I J C T A, 9(6), 2016, pp. 2799-2808 ISSN: 0974-5572
© International Science Press

Design of Malicious Activity Detection & Prevention System for Enterprise Network

Kavitha P.*, S. Tamilarasi and V. Cyrilraj*****

ABSTRACT

In the domain of network security, detection of malicious activity is one of the challengeable tasks for network defenders. To identify unwanted network traffic details, it is typically classified as a signature-based intrusion detection systems or anomaly-based detection systems. In existing systems, some researcher worked to detect or predict malicious or infected nodes strictly based on association with other known malicious nodes. However, these system follow community based detection where the resultant inference has inaccuracy problem which leads meaningless probabilistic output. To provide solution for these issues, Malicious Activity Detection and Prevention algorithm is proposed to detect and classify the types of malicious activity in enterprise networks. This system evaluates the malicious activity level on every node and tracks the all activities accurately. Proposed mechanism assists network to contribute the data to client Local PC packet by packet in cipher text mode. This process ensures that in worst cases, external attack happen then also legitimate users are unable to get complete data and view the original content. To evaluate the system performance, proposed approach is compared with various existing approach in terms of malicious detection rate and wrongly attack detection rate. MADP increases the detection classification accuracy (MDR) by 1% and decreases the wrongly attack detection rate accuracy by 1% approximate with closest existing approach. Finally, this system state that proposed MADP approach is best approach for overall datasets.

Keywords: Malicious Activity, Enterprise Network, Malicious detection rate, wrongly attack detection rate, Attack detection and prevention.

1. INTRODUCTION

In the domain of network security analysis, detection of malicious activity is one of the critical concerns for network defenders. To detect unwanted network traffic details, it is typical to categorize a signature-based intrusion detection systems or anomaly-based detection systems. The activity of multiple hosts can summative performs spatial anomaly detection which considers the relationship between one to another host. There are some existing methods available to decrease the propagation of threat level exponentially in all connected nodes. But, it does not cover the same value as a byproduct of this dampening. However, the resultant inferences have inaccuracy which leads a meaningless probabilistic output.

In literature, many researcher worked behalf of threat propagation to detect or predict malicious or infected nodes or users that are strictly based on union with other known malicious nodes. This system is highly prevalent in the terms of graph analytics. However, this system is followed community based detection system. This system is unable to track the provenance of the propagated threat which leads to inaccurate inference. Some works are implemented based on threat prioritization or identifying "Top *N*" and it can be applicable for rank-ordered list.

* Research Scholar, Department of CSE, Dr. M.G.R. Educational and Research Institute, Chennai, Tamilnadu, India, *Email: kavithamgru.phd@gmail.com*

** Professor & Head, Department of CSE, Jaya College of Engineering and Technology, Chennai, Tamilnadu, India, *Email: tamilarasisu@gmail.com*

*** Dean, Engineering & Technology, Dr. M.G.R. Educational and Research Institute, Chennai, Tamilnadu, India, *Email: cyrilraj@hotmail.com*

However, this system just sends alert or notification to user on behalf of threat detections. Existing methods are unable to maintain the good scalability in terms of threat detections and does not consider malicious preventions whose output score exceeds a threshold. To overcome these issues, Malicious Activity Detection and Prevention Algorithm is proposed to detect malicious activity that is injected by user in monitored network. It performs iterative propagation which allows for asymmetric weighting factors. This system can be applied for malicious web domains and proactively expand blacklists for new application. This system works to identify the types of attacks and prevent them based on their accessibility and authentication. Proposed system establishes secure and reliable communication between network and user during content transmissions. Proposed system enhances the privacy to remove the adverse effect of cyclic propagation which is byproduct of current methods. This system blocks the malicious activity which is performed by malicious user. This system assists network to share content with client Local PC in packet by packet in cipher text mode. This process protects user data from legitimate user. Even though, network compromised with malicious, they are unable to get complete data and view the original content**.** It ensures content transmission efficiency through packet by packet distribution and also accurately classifies the malicious detection rate and wrongly attack detection rate. The paper works are followed as:

- To establish a secure and reliable communication to transmit the data from server to client systems.
- To detect the server and client activity in step by step process to differentiate between right user and legitimate user.
- To detect the malicious activity or involved user and blocks their accessibility in enterprise network.
- Distribute the content in packet by packet to client in cipher text mode to avoid external attacks.
- Improve the accuracy of malicious detection rate and wrongly attacked detection rate

The rest of this paper organizations are followed as: Section 2 expresses the related work. It explains all research papers concept which are close to proposed mechanism. Section 3 explores the information about proposed system methodology with proposed algorithm elaboration. Section 4 discusses about proposed system implementation details along with performance. Section 5 summarizes the overall work with future work.

2. RELATED WORK

In paper [1], authors developed conditional random field framework for building probabilistic models to segment and label sequence data. Conditional random field produce several features over Markov hidden models and stochastic grammars for including the ability to relax strong independence assumptions. Paper [2] investigated a simple label propagation algorithm that used network structure alone as its guide. It does not require optimization of a predefined objective function and also no prior information about the communities. Paper [3] focused on communities based malicious detection in sociology, biology and computer science disciplines where systems are often executed in graph. Paper [4] implemented EMBER (Extreme Malicious Behavior viewer) technique for analyzing and displaying the malicious activity at city level. EMBER used a metric namely as Standardized Incidence Rate (SIR) which represents the number of hosts exhibiting malicious behavior per 100,000 available hosts. Paper [5] developed protocol to detect the anomalous network activity without providing any historical context of collection. These anomalies can be further described with a few high-level characterizations.

In this paper [6], authors implemented a method for detecting malicious activity within networks of interest. They leverage prior community detection which is worked by propagating threat probabilities across graph nodes, given an initial set of known infected nodes. Paper [7] designed new perspective approach for activity-based community detection, where a community is defined as a group of active users engaged in correlated activities over time. Paper [8], authors provided a framework and empirical results that elucidate a “detection theory” for graph-valued added database. It focuses on detection of anomalies in un-weighted and undirected graphs through L1 properties of the eigenvectors graph, called as a modularity matrix. Paper [9] developed a simple method to identify the potential members of an unstructured P2P botnet in a network starting from a known peer. Paper [10] studied about malicious activity to find out the source of computer virus in a network. This model is designed for virus spreading in a network with a variant of SIR model and then constructs an estimator for virus source.

In paper [11] authors focused on measurement of local community structure and an algorithm inferred the hierarchy of communities that enclose a given vertex by exploring the graph at a time**.** In paper [12], authors designed an approach based on comparative metrics to incentivize ISPs for mitigating botnets. Paper [14] developed a Notos, dynamic reputation system for DNS. Utilization of DNS has unique characteristics and it can be distinguished from legitimate, professionally provisioned DNS services. In paper [14], authors studied about Bayesian methods which have grown from a specialist niche to become main stream, while graphical models combined with approach to describe and apply probabilistic models. In paper [15], authors expressed about relational data which has two characteristics: first, statistical dependencies exist between the entities and second, each entity often has a rich set of features that can be aided in classification.

Paper [16] designed a general detection framework that is independent of botnet C&C protocol which is not required any prior information of botnets. In paper [17], authors explored online learning method for detecting malicious infected Web page using lexical and host-based advantages associated URLs. This approach contains large volume of training data to provide good results. Paper [18] introduced network quality term un-cleanliness: an indicator of the propensity for hosts in a network to find the compromised hosts which are compromised with external parties. In paper [19], authors studied about many types of reputation based threat like a block unsolicited email, or spam. Authors proved that these blacklists contain exhibit non-trivial false positives and false negatives threat. Paper [20] introduced EXPOSURE to employ large-scale and passive DNS analysis techniques for detecting malicious activity involved domains.

In paper [21], authors studied various aspects of the inner workings binary variants A and B which were the first in a chain of recent revisions. Their objective was to keep this epidemic resistant to ongoing eradication attempts. Paper [22] introduced a method to evaluate the network intrusion detection systems using an observable attack space. Paper [23] provided a layer protection which monitors network traffic for predefined suspicious activity or patterns, and alert system administrators when potential hostile traffic is detected. Paper [24] explained an alternative blacklist generation strategies to produce higher-quality of results for an individual network. Paper [25] developed PTP techniques to propagate trust across network. It measures the trustworthiness of incoming connection request.

3. PROPOSED SYSTEM METHODOLOGY

This section expresses the workflow of proposed mechanism along with implementation steps. In details, this phase elaborates proposed algorithm in the application a way to understand the work flow of proposed techniques in regularly used application which is displayed in figure 1(a) & (b). Implementation approach is divided in following modules namely as server, client, content upload and content download along with proposed algorithm.

3.1. Server

Here, server is a kind of user who wants to upload some content in storage server. Before doing any activity, server has to register first and next he/she should have to complete authentication process. After giving login details, proposed technique will verify the server credential information. If server found that the user is genuine then it will allow performing the operation otherwise it will be treated as malicious user.

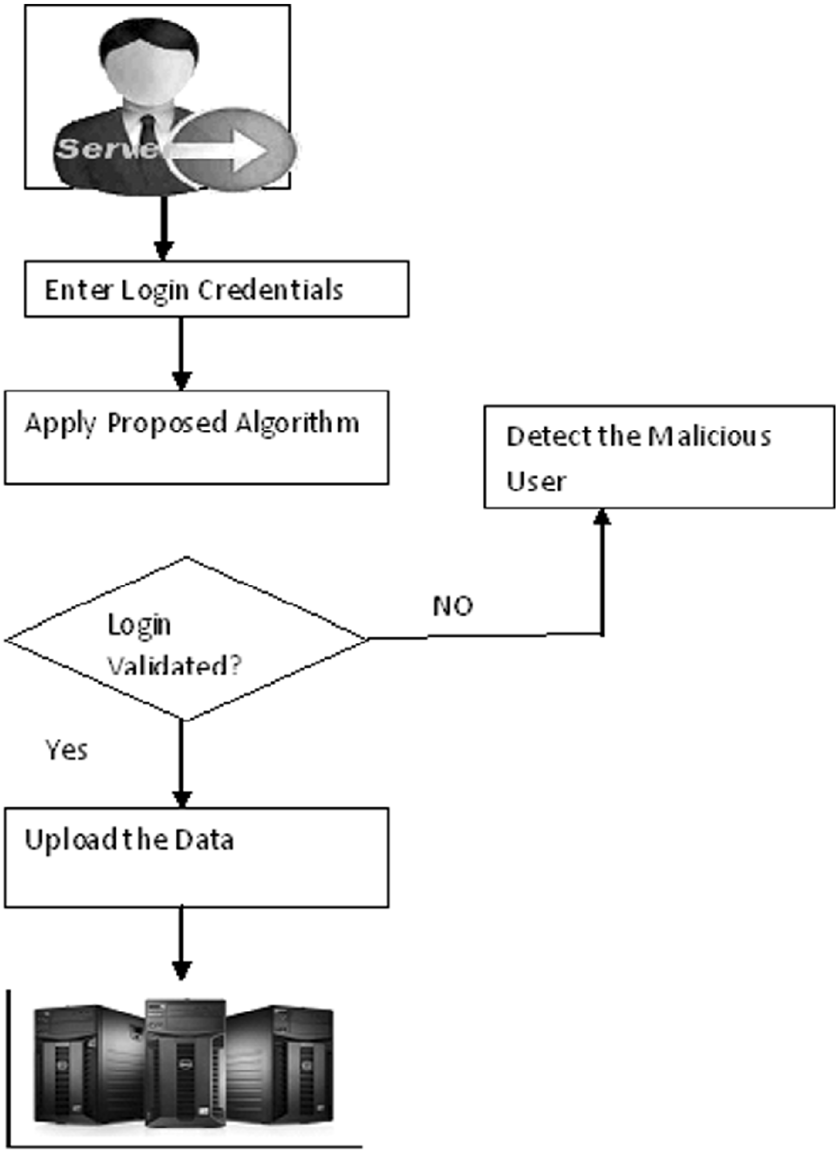

Figure 1 (a): Workflow of proposed architecture for Server Side

3.2. Content Upload

After processing the login credential, server can upload or share his/her data to storage server. Here, server can upload any types of contents in secure and efficient manner. All upload content filename will be visible for genuine clients.

3.3. Client

In this module, client is treated user who wants to access the content from storage server. Initially, client has to register with his/her personal information in this framework. After performing login authentication, client can access any kinds of content from storage server in secure and fast way.

3.4. Content Download and View

This module is feasible to user for downloading the content from storage server. Before downloading the content, client has to enter secret key which is verified by proposed approach. If clients given details are invalid then client will be treated as a malicious user and service page will be blocked for malicious user. If secret key is validated and if client try to do some abnormal activity then in this case also client will be

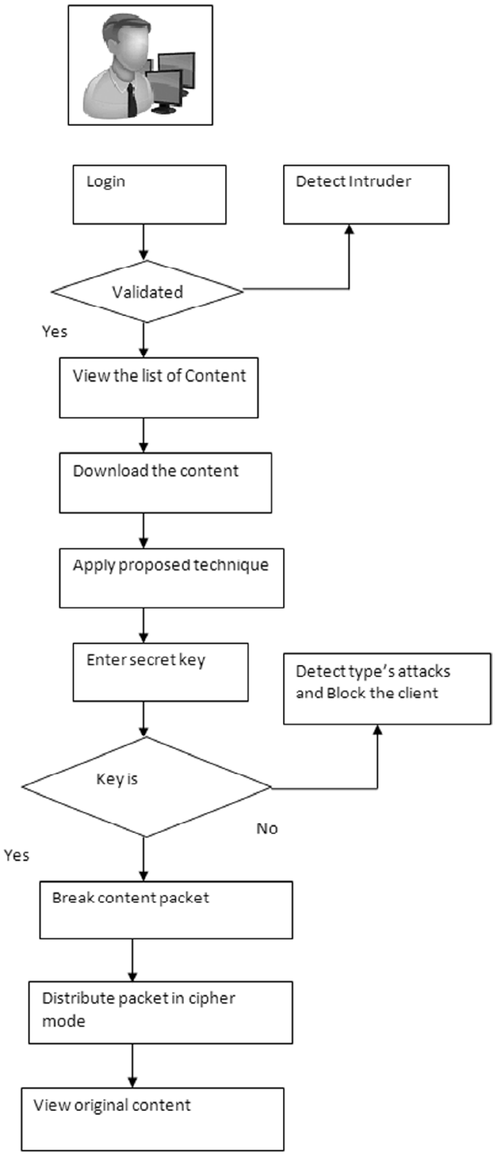

Figure 1 (b): Workflow of proposed architecture for Client Side.

treated as a malicious user. Client can download the content only one condition, if he/she produces all the information correctly and system understand that he/she is genuine user. Here, content will be divided in packet by packets in cipher text mode and distribute to client. Once client receive the entire packet then he/she can view the original content.

3.5. Malicious Activity Detection and Prevention (MADP) Algorithms

Malicious Activity Detection and Prevention algorithm is implemented to detect the various kinds of malicious activity in enterprise network. It does not only detect but it also prevents them and provides secure content transmission in networks. This approach is capable to detect all kinds of attacks like signature based, anomaly based attacks, distributed service denial attacks and Intrusion detection based attacks. These systems allow user to share the data, after completion of all credential information. This system does not think that malicious user will be outsider or external user but it may be known or registered user who wants to access the content ill-legal way. It detects and classifies the various kinds of attack. This system evaluates the malicious activity level on every user and tracks as well accurately. Hence, the level of malicious activity on every node is equal to the weighted sum of the threat of neighboring nodes, discounting the level of malicious. It defines the probability of being in the malicious community. Proposed mechanism helps to develop the application for contributing the data in packet by packet in cipher text mode to client. This process protects user data from legitimate user. Even though, network compromised with malicious, they are unable to get complete data and view the original content. This system also improves the classification accuracy of malicious detection rate and wrongly classified malicious activity

Information deviation ($D_{i)}$ is a non-commutative measure of the difference between two probability distributions P and Q. P typically represents the – True distribution of data, and Q usually represents a theory, model, or approximation of P. Initially it computes deviation for ($P \parallel M$) and ($Q \parallel M$) to achieve the total deviation for the (i^{th}) IP. If the (i^{th}) IP, Di is more than studied profile (Di Ã γ) then two probabilities have deviate. Therefore, P and Q denote the behavior of different entities. But if D is equal to 0.0 then it indicates that there is a possibility of malicious attack. It evaluates the packets which are stored in an intermediate buffer for malicious attack by using frequency counter. Each incoming packet is compared with the identity of blacklisted IP for the exact similarity. If an exact match is identified the frequency count of packet is incremented by 1. Since in a Secure Path Identification Attack is detected for large volume of packets surround the victim in a short period of time. There is a very high possibility that the attacker sends similar packet many times which almost happens in a malicious attack. The frequency count of each packet is checked. If it exceeds the threshold value for a particular IP, the system indicates an attack and that IP is identified as the attacker. The pseudo code of proposed approach is explained below in detail.

```
Input: The user credentials (UC) and Data Packets (DP)
Output:   Attack Identification (AI) and secure data transmission (SDT)
Procedure:
            For each sample $(t)$
          If studying period
        Define probabilities of each value for header Attributes
        For every $IP$ and store them;
        Else
      Define probabilities of each value for the header Attributes for every IP;
    Define the $D$  $for$ $IP$;
  If $Di \approx 0.0$
  Chance for malicious attack
  Check for malicious detection using frequency counter;
  If flooding attack ($frequency$ $counter > threshold$)
Secure path identification attack identified
  Delete matching packets;
  Else
  Allow to view the content;
  End If
  Else
  Identify the attacker and block the user;
  End If
End If
End For
End
```

Figure 2: Pseudo Code for Malicious Activity Detection and Prevention Approach

4. RESULT AND DISCUSSION

4.1. Programming Environments

This work is deployed with Intel Dual Core Processor with 1GB RAM running with windows7 ultimate. Here, proposed approach is implemented in NetBeans 8.0 along with MYSQL 5.5 database. To evaluate proposed mechanism with existing approach, Weka 3.7.3 open source tool is utilized. This is evaluated with three kinds of dataset namely as Trojan, Virus and Worm datasets which is taken from internet resource (http://nexginrc.org/Datasets/Default.aspx) ICARIS09-dataset directory.

4.2. Simulated Result

In this phase, proposed systems represent mathematical model to enhance the malicious detection rate and wrongly attack detection rate, to classify the malicious or virus infected data. Here, this system can classify the types of attack and provide enhanced security for user content. It expresses the flowing performance matrix separately namely as malicious detection rate (MDR), and wrongly attack detection rate (WADR).

4.2.1. Malicious Detection Rate (MDR)

In this section, proposed approach explains mathematical model in equation (1) to correctly detect the malicious data (%). The Malicious Detection Rate (MDR) is calculated as correctly classified malicious data with respect of overall data.

$$MDR = \frac{T_p}{T_P + F_N} \tag{1}$$

Where T_P is true positive and F_N is false negative with respect of malicious data.

4.2.2. Wrongly Attack Detection Rate (WADR)

In this section, proposed method describes mathematical model for wrongly malicious classified data (%) in equation (2). Proposed method calculates the wrongly malicious classified data with respect of total instance. Wrongly Attack Detection Rate (WADR) is calculated as:

$$WADR = \frac{F_p}{F_P + T_N} \tag{2}$$

Where F_P is false positive and T_N is true negative behalf of incorrectly classified malicious data. Here, the performance evaluation of proposed approach is tested in Weka 3.7.2 GUI tools with default parameters setting. For classification, 10 cross fold validation test is conducted to measure the precision, recall and F1 score for following classifier namely as a Bagging, Random Forest, PART, SMO (Sequential Minimal Optimization) and NB (NavieBayes) classifier which results are displayed in table 01. Here, SMO classifier function is integrated with Malicious Activity Detection and Prevention (MADP) approach. Based on the, classification result of precision, recall and F1 score, it noticed that almost classifier have 90% classification accuracy. Table 1 display the precision, recall and F1 score classification for Trojan, Virus and worm datasets. This systems display their average values for respective parameter with respective datasets. According to table 1 result, it noticed SMO performed well on Trojan and virus dataset. Finally, it claims that SMO is best classifier for overall datasets.

According to Figure 3 and 4, MADP is performed well on overall datasets. In the terms of malicious detection rate, proposed approach display best result for Trojan, virus dataset but proposed approach is outperformed by PART approach in worm datasets. However, PART classifier has very low result in

Table 1
MDR Precision, Recall and F1 Score for Trojan, Virus and Worm Datasets

Learning		*Trojan*			*Virus*			*Worm*	
Algorithms	*Precision*	*Recall*	*F1*	*Precision*	*Recall*	*F1*	*Precision*	*Recall*	*F1*
Bagging	0.935	0.934	0.934	0.966	0.966	0.965	0.957	0.957	0.956
PART	0.916	0.915	0.915	0.876	0.876	0.876	0.953	0.952	0.952
NB	0.916	0.915	0.916	0.978	0.978	0.978	0.957	0.957	0.957
SMO	0.968	0.967	0.967	0.989	0.989	0.989	0.961	0.961	0.961

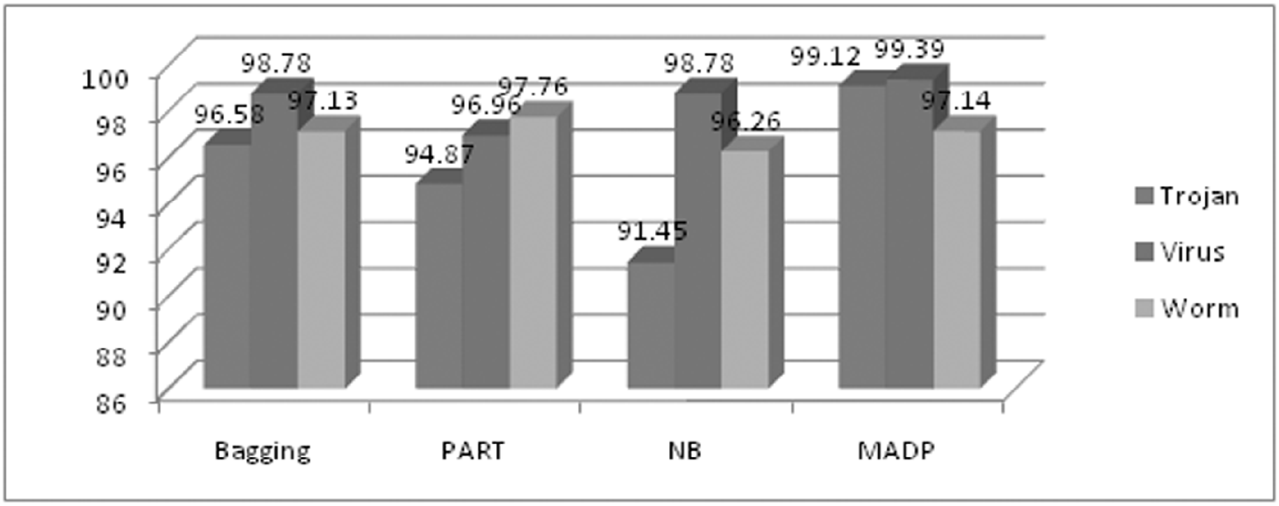

Figure 3: Detection Rate (MDR) in % for Trojan, Virus and Worm Datasets

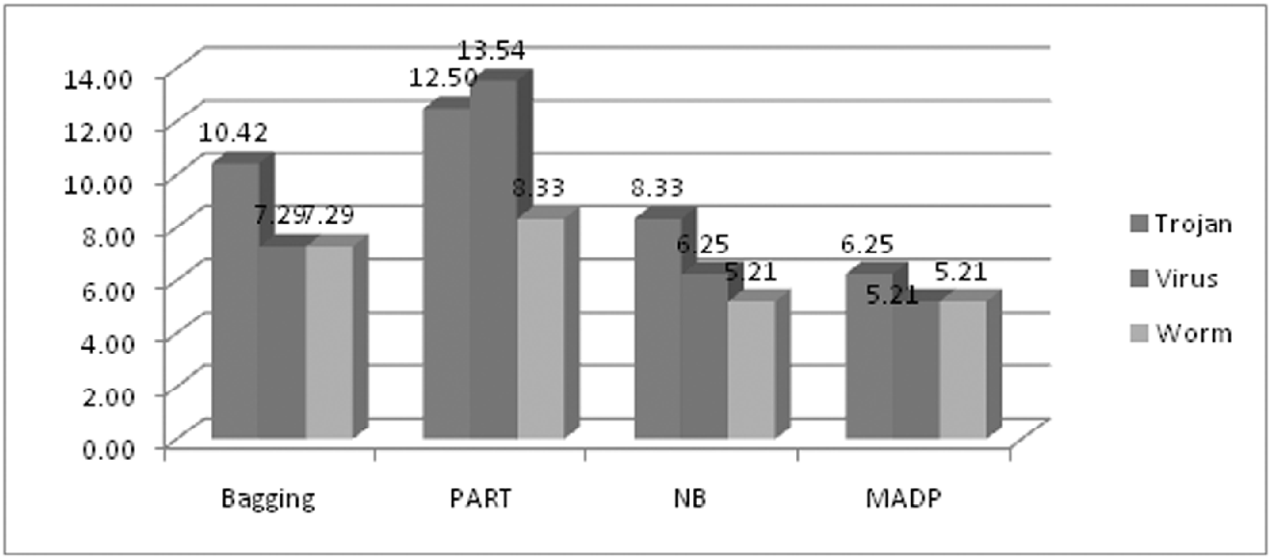

Figure 4: Wrongly Attack Detection Rate (WADR) in % for Trojan, Virus and Worm Datasets

Trojan and virus datasets. In the terms of WADR (wrongly attack detection rate), proposed MADP performs well in overall dataset. Behalf of worm dataset, proposed approach and NaiveBayes display same value. However, NaiveBayes has high wrongly classified rate in other datasets. MADP is increase the detection classification accuracy (MDR) by 1% and decrease the wrongly detection accuracy by 1% approximately with closest existing approach. Finally, this system stated that proposed MADP approach is best approach for overall datasets.

5. CONCLUSION

This paper presents Malicious Activity Detection and Prevention algorithm to detect and classify the types of malicious activity in enterprise networks. This system evaluates the malicious activity level on every node and tracks accurately. Proposed mechanism assists network for transmitting the data in packet by packet in cipher text mode to client. This process protects user data from legitimate user. Even though, network compromised with malicious, they are unable to get complete data and view the original content**.** This system ensures that content transmission efficiency through packet by pack distributions. This system evaluates the malicious activity level on every user and tracks as well accurately. MADP is performed well on overall datasets in the terms of malicious detection rate for Trojan, Virus and Worm dataset. But, proposed approach is outperformed by PART classifier in Worm datasets.

However, PART classifier has very low result in Trojan and virus datasets dataset. In the terms of WADR (wrongly attack detection rate), proposed MADP performs well on overall dataset. With respect to worm dataset, proposed approach and NaiveBayes display the same value. However, NaiveBayes has high wrongly classified rate in other datasets. MADP increases the detection classification accuracy (MDR) by 1 % and decrease the wrongly detection accuracy by 1% approximately with closest existing approach. Finally, this system stated that proposed MADP approach is a best approach for overall datasets. In future, this paper work can be extended for military system in remote areas. This work can also be integrated with cloud to store large amount of contents.

REFERENCES

[1] Lafferty, J., McCallum, A., & Pereira, F. C., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", *In Proceedings of the 18th International Conference on Machine Learning*, pp. 282-289, 2001.

[2] Raghavan, U. N., Albert, R., & Kumara, S., "Near linear time algorithm to detect community structures in large-scale networks" *Physical Review. E, 76*(3), 036106, pp. 1-12, 2007.

[3] Fortunato, S., "Community detection in graphs", Physics *reports*, pp. 75-174, 2010.

[4] Yu, T., Lippmann, R., Riordan, J., & Boyer, S. "Ember: a global perspective on extreme malicious behavior" In *Proceedings of the seventh international symposium on visualization for cyber security*, pp. 1-12, 2010.

[5] Carter, K. M., Lippmann, R. P., & Boyer, S. W., "Temporally oblivious anomaly detection on large networks using functional peers" In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pp. 465-471, 2010.

[6] Carter, K. M., Idika, N., & Streilein, W. W., "Probabilistic threat propagation for malicious activity detection", *2013 IEEE International Conference on* Acoustics, *Speech and Signal Processing (ICASSP)*, pp. 2940-2944, 2013.

[7] Philips, S., Yee, M., Kao, E., & Anderson, C., "Detecting activity-based communities using dynamic membership propagation". 2012 IEEE International Conference on *Speech and Signal Processing (ICASSP)*, pp. 2085-2088, 2012.

[8] Miller, B., Bliss, N., & Wolfe, P. J., "Sub graph detection using eigenvector L1 norms", In *Advances in Neural Information Processing Systems*, pp. 1633-1641, 2010.

[9] Coskun, B., Dietrich, S., & Memon, N., "Friends of an enemy: identifying local members of peer-to-peer botnets using mutual contacts", In *Proceedings of the 26th Annual Computer Security Applications Conference*, pp. 131-140, 2010.

[10] Shah, D., & Zaman, T., "Detecting sources of computer viruses in networks: theory and experiment", In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 38, No. 1, pp. 203-21, 2010.

[11] Clauset, A., "Finding local community structure in networks", Physical *review* 2008 *E*, *72*(2), 026132, pp. 1-7, 2005.

[12] Lone, Q., Moura, G. C., & Van Eeten, M., "Towards incentivizing ISPS to mitigate bonnets", Springer Berlin Heidelberg, pp. 57-62, 2014.

[13] Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., & Feamster, N., "Building a Dynamic Reputation System for DNS", In *USENIX security symposium*, pp. 273-290, 2010.

[14] Anzai, Y., "Pattern Recognition & Machine Learning", Elsevier, pp. 139-152, 2012.

[15] Sutton, C., & McCallum, A., "An introduction to conditional random fields for relational learning", *Introduction to statistical relational learning*, pp. 93-128, 2010.

[16] Gu, G., Perdisci, R., Zhang, J., & Lee, W., "BotMiner: Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection", In *USENIX Security Symposium* Vol. 5, No. 2, pp. 139-154, 2008.

[17] Ma J., Saul, L. K., Savage, S., & Voelker, G. M., "Identifying suspicious URLs: an application of large-scale online learning", In *Proceedings of the 26th annual international conference on machine learning,* pp. 681-688, 2009.

[18] Collins, M. P., Shimeall, T. J., Faber, S., Janies, J., Weaver, R., De Shon, M., & Kadane, J., "Using uncleanliness to predict future botnet addresses", In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement,* pp. 93-104, 2007.

[19] Sinha, S., Bailey, M., & Jahanian, F., "Shades of Grey: On the effectiveness of reputation-based "blacklists", in 2008 *3rd International Conference on Malicious and Unwanted Software*, pp. 57-64, *2008*.

[20] Bilge L., Kirda, E., Kruegel, C., & Balduzzi, M., "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis, In *NDSS,* pp. 1-17, 2011.

[21] Porras, P. A., Saïdi, H., & Yegneswaran, V., "A Foray into Conficker's Logic and Rendezvous Points", In *LEET, pp*. 1-9, 2009.

[22] Collins, M. P., & Reiter, M. K. "On the limits of payload-oblivious network attack detection", In *Recent Advances in Intrusion Detection,* pp. 251-270, 2008.

[23] Roesch, M., "SNORT: Lightweight Intrusion Detection ofr Networks", In Library and Information Science Abstracts, Vol. 99, No. 1, pp. 229-238, 1999.

[24] Zhang, J., Porras, P. A., & Ullrich, J., "Highly Predictive Blacklisting", In *USENIX Security Symposium* pp. 107-122, 2008.

[25] Pathak, M. N., & Apare, R. S., "Survey On A PTP Approach to Provide Trust in Network Security for Misbehavior Detection," *International Journal of Innovations & Advancement in Computer Science, Vol. 3, Issue 9* pp. 100 –102, 2014.