Future Generation Computer Systems 95 (2019) 841–854

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Rapid detection of spammers through collaborative information sharing across multiple service providers

Muhammad Ajmal Azad [a,*,1], Ricardo Morla [b]

[a] School of Computing Science, Newcastle University, United Kingdom
[b] Faculty of Engineering, University of Porto, Portugal

HIGHLIGHTS

- Spammers/telemarketers target a very large number of recipients usually dispersed across many Service Providers.
- Collaboration among service providers would increase the detection accuracy but has the challenge of privacy and system resources.
- This paper proposes a collaborative system for the early detection of spammers in a network.

ARTICLE INFO

Article history:
Received 24 April 2017
Received in revised form 27 November 2017
Accepted 22 December 2017
Available online 16 January 2018

Keywords:
SPIT
Collaboration
VoIP
Network operations
Reputation

ABSTRACT

Spammers and telemarketers target a very large number of recipients usually dispersed across many Service Providers (SPs). Collaboration and Information sharing between SPs would increase the detection accuracy but detection effectiveness depends on the amount of information shared between SPs. Having service provider's exchange call detail records would arguably attain the best detection accuracy but would require significant network resources. Moreover, SPs are likely to feel uncomfortable in sharing their call records because call records contain user's private information as well as operational details of their networks. The challenge towards the design of collaborative Spam over Internet Telephony (SPIT) detection system is two-fold: it should attain high detection accuracy with a small false positive, and should fully protect the privacy of users and their service providers. In this paper, we propose a COllaborative Spit Detection System (COSDS)—a collaborative SPIT detection system for the Voice over IP (VoIP) network where service providers collaborate for the effective and early detection of SPIT callers without raising privacy concerns. To this extent, COSDS relies on a trusted Centralized Repository (CR) and exchange of non-sensitive reputation scores. The CR computes global reputation of users by aggregating the reputation scores provided by the respective collaborating SPs. The data exchanged to the CR is not sensitive regarding users privacy, and cannot be used to infer the relationship network of users. We evaluate the performance of our system using synthetic data that we have generated by simulating the realistic social behavior of spammers and non-spammers in a network. The results show that the COSDS approach has better detection accuracy as compared to the traditional stand-alone detection systems. For instances, in a setup where spammers are making calls to recipients of many SPs, COSDS successfully identifies spammers with the True Positive (TP) rate of around 80% and false positive (FP) rate of around 2% on a first day, which further increases to 100% TP rate and zero FP rate in three days. COSDS approach is fast, requires a small communication overhead, ensures privacy of users and collaborating SP, and requires only few iterations for the reputation convergence within the SP.

1. Introduction

For many years, E-mail and other online networks (social networks, websites, blogs) have been widely used by scammers to target users with the unwanted content (advertisements, malware etc.). The trend has been changed from the last few years, because of cheap telephony and its larger customers base, telephone call has become the preferred method used by fraudsters to distribute

* Corresponding author.
E-mail addresses: muhammad.azad@ncl.ac.uk (M.A. Azad), ricardo.morla@fe.up.pt (R. Morla).

[1] Work was done when the author was at INSEC TEC and University of Porto, Portugal.

https://doi.org/10.1016/j.future.2017.12.026
0167-739X/

the unsolicited and advertisement content [1]. Unwanted communication in telephony is more annoying than the traditional email or text spamming as it requires an immediate response from the call recipient. Moreover, unwanted pre-recorded messages unnecessarily overwhelm the voice mail-box with the lengthy unwanted speech stream, that later requires a considerable amount of time to clean it [2]. Apart from annoyance, these calls can cause financial loss to the telephony users and the service providers [3].

Telecommunications or VoIP service providers (SPs) deploy standalone SPIT detection systems [4–7] in their networks for protecting their subscribers from unsolicited calls. These systems decide about the behavior of the user by considering the metadata from the single source. Stealthy or low rate spammers can easily evade these standalone systems by simply making a low rate spam calls to recipients of several service providers without overwhelming any single service provider with the spam calls. By doing so, spammers remain undetected for a longer period, since service provider does not have enough evidence to characterize these callers as spammers because of small calling rate. Standalone systems (SASs) may identify stealthy spammers over the time, but this detection is very late, as spammers have already called a large number of users across several SPs. Traditional standalone approaches could improve the detection rate by simply asking callers to solve the CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) [8,4] test for every call they originate, but CAPTCHA test requires extensive system resources for the real-time authorization, and is also intrusive to the caller.

The collaboration among multiple service providers would increase the chances of detecting low rate spammers timely and effectively, however, its detection performance depends on the type of information exchanged between collaborators. There are two key challenges in the design of collaborative SPIT detection system: firstly, what information should be exchanged as the course of the collaboration process; and secondly, to whom this information should be made available. The collaborative solution can be either the distributed—where the information from each SP is shared and processed in a completely distributed fashion and the centralized—where all information from the SP is reported to the single centralized location for analysis. The service providers are not willing to share sensitive data of their customers directly with each other, because they are business competitors, and are concerned about the privacy of their customers and own network configurations. Generally, a better detection accuracy is expected when collaboration is achieved through the exchange of complete call records, but at the cost of privacy and system resources. On the other hand, SPs feel more comfortable in exchanging information that represents the aggregate behavior of users in their network and does not leak any information that could be used to infer the social relationship network of users. The exchange of summarized information could protect the privacy of users but at the cost of detection accuracy and privacy. The challenge is designing a system that minimizes the detection time, show improved detection rate, and does not pose any threat to the privacy of collaborators and users.

In this paper, we propose a Collaborative Spit Detection System (COSDS) for an accurate and early detection of SPIT caller, that leverages collaboration among many autonomous SPs. COSDS system does not require direct collaboration among SPs, instead, the collaboration is carried out with the exchange of non-sensitive summarized information with the trusted CR. The design achieves two objectives: ensuring the privacy of users, and reducing the network load required for the collaboration. Particularly, the collaborating SP computes and submits the Local Reputation (*LR*) scores (summarized information) of users/customers to the trusted CR. This reputation score represents the aggregate behavior of user within the SP, and have been computed from the user's past call transactions with others. The CR is responsible for computing the global reputation of the user by aggregating the *LR* scores, and the computation of the classification threshold below which user is flagged as the global spammer. The CR responds collaborating SP with the global reputation (*GR*) score and classification result. The SP either choose CR recommendations or act independently against spammers by making his own decision using global reputation score along with other social network features. Each collaborating SP interacts directly with the CR and requires only two transmission cycles for getting the *GR* of his users i.e. one cycle for sending the *LR* to the CR, and one cycle for receiving the *GR* from the *CR*. From the perspective of deployment in a real scenario, the COSDS approach does not require changes in the call setup messages, and would easily convince SP to take part in a collaboration.

We evaluate our system using the synthetic data that have been generated through models of spammers and non-spammers social behavior. The evaluation results demonstrate that the COSDS system outperforms standalone detection systems in terms of detection accuracy and detection time. Specifically, for a network having a large number of spammers, COSDS managed to achieve a zero FP rate and blocked all spammers within 3 days.

The proposed approach is an extension of the SP level SPIT detection system presented in [5]. In this paper, we establish cooperation among SPs and focus on defining the components and mechanism for the collaborative SPIT detection. This enables early and accurate detection of the spammer while considering the local reputation scores of the caller in many collaborating SPs.

In a summary, this paper makes the following contributions:

- It presents the design of COSDS, a system for the collaborative SPIT detection that enables SPs to part in the collaboration with the exchange of non-sensitive summarized information. COSDS is typically more efficient in detection accuracy than the standalone detection systems, and more efficient in system resources than collaboration with the exchange of call records. Further, it ensures privacy of users, yet achieving the high true positive rate and the small false positive rate.
- It gives a detailed implementation and evaluation of the approach over the synthetic call detailed record (CDR). Particularly, the evaluation is performed for the different number of collaborators, different percentage of spammers, and for the following metrics: true positive rate (TPR), false positive rate (FPR) and accuracy. It also compares the performance of COSDS to a system where collaboration is carried out through the exchange of non-summarized information.

The rest of the paper is structured as follows. Section 2 presents the background on voice spamming and its difference from the email spamming. Section 3 reviews the works from the other researchers and provides motivation for our approach. In Section 4, we describe the architecture of the collaborative SPIT detection system, describe how the privacy of the user and collaborating SP is protected in Section 5. The experimental setup is presented in Section 6. Section 7 presents the performance evaluation for different performance metrics is presented. Finally, we conclude the paper in Section 8.

2. VoIP spam

The advent of VoIP technology enables telecommunication service providers to converge their legacy circuit-switched networks into a single all IP-based network to minimize their operational expenses. The advent of VoIP network has also benefited users to

enjoy the low rate telephone calls domestically and internationally. The number of telephony subscriber (VoIP, mobile and Fixed) are more than 6 billion with trillions of calls per year worldwide. The scammer and telemarketer also find telephony an attractive and less costly medium to target a large number of subscribers with the unwanted content. Currently, complaints made to Fraud.org and FTC indicates that the telephone was the initial method of contact for most of the scam [1,9]. The number of complaints about unwanted calls (telemarketing, scams, robo) received to the FTC has increased from 3.6 million during the year 2015 to over 5.3 million during the year 2016. The average monthly complaints have also increased to more than 250 thousand in 2016, an increase of almost four times as compared to previous years [3]. The unwanted calls not only disturb users, but they also bring financial loss to recipients and service providers [10,11]. The estimated annual telephone fraud loss is around $40.1 Billion (USD) [9] around the globe, of which subscribers directly lose more than $8.6 billion annually in the United States alone.

SPIT is similar to the email spamming, however, it causes a serious discomfort to the recipients due to its more interactive and intrusive nature. A fundamental difference between the voice and the email spamming is that emails can be held and processed relatively for a long time before being sent to users, whereas the voice call requires an immediate response from the operator, and has to be processed in a real-time before alerting call recipients. Another important difference is that in an email, content is available inside the text body prior to its delivery, but in a voice, content is only available after establishing the connection between the caller and the caller. Further, the content of voice call is speech stream that is difficult to be processed in a real-time as compared to processing the text in case of emails. From the user's perspective, classifying a SPIT content in a voice mail-box consumes more time than the email spam. The email user categorizes email in an inbox as a spam or non-spam on a first look by checking the header and subject information, however, in case of a SPIT-recorded call, the receiver has to listen few seconds of recording before categorizing it as an unwanted content [2].

3. Related work and motivation

This section summarizes related work in the area of SPIT detection. Specifically, we discuss limitations of standalone systems, examine how existing collaborative system works, and describe the motivation of this work.

3.1. Standalone anti-SPIT systems

Several solutions have been proposed for blocking the SPIT caller. These solutions operate independently as the standalone systems, and can be grouped into several classes: black-list or white-list based systems [12–14], systems analyzing the social behavior and reputation of the caller [15–24], authenticating the caller by challenging him in the form of CAPTCHA and Turing test [25–27,8], imposing extra cost on the caller if he is flagged as unwanted [28] by recipients of call, systems processing speech content [29–32], analyzing the linguistics from the speech streams [33,34], and statistical systems that analyzes the flow of packets during the call setup phase [35,36] or analyze caller's behavior from the logged CDRs [37,38]. A single standalone SPIT solution can also be employed by combining many individual systems in the form of a collaborative multistage system [28,39–42]. Recently, new detection systems have been proposed to fight against the unwanted callers. These systems include (1) deploying mobile and fixed telephony Honeynets [43–46] for collecting the call records, to be used for detecting and characterizing the behavior of unwanted callers, (2) the chatbot system [47], which connects back to the caller with an automated phone bot, and (3) the identity linking based system [48] that first connects identities that belong to one physical caller and then performs aggregate detection.

3.2. Limitations of standalone detection systems

Stand-alone SPIT detection systems are currently major systems for thwarting SPIT callers. These systems are typically placed within SP, and utilize meta-data from one source for deciding about the behavior of the caller. Since there is no cooperation among SP, no data about the caller is pass between SPs except the call handling messages. These systems prolong detection when spammers make a low rate spam calls to recipients of several SPs. The standalone systems, in this attack, would not have enough evidence to block the spammers in a timely manner. Particularly, the stand-alone system performs well when the number of calls from the same caller spikes. However, this is problematic because the spammer has already reached to a large number of users. The SAS improves their detection capability by combining several standalone detection approaches as a single system or asks caller to solve the CAPTCHA challenge. However, these implementations have following limitations. First, it involves caller for solving the CAPTCHA challenge and is also resource intensive. Second, it requires that call request should be pass through many detection components thus would increase the call setup delay. Third, it requires relatively a large number of calls from the caller for making the final recommendation, which still allows spammers to reach several subscribers.

3.3. Collaborative detection systems

Several collaborative detection systems have been proposed for detecting the spammers and intruders over the Internet by incorporating collaboration between domains and Internet service providers [49]. These systems normally collaborate by sharing the message content, reputation score or IP headers, thus raise privacy concerns. The privacy of the user can be protected by exchanging the hash of the message [50–52], and by using a trusted centralized system [53,54]. A content-based collaborative systems are not feasible in perspective of speech content because of resource-consuming processing and hashing of speech stream in a real-time. Collaboration can be achieved through the exchange of single reputation scores of between users or between users and the CR [53,55,56].

Very few works have been proposed for the collaborative spam detection in a telecommunication and VoIP network. The existing collaborative approaches are mainly based on the process of internal collaboration in the form of multistage systems [28,39,41,57]. In [39] authors present a collaborative multistage solution that integrates feedback from multiple modules to decide about the behavior of the caller. Systems like [28,41] require internal collaboration among the black-list module, reputation module, statistical and the user feedback module for the classification of the caller. The bulk information about caller's behavior resides in his home network. In [58] authors propose a collaborative system that allows receiving SP to access the quality of SPIT algorithm used by the home SP of the caller, but this approach does not rate the caller. SPACEDIVE [59] performs collaborative intrusion detection within the VoIP network domains by correlating the local and remote rules at the individual and across different components. A distributed cooperative detection method has been proposed in [60] for identifying SPIT callers through collaboration between several VoIP servers. The proposed approach did not discuss how feedback from the collaborators is aggregated for the final status of the caller. SDRS [57] provides different reaction mechanisms against SPIT callers by having collaboration between various standalone SPIT detection systems.

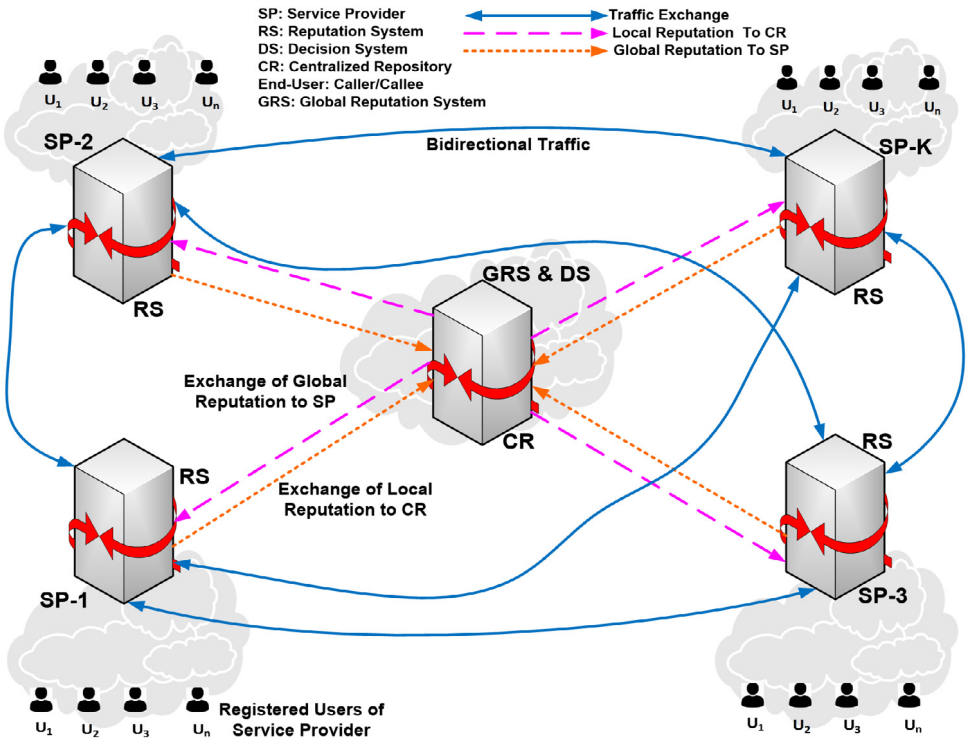


Fig. 1. Building block of collaborative SPIT detection system.

3.4. Motivation

As discussed, stealthy spammers can easily evade the system by making the low rate spam calls. However these spammers need to reach a large number of recipients for the greater financial benefit, thus distribute their calls to recipients of many SPs simultaneously or turn by turn. Therefore, observing call patterns of the caller across many SPs in a collaborative way would minimize the effect of stealthy spammers. The collaborative process brings the challenge of privacy protection, and the trade-off between the detection accuracy and the information shared during the process of collaboration. As a first step towards the collaborative detection, raw data or the social graph is exchanged among the collaborators, but privacy concerns and trust issues limit SPs to take part in this collaboration process, respectively. In order to have a collaborative view of the caller, we need to have a system that computes reputation of the caller without compromising the privacy of collaborating SPs and their customers, yet achieving a high detection accuracy with small communication overheads.

4. Collaborative SPIT detection system: the design

In this section, we present the design of COSDS system that computes global reputation of the user by aggregating the scores provided by the collaborating SP. It performs all operations without requiring high communication overheads and without posing any threat to the privacy of collaborators and their users. We describe the system architecture of COSDS system in Section 4.1 and present a procedure to calculate the aggregated reputation in Section 4.5. We introduce the procedure for calculating the classification threshold in Section 4.6. Later in Section 4.8 we describe the design choices for the collaboration process.

4.1. System components

COSDS system consists of three main components as illustrated in a Fig. 1: users, the collaborating SPs or VoIP operators, and the CR. The user is the caller or the callee involved in the bidirectional communication using services from the SP. The SP handles user's in-coming and out-going call request using signaling protocols (SIP, H323, SS7 etc.), and records call transactions in a call detail record (CDR) database. The SP has a local reputation system for computing the reputation of callers using information from the call records or explicitly asking callee for the feedback about the caller recent called him. The core component of COSDS is a trusted CR system that is responsible for computing the global reputation of the caller by aggregating the local scores provided by the cooperating SPs. The CR is also responsible for computing the classification threshold below which the caller is classified as a spammer, and finally, update collaborating SPs with the global reputation score and the classification result. The design choice of trusted CR and use of non-sensitive aggregated scores ensures privacy protection without consuming extensive resources.

4.2. Assumptions

We consider following call behavior assumptions in the design of COSDS system: (1) people calling behavior can change over the time (they add or remove links, have different calling behavior with family and friends etc.) [61]; (2) the calling behavior of the legitimate caller is different from the calling behavior of spammer, (e.g. frequency and duration of interactions, number of unique callees etc.) [16,37,62]; (3) the detection approach based on the collaboration between SPs is likely to have a better detection accuracy, and a small detection time than that of standalone systems [63,64].

Further, we also assume the following. We assume that a CallerREP reputation management system has been deployed in a SP for computing the local reputation of callers, and SP is agreed on exchanging the local reputation scores to the trusted CR. Further, we assume that spammers distribute their calls to recipients of different SPs without overwhelming any single SP. For the anonymized callers, the SP also agreed to provide the call-id of the caller. We also assume that collaborating SP requires global reputation scores and status of only those callers registered in other SPs.

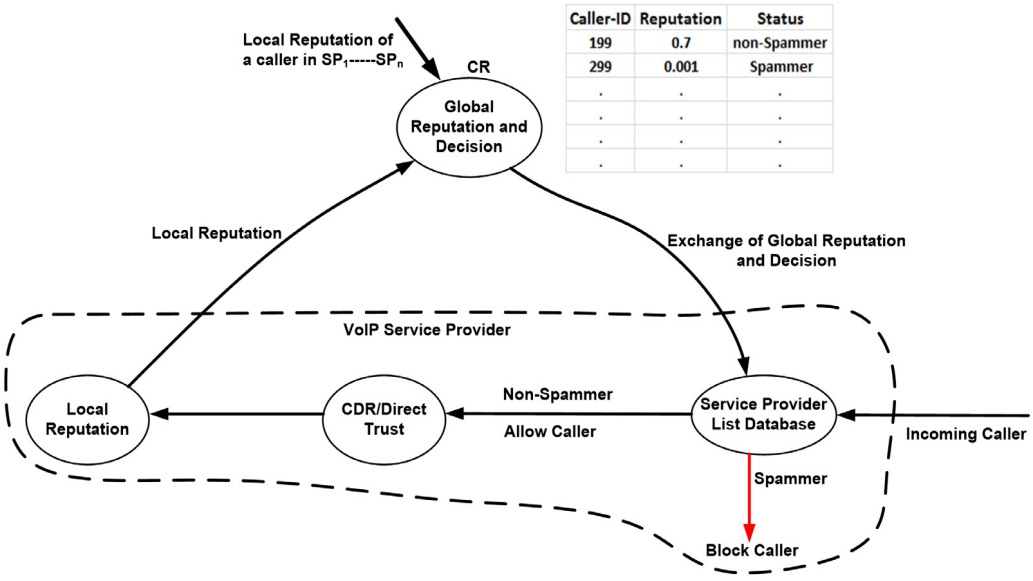


Fig. 2. SP's level working of collaborative SPIT detection system.

4.3. Exchanging reports

The SP exchanges scores to the CR in the following format: [Caller ID, Local Reputation Score, Trust for the SP]. The first argument is the unique identity of a caller (i.e. IP address or telephone number), the second argument is the local reputation score of the caller, the third argument is optional and represents trust value of SP on others. We are using the telephone number as the calling identity. Each exchange record is represented as the row and value of reputation score is normalized between 0 and 1. The CR responds back SP with the global reputation score in the following format: [Caller ID, Global Reputation Score, Decision]. The global reputation is the aggregated reputation of the caller, and the decision is the classification result (spammer or non-spammer).

4.4. Work-flow of collaborative system

Fig. 2 illustrates SP's reactions for the call request it receives from the caller. Upon receiving the call request, the SP first check the trustworthiness of caller using its own local reputation table. If a caller is blacklisted then the SP immediately block the caller, and if the caller is not blacklisted then the SP allows the caller to pass through the network. After the conversation ends, the SP updates the local reputation of a caller, sends this score to the CR for the updated global view in next aggregation cycle. The CR computes the global reputation of the caller for the aggregation cycle considering new scores and responds SPs with the global reputation score and the classification result. The SP then update its database using classification recommendation from CR or uses global reputation scores in combination with local social and call features of a caller.

4.5. Global reputation of a caller

The global reputation of a caller is computed in two steps. First, a SP computes local reputation of the caller and exchanges it to the CR, and secondly, a CR computes global reputation by performing a weighted aggregating on the *LR* score received from several collaborating SPs.

In a service provider, the *LR* of the caller is computed in two steps. First, a direct trust between a caller and the callee is computed from the caller's past call transactions with the callee, secondly, a local reputation of the caller is computed using modified

Algorithm 1 Global Reputation of Caller S At CR

procedure AGGREGATING REPUTATION OF CALLER S ()
  INPUT: Trust Matrix $Trust_{SR}$ of Caller S With Callee R within the SP Using Eq. (1).
  OUTPUT: Global Reputation GR_S of the Caller$S$.
  for for each SPs **do**
    for for All User within the SP **do**
      Initially $GR_S = 1/PO_S$
      Iterate until convergence
      while $\delta < \epsilon$ **do**
        $LR_S \leftarrow Trust_{SR} \times GR_S$
        $GR_S \leftarrow LR/\|LR\|$
        $gr \leftarrow \|LR\|$
        $\delta \leftarrow \left|\frac{gr - gr_{previous}}{gr}\right|$
        $gr_{previous} \leftarrow gr$
      end while
    end for
  end for
  Send the Local Reputation Scores [$CallerID, LR_S, Trust\ for\ SP$] to the CR.
  for *All caller S* **do**
    $GR_S = \frac{\sum_{SP=1}^{N} W_{SP} \times LR_S^{SP}}{N}$
  end for
  Exchange of Global Reputation GR_S of a caller to each Collaborating SP.
end procedure

Eigen Trust algorithm [5]. The direct trust can be computed either in an intrusive way—asking callee for the explicit feedback [15,40] of caller behavior with him, or computed in a non-intrusive way—using the average call duration information from the CDR [4,37]. The former approaches require feedback from callee and changes in the handset, whereas the later approaches would allow spammers that had only a few good duration out-going calls out of a large number of called recipients. The collective use of several features in a non-intrusive way would better characterize the trust relationship between caller and the callee. In a combined approach, a direct trust between the caller and the callee is computed by collectively considering three features: the number of in-coming

and out-going calls between the caller and the callee, the call duration of calls made and received between the caller and the callee, and the number of unique callees of the caller [5]. These features have been considered because of the fact that legitimate and spam caller exhibits different calling behavior. The legitimate caller usually has a long duration, bi-directional repetitive calling behavior with his friends and family members has a small duration bidirectional calls with only a very few callees, and a very small number of unique callees. On the other hand, the spammer or the advertiser usually targets a large number of callees, that normally results in a short duration calls to a large number of callees. This unbalanced calling behavior would result in a small direct trust score for the spammer, a high trust score for the legitimate caller. In a service provider SP, the direct trust $Trust_{SR}^{SP}$ between a caller S and the callee R is computed by using Eq. (1).

$$Trust_{SR}^{SP} = \frac{CD_{SR}^{SP} \times Call\,Rate_{SR}^{SP} + CD_{RS}^{SP} \times Call\,Rate_{RS}^{SP}}{PO_{S}^{SP}} \tag{1}$$

In Eq. (1), CD is the call duration between a caller and the callee in a specific time interval, $Call - Rate$ is a calls frequency made between the caller and the callee in a specific time interval, and PO is the out-degree of the caller. The trust of all callers is represented as the sparse trust matrix of dimensions $N \times N$, where N is the total number of users within the SP. If there is no interaction between the caller and the callee then $Trust_{SR}^{SP}$ is set to be zero. The direct trust between the caller and the callee is asymmetric as the caller and the callee might have different number of callees.

Finally, the SP applies modified Eigen Trust algorithm to the direct trust matrix to compute the local reputation score. The direct trust are normalized as following: $Trust_{SR}^{SP} = Trust_{SR}^{SP} / \sum Trust_{S}^{SP}$. The caller's reputation in the SP is represented as LR_{SR}^{SP}, and is computed iteratively by multiplying trust scores with the local reputation scores ($LR_{S}^{SP} = Trust_{SR}^{SP} \times GR_S$). Where GR_S represents the global reputation of the caller after the collaboration. Initially, the GR_S of a caller is set equal to $1/PO_{S}^{SP}$ [5]. The computation of local reputation of a caller is an iterative process, and is continued until the average relative error between δ is less than ϵ as shown in algorithm 1(lines 8–16). On each aggregation cycle, the SP sends local reputation LR_{S}^{SP} of a caller to the CR, and receives the aggregated global reputation GR_S of the caller from the CR.

The trusted CR computes global reputation of a caller using weighted average algorithm as represented in a Eq. (2).

$$GR_S = \frac{\sum_{O=1}^{N} W_{SP} \times LR_{S}^{SP}}{N} \tag{2}$$

In Eq. (2), N is the total number of SP participating in the collaboration, W_{SP} is the trust score of SP sending calls to reporting SP, and LR_{S}^{SP} is the local reputation of the caller in a SP.

The weighted average aggregation allows CR to give a different importance to the different collaborating SPs. The weighted aggregation minimizes the effect made by SPs deliberately supporting spammers for an extra financial benefit. The lower the local reputation of the caller in several SPs, the lower would be his global reputation—despite the caller is having the trustworthy reputation in only one SP. In other words, if the majority of SPs assign small reputation value to the caller S then the caller S's global reputation would bend towards the reputation of S in the majority of SPs. The call-receiving SP assigns weights to the other SPs on the basis of the fraction of callers classified as a spammer to the total number of calls coming from the SP.

$$W_{ij} = 1 - \frac{No.of\ \ Callers\ from\ SP_{ji}\ identified\ as\ spammers}{Total\ No.\ of\ \ Unique\ Callers\ from\ SP_{ji}} \tag{3}$$

In Eq. (3),SP_i receives calls from SP_j and W_{ij} is the trust weight of SP_i for the SP_j.

The computation of the global reputation is not resource intensive. In COSDS, convergence is performed locally. The SPs receive global reputation of the caller in two transmission cycles: one transmission cycle for exchanging local reputation to the CR, and one cycle for receiving the global reputation from the CR. The communication overhead of COSDS is much less than the communication overhead required for computing global reputation by having a collaboration in a distributed way [65–67]. COSDS is also independent of the number of pre-trusted users, and the traffic overhead remains constant regardless of the out-degree of the user and number of collaborators.

4.6. Detection of SPIT caller

The CR maintains a vector of global reputation score of each caller, and has a value between 0 and 1. We now describe the procedure for determining the classification threshold value T below which the caller is considered as a spammer. There can be two methods for the classification threshold: (1) a fixed threshold—that is estimated based on TP or FP tolerance policy of SP; and, (2) a dynamic threshold—estimated from the current and past calling behavior of callers. The choice of a fixed threshold is straightforward, but it does not necessarily represent the dynamic calling behavior of callers, and the dynamic threshold requires analysis of the present and past behavior of all callers.

The design of COSDS expects that callers with the small global reputation score are likelier to be considered as spammers, and callers with a high global reputation score are likelier to be the legitimate callers. The spam callers usually have the similar call pattern and almost similar global reputation, small reputation scores, and their reputation is much different from the global reputation scores of the legitimate callers. In addition, the SP may also require blocking the top spamming identities because of revenue. Considering these facts, the COSDS system adopted the percentile-based dynamic threshold method for the classification threshold. The procedure for classifying a caller as spammer or non-spammer is presented in algorithm 2. In algorithm 2, the 25th percentile of GR (global reputation vector) is computed first, and then the mean m of the global reputation score of all callers below the 25th percentile is used as a final threshold T. The caller can be classified as legitimate 1 or non-legitimate -1 based on the following rule:

$$Caller_S = \begin{cases} Spammers & if\ GR_S < \beta \times T \\ non - Spammers & if\ GR_S > \beta \times T \end{cases}$$

The spammer normally increases his number of callees over the time, and it might be possible that he bypass COSDS defense on a first aggregation cycle, but over the time or after a number of iterations he would not be able to by-pass the detection system. The FP rate (non-spammer classified as a spammer) of COSDS system under high spamming rate is almost zero. This would maximize the profit of SP by not blocking the legitimate callers. The 25th percentile threshold would not provide optimum detection when the percentage of spammer exceeds 30 percent. A small adjustment in the threshold could improve the detection performance. In order to maximize the true positive and true negative rates, a SP's parameter β (greater or less than 1) is defined along with the threshold T. Alternatively, classification can also be performed using supervised and unsupervised machine learning techniques. However, the challenges in such classification are the initial labeling of reputation vector and deciding the number of clusters. The problem can be solved by initially labeling some of the callers manually or clustering them during first aggregation cycle, and then using this information for classification and updating of labels. Further, the classification process requires manual input from the SP, and the initial clustering requires the analysis of behavior of callers in spammer and non-spammer clusters, respectively.

Finally, the CR responds SP with a global reputation vector and its classification decision. The SP can either accept the CR recommendation or use the recommendation along with caller local calling behavior. For example, using local features such as the inter-arrival time of the call request, and the call rate along with the global reputation for the final classification. The SP can also ask the callee by sending him the recommendation as the part of call request message, and the callee then decides whether to accept or reject the call.

Algorithm 2 Detecting Spammer and Updating Service provider Trust

```
1: procedure SPIT Caller ()
2:     INPUT: Reputation GR_S and threshold β
3:     SP−defined parameter ← β (β = 1 if SP has no preference)
4:     m = 1st − quartile(GR_S)
5:     T ← mean(GR < m)
6:     for All caller S do
7:         if (GR[S]<β× T) then
8:             Caller S is Spammer
9:         else
10:            Caller S is non-Spammer
11:        end if
12:    end for
13:    Update Weights of SP using Eq. (3)
14: end procedure
```

4.7. Communication overheads

The communication overheads between a collaborator and the CR depends on the amount of information exchanged between them. In COSDS, the SP stores and exchanges *LR* of his users to the CR, that requires only 22 bytes for a one user (14 Bytes for the Caller-Id and 8 Bytes for the Reputation score). The CR computes *GR*, makes the decision, and exchanges the result to the collaborating SP, this requires 23 bytes (14 bytes for the Caller-ID, 8 Bytes for the Global Reputation, and 1 Byte for the decision). The overall communication overhead required for sending scores to CR is $n * 22$ Bytes (where n is the total number of users in a SP) and communication overhead requires for responding the collaborating SP is $k * 23$ Bytes per SP (where k is a total number of users from all SP). The exchange of CDRs and direct trust scores require high communication and memory overheads because of a large amount of data.

4.8. Design options for information summarization and collaboration

The service provider holds different types of call data that can be used in a variety of ways. One such data is the Call Detail Record (CDR) database that records calling history of the caller. It contains a diverse set of information including caller–callee unique identifiers, IP addresses of the caller and the callee, duration and time of a call, and call status (successful, failed, busy), and is mainly used for billing and network management. The CDRs can be used for characterizing the social behavior of users. In this section, we discuss three design options that collaborators can use for the collaboration. The major challenge is to achieve a trade-off between privacy, detection accuracy and communication overhead. The trade-off between accuracy and privacy can be achieved through following design options: (1) design having no privacy protection, (2) design having partial privacy protection, and (3) design having absolute privacy protection.

In a first design option, the SP exchanges the call records containing caller–callee identities, call duration, and call time to the CR. The CR aggregates CDRs from all collaborating SPs, and computes global reputation of the caller using Caller-REP approach. Although the trusted CR guarantees protection of sensitive information provided by SPs but curious CR and intruder at CR would still learn the real identity and relationship network of callers. This design option may provide better detection accuracy because of the availability of complete information about the behavior of the caller, but has privacy leakage, and requires extensive communication overheads. Further, this collaboration process also increases the computation load on the CR because of processing of millions of call records from each collaborating SP.

In a second design option, the SP sends caller–callee direct trust scores to the CR. In this case, the SP computes direct trust between the caller and the callee using Eq. (1), and sends the normalized trust matrix to the CR. The CR then performs three functions: aggregates the direct trust scores of callers from different sources, computes the global reputation of callers, and finally classifies them as spammers or non-spammers using approach [5] and algorithm 2. This design option hides end-user critical personal information such as call-rate and call duration, but it still provides relationship network information and trust score of callers on their callees.

In a third design option, the SP locally computes the reputation of a caller and exchanges these reputation scores to the CR. The CR then computes global reputation of the caller by aggregating local reputation scores and classifies them. The exchange of reputation scores cannot be used to recreate the social and calling network of the caller. This approach not only ensures privacy of users and their service providers but also has small communication overhead and computational load.

Generally, protecting the privacy of the end user has the following perspectives: data privacy where no entity is able to learn the personal information such as the name or age of the user; and secondly, calling network privacy, where no entity can learn the social network and calling behavior of users. In first and second design option, neither data privacy nor network privacy is protected; however, in a third option, both are protected from the breach. In COSDS, the privacy of the caller is well protected and cannot be misused to infer any information about users. Further, the exchange of the single reputation score also convinces SPs to have collaboration.

5. Discussion on the privacy protection

SP needs to protect privacy of his customers in two aspects: (1) Protection of user's pseudonymized Identity: preventing the adversary having some auxiliary (AUX) information to find the anonymized identity of his target; and, (2) Social Relationship Network Protection: the existence and strength of social relationship between the target user and his friends should not be learned by the adversary.

5.1. Adversary model and the privacy breach

We assume that users are not intrusive, and SPs are not misusing recorded CDRs. We consider the honest but curious model, i.e. the CR performs its functionality honestly, but adversary at CR or CR itself tries to infer the relationship network of the user. The adversary has some auxiliary information about the target user, which may include time of few calls, a call duration of few calls, and the call rate. The objective of the adversary is to use this auxiliary information to infer social relationships of the target user in a specific SP. We assume that the communication between CR and collaborating SPs is secure. The Probability that an adversary can

breach the privacy, and gets true records given AUX information is presented as:

$$Pr(PrivacyBreach|AUX) = \begin{cases} 1/X & ; if\ X > 0 \\ 0 & ; if\ X = 0 \end{cases} \quad (4)$$

Where X is a number of users returned for the AUX information.

5.2. Privacy protection at SP

The SP processes CDRs for the computation of the *LR* score. The adversary has following auxiliary information and tries to break privacy of user during the computation of *LR* at SP.

AUX1: An adversary knows call related information of the target user, and wants to find an anonymized identity of the target user. For example, an adversary knows target user called some known person at 11:20 am.

AUX2: An adversary knows out-degree of the target user along with AUX1. For example, an adversary knows call times of calls made by the target user, and a number of friends of the target user.

AUX3: An adversary knows the calling behavior of target user along with AUX1 and AUX2. For example, an adversary knows call rate and call duration of the target user's few calls, and wants to learn the complete relationship network of the target user.

COSDS protects the privacy of the user within the SP by setting the following best practices. (1) The SP shall protect users records from the unauthorized access using strong authentication processes, (2) The SP shall provide an opt-op option to the subscriber if his out-degree is small, and (3) The SP shall pseudonymized identity of the user for further reducing the risk of misuse of the data. Pseudonymized identities can provide one level of protection, but adversary can still find the pseudonymized identity of the target by using single AUX or correlating multiple AUX. We use the following mechanism for the CDR anonymization at the SP:

P1: We strip the minutes and seconds information from the date and call time of the CDR record. By doing this, the probability of inferring the pseudonymized identity is extremely very small for the AUX1.

P2: The out-degree of the users in the CDR is k-anonymized. For users having unique out-degree, the random noisy user can be generated which has the same out-degree but with different pseudo identity. This k-anonymization would affect the detection accuracy but provides privacy protection for the AUX2.

The adversary knows AUX1 of his target subscriber; for example, adversary learns from media that presidents of two countries talk to each other for some duration on some specific time. In this case, the adversary wants to learn pseudo identities associated with both presidents. The adversary can possibly find a small candidate-set if time information in the CDR is not properly anonymized, and by correlating more information adversary can find the correct identities of both presidents. However, in our scheme, striping minutes and seconds minimizes the risk of de-identification. In some scenario, the adversary can make some calls to the target user with the intentions to use this information for inferring friends of the target user. Similarly, like above case, the adversary knows call duration and call time of all his calls to the target subscriber. If timing information is not anonymized then the adversary can learn the target's pseudonymized identities, and so his friendship network. The adversary can also correlate multiple AUX to reduce the size of the candidate set. However, our proposed anonymization approach significantly reduces the risk but adversary can breach privacy by making some large number of bi-directional links which are normally not under his control.

5.3. Privacy protection at CR

The *CR* computes *GR* of users by aggregating the statistical information it received from the collaborating SP. The trusted CR ensures that provided information would not be misused and perform operations correctly, but it still has the possibility of privacy breach attack by the adversary. The exchange of reputation scores to the trusted CR is not revealing any information that might be used to infer the relationship network of users. However, in some circumstances, adversary or other SP can try to infer some information about the target by correlating local and global view of the user. In this scenarios, the SP can itself become adversary and wants to learn relationship network of target belonging to other SP from the received *GR*, and locally recorded CDR of the target. The adversary has the following AUX information at CR:

AUX4: The adversary knows *LR* of the target user and partial information about target user relationship network. Furthermore, the adversary also knows that target user only interacts with highly reputed subscribers or subscribers having similar reputation scores. The goal of the adversary is to predict possible relationship network of the target user in a target SP.

The exchange of single reputation score protects privacy protection against AUX 1, 2 and 3 ; however, the adversary SP can make a partial guess about relationship network of the target subscriber given AUX 4 but the probability of breach is extremely very small. The computational cost for such attack is also very high.

6. Experimental methodology

In this section, we present the methodology for creating the synthetic data and evaluation metric used to evaluate the performance of COSDS.

6.1. Synthetic dataset and methodology

We validated accuracy and correctness of our system using synthetic dataset that has been generated by simulating the realistic calling behavior of spammers and non-spammers. Fig. 3 illustrates our simulation model. The calling behavior of the user is modeled using fundamental call parameters that are call-rate, call duration, and a number of unique callees of the user. We considered learned distribution for these three call parameters for the legitimate and SPIT callers, respectively. Legitimate callers normally have high duration calls with a large number of their recipients; whereas SPIT callers have a large number of short duration calls [37,62] with the majority of them. SPIT callers typically do not call same recipients, again and again, this behavior is different from normal callers who exhibits repetitive calling behavior [62]. Finally, SPIT callers target a large number of recipients thus exhibit high out-degree, which is different from the normal callers, which has a small number of unique callees.

We generated the legitimate caller as following: (1) a power-law out-degree distribution is considered for generating the out-degree of the caller with the average out-degree of 10 [68,69]; (2) an exponential distribution with the average call duration of 360 s is used for the call duration [70], and (3) a Poisson distribution with mean value of 5 calls per day is used for the call rate. For generating spammers, we used the following configurations. The call duration of the spammer follows an exponential distribution but with the average duration of 90 s towards few callees, and the average duration of 40 s with a large number of callees [37,62], the out-degree of the spammer is randomly chosen between 500 and 2000 unique callees and the average call rate is 1.5.

The calls are distributed among all 5-service providers (70% calls made by legitimate users to users registered on the same

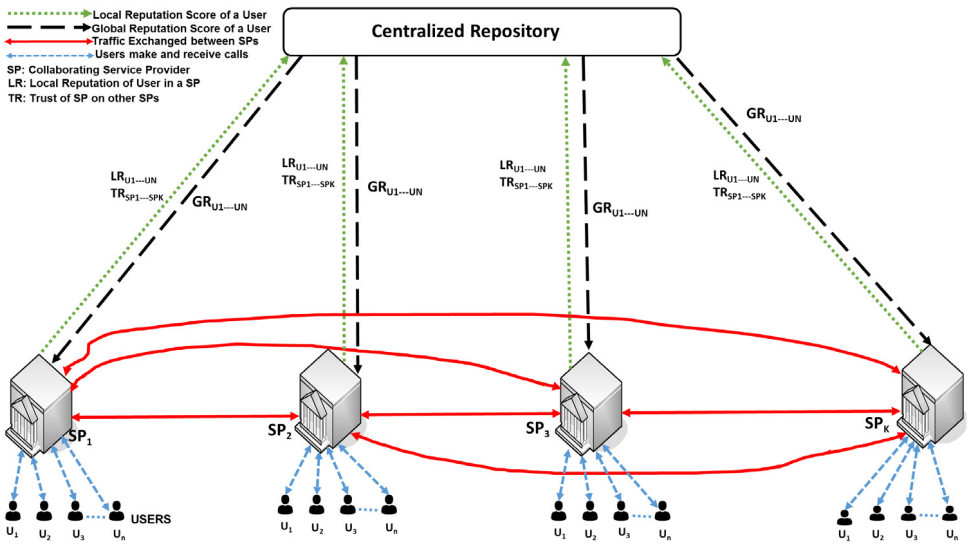
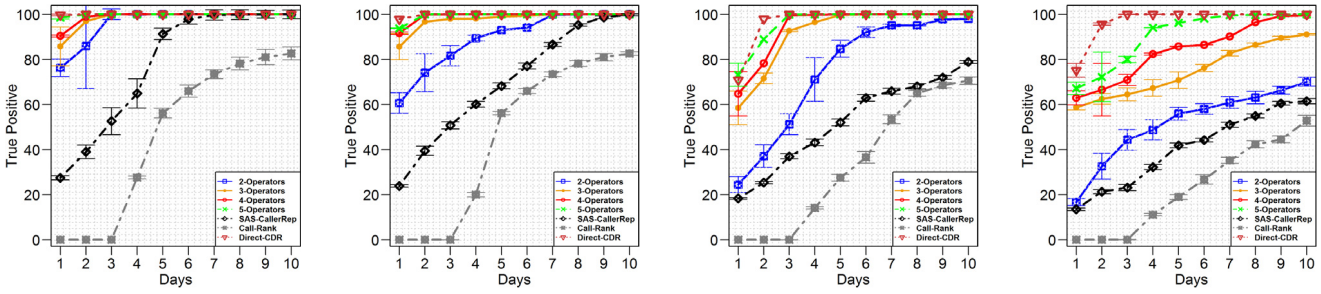


Fig. 3. Collaborative simulation model.



(a) 5% spammer. (b) 10% spammer. (c) 20% spammer. (d) 30% spammer.

Fig. 4. Detection rate of COSDS for SP trust=1 and β threshold=1.

Table 1
Confusion matrix.

Predicted/Actual	Spam	Not-spam
Spam	True positive (TP)	False positive (FP)
Not-spam	False negative (FN)	True negative (TN)

network while remaining 30% are equally distributed among other service providers). The number of legitimate users is fixed (50K) in each SP and the number of spammers varies from 5% to 30%. Each collaborating SP computes reputation score and reports the score to CR on daily basis. We repeated experiments for 10 times and reported average and standard deviation results.

6.2. Evaluation metrics

We used the following information retrieval performance metric to evaluate the performance of proposed approach: the detection or TP rate (TPR), the FP rate (FPR) and the detection accuracy (ACC). The detection rate or TPR is the average of a number of spam callers correctly identified as spam callers to the total number of spammers. The FPR is defined as the average of a total number of legitimate callers incorrectly classified as a spammer to the total number of legitimate callers. The evaluation metrics can be explained through the confusion matrix illustrated in a Table 1. The TPR, FPR and accuracy is computed as TPR= TP/(TP+FN), FPR = FP/(TN+FP) and ACC=(TP+TN)/(TP+TN+FN+FP).

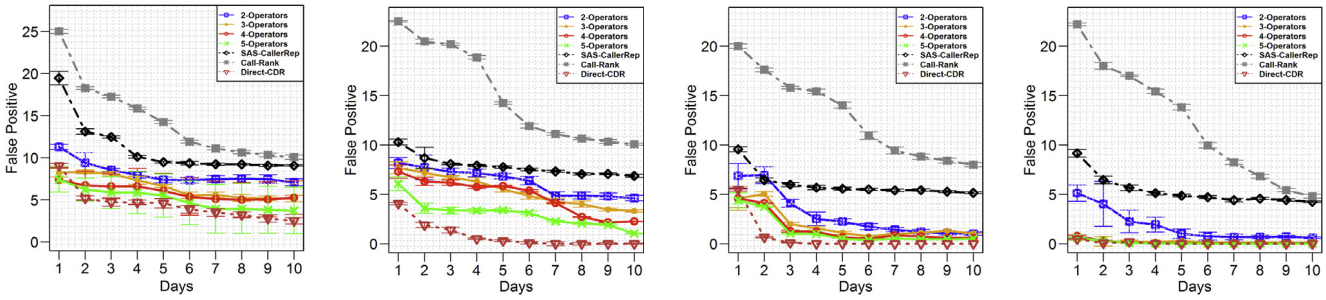
7. Performance evaluation

In this section, we provide performance results of COSDS and compare it with the Caller-REP and the Call-Rank system. We also compare its performance to two other collaboration options i.e. exchange of CDRs and direct trust scores to CR.

7.1. Detection rate

We evaluate the detection rate of COSDS system for three parameters: detection rate over the time, detection rate when the number of spammers varies from 5% to 30%, the detection rate with a different number of collaborators. Fig. 4 shows the detection rate of COSDS when the number of spammers varies from 5% to 30% and the number of collaborators varies from 2 to 5. We observe that COSDS shows effective resistance against spammers and blocks around 99% spammers within 3 days. Specifically, on a first day, COSDS manages to block 80% of spammers, that increases further to 100% detection with the time regardless of the number of spammers in the network. On the other hand, the non-collaborative systems show a slow resistance against stealthy spammers, that still allows these spammers even after 5 days. Specifically, standalone systems show effective resistance when the number of spammers is small (block 97% spammers in 5 days).

The improved performance of COSDS is attributed to the collaboration among SPs and method used for computing local reputation of the caller. The TPR of COSDS increases by a certain percentage with the number of collaborators as shown in a Fig. 4.



(a) 5% spammer. (b) 10% spammer. (c) 20% spammer. (d) 30% spammer.

Fig. 5. False positive rate of COSDS for SP trust=1 and β threshold=1.

Our results show that 4 SPs are enough for blocking above 98% of SPIT caller regardless of spamming rate. The collaboration with the exchange of reputation scores may result in a loss of some information about the behavior of the caller, but it still provides optimum detection rate as compared to a system having collaboration with CDR and trust score.

From Fig. 4, we also observe that Call-Rank has a degraded detection rate at the SP level compared to Caller-REP system. However, both Call-Rank and Caller-REP has worst detection capability when compared to the COSDS system. We attribute detection rate of both to the following. Call-Rank computes reputation of the caller using only one feature i.e. average call duration which can be easily evaded by spammers by managing a small number of long duration calls within his circle. On the other hand, Caller-REP utilizes three features collectively for computing reputation and are difficult to be evaded by the spammer.

7.2. False alarms

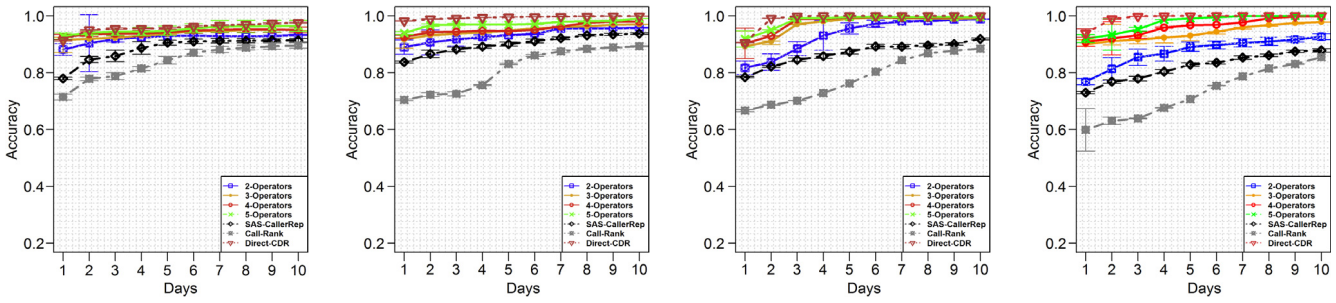
Although TPR is the key performance measure for evaluating the performance of any SPIT detection system, however, it should have ideally zero FPR. The false classification of the legitimate caller as the spammer not only annoys legitimate callers but it also results in a revenue loss for the SP because of blocking of legitimate callers. Fig. 5 represents the FPR of COSDS for different percentages of spammers and collaborators. COSDS achieves small FPR as compared to standalone systems. Particularly, it has 0% FPR in 3 days as compared to the non-collaborative approach which still has a high FPR even after 5 days shown in a Fig. 5. Similar to TPR, the FPR rate decreases with the number of collaborators. COSDS with five collaborators achieves a FPR of less than 5% during first three days for any percentage of spammers. Specifically, in a network with a small percentage of spammers such as 5% and 10%, COSDS misclassifies a large number of legitimate callers as spammers during first few days, and then have less than 5% FPR rate within 3 days shown in Fig. 5(a) and 5(b). Under a high spamming rate COSDS manages to have almost zero FP rate within 2 days shown in Fig. 5(c) and 5(d). The FP rate of non-collaborative approach and Call-Rank is not acceptable as both have FPR greater than 5% even after 5 days. Although COSDS only uses reputation scores from SP for global reputation and decision, other behavioral features can also be used to minimize the FP rate. One such feature is the out-degree of the caller because a small out-degree or a small number of calls in a defined time window do not characterize that caller is a spammer. The FP rate can also be minimized by using a fixed threshold β defined by the SPs according to their requirements. The FPR of COSDS and collaboration with the Direct-CDR are almost same.

7.3. Detection accuracy

The detection accuracy is the proportion of true identification (both true positives and true negatives) to the total number of callers (either spammer or legitimate). It characterizes system's capability of making the correct decision about callers. Under a small spamming rate, the COSDS approach misses a significant number of spammers and considerably blocks a high number of legitimate callers. However, under a high spamming rate, COSDS effectively detects most of spammers with a small FPR. Fig. 6 shows the accuracy of COSDS and other approaches when the number of spammers varies from 5% to 30%. Our experimental results show that the accuracy of the COSDS with five collaborator reaches to 100% in 4 days for any spamming rate which is much better than the non-collaborative system as shown in a Fig. 6. Specifically, we observe that COSDS reaches an overall accuracy of 99% in 5 days when the number of spammers is small ($<$10%), and reaches to the overall accuracy of 99% in three days when a number of spammers are high ($>$10%). This is due to the fact that at a small spamming rate COSDS misclassifies many legitimate callers as a spammers i.e. about 7% in three days, but FPR goes to less than 2% in 5 days. CDR based collaboration indicates a high detection accuracy than the COSDS approach. The detection accuracy can also be improved by incorporating other social network features (out-degree, clustering coefficient) along with the global and local reputation scores.

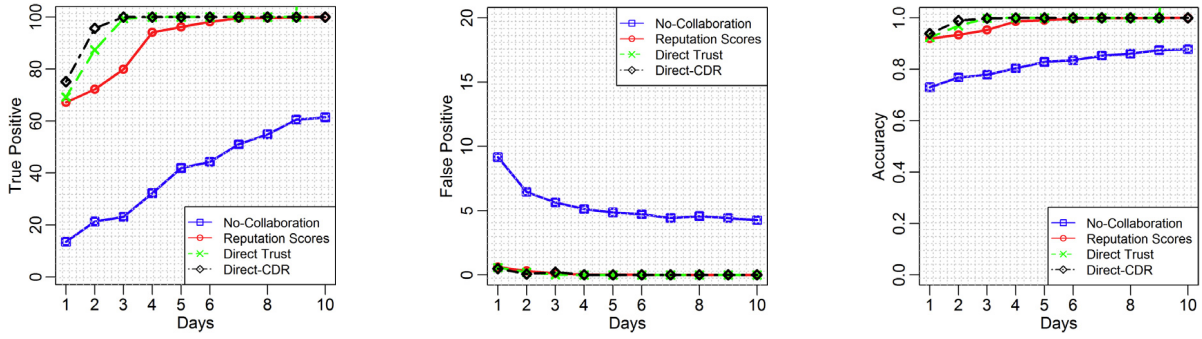
7.4. Privacy and system performance

The TPR increases and FPR decrease with the amount of information that is being exchanged for the collaboration. However, revealing a large amount of information (e.g call record data) would not only pose threat to the privacy of users but also require high communication overheads. Fig. 7 shows the TPR, the FPR and the accuracy of COSDS system and other design options that can be used for the collaboration. The results are shown for 30% spammers and for the 5 collaborators. The detection accuracy is opposite to that of privacy. If SP requires a complete privacy protection of their users then they would not collaborate with other SPs, and performs local detection only. This is similar to the standalone spam detection that has an absolute privacy protection but poor detection accuracy. We considered three different collaborative mechanisms in our experiments: (1) Collaboration with the complete CDRs; (2) Collaboration with the exchange direct trust scores of callers with their callees; and (3) collaboration with the exchange of reputation score of callers. The CDR based collaboration achieves best detection accuracy but at the cost of privacy and system resources. The direct trust-based collaboration though hides some sensitive information (time of call and duration), but it would not hide the social network of the caller. This approach achieves 100% TPR



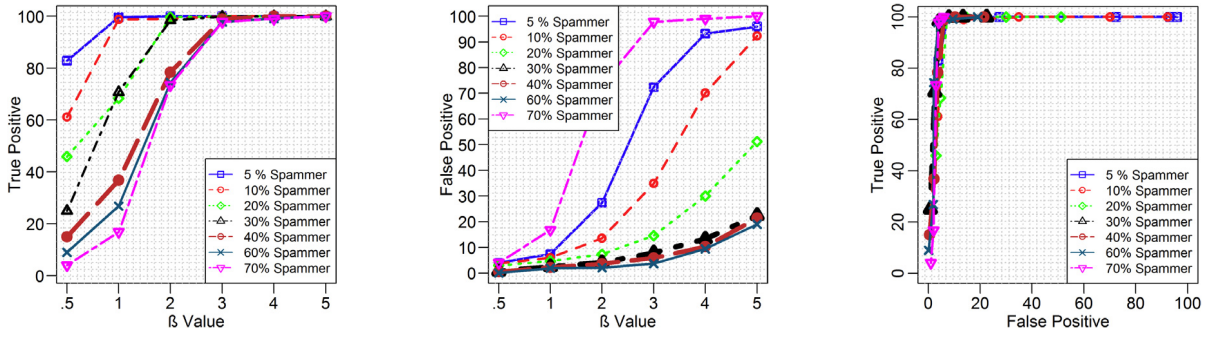
(a) 5% spammer. (b) 10% spammer. (c) 20% spammer. (d) 30% spammer.

Fig. 6. Detection accuracy for COSDS and non-collaborative system for SP trust=1 and β threshold=1.



(a) True positive rate. (b) False positive rate. (c) Accuracy.

Fig. 7. Privacy and true-positive, false positive and accuracy trade-off for different collaboration methods. The data for these plots has been taken from Figs. 4–6.



(a) True positive rate. (b) False positive rate. (c) ROC curve.

Fig. 8. The effect of threshold β for TP and FP Rates for 5 collaborators and for First Day.

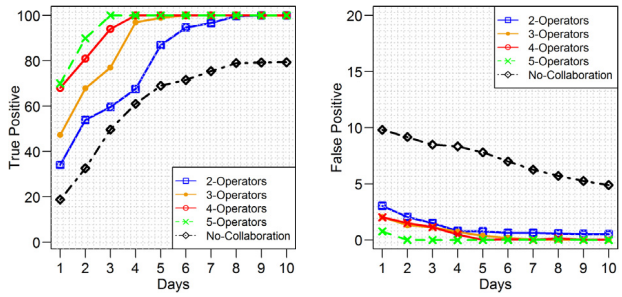
and zero FPR in 3 days, but relationship privacy is not protected. In comparison to first two approaches, COSDS not only protect privacy but also achieves comparable detection accuracy over the time.

7.5. Effect of threshold on performance

Earlier, we have discussed the performance of COSDS results for the threshold based on 25th percentile. Now we discuss the effect when the threshold is used along with the β parameter. The service provider set threshold value β according to his own requirements. The optimal choice for the β parameter depends on the relative trade-off between the TPR and the FPR. The major challenge to choose such value for β value that would incur a small FPR and a high TPR. The TPR and FPR of COSDS for the different β values are shown in a Fig. 8(a) and 8(b). The number of spammers varies from as low 5% to as high 70%. It can be seen from Fig. 8(a) that COSDS does not have high TPR when the β value is small. Specifically, in a network with more than 30% spammers, it shows poor resistance against spammers but provides better FPR as shown in Fig. 8(b). The choice of high β value could improve the TPR to 100% for any spamming rate, but it would relatively has high FPR for a network having spammers less than 10%. The FPR could be improved by using β value along with the number of unique callees of the caller. We recommend that β value should be between 2 and 3, should be used in conjunction with the number of unique callees of the subscriber. The trade-off between TPR and FPR for varying thresholds is shown in a Fig. 8(c). It can be seen from Fig. 8(c) that TPR of COSDS increases with the increase in FPR, that then increases to 100% TPR with relatively a small FPR.

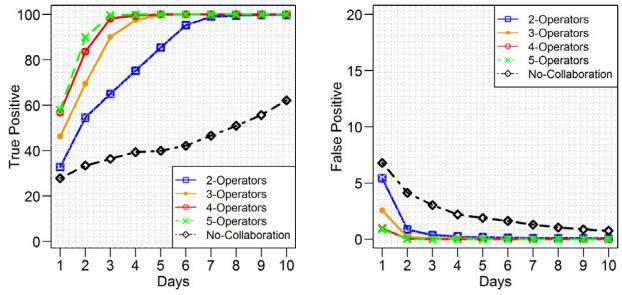
7.6. Resilience against different spam calling behaviors

In a real network scenario, it is possible that spammers have different calling behaviors. In this simulation setup, we evaluate the performance of COSDS system for three types of callers [37]. (1)



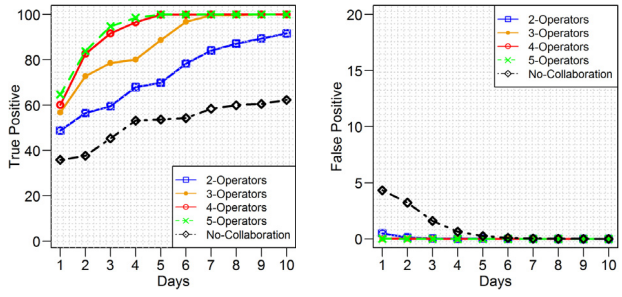
(a) True positive rate. (b) False positive rate.

Fig. 9. System behavior against spammers having high out-degree and high duration calls .



(a) True positive rate. (b) False positive rate.

Fig. 10. System behavior against spammers having small out-degree and small duration calls.



(a) True positive rate. (b) False positive rate.

Fig. 11. System behavior against spammers having small out-degree and long duration calls.

callers calling a large number of unique callees, all their successful calls have a good call duration, but callers do not receive calls from their callees. This would be a representative behavior of telemarketers and prank callers because of their high out-degree. (2) Callers calling a small number of callees per day, having only a few good duration calls, and also not receiving any call from the callees. These callers always try to call a limited number of new callees within a specific time. This characterizes the behavior of intelligent spammers. (3) Callers calling a small number of callees per day, and also manage to have high duration calls with many of them. This characterizes the behavior of caller where user wishes to interact with certain spammers marketing products callees like.

In the first experiment of this series, we generated the data with the following settings. The experiment setup consists of 15000 spammers, and 50000 legitimate callers equally distributed across five SPs. Each spam caller randomly chooses a callee and calls to 50% of the total number of callees in a SP. The average call duration of the caller varies from 180 s to 400 s with the average call duration of 220 s. The legitimate callers in this experiment follow the same distribution as provided in Section 6. The FPR and TPR of COSDS (collaboration with 3, 4 and 5 collaborators) are shown in Fig. 8(a) and 8(b). The FPR in this condition decreases with the number of collaborators and five collaborators are enough to have a FPR rate less than 0.5%. COSDS achieves a TPR of above 95% in six days much later than what it able to achieve with spammer model defined in Section 6. This is because of the long duration calls of the spammer to a large number of callees. Despite having high duration calls to a large number of callees, these callers are still identified as a spammer because of their high out-degree and non-repetitive calling behavior.

In the second experiment of this series, we simulated the user data with the following settings. The spammer controls his out-degree and calls only 15–25 unique callees per day. The call duration varies from 90 s to 200 s with the average duration of 150 s. The TPR and FPR rates of COSDS is shown in Fig. 10(a) and 10(b). It is clear from figures that COSDS is able to block all spammers in 4 days, that means controlling the out-degree per day would not allow the spammer to evade the system for a long time. As the caller increases his out-degree over time, its reputation starts decreasing because of high degree and few incoming calls.

In the final experiment of this series, we simulated the data with the following parameters. The spammers only call 35–50 unique callees per day, and the duration of each call is greater than 300 s, with the average duration of 350 s. The TPR and FPR rate of COSDS in this setup is shown in a Fig. 11(a) and 11(b). COSDS does not show effective resistance on the first few days because of the fact that call duration of spammer is almost same as the call duration of legitimate users. However, COSDS blocks almost all spammers after 6 days. The FPR rate in this scenario is also decreasing and remains less than .5%.

From Figs. 9–11, it can be seen that non-collaborative systems behave poorly in terms of the TPR and FPR for these types of spammers. From Figs. 9–11, we also conclude that spammers would be able to bypass the COSDS system for a relatively long time by simply managing long duration calls without controlling their number of callees, otherwise control out-degree to remain undetected and have repetitive call behavior.

8. Conclusions

In this paper, we presented a COSDS, a collaborative system that implies collaboration among telecommunication or VoIP SPs for an effective and rapid identification of stealthy spammers. In COSDS, the local reputation of a caller is computed by collectively using a number of social network and call features, and the global reputation is computed by aggregating the local reputation scores reported by the collaborating SPs. COSDS provides strong privacy protection to the collaborating SPs and their customers through the exchange of a single non-sensitive summarized information to the trusted CR. The SP, the trusted CR, and an adversary with some background information would not be able to infer the relationship network of users. The effectiveness of COSDS has been demonstrated using realistic synthetic call detailed records. The performance results show that COSDS outperforms non-collaborative system in terms of detection accuracy and detection time. The collaboration overheads increase with the number of users, and a number of collaborating SPs, that proves COSDS is a light weighted approach. As a part of future work, we are planning to investigate the design of a secure protocol for the exchange of reputation scores, a procedure for joining and leaving the CR, and combining different reputation based SPIT detection approaches.

Acknowledgments

The authors would like to acknowledge the financial support from the FCT (Portuguese Foundation for Science and Technology) with the associate laboratory contract INESCTEC under grant SFRH/BD/80135/2011. We also thank anonymous reviewers for their constructive comments.

References

[1] Consumer sentinel network reports, 2017. [Online]. Available: https://www.ftc.gov/enforcement/consumer-sentinel-network/reports. (Accessed March 2017).
[2] H. Tu, A. Doup, Z. Zhao, G. Ahn, Sok: Everyone hates robocalls: A survey of techniques against telephone spam, in: Proc. of 2016 IEEE Symposium on Security and Privacy, SP, 2016, pp. 320–338.
[3] Ftc issues fy 2016 national do not call registry data book, 2017. [Online]. Available: https://www.ftc.gov/news-events/press-releases/2016/12/ftc-issues-fy-2016-national-do-not-call-registry-data-book. (Accessed March 2017).
[4] V. Balasubramaniyan, M. Ahamad, H. Park, CallRank: Combating SPIT using call duration, social networks and global reputation, in: Proc. of The Fourth Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, CEAS, 2007.
[5] M.A. Azad, R. Morla, Caller-Rep: Detecting unwanted calls with caller social strength, Comput. Secur. 39 (Part B) (2013) 219–236.
[6] S. Roman, N. Saverio, T. Sandra, B. Marcus, SPam over Internet Telephony (SPIT) prevention framework, in: Proc. of IEEE Global Communications Conference, GLOBECOM, 2006, pp. 1–6.
[7] M. Mahoney, Dialing back: How phone companies can end unwanted robocalls, ConsumersUnion Policy and Actions from consumer reports, Tech. Rep. 2015. [Online]. Available: https://www.consumersunion.org/wp-content/uploads/2015/02/Dialing-Back-Complete-Report-11.16.2015.pdf.
[8] A. Tasidou, P.S. Efraimidis, Y. Soupionis, L. Mitrou, V. Katos, Privacy-preserving, user-centric voip captcha challenges: an integrated solution in the sip environment, Inf. Comput. Secur. 24 (2016) 2–19.
[9] Communications Fraud Control association (CFCA) Announces Results of Worldwide Telecom Fraud Survey published in 2016, 2016.
[10] M. Sahin, A. Francillon, P. Gupta, M. Ahamad, Sok: Fraud in telephony networks, in: Proc. of 2017 IEEE European Symposium on Security and Privacy, EuroS P, 2017, pp. 235–250.
[11] M. Sahin, A. Francillon, Over-the-top bypass: Study of a recent telephony fraud, in: Proc. of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 1106–1117.
[12] M. Hansen, M. Hansen, J. Möller, T. Rohwer, C. Tolkmit, H. Waack, Developing a legally compliant reachability management system as a countermeasure against SPIT, in: Proc. of The Third Annual VoIP Security Workshop, 2006, pp. 1–7.
[13] D. Shin, J. Ahn, C. Shim, Progressive multi gray-leveling: A voice spam protection algorithm, IEEE Netw. 20 (2006) 18–24.
[14] J. Rosenberg, C. Jennings, The Session Initiation Protocol (SIP) and Spam. RFC 5039, January 2008.
[15] Y.-S. Wu, S. Bagchi, N. Singh, R. Wita, Spam detection in voice- over-IP calls through semi-supervised clustering, in: Proc. of 39th Annual Conference on Dependable Systems and Networks, DSN, 2009, pp. 307–316.
[16] H. Bokharaei, A. Sahraei, Y. Ganjali, R. Keralapura, A. Nucci, You can SPIT, but You can't hide: Spammer identification in telephony networks, in: Proc. of 2011 IEEE International Conference on Computer Communications, 2011, pp. 41–45.
[17] H. Sengar, X. Wang, A. Nichols, Thwarting spam over internet telephony (SPIT) attacks on VoIP networks, in: Proc. of 19th ACM International Symposium on Quality of Service, IWQos, 2011, pp. 1–3.
[18] H. Sengar, X. Wang, A. Nichols, Call behavioral analysis to thwart SPIT attacks on VoIP networks, in: Proc. of Security and Privacy in Communication Networks, 2012, pp. 501–510.
[19] M.A. Azad, R. Morla, Mitigating spit with social strength, in: Proc. of The 12th IEEE International Conference On Trust, Security and Privacy in Computing and Communications, IEEE TrustCom, 2012, pp. 1393–1398.
[20] T. Kusumoto, E.Y. Chen, M. Itoh, Using call patterns to detect unwanted communication callers, in: Proc. of International Symposium on Applications and the Internet, SAINT, 2009, pp. 64–70.
[21] N. Chaisamran, T. Okuda, G. Blanc, S. Yamaguchi, Trust-based VoIP spam detection based on call duration and human relationships, in: Proc. of IEEE/IPSJ International Symposium on Applications and the Internet, 2011, pp. 451–456.
[22] R. Zhang, A. Gurtov, Collaborative reputation-based voice spam filtering, in: Proc. of International Conference on Database and Expert Systems Applications, DEXA, 2009, pp. 33–37.
[23] K. Toyoda, M. Park, N. Okazaki, T. Ohtsuki, Novel unsupervised spitters detection scheme by automatically solving unbalanced situation, IEEE Access 5 (2017) 6746–6756.
[24] R. Jabeur Ben Chikha, T. Abbes, W. Ben Chikha, A. Bouhoula, Behavior-based approach to detect spam over ip telephony attacks, Int. J. Inf. Secur. 15 (2016) 131–143.
[25] J. Lindqvist, M. Komu, Cure for spam over internet telephony, in: Proc. of The 4th Annual IEEE Consumer Communications & Networking Conference, 2007, pp. 896–900.
[26] J. Quittek, S. Niccolini, S. Tartarelli, R. Schlegel, On spam over internet telephony (SPIT) prevention, IEEE Commun. Mag. 46 (2008) 80–86..
[27] J. Quittek, S. Niccolini, S. Tartarelli, M. Stiemerling, M. Brunner, T. Ewald, Detecting SPIT calls by checking human communication patterns, in: Proc. of IEEE International Conference on Communications, 2007, pp. 1979–1984.
[28] Y. Rebahi, D. Sisalem, T. Magedanz, SIP spam detection, in: Proc. of International Conference on Digital Telecommunications, ICDT, 2006, pp. 68–74.
[29] Y. Hong, S. Kunwadee, Z. Hui, S. ZonYin, S. Debanjan, Incorporating active fingerprinting into SPIT prevention systems, in: Proc. of The 3rd Annual VoIP Security Workshop, 2006.
[30] D. Lentzen, G. Grutzek, H. Knospe, C. Porschmann, Content-based detection and prevention of spam over IP telephony - system design, prototype and first results, in: Proc. of IEEE IEEE International Conference on Communications, ICC, 2011, pp. 1–5.
[31] G. Zhang, S. Fischer-Hübner, Detecting near-duplicate spits in voice mailboxes using hashes, in: Proc. of 14th International Conference on Information Security, 2011, pp. 152–167.
[32] A. Seyed, S. Hemant, W. Haining, A voice spam filter to clean subscriber's mailbox, in: Proc. of 8th International Conference on Security and Privacy in Communication Networks, 2012, pp. 349–367.
[33] J.L. Tabron, Linguistic features of phone scams: A qualitative survey. in: Proc. 11th Annual Symposium on Information Assurance, ASIA'16, 2016.
[34] V.A. Balasubramaniyan, A. Poonawalla, M. Ahamad, M.T. Hunter, P. Traynor, PinDr0P: Using single-ended audio features to determine call provenance, in: Proc. of the 17th ACM Conference on Computer and Communications Security, 2010, pp. 109–120.
[35] M. Robert, V. Dmitri, Detection and mitigation of spam in IP telephony networks using signaling protocol analysis, in: Proc. of IEEE/sanoff Symposium on Advances in Wired and Wireless Communications, 2005, pp. 49–52.
[36] M.A. Akbar, M. Farooq, Securing SIP-based voip infrastructure against flooding attacks and spam over IP telephony, Knowl. Inf. Syst. 38 (2014) 491–510.
[37] S. Chiappetta, C. Mazzariello, R. Presta, S. Romano, An anomaly-based approach to the analysis of the social behavior of voip users, Comput. Netw. 57 (2013) 1545– 1559.
[38] J. Seedorf, N. d'Heureuse, S. Niccolini, M. Cornolti, Detecting trustworthy real-time communications using a web-of-trust, in: Proc. of IEEE Global Communications Conference, GLOBECOM, 2009, pp. 1–8.
[39] P. Kolan, R. Dantu, Socio-technical defense against voice spamming, ACM Trans. Auton. Adapt. Syst. 2 (2007).
[40] R. Dantu, P. Kolan, Detecting spam in VoIP networks, in: Proc. of The Steps to Reducing Unwanted Traffic on the Internet, 2005, pp. 31–37.
[41] M.A. Azad, R. Morla, Multistage SPIT Detection in transit VoIP, in: Proc. of The IEEE 19th International Conference on Software, Telecommunications and Computer Networks, SoftCOM, 2011, pp. 1–9.
[42] G. Vennila, M.S.K. Manikandan, M.N. Suresh, Detection and prevention of spam over internet telephony in voice over internet protocol networks using Markov chain with incremental svm, Int. J. Commun. Syst. 30 (2017).
[43] P. Gupta, B. Srinivasan, V. Balasubramaniyan, M. Ahamad, Phoneypot: Data-driven understanding of telephony threats, in: Proc. of 20th Network & Distributed System Security Symposium, NDSS, 2015.
[44] M. Balduzzi, P. Gupta, L. Gu, D. Gao, M. Ahamad, MobiPot: Understanding mobile telephony threats with honeycards, in: Proc. of 11th ACM ASIA SIGSAC Conference on Computer and Communications Security, ASIA CCS, 2016.
[45] A. Marzuoli, H. Kingravi, D. Dewey, R. Pienta, Uncovering the landscape of fraud and spam in the telephony channel, in: Proc. of 15th IEEE International Conference on Machine Learning and Applications, ICMLA, 2016, pp. 853–858.
[46] M. Najmeh, S. Oleksii, N. Nikiforakis, Dial one for scam: A large-scale analysis of technical support scams, in: Proc. of the 24th Network and Distributed System Security Symposium, NDSS, 2017.
[47] M. Sahin, M. Relieu, A. Francillon, Using chatbots against voice spam: Analyzing lenny's effectiveness, in: Proc. of Thirteenth Symposium on Usable Privacy and Security, SOUPS, 2017, pp. 319–337.
[48] M.A Azad, R. Morla, Early identification of spammers through identity linking, social network and call features, J. Comput. Sci (2016).
[49] Y. Chen, K. Hwang, W.-S. Ku, Collaborative detection of ddos attacks over multiple network domains, IEEE Trans. Parallel Distrib. Syst. 18 (2007) 1649–1662.
[50] E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, P. Samarati, P2p-based collaborative spam detection and filtering, in: Proc. of Fourth International Conference on P2P Computing, 2004, pp. 176–183.
[51] K. Li, Z. Zhong, L. Ramaswamy, Privacy-aware collaborative spam filtering, IEEE Trans. Parallel Distrib. Syst. 20 (2009) 725–739.

[52] M.A. Azad, S. Bag, S. Tabassum, F. Hao, privy: Privacy preserving collaboration across multiple service providers to combat telecoms spam, IEEE Trans. Emerg. Top. Comput.
[53] M. Sirivianos, K. Kim, X. Yang, Socialfilter: Introducing social trust to collaborative spam mitigation, in: Proc. of Workshop on Collaborative Methods for Security and Privacy, collSec, 2010.
[54] Distributed checksum clearinghouses. [Online]. Available: http://www.rhyolite.com/dcc/.
[55] G. Singaraju, B.B. Kang, Repuscore: Collaborative reputation management framework for email infrastructure, in: Proc. of 21st conference on Large Installation System Administration, 2007, pp. 1–9.
[56] J. Kong, B. Rezaei, N. Sarshar, V. Roychowdhury, P. Boykin, Collaborative spam filtering using e-mail networks, IEEE Comput. 39 (2006) 67–73.
[57] B. Mathieu, S. Niccolini, D. Sisalem, SDRS: A voice-over-IP spam detection and reaction system, IEEE Secur. Priv. 6 (2008) 52–59.
[58] C. Sorge, J. Seedorf, A Provider-level reputation system for assessing the quality of SPIT mitigation algorithms, in: Proc. of IEEE IEEE International Conference on Communications, ICC, 2009, pp. 1–6.
[59] Y.-S. Wu, V. Apte, S. Bagchi, S. Garg, N. Singh, Intrusion detection in voice over IP Environments, Int. J. Inf. Secur. 8 (2009) 153–172.
[60] A. Gazdar, Z. Langar, A. Belghith, A distributed cooperative detection scheme for SPIT attacks in SIP based systems, in: Proc. of Third Network of the Future, NOF, 2012, pp. 1–5.
[61] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, J. Leskove, Mobile call graphs: Beyond power-law and lognormal distributions, in: Proc. of 14th ACM Special Interest Group on Knowledge Discovery in Data, SIGKDD, 2008, pp. 596–604.
[62] N. d'Heureuse, S. Tartarelli, S. Niccolini, Analyzing telemarketer behavior in massive telecom data records, in: Springer Trustworthy Internet, 2011, pp. 261–271.
[63] A. Ramachandran, N. Feamster, S. Vempala, Filtering spam with behavioral blacklisting, in: Proc. of 14th ACM CCS, 2007, pp. 342–351.
[64] A. Ramachandran, N. Feamster, Understanding the network-level behavior of spammers, in: SIGCOMM Computer Communication Review, vol. 36, 2006, pp. 291–302.
[65] R. Zhou, K. Hwang, M. Cai, Gossiptrust for fast reputation aggregation in peer-to-peer networks, IEEE Trans. Knowl. Data Eng. 20 (2008) 1282–1295.
[66] R. Zhou, K. Hwang, Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing, IEEE Trans. Parallel Distrib. Syst. 18 (4) (2007) 460–473.
[67] S.D. Kamvar, M.T. Schlosser, H. Garcia-Molina, The eigentrust algorithm for reputation management in P2P networks, in: Proc. of 12th International Conference on World Wide Web, 2003, pp. 640–651.
[68] A.A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, A. Joshi, On the structural properties of massive telecom call graphs: Findings and implications, in: Proc. of 15th ACM International Conference on Information and Knowledge Management, CIKM, 2006, pp. 435–444.
[69] W. Henecka, M. Roughan, Privacy-preserving fraud detection across multiple phone record databases, IEEE Trans. Dependable Secure Comput. 12 (2015) 640–651.
[70] P. De Melo, L. Akoglu, C. Faloutsos, A. Loureiro, Surprising patterns for the call duration distribution of mobile phone users, in: Proc. of The 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part III, 2010, pp. 354–369.

Muhammad Ajmal Azad received the Ph.D. (2016) degree in Electrical and Computer Engineering from the University of Porto, Portugal and M.S. (2008) in Electronics Engineering from the International Islamic University Pakistan. He is currently research associate at department of computing sciences in Newcastle University, United Kingdom. His research interests include privacy-aware collaboration, reputation aggregation, privacy protection, privacy aware outsourcing of network logs and spam detection in telecommunication network.

Ricardo Morla received the Ph.D. degree in Computing from the Lancaster University. He is currently an assistant professor in the Department of Electrical and Computer Engineering U.Porto. His research interests include distributed computer systems and computer networks, with an emphasis on P2P content delivery networks, network management and application of machine learning to the management of IT systems.