International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 12
ISSN: 2454-4248
102 – 104

Intrusion Detection System Using Feature Selection and classifier based Algorithm

Pradeep Laxkar
PhD Scholar , CSE Department
SPSU Udaipur
Udaipur, India
pradeep.laxkar@gmail.com

Prof. Prasun Chakrabarti
HOD , CSE Department
SPSU Udaipur
Udaipur, India
prasun.chakrabarti@spsu.ac.in

***Abstract*—** With the enlargement of web, there has been a terrific increases in the number of attacks and therefore Intrusion Detection Systems (IDS's) has become a main topic of information security. The purpose of IDS is to help the computer systems to deal with attacks. The feature selection used in IDS helps to reduce the classification time. In this paper, the IDS for detecting the attacks efficiently has been proposed. We have proposed an algorithm based on associan rule to detect intrusion. We have combined algorithm with feature selection to improve efficiency of IDS.The proposed feature selection and associan rule algorithms enhance the performance of the IDS in detecting the attacks.

Keywords*—IDS , Feature selection , attack , threats.*

*****

I . INTRODUCTION

An intrusion detection system (IDS) monitors network traffic and monitors for suspicious activity and alerts the system or network administrator. In some cases, the IDS may also respond to anomalous or malicious traffic by taking action such as blocking the user or source IP address from accessing the network[6].

IDS come in a variety of "flavors" and approach the goal of detecting suspicious traffic in different ways.

There are network based (NIDS) and host based (HIDS) intrusion detection systems. There are IDS that detect based on looking for specific signatures of known threats- similar to the way antivirus software typically detects and protects against malware- and there are IDS that detect based on comparing traffic patterns against a baseline and looking for anomalies. There are IDS that simply monitor and alert and there are IDS that perform an action or actions in response to a detected threat. We'll cover each of these briefly.

NIDS

Network Intrusion Detection Systems are placed at a strategic point or points within the network to monitor traffic to and from all devices on the network. Ideally, you would scan all inbound and outbound traffic, however doing so might create a bottleneck that would impair the overall speed of the network.

HIDS

Host Intrusion Detection Systems are run on individual hosts or devices on the network.

An HIDS monitors the inbound and outbound packets from the device only and will alert the user or administrator of suspicious activity is detected

Signature Based

A signature based IDS will monitor packets on the network and compare them against a database of signatures or attributes from known malicious threats.

Anomaly Based

An IDS which is anomaly based will monitor network traffic and compare it against an established baseline. The baseline will identify what is "normal" for that network- what sort of bandwidth is generally used, what protocols are used, what ports and devices generally connect to each other- and alert the administrator or user when traffic is detected which is anomalous, or significantly different than the baseline.

Passive IDS

A passive IDS simply detects and alerts. When suspicious or malicious traffic is detected an alert is generated and sent to the administrator or user and it is up to them to take action to block the activity or respond in some way.

Reactive IDS

A reactive IDS will not only detect suspicious or malicious traffic and alert the administrator but will take pre-defined proactive actions to respond to the threat. Typically this means blocking any further network traffic from the source IP address or user[6].

Feature Selection

As many pattern recognition[7] techniques were originally not designed to cope with large amounts of irrelevant features, combining them with FS techniques has become a necessity in many applications.. The objectives of feature selection are manifold, the most important ones being:

(a) to avoid overfitting and improve model performance, i.e. prediction performance in the case of supervised

 ISSN: 2454-4248
102 – 104

classification and better cluster detection in the case of clustering.
(b) to provide faster and more cost-effective models.
(c) to gain a deeper insight into the underlying processes that generated the data.

However, the advantages of feature selection techniques come at a certain price, as the search for a subset of relevant features introduces an additional layer of complexity in the modelling task. Instead of just optimizing the parameters of the model for the full feature subset, we now need to find the optimal model parameters for the optimal feature subset, as there is no guarantee that the optimal parameters for the full feature set are equally optimal for the optimal feature subset .As a result, the search in the model hypothesis space is augmented by another dimension: the one of finding the optimal subset of relevant features. Feature selection techniques differ from each other in the way they incorporate this search in the added space of feature subsets in the model selection.
In the context of classification, feature selection techniques can be organized into three categories, depending on how they combine the feature selection search with the construction of the classification model: filter methods, wrapper methods and embedded methods.

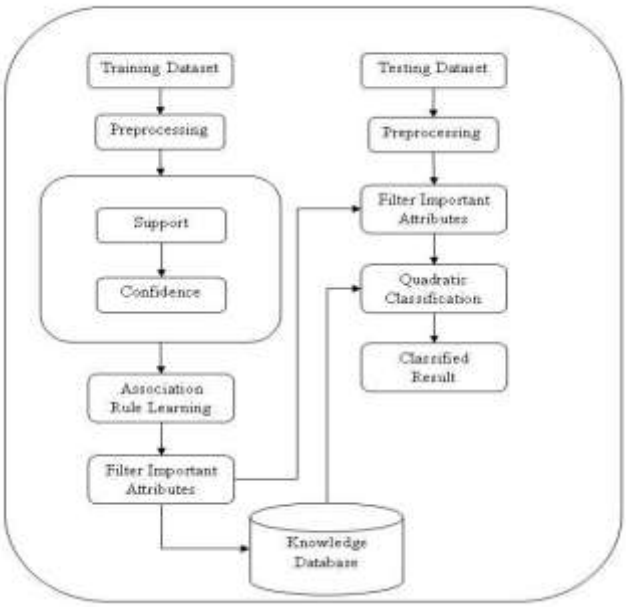
Association Rules Mining
Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Rakesh Agrawal et al. introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule $\{\text{onions}, \text{potatoes}\} \Rightarrow \{\text{burger}\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection and bioinformatics. As opposed to sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

.

[1] LITERATURE REVIEW

1. Amiri, Fatemeh & Rezaei Yousefi, Mohammadmahdi & Lucas, Caro & Shakery, Azadeh & Yazdani, Nasser[1] proposed a feature selection phase, which can be generally implemented in any intrusion detection system. In this work, authors proposed two feature selection algorithms and study the performance of using these algorithms compared to a mutual information-based feature selection method. These feature selection algorithms require the use of a feature goodness measure. They investigate using both a linear and a non-linear measure—linear correlation coefficient and mutual information, for the feature selection.
2. In this paper[2] authors evaluated three fuzzy rule-based classifiers to detect intrusions in a network. Results are then compared with other machine learning techniques like decision trees, support vector machines and linear genetic programming. Authors also modeled Distributed Soft Computing-based IDS (D-SCIDS) as a combination of different classifiers to model lightweight and more accurate (heavy weight) IDS.
3. Authers[3] investigated the performance of two feature selection algorithms relating Bayesian networks (BN) and Classification and Regression Trees (CART) and an ensemble of BN and CART.The discussed that significant input feature selection is significant to design an IDS that is lightweight, efficient and effective for real world detection systems.
4. The intention of authors in this paper[4] is to develop effective machine learning or data mining techniques based on flexible neural tree FNT. Based on the pre-defined instruction/operator sets, a flexible neural tree model can be created and evolved. They applied for two real-world problems involving designing intrusion detection system (IDS) and for breast cancer classification. The IDS data has 41 inputs/features and the breast cancer classification problem has 30 inputs/features. Experiential results indicate that the proposed method is efficient for both input feature selection and improved classification rate.
5. In this paper[5], a new hybrid intrusion detection method that hierarchically integrates a misuse detection model and an anomaly detection model in a decomposition structure is proposed. The proposed hybrid intrusion detection method was evaluated by conducting experiments with the NSL-KDD data set, which is a modified version of well-known KDD Cup 99 data set.

[2] PROPOSED SYSTEM



 ISSN: 2454-4248
102 – 104

In proposed system, first of all training dataset is given as input.Dataset is preprocessed. In preprocessing noise data are removed from the dataset. Noise removal is used to improve the classification. After that support and confidence values are calculated. Support and confidence and are used to find the association rules. Association rules are used to find the relationship between the attributes. Strong association rules are considered for the attribute filtering. So the attributes have high association rules are filtered. And the filtered data for the attributes are stored in the knowledge database. Then testing dataset is given as input to the system. Preprocessing is also done for the testing data. Then filter the attributes from the testing data. Filtered testing data and knowledge database are given as input to quadratic classification. Classification gives result as intrusion.

IV. IMPLEMENTATION

For implementation we will use NSL KDD data set. For real timr data set we will use wireshark. Wire shark is tool which collect data and store in .pcap formate. We will convert data .pcap to .csv formate using tshark tool.Implementation will be using java. We will implement our algoriyhm in java.

V . CONCLUSION

In this work, a new IDS has been proposed and under implementation by merging an l Feature Selection algorithm and classifier techniques for intrusion detection. We can icrease intrusion detection system accuracy up to 95% using this approach .We can also reduces the false positive rates and also reduces the computation time.

FUTURE WORK

We can use spark's machine learning library to process large amount of data. Same algorithm can be used in spark enviorment for better results.

REFERENCES

[1] Amiri, Fatemeh & Rezaei Yousefi, Mohammadmahdi & Lucas, Caro & Shakery, Azadeh & Yazdani, Nasser. (2011). Mutual information-based feature selection for intrusion detection systems. J. Network and Computer Applications. 34. 1184-1199. 10.1016/j.jnca.2011.01.002.

[2] Abraham, Ajith. (2007). D-SCIDS: Distributed Soft Computing Intrusion Detection Systems. Journal of Network and Computer Applications. 30. 81-98. 10.1016/j.jnca.2005.06.001.

[3] Chebrolu, Srilatha & Abraham, Ajith & P. Thomas, Johnson. (2005). Feature deduction and ensemble design of intrusion detection systems. Computers & Security. 24. 295-307. 10.1016/j.cose.2004.09.008.

[4] Chen, Yuehui & Abraham, Ajith & Yang, Bo. (2006). Feature selection and classification using flexible neural tree. Neurocomputing. 70. 305-313. 10.1016/j.neucom.2006.01.022.

[5] Kim, Gisung & Lee, Seungmin & Kim, Sehun. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications. 41. 1690–1700. 10.1016/j.eswa.2013.08.066.

[6] Tony Bradley, CISSP, MCSE2k, MCSA, A+(2017), Introduction to Intrusion Detection Systems (IDS) , https://www.lifewire.com/introduction-to-intrusion-detection-systems-ids-2486799.

[7] Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics, Bioinformatics , 2007, vol. 23 (pg. 2507-17)