Prevention of Spam over IP Telephony (SPIT)

Juergen QUITTEK, Saverio NICCOLINI, Sandra TARTARELLI, Roman SCHLEGEL

Abstract

Spam over IP Telephony (SPIT) is expected to become a serious problem in the near future. It has the potential to become an even bigger problem than email spam, because the callee will be disturbed by each received SPIT call. This paper describes a new SPIT prevention method that is effective and acceptable for the call participants because it does not affect the callee at all and limits the interaction with the caller to an acceptable minimum. As foundation for building general SPIT prevention systems with this and other innovative methods, this paper proposes a reference architecture for SPIT prevention systems.

Keywords

spam, VoIP, Internet telephony, SPIT

1. Introduction

Spam is defined as the transmission of bulk unsolicited mails; it is considered to be one of the biggest problems the Internet has ever faced. With the increasing deployment of Internet telephony solutions, often referred to as Voice over IP (VoIP), it is commonly expected that a similar form of spam will affect also this area. This threat is known as SPIT (Spam over Internet Telephony) and it is defined as the transmission of unsolicited calls over Internet telephony.

The potential of SPIT to reduce productivity is much higher than email spam, because each SPIT call immediately disturbs the callee by the ringing phone. For ensuring the success of VoIP it will be crucial to provide effective prevention, particularly in public networks and at gateways between public and enterprise networks.

The transmission of unsolicited calls already exists in the traditional Public Switched Telephone Network (PSTN), where such calls are mostly initiated by telemarketers. However, the high cost of PSTN calls compared to email or VoIP communications limits the attractiveness of this form of advertisement for telemarketers. On the contrary, the costs a spammer would encounter using Internet telephony are substantially lower. A recent study[1)] reported that SPIT is roughly three orders of magnitude cheaper to send than traditional PSTN telemarketer calls.

Unfortunately, the nature of SPIT is too different from email spam for using spam prevention methods effectively on SPIT. Particularly, the content of SPIT cannot be checked before the callee is disturbed by the ringing phone.

This paper proposes a novel technique to identify SPIT calls based on typical voice communication patterns. However, as for email spam, a single method will not be sufficient to prevent sufficiently from all SPIT attacks. Therefore, this paper proposes a generic SPIT prevention system architecture that allows to flexibly integrate different SPIT prevention methods in order to minimize the inconvenience caused to initiator and recipient of the voice call (referred to as caller and callee) by the SPIT prevention system, while maximizing its effectiveness in allowing only legitimate calls to reach the callee.

2. Comparison of Traditional Spam versus SPIT

SPIT is a much bigger threat for users than email spam since it will interrupt the users immediately. A SPIT call makes the phone ringing and disturbs the callee. Email spam can be queued in the email program of the receiving user without disturbing the user until he looks at it; even when the mails are checked, the user can process a big number of them in a short amount of time identifying quickly the spam without the need of giving much attention to it. The disturbance in the case of SPIT calls is instead repeated multiple times.

Sending SPIT is technically eased by the fact that Internet Telephony protocols and systems have poor identity management (the same technical problem that is present in the mail systems).

An additional problem with SPIT is that most available tech-

nologies (coming from email spam prevention) are not useful since:

- the time scale is much different (mails are non-real-time communications while Internet Telephony calls are real-time communications);
- one of the most effective methods (namely content filtering) is not really applicable since the call has to be answered before the content is delivered. Furthermore, automatic methods based on speech recognition are currently too complex and language dependent to be deployed for VoIP calls.

In addition to these simple considerations that give us an overview of the potentiality of the SPIT threat, SPIT will mainly occur in the future because of the reduced costs "SPITters" would encounter in using the Internet Telephony with respect to the PSTN. A simple cost analysis shows how much difference in costs occurs between calls delivered using the PSTN and the public Internet. There are three layers at which we can expect differences in costs between spam over PSTN and over Internet Telephony:

- the costs of the system in terms of software;
- the costs of the system in terms of hardware;
- the costs per spam call;

The costs of the system in terms of software are basically not varying between the two different forms of voice spam (the software could be basically the same, it is just the hardware needed to connect to the network which changes). The costs of the system in terms of hardware are clearly in disfavor of the PSTN spammer (PSTN cards are much more expensive than network interface cards). As for the costs per spam call, they are in disfavor of the system for sending spam over PSTN because of the higher costs of the PSTN connections; a rough analysis speaks of three order of magnitude lower costs for a SPIT system. **Table 1** resumes the costs comparison and clearly shows the costs saving that SPIT systems are offering to possible telemarketers.

Table 1 Resume of costs comparison for call spam systems.

Costs	SPAM over PSTN	SPIT	Additional considerations
Software costs	X	X	X is depending on the signaling protocol
Hardware costs	10 Y - 100 Y	Y	Y is independent of the signaling protocol
Costs per spam call	about 1000 Z	Z	Z is independent of the signaling protocol

3. A Generic SPIT Prevention System Architecture

A SPIT prevention system has to meet some basic requirements in order to be effective.

- It must minimize the probability of blocking legitimate calls.
- It must maximize the probability of blocking SPIT calls.
- It should minimize the interaction required to the callee to determine whether a call is SPIT.
- It should limit the inconvenience caused to the caller that tries to place a legitimate call.
- It should be general enough to apply to different types of environments (e.g. office, home etc.), different cultures, and languages and so on.

In the literature, several methods have been proposed to prevent SPIT calls; however none of them meets all of these requirements. Besides, most effective methods in preventing SPIT require interaction with the caller and are therefore too intrusive, so that the caller might decide to tear down the call causing the callee to possibly miss important calls. Even worse, other methods require a feedback from the callee. An effective SPIT prevention system must therefore combine the capabilities offered by different component methods, so that the resulting system is able to efficiently block SPIT calls while requiring the least possible interaction with the caller and the callee.

Furthermore, we believe that, being the caller the one that starts the action, he or she is probably more willing to accept a certain level of inconvenience compared to the callee. For this reason, we consider a method involving the caller less intrusive than one requiring feedback from the callee.

Based on the above assumptions, we propose a generic architecture for SPIT prevention systems consisting of five stages with increasing intrusiveness, see **Fig. 1**.

At the first stage, prevention methods act invisible to the caller and callee. At stage two the prevention methods interact with the caller or at least with the caller's terminal. Stage three requires a feedback from the callee before the call is actually established, while stage four allocates those methods that judge a call while it is active. Finally, at stage five, feedback from the callee occurs after the call has been terminated and it contributes to blocking SPIT in the future.

At all stages either the automated mechanisms peculiar to the stage or the callee provide a feedback to the SPIT prevention system, which requires such knowledge as input for some of the modules in the first stage.

Furthermore, an incoming call will not have necessarily to pass through all stages. For instance, a call which has already

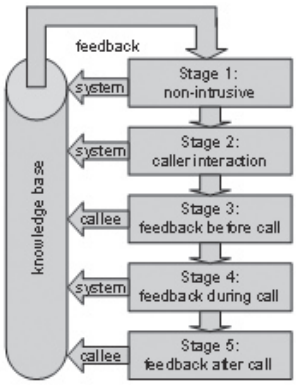

Fig. 1 Generic SPIT prevention system architecture.

been recognized as legitimate by the first stage does not need to be further controlled and can directly be established, i.e. passed to stage four. In general, the actual path followed by a call depends by implementation specific factors, like the level of intrusiveness accepted by the system.

The discussion of methods for different stages in the next section shows that, in general, higher effectiveness can be achieved by higher stage methods which are more intrusive by nature. We assume that a very good SPIT prevention system will have to be composed of multiple methods and cover all stages while balancing well intrusiveness and effectiveness.

4. Building Blocks for SPIT Prevention

Several methods are under discussion as potential building blocks of SPIT prevention systems, see Reference 1) and references therein. In this section, we provide an overview of known methods. We structure the section by mapping each method to the stage of our generic architecture in Fig. 1 at which the method can be applied. Each method is detailed with a short description and with comments whether the prerequisites (e.g. in terms of infrastructure, standardization activities, etc.) for a given module already exist, whether it is realistic to achieve them in the near future. Furthermore we provide a rough estimate of how difficult it would be to implement a module.

4.1 Stage 1: No Interactions with Call Participants

Methods in this section do not require interaction with the caller and are completely invisible to him/her. The details are shown in **Table 2**.

Table 2 Stage 1 methods.

Stage 1 Module	Details
Lists	The identity of a caller is compared to a set of stored identities to decide whether to accept or reject a call. Two types of lists are considered: white and black. Identities on a white list are the ones which are allowed to call while calls from identities which are on a black list should be rejected. Such a methodology has no obstacles to its implementation since related technology is mature and implementation is easy.
Circles of Trust	Trusted inter-domain connections are checked before a call is forwarded to the intended recipient. The rationale is that each domain controls its own users and the domains agree not to send SPIT/SPAM to each other. Such a methodology has no obstacles to its implementation since related technology is mature and object of current standardization; the implementation itself has medium complexity.
Pattern/Anomaly Detection	This method tries to detect suspicious patterns in traffic to identify SPIT calls. Suspicious patterns can be either defined by deterministic or by statistical rules. Such a methodology has no obstacles to its implementation since related technology is mature but has never been adapted to voice calls; the implementation itself has medium complexity.
Greylisting	It is a technique referred to as PMG (Progressive Multi Grey-leveling). This method monitors calls and attribute a grey level to each caller. If a caller keeps trying to place calls in a certain time span the grey level increases. If the caller stops placing calls the grey level slowly decreases. Such a methodology has already been implemented by other researchers[3)] and the implementation itself has medium complexity.

4.2 Stage 2: Interaction with Caller

Methods in this section require either an interaction with the caller's terminal (computational puzzles, sender checks) or with the caller (Turing test). The details are shown in **Table 3**.

4.3 Stage 3: Callee Interrupted by SPIT Call

Methods in this section require - at least sometimes - an action by the callee on arrival of a SPIT call. The details are shown in **Table 4**.

4.4 Stage 4: Callee Receives Call

Methods in this section require that the callee receives the call and operate while the call is active. From the current literature, only the content filtering method falls into this category, which as explained below is not suited for SPIT. However,

Table 3 Stage 2 methods.

Stage 2 Module	Details
Computational Puzzle	This technique gives the caller's terminal a resource-consuming task to perform before establishing the call. Computational puzzles in conjunction with SIP are currently being standardized by the IETF and implementation has medium complexity.
Sender Check	The idea behind this method is to verify that a caller is a valid sender for the domain he is calling from. Such a methodology, even if the technology is mature, is not easily applicable to real-time communications and thus to SPIT prevention because of the time it takes to perform the sender checks.
Turing Test	Turing tests are a method to tell humans and computers apart. These tests are also called CAPTCHAs (Completely Automated Public Turing Test to Tell Computers and Humans Apart). Audio CAPTCHAs are indeed suited to prevent SPIT even if technology is prone to errors; the implementation itself has medium complexity.

Table 4 Stage 3 methods.

Stage 3 Module	Details
Consent-Based Communication	This solution requires user A to authorize user B, the first time user B tries to contact user A. A framework for consent-based communications combined with lists is currently being standardized by the IETF for the SIP protocol; the implementation itself has medium complexity.

Table 5 Stage 4 methods.

Stage 4 Module	Details
Content Filtering	Content analysis cannot be applied to prevent SPIT. The content is very different from email (ASCII text versus coded speech) and voice recognition is not yet fully solved and consuming a lot of computational resources. Moreover, the content is not available when the check needs to be performed (the content is delivered online after the receiver has been disturbed by a ringing phone and after the receiver accepted the incoming call).

potentially other methods might be proposed in the future that fit in this category; therefore we deem it useful to allocate a stage in the proposed architecture to take them into account. The details are shown in **Table 5**.

4.5 Stage 5: Feedback from Callee after Call

Methods in this section require that the callee gives feedback on calls received. The details are shown in **Table 6**.

Table 6 Stage 5 methods.

Stage 5 Module	Details
Reputation System	A reputation system works by attaching a reputation score to a contact indicating if this contact has been showing good or bad behavior. This score can be most effectively evaluated based on user feedback but it could also be tied to other building blocks. Such a methodology has no obstacles to its implementation since technology is mature but the feedback framework needs to be standardized; the implementation itself has medium complexity.
Limited-Use Addresses	The limited use of addresses is a mechanism which tries to defeat spam by changing the address as soon as the first spam messages arrive at the address. Such a methodology has no obstacles to its implementation since the technology is mature but the implementation has high complexity because of the need to change addresses as soon as the first SPIT call is received.
Payments at Risk	Payments at risk works by charging a fee for the first contact and refunding that fee if the call was not SPIT and adding the caller to a white list. This technique requires a feedback mechanism to indicate whether a call was SPIT or not and a payment infrastructure for micro-payments. Given the second prerequisite this methodology seems to be quite unrealistic and also implementation is estimated to have high complexity.
Legal Action	This method works by introducing legislation in all countries to prohibit the distribution of spam over VoIP. Even if implementation is foreseen to be relatively easy the methodology is quite unrealistic because of the lack of a global legislation framework.
First-Contact Feedback	This method relies on a mechanism where the callee can provide a feedback. The idea is that an unknown caller identity is allowed to call the callee exactly once and then the callee has to provide a feedback on this call. Such a methodology has no obstacles to its implementation since technology is mature but the feedback framework needs to be standardized; the implementation itself has medium complexity.

5. SPIT Prevention at Stage 2

This section describes a novel SPIT prevention method based on the analysis of human communication patterns. Its design and integration into a SPIT prevention system is based on considerations stated in Section 3.

The major goal of a SPIT prevention system is protecting the callee from being disturbed by SPIT calls while ensuring that the callee does not miss legitimate calls. Ideally, this is achieved in a way which is friendly to the caller. However, methods suitable for stage one, i.e. invisible to both caller and callee in many cases are not effective enough; therefore, a certain level of intrusiveness has to be taken into account.

In order to avoid disturbing the callee by any SPIT call, we focused our contribution on an innovative Turing test (Stage 2 method) for SPIT prevention. At stage 2, the SPIT prevention system accepts an incoming call on behalf of the callee and performs a check. Depending on the result, the call is then either forwarded to the callee or terminated, see **Fig. 2**.

Assuming that SPIT calls would rather be performed by computer programs than by human callers, we searched for a strong method that separates human callers from machines. Such a Turing test should meet the following seven requirements:

1) be polite enough not to offend the caller,
2) be quick enough not to require too much patience from the caller,
3) work well with callers that have different kinds of background knowledge,
4) work well with caller using different kinds of pronunciation,
5) work well with callers speaking different dialects or languages,
6) be simple enough to be implemented on a relatively cheap devices at the callee's side,
7) create a resource-intensive and complex task for a machine imitating a caller.

We developed a Turing test based on checking human communication patterns that meets all these requirements and provides a high degree of protection for the callee, while at the same requiring only a very limited and adjustable level of inconvenience for the caller.

5.1 Checking Communication Patterns

Our Turing test is based on the assumption that human conversation follows certain activity patterns. There are available studies that demonstrated that such communications patterns are identifiable, see for example 2). When a human caller calls another human being, then there are certain conventions that both follow. After the callee accepts the call, the callee is the first one to speak. During the call, typically one speaker is silent while the other one is speaking. The proposed Turing test checks if the caller follows these conventions.

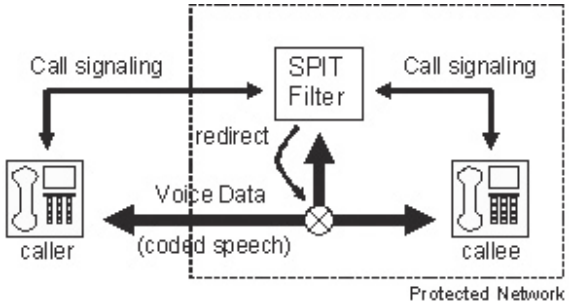

Fig. 2 SPIT filter at stage 2.

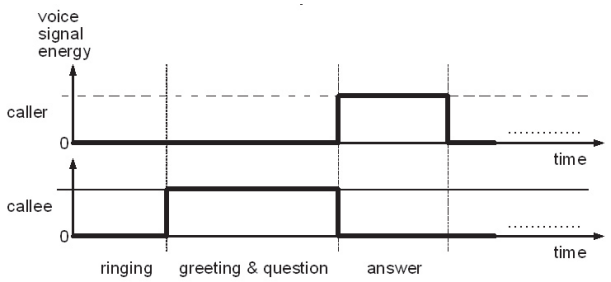

Fig. 3 Basic voice signal energy pattern from caller to callee (top) and from callee to caller (bottom) at the beginning of a call.

Fig. 3 shows an example communication pattern indicated by the voice signal energy of the two voice connections involved: caller to callee (top) and callee to caller (bottom). While the phone is ringing, the energy on both voice connections is low or zero.

This pattern concerns the voice signal energy only. It is independent of the content of the voice signal and can therefore be implemented easily and be performed with very few resources.

5.2 SPIT Detection Scheme

When the SPIT prevention system accepts the call it sends a pre-recorded greeting message to the caller that can be adapted to the level of intrusiveness that is assumed to be acceptable.

If the caller interrupts the greeting he is either impolitely not following the common communication pattern, or it is a machine that immediately starts its SPIT message. In both cases, the SPIT prevention system would classify the call as SPIT and terminate it, optionally after sending a pre-recorded voice message explaining this.

The intrusiveness of this method can be minimized by the SPIT detection system accepting the call, but instead of a greeting sending a ring tone. A human caller would then assume that the call has not yet been established, while a SPIT engine that does not analyze the greeting message would assume that it can start sending the SPIT message, because an established connection was signaled.

More intrusive, but still assumed to be well acceptable, is a greeting message that tells the caller that the call is being forwarded and will be established soon.

For a stronger check, the greeting can be followed by a quick simple question, for example, for the name of the called person.

Such a question should be made such that a short answer can be expected with high probability. Then the SPIT prevention system can check if first the caller starts speaking briefly after the question was made and stops talking and remains silent for at least a short time after answering the questions. For both checks, no speech recognition is necessary. Detection of the voice energy level switching from 'low' to 'high' and back to 'low' is sufficient. If this energy pattern cannot be observed, then the SPIT filter assumes that the caller is a machine and terminates the call.

5.3 Evaluation and Practical Experiences

This procedure meets all seven requirements listed in the previous section. It is sufficiently polite and quick, independent of the caller's background knowledge and pronunciation and can be adapted to the caller's preferred language (if signaled by the caller at call setup). The voice energy analysis is computationally simple while the effort required on the SPIT engine side for passing this test is high. Furthermore, the level of intrusiveness can be selected according to the preferences of the callee. Experiments with a prototype implementation have shown that voice signal energy detection cannot just distinguish between 'full silence' and 'some signal energy,' because the caller may be in an environment creating background noise (in a bus, at a train station). Therefore, the detection system needs to use experimentally confirmed threshold levels for 'low' and 'high' voice energy. Also short peaks of signal energy originating in the caller's environment should be filtered and not detected as 'high' voice signal level.

A SPIT prevention system using this method should still use stage 1 methods in addition. Particularly, white lists and black list should be maintained. A white list is convenient because it avoids that known callers are checked. Also the results of the Turing test can be fed back to white and black lists. A caller that has been rejected by the Turing test for a given number of times in a row should be added to the black list, while callers passing the test one or more times should eventually be added to the white list.

6. Conclusions

This paper proposes a novel technique to identify SPIT calls based on typical voice communication patterns. For integrating it with other SPIT prevention methods, the paper further proposes a generic SPIT prevention system architecture that allows to flexibly integrate different SPIT prevention methods in order to minimize the inconvenience caused caller and callee. The architecture distinguished different stages that differ in the level of interaction with caller and callee. The stages model was used to classify the set of already known SPIT prevention methods.

We elaborated requirements for realizing such a recommended system and described a SPIT prevention system design. As innovative component of this design that meets all requirements we developed a stage 2 Turing test for distinguishing human callers from SPIT engines. The test detects the reaction of the caller on a greeting message and compares it to common human communication patterns when communicating over the phone. The method limits interaction with the caller in a configurable way and can be performed with low computational resources.

References

1) J. Rosenberg et al., "The Session Initiation Protocol (SIP) and Spam," draft-ietf-sipping-spam-01.txt, July 2005. Work in progress.
2) F. Hammer et al., "Elements of Interactivity in Telephone Conversations," Proc. 8th International Conference on Spoken Language Processing (ICSLP/INTERSPEECH 2004), Vol.3, pp.1741-1744, Jeju Island, Korea, Oct. 2004.
3) D. Shin, and C. Shim, "Voice Spam Control with Gray Leveling," Proc. of 2nd VoIP Security Workshop, Washington DC, June 1-2 2005.

Authors' Profiles

Juergen QUITTEK
Senior Manager,
Network Laboratories,
NEC Europe Ltd.

Saverio NICCOLINI
Research Staff,
Network Laboratories,
NEC Europe Ltd.

Sandra TARTARELLI
Senior Research Staff,
NEC Europe Ltd.

Roman SCHLEGEL
Swiss Federal Institute of Technology