Detection and Filtering Spam over Internet Telephony - A User-behavior-aware Intermediate-network-based Approach

Yan Bai[1], Xiao Su[2] and Bharat Bhargava[3]

[1]Institute of Technology, University of Washington Tacoma, Tacoma, WA 98402, USA
[2]Department of Computer Engineering, San Jose University, San Jose, CA 95192, USA
[3]Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA

ABSTRACT

VoIP applications have gained popularity due to largely reduced cost and wider range of advanced services, as compared to traditional telephone networks. However, SPIT (Spam over Internet Telephony), known as unsolicited bulk calls sent via VoIP networks, is becoming a major problem that would undermine the usability of VoIP. Unlike detection and filtering of e-mail spam, countermeasures against SPIT face great challenges on how to identify and filter SPIT in real time. In this paper, a user-behavior-aware anti-SPIT technique implemented at the router level for detecting and filtering SPIT is proposed. The rationale for the technique is that voice spammers behave significantly different from legitimate callers because of their revenue-driven motivations. The technique defines and combines three features developed from user behavior analyses to detect and filter spam calls. Compared to existing SPIT defending techniques, it is simple, fast and effective. Other advantages of our approach are that it is applicable for detecting and filtering both machine-initiated and human-initiated spam calls, better protects VoIP calls against sybil attacks and spammer behavior changes.

Keywords
Spam, filtering, SPIT, spam over Internet Telephony, Voice over IP

1. INTRODUCTION

VoIP applications will become as ubiquitous as emails. ABI Research predicts that VoIP subscribers worldwide will reach to 267 million in 2012. For spammers, sending commercial messages over VoIP networks is appealing because it can be done quickly and inexpensively.

However, for VoIP users, receiving spam calls is extremely annoying, as the unwanted messages clutter their voice mailboxes and interferes with their normal activity every time the phone rings. Furthermore, spam calls waste a large amount of network bandwidth and would cause the delays for other network traffic.

In this paper, we present a user-behavior-aware anti-SPIT technique implemented at the router level for detecting and filtering SPIT. Section 2 describes related works, Section 3 discusses the proposed technique, Section 4 provides the simulation results, and Section 4 summarizes the work and describes future work.

2. RELATED WORKS

A significant amount of work has been done to defend against e-mail spam. Most of existing e-mail spam filters, such as SpamAssassin, SpamBouncer, or Mozilla Junk Mail Control, employ content-based filtering. The main idea of content-based techniques is to classify an e-mail into an unsolicited or a good one by checking some features in its content and filters a spam message after it has been delivered and stored in the receiver's mail server [1]. This approach is not suitable for combating SPIT because the requirements for detection and filtering of SPIT are significantly different from defending e-mail spam. Typical VoIP calls use Session Initialization Protocol (SIP) and comprise of two phases, a call setup phase and a media session. Any call handling decision must be made in real-time before actual media session starts. Furthermore, VoIP calls must be delivered to the user synchronously, so there is no entity like email server that stores the voice data for anti-SPIT processing.

Current anti-SPIT technologies can be classified into list-based filtering, reputation-based filtering, Turing test and pattern-based filtering approaches. All these anti-SPIT techniques contribute to the reduction of the quantity of SPIT that users receive. However, they are still in a primitive stage [11].

- List-based Filtering: allowing good calls from the white list or blocking the spam calls from the black list, or

temporarily rejecting unclassified calls from the grey list [2, 3, 4, 5]. This approach is susceptible to sybil attacks [5].

- Reputation-based Filtering: using buddy list and user ratings generate reputation scores, which will determine the acceptance or rejection of the caller [6]. This scheme restricts the scope of good calls to users' social network linkage, requires users' feedback to form ratings, and relies on the selection of an appropriate reputation threshold.

- Turing Test: distinguishing human callers from automatic SPIT generators running on a botnet [8]. It performs Turing tests based on human conversation patterns, and is not applicable for human-initiated SPIT.

- Pattern-based Filtering: monitoring the call patterns, such as the call frequency, and connecting a call if current call pattern matches the previous call pattern of the caller, otherwise, blocking a call [9, 10]. It cannot work effectively when spammers evade filters by changing their call patterns.

Furthermore, all the techniques described in the above are deployed at the sending end or receiving end, which incurs a large amount of end-to-end delay. This delay can result in either the call being answered or ending up in the callee's voice mail.

To address these challenges, we have focused on designing a SPIT defending technique that is applicable for detecting and filtering both machine-initiated and human-initiated spam calls, better protects VoIP calls against sybil attacks and spammer behavior changes. Moreover, our anti-SPIT technique can be implemented at a router, thus reducing the delay time and decreasing the probability of accepting spam calls.

3. THE PROPOSED SCHEME

The proposed scheme for detecting and filtering SPIT is called User-behavior-aware anti-SPIT technique. It consists of user-behavior-aware analysis and filtering.

Intuition

To design an anti-SPIT technique, we first analyze the caller/callee interaction behaviors under normal calls and spam calls. We found out that there is a fundamental difference between legitimate users and spammers on making and receiving calls. A legitimate caller typically makes and receives calls, while a spammer makes a large number of calls but seldom receives a call. Apparently, a ratio of answered calls and dialed calls can be used to distinguish a legitimate caller and a spammer.

The other key observation is the caller/callee historical behaviors that a legitimate user makes calls to their buddies and, typically, calls the same number more than once, while a spammer calls as many callees as possible and thus seldom repeats dialing the same number. Clearly, a ratio of repeated calls and distinct calls can be used to differentiate a legitimate caller and a spammer. A distinct call refers to a call number that has been dialed only once.

Our third observation is caller/callee social behaviors. A legitimate caller usually calls their buddies, while a spammer often calls a large number of unknown callees. An unknown callee refers to an individual who does not call back the caller. Thus, a ratio of calls to unknown users and the total number of callees can be used to classify a legitimate caller and a spammer.

User-behavior-aware Anti-SPIT

Based on the above analysis into the different behaviors between a legitimate caller and a spammer, we propose three features to identify spam calls:

- Interaction Ratio (IR): is defined as the ratio of answered calls and dialed calls of a caller.

- Historical Ratio (HR): is defined as the ratio of repeated calls and distinct calls of a caller.

- Social Ratio (SR): is defined as the ratio of unknown callees and total number of callees of a caller.

Combing the three features in our anti-SPIT technique, our design objectives are achieved. Firstly, Interaction Ratio is suitable for detecting both machine- initiated and human initiated spam calls. In either case, a spammer produces a smaller value of IR compared to a legitimate caller. Secondly, both Historical Ratio and Social Ratio should be kept unchanged in a time span because spammers have strong revenue-driven motivations. They do not have a lot of incentives to change either Historical Ratio or Social Ratio in a certain period of time. If either ratio is modified too quickly, spammers will definitely reduce their net profits. Fig 1 shows the procedure of our User-behavior-aware filtering.

Procedure ***User-behavior-aware filtering***
A new call request is passed at a router,
Calculate IR, HR and SR for the caller
if IR < X
    then the new call is rejected;
else if HR < Y;
    then the new call is rejected;
else if SR > Z;
    then the new call is rejected
else the new call is admitted;
end ***User-behavior-aware filtering***

Fig.1. Procedure of User-behavior-aware Filtering

In Fig. 1, X, Y, and Z are predetermined values to distinguish a legitimate caller and a spammer.

Implementation

To our best knowledge, most anti-SPIT techniques are deployed at the sending or receiving side, which incurs a large amount of end-to-end delay. Consequently, they might unintentionally falsely accept a relatively large number of spam calls as good calls. In our approach, we attempt to reduce the delay time and decrease the probability of accepting spam calls. Unlike all the existing anti-SPIT techniques, our User-behavior-aware anti-SPIT technique operates inside the network, at the router level.

A VoIP call involves either a request from a caller to a callee or a response from a callee to a caller. Both request and response are passed through the routers. This suggests that, by a simple analysis of request and response messages, a router could control SPIT by monitoring all Session Initiation Protocol (SIP) sessions passing through a router and classifying each SIP session as spam or legitimate and finally block the SPIT traffic.

By recording the number of SIP connections established by each host on the router, and the number of SIP servers each host connects to, we can easily generate a filtering information database at the router according to the three ratios described in our User-behavior-aware filtering technique.

Although the router-level implementation may increase the amount of processing at a router, we believe it is a viable approach because of largely reduced delay, and thus, the critical real-time requirements of VoIP calls can be met. Our implementation can benefit from special-purpose high-speed network processors [4].

4. EVALUATION

The proposed User-behavior-aware anti-SPIT technique was evaluated through simulations. Since there are few SPIT data in the public domain, we generate random distribution of good and spam calls to evaluate our technique.

To simulate router-level implementation, an additional module is added to a router. The proposed anti-SPIT algorithm is installed at the module prior to the start of simulation. During the simulation, the voice data are passed to the module for processing. Once processed, the VoIP calls are either blocked or accepted according to the algorithm. As for comparison, we also implement the proposed technique at user-level, in which no additional module is added to the router. We evaluate the technique based on the following performance metric:

- Accuracy: the effectiveness of correctly classified good calls and spam calls. It is calculated according to the following formula:

$$\frac{1}{2}\left(\frac{\text{correctly classified good calls}}{\text{Total number ofgood calls}}+\frac{\text{correctly classified spam calls}}{\text{Total number ofspam calls}}\right)$$

Router-level Implementation

Fig. 2-3 show the results of router-level implementation of our user-behavior-aware anti-SPIT technique for a simulated VoIP data set. The set contains 10% spam calls.

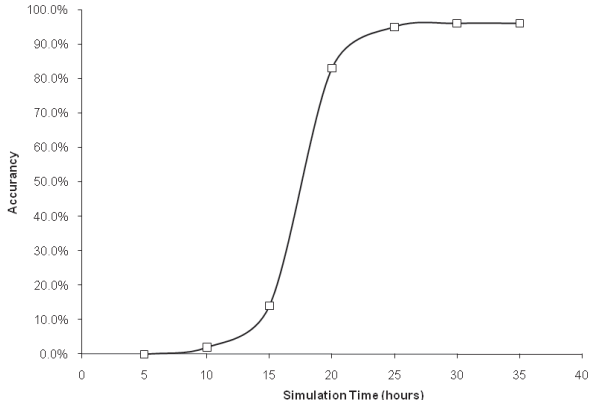

Fig. 2. Effectiveness of the Proposed Technique (Call Density of 30 calls/hour)

In Fig. 2, initially, we have no/less knowledge about spammers. Many spam calls cannot be detected, while good calls are blocked. The misclassification produces a lower filtering accuracy. However, one-day after the user-behavior-aware anti-SPIT technique is applied, 96% of calls are correctly classified.

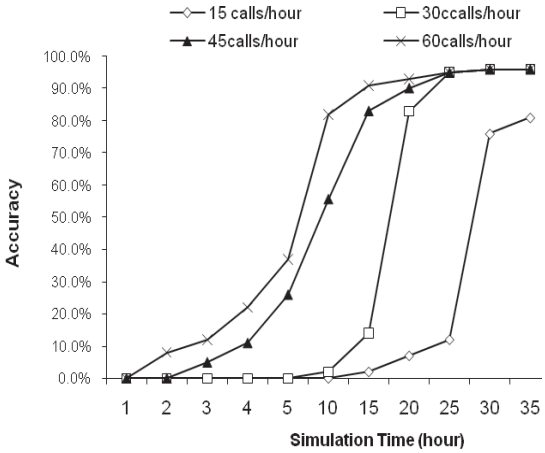

Fig. 3. Varing the Call Density.

In Fig. 3, over 80% calls are correctly classified for a call density of 45calls/hour and 60calls/hour in about half

day, while less than 10% calls are correctly classified for a call density of 15calls/hour and 30calls/hour in the same time span. So, the anti-SPIT technique needs less time to achieve high filtering accuracy when spammers generate more spam calls in a short period of time.

Comparison of Router-level and User-level Implementations

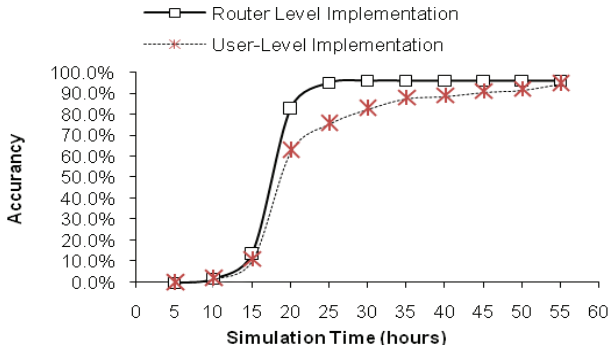

Fig. 4. Comparison of Router-level and User-level Implementations

Fig. 4 compares the results of router-level and user-level implementations of our user-behavior-aware anti-SPIT technique for a simulated VoIP data set with a call density of 30 calls/hour. As shown in Fig. 4, router-level implementation uses 50% less time than the user-level implementation to reach the high filtering accuracy (i.e., 96%). The training period decreases because the transmission delay from caller to a router is relatively small in comparison from caller to callee and router has access to more information on calls than a single user does. The results suggest than with intermediate assistance, our technique can detect and filter spam call effectively and efficiently, which meets the real-time requirements of VoIP calls.

We have also been performed simulation experiments for random divisions of good calls and spam calls. The results agree with the above conclusions.

5. CONCLUSIONS

Given the challenges of anti-SPIT, it is unlikely that current approaches are able to totally eliminate unsolicited bulk calls. In this paper, the behavior characteristics of call participants under normal calls and spam calls are analyzed. A user-behavior-aware filtering method is proposed, by which the Interaction Ratio, Historical Ratio, and Social Ratio are calculated and used to identify and block spam calls. Simulations are performed to validate its efficacy. The results show that the proposed technique can eliminate 96% of spam calls while using 50% less time than the user-level implementation.

The major contributions of this paper are two-fold. Firstly, the proposed technique provides a simple way to reduce SPIT: the acceptance or rejection of a new call request is determined by three ratios. Secondly, the technique is implemented at a router, rather than at the receiving end, thus reducing consuming network bandwidth and improving delay performance.

Like other anti-spam techniques, the performance of the proposed anti-SPIT technique also depends on the size of training dataset. We are currently investigating how to optimize the training dataset thus decreasing both false positive rate and false negative rate and further improving the overall accuracy of our anti-spit technique.

REFERENCES

1. A. Khorsi, "An Overview of Content-based Spam Filtering Techniques", *Informatica*, vol. 31, no. 3, October 2007, pp 269-277.
2. J. Rosenberg and C. Jennnings, "The Session Initiation Protocol and Spam", *IETF Draft*, draft-ietf-sipping-spam-04.txt, February 2007.
3. M. Hasen and et.al., "Developing a Legally Compliant Reachability Management System as a Countermeasure Against SPIT", *Third Annual VoIP Security Workshop*, Berlin, Germany, June 2006.
4. B. Agrawal, and et.al., "Controlling Spam Emails at the Routers", *IEEE International Conference on Communications (ICC*), Seoul, Korea, May 2005..
5. B. N. Levine and et.al., "A Survey of Solutions to the Sybil Attack", *Tech report 2006-052*, University of Massachusetts Amherst, Amherst, MA, October 2006.
6. Y. Rebahi and D. Sisalem, "SIP Service Providers and the Spam Problem", *the 2nd Workshop on Securing Voice over IP*, Washington DC, USA, June 2005.
7. A. Balasubrmaniyan and et.al., "CallRank: Combating SPIT Using Call Duration, Social Networks, and Global Reputation", *Conference on Email and Anti-Spam (CEAS2007)*, Mountain View, CA, August 2007.
8. J. Quittek and et.al., "Detecting SPIT Calls by Checking Human Communication Patterns", *IEEE International Conference on Communications (ICC)*, Glasgow, Scotland, June 2007.
9. D. Shin and C. Shim, "Voice Spam Control with Gray Leveling", the 2nd Workshop on Securing Voice over IP, Washington DC, USA, June 2005.
10. B. Sterman, "A Security Model for SPIT Prevention", *the 2nd Workshop on Securing Voice over IP*, Washington DC, USA, Jun. 2005.
11. V. M. Quinten, and et.al., "Analysis of Techniques for Protection Against Spam over Internet Telephony", *Lecture Notes in Computer Science*, Springer-Verlag, No. 4606, 2007, pp.70-77.