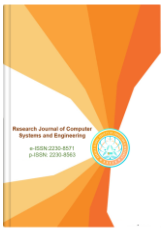
Research Journal of Computer Systems and Engineering

ISSN: 2230-8571, 2230-8563

Volume 02 Issue 01 - 2021 (January to June)

Page 11:15



Bot Net Detection for Social Media Using Segmentation with Classification Using Deep Learning Architecture

Dr. Prakash Pise
Sinhgad Institute of Management,
SPPU PUNE UNIVERSITY, India
prakash.pise532@gmail.com

Article History

Received: 22 January 2021
Revised: 14 April 2021
Accepted: 19 May 2021

Abstract

The use of bots on social media raises serious questions about the reliability and authenticity of the content. Currently, there are numerous methods for detecting bots. However, there is still room for improvement in the detection's accuracy.This research propose novel technique in BotNet detection for social media networks based on segmentation and classification utilizing DL technique. Input is collected as BotNet based social media and processed for noise removal and smoothening. The segmentation of processed data is carried out using Fuzzy-C means clustering and feature extracted using Multi layered Convolutional Neural Network (MLCNN). The experimental analysis has been performed in terms of RMSE, F 1 score, recall, accuracy, and precision.The suggested solution offers a machine learning-based bot identification technique that is more precise and efficient. The research makes use of a number of strategies and methods that improve the effectiveness of bot detection and removal.

Keywords: Bot Net, social media networks, segmentation, classification, deep learning

1. Introduction:

The term "bot," which denotes that the victim is under the control of an attacker, is where the concept of the botnet originates. Recently, there has been a sharp growth in the use of botnets. Botnets are collections of machines that are connected to the internet and have a lot of computational power and bandwidth together.The attacker, also referred to as the botmaster, can direct sizable botnet networks from several locations in order to conduct attacks. Distributed denial of service assaults, email spam, key logging, and password cracking are characteristics of botnets. Botnets are one of the biggest risks to the Internet right now [1].Botnets have a number of different component strategies that practically make them distinct in terms of their capabilities and technical implementation [2].However, there is always a botmaster, at least one, but frequently thousands of command and control servers, and controlled nodes. A botnet is a collection of infected nodes that act as an internet worm by executing commands while attempting to avoid detection by anti-malware software [3].

2. Related works:

An extensive body of research has been produced in the domain of bot (net) detection. The majority of currently used bot detection approaches use techniques for C2 channel detection based on statistical characteristics of packets and flows [4].Solutions like [5] have a narrow scope and are concentrated on particular communication protocols like IRC. Contrarily, Botminer [6] is a protocol-independent approach that makes the assumption that bots connected to the same botnet have common harmful traits and communication patterns.This presumption is oversimplified because botnets frequently randomise their communication patterns and topologies, as seen in more recent malware like Mirai [8]. It follows that a nonprotocol-specific, less evasive detection technique is preferred.Additionally, [9,10] have taken advantage of traffic-based statistical features and ML-driven anomaly detection to detect known and unknown assaults with minimal mistake rates.

3. System model:

This section discussnovel technique in BotNet detection for social media networks based on segmentation and classification utilizing DL technique. Input is collected as BotNet based social media and processed for noise removal and smoothening. The segmentation of processed data is carried out utilizing Fuzzy-C means clustering as well as feature extracted using MLCNN.

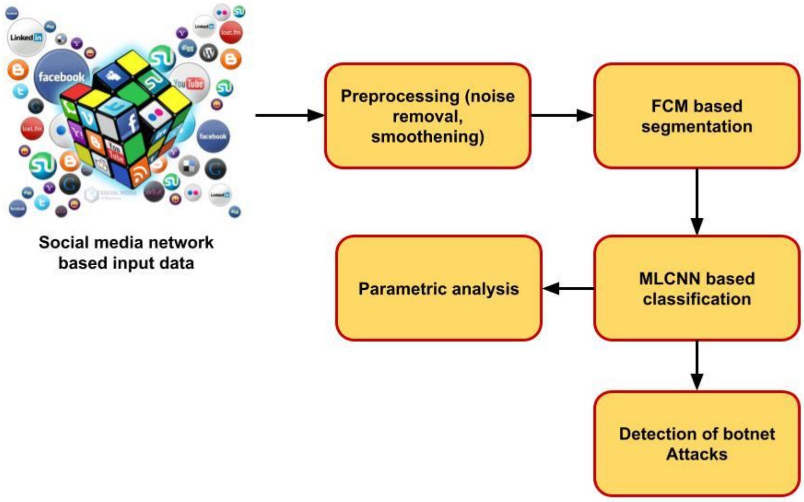

Social media network based input data
Preprocessing (noise removal, smoothening)
FCM based segmentation
MLCNN based classification
Parametric analysis
Detection of botnet Attacks

Figure 1: Overall proposed architecture

3.1Fuzzy-C means clustering based segmentation

The FCM algorithm works by assigning each data point's membership to a corresponding CC based on the data point and cluster distance; its key benefit is that it consistently produces excellent results even when the data are overlapped, and it also assigns each data point to several clusters.

Let us consider the dataset $\mathrm{Z} = \{\mathrm{z}_1, \mathrm{z}_2, \dots, \mathrm{z}_\mathrm{q}\}$ with cluster set $\mathrm{x} = \{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{p}\}$ and membership set $\mathrm{W}= \left\{ \begin{matrix} \mathrm{w}_{\mathrm{k}1} \mid 1 \leq \mathrm{k} \leq \mathrm{e}, 1 \leq \mathrm{l} \leq \\ \mathrm{p} \end{matrix} \right\}$ further considering these three FCM can be formulated by eq. (1).

$$\min: \sum_{k=1}^{e} \sum_{l=1}^{p} w_{kl}^{o} \| z_l - x_k \|^2 \qquad (1)$$

We create modified-FCM, or Modified FCM, in the equation below to prevent spurious clustering by eq. (2).

$$L_o(W,X) = \sum_{k=1}^{e} \eta_i \sum_{k=1}^{p} (1 - u_{kl}^{o} \quad ^{\circ} \qquad (2)$$

In order to update the membership matrix and cluster centres, the equation can be optimised as eq. (3):

$$x_k = \sum_{l=1}^{p} w_{kl}^{o} z_l / \sum_{l=1}^{p} w_{kl} \qquad (3)$$

Membership matrix by eq. (4)

$$w_{kl} = \left(1 + \left(\frac{e_{kl}}{\eta_k}\right)^{-1/(o-1)}\right)^{-1} 0 \qquad (4)$$

3.2 *Multi layered Convolutional Neural Network (MLCNN):*

Convolution and pooling operations alternate in MLCNN. In order to speed up computation and increase spatial and configuration invariance, convolutional layers are frequently alternated with pooling layers. The final few levels (near the outputs) will be fully linked 1-dimensional layers. A feed-forward neural network can be thought of as a function of mapping data x in greater in eq. (5), (6):

$f(x) = f_L(\ldots f_2(f_1(x_1, w_1), w_2)\ldots, w_L)$.(5)

L(w)=1/n×(∑i=1nl(zi,f(xi;w))),(6)

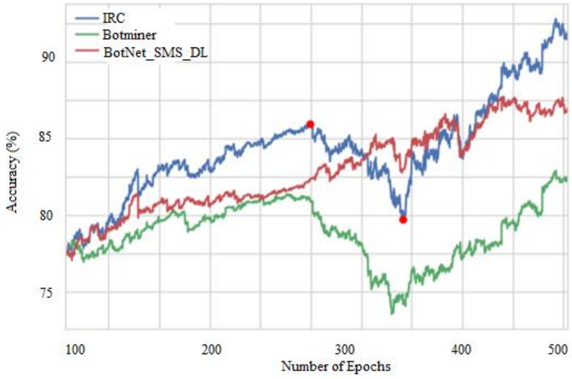
where n is the sample count and zi is sample i's actual label. A neural network can be trained to minimise the loss function L in order to solve the training problem. To get a better depiction of the original image, the results of these groupings are then tiled so that they overlap (such as edges in the image).

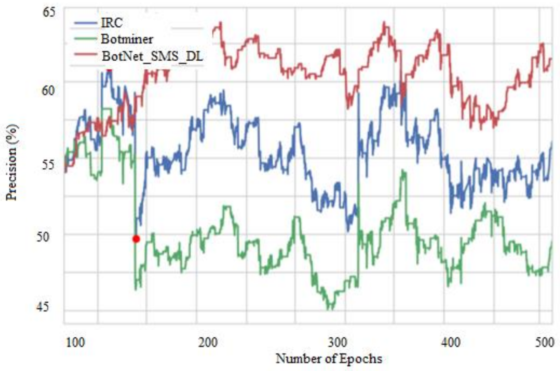
4. Performance analysis:

This paper's preprocessing was put into practise using MATLAB 2021a. Based on Python 3.6, the enhanced SVM network model was created. The operating system was Windows 64-bit, and it had an Intel(R) Core(TM) i7-5500U CPU running at 2.40 GHz and 8.00 GB of RAM. The CTU-13 [42] dataset is the foundation for our analysis. The 13 individual subset datasets (DS) that make up CTU13 include captures from 7 different types of malware that perform activities such as port scanning, DDoS, click fraud, spamming, and more.

Table-1 Comparative analysis of CTU13 dataset

Parameters	**IRC**	**Botminer**	**BotNet_SMS_DL**
Accuracy	78	85	92
Precision	56	61	65
Recall	61	65	68
F-measure	48	51	53
RMSE	52	56	58





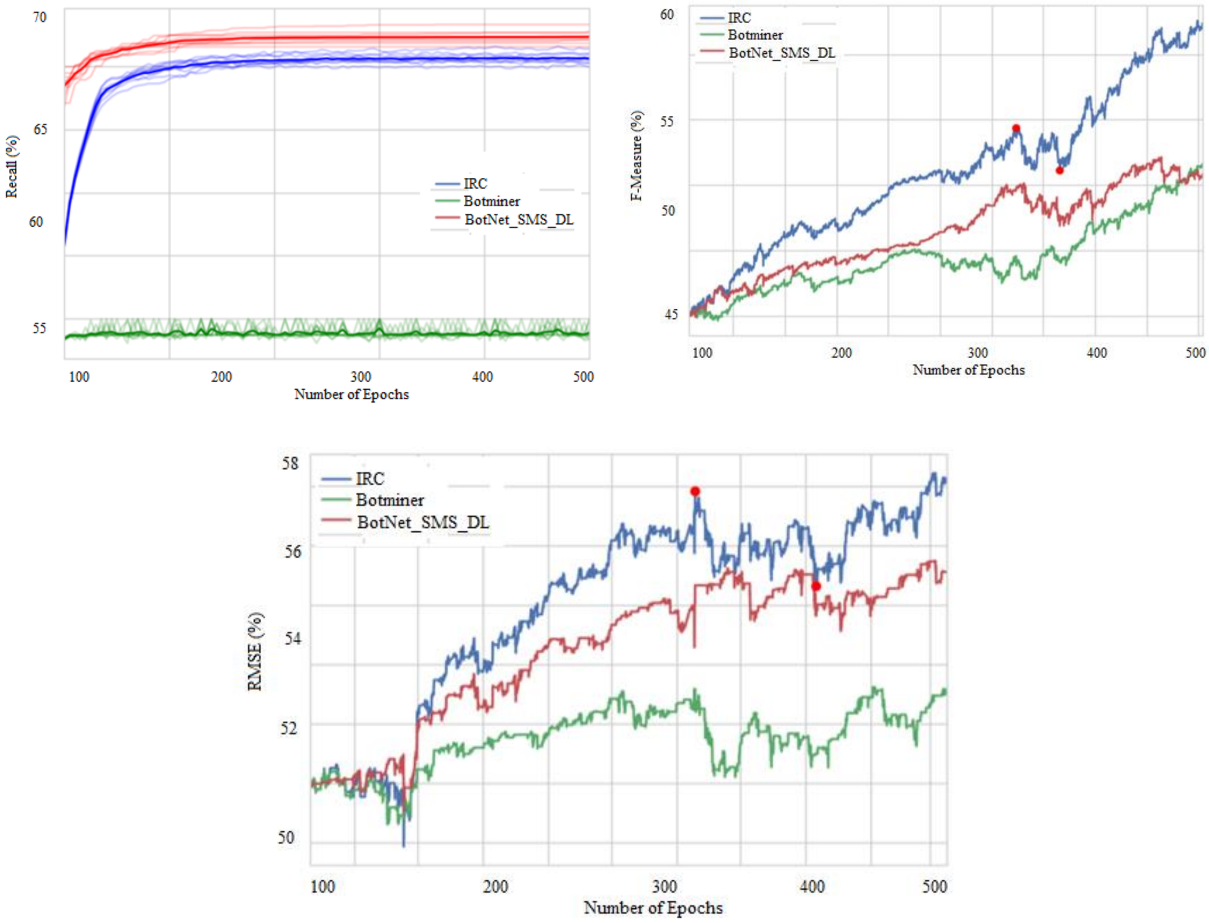


Figure- 3 Comparative analysis of CTU13 dataset in terms of (a) Accuracy, (b) precision, (c) Recall, (d) F-measure, (e) RMSE

The above table-1 and figure 3 (a)- (e)shows the comparative analysis for CTU13 dataset based on accuracy, precision, recall, F_ measure and RMSE. Here proposed technique attained accuracy of 92%, precision of 65%, recall of 68%, F_ measure of 53% and RMSE of58% which is optimized when compared to existing technique for MIT-BIH dataset. One metric for measuring classification model performance is accuracy. Informally, accuracy is percentage of predictions that our method correctly predicted. Accuracy is defined as follows in formal language: Accuracy is the quantity of accurate forecasts. sum of all projections. How frequently an algorithm successfully classifies a data point can be determined, for example, by looking at the accuracy of the algorithm. The percentage of projected data points that really occurred is known as accuracy. One measure of a ML methods effectiveness is precision, or standard of a successful prediction the model makes. Ratio of overall number of true positives to total number of positive forecasts is known as precision. F1-score combines precision and recall of a classifier into one metric by computing their harmonic means. It is mostly used to evaluate the performance of two classifiers.

5. Conclusion:

This research proposenovel technique in BotNet detection for social media networks based on segmentation and classification utilizing DL technique. Input is collected as BotNet based social media and processed for noise removal and smoothening. The segmentation of processed data is carried out using Fuzzy-C means clustering and feature extracted using Multi layered Convolutional Neural Network (MLCNN). the proposed technique attained accuracy of 92%, precision of 65%, recall of 68%, F_ measure of 53% and RMSE of58%. Additionally, feature normalisation greatly enhances the models' spatial stability. The system is resistant to unidentified threats and cross-network ML model inference and training. It works well with big data and can find bots that use various protocols.

Reference:

[1] McDermott, C. D., Majdani, F., &Petrovski, A. V. (2018, July). Botnet detection in the internet of things using deep learning approaches. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

[2] Vinayakumar, R., Soman, K. P., Poornachandran, P., Alazab, M., &Jolfaei, A. (2019). DBD: deep learning DGA-based botnet detection. In *Deep learning applications for cyber security* (pp. 127-149). Springer, Cham.

[3] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, “A comprehensive survey on machine learning for networking: evolution, applications and research opportunities,” Journal of Internet Services and Applications, vol. 9, no. 1, pp. 1–99, 2018.

[4] Pektaş, A., &Acarman, T. (2019). Deep learning to detect botnet via network flow summaries. *Neural Computing and Applications*, *31*(11), 8021-8033.

[5] Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., &Tesconi, M. (2019, June). Rtbust: Exploiting temporal patterns for botnet detection on twitter. In *Proceedings of the 10th ACM conference on web science* (pp. 183-192).

[6] Sperotto, A.; Schaffrath, G.; Sadre, R.; Morariu, C.; Pras, A.; Stiller, B. An overview of IP flow-based intrusion detection. IEEE Commun. Surv. Tutor. 2010, 12, 343–356.

[7] Bridges, R.A.; Glass-Vanderlan, T.R.; Iannacone, M.D.; Vincent, M.S.; Chen, Q. A survey of intrusion detection systems leveraging host data. ACM Comput. Surv. (CSUR) 2019, 52, 1–35.

[8] Letteri, I., Penna, G. D., &Gasperis, G. D. (2019). Security in the internet of things: botnet detection in software-defined networks by deep learning techniques. *International Journal of High Performance Computing and Networking*, *15*(3-4), 170-182.

[9] Koroniotis, N., Moustafa, N., Sitnikova, E., & Slay, J. (2017, December). Towards developing network forensic mechanism for botnet activities in the IoT based on machine learning techniques. In *International Conference on Mobile Networks and Management* (pp. 30-44). Springer, Cham.

[10] Maeda, S., Kanai, A., Tanimoto, S., Hatashima, T., & Ohkubo, K. (2019, January). A botnet detection method on SDN using deep learning. In *2019 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1-6). IEEE.