ROBO-SPOT: Detecting Robocalls by Understanding User Engagement and Connectivity Graph

Muhammad Ajmal Azad, Junaid Arshad and Farhan Riaz

***Abstract*—Robo or unsolicited calls have become a persistent issue in telecommunication networks, posing significant challenges to individuals, businesses, and regulatory authorities. These calls not only trick users to disclose their private and financial information but also affect their productivity through unwanted phone ringing. A proactive approach to identify and block such unsolicited calls is essential to protect users and service providers from potential harm. Therein, this paper proposes a solution to identify robo-callers in the telephony network utilising a set of novel features to evaluate the trustworthiness of callers in a network. The trust score of the callers is then used along with machine learning models to classify them as legitimate or robo-caller. We used a large anonymized data set (call detailed records) from a large telecommunication provider containing more than 1 billion records collected over 10 days. We have conducted extensive evaluation demonstrating that the proposed approach achieves high accuracy and detection rate whilst minimizing the error rate. Specifically, the proposed features when used collectively achieve a true-positive rate of around 97% with a false-positive rate of less than 0.01%.**

***Index Terms*—Social Network Analysis, Reputation, SPIT, Unwanted Calls, Robo-callers, Telephone Network**

I. Introduction

Telephone networks (Mobile, Voice over IP (VoIP) and Fixed landline networks) play an important role in modern communication, allowing people and businesses to connect. Individuals and businesses benefit from telephony services in a variety of ways, including improved real-time communication, connectivity, and overall ease. The number of telephone subscribers (Mobile, VoIP, Landline) across the world exceed more than 8 billion [1]. These networks have also attracted unwanted callers, also known as robocalls or spam calls, targeting the users with a nuisance and unwanted calls. These calls can severely affect the productivity of individuals and businesses, cause anxiety and annoyance, and waste valuable time. Moreover, unwanted calls can also be used for fraudulent purposes such as phishing scams, identity theft, and other forms of social engineering. Scammers often use robocalls as the first or medium to trick people into providing personal information to be used for financial fraud. Recent statistics on telephony spam reveal that answering spam calls has an adverse impact on worker productivity losing 20 million man-hours resulting in a financial loss of about $475 million annually [2]. Consequently, unwanted calls have become a major concern for regulators, telecommunication operators, and law enforcement agencies as they receive hundreds of thousands of consumer complaints about unsolicited, unauthorized and fraudulent calls. For instance, it is estimated that consumers in the United States have received more than 50 billion spam phone calls in 2022 [3]. Specifically, in 2021, FTC (Federal Trade Communication) received more than 1.8 million complaints about robocalls [4] and citizens lost around $39.5 billion to phone scams in 2022 [5].

Several solutions have been proposed to deal with the challenge of spamming in telecommunication networks. List-based systems such as [6], [7] are among the simplest solutions that manage a database of *white* (callers who are allowed to call) and *blacklisted callers* (identities barred from calling). These systems required dynamic updates of the list database and are susceptible to identity spoofing attacks [8]. Blacklisting or whitelisting all new identities is also not an optimal solution as it would block many legitimate callers whilst allowing many spam callers. Further, a list-based system needs to collaborate with other approaches e.g. reputation or machine-learning-based systems to decide whether a caller should be included in a white or blacklist [9]. Reputation-based systems such as [10]–[13] compute the reputation of a caller based on the feedback from callees of the call or using call-related information such as call detailed records (CDRs) or signalling information. Although reputation-based systems could block spam calls, their performance depends on features used for creating a reputation model. Another popular approach to identify spam callers is to involve multiple service providers in the collaboration process [14]–[16]. Telecommunication operators can also collaborate [17], [18] by providing information about the behaviour of a caller within their respective networks to identify the stealthy and slow-rate spammers.

Legitimate subscribers usually develop social circles with their friends, family, and colleagues with whom they interact more frequently whereas, malicious subscribers exercise massive spamming to a large number of subscribers, which normally results in a non-connected social network. For example, telemarketers usually call or send messages to a large number of subscribers which often results in a majority of small-duration calls. On the other hand, legitimate callers normally have connected social circles and developed a strong relationship network with a large number of users with high-duration

Muhammad Ajmal Azad (e-mail: muhammadajmal.azad@bcu.ac.uk) and Junaid Arshad (junaid.arshad@bcu.ac.uk) are with School of Computing and Digital Technology, Birmingham City University, United Kingdom. Farhan Riaz (e-mail: friaz@lincoln.ac.uk) is with the School of Computer Science, University of Lincoln, United Kingdom.

Manuscript received ..

calls. Consequently, legitimate callers usually spent 80% of their talk time with only a few strongly connected friends [19] whereas spam callers exceptionally have a high number of callees. Therefore, the communication behaviour of spammers is different from that of legitimate calls whose interactions are restricted to a social group. By analyzing the patterns of calls and social connections between phone numbers, the service provider can protect consumers from unwanted calls and other forms of telephone-based fraud. Detecting unwanted calls using social network analysis can involve analyzing the calling patterns of callers in the telecommunication network [20], [21], [22]. Within telecommunication networks, the social behaviour of users can be computed from call duration, call intensity, and callee feedback [10], [12], [23], [19]. However, these studies use one feature to compute the trustworthiness of the caller, which can be easily circumvented by spammers and telemarketers.

The use of reputation-based systems along with machine learning could be a robust approach to block robo-callers, improve detection accuracy and minimize false positives. In this paper, we present a system called ROBO-SPOT that automatically classifies a caller as malicious or legitimate based on his relationship network. To this extent, first, we analyzed user behavioural attributes from the real call detailed records (CDRs) and evaluate his reputation through a diverse set of features. Secondly, we integrated machine learning into these novel features for the automatic classification of the caller as legitimate or malicious. We believe this is the first study that analyses a real labelled dataset from a telecommunication operator to characterize the behaviour of spammers and non-spammers. Previous studies that focused on analyzing CDRs use data collected at a honeynet [24], [25] and apply machine learning models over a manually labelled dataset [26], [27]. Our work is different from the previous work both in terms of data analysis as well as the novelty of the features used for the computation of the reputation of the caller.

The major contributions of this work are:

- We analyzed CDRs obtained from a large telecommunication operator and design a suitable reputation system based on the communication behaviour of users towards others. We specifically used user connectivity, degree distribution, call duration, and relationship network to compute the overall reputation of the caller.
- We devised and deployed a machine learning model on the reputation scored in order to classify the caller as malicious or non-malicious. Our system has demonstrated high accuracy and can detect spammers with a very small false positive rate.

The rest of the paper is organized as follows. Section II presents the literature review and the motivation for this work. Section III describes the behavioral properties of spammers and non-spammers. Section IV defines the problem and presents a reputation framework. Section V presents a discussion on the spam detection process. Section VI evaluates the performance using various machine learning methods. Section VII concludes the paper.

II. Related Work

In this section, we briefly discuss work that has been proposed for detecting spammers in telephone networks. Overall, such detection systems can be categorized into the following types: i) black-and-white list-based systems which assign the identity of a caller to the respective database based on the behaviour of the caller, and ii) systems employing behavioural and social graph-based mechanisms to quantify the reputation of a caller within the network, processing speech streams and blocking caller if the caller speech stream matches the known spam words. A comprehensive taxonomy of spam detection systems proposed for telecommunication and Voice over IP (VoIP) networks can be found in [28], [29]. In this paper, we discuss related works that employ machine-learning approaches to classify a caller as a spammer or non-spammer.

Content-based detection systems analyze speech content exchanged between a caller and a callee and block spammers if the speech contains known spam words [30]–[33]. However, the application of content-based systems in real-time communication has several limitations. Firstly, it introduces some noticeable delays between conversations of subscribers. Secondly, the operators require sophisticated software and hardware resources to process the speech streams in real time. Thirdly, and most importantly, content-based systems decide about the caller after the call has already been established and the user has already been annoyed with the call. Furthermore, the privacy of the subscriber is not ensured and speech processing is prohibited by law in many countries.

List-based approaches are identity-based detection systems that maintain a database of black, white and grey identities [6], [7]. The call processing engine consults the list database during the call setup phase and allows or blocks the caller. A list database can be implemented in a personalized setting, applicable to users only, and a global setting, in which one list is used for all subscribers of the network. A grey list can also be used for maintaining the list of subscribers to be observed for a further time period. List-based approaches need to be implemented along with other approaches [34], [35]. A common problem with list-based systems is to manage the fast-growing list database. Further, the list-based system can be easily circumvented by spoofing the identities of legitimate subscribers.

Several behavioural-based approaches have also been proposed that estimate the trustworthiness of a subscriber based on the calling behaviour and social connections of the subscriber. For example, CallRank [10] estimates the trustworthiness of a subscriber in two steps: 1) computing the direct trust between a subscriber and his callee using the average call duration, and 2) estimating the global reputation of the subscriber using the Eigen trust algorithm. The system requires assistance from the callee to decide whether to accept the call or not by providing the reputation score of the subscriber to the callee. Similarly, CallREP [13] estimates the reputation of the subscriber by collectively using several social features together i.e. call duration, call rate and out-degree of the caller. The system blocks the spammer based on a fixed or automated classification threshold. Zhang et al. [36] used call duration as the feature to estimate the reputed behaviour of a subscriber. Kolan et al. [11], [12] proposed a multistage system that consists of three stages. The first stage computes the trust score of the subscriber with others by getting feedback from the callees of the subscriber. The second stage computes the global reputation of the subscribers by applying the Bayesian network algorithm. The third state compares the identity of the subscriber with the list database that is being updated using the first and second stages. Gupta et al. [24] deploy a large-scale telephone honeypot system for analyzing the social behaviour of subscribers making calls to these honey-phones. The study assumes that only spammers call these identities, but it does not have information about how these spammers behave with other network users. Balduzzi [25] deploys a mobile honeypot for collecting fraudulent calls and short messages. These calls and SMS are then analyzed for studying the mechanism used by the spammers for collecting target identities and their calling patterns.

Several machine learning-based detection systems have been proposed for detecting spammers in VoIP and telecommunication networks. Yu-Sung et al. [37] use the extended K-mean clustering algorithm based on the call parameters (messages exchanged during call setup, and termination) along with the callee feedback about the behaviour of the subscriber. Azad et al. [38] utilize the K-mean cluster algorithm and use social and behavioural features of the caller to mark the caller as a spammer or a non-spammer. Liu et al. [27] discovered the telephone numbers involved in spam campaigns by using unsupervised and supervised machine learning methods along with the known spam phone numbers to find out new spammers. Sharbani et al. [39] estimate the effectiveness of spam blacklists by measuring their ability to block future unwanted phone calls. Li et al. [26] use 29 features along with machine learning algorithms to predict whether the subscriber is a legitimate user or a spammer. Chiappetta et al. [40] used an unsupervised clustering algorithm i.e., the K-Means algorithm to group users based on the behavioural model.

Towards the design of robust techniques to stop the spammers at the edge of the network without adding any intrusiveness to the user, this paper analyses the social behaviour of the callers in the large telecommunication service providers, which log the complete call data of all users. We perform a detailed analysis of the call detailed data records for different network and social features of the spammers and non-spammers and utilize the findings along with machine learning to classify them into different classes. The novelty of this work is that it utilizes unique social network features for characterizing the behaviour of the users and calculation of reputation scores and uses neural networks to classify callers as spammers or non-spammer.

III. Data Representation and Measurements

In this section, first, we describe our data set and then we analyze the behavioural characteristics and properties of subscribers.

A. Data Set

Telecommunication operators record call transactions of their customers in Call Detail Records (CDRs), which are primarily used for billing purposes and network management. Telecommunication operators can also utilize these records for characterizing subscribers for other purposes such as marketing, identification of disease outbreaks and identification of malicious subscribers. A CDR usually contains meta-data of call transactions without any speech content. A typical CDR consists of a number of parameters. These include identities of the caller and the callee (subscriber of the network), initiation time of the call, disconnection time of the call, call duration, disconnecting party, call type (voice, SMS, MMS) and status of the call (successful or failed).

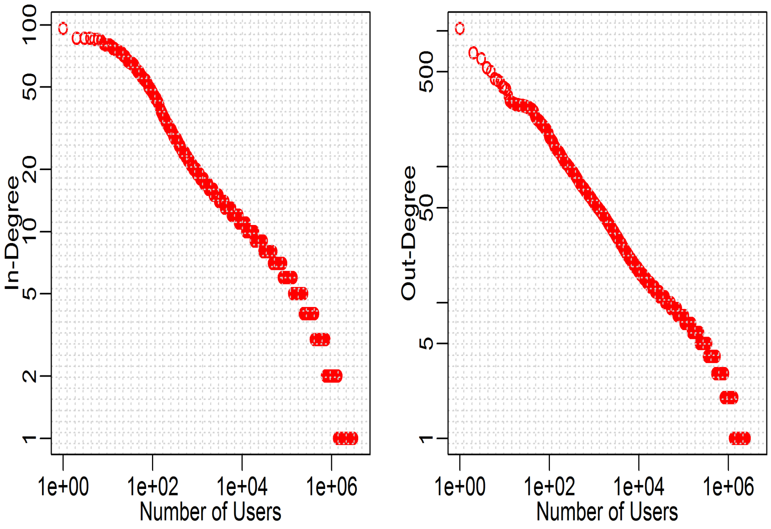
In order to evaluate real-life data, we worked with the largest telecommunication provider and collected anonymized CDRs containing 1 billion call records related to 3 million subscribers across 10 days. The privacy of the subscribers in this data set is ensured by assigning a random anonymized identity to each subscriber. Furthermore, the call time of the caller is also rounded to the nearest hour in order to minimize the risk of de-identification [41]. The average calling rate of subscribers per second is 10 to 280 calls between midnight and mid-day. The average number of calls made by a subscriber to other subscribers is 2.8 in a day. Figure 1 B and C represent the in-degree and out-degree distribution of subscribers for one day. Figure 2 A and B represent the behavioural features (average call rate and average call duration) of all subscribers in the dataset. In addition to the data set, the telecommunication operator has also provided the pseudonyms of confirmed classified spammers (15K). In our study, we first analyze the calling behaviour of labelled legitimate and spam subscribers followed by proposing a novel feature set to improve the automatic classification of unlabelled subscribers as legitimate users or spammers.

B. Data-Representation

The CDR data of subscribers can be represented as a connected weighted social graph. The weighted call graph G is modelled as (V, E, W) where a node V represents the identity of the subscriber (caller or callee), an edge E is drawn if users interacted with each other at least once, and the weight W on the edges defines the connectivity strength between subscribers. This weight can be derived from the frequency of interactions and the duration of the interaction, simultaneously. We model the incoming and outgoing calls as a separate edge between subscribers. The call graph of the subscribers can also be represented as a sparse adjacency matrix, where 1 represents caller S interacted with callee R and 0 represents no interaction between the caller and the callee. A $n \times n$ adjacency matrix A is represented as follows.

$$A_{ij} = \begin{cases} 1; & \text{if } i \text{ interacted } j \\ 0; & \text{Otherwise} \end{cases} \tag{1}$$

In the case of the weighted call graph, A_{ij} is replaced by the weights determined from the frequency of interaction and



(a) In-Degree Distribution

(b) Out-Degree Distribution

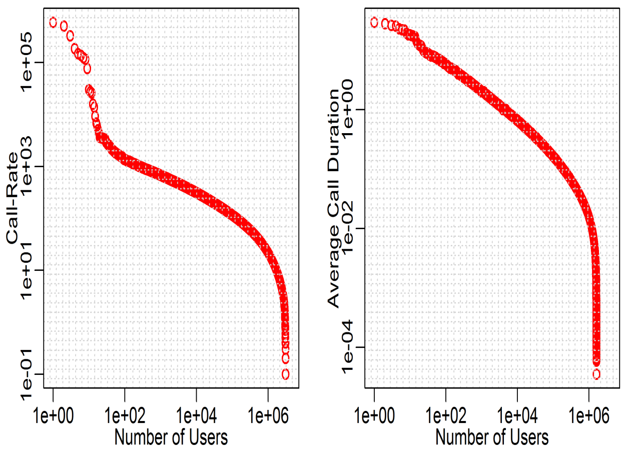
Fig. 1: Distribution representation of subscribers relationship network from the Call detailed records. The call duration in B is in seconds.

the call duration of interactions. In Equation 1, i represents the caller whereas the callee is represented by j. The weight also represents how much trust subscribers have in each other. In this paper, we modelled the weighted call graph using the following information from the CDRs.

Call Duration: Call duration represents the length of time that two subscribers spoke to each other. Specifically, the call duration of caller i to callee j is the sum of the duration of all calls made by caller i to callee j. The aggregated call duration, therefore, is the sum of call durations of all calls made and received by a subscriber A. The call duration distribution of all subscribers is shown in Figure 2 B.

Call-Rate: Call-Rate represents the frequency of interaction between the caller and the callee. Specifically, the call-rate between the caller i and the callee j is the sum of all calls made from the caller i to the callee j. The aggregated call-rate, therefore, is the sum of all calls made and received by the subscriber A. The call-rate distribution of subscribers is shown in Figure 2 A.

Degree: Each subscriber has some incoming links (total number of unique callers who initiated a call to a subscriber), and some outgoing links (unique callees a certain subscriber has initiated calls to). The number of outlines of a caller i is its out-degree, whereas the number of in-links is the in-degree of the caller i. Total Degree is the sum of in-degree and out-degree. The in-degree of the caller i is represented as ID_i and the out-degree of the caller i is represented as OD_i. The outgoing Interactions represent that a subscriber is more important to some subscribers than those to whom he did not initiate any call. The in-degree and out-degree distribution of subscribers for the dataset is shown in Figure 1 B and C, respectively.



(a) Average Call Rate

(b) Average Call Duration

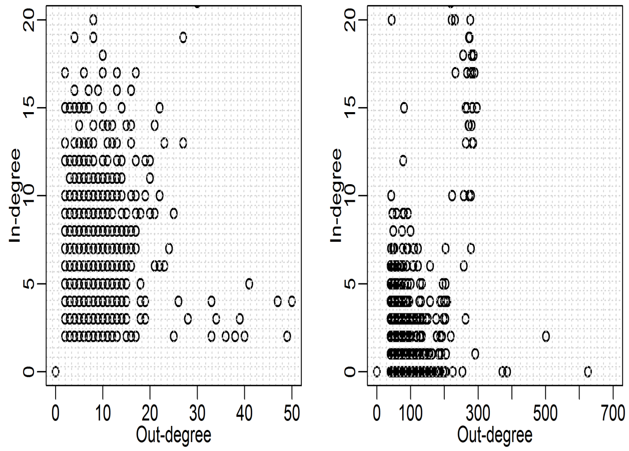
Fig. 2: Distribution representation of Behavioral features of all subscribers from the call detailed records.

C. Ethics

The data used in this research is provided by a telecommunication operator. The identities of the caller, callee and the time stamp of call records have been anonymized by the operator. The region of the provided data is not disclosed, thus it cannot be deanonymized by the data handler. Further, data is seen only by one author based and is not moved outside the country.

D. Data Measurement and Analysis

Legitimate users and spammers have different goals in the system. Hence, we expect they also differ in how they behave



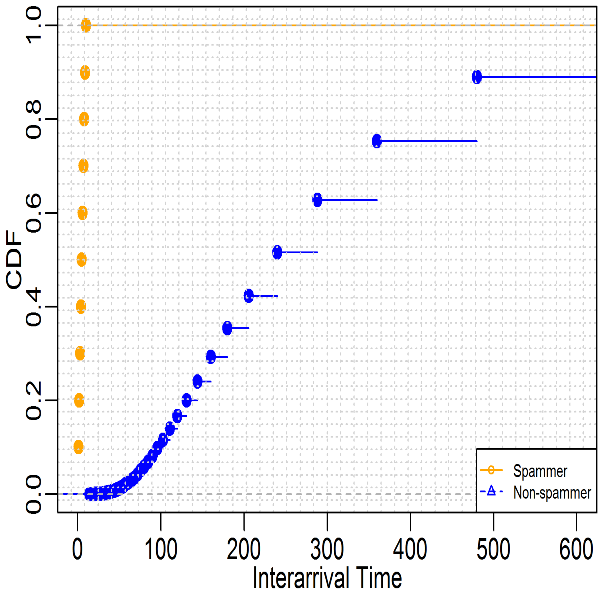
(a) non-spammer Degree Ratio (b) Spam Degree Ratio

Fig. 3: Representation of subscribers Friendship network of legitimate users and spammers from the call detailed records.

in the network. In this section, we analyzed the labelled dataset for the following features.

Friendship Network: We analyzed the friendship network of users for two categories i.e. spammers and non-spammers. Specifically, we analyzed it from the following perspectives: 1) whether the telemarketers or robocaller call a large number of subscribers, 2) what are the friendship characteristics of legitimate callers, and 3) Do spammers also receive calls from their callees? This analysis requires understanding the in-degree and out-degree of subscribers during the observed time period. For the analysis, we fixed the analysis window to 10 days, however, a smaller time window can also be used. A spammer is expected to target a large number of callees than normal callers. It is also expected that spammers receive a very low number of calls from other subscribers as well. In our characterization analysis, we observed the same patterns i.e. telemarketers called a large number of subscribers, and in return, only a few subscribers originated calls to them. Figure 3 shows the scatter plots of the out-degree to the in-degree ratio for the spam and legitimate callers. Figure 3 clearly shows that human subscribers normally have a balanced out-degree to in-degree ratio, whereas a spammer (telemarketers or robocallers) normally has unbalanced out-degree to in-degree ratios. This is because spammers typically target a large number of subscribers for the large footprint and outreach, whereas a legitimate caller does not change their callees too often over time.

Repetitive Index: The second feature we analyzed is the repetitive calling behaviour of subscribers. A typical legitimate subscriber is expected to develop a relationship network with a set of users which is consistent over time with minimal updates. On the contrary, spammers normally target new callees for their calls. Repetitive call behaviour represents the strength of connection among subscribers. In this context, we analyzed the repetitive calling behaviour of spammers and non-spammers. The ratio of the total number of calls to the out-degree of the subscriber is represented in Figure 5. Figure 5 A shows that a human caller typically has a repetitive calling behaviour, and the ratio of out-degree to the total number of calls is normally between 0.05 to 0.6. However, there still exist some subscribers in the legitimate caller list that have had a high non-repetitive call ratio but their out-degree and call duration is within the behaviour of legitimate callers. On the other hand, spammers generally do not exhibit repetitive calling behaviour and have a ratio between 0.7 to 1 which characterizes that a spammer has almost the same number of unique calls to the total number of calls i.e. *every call made by a spammer is to a new callee*. Our analysis shows that a small number of spam callers also have repetitive calling behaviour (perhaps an attempt to impersonate a legitimate caller or trying callee again) however they are unable to control their overall out-degree and call duration.

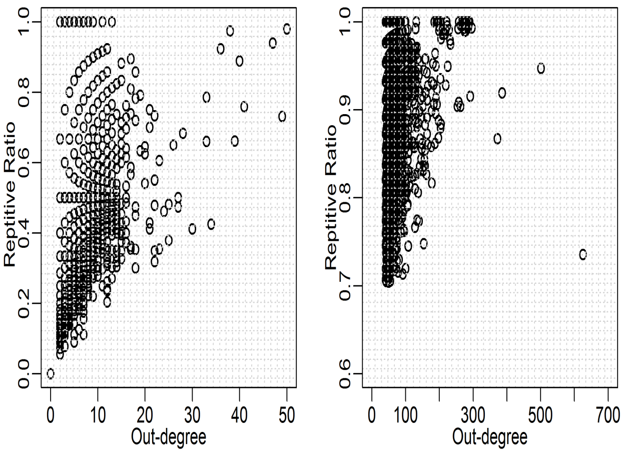


(a) Inter-arrival Time

Fig. 4: Inter-arrival time distribution of spammers and non-spammers

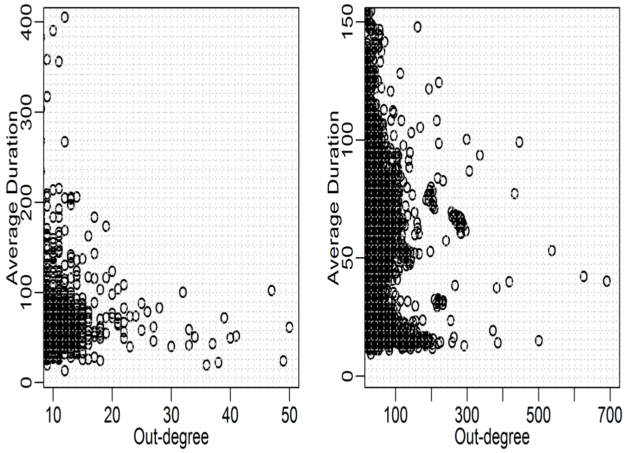
Inter-arrival Time: The inter-arrival time of calls from the subscriber represents the activity behaviour of the subscribers. The larger the time, the less active the user is. Figure 4 presents the cumulative distributions of the inter-arrival time of calls from the subscriber in each user class. It shows a clear distinction between spammers and non-spammers. The time-lapse between calls from the spammer is very much less than the time-lapse between calls from a legitimate subscriber. Since spammers have to do massive calling in order to have massive advertisements or marketing campaigns. Therefore they generate a large number of calls per hour. By contrast, the non-spammers normally call a few users during specific time periods (such as daytime or evenings), thus usually having a larger gap between calls to their callees.

Engagement: The last feature we analyzed is the engagement of subscribers during the observation time window which



(a) non-spammer Repetitive Behaviour (b) Spammmer Repetitive Behaviour

Fig. 5: Representation of subscribers repetitive calling behaviour with respect to out-degree from the Call detailed records.



(a) non-spammer Call Duration (b) spammer Call Duration

Fig. 6: Representation of subscribers Relationship network from Call detailed records using call duration and out-degree feature.

is computed through call duration. In this respect, spammers are envisaged to have high aggregate call duration as they target a very large number of subscribers [42] of which many are of short duration, however, they still manage good duration to a large number of callees. Therefore, using call duration alone as an identification feature is not expected to provide optimal detection accuracy. However, it is necessary to analyze the call duration feature along with the out-degree of the caller. Figure 6 represents the scatter plot of the average call duration with the out-degree of the caller. It is observed that the average call duration of spammers is around 40 seconds with only a few calls resulting in good duration, whereas the average call duration of the legitimate subscribers is around 100 seconds. Furthermore, some of the callers have an average duration of 20 seconds; this is because a caller has only made one call to a single callee during the analysed period. The higher-degree legitimate callers also have a small average call duration similar to the spammers but they exhibit some incoming calls as well as shown in Figure 3.A.

IV. Proposed FRAMEWORK

In this section, we define the problem and present the framework for classifying the caller as legitimate or malicious.

A. Problem Definition

In telecommunication networks, there are a set of n users $U = \{u_1, u_2, \ldots, u_k\}$. Each user u_i has a public identity i.e. the telephone number. The user provides this number to friends, family members or peers enabling them to reach out to him. The user developed social relationships and communities over a period of time. The telephone spam detection problem is to predict whether u_i has a behaviour that resembles a spammer or a legitimate caller, through applying a machine learning method ML to the set of features F extracted from call detailed records. Given a set of features of user u_i: $F = \{F_1, F_2, \ldots, f_m\}$, a machine learning method ML would predict whether u_i is a spammer or not. In this work, we estimated these features by modelling CDRs and computing a different set of reputation scores (Section IV-C).

B. Overall Framework

Figure 7 presents an overview of the proposed spam detection framework. First, a social call graph of the users is constructed from the streamed call detailed records. Second, the reputation score of the user is computed using the semantics of the call graph and user behaviour. Finally, the reputation scores are used along with the ML model for classification purposes.

C. Reputation System

Based on the observations learned in Section III-D, in this section, we present three ways to compute the reputation of the user in telecommunication networks. To this extent, we collectively use features such as in-degree, out-degree, call duration, and repetitive call behaviour. We have observed that legitimate subscribers normally have a stable call pattern e.g. having a good-duration call to a small number of unique callees, receiving calls from their callees as well and having a relationship with a small number of callees. On the other hand, spammers have dramatically different call behaviour. They receive a fewer number of calls from their callees and make calls to a large number of users who often span over a small duration as well. We used three features to define the reputation of the user in the network as described below.

The first feature is to estimate reputation is the repetitive index of the caller. It has been observed that spammers target new recipients for their calls, and hardly repeat the callee for the call. The repetitive index of the caller i can be computed

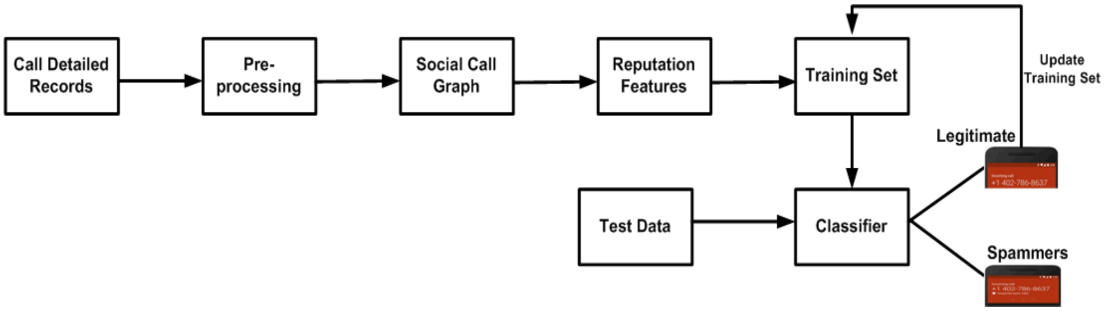
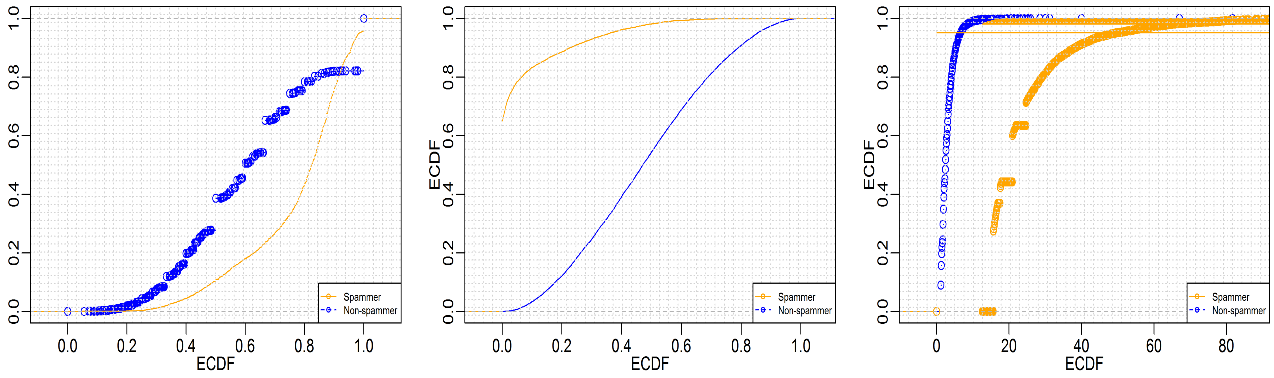


Fig. 7: Overall Framework of Reputation-based Approach



(a) Repetitive Behavior

(b) Duration Behavior

(c) Out-Degree Behavior

Fig. 8: Commutative Probability Distribution of Spammers and Non-spammers for the proposed Features.

as follows :

$$R_i = \frac{\text{Out-Degree of } u_i}{\sum Calls_i} \quad (2)$$

The equation 2 would result in a high repetitive index (near to 1) for the spammer and a small one for the non-spammer because of their repetitive call behaviour. Figure 8 A shows the CDF (Commulative Distribution Function) of the Repetitive index of spammers as well as non-spammers. Figure 2 clearly shows that a large number of spammers have non-repetitive behaviour and their value is close to 1. A small number of non-spammers also have non-repetitive behaviour. This is because some users are new to the system and have only interacted with a small number of users.

Another property that can differentiate spammers from non-spammers is the engagement of the users which can be computed from the user's call duration. Legitimate callers normally have good-duration incoming calls as well as good-duration outgoing calls. On the other hand, spammers do not have a large number of good-duration incoming and outgoing calls. Considering this fact about the behaviour of spammers and non-spammers, we define the reputation measures as the duration index of incoming calls. This can be computed by averaging the duration of incoming calls to the sum of the total duration of all calls. The engagement index E_i of user i is computed as:

$$E_i = \frac{\sum_{j=1}^{n} CD_{ju}}{\sum_{j=1}^{n} CD_{ju} + \sum_{j=1}^{n} CD_{uj}} \quad (3)$$

Where CD_{ji} is the sum of the duration of incoming calls to user i and CD_{ij} is the sum of the duration of the calls made by the user i to its callees. Spammers would always result in a small duration index, whereas the non-spammer would have a relatively good duration index reputation score because of the number of good-duration incoming calls. Figure 8 B shows the CDF of the duration index of spammers and non-spammers based on the call duration.

The third feature we used is the number of unique callees of the caller (out-degree of the caller) and the total number of calls. Besides, we used two variables α and β to give importance to certain features. The value of these variables can be between 0 and 1. The greater the value, the higher the importance given to the features. The Degree index of caller i is computed as follows:

$$D_i = \alpha \times (O_i) + \beta \times (UC_i) \quad (4)$$

Where O_i is the number of calls made by caller i and UC_i is the number of unique callees of user i. We suggest a small value for the α and a higher value for the β in order to give more importance to the number of unique callees of the caller than the total number of calls. Figure 8 C represents the degree index of spammers and non-spammers in the labelled data set.

Existing systems mainly used the average call duration as the measure for the reputation of the caller [10], [13], [43]. However, spammers could easily bypass this feature by having good duration calls by creating a set of Sybil identities. Similarly, the trust between a caller and a callee can be computed by getting the feedback from the callee, but again this can be circumvented by developing Sybil identities as well and also this solution requires changes in the handset to report the feedback. The features we reported in the paper are useful in two aspects: 1) the features are proposed based on the study of a large set of real call detailed records, and 2) the mentioned features have not been used before for blocking spammers in the network.

V. Detecting Spammers

We believe the challenge of identifying telephony spammers can be regarded as a classical classification problem. Therefore, we apply the supervised machine learning algorithm along with the designed features discussed in the previous section. The task is to classify the subscriber either as a spammer or a non-spammer. In this setup, each subscriber is represented as the vector of feature values along with the classification. The classification algorithm learns the model based on the pre-labelled data and then applies the model to classify new subscribers. Our goal is to evaluate the performance of proposed features for identifying spammers in a timely and effective way. Figure 9 presents the architecture of Neural Networks proposed for the classification task.

Neural networks (NNs) are a well-established domain within artificial intelligence and machine learning research. Neural networks are inspired by the human brain to identify patterns within complex datasets. NNs are primarily a clustering technique i.e. they are aimed at achieving segregation within a dataset based on inherent characteristics identifying groups of data items sharing similar characteristics. Neural networks primarily operate on numerical data and therefore real-life data such as images and textual data require to be transformed into a numerical form. As NNs are a supervised machine learning technique, an explicit training phase is required which enables a neural network to *learn* patterns within the dataset. This is followed by a *testing* phase during which the model created as a result of the training phase is envisaged to identify or classify patterns previously unseen by the model. Let us assume that we have an n-dimensional feature vector $X \in \mathbb{R}^{1\times n}$ and $\mathbf{W}_1 \in \mathbb{R}^{n\times n}$ is the weight matrix at the input layer of the NN. If $\mathbf{b}_1 \in \mathbb{R}^{1\times n}$ is the bias vector at the input layer and the output of the layer can be represented as follows:

$$z_1 = \mathbf{W}^{(1)}X + \mathbf{b}^{(1)} \tag{5}$$

Given that the ReLU activation is presented as

$$\text{ReLU}(z) = \begin{cases} z, & \text{if } z \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

The output z_1 when subjected to ReLU layer can be written as follows

$$\mathbf{h}^{(1)} = \arg\max(0, z_1) \tag{6}$$

The two hidden layers of the NN are similarly defined. The output layer takes the activations from the second hidden layer $\mathbf{h}^{(2)}$ and produces a scalar output y. The output is computed as follows:

$$y = \mathbf{W}^{(o)}\mathbf{h}^{(2)} + b^{(o)} \tag{7}$$

where $\mathbf{W}^{(o)}$ is the weight vector for the output layer, $b^{(o)}$ is the bias for the output layer. Overall, the neural network can be represented by the set of parameters $\Theta = \mathbf{W}^{(1)}, \mathbf{b}^{(1)}, ...\mathbf{W}^{(o)}, \mathbf{b}^{(o)}$. The output of the neural network is a function of the input X and the parameters Θ.

To train this network, we make use of a stochastic gradient descent algorithm, minimized using the mean squared error between the predicted and expected outputs.

$$\mathbf{L}(\mathbf{\Theta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{8}$$

The model parameters are recursively updated using a backpropagation algorithm where the update criteria are as follows:

$$\Theta_{t+1} = \Theta_t + \alpha\nabla_{\Theta}\mathbf{L}(\mathbf{\Theta_t}) \tag{9}$$

We used batch normalization, with the number of epochs set at 500. A feature vector was provided at the input of the network. It was fed forwarded through the densely connected layers. The mean squared error is computed at the output layer of the network that was back-propagated for adjustment of the weights. After performing 500 epochs, our network can learn the abstract representation of the features.

The classification experiments are performed using 10-fold cross-validation. In each test, the original sample is divided into 10 sub-samples, out of which nine are used as the training dataset, and the remaining one is used for testing the classifier. The process is then repeated 5 times, with each of the 10 sub-samples used exactly once as the test data, thus producing 10 results. The entire 10-fold cross-validation was repeated 5 times with different seeds used to shuffle the original data set, thus producing 50 different results for each test. The results reported are averaged over the 5 runs.

A. Evaluation Metric

The proposed approach is evaluated using the following four standard metrics, namely, precision, recall, the F-score and accuracy. The recall represents the true positive rate (TPR) of the system is the fraction of spammers classified as a spammer from the set of all spammers, and it is defined using Equation 10. The confusion matrix is presented in Table I. The TP represents the number of actual spammers classified as spammers, and FN (False Negative) represents the number of actual spammers misclassified as legitimate user. The Precision represents the ratio of correctly identified spammers to the total number of users identified as spammers and is computed using

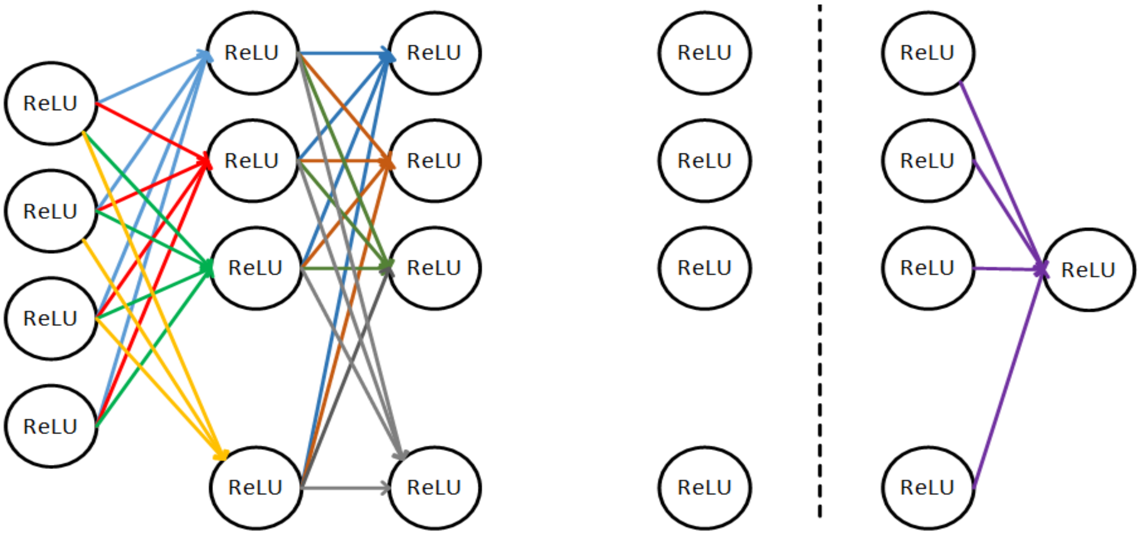
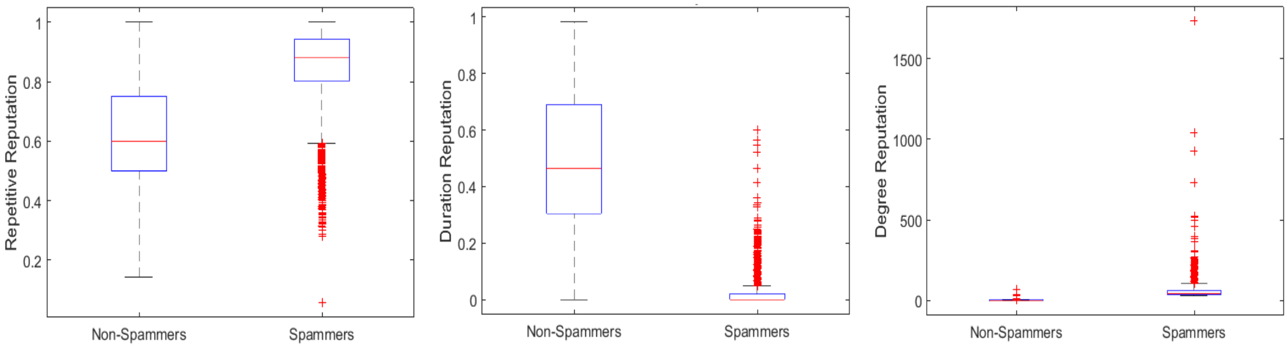


Fig. 9: Architecture of the proposed Neural Network used for the classification task.



(a) Repetitive Reputation (b) Duration Reputation (c) Out-Degree Reputation

Fig. 10: Box plots indicating the ranges of values for individual features.

Equation 11. The FP (False positive) in the equation is the number of legitimate users classified as spammers. Precision is the crucial measure for the evaluation of classifiers as it also considers the FP in computation. The F-Score is the harmonic mean of precision and recall as given in Equation 12. Finally, accuracy is the fraction of true classification i.e. spammer as spammer and non-spammer as non-spammer. The accuracy of the model is represented as in Equation 13.

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{12}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

VI. Classification Results

In this section, we evaluate the performance in different aspects.

		Predicted		
		Spammer	non-Spammer	Total
Actual	Spammer	TP	FN	$TP + FN$
	non-Spammer	FP	TN	$FP + TN$
	Total	$TP + FP$	$FN + TN$	N

Table I: Confusion Matrix.

A. Statistical Significance of Features

In this section, we statistically analyze the significance of the proposed reputation features. To perform this analysis, we have considered the strength of every feature individually towards its discriminative power in the identification of legitimate subscribers and spammers. This testing is being done using the paired t-test, and the p-values of the test are used to determine if the features are statistically different from each other for both the legitimate and the spammer classes. A visual analysis of the features is also presented in the form of box plots (Fig. 10). Based on our experiments, we have concluded that all the features that have been considered in this analysis are statistically significant with p-values significantly lower than 0.05, thus indicating that the mean difference between

Feature	TPR	FPR	Accuracy	Precision	Recall	F-Measure
F1	0.821	0.007	0.981	0.892	0.821	0.855
F2	0.872	0.006	0.986	0.910	0.872	0.890
F3	0.938	0.003	0.996	0.994	0.938	0.965
F4	0.967	0.001	0.997	0.992	0.967	0.979

Table II: Classification results using Neural Networks for the proposed features. F1- Repetitive Reputation, F2-Duration Reputation, F3-Degree Reputation, F4-All Combined

the features for the spammer and non-spammer classes is significantly different from zero. It can also be seen visually (Fig. 10) that the ranges of values for all the features including *Repetitive Calling Behavior*, *Call Duration* and *Degree* are appearing in very different ranges for both the non-spammers and the spammers. Given these experiments, it is safe to conclude that the feature set used in our study is very strong for effectively differentiating between the two considered classes without putting much load on the classification methods. This means that simple classification methods that work linearly in Euclidean spaces (such as nearest neighbours) should work reasonably well when these features are used for the proposed classification task using other machine learning methods.

B. Basic Classification Result

We evaluate the performance of our proposed features using the Neural Network classification method. There are four types of features used to represent the behaviour of the subscribers for classifying them as spammers and non-spammers. The performance results for our defined performance metrics and four features are shown in Table II. It is observed that feature F3 (Reputation based on the degree distribution) achieves the highest precision as compared to other features, however, it also misclassifies a relatively small number of legitimate subscribers as spammers as indicated by the FPR. Furthermore, it is also observed that features F1 (Reputation based on repetitive calling behaviour) and F2 (Reputation based on the call duration) allow a large number of spammers to call the subscribers and also it blocks legitimate callers at a slightly higher ratio. For these experiments, a small fraction of spammers were misclassified as non-spammers, because spammers aim to achieve a behaviour pattern similar to that of a legitimate caller to avoid detection. Although the misclassification of spammers as legitimate users is not expected to affect the revenue of the telecommunication operators, it can be a cause of unease for the callee and can affect the reputation of the operator.

From the perspective of the telecommunication operator, the misclassification of legitimate callers as the spammer is considered more damaging as this would not only bring a financial loss to the operator but would also make resources unavailable to legitimate users which can cause displeasure as well as reputation damage. From the results, it is clear that individual features do not have a high detection rate, so we analyzed the performance by considering all the reputation features together. This has not only decreased the false positive rate to an acceptable rate but has also shown effective detection of spammers as well. Additionally, we also report the performance results for another performance metric to better understand the behaviour of the proposed features.

C. Results on the Datasets With Varying Spammers and Legitimate Users Ratio

In this analysis, we analyzed the performance of using the proposed features under the condition that the number of spammers and non-spammers varies. We conducted experiments for the following scenarios: 1) fixing the number of spammers and varying the number of non-spammers, and 2) fixing the number of non-spammers and varying the number of spammers. We repeated experiments for the following different ratios of spammers and legitimate subscribers, i.e. 1:1, 1:2, 1:5, and 1:10 and presented the evaluation results in terms of TPR, FPR, and precision, for the classifier Neural Networks in Tables III and IV. The dataset with spammers and legitimate subscribers with the ratio 1:1 has 500 spammers and 500 legitimate subscribers, whereas the dataset with a ratio 1:2 has 500 spammers and 1000 legitimate subscribers, respectively. Similarly, for the second scenario, we fixed the number of legitimate subscribers to 200 and varied the number of spammers from 200 to 2000 accordingly. Table III represents the results for our proposed features set for a scenario when the number of legitimate users varies whereas the number of spammers is fixed at 500. Through our analysis, we observed that the dataset ratio shows a correlation with efficiency metrics when the number of legitimate subscribers in the dataset increases. Specifically, considering feature F4 in our analysis, the true positive rate slightly increases when the ratio of legitimate subscribers in the dataset increases. Further, the false-positive rate also decreases with the increase in the number of legitimate subscribers. These results also show that the proposed features, when used collectively with unbalanced distribution, would achieve the maximum true positive rate with very small false positives. Comparing the features individually, it is clear from Table III that feature F3 out-performs others in terms of both detection rate and false positive but it still has a small false-positive rate which can be effectively minimized by combining features together.

Table IV represents the results for the second scenario. The results show that the false positive rate increases with the increase in the number of spammers in the dataset. Similarly, the false-positive rate decreases with the increase of legitimate subscribers in the dataset as shown in Table III. The true positive rate in both scenarios stays the same even if the number of spammers or legitimate subscribers increases.

D. Comparisons With Other Methods

In this section, we compare our approach with other machine learning models. We performed experiments for 4 supervised machine learning models: namely, Support Vector Machines (SVM), Decision Tree (DT), K-nearest neighbours (KNN) and Naive Bayes (NB). Table V represents the comparisons of different machine learning models over proposed features.

- **KNN:** it takes an unlabeled object and labels it based on the majority of k nearest objects in the training set.

Ratio	F1			F2			F3			F4		
	TPR	FPR	Precision	TPR	FPR	Precision	TPR	FPR	Precision	TPR	FPR	Precision
1:1	0.892	0.315	0.739	0.968	0.084	0.920	0.980	0.020	0.980	0.988	0.016	0.984
1:2	0.735	0.210	0.637	0.960	0.047	0.911	0.970	0.015	0.970	0.988	0.011	0.978
1:5	0.250	0.000	1.000	0.946	0.024	0.889	0.978	0.005	0.976	0.988	0.001	0.994
1:10	0.200	0.000	1.000	0.922	0.015	0.864	0.974	0.002	0.978	0.994	0.001	0.994

Table III: Performance for the Fixed number of spammers and a varying number of legitimate subscribers using Neural Network.

Ratio	F1			F2			F3			F4		
	TPR	FPR	Precision	TPR	FPR	Precision	TPR	FPR	Precision	TPR	FPR	Precision
1:1	0.900	0.313	0.741	0.970	0.043	0.960	0.970	0.043	0.960	0.975	0.025	0.975
1:2	0.925	0.388	0.826	0.980	0.134	0.936	0.985	0.045	0.978	0.985	0.040	0.980
1:5	0.961	0.677	0.876	0.987	0.189	0.963	0.991	0.200	0.961	0.991	0.046	0.991
1:10	0.975	0.775	0.926	0.994	0.229	0.977	0.994	0.200	0.980	0.994	0.174	0.983

Table IV: Performance for the Fixed number of legitimate subscribers and a varying number of spammers using Neural Network.

Ratio	SVM			Naive Bayes			Random Forest			KNN		
	TPR	Precision	Recall	TPR	Precision	Recall	TPR	Precision	Recall	TPR	Precision	Recall
F1	0.703	0.857	0.703	0.791	0.748	0.791	0.739	0.857	0.739	0.791	0.956	0.791
F2	0.891	0.891	0.891	0.858	0.787	0.858	0.828	0.832	0.828	0.828	0.831	0.828
F3	0.763	0.929	0.763	0.926	0.908	0.926	0.885	0.944	0.885	0.907	0.996	0.907
F4	0.946	0.964	0.946	0.953	0.965	0.953	0.950	0.964	0.950	0.981	0.992	0.981

Table V: Performance of Different Machine Learning algorithm with respect to our feature set.

A neighbour is deemed close if it has a small distance based on a distance metric, most commonly the Euclidean distance.

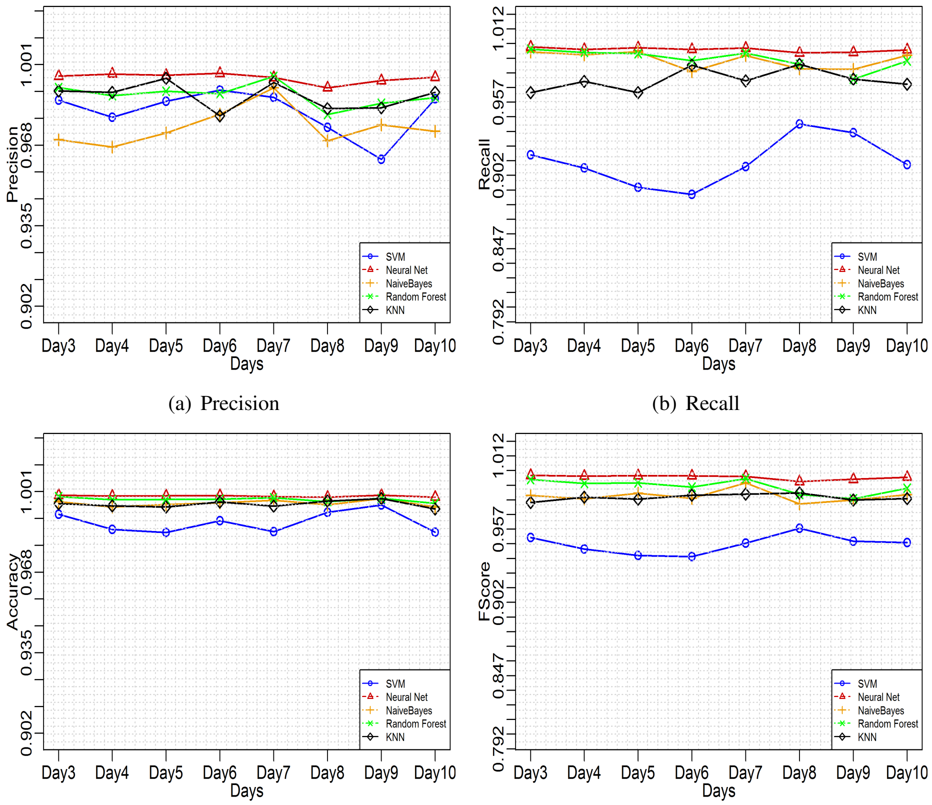
- **NB:** It is a simple probabilistic classifier after applying Bayes' theorem using strong independence assumptions. This classifier works on the probability of a class based on the number of instances that occur in that class.
- **DT:** It makes use of trees by forming a set of rules to figure out the label of a given input. The output of this classifier is a tree that contains the rules to predict the target output variable.
- **SVM:** It is based on the concept of a hyperplane that maximizes the margin of separation between the two classes. A kernel function is used to transform the data into higher dimensions to increase the odds of finding the relevant hyperplane.

This is a nice blend of different machine learning algorithms belonging to conceptually different approaches towards addressing the classification problem including the probabilistic method (NB), tree-based classifiers (DT), simple Euclidean space method (KNN) and a method based on the decision planes (SVM) with a linear kernel. Our choice is motivated by the fact that our main objective is to assess the strength of the feature sets rather than carrying out a comparison of the classifier which is not the main focus of this paper. Consequently, using simpler classifiers such as KNN will help us in quantifying the strength of the features, as a linearly discriminant feature set will not require a very strong machine-learning framework to achieve better classification results.

Our experiments show that when using F1 and F4, KNN outperforms the other methods that have been considered in this paper. The overall performance is relatively low when using F1 in any machine learning method, which indicates that it is not a very strong feature set that can be used reliably for the identification of spammers. The combination of all features together turns out to be a strong feature set, and even the usage of the most simple algorithm (KNN) yields very good results for the subject task. The observations are consistent for all the metrics that have been considered in this analysis. In the case of F4, irrespective of the classifier used, very good performance results are obtained which indicates the relative strength of F4 as compared to the other feature sets. In general, the experiments show that the feature sets proposed in this paper can be reliably used to identify spammers with a high detection accuracy and small false positives.

E. Performance over Time

In this section, we evaluate the performance of the proposed features concerning all machine learning methods using the training and testing data set from two different periods. The number of non-spammers on each day is fixed at 12000 on a respective day, whereas the number of spammers vary from 500 to 2000 depending on the number of spammers found in the data on a respective day. Figure 11 shows the precision, recall, detection accuracy and F-score for different machine learning methods for the feature F4, and when data from Day 1 and Day 2, is used from the training set, and data from Day 3 to Day 10 is used for the testing set. We can see from Figure 11.C that except SVM, other machine learning methods provide high detection accuracy which slightly changes with time. We can see that SVM also has a small recall ratio as compared to other machine learning methods. Specifically, the NN method along with feature F4 would achieve higher detection accuracy as well as a higher precision ratio. Figure 11.D shows the F-score using the F4 features. We can see that the F-score of all approaches slightly increases or decreases with the day. For the Random Forest, Naive base, and KNN



(a) Precision

(b) Recall

(c) Accuracy

(d) F-Score

Fig. 11: Features Performance over time.

Model, it slightly decreases over time, however, it remains stable for the Neural network models.

In summary, the results indicate that all machine learning approaches besides SVM achieve acceptable precision, recall, accuracy, and F1 score. Furthermore, neural networks achieve stable true positive and false positive rates over time.

VII. Conclusion

This paper investigates a spam detection framework for telecommunication users, based on the social connection of users along with selected machine learning techniques to classify subscribers as spammers or non-spammers. On the developed social call graph of subscribers, several reputation features are proposed to evaluate and rank the trustworthiness of subscribers and classify them using supervised machine learning methods. The proposed reputation features perform well when combined even in the presence of a large number of legitimate subscribers and a small percentage of spammers. The framework also makes it difficult for spammers to bypass the system, as modifying a number of features to bypass the system incurs a significant cost to the spammers. Currently, we have computed reputation based on three features. A possible extension is to explore additional features such as the inter-arrival time between call requests, the uniqueness in the numbering patterns of the callees, and attack groups over time. Another extension for future research is to analyze the performance of the system when classification and clustering are combined together over an unseen dataset.

References

[1] Number Of Global Mobile Subscribers To Surpass Five Billion This Year, Finds New GSMA Study. [Online]. Available: https://goo.gl/XCgA8G

[2] Spam Phone Calls Cost U.S. Small Businesses Half-Billion Dollars in Lost Productivity, Marchex Study Finds. [Online]. Available: http://goo.gl/jTrgp3

[3] U.S. Phones Were Hit by More Than 50 Billion Robocalls in 2021. [Online]. Available: https://shorturl.at/lyRY4

[4] C. K. JENNIFER. (2017) FTC Releases Updated Do Not Call Registry Data Book; Impersonator Fraud Tops List of Consumer Complaints. [Online]. Available: https://shorturl.at/osAP1

[5] (2017) Truecaller Insights 2022 U.S. Spam & Scam Report. [Online]. Available: https://shorturl.at/cimtH

[6] M. Hansen, M. Hansen, J. Möller, T. Rohwer, C. Tolkmit, and H. Waack, "Developing a Legally Compliant Reachability Management System as a Countermeasure against SPIT," in *Proceedings of 3rd Annual VoIP Security Workshop*, 2006.

[7] D. Shin, J. Ahn, and C. Shim, "Progressive Multi Gray-Leveling: a Voice Spam Protection Algorithm," in *IEEE Network*, 2006, pp. 18–24.

[8] M. A. Azad, S. Bag, C. Perera, M. Barhamgi, and F. Hao, "Authentic caller: Self-enforcing authentication in a next-generation network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3606–3615, 2020.

[9] M. A. Azad and R. Morla, "Multistage spit detection in transit voip," in *SoftCOM 2011, 19th International Conference on Software, Telecommunications and Computer Networks*, 2011, pp. 1–9.

[10] V. Balasubramaniyan, M. Ahamad, and H. Park, "CallRank: Combating SPIT Using Call Duration, Social Networks and Global Reputation," in *Proceedings of Fourth CEAS2007.*, 2007.

[11] P. Kolan and R. Dantu, "Socio-Technical Defense Against Voice Spamming," *ACM Trans. Auton. Adapt. Syst.*, vol. 2, no. 1, 2007.

[12] R. Dantu and P. Kolan, "Detecting Spam in VoIP Networks," in *Proceedings of the Steps to Reducing Unwanted Traffic on the Internet ,Berkeley, CA, USA*. USENIX, 2005, pp. 31–37.

[13] M. A. Azad and R. Morla, "Caller-Rep: Detecting unwanted calls with caller social strength," *Computers & Security*, vol. 39, Part B, pp. 219–236, 2013.

[14] R. Schlegel, S. Niccolini, S. Tartarelli, and M. Brunner, "SPam over Internet Telephony (SPIT) Prevention Framework," in *Proceedings of IEEE GLOBECOM '06*, 2006.

[15] D. Gritzalis and Y. Mallios, "A sip-oriented {SPIT} management framework," *Computers & Security*, vol. 27, pp. 136 – 153, 2008.

[16] M. A. Azad and S. Bag, "Decentralized privacy-aware collaborative filtering of smart spammers in a telecommunication network," in *Proceedings of the Symposium on Applied Computing*, ser. SAC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 17111717. [Online]. Available: https://doi.org/10.1145/3019612.3019792

[17] M. Ajmal, S. Bag, S. Tabassum, and F. Hao, "privy: Privacy preserving collaboration across multiple service providers to combat telecoms spam," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, 2018.

[18] W. Henecka and M. Roughan, "Privacy-preserving fraud detection across multiple phone record databases," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 6, pp. 640–651, Nov 2015.

[19] H. Bokharaei, A. Sahraei, Y. Ganjali, R. Keralapura, and A. Nucci, "You can SPIT, but You can't hide: Spammer Identification in Telephony Networks," in *Proceedings of 2011 IEEE INFOCOM*, 2011, pp. 41–45.

[20] P.-A. Chirita, J. Diederich, and W. Nejdl, "MailRank: Using Ranking for Spam Detections," in *Proceedings of 14th ACM international conference on Information and knowledge management*, ser. CIKM '05, 2005, pp. 373–380.

[21] H. Lam and D. Yeung, "A Learning Approach to Spam Detection Based on Social Networks," in *Proceedings of Fourth Conference on Email and Anti-Spam (CEAS2007)*, 2007.

[22] P. O. Boykin and V. P. Roychowdhury, "Leveraging Social Networks to Fight Spam," in *IEEE Computer*, 2005, no. 38, pp. 61–68.

[23] H. Sengar, X. Wang, and A. Nichols, "Call Behavioral Analysis to Thwart SPIT Attacks on VoIP Networks," in *Security and Privacy in Communication Networks*, vol. 96, 2012, pp. 501–510.

[24] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad, "Phoneypot: Data-driven Understanding of Telephony Threats," in *Proceedings of 20th NDSS*, 2015.

[25] M. Balduzzi, P. Gupta, L. Gu, Gao.D, and M. Ahamad, "MobiPot: Understanding Mobile Telephony Threats with Honeycards," in *Proceedings of 11th ACM ASIACCS*, 2016.

[26] H. Li, X. Xu, C. Liu, T. Ren, K. Wu, X. Cao, W. Zhang, Y. Yu, and D. Song, "A machine learning approach to prevent malicious calls over telephony networks," in *Proceedings of 2018 IEEE Symposium on Security and Privacy (SP)*, May 2018, pp. 53–69.

[27] J. Liu, B. Rahbarinia, R. Perdisci, H. Du, and L. Su, "Augmenting telephone spam blacklists by mining large cdr datasets," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ser. ASIACCS '18, 2018, pp. 273–284.

[28] H. Tu, A. Doup?, Z. Zhao, and G. J. Ahn, "Sok: Everyone hates robocalls: A survey of techniques against telephone spam," in *2016 IEEE Symposium on Security and Privacy (SP)*, May 2016, pp. 320–338.

[29] M. A. Azad, R. Morla, and K. Salah, "Systems and methods for spit detection in voip: Survey and future directions," *Computers & Security*, vol. 77, pp. 1 – 20, 2018.

[30] S. Iranmanesh, H. Sengar, and H. Wang, "A Voice Spam Filter to Clean Subscribers? Mailbox," in *Proceedings of Security and Privacy in Communication Networks*. Springer, 2013, pp. 349–367.

[31] J. Strobl, B. Mainka, G. Grutzek, and H. Knospe, "An Efficient Search Method for the Content-based Identification of Telephone-SPAM," in *Proceedings of 2012 IEEE ICC*, 2012, pp. 2623–2627.

[32] D. Lentzen, G. Grutzek, H. Knospe, and C. Porschmann, "Content-Based Detection and Prevention of Spam over IP Telephony - System Design, Prototype and First Results," in *Proceedings of 2011 IEEE ICC*, 2011, pp. 1–5.

[33] V. A. Balasubramaniyan, A. Poonawalla, M. Ahamad, M. T. Hunter, and P. Traynor, "Pindr0p: Using single-ended audio features to determine call provenance," in *Proceedings of the 17th ACM Conference on Computer and Communications Security*, ser. CCS '10, 2010, pp. 109–120.

[34] P. Kolan and R. Dantu, "Socio-Technical Defense Against Voice Spamming," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 2, 2007.

[35] K. Ono and H. Schulzrinne, "Have I met you before?: using Cross-Media Relations to Reduce SPIT," in *Proceedings of 3rd IPTCOMM*, 2009, pp. 1–7.

[36] R. Zhang and A. Gurtov, "Collaborative Reputation-based Voice Spam Filtering," in *Proceedings of DEXA 09.*, 2009, pp. 33–37.

[37] Y.-S. Wu, S. Bagchi, N. Singh, and R. Wita, "Spam Detection in Voice-Over-IP Calls through Semi-Supervised Clustering," in *Proceedings of 39th Annual IEEE/IFIP DSN, Portugal*, 2009, pp. 307–316.

[38] M. A. Azad, R. Morla, J. Arshad, and K. Salah, "Clustering voip caller for spit identification," *Security and Communication Networks*, vol. 9, no. 18, pp. 4827–4838, 2016.

[39] S. Pandit, R. Perdisci, M. Ahamad, and P. Gupta, "Towards measuring the effectiveness of telephony blacklists," in *Proceedings of 23rd NDSS*, 2015.

[40] S. Chiappetta, C. Mazzariello, R. Presta, and S. Romano, "An anomaly-based approach to the analysis of the social behavior of voip users," *Computer Networks*, vol. 57, no. 6, pp. 1545 – 1559, 2013.

[41] M. A. Azad and R. Morla, "Rapid detection of spammers through collaborative information sharing across multiple service providers," *Future Generation Computer Systems*, 2018.

[42] S. Chiappetta, C. Mazzariello, R. Presta, and S. P. Romano, "An anomaly-based approach to the analysis of the social behavior of voip users," *Computer Networks*, vol. 57, no. 6, pp. 1545–1559, Apr. 2013.

[43] N. Chaisamran, T. Okuda, G. Blanc, and S. Yamaguchi, "Trust-Based VoIP Spam Detection Based on Call Duration and Human Relationships," *Proceedings of IEEE/IPSJ International Symposium on Applications and the Internet,*, vol. 0, pp. 451–456, 2011.