International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

# DETECTION OF MALICIOUS SOCIAL BOTS USING LEARNING AUTOMATA WITH URL FEATURES IN TWITTER NETWORK

**Koppula Vijaya Ramya Sri** (MCA Scholar), B V Raju College, Vishnupur, Bhimavaram, West Godavari District, Andhra Pradesh, India, 534202.
**G. Ramesh Kumar**, B V Raju College, Vishnupur, Bhimavaram, West Godavari District, Andhra Pradesh, India, 534202.

**Abstract**
Malicious social bots generate fake tweets 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 tweet 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 in the Twitter network. 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 learning automata-based malicious social bot detection (LA-MSBD) algorithm is proposed by integrating a trust computation model with URL-based features for identifying trustworthy participants (users) in the Twitter network. The proposed trust computation model contains two parameters, namely, direct trust and indirect trust. Moreover, the direct trust is derived from Bayes' theorem, and the indirect trust is derived from the Dempster– Shafer theory (DST) to determine the trustworthiness of each participant accurately. Experimentation has been performed on two Twitter data sets, and the results illustrate that the proposed algorithm achieves improvement in precision, recall, F-measure, and accuracy compared with existing approaches for MSBD.

## 1. INTRODUCTION

MALICIOUS social bot is a software program that pretends to be a real user in online social networks (OSNs) [1], [2]. Moreover, malicious social bots perform several malicious attacks, such as spread social spam content, generate fake identities, manipulate online ratings, and perform phishing attacks [1]. In Twitter, when a participant (user) wants to share a tweet containing URL(s) with the neighboring participants (i.e., followers or followers), the participant adapts URL shortened service (i.e., bit.ly [3]) in order to reduce the length of URL (because a tweet is restricted up to 140 characters). Moreover, a malicious social bot may post shortened phishing URLs in the tweet [4]. As shown in Fig. 1, when 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 Twitter network suffering from several vulnerabilities (such as phishing attack). Several approaches

have been proposed to detect spam in the Twitter network [5]–[8]. These approaches are based on tweet-content features, social relationship features, and user profile features.

However, the malicious social bots can manipulate profile features, such as hashtag ratio, follower ratio, URL ratio, and the number of re tweets. The malicious social bots can also manipulate tweet-content features, such as sentimental words, emoticons, and most frequent words used in the tweets, by manipulating the content of each tweet [9]. 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. However, extracting social relationship-based features consumes a huge amount of time due to the massive volume of social network graph [10]. Therefore, identifying the malicious social bots from the legitimate participants is a challenging task in the Twitter network. The existing malicious URL detection approaches [11], [12] are based on DNS information and lexical properties of URLs. The malicious social bots use URL redirections in order to avoid detection [13]. However, for detectors, identification of all malicious social bots is an issue because malicious social bots do not post malicious URLs directly in the tweets. Thus, it is important to identify malicious URLs (i.e., harmful URLs) posted by malicious social bots in Twitter. Most of the existing approaches [14], [15] 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 [16].

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 [17], [18]. This motivated us to consider one of the reinforcement learning techniques (such as the learning automata (LA) model) to handle temporal data patterns. In this work, we design an LA model to detect malicious social bots with improved precision and recall. In this article, the malicious behavior of participants is analyzed by considering features extracted from the posted URLs (in the tweets), such as URL redirection, frequency of shared URLs, and spam content in URL, to distinguish between legitimate and malicious tweets. To protect against the malicious social bot attacks, our proposed LA-based malicious social bot detection (LA-MSBD) algorithm integrates a trust computational model with a set of URL-based features for the detection of malicious social bots. The proposed trust computational model contains two parameters, namely, direct trust and indirect trust. The direct trust value is derived from the Bayesian learning [19] (by considering URL-based features) to determine the trustworthiness of tweets posted by each participant. In addition to the direct trust, belief values (i.e., indicators for determining indirect trust) are collected from multiple neighbors of a participant. This is due to

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

the fact that in case the neighbors of a participant are trustworthy, the participant is likely to be trustworthy. Furthermore, Dempster's combination rule [20] aggregates the belief values provided by multiple one-hop neighboring participants in order to evaluate the indirect trust value of participants in the Twitter network.

## 2. EXISTING SYSTEM

Besel et al. [23] analyzed social botnet attack on Twitter. The authors have presented that social bots use URL shortening services and URL redirection in order to redirect users to malicious web pages. Echeverria and Zhou [24] presented methods to detect, retrieve, and analyze botnet over thousands of users to observe the social behavior of bots. In [25], a social bot hunter model has been presented based on the user behavioral features, such as follower ratio, the number of URLs, and reputation score. In [26], a trust model has been designed to detect malicious activities in an OSN. The authors analyzed that the low trust value of a user indicates that the information spread by the user is considered as untrustworthy.

In [1], an MSBD approach has been proposed by considering user behavioral features, such as commenting, liking, and sharing. Madisetty and Desarkar [5] have developed five different convolutional neural network models by considering tweet features. In [27], a social botnet detection algorithm is proposed by considering spam content in tweets and trust to identify social bots. Gupta et al. [7] designed a framework for detecting spammers in the Twitter network using different machine learning algorithms. In this article, we focus to detect malicious social bots (who perform phishing attacks) by considering various URL-based features using an LA model.

Several spam-detection approaches have been proposed in the Twitter network to distinguish non spam accounts and spam accounts [5]–[8]. Moreover, these studies consider user profile features, which can easily be modified by malicious bots. To avoid feature manipulation, Yang et al. [28] considered social relationships between malicious users and with their neighboring users based on closeness centrality. Moreover, profile features and social interaction features may not help in detecting malicious URLs that are posted by the participants.

To address the above-mentioned problem, Janabi et al. [29] considered URL-based features (such as URL length, Http-302 status code, and disabling right click) to distinguish legitimate URLs from suspicious URLs. In [30], a URL-based approach is proposed to detect spam tweets in Twitter based on the tweet content and URL redirection chains. Patil and Patil [31 used decision tree classifiers with statistical features in order to detect malicious URLs. Moreover, social bots may use malicious URL redirections in order to avoid detection.

Disadvantages

In the existing work, the system considers user profile features, which can easily be modified by malicious bots.

This system aims to profile features and social interaction features which may not help in detecting malicious URLs that are posted by the participants..

## 3. PROPOSED SYSTEM

The proposed LA-MSBD algorithm helps to detect malicious social bots accurately

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

(in terms of precision, recall, F-measure, and accuracy) in Twitter. The major contributions are as follows.

Analyze the malicious behavior of a participant by considering URL-based features, such as URL redirection, the relative position of URL, frequency of shared URLs, and spam content in URL.

Evaluate the trustworthiness of tweets (posted by each participant) by using the Bayesian learning and Dempster–Shafer theory (DST).

Design of an LA-MSBD algorithm by integrating a trust model with a set of URL-based features.

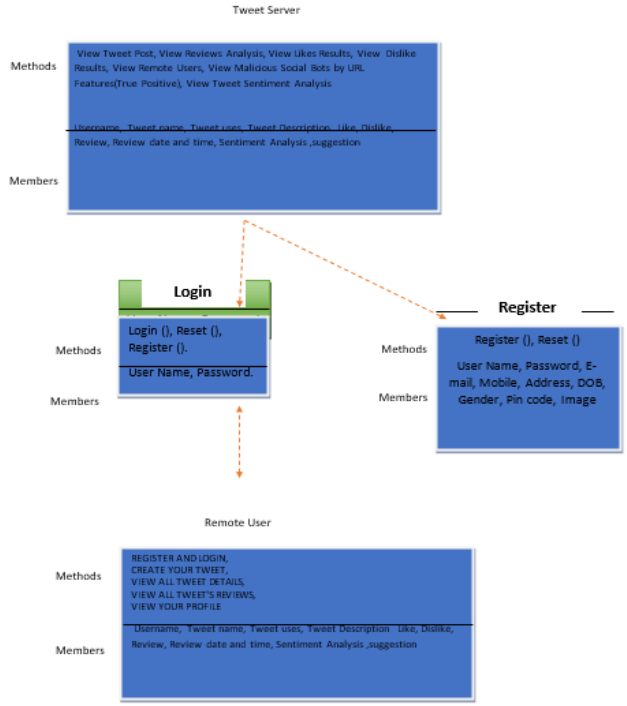
Performance evaluation of the proposed LA-MSBD algorithm using two Twitter data sets, namely, The Fake Project data set [21] and Social Honeypot data set [22] in terms of precision, recall, F-measure, and accuracy for MSBD in the Twitter network.

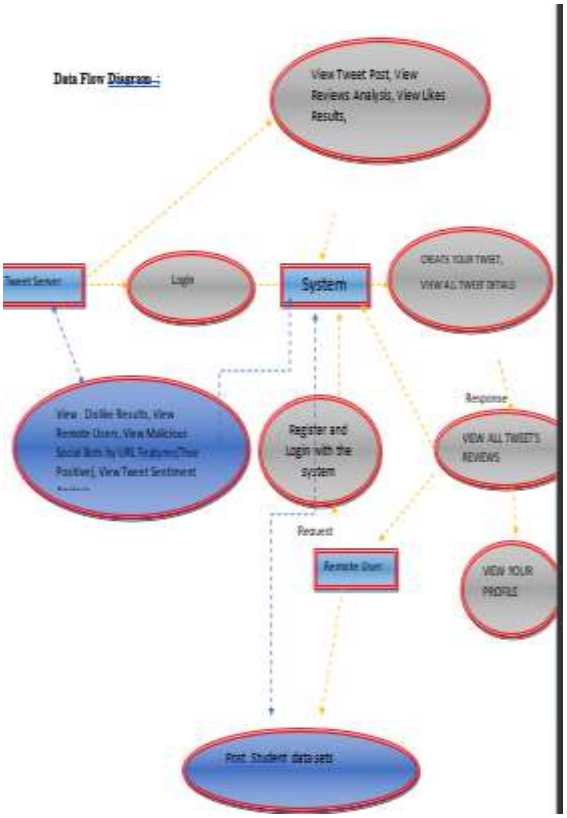
**Advantages**

The malicious behavior of participants is analyzed effectively by considering features extracted from the posted URLs (in the tweets), such as URL redirection, frequency o shared URLs, and spam content in URL, to distinguish between legitimate and malicious tweets.

To protect against the malicious social bot attacks, our proposed LA-based malicious social bot detection (LA-MSBD) algorithm integrates a trust computational model with a set of URL-based features for the detection of malicious social bots.

**Class Diagram :**



## 4. CONCLUSIONS

This article presents an LA-MSBD algorithm by integrating a trust computational model with a set of URL-based features for MSBD. In addition, we evaluate the trustworthiness of tweets (posted by each participant) by using the Bayesianlearning and DST. Moreover, the proposed LA-MSBD algorithm executes a finite set of learning actions to update action probability value (i.e., probability of a participant posting malicious URLs in the tweets). The proposed LA-MSBD algorithm achieves the advantages of incremental learning. Two Twitter data sets are used to evaluate the performance of our proposed LA-MSBD algorithm. The experimental results show that the proposed LA-MSBD algorithm achieves up to 7% improvement of accuracy compared with other existing algorithms. For The Fake Project and Social Honeypot data sets, the proposed LA-MSBD algorithm has achieved precisions of 95.37% and 91.77% for MSBD, respectively. Furthermore, as a future research challenge, we would like to investigate the dependence among the features and its impact on MSBD**.**

## 5. 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] S. Madisetty and M. S. Desarkar, "A neural network-based ensemble approach for spam detection in Twitter," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 4, pp. 973–984, Dec. 2018.

[6] H. B. Kazemian and S. Ahmed, "Comparisons of machine learning techniques for detecting malicious webpages," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1166–1177, Feb. 2015.

[7] 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.

[8] T. Wu, S. Liu, J. Zhang, and Y. Xiang, "Twitter spam detection based on deep learning," in *Proc. Australas. Comput. Sci. Week Multiconf. (ACSW)*, 2017, p. 3.

[9] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "Key challenges in defending against malicious socialbots," Presented at the 5th USENIX Workshop Large-Scale Exploits Emergent Threats, 2012, pp. 1–4.

[10] G. Yan, "Peri-watchdog: Hunting for hidden botnets in the periphery of online social networks," *Comput. Netw.*, vol. 57, no. 2, pp. 540–555, Feb. 2013.

[11] D. Canali, M. Cova, G. Vigna, and C. Kruegel, "Prophiler: A fast filter for the large-scale detection of malicious Web pages," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 197–206.

[12] A. K. Jain and B. B. Gupta, "A machine learning based approach for phishing detection using hyperlinks information," *J. Ambient Intell.*