) are unlikely to be successful because they require cooperation and agreement between domains, and often require all calls to be routed through a spam-checking server. These central servers are likely to be a cost and throughput bottleneck.

Client-based approaches are more likely to succeed, and are a better fit to the naturally distributed processing model of the Internet. Taking the framework described in section 4 above as an example, the costs of implementing a viable, distributed spam-prevention framework are quite reasonable.

A white list can be checked very quickly, even in the presence of strong authentication mechanisms that require checking of a digital signature. It would be expected that most calls would be in this category. Black lists may be somewhat longer, but techniques such as database hashing allow the lists to be checked quickly. Those calls that are not on either list need to pass through a number of tests, such as a Turing test, or a consent-based communications system. These impose a larger load on the caller, but this will only occur once if that caller is then added to the receiver's white list, but will occur for every spam call and will often be sufficient to discourage spamming. The cost to the receiver is the need to accept the communication request, and the need to provide feedback about whether the caller is added to black or white lists. This cost is moderate.

Overall, the costs involved in a client-based spam-prevention framework are quite manageable, and are not seen as an obstacle to adoption. There is still some work to do on appropriate standards about how to incorporate these techniques within the existing SIP protocol framework, such as exactly how to incorporate Turing tests in the SIP protocol. However, a major issue is that there is still a lack of clear evidence about the cost-benefits of many of the proposed techniques.

6. Conclusions

Spam over IP telephony remains a significant obstacle to the availability of low-cost, widely available, open VoIP systems. To date, the majority of VoIP uptake has either been as vendor-specific PABX replacements which use VoIP internally and connect more widely via the PSTN, or it has been in the form of closed-protocol systems such as Skype. The former systems depend on PSTN charges to prevent external spam, and systems like Skype use consent-based communications which limit their usefulness.

Many individual algorithms, systems and approaches have been proposed for reducing the impact of SPIT and these have been reviewed above. No single technique is ideal and the most realistic approach is the development of frameworks which are able to combine various techniques in a layered set of filters. Ideally, such a framework should provide broad support for all stages of VoIP spam including spam contact requests and voicemail.

Finally, there is lack of good experimental data on the cost, accuracy and ease-of-use of such complex, multi-module SPIT filtering systems.

References

Baset, S. A., & Schulzrinne, H. (2006). An analysis of the skype peer-to-peer internet telephony protocol. *25th IEEE International Conference on Computer Communications: INFOCOM 2006,* 1-11. http://dx.doi.org/10.1109/INFOCOM.2006.312

Baumann, R., Cavin, S., & Schmid, S. (2006). *Voice Over IP-Security and SPIT.* Swiss Army, FU Br 41, KryptDet Report. Retrieved from http://www.rainer.baumann.info/public/voip.pdf

Browne, R. (1991) The Turing Test and non - information flow. *IEEE Computer Society Symposium on Research in Security and Privacy,* 373-385. http://dx.doi.org/10.1109/RISP.1991.130804

Chellapilla, K., & Simard, P. Y. (2004). Using machine learning to break visual human interaction proofs (HIPs). In Saul, K., Weiss, Y., & Bottou, L. (Eds), *Advances in Neural Information Processing Systems, 17*, Cambridge MA: MIT Press, pp. 265-272.

Clayton, R., & Laurie, B. (2004). Proof of Work Proves not to Work. *3rd Annual Workshop on Economics and Information Security*, 1-9. Retrieved from http://oinvite.googlecode.com/files/ProofofWorkNoWork.pdf

Dantu, R., Fahmy, S., Schulzrinne, H., & Joao Cangussu, J. (2009). Issues and challenges in securing VoIP. *Computers & Security, 28*, 743-753. http://dx.doi.org/10.1016/j.cose.2009.05.003

Davidson, J., Peters, J., Bhatia, M., & Kalidindi, S. (2006). *Voice over IP Fundamental* (2nd ed.). Cisco Press.

Day, M., Rosenberg, J., & Sugano, H. (2000). *A Model for Presence and Instant Messaging, RFC 2778*. Retrieved from http://www.ietf.org/rfc/rfc2778.txt

Dierks, T., & Rescorla, E. (2008). *The Transport Layer Security (TLS) Protocol Version 1.2, RFC 5246*. Retrieved from http://www.ietf.org/rfc/rfc5246.txt

Hung, P. C., & Martin, M. V. (2006). Security Issues in VOIP Applications. *Canadian Conference on Electrical and Computer Engineering, CCECE '06*, 2361-2364. http://dx.doi.org/10.1109/CCECE.2006.277789

Jennings, C., Peterson, J., & Watson, M. (2002). *Private Extensions to the Session Initiation Protocol (SIP) for Asserted Identity within Trusted Networks, RFC 3325*. Retrieved from http://www.ietf.org/rfc/rfc3325.txt

Khasnabish, B. (2003). *Implementing Voice over IP*. Lexington. MA: John Wiley & Sons. http://dx.doi.org/10.1002/0471225274

Kolan, P., Vaithilingam, R., & Dantu, R. (2007). Automatic Calibration Using Receiver Operating Characteristics Curves. *2nd International Conference on Communication Systems Software and Middleware: COMSWARE 2007*, 1-8. http://dx.doi.org/10.1109/COMSWA.2007.382484

Lindqvist, J., & Komu, M. (2007). Cure for Spam over Internet Telephony. *4th IEEE Consumer Communications and Networking Conference: CCNC 2007*, 896-900. http://dx.doi.org/10.1109/CCNC.2007.181

Liu, S., & Cui, K. (2009). Applications of Support Vector Machine Based on Boolean Kernel to Spam Filtering. *Modern Applied Science, 3*(10), 27-31.

MacIntosh, R., & Vinokurov, D. (2005). Detection and mitigation of spam in IP telephony networks using signalling protocol analysis. *IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, 49-52. http://dx.doi.org/10.1109/SARNOF.2005.1426509

Markkola, A., & Lindqvist, J. (2008). Accessible voice CAPTCHAs for internet telephony. *Symposium on Accessible Privacy and Security (SOAPS)*, 1-2. Retrieved from http://cups.cs.cmu.edu/soups/2008/SOAPS/markkola.pdf

Park, S. Y., Kim, J. T., & Kang, S. G. (2006). Analysis of applicability of traditional spam regulations to VoIP spam. *8th International Conference in Advanced Communication Technology: ICACT 2006.* 1215-1217. http://dx.doi.org/10.1109/ICACT.2006.206189

Peterson, J., & Jennings, C. (2006). *Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP), RFC 4474*. Retrieved from http://www.ietf.org/rfc/rfc4474.txt

Quittek, J., Niccolini, S., Tartarelli, S., & Schlegel, R. (2008). On Spam over Internet Telephony (SPIT) Prevention. *IEEE Communications Magazine, 46*(8), 80-86. http://dx.doi.org/10.1109/MCOM.2008.4597108

Quittek, J., Niccolini, S., Tartarelli, S., Stiemerling, M., Brunner, M., & Ewald, T. (2007). Detecting SPIT calls by checking human communication patterns. *IEEE International Conference on Communications: ICC'07*, 1979-1984. http://dx.doi.org/10.1109/ICC.2007.329

Rebahi, Y., Sisalem, D., & Magedanz, T. (2006). SIP Spam Detection. *International Conference on Digital Telecommunications, ICDT '06,* 68-68. http://dx.doi.org/10.1109/ICDT.2006.69

Rosenberg, J. (2009). *A Framework for Application Interaction in the Session Initiation Protocol (SIP), RFC 5629.* Retrieved from http://www.ietf.org/rfc/rfc5629.txt

Rosenberg, J., & Jennings, C. (2008). *The Session Initiation Protocol (SIP) and Spam, RFC 5039*. Retrieved from http://www.ietf.org/rfc/rfc5039.txt

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., … Schooler, E. (2002). *SIP: Session Initiation Protocol, RFC 3261*. Retrieved from http://www.ietf.org/rfc/rfc3261.txt

Schlegel, R., Niccolini, S., Tartarelli, S., & Brunner, M. (2006). ISE03-2: SPam over Internet Telephony (SPIT) Prevention Framework. *IEEE Global Telecommunications Conference: GLOBECOM '06*, 1-6. http://dx.doi.org/10.1109/GLOCOM.2006.192

Schulzrinne, H., & Rosenberg, J. (1998). Signaling for Internet Telephony. *Sixth International Conference on Network Protocols*, 298-307. http://dx.doi.org/10.1109/ICNP.1998.723751

Shin, D., Ahn, J., & Shim, C. (2006). Progressive multi gray - leveling: a voice spam protection algorithm. *IEEE Network, 20*(5), 18-24. http://dx.doi.org/10.1109/MNET.2006.1705879

Skype. (2013). *Skype Terms of Use*. Retrieved from http://www.skype.com/en/legal/tou/

Varshney, U., Snow, A., McGivern, M., & Howard, C. (2002). Voice over IP. *Communications of the ACM, 45*(1), 89-96. http://dx.doi.org/10.1145/502269.502271

Von Ahn, L., Blum, M., & Langford, J. (2004). Telling humans and computers apart automatically. *Communications of the ACM, 47*(2), 56-60. http://dx.doi.org/10.1145/966389.966390

Wang, F., Mo Y., & Huang, B. (2007). P2P-AVS: P2P Based Cooperative VoIP Spam Filtering. *IEEE Wireless Communications and Networking Conference*, 3547-3552. http://dx.doi.org/10.1109/WCNC.2007.650

Copyrights