INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND APPLICATIONS
VOLUME XVI, ISSUE IV, APRIL 22, WWW.IJCEA.COM, ISSN 2321-3469

DETECTION OF MALICIOUS SOCIAL BOTS USING URL FEATURES

Aarti B. Mastud[1], Svara R. Masurekar[1], Adarshsingh M. Mokashi[1], Aarti Abhyankar[2]
[1]Research Scholar, [2]Professor, Department of Computer Engineering, New Horizon Institute of Technology and Management, University of Mumbai, Thane, India

ABSTRACT:

Malicious social bots generate fake messages and automate their social relationships either by pretending like a follower or by creating multiple fake accounts with malicious activities. Moreover, malicious social bots post shortened malicious URLs in the message in order to redirect the requests of online social networking participants to some malicious servers. Hence, distinguishing malicious social bots from legitimate users is one of the most important tasks. To detect malicious social bots, extracting URL-based features (such as URL redirection, frequency of shared URLs, and spam content in URL) consumes less amount of time in comparison with social graph-based features (which rely on the social interactions of users). Furthermore, malicious social bots cannot easily manipulate URL redirection chains. In this article, a Logistic Regression algorithm is proposed by integrating a trust computation model with URL-based features for identifying trustworthy participants (users) in the social media network. This will help to determine the trustworthiness of each participant accurately. Experimentation has been performed on two social media network data sets, and the results illustrate that the proposed algorithm achieves improvement in precision, recall, and accuracy compared with existing approaches for MSBD.

Keywords: Logistic Regression Model, malicious social bots, trust, legitimate, Non-Legitimate

[1] INTRODUCTION

A malicious social bot is a software program that pretends to be a real user in online social networks. They perform several malicious attacks. Moreover, a malicious social bot may post

shortened phishing URLs in the message a participant clicks on a shortened phishing URL, the participant's request will be redirected to intermediate URLs associated with malicious servers that will redirect the user to malicious web pages. Then, the legitimate participant is exposed to an attacker. Distinguishing malicious social bots from legitimate users is one of the most important tasks in social networking apps. To detect malicious social bots, extracting URL-based features consumes less amount of time in comparison with social graph-based features (which rely on the social interactions of users). Malicious social bots cannot easily manipulate URL redirection chains. We will build a learning automata-based malicious social bot detection (LA-MSBD) algorithm by integrating a trust computation model with URL-based features for identifying the trustworthy users in the social network. Experimentation will be performed on two social media network data sets.[2]

[1.1] SOCIAL BOTS:

A social bot is an agent that communicates more or less autonomously on social media, often with the task of influencing the course of discussion and/or the opinions of its readers. It is related to chatbots but mostly only uses rather simple interactions or no reactivity at all. The messages (e.g. messages) it distributes are mostly either very simple, or prefabricated (by humans), and it often operates in groups and various configurations of partial human control (hybrid). It usually targets advocating certain ideas, supporting campaigns, or aggregating other sources either by acting as a "follower" and/or gathering followers itself. In this very limited respect, social bots can be said to have passed the Turing test. If social media profiles are expected to be human, then social bots represent fake accounts. The automated creation and deployment of many social bots against a distributed system or community is one form of Sybil attack. Twitter Bots are already well-known examples, but corresponding autonomous agents on Facebook and elsewhere have also been observed. Nowadays, social bots are equipped with or can generate convincing internet personas that are well capable of influencing real people, although they are not always reliable. Social bots, besides being able to (re-)produce or reuse messages autonomously, also share many traits with spambots with respect to their tendency to infiltrate large user groups. Using social bots is against the terms of service of many platforms, especially Twitter and Instagram. However, a certain degree of automation is of course intended by making social media APIs available.[1,2]

[1.3] MALICIOUS SOCIAL BOTS

A malicious social bot is a software program that pretends to be a real user in online social networks (OSNs).[1] Moreover, malicious social bots perform several malicious attacks, such as containing URL(s) with the neighboring participants (i.e., followers for followers), the participant adapts URL shortened service (i.e., bit.ly) in order to reduce the length of the URL (because a message is restricted up to 140 characters). Moreover, a malicious social bot may post shortened phishing URLs in the message. When a participant clicks on a shortened phishing URL, the participant's request will be redirected to intermediate URLs associated with malicious servers that, in turn, redirect the user to malicious web pages. Then, the legitimate participant is exposed

to an attacker. This leads to social networks suffering from several vulnerabilities (such as phishing attacks). Several approaches have been proposed to detect spam in social networks. These approaches are based on message-content features, social relationship features, and user profile features. However, malicious social bots can manipulate profile features, such as hashtag ratio, follower ratio, URL ratio, and the number of messages again. The malicious social bots can also manipulate message-content features, such as sentimental words, emoticons, and most frequent words used in the messages, by manipulating the content of each message. The social relationship-based features are highly robust because the malicious social bots cannot easily manipulate the social interactions of users in the social network. However, extracting social relationship-based features consumes a huge amount of time due to the massive volume of social network graphs. Therefore, identifying the malicious social bots from the legitimate participants is a challenging task in the social network. The existing malicious URL detection approaches are based on DNS information and lexical properties of URLs. The malicious social bots use URL redirections in order to avoid detection. However, for detectors, identification of all malicious social bots is an issue because malicious social bots do not post malicious URLs directly in the messages. Thus, it is important to identify malicious URLs (i.e., harmful URLs) posted by malicious social bots in the
Networks.

[2] LITERATURE REVIEW

[2.1] LITERATURE BACKGROUND:

The existing methods and systems help us in providing us with the basic knowledge of how we can implement our project. We learn from various elaborate explanations and intend to improve the existing methodologies and hence come up with our proposed system. Following are the various insights gathered from different sources which have proved helpful in our literature survey. The first paper we referred to was based on the clickstream sequences of the social bots by P. Shi, Z. Zhang, and K.-K.-R. Choo. A novel method of detecting malicious social bots, including both feature selection based on the transition probability of clickstream sequences and semi-supervised clustering, is presented in this paper. This method not only analyzes the transition probability of user behavior clickstreams but also considers the time feature of behavior. There was also major use of URL shortening services by these malicious social bots. For that purpose, we discussed the second paper which majorly focussed on the detection of social bots which used URL shortening services authored by S. Lee and J. Kim. Here, botnet models based on USSes were used by the bots to prepare for new security threats before they evolve. Specifically, using USSes for alias flux to hide botnet command and control (C&C) channels. In alias flux, a botmaster obfuscates the IP addresses of his C&C servers, encodes them as URLs, and then registers them to USSes with custom aliases generated by an alias generation algorithm. Later, each bot obtains the encoded IP addresses by contacting USSes using the same algorithm. For USSes that do not support custom aliases, the botmaster can use shared alias lists instead of the shared algorithm. DNS-based botnet detection schemes cannot detect an alias flux botnet, and network-level detection and blacklisting of the fluxed aliases are difficult. We also

discuss possible countermeasures to cope with these new threats and investigate operating USSes.

Thirdly, we referred to a neural network-based spam detection model which was done on a Twitter network. Most of the current spam filtering methods in Twitter focus on detecting the spammers and blocking them. However, spammers can create a new account and start posting new spam messages again. So there is a need for robust spam detection techniques to detect the spam at message level. These types of techniques can prevent spam in real time. To detect the spam at message level, often features are defined, and appropriate machine learning algorithms are applied in the literature. Here, they combine both deep learning and traditional feature-based models using a multilayer neural network which acts as a meta-classifier. The evaluation was done on two data sets, one data set is balanced, and another one is imbalanced. The experimental results show that our proposed method outperforms the existing methods.[1,2,3]

[2.2] EXISTING SYSTEM:

The existing malicious URL detection approaches are based on DNS information and lexical properties of URLs. The malicious social bots use URL redirections in order to avoid detection. The authors have presented that social bots use URL shortening services and URL redirection in order to redirect users to malicious web pages. These approaches are based on message-content features, social relationship features, and user profile features. However, malicious social bots can manipulate profile features, such as hashtag ratio, follower ratio, URL ratio, and the number of messages. The malicious social bots can also manipulate message-content features, such as sentimental words, emoticons, and most frequent words used in the messages, by manipulating the content of each message. The social relationship-based features are highly Therefore, identifying the malicious social bots from the legitimate participants is a challenging task in the Twitter network. The malicious social bots use URL redirections in order to avoid detection. However, for detectors, identification of all malicious social bots is an issue because malicious social bots do not post malicious URLs directly in the messages. Thus, it is important to identify malicious URLs (i.e., harmful URLs) posted by malicious social bots on Twitter. robust because malicious social bots cannot easily manipulate the social interactions of users in the Twitter network. However, extracting social relationship-based features consumes a huge amount of time due to the massive volume of the social network graph. Most of the existing approaches are based on supervised learning algorithms, where the model is trained with the labeled data in order to detect malicious bots in OSNs. However, these approaches rely on statistical features instead of analyzing the social behavior of users. Moreover, these approaches are not highly robust in detecting the temporal data patterns with noisy data (i.e., where the data is biased with untrustworthy or fake information) because the behavior of malicious bots changes over time in order to avoid detection.

[2.3] LIMITATION OF EXISTING SYSTEM:

The malicious social bots can manipulate profile features, such as hashtag ratio, follower ratio, URL ratio, and the number of messages. The malicious social bots can also manipulate

message-content features, such as sentimental words, emoticons, and most frequent words used in the messages, by manipulating the content of each message. The social relationship-based features are highly robust because the malicious social bots cannot easily manipulate the social interactions of users in the Twitter network. The existing approaches rely on statistical features instead of analyzing the social behavior of users. Moreover, these approaches are not highly robust in detecting the temporal data patterns with noisy data (i.e., where the data is biased with untrustworthy or fake information) because the behavior of malicious bots changes over time in order to avoid detection.

[3] PROBLEM DEFINITION

Malicious social bots generate fake messages and automate their social relationships either by pretending to be a follower or by creating multiple fake accounts with malicious activities. This problem can be prevented by identifying malicious social bots from legitimate participants.

[4] PROPOSED SYSTEM OVERVIEW

We first propose a framework for analyzing the messages posted by participants in the social media network. In addition, we present a trust model with several features that are extracted from URLs (which are posted by the participants in the posts) for evaluating the trust value of each participant. Finally, a Logistic Regression algorithm is proposed to identify malicious social bots.

The proposed framework consists of three components: data collection, feature extraction, and LA model. The data collection component consists of three subcomponents (i.e., subphases): reading posts, collecting posts, and URLs. Moreover, the collected posts and collected URLs are stored in a repository. The feature extraction consists of two subcomponents: expanding shortened URLs and extracting feature sets. Whenever a feature extraction component obtains a shortened URL from the repository, it is converted into a long URL using URL shortened services (such as t.co, bit.ly, and tinyurl.com). For each URL, we extract several features that are based on the lexical properties of URLs (such as spam content and the presence of -, @, and # symbols in the domain name) along with the features of URL redirection (such as URL redirection length and relative position of initial URL). Furthermore, we use these features as input to the proposed LA model for MSBD. The proposed LA model is integrated with a trust evaluation model. Moreover, the trust model determines the probability of posts containing any malicious information (such as URL redirection, frequency of URLs, and spam content in URL). Finally, after evaluating the malicious behavior of a series of messages posted by a participant, we classify posts as malicious and legitimate posts. However, malicious posts are likely to be posted by malicious social bots. This helps in distinguishing malicious social bots from benign participants.

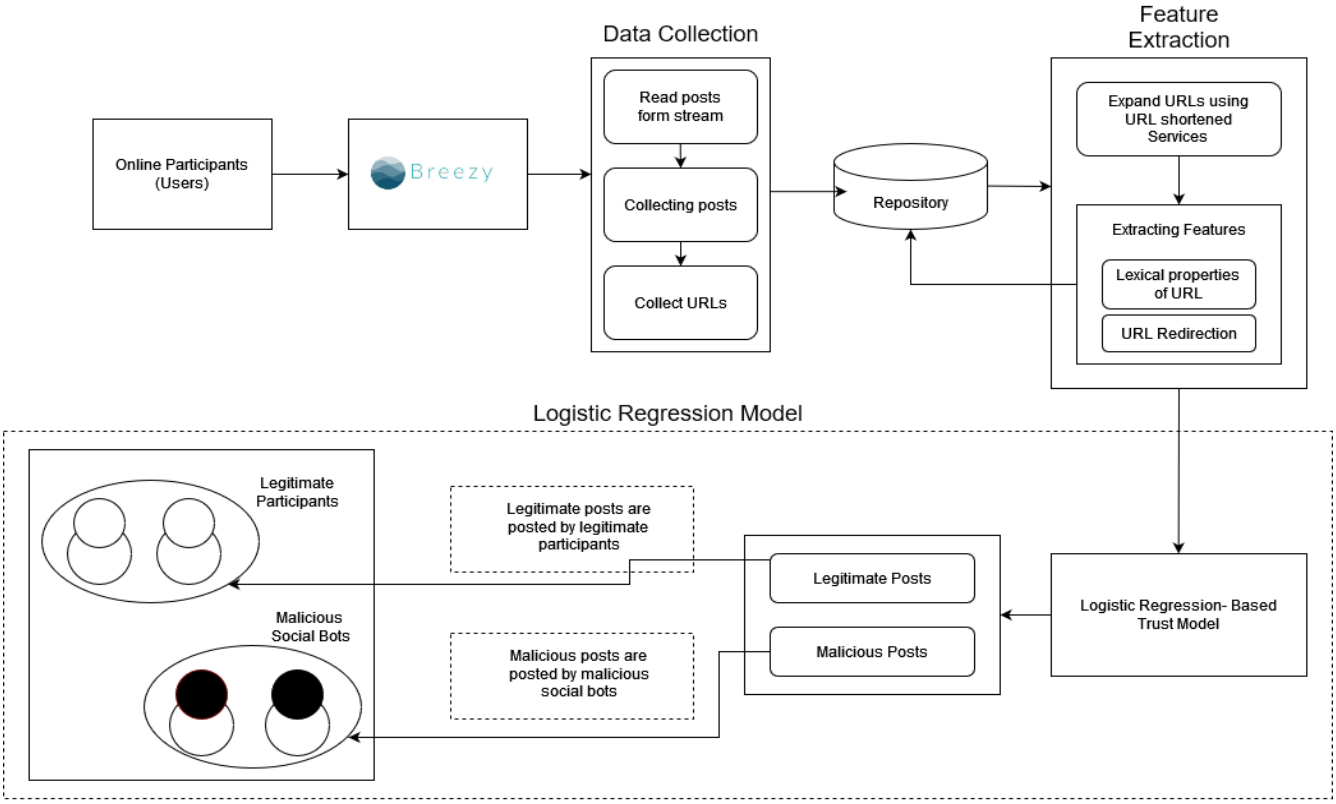

Figure. 1 Architecture Diagram.

[4.1] ALGORITHM

[I] Data Collection:

- A framework where the users can post messages containing URLs will be created.
- Post will be here from the stream.
- Collection of posts will be done.
- URLs will be collected.

[II] Extraction of Features:

- Shortened URLs will be expanded using URL shortening services.
- Lexical properties of the URLs will be extracted.

[III] Data Storage:

- The Extracted features will be stored in the repository.
- Dataset will also be stored in this repository.

[IV] Logistic Regression Model:

- Dataset will also be stored in this repository.
- We have created a logistic regression based trust model wherein, using the extracted features of the URLs collected and inserted dataset a binary classification will be done.

- In this classification, the URLs posted by the user will be categorized into Legitimate URLS or Malicious URLs.
- Using this categorization, the users will be classified as Legitimate users or Malicious Social Bots.

[5] RESULTS AND DISCUSSIONS

To achieve the desired results, we used a number of various approaches to make sure our project model is as accurate as possible. We tried using several different algorithms such as, 'Bag of Words Algorithm' to classify Legitimate and Malicious Bots, which had an accuracy of approximately 72%. In our second attempt, we tried classifying the users using 'SVC Model', which gave an accuracy of 84% approximately. Since, the accuracies of the above mentioned models are significantly lower than what we expected, we tried using 'Logistic Regression Model' using user-defined vectorization tokens. The accuracy of this model was approximately 94%. Lastly, we tried using the 'Logistic Regression Model' using the tfidfvectorizer module and we got an accuracy of 99%. This model was able to classify the users into Legitimate users and Malicious Bots and furthermore, such bots can be blocked from using the platform, to enhance the security of the system.

[6] CONCLUSION

The existing models use only the URL features for analysis of malicious users. Along with URL features we also analyzed the behavioral pattern of the user. Lexicals properties, such as use of symbols like: &,!_,-,*,% and numbers, also be analyzed and the detection will be done. We made a trust computational model for the same. At the end, if we come across any suspicious bot, we can also block the user from posting any further messages that can be harmful for other users. The proposed system helps to detect malicious social bots accurately. We analyze the malicious behavior of a participant by considering URL-based features. The model executes for a finite set of learning actions to update the action probability value and achieves the advantages of incremental learning.

REFERENCES

[1] P. Shi, Z. Zhang, and K.-K.-R. Choo, “Detecting malicious social bots based on clickstream sequences,” IEEE Access, vol. 7, pp. 28855–28862, 2019.

[2] G. Lingam, R. R. Rout, and D. V. L. N. Somayajulu, “Adaptive deep Q-learning model for detecting social bots and influential users in online social networks,” Appl. Intell., vol. 49, no. 11, pp. 3947–3964, Nov. 2019.

[3] D. Choi, J. Han, S. Chun, E. Rappos, S. Robert, and T. T. Kwon, “Bit.ly/practice: Uncovering content publishing and sharing through URL shortening services,” Telematics Inform., vol. 35, no. 5, pp. 1310–1323, 2018.

[4] S. Lee and J. Kim, “Fluxing botnet command and control channels with URL shortening services,” Comput. Commun., vol. 36, no. 3, pp. 320–332, Feb. 2013.

[5] H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, “A framework for real-time spam detection in Twitter,” in Proc. 10th Int. Conf. Commun. Syst. Netw. (COMSNETS), Jan. 2018, pp. 380–383.

Author[s] brief Introduction

Aarti Mastud

Aarti Mastud is a Computer Engineering Research Scholar at New Horizon Institute of Technology and Management, Thane. Her research interests include Machine Learning, Full Stack Development. She is the member of CSI.

Svara Masurekar

Svara Masurekar is a Computer Engineering Research Scholar at New Horizon Institute of Technology and Management, Thane. Her research interests include Machine Learning and Deep Learning. She is a member of CSI

Adarshsingh Mokashi

Adarshsingh Mokashi is a Computer Engineering Research Scholar at New Horizon Institute of Technology and Management, Thane. His research interests include Development of Gaming apps.

Aarti Abhyankar

Mrs. Aarti Abhyankar is Assistant Professor in computer Engineering at New Horizon Institute of Technology and Management, Thane. She has pursued her engineering education at University of NMU and completed ME from Mumbai University. She has published 5 research papers in the International journal/conference. Her research interests include Data Security.