Multimed Tools Appl (2017) 76:6079–6096
DOI 10.1007/s11042-016-3555-3

A multi-feature approach to detect Stegobot: a covert multimedia social network botnet

Natarajan Venkatachalam[1] · R. Anitha[1]

Received: 28 February 2015 / Revised: 27 February 2016 / Accepted: 26 April 2016 /
Published online: 23 May 2016

Abstract Online Multimedia Social Networks(OSNs) are popular and efficient medium for millions of users. Unfortunately, in wrong hands, they are also effective medium for spreading social malware and propagation of social botnet. A newly proposed multimedia social network threat, Stegobot masks crucial information in a digital image by using a technique known as steganography. Stegobot works by first infecting a computer and then communicates the stolen information, which could be login passwords, bank account details or credit card numbers. Also it efficiently utilizes the advantage of image steganography to hide the presence of communication within the image sharing behavior of OSNs. Since these social bots exhibit unobservable communication channels, existing botnet detection mechanisms cannot be applied to such botnets. In this paper, we present a novel host based method for detecting and differentiating Stegobot profiles. Also the proposed method shows the ability to detect Stegobot network traffic which is inherently different from legitimate multimedia social network traffic. The best performance of our detection system is demonstrated on different social networks data set with different evaluation metrics. Multiple aspects of multimedia attributes proposed in this study help to explore the hidden communication structure of botnet. Stegobot profiles mimic genuine users and compromise other vulnerable users in social network. By using single view features alone it is very difficult to detect bot profiles as well as Stegobot communications and hence in this work a multi-feature approach is considered. Also, this work attempts to help network security experts and forensic analysts to understand the Stegobot communication and the key profiles inside the malicious network.

Keywords Multimedia Social Network · Botnet · Stegobot traffic · Image steganography · Malicious profile · Botnet detection

✉ Natarajan Venkatachalam
Natarajan.V.IN@ieee.org

[1] Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Coimbatore, 641004, India

1 Introduction

In recent years, Online Multimedia Social Networking (OMSN) sites facilitate a high degree of user personalization and user intercommunication [4]. The enormous growth of OMSN has also resulted in an increase in their use for significant criminal activities including identity theft, piracy, illegal trading, cyber stalking and cyber terrorism [34]. Cyber criminals are becoming increasingly sophisticated in attempting to use social networking based technology in order to evade detection and perform criminal acts. This happens in virtual environment using social network as a communication medium and it gives attackers to increase the chance of attacking systems [1, 7, 15, 36]. In addition, combating this growing level of crime is challenging due to the ever increasing scale of today's OMSN.

Due to emerging social network activities, malware trends are now shifting towards new direction. The challenge faced by stealthy malcode is to reach and stay on the vulnerable hosts for a longer period. The longer a threat remains undiscovered in the wild, the more opportunity it has to compromise systems before measures can be taken to protect against it. Further more, its ability to steal information increases the longer it remains undetected on a compromised computer [22].

Recently Nagaraja et al. [20] proposed a new type of stealthy multimedia social network threat called Stegobot. Stegobot was created to show how easy it would be for an intruder to hijack Facebook photos to create a secret communication channel that is very difficult to detect. Stegobot gains control of computers by making users to open or download malicious images. Instead of contacting the botmaster directly, it takes the advantage of social network activity to communicate with the botmaster.

Social network based security has drawn lots of research interest and various systems have been developed with the support of machine learning techniques. Recently proposed techniques [2, 26, 32] aim to detect spamming in online social networks. Cao et al. [5] have proposed a method to detect fake accounts created in large scale online social networks. Stein et al. [31] have proposed the Facebook Immune System (FIS) to protect the users. They have built an adversarial learning system that performs real time detection of malicious activities from the regular activity on Facebook's database. This system contains historical data related to Facebook user's malicious activities. But it is very difficult to identify and block the socialbots. Socialbots are designed to appear as a normal Facebook user. Viswanath et al. [33] have proposed graph theoretic techniques for defensing against sybil attacks, which is an alternative to machine learning based detection learning systems. But these methods are not efficient enough to detect Stegobots. This reveals a new security challenge in the domain of multimedia social networks.

In our previous work [23], we proposed a multilevel detection mechanism that analyzes user profiles and finds out the malicious ones. The idea behind this method is that each profile content with huge volumes of image data is to be analyzed independently. It is an extremely difficult task to analyze a huge amount of image data. Also the real time deployment of this technique is a challenging one. The major challenge with most of the existing techniques is the detection of socialbots using single view features. Since Stegobot profiles in OMSN look like genuine profile, these approaches cannot detect new kind of bots [20]. Due to these difficulties, it would be desirable to develop additional methodologies to overcome these issues. In this paper, we present a new technique for analysis and detection of Stegobot network traffic. Multiple aspects of social attributes proposed in this paper help to explore the hidden communication structure of botnet. Further, we enhance our method towards OMSN security that stands against Stegobot profiles which mimic genuine users.

Also, this work attempts to help network detectives and forensic analysts to understand the structure of Stegobot and the key profiles inside the malicious network.

The rest of this paper is organized as follows: Section 2 describes Stegobot network along with threat model analysis. Section 3 introduces our proposed method for social botnet detection. The experimental study, observations and performance are presented in Section 4 and Section 5 concludes the paper.

2 Analysis of Stegobot network

This section briefs a conceptual overview of Stegobot Network (SbN), and gives a brief outline of the adversarial objectives behind maintaining such a network. This is followed by a short note about the SbN design goals and its construction details.

2.1 Design goals of Stegobot

Stegobot Network is a set of bots that are owned and maintained by a human controller called botherder. A SbN consists of three components: the botmaster, Stegobot and the command and control channel. Each Stegobot controls a profile in a targeted OMSN, and is capable of executing commands that results in operations related to social interactions. These commands are either sent by the botmaster or predefined locally on each Stegobot. The data stolen by the Stegobots are called botcargo and are sent back to the botmaster through the covert channel.

Botnet command and control channels have traditionally been carried over protocols such as IRC (Internet Relay Chat) or the various P2P (Peer To Peer) networks. The ability to coordinate and upload new commands to bots gives the botmaster high power when performing criminal activities like sending spam, perform Distributed Denial of Service (DDoS) attacks and phishing etc.

A peculiar property of Stegobot [3] is the design of the communication channel between the bots and the botmaster. The goal of Stegobot is to introduce probabilistically unobservable communication channels connecting the bots and botmaster. If the command and control communication is unobservable, then botnet detection can be more difficult than the detection of normal botnets [8, 17, 25].

Figure 1 shows a conceptual model of Stegobot. Each node in the OMSN represents a profile. The Stegobots are marked in black. Infiltrated profiles are marked in gray. Edges between nodes represent social connections. The dashed arrow represents connection requests. The small arrows represent social interactions [13, 21].

Stegobot is designed to infect users connected to each other via social links such as an email communication or an online social network that allows friends to share images. The propagation of bot binaries take place via social malware attacks [22]. These bots contain pre-programs to perform malicious activities such as harvesting email addresses, stealing passwords, credit card numbers and keylogging. Alternatively, in a more flexible design, the bots execute commands received from the botmaster. In Stegobot, the images shared by social network users are utilized as a medium for building up the command and control channel. Specifically, image steganography is used to setup a communication channel within the social network, which serves as botnet's command and control channel. The information exchanged between bots must be transferred only using steganographic channels.

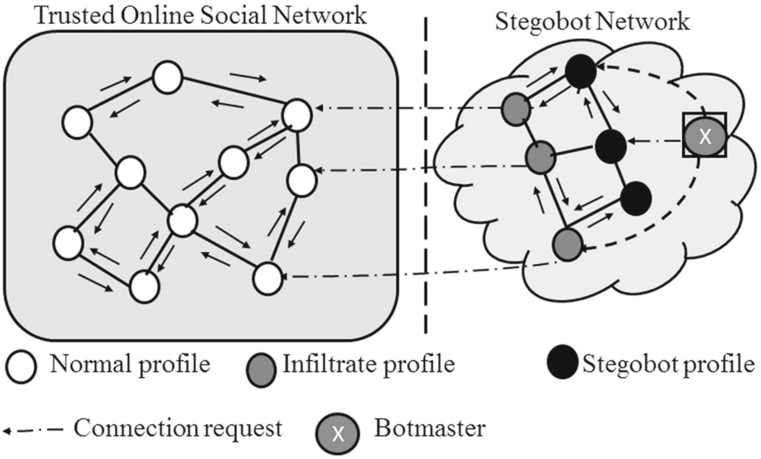

Fig. 1 Graphical representation of Stegobot network in OMSN

2.2 Behavior analysis

The behavioral characteristics [2, 11] of profiles are analyzed in order to distinguish normal profile activities and malicious activities. Botmaster can obtain their malicious profiles in any one of the two ways: compromising existing genuine profiles and creating fake profiles. In the first case, the botmaster takes control of a legitimate profile following attacks via social links such as online multimedia social network or an e-mail communication that allows users to exchange email [22]. This technique is attractive for an attacker, because legitimate user already has a significant number of friends. Also these profiles are trusted or at least known to their friends, social malware attacks from these accounts are more likely to succeed. Additionally, profile user may lose control over these accounts at any time after successful infection of Stegobot. Alternatively, botmaster may create a new fake account, in the sense that they do not represent a real user in OMSN. Despite the use of mechanisms like user account registration, authentication and CAPTCHAs(Completely Automated Public Turing test to tell Computers and Humans Apart) still it is easy to automate and botmaster can efficiently create large number of fake profiles.

It is very important to analyze the Stegobot profile covert communication patterns and its different characteristics of compromised profiles. Legitimate profiles and Stegobot profiles have different goals in the OMSN system and also differs in behavior to achieve their purposes. It is important to analyze a large set of attributes that reflect profile behavior and investigating their relative discriminatory power to distinguish between malicious and normal profiles. In this paper, we considered three set of features, namely image features, profile features and social network features for the detection of Stegobot.

3 Stegobot detection

The goal of the host based detection method is to detect the presence of Stegobot command and control (C&C) traffic on the monitored host. A functional block diagram of the proposed method is shown in Fig. 2. To facilitate the Stegobot detection as well as the detection

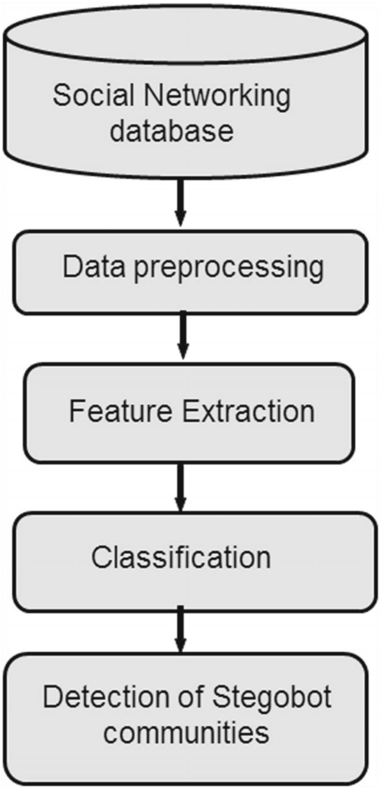

Fig. 2 Functional block diagram of the proposed Stegobot detection

of stego communication between two profiles, we use social network features, graph based features and image content features. Based on the Stegobot policy on OMSN, various graph based features are extracted from users social graph and most recent activities. Traditional classification algorithms are applied to detect suspicious behaviors of Stegobot.

3.1 Preprocessing

To build a Stegobot detection framework, we have to create a standard social profile object model within the social network. The object model of a profile is a schema containing most common features of social network user account. A web crawler using different social network APIs are utilized to collect real data set from publicly available information in OMSN user account. These techniques may have challenges due to their large volume of data, although they significantly improve the time taken to classify a profile object as infected profile or legitimate profile.

3.2 Feature extraction

The attributes consist of individual characteristics of user profiles and behaviors. Both genuine and bot profiles have certain kind of patterns. For example, genuine users have many legitimate friends, while bots are fake profiles and never reply to comments. The bot profiles may have many attractive images and celebrity photos and attract other users to accept itself as a friend by initiating friend request or image sharing. The selection of features for detection of Stegobot is based on the following assumptions: Any infected profile has equal probability to send request or to infect any susceptible profile in a social network community. The vulnerable profiles are infected when it has friendship with or follow the bot infected profile. It is suitable to investigate the various features like vulnerable image contents, social graph and profile based features. The extracted features are based on basic social network, steganography and graph theory concepts and their underlying design principles.

3.2.1 Image features

Stegobot transmits the malicious images and stego images with confidential information to botmaster using the image sharing behavior of multimedia social networks. The Stegobot communication channels are built by leveraging image steganography and the social image sharing behavior of users. In order to achieve better detection accuracy, the attributes are derived from social network based features and image features in a combined manner.

Image attributes capture specific properties of the image uploaded by the user. Each profile has a set of images in the user account, each image with its features that serve as a main parameter for detection of malicious images. In particular, each image is characterized by its stegnographic features along with its size, quality, number of views and number of likes, number of times the image is selected as a favorite. The Stegobot communication channels are built by leveraging image steganography and the social image sharing behavior of users. In order to achieve better detection accuracy, the attributes were derived from social network based features and image features in a combined manner.

For each profile we select images having maximum size, high quality, maximum number of likes and favorite for many users and extract the features based on the steganalysis point of view.

For each profile and its associated images in the data set, we extract the features based on the steganalysis point of view. The choice of the image steganography based features for Stegobot identification is determined based on our previous positive experience in the domain of image steganalysis. This detection can be extended in a straightforward manner to color images by considering the color image as three gray scale images and fusing the results from each channel. All the 24 content features utilized in our previous work [23] are considered as image based features in this work.

Fridrich et al. [10] have investigated the steganalysis problem using discrete cosine transform based features. Natarajan et al. [24] later proposed different set of features based on the contourlet transform and subband coefficient modeling using Gaussian distribution. These methods are used towards identifying and extracting secret message from stego images embed by the well known image steganographic algorithms. In this work, we focus on detection of malicious content inside the image and bot binaries within the image. With this objective, each image is characterized using a total of 24 features calculated directly from the carries object. The detailed process of calibration used to construct image content based features are reported in [10, 17, 25]. These steganalysis based features are efficiently utilized for detection of Stegobot communication through images. The image based feature set

$$F^{(I)} = < lm_k, hm_k, SSIM, H, H_{R_{cd}}, H_{C_{cd}}, D_{Row_{KL}}, D_{Col_{KL}}, I_{row}, I_{col}, H_{\_Row}, H_{\_Col}, dc_l, dc_h, dc_m, wc_l, wc_h, wc_m > \tag{1}$$

where lm_k and hm_k (k=1,2,3,4) are the moments of low frequency and high frequency contourlet subbands of the image respectively. $SSIM$ is the mean structural similarity of the whole image. H denotes the image entropy, $H_{R_{cd}}$, $H_{C_{cd}}$ are row wise and column wise conditional image entropy. $D_{Row_{KL}}$ and $D_{Col_{KL}}$ are row wise and column wise K-L divergence respectively. Row wise mutual image entropy and column wise mutual image entropy are denoted by I_{row}, I_{col}. $H_{\_Row}$, $H_{\_Col}$ are row wise and column wise Rényi entropy. Low, high and medium DCT frequency subbands features dc_l, dc_h and dc_m. wc_l, wc_h and wc_m are low, high and medium wavelet based frequency subbands.

3.2.2 User profile features

The profile based analysis gives individual characteristics of profile behavior. Malicious profile users spend more time doing activities such as sending abnormal requests to random profiles, adding more images as favorite and abnormal sharing of images to others. Liben-Nowell et al. [18] proposed link prediction method using machine learning techniques to predict links between users in different online social networks. Stringhini et al. [32] proposed a technique for spammer profiles detection by using supervised learning algorithms. Altshuler et al. [27] have investigated different users properties, such as origin and ethnicity, inside the social relationship, which helps to predict malicious user in OSN. Recently, Fire et al. [9] used the online social networks topological features to identify fake users in different online social networks. As part of this research we proposed a method for investigating the social network users which of their friends might be a malicious profile. Our proposed features are based on the various Stegobot characteristics and its connection properties between OSN users. This type of problem is to some degree similar to the problem of investigating Stegobot profile in multilevel analysis, studied by Natarajan [23].

In this study, we also used different type of graph relationship between users, similar to the study of Buscarino et al. [4] and Mislove et al. [33]. In addition, our study contrasts previous studies [37], because we construct our feature set by using only variations of the data collected in real-time from the Stegobot attacker activity point of view. Hence the following six features used in our previous work [23] are included based on the profile behavior analysis.

- Following/Follower Ratio(R).
- URL ratio(U).
- Message Similarity(M).
- Ratio of Trusted Friends(RF).
- Message/Image Shares(S)
- Female/Male Friend Number(N).

The social network profile based feature vector is denoted by

$$F^{(S)} =< R, U, M, RF, S, N > \tag{2}$$

3.2.3 Social graph based features

We have utilized image features and user profile features alone to detect stegobot. But that cannot identify Stegobot communications. Hence social graph features focusing on interaction between profiles are combined to form the feature set. The social network attributes capture the social graph based relationships established between profiles via image response interaction, which is one of the several possibilities in online multimedia social network [4, 19]. The idea is that these features might capture specific interaction pattern that could help to differentiate genuine profiles and bot profiles. The following profile features are extracted from the image response user graph, which captures the level of interaction of corresponding profiles. The features include cluster coefficient, betweenness, reciprocity, assortativity and page rank.

Cluster coefficient(transitivity) The clustering coefficient of social graph measures the degree of interconnection of a network. In other words, it measures the tendency of two nodes that are not adjacent but share an acquaintance, to get themselves in contact. High clustering coefficients mean the presence of high number of triangles in the network. The clustering coefficient $c(i)$ for a node v_i is the number of directed links divided by the number of possible directed links that could exist between the nodes neighbors. If a node v_i's neighbors have n directed links between them, then the clustering coefficient

$$c(i) = \frac{n}{d_i(d_i - 1)} \tag{3}$$

where d_i is the number of links the node v_i has to other nodes.

It is well known in literature [11], that social network shows high values of clustering coefficient since they reflect the underlying social structure of contact among friends. It provides the possibility of computing both global clustering coefficient for any social network and the local clustering coefficient of any given node.

Betweenness Betweenness is a measure of the nodes centrality in the graph , that is nodes appearing in the large number of shortest paths between any two nodes have higher betweenness than others. The betweenness centrality of a node v can be defined as

$$B_C(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4}$$

where σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through a node v.

Modularity When examining communities in networks, we require an objective metric to evaluate how good a particular network into communities. Modularity is one measure of the structure of networks or graphs. It is used to measure the strength of division of network into modules. It is defined as

Q= Number of edges within communities - expected number of such edges

Reciprocity A traditional way to define the reciprocity R_r is using the ratio of the number of links pointing in both directions $L^{<->}$ to the total number of links L.

$$R_r = \frac{L^{<->}}{L} \tag{5}$$

With this definition, $R_r = 1$ for a purely bidirectional network and $R_r = 0$ for a purely unidirectional one. Real networks have an intermediate value between 0 and 1.

Assortativity It is a measure of the likelihood for nodes to connect to other nodes with similar degrees. The assortativity A_r is defined as the Pearson correlation coefficient between the degrees of all pairs of nodes connected by an edge. Thus, the assortativity coefficient A_r ranges between -1 and 1. A high A_r means that nodes tend to connect to nodes of similar degree, while a negative coefficient means that nodes likely connect to nodes with very different degree from their own [6, 19].

Page rank The page rank (PR) algorithm is commonly used to assess the popularity of a webpage. The computed metric, which we refer to as user rank, indicates the degree of participation of user in the system through interactions via image responses.

The social graph based feature vector is denoted by

$$F^G =< c(i), B_C(v), Q, R_r, A_r, PR > \tag{6}$$

3.3 Classification

The Weka Machine Learning Java library is used to build the classifier. During the training phase, all generated instances are labeled as Stegobot and normal indicating the type of communication channel. Labeling is based on prior knowledge of Stegobots used to generate social network traces.

Different classification methods, such as Naive Bayes, Support Vector Machines (SVM), Decision Tree and k-Nearest Neighbors (k-NN) are used to identify the bots. Among these algorithms, Bayesian classifier has the best performance for several reasons. First, Bayesian classifier is noise robust. Another reason that Bayesian classifier has a better performance is that the class label is predicted based on user's specific pattern. A Stegobot detection probability is calculated for each individual user based on its behaviors, instead of giving a general rule. Also, Bayesian classifier is a simple and very efficient classification algorithm.

In order to evaluate the accuracy of proposed detection algorithm, 10-fold cross validation technique is used to split the data set into ten random subsets, out of which nine sets are used for training and other one is used for testing. The same procedure is repeated until all ten subsets have been utilized as the testing set exactly once. The results reported are an average of the results of all ten runs.

3.4 Stegobot detection algorithm

In this section, we formally introduce the generalized algorithm for detection of compromised profiles in OMSN. Let $G = (U, E)$ be a time stamped multimedia social network. U be the set of all profiles of users, who shared image responses and posts until a certain instant of time. E denotes the set of all relations between the profile users. We denote real time social network dataset as D. Three different views of social profiles namely image content based features $F^{(I)}$, social profile activity based features $F^{(S)}$, social graph based features $F^{(G)}$ as detailed in Section 3.2 are considered. Ideally, the different views of social profiles are conditionally independent. For a given unknown user profile $u_k \in U$ corresponding to the kth instance $I_{u_k} \in D$ is represented by the feature vector $\mathbf{F}_{u_k} = (F_{u_k}^{(I)}, F_{u_k}^{(S)}, F_{u_k}^{(G)})$. The classification problem can be formulated as follows: Given a new user profile $u_i \in U$

represented by the calibrated feature vector $\mathbf{F}_{u_i}$, the decision maker determines the class C to which the user belongs to. Namely,

$$U = (F^{(I)}_{u_k}, F^{(S)}_{u_k}, F^{(G)}_{u_k}) \rightarrow C = \{\text{Normal, Stegobot}\} \tag{7}$$

We select any efficient machine learning algorithm, and implement the decision maker based on that. Further, $p^N_{u_i}$, $p^B_{u_i}$ are the calibrated probabilities for user u_i to be predicted as normal, Stegobot respectively. For completeness, the overall working mechanism of the Stegobot detection algorithm is shown in Algorithm 1.

Algorithm 1 Stegobot Detection Algorithm

```
Input: A set $U$ of all OMSN users
Output: $\mathbf{C}$ Set of all compromised profiles
Initilize: $\mathbf{C} \leftarrow \emptyset$
Construct: Training Model $\mathbf{Q}$    ▷ training model is constructed using traning dataset
while $U \neq \emptyset$ do
    $CL \leftarrow \{u_i\}$ where $u_i \in U$
    Collect info of $u_i$'s and list of images
    Compute: $\mathbf{F}_{u_i} = (F^{(I)}_{u_i}, F^{(S)}_{u_i}, F^{(G)}_{u_i})$
    $[p^N_{u_i}\ \ p^B_{u_i}]$ = decisionmaker($\mathbf{Q}, F_{u_i}$)
    if $p^N_{u_i} < p^B_{u_i}$ then $\mathbf{C} = \mathbf{C} \cup CL$
    end if
    $u_i \leftarrow u_{i+1}$
end while
return $\mathbf{C}$
```

The proposed Stegobot detection algorithm works as follows: Given social network G, first choose set of profiles U that needs to be tested. The critical part of the algorithm is the construction of feature vector as discussed in Section 3.2.

For each profile, suitable attributes (features) are selected on which the classification algorithm is applied. The extracted features are passed to the selected classifier in order to construct trained model $\mathbf{Q}$. The decision maker calibrates the probabilities $p^N_{u_i}$, $p^B_{u_i}$ respect to the prediction of normal and Stegobot using training model $\mathbf{Q}$. This process is repeated until the set U becomes empty. Initially, this algorithm analyzes all the $|U|$ nodes as an individual user. In the detection process $|\mathbf{C}|$ users are detected as compromised ones. For each user, feature vector is computed with time complexity d, where d is the dimension of the feature set. From the observation $|\mathbf{C}|$ is at most as large as $|U|$, the worst case time complexity is $d * |U|$.

4 Experiments and results

In order to evaluate the performance of the proposed technique it is deployed in a real time environment. We conducted different experiments which are detailed below:

4.1 Dataset collection

There are many challenges and practical difficulties in obtaining real world datasets of Stegobot network traces. Some of the publicly available datasets consist of information collected from social spam bots, which may not reflect real Stegobot network behavior. However, in order to evaluate our system we attempt to collect considerable amount of social bot network traffic traces from different sources. We used Barracude labs [38] dataset (**D1**) for large scale infiltration networks. Barracude labs is currently working on various security and privacy applications in social networks. We have collected ICWSM-13 [40] dataset (**D2**), which is released as openly available community resources for social media researchers. Further, we captured the bot and normal traces (**D3**) based on the experimental guidelines [39] and similar works [3, 20] in public network.

Our raw-image database contains 5150 color images and they are never compressed. Due to the storage in social networks, we have converted these raw images into JPEG images and build vulnerable and steganalysis databases for experimental purposes.

In bench-marking image steganographic techniques, one of the important components is the image dataset employed. Our goal was to use a dataset of images which would include a variety of texture,qualities and sizes, at the same time we wanted to have a set which would represent the type of images found on public domains. Obtaining images by INRIA Holidays dataset [16, 28], would provide diverse subjects such as natural scenes and artificial objects. Although the original images were color images of different sizes, all images have been changed into 512 X 512 gray-level images and saved as JPEG files with a quality factor 85 with the JPEG compression. The overall qualitative assessment of images have been carried out using automated qualitative assessment of multi-modal distortions method [12]. In each experiment,an image is adapted to OSN constraints, and then the infected image data set is created by NERGAL tool [29], YASS scheme [30] and F5 [35]. The infected image is uploaded into some of the social network profiles through user account, and then downloaded from another account. Finally, downloaded images are used for the experiments. For each method, different image data sets with different payloads and file types are generated. Table 1 provides the general information about datasets used in this paper.

4.2 Feature evaluation

Feature analysis is very important to construct effective feature vector to design a detection technique with high efficiency. The main challenge of this work is to identify the most influential features for detecting malicious profiles. The detection of Stegobot on OSNs is a form of adversarial challenge between normal and malicious profiles. Thus, it is important to understand whether different sets of features could lead our approach to accurate detection.

In order to study the importance of selected attributes we use well known feature selection method available on Weka [14]. We assessed the relative power of the 36 selected features [23] in discriminating Stegobot profile and legitimate profile by applying χ^2 test. Table 2 presents the 10 top most features selected from the feature vector according to the ranking estimated by χ^2 test. From Table 2, one can note that the most important attributes are the clustering coefficient and betweenness. The importance of the attributes highlights

Table 1 Summary of Datasets

Source	Type of Network	Number of Instances	No. of nodes	No. of edges
Barracuda lab	Facebook, Flickr	476667	931585	6930045
ICWSM 2013	Facebook, Titter	6426585	1976454	1665369
	Facebook			
Public Network	Flickr, Google+	22345	6150	29983

Table 2 Summary of selected OMSN based feature attributes

χ^2 Rank /Position	Selected Attribute
1	Clustering coefficient
2	Betweenness
3	URL ratio
4	Message/Image Shares
5	Message Similarity
6	Contourlet high frequency subband
7	Page Rank
8	Entropy
9	Contourlet medium frequency subband
10	DCT high frequency subband

an important aspect to detect Stegobot communication. Sometimes Stegobots mimic as a real OMSN user and in this scenario it may escape from the content based analysis. Also it is observed from the ranking, URL ratio is one of the significant features and it is true since Stegobots are most interested in sharing URLs, similar messages and images. We can also note that contourlet high frequency subbands, entropy, DCT high frequency subbands are efficient in the detection of Stegobot profiles.

The detection of stegobot on OMSNs is a form of adversarial challenge between normal and malicious profiles. Thus, it is important to understand whether different sets of features could lead our approach to accurate detection scheme. We estimate the classification results considering different subsets of 10 features that occupy adjacent positions in the ranking are used. Table 3 presents the number of features from each set(image,profile,graph) in the top 10,20,30 and 40 most discriminative features according to the ranking estimated by chi-square test.

4.3 Performance evaluation

In the evaluation of Stegobot detection algorithm, we perform cross-validation using the datasets **D1**, **D2** and **D3** as described in Section 4.1. Different characteristics of Stegobot are calibrated and translated into feature vector. The proposed Stegobot detection algorithm is evaluated with many of the classifiers in the Weka tool using default values for all parameters. In this work, the efficiency of four different families of classification methods; Naive Bayes(NB), Decision Trees(DT), k- Nearest Neighbor(k-NN) and Support Vector Machine(SVM) with respect to accuracy have been compared.

More robust models can be achieved by locating k, where $k > 1$, neighbors and letting the majority vote decide the outcome of the class labeling. A higher value of k results in a smoother, less locally sensitive function. The nearest neighbor classifier can be regarded as a special case of the more general k-nearest neighbors classifier, hereafter referred to

Table 3 Number of features at top positions in χ^2 ranking

Feature Set	Top 10	Top 20	Top 30	Top 40
Image	6	13	21	24
Profile	4	5	7	10
Graph	0	2	2	6

Table 4 Computed performance metrics for different classification methods with Facebook dataset and Flickr dataset

	Facebook Dataset					Flickr Dataset				
Classifier	TP	TN	FP	FN	ACC	TP	TN	FP	FN	ACC
Naive Bayes	96.13	100	0.0	3.87	96.24	96.23	100	0	3.77	96.26
Decision Tree	92.28	97.51	2.49	7.72	94.30	94.37	95.89	4.44	5.63	94.98
k-NN	93.15	98.54	1.46	6.85	93.61	92.36	96.42	3.58	7.65	93.68
SVM	90.42	97.83	2.17	9.58	89.93	89.64	100	0	10.36	89.93

as a k-NN classifier. The essential assumption of the method is that malicious profiles are surrounded by malicious and infected profiles.

During the practical usage of our proposed system, we build classifiers from different families using the datasets **D1**, **D2** and test the results using dataset **D3**. The results in terms of true positive(TP), true negative(TN), false positive(FP) , false negative(FN) and Accuracy(ACC) are investigated in different scenarios. Accuracy is the most famous metric in machine learning, which is used to measure correct classification of all instances to their actual class. The top 4 classifiers and their performance are presented in the Table 4. It can be seen that, the Naive Bayesian classifier has the best overall performance in comparison with other techniques.

To perform the cross validation, we have conducted additional experiments with the same data sets. First we train the same set of classifiers using the public network dataset(**D3**). Then the trained models are tested against datasets **D1** and **D2**. The results of these experiments are summarized in Table 5. As can be seen, again Naive Bayesian classifier gives the best results, with an average of 96.26 % and a 4.25 % false positives.

In addition, we have performed a set of experiments to assess the generalization ability of our proposed method with Naive Bayesian classifier. We again trained the Naive Bayes classifier multiple times with randomly selected instances from all the three datasets. Every time, we left out specific type of social network traffic from training. For example, first we train the classifier while leaving out all Facebook traffic from the training dataset, and then we test the obtained trained classifier on the Facebook traffic. The results of this set of experiments are reported in Table 6. The results show that the proposed method can detect almost all types of social network profiles. The best algorithm, Naive Bayes, even

Table 5 Computed performance metrics for different classification methods with Barracuda labs dataset and ICWSM - 2013 dataset

	Barracuda labs Dataset					ICWSM - 2013 Dataset				
Classifier	TP	TN	FP	FN	ACC	TP	TN	FP	FN	ACC
Naive Bayes	95.44	98.76	1.24	4.56	92.24	95	98.93	1.07	5.0	96.26
Decision Tree	94.55	98.76	1.24	5.45	92.30	94.53	99.21	0.79	5.47	97.9
k-NN	92.68	98.75	1.25	7.32	96.1	91.92	99.05	0.95	8.08	97.9
SVM	88.40	98.84	1.16	11.6	87.12	96.1	87.12	0.59	12.08	96.3

Table 6 Computed performance metrics for Naive Bayes classification method with different dataset

Dataset	TP	TN	FP	FN	ACC
Facebook	96.15	96.34	3.66	3.85	96.30
Flickr	96.18	94.54	5.46	3.82	95.56
Twitter	95.17	95.44	4.56	3.84	95.93
Google+	94.37	95.89	4.11	5.63	94.98

without any tuning reaches almost 96 % true positive and 98 % true negative. All other classification methods perform well, with significant true positive rates and false positive rates. This attests the effectiveness of our proposed method.

Some times, accuracy does not provide any details about misclassification of instances where a class tends to be misclassified. In Stegobot detection, it is worse if a genuine profile is being wrongly classified as a bot than if a bot is misclassified as a legitimate profile. Therefore, we used the precision which is used to determine the fraction of actual positives in the group of instances classified as positives. Precision will be high if the number of correctly classified infected profile is high and the number of false positives is low. Recall measures how many elements of a class are correctly classified. In this case, we use it to measure how many of all actual bots are detected. In addition, F measure is used, which is the harmonic mean between precision and recall.

To compare the performance of classification with the top 10 combined features and that of individual features some experiments are done and the results are shown in Table 7. Experimental results show an increase in detection rate in combined feature set based detection. This leads to the conclusion that, doing classification in multi view feature set can increase the botnet detection rate.

Finally Receiver Operating Characteristic (ROC) analysis is used to measure the goodness of the proposed algorithm. The x-axis represents a cumulative false positive rates while the y-axis represents the cumulative true positive rates. False positive rate also known as the false alarm ratio.

Figure 3 shows the comparison of the accuracy of individual feature set and combined feature set. From the results, it is clear that combined feature based Stegobot detection algorithm gives better accuracy than other single feature set.

Further, in the multi view feature analysis we identify the parameters which are responsible for evasion of detection. Then, the change in values of these parameters are correctly identified by the help of Navie Bayes classifier. The main idea is to exploit the strength of selected parameters and Navie Bayes to obtain a robust detection mechanism which ensures resiliency of evasion of detection.

Table 7 Performance comparison of individual feature with combined feature set

	Image features			Social network features			Combined features		
Classifier	Precision	Recall	F measure	Precision	Recall	F measure	Precision	Recall	F measure
Naive Bayes	1	0.74	0.84	0.78	0.65	0.71	0.96	0.98	0.97
SVM	0.95	0.62	0.75	0.71	0.62	0.62	0.93	0.94	0.93
Decision Tree	0.94	0.68	0.78	0.66	0.46	0.54	0.91	0.94	0.92
k-NN	0.89	0.81	0.85	0.59	0.55	0.57	0.90	0.87	0.88

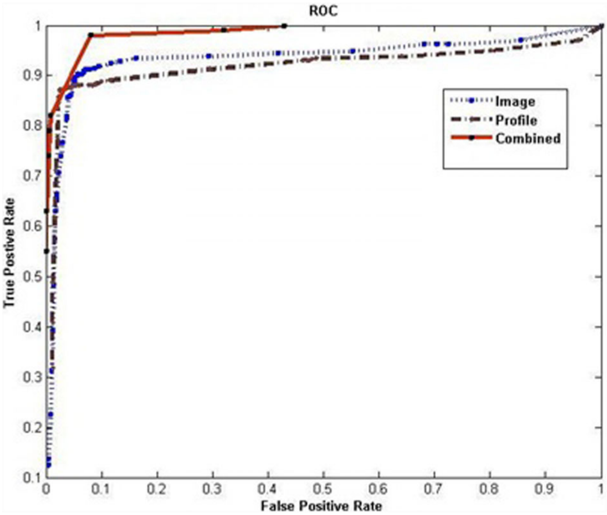

Fig. 3 Performance variance of Naive Bayes classifier with respect to different feature set

The experiments ended with very promising results, showing that the model of Stegobot detection is possible with some scalability challenges. Further research should concentrate on deploying in real time online social network.

5 Conclusion

Stegobot detection differs from the detection of traditional botnets since Stegobot traffic does not introduce new communication endpoints between bots. In this paper a new approach to detect Stegobot profiles as well as Stegobot communication in online multimedia social network is presented based on multiview features. This method uses a composition of multimedia image content based features, profile based features and social graph theoretic features to identify profiles that experience a sudden change in behavior. This detection can be extended in a straightforward manner to color images by considering the color image as three grayscale images. Traditional classification algorithms are applied to detect suspicious behaviors of Stegobot. Data collected from four popular multimedia social networking sites Facebook, Flicker, Twitter and Google+ are utilized for this study. Our study shows that some of the identified attributes are significant for classification of data and can be useful for a network forensic analyst to develop better prevention strategies. So far in this work, machine learning algorithms have been used for the detection of Stegobot. The main focus of this work is to detect whether a user profile in a multimedia social network community is a bot profile or not. The social graph based features could identify Stegobot communication. In future, large scale infiltration in OMSNs is a major cyber threat and defending against such threats is a challenge.

References

1. Angelopoulou O (2007) ID Theft: A computer forensics' investigation Framework. School of Computer and Information Science. Edith Cowan University, Perth

2. Benevenuto F, Rodrigues T, Almeida V, Almeida J, Gonalves M (2009) Detecting spammers and content promoters in online video social networks. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pp 620–627
3. Boshmaf Y, Muslukhov I, Beznosov K, Ripeanu M (2013) Design and analysis of a social Botnet. Comput Netw 57(2):556–578
4. Buscarino A, Frasca M, Fortuna L, Fiore A. S (2012) A new model for growing social networks. IEEE Syst J 6(3):531–538
5. Cao Q, Sirivianos M, Yang X, Pregueiro T (2012) Aiding the detection of fake accounts in large scale social online services. In: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (pp. 15-15). USENIX Association
6. Castillo C, Donato D, Gionis A, Murdock V, Silvestri F (2007) Know your neighbors: Web spam detection using the web topology. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp 423–430
7. Ellison NB (2007) Social network sites: Definition, history, and scholarship. J Comput.-Mediat Commun 13(1):210–230
8. Fedynyshyn G, Chuah MC, Tan G (2011) Detection and classification of different Botnet C & C channels. In: Autonomic and Trusted Computing. Springer, Berlin, pp 228–242
9. Fire M, Katz G, Elovici Y (2012) Strangers intrusion detection-detecting spammers and fake proles in social networks based on topology anomalies. HUMAN 1(1):26
10. Fridrich J, Goljan M, Hogea D (2003) Steganalysis of JPEG images: Breaking the F5 algorithm. In: Information Hiding. Springer, Berlin, pp 310–323
11. Gao H, Hu J, Wilson C, Li Z, Chen Y, Zhao BY (2010) Detecting and characterizing social spam campaigns. In: Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, pp 35–47
12. Gowacz A, Grega M, Gwiazda P, Janowski L, Leszczuk M, Romaniak P, Romano S. P (2010) Automated qualitative assessment of multi-modal distortions in digital images based on GLZ. Ann Telecommun-annales des tlcommunications 65(1-2):3–17
13. Perdisci GR, Zhang J, Lee W (2008) Botminer: Clustering analysis of network traffic for protocol-and structure-independent Botnet detection. In: USENIX Security Symposium, vol 5, pp 139–154
14. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. H (2009) The WEKA Data Mining Software: An update. ACM SIGKDD Explor Newsl 11(1):10–18
15. Hughes D, Rayson P, Walkerdine J, Lee K, Greenwood P, Rashid A, Brennan M (2008) Supporting law enforcement in digital communities through natural language analysis. In: Computational Forensics. Springer, Berlin, pp 122–134
16. Jegou H, Douze M, Schmid C (2008) Hamming embedding and weak geometric consistency for large scale image search. In: Computer VisionECCV 2008. Springer, Berlin, pp 304–317
17. Kodovsk J, Fridrich J (2012) Ensemble classifiers for steganalysis of digital media. IEEE Trans Inf Forensics Secur 7(2):432–444
18. LibenNowell D, Kleinberg J (2007) The linkprediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031
19. Mislove AE (2009) Online social networks: measurement, analysis, and applications to distributed information systems. ProQuest
20. Nagaraja S, Houmansadr A, Piyawongwisal P, Singh V, Agarwal P, Borisov N (2011) Stegobot: a covert social network Botnet. In: Information Hiding. Springer, Berlin, pp 299–313
21. Nagaraja S, Mittal P, Hong CY, Caesar M, Borisov N (2010) BotGrep: Finding P2P Bots with Structured Graph Analysis. In: USENIX Security Symposium, pp 95–110
22. Nagaraja S, Anderson R (2009) The snooping dragon: social-malware surveillance of the Tibetan movement. University of Cambridge Computer Laboratory
23. Natarajan V, Sheen S, Anitha R (2014) Multilevel Analysis to Detect Covert Social Botnet in Multimedia Social Networks. The Computer Journal, bxu063
24. Natarajan V, Sheen S, Anitha R (2012) Detection of Stegobot: A covert social network Botnet. In: Proceedings of the First International Conference on Security of Internet of Things, pp 36–41
25. Natarajan V, Anitha R (2012) Universal steganalysis using contourlet transform. In: Advances in Computer Science, Engineering & Applications. Springer, Berlin, pp 727–735
26. Pitsillidis A, Levchenko K, Kreibich C, Kanich C, Voelker GM, Paxson V, Savage S (2010) Botnet Judo: Fighting Spam with Itself. In: NDSS

27. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web, pp 851–860
28. Schaefer G, Stich M (2003) UCID: An uncompressed color image database. In: Electronic Imaging 2004 (pp. 472-480). International Society for Optics and Photonics
29. Shafiq MZ, Khayam SA, Farooq M (2008) Embedded malware detection using markov n-grams. In: Detection of Intrusions and Malware, and Vulnerability Assessment. Springer, Berlin, pp 88–107
30. Solanki K, Sarkar A, Manjunath BS (2007) YASS: Yet another steganographic scheme that resists blind steganalysis. In: Information Hiding. Springer, Berlin, pp 16–31
31. Stein T, Chen E, Mangla K (2011) Facebook immune system. In: Proceedings of the 4th Workshop on Social Network Systems, p 8
32. Stringhini G, Kruegel C, Vigna G (2010) Detecting spammers on social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference, pp 1–9
33. Viswanath B, Post A, Gummadi KP, Mislove A (2011) Analysis of social network-based sybil defenses. ACM SIGCOMM Comput Commun Rev 41(4):363–374
34. Wasserman S, Faust K (1994) Social network analysis: Methods and applications (Vol. 8). Cambridge university press
35. Westfeld A (2001) F5 A steganographic algorithm. In: Information hiding. Springer, Berlin, pp 289–302
36. Zainudin NM, Merabti M, Llewellyn-Jones D (2010) Digital forensic investigation model for online social networking. In: Proceedings of the 11th Annual Conference on the Convergence of Telecommunications, Networking & Broadcasting, Liverpool, pp 21–22
37. Zheng X, Zeng Z, Chen Z, Yu Y, Rong C (2015) Detecting spammers on social networks. Neurocomputing 159:27–34
38. Barracuda Labs https://barracudalabs.com/
39. Social Computing Research Group http://socialnetworks.mpi-sws.org/datasets.html
40. ICWSM http://www.icwsm.org/2014/datasets/datasets/

Natarajan Venkatachalam received his B.Sc degree in Mathematics from Bharathiar University in 2008, and M.Sc. degree in Applied Mathematics from Anna University in 2010. He is currently pursuing Ph.D degree in steganalysis and its applications at Anna University, Chennai, India. From July 2010 to May 2012, he served as a Research Associate in Smart and Secure Environment Projects, funded by National Technical Research Organization, Government of India, New Delhi. In June 2012, he joined Cognizant Research and Development , Chennai, as a Graduate Research in Intern-Technology. Natarajan is an active student member of IEEE Computer Society. His research includes cyber security, steganography , steganalysis, coding theory and image processing.

R. Anitha is an Associate Professor with the Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Coimbatore, India. She received her M.Sc. degree in Mathematics from Madurai Kamaraj University, Madurai, India and got her M.Phil. and Ph.D. degrees in Mathematics from Bharathiar University, Coimbatore, India. Her research interests include Cryptography, Information Security, Digital Watermarking, Botnet and Malware Detection.