IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

CYBER ATTACK DETECTION

[1]Munish Katoch ,[2]Sushanth Jakkani , [3]Rohith Thokala
[1]Associate Professor, Dept. of CSE. Lovely Professional University, Jalandhar, Punjab
[2,3] B.Tech Students, Dep. of CSE. Lovely Professional University, Jalandhar, Punjab

Abstract:
Cyber-crime is proliferating everyplace exploiting all types of vulnerability to the computing atmosphere. moral Hackers pay additional attention towards assessing vulnerabilities and recommending mitigation methodologies. the event of effective techniques has been associate pressing demand within the field of the cyber security community. Most techniques utilized in today's IDS aren't able to modify the dynamic and complicated nature of cyber-attacks on pc networks. Machine learning for cyber security has become a difficulty of nice importance recently thanks to the effectiveness of machine learning in cyber security problems. Machine learning techniques are applied for major challenges in cyber security problems like intrusion detection, malware classification and detection, spam detection and phishing detection. though machine learning cannot alter an entire cyber security system, it helps to spot cyber security threats additional expeditiously than alternative software-oriented/ methodologies, and therefore reduces the burden on security analysts. Hence, economical adaptative ways like varied techniques of machine learning may end up in higher detection rates, lower warning rates and affordable computation and communication prices. Our main goal is that the task of finding attacks is essentially completely different from these alternative applications, creating it considerably more durable for the intrusion detection community to use machine learning effectively.

Keywords: **Machine Learning, Intrusion detecting system.**

I. INTRODUCTION

Great applied technology sometimes needs sanctionative partner technology, and it will struggle to form headway until that partner looks. for several years, Intrusion Detection System (IDS) technology struggled to deliver economical, high-quality intrusion observance, associated is barely presently experiencing success with the arrival of AN unintentional sanctionative partner technologies like cloud computing, machine learning & AI.

During the late 1980's, with a growing kind of shared networks, enterprise system directors all over the world began adopting intrusion detection systems. However, IDS conferred a couple of problems. First, it's going to exclusively alert on wonderful issues that had been classified as threats on a signature list; zero-day attacks might compromise a network's security. Second, the constant scanning and alter of a signature list was cumbersome and necessary resource drain.

In the 1990's, IDS technology improved to handle the increasing selection and sophistication of network attacks. This new technique, named anomaly detection, relied on distinctive uncommon activity patterns on the network, and provided alerts for any notable abnormality.

Unfortunately, the inconsistent nature of networks through the 1990's and early 2000's resulted in AN extremely high kind of false positives, and plenty of administrators thought IDS to be unreliable, and headed for a slow death.

The advent of cloud computing & machine learning, however, has brought new affiliation to IDS systems, resulting in a surge among the IDS market. a very important a part of today's security best practices, IDS systems square measure designed to sight attacks which will occur, despite preventative measures. In fact, IDS is presently one in all the best mercantilism security technologies and foreseen to still gain momentum. After all, security — cloud security specially – is far too advanced to be monitored manually.

Big information to boot plays a significant role among the expansion and importance of intrusion detection these days. The world's info doubles every twenty months, and as cloud hosted databases expand exponentially, it's no marvel IDS may be a ton of important than ever.

By combining AI, machine learning, neural networks and similar advances in programming, we have a tendency to square measure able to forestall to Intrusion Detection Systems that will not exclusively raise the alarm, but take the suitable action to thwart the attack.

II. LITERATURE SURVEY

The intrusion detection technology can be divided into three major categories: pattern matching methods, traditional machine learning methods and deep learning methods. At the beginning, people mainly use pattern matching algorithms for intrusion detection. Pattern matching algorithm is the core algorithm of intrusion detection system based on feature matching. Most algorithms have been considered for use in the past. In, the authors make a summary of pattern matching algorithm in Intrusion Detection System: KMP algorithm, BM algorithm, BMH algorithm, BMHS algorithm, AC algorithm and AC-BM algorithm. Experiments show that the improved algorithm can accelerate the matching speed and has a good time performance. In, Naive approach, Knuth-Morris Pratt algorithm and Rabin Karp Algorithm are compared in order to check which of them is most efficient in pattern/intrusion detection. Pcap files have been used as datasets in order to determine the efficiency of the algorithm by taking into consideration their running times respectively.

These traditional pattern recognition algorithms have serious defects, which cannot achieve the effect of intrusion detection. Finding an efficient algorithm that reaches high efficiency and low false positive rates is still the focus of current work. With the development of artificial intelligence, the application of intelligent algorithms for intrusion detection has become a new research hotspot. The traffic anomaly detection methods based on machine learning have achieved a lot of success. In the authors propose a new method of feature selection and classification based on support vector machine (SVM). Experimental results on NSL-KDD cup 99 of intrusion detection data set showed that the classification accuracy of this method with all training features reached 99%. In , the authors combine k-mean clustering on the basis of KNN classifier. The experimental results on NSL-KDD dataset show that this method greatly improves the performance of KNN classifier.

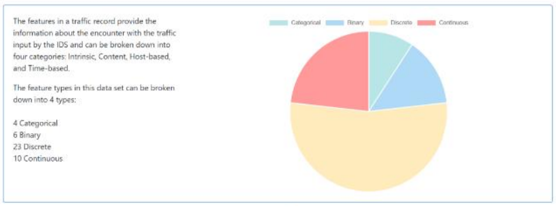
In, the authors propose a new framework to combine the misuse and the anomaly detection in which they apply the random forests algorithm. Experimental results show that the overall detection rate of the hybrid system is 94.7% and the overall false positive rate is 2%. In, the performance of NSL-KDD dataset is evaluated via Artificial Neural Network (ANN). The detection rate obtained is 81.2% and 79.9% for intrusion detection and attack type classification task respectively for NSL-KDD dataset. In, an intrusion detection method based on decision tree (DT) is proposed. Experimental results of feature selection using the relevant feature selection (CFS) subset evaluation method show that the DT based intrusion detection system has a higher accuracy. As described above, machine learning methods have been proposed and have achieved success for an intrusion detection system. However, these methods require large-scale preprocessing and complex feature engineering of traffic data. It is impossible to solve the massive intrusion data classification problem using machine learning methods.

Proposed system

We propose an end-to-end deep learning model logistic regression that is composed of logistic regression and attention mechanism. logistic regression can well solve the problem of intrusion detection and provide a new research method for intrusion detection.

We compare the performance of logistic regression with traditional deep learning methods, the model can extract information from each packet. By making full use of the structure information of network traffic, the logistic regression model can capture features more comprehensively. 4) We evaluate our proposed network with a real NSLKDD dataset. The experimental results show that the performance of algorithm is better than the traditional methods.

III. METHODOLOGY:


The features in a traffic record provide the information about the encounter with the traffic input by the IDS and can be broken down into four categories: Intrinsic, Content, Host-based, and Time-based.

The feature types in this data set can be broken down into 4 types:

4 Categorical
6 Binary
23 Discrete
10 Continuous

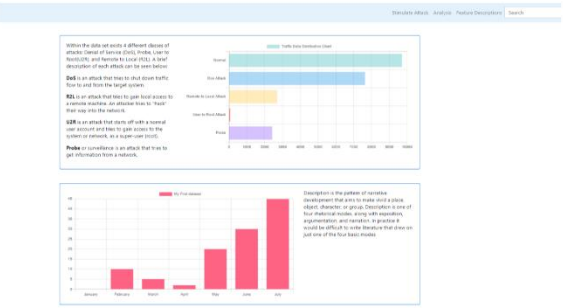
(Fig 7.3. Dataset divided into four parts)

Data Preprocessing Layer: There are three symbolic data types in NSL-KDD data features: protocol type, flag and service. We use one-hot encoder mapping these features into binary vectors. One-Hot Processing: NSL-KDD dataset is processed by one-hot method to transform symbolic features into numerical features. For example, the second feature of the NSL-KDD data sample is protocol type. The protocol type has three values:

tcp, udp, and icmp. One-hot method is processed into a binary code that can be recognized by a computer, where tcp is [1, 0, 0], udp is [0, 1, 0], and icmp is [0, 0, 1]

Normalization Processing: The value of the original data may be too large, resulting in problems such as ‘‘large numbers to eat decimals’’, data processing overflows, and inconsistent weights so on. We use standard scaler to normalize the continuous data into the range [0, 1]. Normalization processing eliminates the influence of the measurement unit on the model training, and makes the training result more dependent on the characteristics of the data itself.

IV.IMPLEMENTATION:



Within the data set exists 4 different classes of attacks: Denial of Service (DoS), Probe, User to Root(U2R), and Remote to Local (R2L). A brief description of each attack can be seen below:

DoS is an attack that tries to shut down traffic flow to and from the target system.

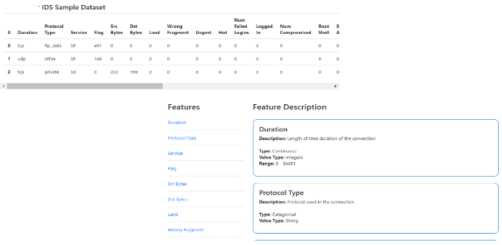
R2L is an attack that tries to gain local access to a remote machine. An attacker tries to "hack" their way into the network.

U2R is an attack that starts off with a normal user account and tries to gain access to the system or network, as a super-user (root).

Probe or surveillance is an attack that tries to get information from a network.

Description is the pattern of narrative development that aims to make vivid a place, object, character, or group. Description is one of four rhetorical modes, along with exposition, argumentation, and narration. In practice it would be difficult to write literature that drew on just one of the four basic modes.

(Fig 7.1. Traffic Data Distribution chart)



IDS Sample Dataset

	Duration	Protocol Type	Service	Flag	Src Bytes	Dst Bytes	Land	Wrong Fragment	Urgent	Hot	Num Failed Logins	Logged In	Num Compromised	Root Shell	S A
0	tcp	ftp_data	SF	491	0	0	0	0	0	0	0	0	0	0	0
1	udp	other	SF	146	0	0	0	0	0	0	0	0	0	0	0
2	tcp	private	S0	0	232	199	0	0	0	0	0	0	0	0	0

Features

- Duration
- Protocol Type
- Service
- Flag
- Src Bytes
- Dst Bytes
- Land
- Wrong Fragment

Feature Description

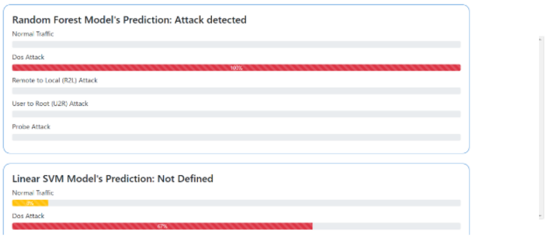
Duration
Description: Length of time duration of the connection
Type: Continuous
Value Type: Integers
Range: 0 - 54451

Protocol Type
Description: Protocol used in the connection
Type: Categorical
Value Type: String

(Fig 7.2. Sample Dataset)

Conclusion

Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. We reviewed several influential algorithms for intrusion detection based on various machine learning techniques. Characteristics of ML techniques makes it possible to design IDS that have high detection rates and low false positive rates while the system quickly adapts itself to changing malicious behaviors. IDS using many Machine Learning Techniques like Random Forest, Decision tree and logistic regression to perform better in various metrics. The IDS should provide the most effective solutions based on the requirements. One thing is sure, any company failing to adopt these techniques now or in the immediate future risk compromising data or worse servers... Future Enhancements: In Future we can detect IoT network attacks by using machine learning methods. In this context, the Bot IoT [9] was used as a dataset because of its regular updates, wide attack diversity, and various network protocols. We used Cyclometer to extract flow-based features from the raw traffic traces. Cyclometer generates 84 network traffic features of the dataset which define the network flow. During the implementation, the importance of weight calculations were made with the Random Forest Regressor algorithm to decide which of the features would be used in the machine learning method.



Random Forest Model's Prediction: Attack detected

Normal Traffic
Dos Attack 100%
Remote to Local (R2L) Attack
User to Root (U2R) Attack
Probe Attack

Linear SVM Model's Prediction: Not Defined

Normal Traffic 8%
Dos Attack 67%

(Fig 7.6. Final Result of Attack Detection)

References:

[1] B. B. Zarpelo, R. S Miani, C. T. Kawakani, and S. C. de Alvarenga, ''A survey of intrusion detection in Internet of Things,'' J. Netw. Comput. Appl., vol. 84, pp. 25– 37, Apr. 2017.

[2] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, ''Network intrusion detection,'' IEEE Netw., vol. 8, no. 3, pp. 26–41, May 1994.

[3] S. Kishorwagh, V. K. Pachghare, and S. R. Kolhe, ''Survey on intrusion detection system using machine learning techniques,'' Int. J. Control Automat., vol. 78, no. 16, pp. 30–37, Sep. 2013.

[4] N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad, ''Survey on SDN based network intrusion detection system using machine learning approaches,'' Peer-to-Peer Netw. Appl., vol. 12, no. 2, pp. 493–501, Mar. 2019.

[5] M. Panda, A. Abraham, S. Das, and M. R. Patra, ''Network intrusion detection system: A machine learning approach,'' Intell. Decis. Technol., vol. 5, no. 4, pp. 347– 356, 2011.

[6] W. Li, P. Yi, Y. Wu, L. Pan, and J. Li, ''A new intrusion detection system based on KNN classification algorithm in wireless sensor network,'' J. Electr. Comput. Eng., vol. 2014, pp. 1–8, Jun. 2014.

[7] S. Garg and S. Batra, ''A novel ensembled technique for anomaly detection,'' Int. J. Commun. Syst., vol. 30, no. 11, p. e3248, Jul. 2017.

[8] F. Kuang, W. Xu, and S. Zhang, ''A novel hybrid KPCA and SVM with GA model for intrusion detection,'' Appl. Soft Comput.., vol. 18, pp. 178–184, May 2018