Optimization Accuracy of Network IDS System using Machine Learning based SVM Technique

Sanjeev Joshi[1], Prof. Suresh. S. Gawande[2], Prof. Satyarth Tiwari[3]
M. Tech. Scholar, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal[1]
Guide, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal[2]
Co-guide, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal[3]

Abstract- These days, intrusion detection system (IDS) is the most arising pattern in our general public. This basically screen network traffic and will alarm the organization chairman of any unordinary action. IDS System work by one or the other searching for marks of known assaults or deviations of typical movement. While there are a few detriments of IDS, for example, low recognition rate and high bogus caution rate. In this paper a mixture IDS (HIDS) strategy dependent on support vector machine (SVM) and evidence theory (ET) has been proposed too different assault recognition method to limit the low bogus alert rate and improve exactness.

Keywords- **IDS, HIDS, PHAD, ALAD, SNORT**

1. INTRODUCTION

Right now, with the improvement of web advances administrations on the planet, the interlopers have been rise quickly. That is the reason there is need of IDS in the security of organization field to shield gatecrashers from approaching the data. Its primary focus on identify possible attacks, get information about them and report attempts. IDS is equivalent to a house alarm which will sound an alarm if an intruder attempts to break into window or door. Its types are varying in circumstance from respective individual computers to large networks. The proposed IDS are a HIDS. It comprises of irregularity and abuse identification module. The misuse detection first identifies the abnormal behavior for detecting and then defines all other behavior as normal. In contrast, due to rapid increase of malware the abnormality discovery is utilized to distinguish the obscure assaults which first identify all normal system behavior and then identify abnormal behavior. To deal with these attacks machine learning (ML) based SVM algorithm is used. This is administered ML calculation which is utilized for both relapse and arrangement challenges. Where, the AI (ML) calculation is additionally utilized in the field of IDS. ML is the strategy for information dissecting that robotizes scientific model building. These uses the computerized technique to "learn" instruction directly from facts without depend on prearranged equation as a model.

In this paper we have restored SNORT as abuse based interruption recognition framework just as application layer anomaly detection (ALAD), packet header anomaly detector (PHAD) as inconsistency based factual calculation. Where, PHAD (bundle header abnormality location) is the peculiarity identification strategies which first model the conventions and afterward time put together model depending with respect to the fast changes of organization. Grunt is an open source network interruption forestalls framework, capable of performing packet logging on IP network and real time traffic analysis. ALAD is anomaly detection based on application layer. Experiments and result shows that the proposed strategy improves the exactness as well as decrease the false alarm rate.

2. ML ALGORITHM

Supervised learning is two stage forms, in the initial step: a model is fabricate depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset.

2.1 Generalized Delta Learning Rule

The all out squared mistake of yield figured by net is least by inclination exact technique known as back proliferation or summed up delta rule.

2.2 Derivation

Consider the discretionary initiation work f(x). The inference of the initiation work is signified by F(x).

$$y_{-ink} = \sum_i z_i w_{jk} \quad (1)$$

$$Z_{-inJ} = \sum_i V_{ij} X_i \quad (2)$$

$$Y_k = f(y_{-ink}) \quad (3)$$

The error to be minimized is

$$E = 0.5 \sum_k [t_k - y_k]^2 \quad (4)$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}} 0.5 \sum_k [t_k - y_k]^2 \quad (5)$$

$$= \frac{\partial}{\partial w_{jk}} (0.5[t_k - t(y_{ink})]^2) \quad (6)$$

$$= -[t_k - y_k] \frac{\partial}{\partial w_{jk}} f(y_{-ink}) \quad (7)$$

$$= -[t_k - y_k] f(y_{-ink}) \frac{\partial}{\partial w_{jk}} (y_{ink}) \quad (8)$$

$$= [t_k - y_k] f^1(y_{ink}) Z_j \quad (9)$$

Let us define

$$\delta_k = -[t_k - y_k] f^1(y_{-ink}) \quad (10)$$

Weight on connection to the hidden unit Z_j

$$\frac{\partial E}{\partial v_{ij}} = -\sum_k [t_k - y_k] f(y_{ink}) \frac{\partial}{\partial v_{ij}} y_k \quad (11)$$

$$= \sum_k [t_k - y_k] f(y_{ink}) \frac{\partial}{\partial v_{ij}} y_{-ink} \quad (12)$$

$$= \sum_k \delta_k \frac{\partial}{\partial v_{ij}} y_{-ink} \quad (13)$$

A high learning rate prompts fast adapting yet the loads may sway, while a ML rate prompts SVM. Strategies recommended for receiving learning rate as follows;

1. Start with a high learning rate and reliably decrease it. Changes in the weight vector should be little keeping in mind the end goal to lessen motions or any difference.
2. A little proposal is to expand the learning rate with a particular ultimate objective to improve execution and to decrease the learning rate remembering the ultimate objective to orsen the execution.
3. Another technique is to twofold the learning rate until the point that the blunder esteem intensifies.

3. PROPOSED METHODOLOGY

In this section, the brief explanation of overall process of proposed methodology is described.

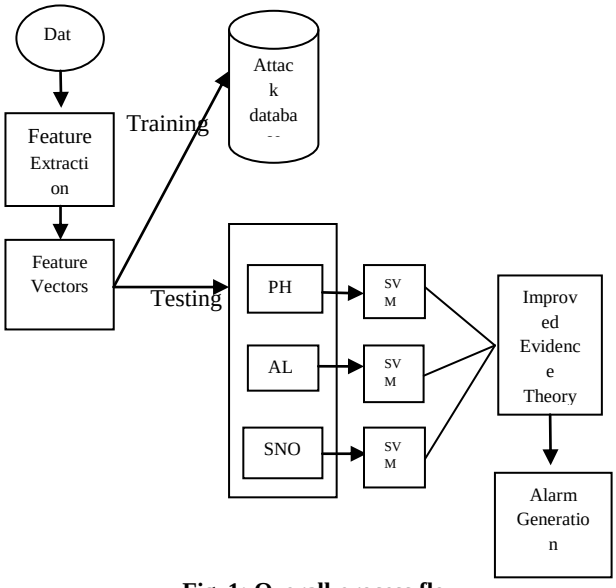

Fig. 1: Overall process flow

The exhibition of HIDS is improved by characterizing the recognition aftereffects of PHAD, ALAD and SNORT utilizing Semi Supervised ML method and a ultimate conclusion making measure is improved by presenting improved proof hypothesis. As in the past work of the exploration, the highlights are extricated from the info dataset utilizing PLS strategy.

At that point the individual choice consequences of PHAD, ALAD and SNORT are grouped based Semi Supervised ML procedure. In the AI method, SVM has gotten one of the famous calculations utilized for interruption discovery because of the capacity to defeat the scourge of dimensionality and their great speculation nature. At that point the arranged outcomes are consolidated utilizing Dempster-Shafer ET for ultimate choice making. Anyway the proportion of contention issue in Dempster-Shafer ET corrupts the general presentation by bringing about more bogus positives. Subsequently an improved ET is proposed. The improved ET at first builds proof uniqueness network utilizing the weighted Euclidean distance. The divergence between the confirmations is estimated and afterward at long last utilizing the disparity network, supporting degree, validity and weight of proof are determined, and the first confirmations are altered. Subsequently the proposed HIDS method with SVM and improved ET improves the identification precision.

HIDS SVM Algorithm

The contention in proof as in Dempster-Shafer technique delivered non-natural outcomes. The contention offered raise to vulnerability in the past methodology. No reasonable end can be attracted one stage when strife emerges. Ultimate choice making needs assortment of extra confirmations. Likewise it is smarter to total all accessible proof without smothering any as if there should

be an occurrence of Dempster-Shafer technique. This is done in this move toward utilizing improved proof hypothesis alongside SVM classifier.

Input: Dataset, Evidence vector E=E_1, E_2.....En
Output: detect attacks

Step 1- Collect the dataset from the network
Step 2- Detect the intrusions in data using PHAD, ALAD and SNORT
Step 3- Classify the results using SVM
Step4- Begin
Step 5- Calculate dissimilarity matrix using equation

$$d_{i,j}(E_i, E_j) = \begin{cases} 0, & C_1 = C_2 \\ (w_1|A_1 - B_1|^2 + w_2|A_2 - B_2|^2 + \\ \ldots + w_n|A_n - B_n|^2)^{1/2}, & C_1 \neq C_2 \end{cases}$$

Where E_i and E_j are two instances evidences of the detection results if discrement Θ, w_i is the evidences, C_1 and C_2 and are the preposition values. There are n evidences for attack detection, then the above formula is used to compute the dissimilarity between E_i and E_j.

Step 6- for i=1 to n//n is the number of evidences
Step 7- Calculate entropy of evidences using equation

$$Entropy_i = SupportingDegree(m_i)_i$$
$$\ln(SupportingDegree(m_i))$$

The data entropy of proof is corresponding to its disparity supporting degree. The validity of the proof is calculated by normalizing the entropy value of the evidence.

Step 8- Calculate credibility of evidence using equation

$$Credibilty_{m_i} = \frac{1/Entropy_i}{\sum_{i=1}^{n} 1/Entropy_i} \quad i = 0,1,2,....n$$

The sum of the $credibility_{mi}$ is equal to 1, which means $credibility_{mi}$ can be viewed as the heaviness of mi. In the wake of getting the heaviness of proof, the proof can be weighted and the acquired essential certain task estimations of improved E_i which is given in equation

Step 9- Calculate the weight of evidences using

$$m'(A) = Credibilty_{m_i}.m(A), m'(\theta) = 1 - \sum_{i=1}^{n} m'(A)$$

Step 10- end for
Step 11- for i=1 to n
Step 12- if n>2
Step 13- $e_x=e_i$
Step 14- $e_y=e_{i+1}$
Step 15- Apply improved combination rule of evidence theory using previous equation
Step 16- if i== n-1
Step 17- return result
Step 18- end if
Step 19- i=i+1
Step 20- end if
Step 21- end for
Step 22- End

4. RESULT ANALYSIS

In this paper training and testing dataset are used to analyses the proposed IDS. In this section, we show the comparison result of the attack types i.e., prob, DOS, R2L, U2R and IDS attack types i.e., PHAD, ALAD, SNORT, HIDS SVM. Table 1 represented by four types of attack and detection percentage. Overall 10000 attacks are applied HIDS system and detected 7160 attacks.

Table 1: Attacks detected by HIDS SVM

Attack Type	Total attacks	Attacks detected	% detection
Probe	3456	2422	70.08%
DoS	2052	1345	65.54%
R2L	3078	2135	69.36%
U2R	1414	1167	82.53%
Total	**10000**	**7069**	**70.69%**

Table 2 represented by different parameter in different technique. Recall and precision is calculated true positive (TP) and false positive (FP) [13].
The recall or sensitivity gives the true positive rate TPR which is defined as the proportion of the attacks that were correctly identified and calculated using the following formula

$$\text{Re}call = \frac{TP}{TP + FN}$$

Precision is also termed as positive predictive value which is defined as the proportion of the predicted attacks that were correct and it can be computed by using the following formula

$$\text{Precision} = \frac{TP}{TP + FP}$$

Table 2: Comparison Result

IDS/ attacks types	Total attacked	TP	FP	Recall	Precision
PHAD	10000	3198	6928	0.27	0.21
ALAD	10000	3399	6489	0.37	0.23
SNORT	10000	3608	6288	0.42	0.27
HIDS SVM	10000	3898	6178	0.52	0.32

Hence, by comparing the above two table it is clear that the IDS attacks types gives better result for attack detection and prevention. Table 3.shows the accuracy and false alarm rate (FAR) parameter is to determine the comparison between the previous method and proposed methodology.

Table 3: Comparison base paper

IDS/attack types	Accuracy	FAR
Previous method	99.19%	0.76%
HIDS SVM Algorithm	99.32%	0.69%

The comparison shows that the proposed method gives high accuracy and low false rate compare to previous method.

5. CONCLUSION

HIDS with include choice and extraordinary classifier technique is proposed which improves exactness of interruption location measure by information decrease. At first, the applicable highlights are extricated and those highlights are given to HIDS framework. The greater part of the IDS look at all the highlights that is excess for interruption recognition. This methodology thinks about the significant highlights in the info traffic for interruption discovery framework that is computationally proficient and viable. The most significant highlights of PHAD, ALAD and SNORT are chosen by utilizing LDA method. At that point the chose highlights are given to SVM classifier. This classifier groups the identification results. At long last, the arranged recognition consequences of every IDS are consolidated by utilizing improved proof hypothesis. Through element choice and SVM characterization measure the interruption location with higher review and exactness values are accomplished. The trial result shows that the proposed HIDS LDA- SVM-IEV strategy are and strong in addressing the information as it was capable to lessen the information and henceforth essentially decreases the time needed to recognize the assaults in the organization traffic.

REFERENCES-

[1] Abhinav Singhal, Akash Maan, Daksh Chaudhary and Dinesh Vishwakarma, “A Hybrid Machine Learning and Data Mining Based Approach to Network Intrusion Detection”, Proceedings of the International Conference on Artificial Intelligence and Smart Systems, IEEE 2021.

[2] Lan Liu, Pengcheng Wang , Jun Lin, and Langzhou Liu, “Intrusion Detection of Imbalanced Network Traffic Based on Machine Learning and Deep Learning”, IEEE Access 2020.

[3] A. Raghavan, F. D. Troia, and M. Stamp, ``Hidden Markov models with random restarts versus boosting for malware detection," *J. Comput. Virol. Hacking Techn.*, vol. 15, no. 2, pp. 97107, Jun. 2019.

[4] Zhiyou Zhang and Peishang Pan “A hybrid intrusion detection method based on improved fuzzy C-Means and SVM”, IEEE International Conference on Communication Information System and Computer Engineer (CISCE), pp. no. 210-214, Haikou, China 2019.

[5] Afreen Bhumgara and Anand Pitale, “Detection of Network Intrusion Using Hybrid Intelligent System”, IEEE International Conferences on Advances in Information Technology, pp. no. 167-172, Chikmagalur, India 2019.

[6] Ritumbhira Uikey and Dr. Manari Cyanchandani “Survey on Classification Techniques Applied to Intrusion Detection System and its Comparative Analysis”, IEEE 4th International Conference on Communication $ Electronics System (ICCES), pp. no. 459-466, Coimbatore, India 2019.

[7] Aditya Phadke, Mohit Kulkarni, Pranav Bhawalkar and Rashmi Bhattad “A Review of Machine Learning Methodologies for Network Intrusion Detection”, IEEE 3rd National Conference on Computing Methodologies and Communication (ICCMC), pp. no. 703-709, Erode, India 2019.

[8] S. Sivantham, R.Abirami and R.Gowsalya “Comapring in Anomaly Based Intrusion Detection System for Networks”, IEEE International conference on Vision towards Emerging Trends in Communication and Networking (ViTECon), pp. no. 289-293, Coimbatore, India 2019.

[9] Azar Abid Salih and Maiwan Bahjat Abdulrazaq “Combining Best Features selection Using Three Classifiers in Intrusion Detection System”, IEEE International Conference on Advanced science and Engineering (ICOASE), pp. no. 453-459, Zakho - Duhok, Iraq 2019.

[10] Lukman Hakim and Rahilla Fatma Novriandi “Influence Analysis of Feature Selection to Network Intrusion Detection System Performance Using NSL-KDD Dataset”, IEEE International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), pp. no. 330-336, Jember, Indonesia 2019.

[11] T. Sree Kala and A. Christy, “An Intrusion Detection System Using Opposition Based Particle Swayam Optimization Algorithm and PNN”, IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, pp. no. 564-569, Coimbatore, India 2019.

[12] Xiaoyan Wang and Hanwen Wang "A High Performance Intrusion Detection Method Based on Combining Supervised and Unsupervised Learning", IEEE Smart World, Ubiquitous Intelligence $ Computing Advanced $ Trusted Computing, Scalable Computing, Internet of People and Smart City Innovations, pp. no. 889-897, Guangzhou, China 2018.

[13] P. Singh and M. Venkatesan, "Hybrid Approach for Intrusion Detection System", IEEE International Conference on Current Trends Towards Converging Technologies (ICCTCT), pp. no. 654-659, Coimbatore, India 2018.

[14] M. Tavallaee, E. Bagheri, W, Lu and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 dataset", IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. no. 892-899, Ottawa, India 2018.

[15] Karuna S. Bhosale and Assoc. prof. Maria, "Data Mining Based Advanced Algorithm for Intrusion Detection in Communication Networks", IEEE International Conference on Computational Techniques, Electronics & Mechanical System (CTEMS), Belgaum, India 2018.