IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 8 (August. 2018), ||V (VI) || PP 01-05

Fast and Effective Network Anomaly Detection Technique Using Hybrid Sequential Pattern algorithms

Dr.P.Muhamed Ilyas [1], Arshad PT [2]
Principal, SS Arts and Science College, Area code, India [1]
Assistant Professor, Department of Computer Science, SAFI Institute of Advanced Study, Vazhayoor, India[2]
Corresponding Author: Dr.P.Muhamed Ilyas

Abstract: Anomaly Detection is a common and dynamic approach in the current network scenario. The securing and safe a network (Secured network) is top most major challenging task, as anomalies which indicate suspicious behaviors, attacks, network malfunctions, or network failures. There are several algorithms and techniques have been used to overcome the detecting the anomalous events and attribution. However, the techniques cannot be efficiently detected anomalies when the number of parallel flows is very large and which of the flows are anomalous. To overcome the issue, an effective adaptive hybrid framework is proposed. The technique is named as Hybrid Sequential Pattern Anomaly Detection (H-SPAD). The H-SPAD includes (i).Sequential Pattern analysis algorithms and Modified SVM with semi supervised nature. The technique is quite adequate for measure Anomaly traffic detection. The experiments are carried with several Intrusion Detection Evaluation dataset and achieved detecting the anomalous events accuracy with good results.

Index Terms: Anomaly detection, Network Traffic, Attribution, cross entropy method.

Date of Submission: 16-08-2018 Date of acceptance: 03-09-2018

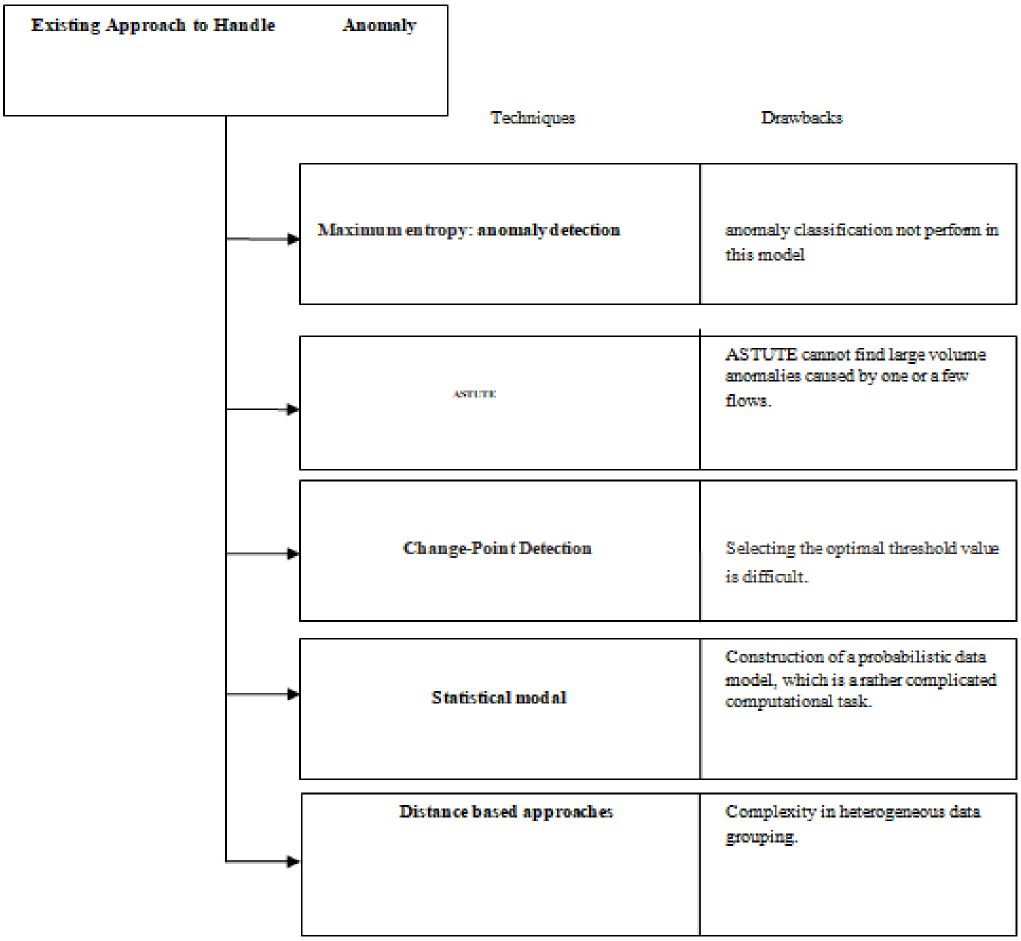
I. INTRODUCTION

Today's communication networks evolve quickly. The same is true for network attacks. New vulnerabilities appear each and every day. This poses a serious securing risk in a network environment. Securing a network which means effectively identify attacks or malicious activities, network scans for vulnerable ports/services, network failures and suspicious behaviors so introduce secure mechanism name called as Anomaly detection [1]. This aims at finding the presence of anomalous patterns in network traffic. This dynamic approach is one of the parts of the network behavior analysis and it offers an additional layer of the security to that provided by traditional anti-threat applications. In data mining, clustering, classification and machine based learning techniques were used to handle anomalies in the network. There are different types of Anomaly detection techniques are proposed, so these techniques are known as maximum entropy [2]. This approach is every time comparing the current network congestion opposes a baseline distribution in the time of network traffic, so it easily detects the network anomalies. Use the maximum entropy framework by estimating the benign traffic (Not traffic congestion) in the time of packet distribution. So, it used to baseline to detect the anomalies. The method is to detect anomalies by analyze only the current traffic instead of an approach called "change point detection approach".

Other widely used techniques ASTUTE, a new method for traffic anomaly detection in network links. This does not need to learn normal traffic behavior from traffic traces. Instead, ASTUTE relies only on empirical traffic properties which hold for highly aggregated network links. ASTUTE is specialized in a class of anomalies (strongly correlated flows); it is accurate but not easier to determine the type of anomaly in case number of parallel flows is very large [3]. The authors of [4] proposed a change-point detection technique for detecting network anomalies. This model based on the multi-cyclic (repeated) Shiryaev–Roberts detection procedure where the likelihood ratio is replaced with the linear-quadratic score. This approach will allow filtering false alarms reducing the false alarm rate to a minimum and simultaneously guaranteeing prompt detection of real attacks. This model Instantaneous detection is not an option, unless the false alarm risk is high.However, very little work has been done on Statistical model this is an interesting approach to detecting anomaly by statistical methods is implemented in the Smart Sifter algorithm [5]. The basic idea of this algorithm is to construct a probabilistic data model based on observations. In this statistical model, the important data from the statistical view is constructed rather than the entire dataset is stored as the training set and then the objects are processed successively. The statistical model that learns while processing each data object, so the data object is describes to be an intrusion of the model changes minimum more after processing it. This process based it introduced the special metrics, the intrusion factor; it is to measure changes in the probabilistic model after adding a new element.

In this paper [6] author has presented the concept of multi-stage feature selection method, it is use the filters and regression wrappers also. This model reduces the original feature vectors from 41 to only 16 features. The cost analyzing is for generating individual features from standard IP Flow Information Export records, available at many routers. This model is to reducing the computational effort. This effort for generated the on-line feature from real-time or live traffic observations at the network nodes and is eliminates very costly features.

The authors in [7] applied Distance based approaches efficient to detect anomaly. The datasets are usually includes continuous and categorical variables. In this model is used to a pair-wise distance that considers both kinds of variables. The distance based approach finds the distance between two objects and calculates the intrusions. This gives the Complexity in heterogeneous data grouping. The below fig 1.0. Depicts, the given working methodologies is includes various techniques, which can be used to detect the network Anomalies.


Existing Approach to Handle Anomaly

Techniques	Drawbacks
Maximum entropy: anomaly detection	anomaly classification not perform in this model
ASTUTE	ASTUTE cannot find large volume anomalies caused by one or a few flows.
Change-Point Detection	Selecting the optimal threshold value is difficult.
Statistical modal	Construction of a probabilistic data model, which is a rather complicated computational task.
Distance based approaches	Complexity in heterogeneous data grouping.

II. PROPOSED WORK

The existing framework requires more training dataset and time. This can be computationally demanding for large data sets. The proposed system provides a framework called Hybrid Sequential Pattern Anomaly Detection (H-SPAD), to minimize the training data collection risk and effectively detect anomalies with classification. The proposed system aims at increasing the classification performance through the hybrid approach along with producing least false alarm rate. This effectively reduces training data by applying the historical data as an input. So this aims at reducing the training overhead. so it is significantly will produce higher results than the other algorithm in training and the detection speed, and have a high enhance of the detection rates of attacking sample.

The proposed system overcomes the above drawbacks by developing a new effective technique **Hybrid Sequential Pattern Anomaly Detection (H-SPAD)**, which is falls under the Pattern analysis and classification approach.

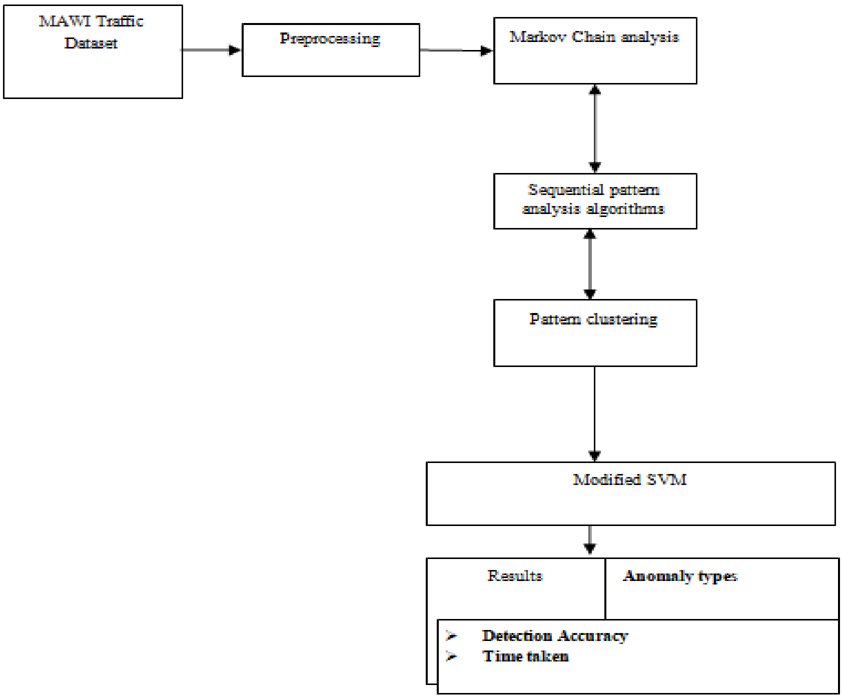


Figure 2.0 steps of H-SPAD

2.1. H-SPAD

2.1.1 Dataset Collection Preprocessing

The first step of H-SPAD is the process of uploading datasets. This network traffic data obtained from the MAWI repository which archives traffic datasets. The dataset will be preprocessed before starting the implementation. This step eliminates the duplicate and missing items in the uploaded dataset.

2.1.2 Markov Chain analysis

This markov chain analysis is a discrete time stochastic progression which is describing a dynamic sequence of possible events in which the probability of each event fully depends only on the state obtained in the previous list of event stage.

Let At denote a random variable representing the state of a system at time t, where t = 0, 1, 2,3, 4,5,6,7,8

Discrete time stochastic process with the following some of the assumptions:

- probability distribution of the state at time t+1 depends on the state at time t, and does not depend on the previous states leading to the state at time t;
- A state transition from time t to time t+1 is independent of time.

Let assume that pij denote the probability that the system is in a state j at time t+1 given the system is in state t at time t interval. If the system has a finite number of states, 1, 2 s, the stationary Markov chain can be defined by a transition probability[p].

An initial probability distribution

$$Q = [q1\ q2\ L\ qs]$$

$$\sum_{j=1}^{j=s} (Pij = 1)$$

The effective transition probability matrix and the initial probability distribution of a stationary Markov chain Can be learned from the observations of the system state in the past interval.

2.1.3. Sequential pattern analysis and pattern grouping

It determines the subsequence's and frequent relevant pattern from the given set of sequences. The algorithm is generates only as many candidates as will fit in memory and the support of the candidate is find out by scanning the dataset. Sequences from these candidates are written to temp and the candidates which are without minimum support are deleted. The same step is repeated until every candidate has been counted. After that this result are grouping effectively to perform effective anomaly detection.

2.1.4 .Modified SVM

Step 1: collect training data samples and test samples.
Step 2: According to data collection, constructs training sample set and test sample set.
Step 3: Set up parameters, initializes the initial support vector object position, every position corresponding a set of attributes (*a1, a2, a3, a4, an*) in model, builds up MSVM prediction model by parameters and samples.
Step4: From the parameters calculate every class threshold value, and then analyze the hyper plane value.
Step5: Randomly select the *P* objects from initial cluster, find out the optimal object position *best X* based on the hyper plane. Set it up as individual target *obj X* .
Step6: The non-optimal objects in the initial cluster moving to target class position and make the overall search.
Step7: The optimal object make overall search according to its neighborhood.
Step8: Update every objects class
Step8: Apply the optimal parameter (*a1, a2…an*) and training sample to build up classification model. This algorithm substantially higher than the previous methods for anomaly class detection.

III. EXPERIMENTAL RESULTS

The traffic data sets is considered in this work consists of attributes like protocol type, service, flag, duration such data are collected from the repositories. To avoid the imbalance of the dataset, the sample set are preprocessed to regenerate new training and test set. In this work H-SPAD with the existing algorithm were compared. The performance evaluation is done based on the performance metrics such as detection time, and accuracy. This performance analysis is represented in the graphical representation. .

The proposed system detects anomaly activities with significant improvement in terms of high detection rate and low false positive rate. This can be analyzed with different set of data's and results are shown in Figures.

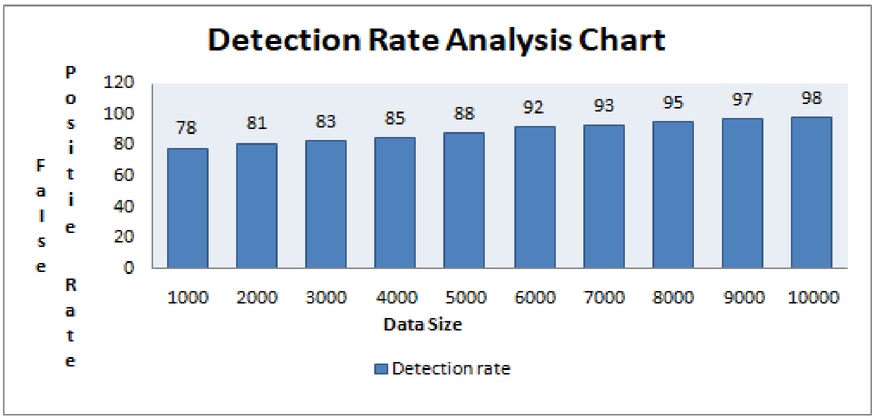


Fig 4.7 Detection rate analysis chart

From the results shown in the graphs, it can be observed that the proposed H-SPAD provided better detection rate and reduced false positive rate when it is analyzed with different number of datasets.

Time comparison analysis chart

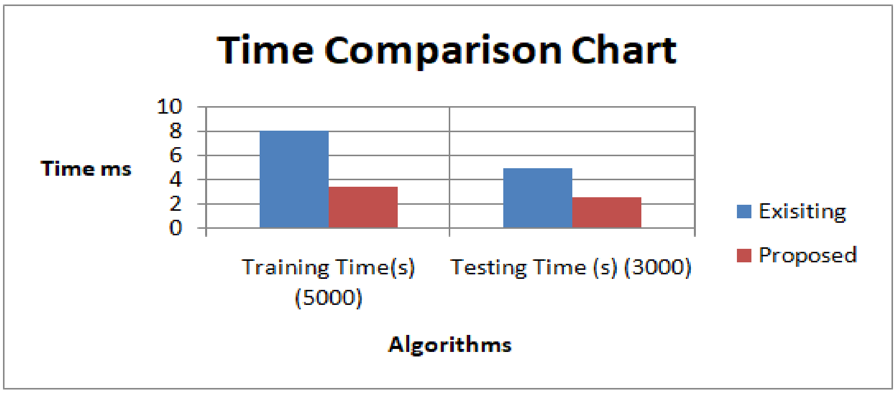


Fig: 5.0 Time comparison between existing and proposed

IV. CONCLUSION

In this paper, the effective framework presented to detect the anomaly especially in traffic dataset. It is done by introducing a methodology called the Hybrid Sequential Pattern Anomaly Detection (H-SPAD) algorithm, which performs enormous and number of parallel flows is very large dataset. Aim of the paper detecting the presence of anomaly from a large amount of data via a classifier and effectively producing least false alarm rate and improving he classification performance. So this aims at reducing the training overhead. The experimental tests were conducted and analyzed and compared with the existing methodology. The performance analysis is made by comparing it with the existing methodology and it is proved that the proposed method improves in its performance over metrics such as detection accuracy, execution time.

REFERENCES

[1]. Casas, Pedro, Pierdomenico Fiadino, and Alessandro D'Alconzo. "Machine-Learning Based Approaches for Anomaly Detection and Classification in Networks." TMA. 2016.

[2]. Y.Gu, A. McCallum, and D. Towsley,“Detecting anomalies in network traffic using maximum entropy estimation,”in Proc. IMC, 2005, p. 32..

[3]. Silveira, F., Diot, C., Taft, N., & Govindan, R. (2011). ASTUTE: Detecting a different class of traffic anomalies. ACM SIGCOMM Computer Communication Review, 41(4), 267-278.

[4]. Tartakovsky, Alexander G., Aleksey S. Polunchenko, and Grigory Sokolov. "Efficient computer network anomaly detection by change point detection methods." IEEE Journal of Selected Topics in Signal Processing 7.1 (2013): 4-11. [4]

[5]. Khan, Rahul Rastogi1 Zubair Khan2 MH. "Network Anomalies Detection Using Statistical Technique: A Chi-Square approach." (2012).

[6]. Iglesias, Félix, and Tanja Zseby. "Analysis of network traffic features for anomaly detection." Machine Learning 101.1-3 (2015): 59-84..

[7]. Prasad, YA Siva, and G. Rama Krishna. "Statistical Anomaly Detection Technique for Real Time Datasets." International Journal of Computer Trends and Technology (IJCTT)–volume 6 (2013)..

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

Dr.P.Muhamed Ilyas “Fast and Effective Network Anomaly Detection Technique Using Hybrid Sequential Pattern algorithms.” IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 8, 2018, pp. 01-05.