Two-stage SPIT Detection Scheme with Betweenness Centrality and Social Trust

Miho Kurata, Kentaroh Toyoda, and Iwao Sasase
Dept. of Information and Computer Science, Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama 223-8522, Japan,
Email: toyoda+kurata@sasase.ics.keio.ac.jp

***Abstract*—Detecting SPIT (Spam over Internet Telephony) is an urgent demand with growing voice communication services. Chaisamran et al. proposed a trust-based SPIT detection scheme and it is superior to other trust-based schemes since it is valid for unknown users. However, this scheme might misdetect a call from low-frequent users as SPIT gradually in time, since they seldom call and thus their social trusts get decreased with time. In this paper, we propose a two-stage SPIT detection scheme using BC (Betweenness Centrality) and social trust to decrease misdetection of a call from low-frequent users as SPIT. BC indicates user's centrality in the entire network and the value of BC against legitimate users gradually increases with time even if users seldom call, whereas it does not increase against spammers since the connection between a legitimate caller and a spammer is hardly established. As a first stage, we use BC as a feature to correctly identify a call request from a low-frequent user. Then we judge whether a call is from a legitimate caller or a spammer by using social trust. By the computer simulation, we show that our scheme improves the false positive rate while maintaining high true positive rate.**

I. Introduction

Various voice communication services are popular with the growing smartphone market. Recently, not only VoIP (Voice over Internet Telephony) service providers but also social network service providers e.g., Facebook and LINE, start voice communication services. They provide telephony services with very low price (even free of charge).

However, it is reported that malicious users or companies may abuse them for advertisement or fraud, which is called SPIT (Spam over Internet Telephony) [1]. Hence, detecting SPIT calls or spammers in voice communication services is an urgent demand for the service providers. Judging whether a call is SPIT or not is more difficult than detecting an e-mail spam since the content of a call cannot be obtained before ringing. Therefore, the system detects SPIT calls or spammers by using users' call histories. In particular, a social trust-based approach is receiving much attention due to growing SNS-based voice communication services. A social trust-based approach judges the legitimacy of a call with a trust value calculated from caller-callee relationships. Although many social trust-based schemes are proposed [2]–[10], we pay attention to the scheme [9] proposed by Chaisamran et al. since it can correctly classify unknown users. They use call duration as the trust value and the longer a user calls to a callee, the higher trust value the callee gives to the caller. This notion comes from the fact that the call duration from a spammer is short and a spammer seldom receives calls in general. They propose a trust inference scheme to deal with the case when a callee receives a call from an unknown user. However we notice that the scheme [9] raises false alarms for low-frequent legitimate users as time goes on. That is, the trust value of low-frequent users gradually decreases since they seldom receive calls. Hence it is necessary to propose a remedy for such low-frequent users.

In this paper, we propose a two-stage SPIT detection scheme with BC (Betweenness Centrality) and social trust. We use BC as a feature to judge whether a low-frequent user is legitimate or not at the first stage. After allowing a call from a low-frequent one at the first stage, we judge the legitimacy of a call by using the social trust-based approach [9]. BC indicates how much a user is gone through shortest paths between paths among other users. The intuition behind utilizing BC is that spammers call towards users while they seldom receive calls and thus spammers tend to be 'isolated' at the edge of the entire network. On the other hand, the value of BC against legitimate users gradually increases with time even if users seldom call. This is because legitimate users are gradually getting involved into the network of legitimate users.

We first show that the value of BC for legitimate users gradually increases with time whereas the that of spitters does not increase. Then, we show that our scheme improves the false positive rate while maintaining high true positive rate by the computer simulation. We also show that calculating BC is not computationally heavy and the BC can be fully computed in the off-line manner. Thus our scheme does not cause delay for the call establishment.

The rest of this paper is constructed as follows: we show the system model in Section II. We then explain related work and the conventional scheme in Section III and Section IV, respectively. The proposed scheme is described in Section V. Simulation results are shown in Section VI. We show our discussion in Section VII.

II. System Model

A. Spammer Model

We define a voice-based spammer as the attacker model throughout the paper. The aim of SPIT is advertisement, voice phishing, and illegal sales. Spammers call towards randomly chosen users but they seldom receive calls from others. In general, the call frequency is higher than that of legitimate

users. Since the contents of SPIT seems to be not interesting to ordinary users, the call duration tends to be much shorter than that of legitimate users.

B. Legitimate User Model

Legitimate users are defined as ordinary users and they communicate with their friends and unknown callees. Legitimate users may subscribe paid contents. There exist from low-frequent users to high-frequent users. Such contents distributors can be identified as spammers from other (non-subscribed) users and thus a subscriber enrolls them into his/her buddy list (friend list) to avoid subscribed call from being blocked.

C. Server Model for SPIT Detection

We assume that a SPIT detection system is deployed in a voice communication service provider and its task is to judge whether a call request should be established or not when receiving a call request from a user. We assume that as many as N_{user} users (including both legitimate users and spammers) in the service provider and the system can access to users' CDR (Call Detail Records) and buddy lists for the inspection.

III. Related Work

Many SPIT detection schemes with call histories are proposed in both industries and academia and they can be classified into the features-based approaches [11]–[16] and the social trust-based approaches [2]–[10]. The feature-based approach classifies users into legitimate callers or spammers based on call features e.g., call frequency and duration. The social trust-based approach classifies a call (or a caller itself) with a trust value calculated from caller-callee relationships. In particular, the latter approach is receiving much attention since it can be extensible for a growing SNS-based voice communication service. CallRank is the first SPIT detection scheme that uses social trust [2]. CallRank uses call duration to establish social network links and reputations. Kusumoto et al. propose to use clustering coefficient to score each user [5]. Seedorf et al propose to use Web-of-Trust as a reputation to mitigate SPIT [3]. Azad and Morla propose the Caller-REP which uses individual trust and global trust [7]. Although many social trust-based schemes are proposed, we pay attention to the scheme [9] proposed by Chaisamran et al., since it can classify a call from an unknown caller.

IV. Conventional Scheme

Chaisamran et al. propose a voice-based SPIT detection scheme with a social trust [9]. This scheme always allows a call from a user in the callee's buddy list. Otherwise, i.e., if a call is from an unknown user, the system judges whether a establishing call is legitimate or not by using an inferred trust value calculated from trust values of other users. Since multiple paths between an unknown caller u and a callee v may exist, the system chooses the maximum inferred trust $T_{u\to v}$ as shown in Eq. (1).

$$T_{u\to v} = \max_{p\in P_{u\to v}} \left(T_{u\to v}^{\text{path}_p} \right), \tag{1}$$

where $P_{u\to v}$ denotes a set of paths between u and v and $T_{u\to v}^{\text{path}_p}$ indicates an inferred trust value calculated between users in a path p and is represented as Eq. (2).

$$T_{u\to v}^{\text{path}_p} = \prod_{i\in p} T_{i(t)}, \tag{2}$$

In order to make a trust value reliable, each user is assigned a trust value from his/her friend depending on the cumulative call duration. This will give a low trust value for spammers since they often call to users but seldom receive calls and the call duration of SPIT is generally shorter than that of the legitimate call. More specifically, a user i in the path p has its own trust value at time t as Eq. (3).

$$T_{i(t)} = \alpha R_{i(t)} + (1-\alpha) T_{i(t-1)}, \tag{3}$$

where α denotes a weight variable whose range is $\alpha \in [0,1]$ and a raw trust value $R_{i(t)}$ at time t is represented as (4).

$$R_{i(t)} = \frac{C_{v(t)}}{\left(\prod_{j=1}^{n} C_{j(t)} \right)^{\frac{1}{n}}}, \tag{4}$$

where $C_{j(t)}$ denotes the cumulative call duration that a user j calls to user v and n denotes the number of v's friends, respectively.

Finally, the system compares the inferred trust value $T_{u\to v}$ and the pre-defined threshold th_T. If $T_{u\to v} > th_T$, the system establishes a call request from a user u to v. Otherwise, it rejects a call request.

A. Shortcomings of the Conventional Scheme

The conventional scheme enables the system to infer the trust of a caller even if he/she does not have direct relationships with a callee. However we notice that the scheme [9] raises false alarms for low-frequent legitimate users as time goes on. That is, the trust value of low-frequent users gradually decreases since they seldom receive calls. Hence it is necessary to propose a remedy for such low-frequent users.

V. Proposed Scheme

Here, we propose a two-stage SPIT detection scheme with BC and social trust in order for the system to correctly identify a call request from low-frequent legitimate users as a legitimate one. We use BC as a feature to allow a call from low-frequent users at the first stage. After that, we judge the legitimacy of a call by using the social trust-based approach [9]. In the following, we first explain why BC mitigates the issue that we pointed out in Section IV and then describe our algorithm.

A. Introduction of BC

In graph theory, BC indicates a user's centrality in the social network [17], [18]. Formally, BC is defined as the ratio of the number of shortest paths from all users to all others that pass through that user. Let σ_{st} denote the number of shortest paths from $s \in U$ to $t \in U$, where U denotes a set of users in the entire network. Let $\sigma_{st}(u)$ denote the number of shortest

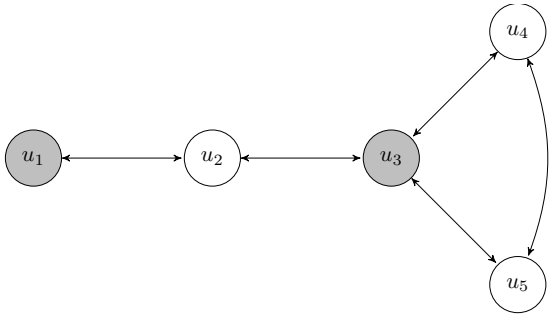


Fig. 1: Toy example of SNS that consists of five users.

paths from $s \in U$ to $t \in U$ that pass through $u \in U$. By using σ_{st} and $\sigma_{st}(u)$, $C_B(u)$, which is the BC of a user u, can be represented as Eq. (5).

$$C_B(u) = \sum_{s \neq u \neq t \in U} \frac{\sigma_{st}(u)}{\sigma_{st}}. \tag{5}$$

The above description is rigid and difficult to understand why Eq. (5) indicates the centrality. Hence we calculate BC with an SNS that consists of five users. Fig. 1 shows an example of such SNS. In this example, we calculate two users' BC, which are u_1 and u_3 and they are represented as

$$\begin{aligned} C_B(u_1) &= \frac{\sigma_{2,3}(u_1)}{\sigma_{2,3}} + \frac{\sigma_{2,4}(u_1)}{\sigma_{2,4}} + \frac{\sigma_{2,5}(u_1)}{\sigma_{2,5}} + \frac{\sigma_{3,4}(u_1)}{\sigma_{3,4}} \\ &\quad + \frac{\sigma_{3,5}(u_1)}{\sigma_{3,5}} + \frac{\sigma_{4,5}(u_1)}{\sigma_{4,5}} \\ &= \frac{0}{1} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1} = 0 \end{aligned}$$

and

$$\begin{aligned} C_B(u_3) &= \frac{\sigma_{1,2}(u_3)}{\sigma_{1,2}} + \frac{\sigma_{1,4}(u_3)}{\sigma_{1,4}} + \frac{\sigma_{1,5}(u_3)}{\sigma_{1,5}} + \frac{\sigma_{2,4}(u_3)}{\sigma_{2,4}} \\ &\quad + \frac{\sigma_{2,5}(u_3)}{\sigma_{2,5}} + \frac{\sigma_{4,5}(u_3)}{\sigma_{4,5}} \\ &= \frac{0}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{0}{1} = 4. \end{aligned}$$

Therefore, $C_B(u_3) > C_B(u_1)$ and thus the user u_3 is located more central than the user u_1. This matches the fact that the user u_1 is located at the edge of the entire network while the user u_3 is located in the center of the network in Fig. 1.

We argue that BC for spammers does not increase. Since spammers call towards users while they seldom receive calls, spammers tend to be 'isolated' at the edge of the entire network and hardly goes through the shortest passes between users. Hence the numerator of Eq. (5) for spammers does not increase well. On the other hand, BC for legitimate users gradually increases with time even if users seldom call. This is because legitimate users gradually make connection with legitimate users and the number of passes that go through legitimate users increases. Therefore, the numerator in Eq. (5) for legitimate users gradually increases.

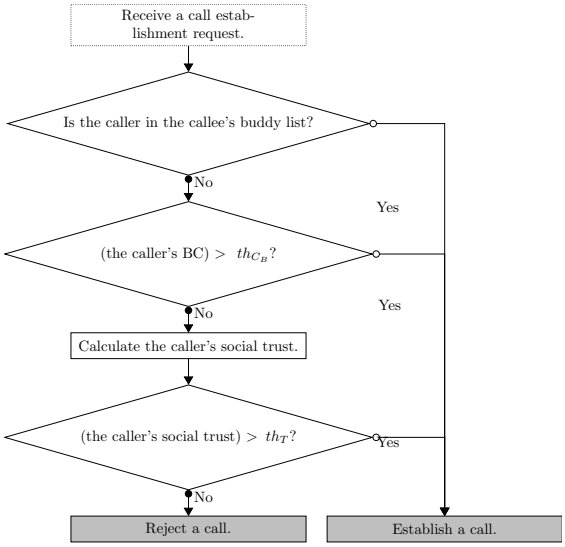


Fig. 2: Flowchart of our SPIT detection scheme.

B. Algorithm

Fig. 2 shows the algorithm of our SPIT detection scheme. When receiving a call establish request from a caller u, the server first checks whether the caller u is in the callee's buddy list. If the caller is in callee's buddy list, they are assumed to be friends and thus the system establishes a call request. Otherwise, the system proceeds to the first detection stage.

If the caller u is not in the callee's buddy list, the system checks whether his/her BC $C_B(u)$ is bigger than a pre-defined threshold th_{C_B}. In order to save the time to calculate $C_B(u)$, the system regularly (e.g., daily or weekly) calculates $C_B(u)$ in the off-line. If $C_B(u) > th_{C_B}$, the system judges that the call is legitimate and establishes the call. Otherwise, the system rejects the call establishment request.

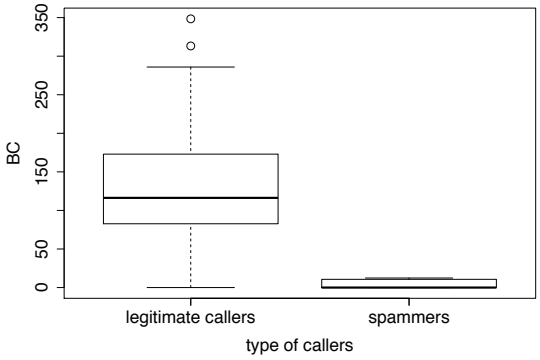
When the user's BC is less than a threshold, the system calculates the inferred trust value $T_{u \to v}$ by using Eqs. (1)-(4) and checks whether the caller's inferred social trust from the callee v $T_{u \to v}$ is bigger than a pre-defined threshold th_T. In order to shorten the time to calculate $T_{u \to v}$, the system may calculate some inferred trusts between users in advance, which is referred to the *landmark-based* method [9], [19]. If $T_{u \to v} > th_T$, the system judges that the call is legitimate and establishes a call. Otherwise, the system rejects the call establishment request.

VI. Simulation Results

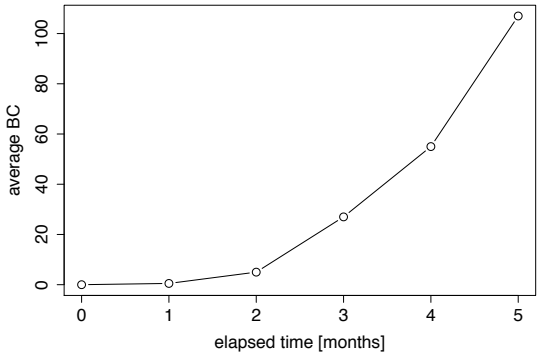
In order to show the effectiveness of our scheme, we first show the characteristics of BC. We then evaluate the detection accuracy of our scheme and the conventional scheme by the computer simulation. We finally compare the required time for the SPIT detection phase for both our scheme and the conventional one. We use R language version 3.1.2 and igraph package [20] for the implementation. TABLE I shows the simulation parameters and they are basically the same as those

TABLE I: Parameter values used in the simulation.

parameter	value
number of users (N_{user})	1,000
ratio of spammers to the entire user (r_{spammer})	1%, 5%, and 10%
distribution of call duration for legitimate users	$N(\mu = 204 \text{ sec}, \sigma = 1)$
distribution of call duration for spammers	$N(\mu = 10 \text{ sec}, \sigma = 1)$
ratio of paid contents subscriber	1 %
th_T	0.25
th_{C_B}	50
α in Eq. (3)	0.2
initial trust value of friends in a buddy list	0.5
initial trust for newcomers	0.4
graph model	random graph with the Erdos-Renyi model [20]
clustering coefficient	0.1



(a) BC versus the type of callers.



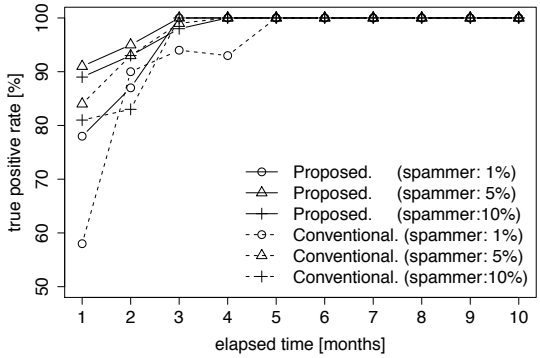
(b) BC versus elapsed times for newcommers.

Fig. 3: Characteristics of BC.

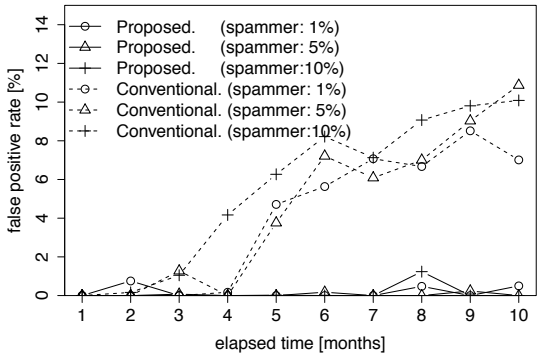
of conventional scheme [9]. If not specified otherwise, we use values specified in TABLE I for the evaluation.

A. BC

We evaluate two characteristics of BC. Fig. 3(a) shows the boxplot of an average BC values against legitimate callers and



(a) True positive rate versus elapsed time.



(b) False positive rate versus elapsed time.

Fig. 4: Detection accuracy.

spammers when five months passed. The boxplot indicates the maximum, the first quartile, the median, the third quartile, the minimum, and some outlier points from the top and thus one can simply grasp the distribution of observed data. As we can see from Fig. 3(a), BC for spammers is concentrated around 0 while that for legitimate callers mostly ranges from 50 to 200. Since spammers call towards users while they seldom receive calls, spammers tend to be 'isolated' at the edge of the entire network. Note that legitimate callers include newcomers in this evaluation and thus BC for newcomers is concentrated around 0. However this is not a problem since the trust value for newcomers is set high enough to avoid a call from being rejected.

Fig. 3(b) shows the average BC for legitimate users with time. As we can see from Fig. 3(b), the BC for legitimate users gradually increases with time. This is because even if a user less frequently calls, he/she can increase BC once he/she connects with multiple friends.

B. Detection Accuracy

Fig. 4(a) and Fig. 4(b) show the true positive rate and false positive rate versus elapsed time, where 'Proposed.' and 'Conventional.' indicate the proposed scheme and the conventional one, respectively. Moreover, 'spammer: 1%', 'spammer: 5%', and 'spammer: 10%' indicate $r_{\text{spammer}} = 0.01, 0.05,$ and, 0.10, respectively. As we can see from Fig. 4(a) and Fig. 4(b), our scheme improves both the true positive rate and false positive rate especially for the false positive rate. We first discuss the true positive rate. From Fig. 4(a), we confirm that our scheme does not degrade the true positive rate.

We then discuss the false positive rate. In Fig. 4(b), the false positive rate against the conventional scheme is getting worse with time. On the other hand, the false positive rate against our scheme is within 2% and does not degrade with time. Therefore we can say that the false positive rate can be remedied by using BC for one of the detection stage. The false positive rate seems to be irrespective of the ratio of spammers. This is because the scheme judges whether each 'call request' (not 'caller') is legitimate or not. The only difference between our scheme and the conventional one is whether to use BC as first detection phase and thus we can say that the BC is effective for judging the legitimacy of call requests from low-frequent callers.

C. Calculation Time

Our scheme tries to filter in a call establishment request from low-frequent callers by using a BC which indicates the centrality of the users in the network. Therefore an extra process is required compared to the conventional scheme. However, we argue that the extra time required for our scheme is negligible. As we mentioned above, a BC can be fully calculated in the off-line manner since it does not involve any callee's information. Note that callee's information (trust value) is required to calculate an inferred trust $T_{u \to v}$. The computation complexity to calculate a BC is $O(N_{\text{user}} N_{\text{link}})$ where N_{link} denotes the number of links by using a faster calculation algorithm proposed by Brandes [18]. We measure the average required time to calculate entire process for our scheme and the conventional scheme with an off-the-shelf computer. The average calculation times for our scheme and the conventional scheme are 54 ms and 51 ms, respectively. From this result, we can say that the time required for the detection stage with BC is negligible.

VII. Conclusion

We have pointed out that calls from low-frequent users are gradually identified as SPIT in the conventional scheme. To remedy this issue, we have proposed a two-stage SPIT detection scheme with BC and social trust. Since the BC can be computed in the off-line manner, our scheme does not cause delay for the call establishment. By the computer simulation, it is shown that our scheme achieves low false positive rate ($< 2\%$) while maintaining high true positive rate. We also show that our scheme takes only 54 ms to judge the legitimacy of a call and thus it is practical.

Acknowledgment

This work is partly supported by the Grant in Aid for Scientific Research (No.26420369) from Ministry of Education, Sport, Science and Technology, Japan.

References

[1] A. D. Keromytis, "A comprehensive survey of voice over IP security research," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 514–537, 2012.

[2] V. A. Balasubramaniyan, A. Mustaque, and P. Haesun, "CallRank: Combating SPIT using call duration, social networks and global reputation," in *Conference on Email and Anti-Spam (CEAS)*, 2007.

[3] J. Seedorf, N. D'Heureuse, S. Niccolini, and M. Cornolti, "Detecting Trustworthy Real-Time Communications Using a Web-of-Trust," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp. 1–8.

[4] R. Zhang and A. Gurtov, "Collaborative Reputation-based Voice Spam Filtering," in *International Workshop on Database and Expert Systems Application (DEXA)*, 2009, pp. 33–37.

[5] T. Kusumoto, E. Y. Chen, and M. Itoh, "Using Call Patterns to Detect Unwanted Communication Callers," *IEEE/IPSJ International Symposium on Applications and the Internet (SAINT)*, pp. 64–70, Jul. 2009.

[6] M. A. Azad and R. Morla, "Mitigating SPIT with Social Strength," in *IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2012, pp. 1393–1398.

[7] ——, "Caller-REP: Detecting unwanted calls with caller social strength," *Computers & Security*, vol. 39, Part B, pp. 219–236, 2013.

[8] N. Chaisamran, T. Okuda, G. Blanc, and S. Yamaguchi, "Trust-Based VoIP Spam Detection Based on Call Duration and Human Relationships," in *IEEE/IPSJ International Symposium on Applications and the Internet (SAINT)*, 2011, pp. 451–456.

[9] N. Chaisamran, T. Okuda, and S. Yamaguchi, "Trust-based VoIP Spam Detection based on Calling Behaviors and Human Relationships," *Journal of Information Processing*, vol. 21, no. 2, pp. 188–197, Apr. 2013.

[10] N. Chaisamran, T. Okuda, Y. Kadobayashi, and S. Yamaguchi, "Trust-based SPIT detection by using call duration and social reliability," in *Asia-Pacific Conference on Communications (APCC)*, 2013, pp. 98–103.

[11] D. Shin, J. Ahn, and C. Shim, "Progressive multi gray-leveling: A voice spam protection algorithm," *IEEE Network*, vol. 20, no. 5, pp. 18–24, Sept.-Oct. 2006.

[12] H. Sengar, X. Wang, and A. Nichols, "Call Behavioral Analysis to Thwart SPIT Attacks on VoIP Networks," in *Security and Privacy in Communication Networks*, ser. LNCS, Social Informatics and Telecommunications Engineering, vol. 96. Springer Berlin Heidelberg, 2012, pp. 501–510.

[13] Y. Bai, X. Su, and B. Bhargava, "Adaptive Voice Spam Control with User Behavior Analysis," in *IEEE International Conference on High Performance Computing and Communications (HPCC)*, 2009, pp. 354–361.

[14] F. Wang, M. Feng, and K. Yan, "Voice Spam Detecting Technique Based on User Behavior Pattern Model," in *IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, 2012, pp. 1–5.

[15] K. Toyoda and I. Sasase, "SPIT callers detection with unsupervised Random Forests classifier," in *IEEE International Conference on Communications (ICC)*, Budapest, Jun. 2013, pp. 2068–2072.

[16] ——, "Unsupervised Clustering-based SPITters Detection Scheme," *Journal of Information Processing*, vol. 23, no. 1, pp. 81–92, 2015.

[17] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[18] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[19] M. Potamias, F. Bonchi, C. Castillo, and A. Gionis, "Fast shortest path distance estimation in large networks," in *ACM Conference on Information and Knowledge Management (CIKM)*, 2009, pp. 867–876.

[20] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006. [Online]. Available: http://igraph.org