On Spam over Internet Telephony (SPIT) Prevention

Juergen Quittek, Saverio Niccolini, Sandra Tartarelli, and Roman Schlegel, NEC Europe Ltd.

ABSTRACT

Spam over IP telephony (SPIT) is expected to become a serious problem in the near future. One of the main requirements for a SPIT prevention system is to avoid the involvement of the callee in the SPIT detection process, because this implies disturbing the callee each time an unclassified potential SPIT call arrives. Further requirements include adaptability to different deployment scenarios and flexibility to quickly react to new kinds of SPIT that bypass existing prevention systems. This article analyzes the requirements for SPIT prevention, provides a thorough classification of currently known SPIT prevention methods, and introduces a reference model for SPIT prevention systems. As an instance of the reference model, we designed and implemented an advanced SPIT prevention system, composed of methods that avoid unnecessary callee interaction, that fulfills a set of requirements while remaining adaptive in order to be customized for different scenarios.

INTRODUCTION

Spam is defined as the transmission of unsolicited email; it is considered one of the biggest problems the Internet has ever faced. Today, far more spam emails than regular emails are transmitted in the public Internet. Among other reasons, the spam problem became so widespread because there were no solutions readily available before the problem arose. Nowadays, there are methods available that are able to counteract this problem using different approaches, but none of these methods constitutes a definitive solution.

With the increasing deployment of Internet telephony solutions, it is commonly expected that a similar form of spam will show up in this area. This threat is commonly referred to as spam over Internet telephony (SPIT) or voice over IP (VoIP) spam. SPIT is defined as the transmission of unsolicited calls over Internet telephony.

Unsolicited calls already exist in the traditional public switched telephone network (PSTN), where such calls are mostly initiated by telemarketers but are limited in number because of the relatively high cost of a PSTN call. Using Internet telephony, these costs are substantially lower; spam software is much easier to program for the Internet Protocol, and a spammer can multiplex a number of calls on a single line. A recent study [1] reported that IP-based SPIT is roughly three orders of magnitude cheaper to send than traditional telemarketing calls.

Taking into account that IP-based applications can infect unprotected machines on the Internet and create botnets, spamming in parallel from huge numbers of these machines, the cost of IP-based SPIT can decrease even more, making SPIT very attractive to telemarketers.

The conclusion of these considerations is a strong need for SPIT prevention systems to be expected in the near future.

This article presents a reference model for SPIT prevention systems and provides a classification for a large set of available prevention methods. Concrete instances of the system and selection of prevention methods supported by an instance may vary significantly depending on the application scenario and location of the system. A SPIT prevention system at a peering point between operators, for example, would have a different structure and use different methods than a SPIT prevention system at a VoIP terminal. As an example instance of the reference model, we describe our implementation of a SPIT prevention system [2, 3].

The remainder of the article is organized as follows. We discuss related work; we suggest a reference model for SPIT prevention systems and section 4 provides a thorough classification of known methods for SPIT prevention. We present considerations on the application of the reference model to specific scenarios while we report on our specific system implementation and we then conclude the article.

RELATED WORK

Ongoing standardization activities attempt to solve the SPIT problem, by promoting the use of strongly authenticated identities [4] together with white lists (to allow known users to call) and a consent framework (to deal with the initial contact problem). A similar approach is presented in [5], which uses reputation-based Session Initiation Protocol (SIP) social networks. However, methods relying on consent-based communication and reputation systems are currently not mature enough, and still require further stan-

 0163-6804/08/$25.00 © 2008 IEEE

dardization work before they can be effectively deployed among interoperable distributed systems.

Other papers investigate methods that do not require wide adoption and can be instead implemented even as standalone solutions.

In [6] the authors propose a technique called Progressive Multi Grey-leveling. In [7] a SPIT firewall that uses fingerprinting to identify the calling devices is presented. A more widely applicable approach is followed in [8], where a SPIT prevention entity receives VoIP data from equipment at the edge of an operator's network and scores suspicious session initiations. Despite the growing interest of the research and standardization communities in solutions for SPIT prevention, a consolidated overview of the available techniques is still missing. The methods reviewed in [1] represent a comprehensive list; however, a more rigorous categorization is required. This article proposes a reference model for SPIT prevention systems, which allows the mapping of the different methods into well defined stages, helping system designers in the development of SPIT prevention systems.

Reference Model for SPIT Prevention Systems

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

- It must have an adaptive strategy for minimizing the probability of blocking legitimate calls.
- It must have an adaptive strategy for maximizing the probability of blocking SPIT calls.
- It should minimize the interactions with 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 (office, home, etc.), different cultures, languages, and so on.
- It should be flexible in order to cope with SPIT getting more sophisticated as email spam does.

None of the methods to prevent SPIT calls proposed in the literature meets all of these requirements. An effective SPIT prevention system must combine the capabilities offered by different methods so that the resulting system is able to efficiently block SPIT calls while requiring the least possible interaction with caller and callee.

Based on the above assumptions, in this article we propose a reference model for SPIT prevention systems that proposes a classification scheme for SPIT prevention methods where the methods are divided into five stages with increasing intrusiveness (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 feedback from the callee before the call is actually established, while stage four employs those methods judging a call while it is ongoing.

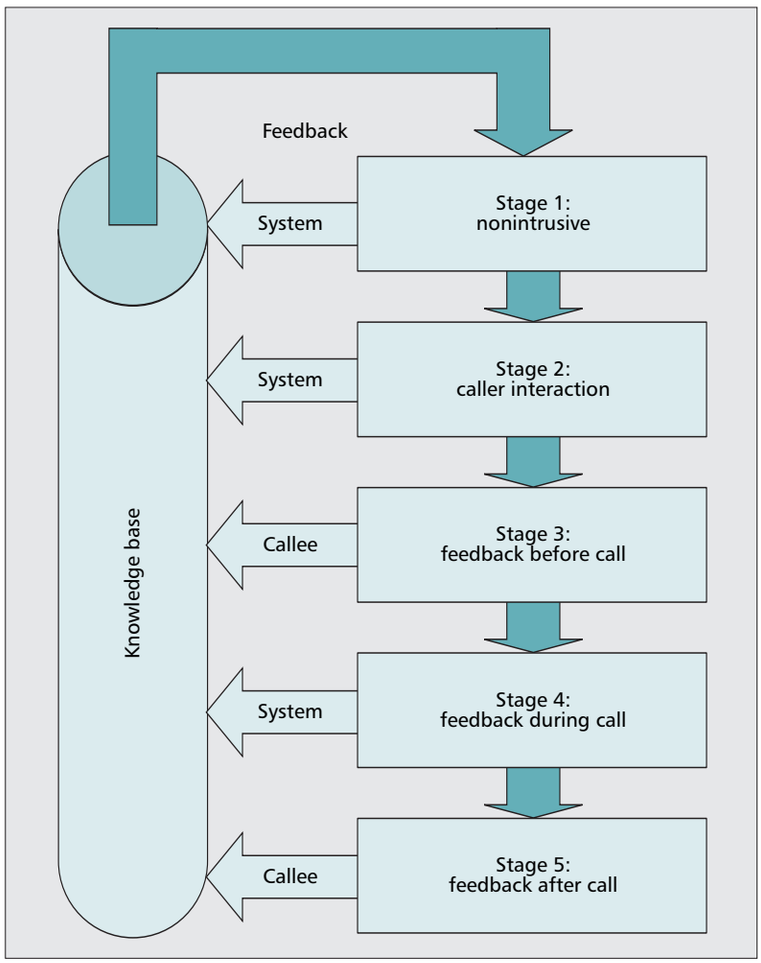

Figure 1. *Reference model for SPIT prevention systems.*

Finally, at stage five, feedback from the callee occurs after the call has been terminated and contributes to blocking SPIT in the future.

At all stages either the system or the callee provides 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 does not necessarily have to pass through all stages. For instance, a call that has already been recognized as legitimate by the first stage does not need to be further inspected and can directly be established (i.e., passed to stage four). In general, the actual path followed by a call depends on 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, a trade-off between intrusiveness and effectiveness can be observed.

Building Blocks for SPIT Prevention

Several methods are under discussion as potential building blocks of SPIT prevention systems [1, references therein]. In this section we provide an overview of known methods. Each method is

Although lists can be applied to SPIT, both black and white lists have some drawbacks. A white list requires explicit permission for every identity which is allowed to call which raises the problem of the initial contact. A black list can easily be circumvented if there is an infinite supply of SIP identities.

mapped to the stages of our reference model in Fig. 1.

Stage 1: No Interactions with Call Participants

Lists — Lists are a simple mechanism where the identity of a caller is compared to a set of stored identities to decide whether to accept or reject a call. There are two different kinds of lists, white and black. Identities on a white list are the ones allowed to call, while calls from identities on a black list should be rejected.

Members of white lists and black lists may be configured manually, but they may also be added as a result of other methods acting on stages 1–5.

Although lists can be applied to SPIT, both black and white lists have some drawbacks. A white list requires explicit permission for every identity allowed to call, which raises the problem of the initial contact. A black list can easily be circumvented if there is an infinite supply of SIP identities.

Lists are in any case more effective if they are used in combination with authenticated identities [1].

Circles of Trust — This solution works by introducing trusted interdomain connections. Each domain controls its own users, and the domains agree not to send SPIT to each other. This method can be implemented in SIP by using authenticated Transport Layer Security (TLS) connections between domains [9].

Pattern/Anomaly Detection — This method works by detecting suspicious patterns in VoIP traffic to identify SPIT calls. Although this approach is in principle very general, we believe that a realistic implementation would try to correlate the arrival time of a call and possibly the identity of the caller with known statistical or deterministic patterns of SPIT calls. Based on this correlation, the module decides whether the incoming call might be SPIT or not. Like all methods working with patterns or statistical data, they suffer from the drawback of possibly generating false positives, resulting in legitimate calls being blocked.

Stage 2: Caller Interactions

Methods in this section require either interaction with the caller's terminal (computational puzzles, sender checks) or with the caller directly (Turing test, or Completely Automated Public Turing test to tell Computers and Humans Apart, CAPTCHA).

Greylisting — With greylisting, the first call from an unknown user is rejected with the message to call again within a given time interval. If this call happens as specified, it will pass the greylisting procedure, and the caller will not be bothered by it for future calls.

Greylisting is a very efficient method for blocking email spam using a built-in feature of the Simple Mail Transfer Protocol (SMTP). But for email, greylisting does not require user interaction, because resending the email is handled by the involved mail servers. For SPIT, greylisting would require user interaction. Asking the caller to call again would require playing a prerecorded message or an extension of existing signaling protocols.

Computational Puzzle — Computational puzzles in conjunction with SIP are currently being discussed in the Internet Engineering Task Force (IETF). The basic idea is giving the caller's terminal a resource-consuming task to perform before establishing the call. This way potential SPIT generators are limited in the number of calls they can initiate in a given time interval. However, botnets could be used to distribute the cost of computing puzzles; therefore, the effectiveness of this method is limited.

Sender Check — The idea behind this method is to verify that a caller is a valid sender for the domain from which he is calling. The required information is, for example, stored in DNS records.

When used with VoIP, this only works when no forwarding is done (i.e., on trapezoid connections [1]).

Turing Test — Turing tests are a conversational method to tell humans and computers apart where the judge is a human being. Today similar tests (still belonging to the class of Turing tests) are used to secure Web sites from being accessed by bots. These tests are also called CAPTCHAs [10] since the judge is a computer instead of a human. Because the Web is first and foremost a visual medium, most CAPTCHAs are visual, although audio CAPTCHAs exist as well. In pure voice systems only audio CAPTCHAs can be used.

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.

Consent-Based Communication — This solution requires user A to authorize user B the first time user B tries to contact user A. It solves the first contact problem but introduces a delay until the first call can be placed. A framework for consent-based communications combined with lists is currently being standardized by the IETF for SIP.

Stage 4: Callee Receives Call

Methods in this section require that the callee receive the call; they operate while the call is active.

Content Filtering — Most of the methods for blocking email spam that are based on content analysis cannot be applied to prevent SPIT. First, the content is very different (ASCII text vs. coded speech), and voice recognition is not yet fully solved and consumes a lot of computational resources. Second, the content is not available when the check needs to be performed.

Stage 5: Feedback From Callee After Call

Methods in this section require that the callee give feedback on calls received.

Stage	Module	Prerequisites	Implementability
1	Lists	Exists	Easy
1	Circles of trust	Realistic	Medium
1	Pattern/anomaly detection	Exists	Medium
2	Greylisting	Realistic	Easy
2	Computational puzzles	Exist (being standardized)	Medium
2	Sender checks	Realistic	Hard
2	Turing test	Realistic	Medium
3	Consent	Exists (being standardized)	Medium
4	Content filtering	Exists	Hard
5	Reputation	Exists	Medium
5	Limited use	Realistic	Hard
5	Payment at risk	Unrealistic	Hard
5	Legal action	Unrealistic	N/A
5	First-time feedback	Realistic	Medium

Table 1. *Classification of SPIT prevention methods.*

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.

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 could also be tied to other building blocks. Reputation systems with negative reputation scores suffer from the same problem as black lists. The system can also be abused if there are large groups collaborating and systematically giving a certain feedback to chosen identities.

Limited-Use Addresses — The limited use of addresses is a mechanism that tries to defeat spam by changing the address as soon as the first spam messages arrive at the address. There are two drawbacks to this method: first, a new address has to be communicated to all existing contacts; and second, a new user has to be able to get the current address of a recipient.

Payments at Risk — Payments at risk works by charging a fee for the first contact, refunding that fee if the call was not SPIT, and adding the caller to a white list.

This technique requires two prerequisites. First, there has to be some kind of feedback mechanism so that the callee can indicate whether or not a call was SPIT. Second, it requires a payment infrastructure for micropayments, and every VoIP user has to have an account for the system to be effective.

A variant of this method is the "interdomain SIP providers" technique [1] where a small number of providers acts as SIP equivalents to the interexchange carriers in the PSTN, charging local SIP providers per message forwarded.

Legal Action — This method works by introducing legislation in all countries to prohibit the distribution of spam over VoIP. The problem here is that law enforcement on an international level is unreliable. There will always be countries where it is legal to send SPIT.

First-Contact Feedback — This method also relies on a mechanism where the user can provide feedback to the server. The idea is that an unknown identity is allowed to call exactly once, and then the callee has to provide feedback. If the callee provides negative feedback, the caller is, for example, put on a black list. If the callee provides positive feedback, the caller is put on a white list.

Summary of Building Blocks

Table 1 summarizes the above survey of existing VoIP prevention methods and indicates 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, or whether it is unlikely that they will exist anytime soon. Furthermore, it provides a rough estimate of how difficult it would be to implement the method. Finally, it specifies the stage to which the method would belong when considering the framework shown in Fig. 1.

INSTANTIATING THE PREVENTION MECHANISMS

The reference model described earlier can be applied in a wide range of scenarios from SPIT prevention at a high-throughput peering point between VoIP operators to SPIT prevention at a VoIP terminal. However, concrete instances of the system for different scenarios may vary significantly. Also, the selection of applied prevention methods strongly depends on application requirements and constraints.

Instances of SPIT prevention systems installed at high-throughput locations in a network (e.g., at VoIP operator peering points or softswitches) are strongly limited concerning the interaction with caller and callee, because of the potentially high resource consumption and increase of connection setup delay.

Particularly suited for high-throughput applications are methods for pattern and anomaly detection. They are still applicable to medium-throughput applications, such as at firewalls and enterprise access points. But for these applications, more customizable stage 2–5 methods can be considered.

Also, close interaction with the callee is possible for medium-throughput systems and enterprise systems. Particularly, the maintenance of per-user black and white lists including user feedback on stages 3–5 is possible.

At a terminal, several SPIT prevention methods are no longer applicable. But prevention methods in stages 3–5 that interact with the callee are very well suited for terminals. The same considerations about limited applicability of lower-stage methods also apply in the case of P2P environments. Also in this case, higher-stage methods seem to be more practical (selected methods from stage 2 as well as from stages 3–5).

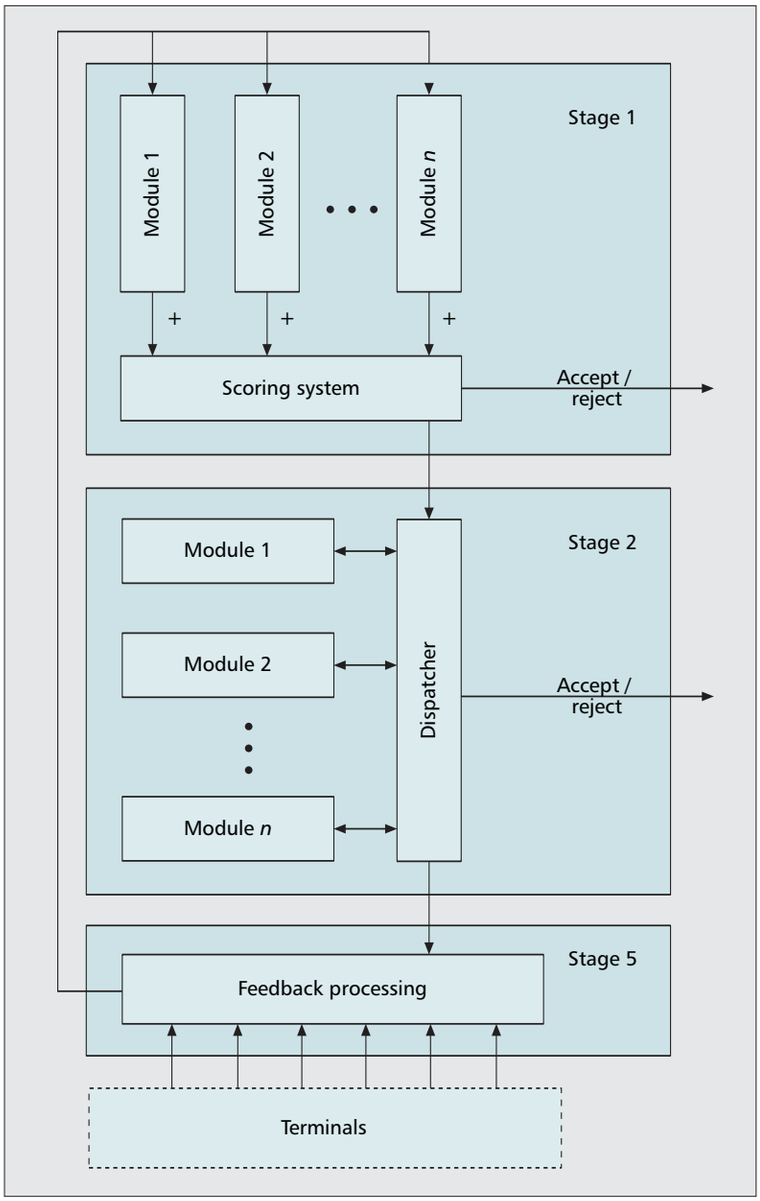

Figure 2. *SPIT prevention system design.*

SPIT PREVENTION SYSTEM IMPLEMENTATION

The SPIT prevention system described here is targeted at an enterprise access point scenario, but the design is extensible to easily be adapted to other scenarios and able to flexibly respond to the expected evolution of sophisticated SPIT attacks.

In order to meet all requirements listed earlier, we used a modular design for our prevention system that flexibly allows linking stages and uses a modular approach for stages 1 and 2. This way, prevention methods can be implemented as modules and added, configured, updated or removed at runtime in order to react quickly on policy changes, user requests, or innovations on the SPIT sender side.

Some of the modules we developed for stages 1 and 2 are innovations themselves designed explicitly to deal with the SPIT threat (Fig. 2).

One important design choice for the enterprise scenario was to prefer a strong newly developed stage 2 method instead of also using stage 3 methods that would have unnecessarily disturbed staff members at work. A callee is only involved in stage 5 if he or she receives a SPIT call despite the efforts of the prevention system.

SPIT PREVENTION AT STAGE 1

Our stage 1 design is based on a scoring system similar to common email spam filters. All modules examine incoming call signaling and produce a score in a normalized range of [–1, 1]. The score indicates the likeliness of the call being SPIT with a high score indicating high likeliness. The total score for each call is the weighted sum of all modules. Weights can be configured per module as well as the thresholds for the overall score.

The total score is compared to two thresholds (a low and a high). If it is below the lower threshold, the call is forwarded to the callee. If it is between the lower and higher thresholds, the detection process is not complete, and the call is forwarded to the second stage modules for further processing. Otherwise, if the total score is above the higher threshold, either the call is

rejected or forwarded to a voicemail system to not suppress the communication (which may be a legal requirement).

SPIT Prevention at Stage 2

Different from stage 1, at stage 2 modules containing prevention methods are called sequentially until a final decision on accepting or rejecting a call has been made. A configurable dispatcher calls the modules, processes their results, and makes the decision.

For this method the dispatcher accepts the call on behalf of the callee and invokes the stage 2 module selected (e.g., a Turing test using voice communication). If the test is successfully passed, the dispatcher forwards the call to the original callee by referring the call, as shown in Fig. 3. Otherwise, the call is either immediately terminated or recorded.

We designed and implemented an efficient and discrete stage 2 module belonging to the "Turing Test" building block. It is based on the assumption that human conversation follows certain activity patterns [3, 11]. When a human caller calls another human being, there are certain conventions that both call participants 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 event of double talk is limited in duration to a fraction of seconds and to only 6 percent of cases on average [12]. The stage 2 module checks if the caller follows these conventions (see Fig. 4 for a graph depicting voice signal energy pattern of conversation participants in a normal case).

When the stage 2 module accepts the call, it sends a prerecorded greeting message to the caller that can be adapted to the level of intrusiveness assumed to be acceptable.

If the caller interrupts the greeting, he or she is either impolitely not following the common communication pattern, or it is a machine that immediately starts its SPIT message. In both cases the stage 2 module would classify the call as SPIT. A call classified as SPIT can be either rejected or recorded depending on configured options.

The intrusiveness of this method can be minimized if, instead of a greeting message, the prerecorded sound of a ringing phone is transmitted. 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 quite acceptable, is a greeting message that tells the caller that his/her call is being forwarded and will be established soon.

For a stronger check, the greeting can be followed by a quick simple question, such as what is 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 stage 2 module can check if 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, the stage 2 module assumes that the caller is a machine.

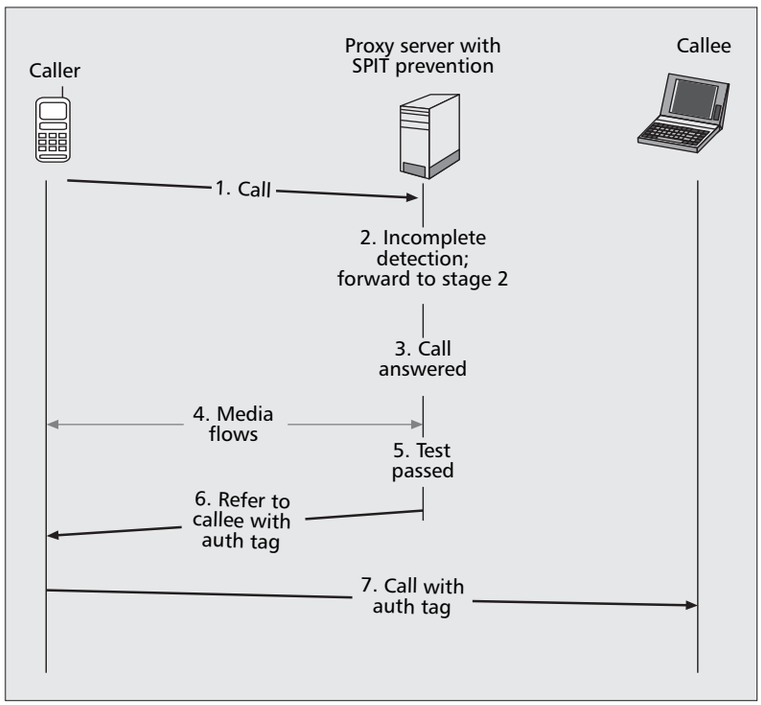

Figure 3. *SPIT detection with stage 2.*

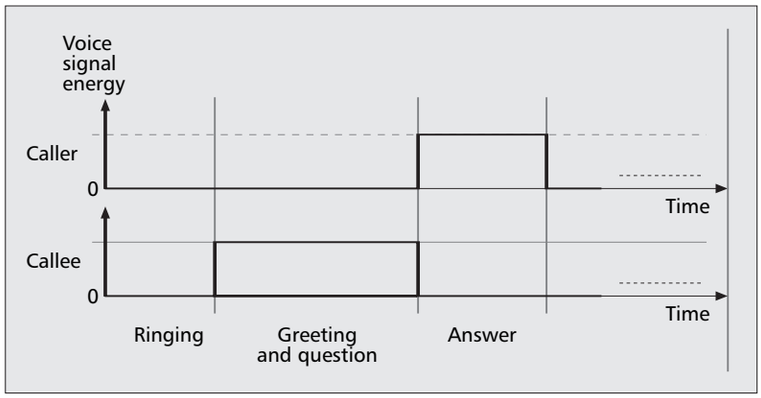

Figure 4. *Basic voice signal energy pattern from caller to callee (top) and from callee to caller (bottom) at the beginning of a call.*

SPIT Prevention at Stage 5

We extended a software client with an additional hang-up button that allows the user to terminate the call, and at the same time indicates to the SPIT prevention system that this was SPIT. Our modified client inserts an additional header to the call termination message for this purpose. Such user feedback is processed by a stage 5 feedback receiver and forwarded to stage 1 modules in order to refine the scoring of subsequent calls. For example, the white/black list module may add the caller identity to the black list.

We extended a software client with an additional hang-up button that allows the user to terminate the call and at the same time indicating to the SPIT prevention system that this was SPIT. Our modified client inserts an additional header to the call termination message for this purpose.

CONCLUSIONS

This article presents a thorough classification and reference model for SPIT prevention building blocks with the purpose of helping newcomers and practitioners in the area to improve their understanding, and better design and implement SPIT prevention systems. Also, recommendations about methods' implementability and instantiation of the building blocks are reported together with an innovative implementation.

REFERENCES

[1] J. Rosenberg and C. Jennings, "The Session Initiation Protocol (SIP) and Spam," IETF RFC 5039, Jan. 2008.
[2] R. Schlegel *et al.*, "SPam over Internet Telephony (SPIT) Prevention Framework," *Proc. IEEE GLOBECOM '06*, San Francisco, CA, Nov.–Dec. 2006.
[3] J. Quittek *et al.*, "Detecting SPIT Calls by Checking Human Communication Patterns," *Proc. IEEE ICC '07*, Glasgow, Scotland, June 2007.
[4] J. Peterson and C, Jennings, "Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP)," IETF RFC 4474, Aug. 2006.
[5] D. Sisalem and Y. Rebahi, "SIP Service Providers and The Spam Problem," *Proc. 2nd VoIP Security Wksp.*, Washington, DC, June 2005.
[6] D. Shin and C. Shim, "Voice Spam Control with Gray Leveling," *Proc. 2nd VoIP Security Wksp.*, Washington, DC, June 1–2 2005.
[7] H. Yan *et al.*, "Incorporating Active Fingerprinting into SPIT Prevention Systems," *Proc. 3rd Annual VoIP Sec. Wksp.*, Berlin, Germany, June 1–2, 2006.
[8] B. Mathieu *et al.*, "SPIT Mitigation by a Network-Level Anti-Spit Entity," *Proc. 3rd Annual VoIP Sec. Wksp.*, Berlin, Germany, June 1–2, 2006.
[9] T. Dierks *et al.*, "The TLS Protocol Version 1.0" RFC 2246, Jan. 1999.
[10] L. von Ahn *et al.*, "Telling Humans and Computers Apart Automatically," *Commun. ACM*, vol. 47, no. 2, Feb. 2004.
[11] F. Hammer *et al.*, "Elements of Interactivity in Telephone Conversations," *Proc. 8th Int'l. Conf. Spoken Language Processing*, vol. 3, Jeju Island, Korea, Oct. 2004, pp. 1741–44.
[12] ITU-T Rec. P.59, "Artificial Conversational Speech."

ADDITIONAL READING

[1] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," IETF RFC 3261, June 2002.
[2] M. Hansen *et al.*, "Developing a Legally Compliant Reachability Management System as a Countermeasure Against SPIT," *Proc. 3rd Annual VoIP Sec. Wksp.*, Berlin, Germany, June 1–2, 2006.

BIOGRAPHIES

JUERGEN QUITTEK (quittek@netlab.nec.de) received his M.S. and Ph.D. degrees in communication engineering at the University of Aachen and Hamburg, Germany. After working as a postdoctoral fellow at the University of California at Berkeley, he joined NEC Laboratories Europe, Germany. There he is the deputy general manager, responsible for the whole networking area. He is a Technical Programming Committee member of several IEEE conferences and chairs two IETF working groups in the area of network management.

SAVERIO NICCOLINI (niccolini@netlab.nec.de) received his M.S. and Ph.D. degrees in telecommunication engineering from the University of Pisa, Italy, in 2000 and 2004, respectively. In 2004 he joined NEC Laboratories Europe, Heidelberg, Germany, where he is currently the manager of the Real-Time Communications group. His research interests are related to voice over IP security, traffic measurements, and P2P networking.

SANDRA TARTARELLI (tartarelli@netlab.nec.de) received her M.S. and Ph.D. degrees in telecommunication engineering, both from the University of Pisa in 1996 and 2001, respectively. In November 2000 she joined NEC Laboratories Europe, where she is currently working as a senior research staff member. Her recent research includes traffic measurements, Voice over IP security, traffic engineering, quality of service provisioning, and load balancing in wireless networks.

ROMAN SCHLEGEL (schlegel@netlab.nec.de) did his Master's thesis at NEC Laboratories Europe on SPIT prevention. He received his M.S. in communication systems in April 2006 from the Swiss Federal Institute of Technology, Lausanne.