International Journal of Engineering &Technology, *7 (3.12) (2018) 1128-1136*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Real-time LAN/WAN and Web Attack Prediction Framework Using Hybrid Machine Learning Model

Mohammad Arshad[1], Md. Ali Hussain[2]

[1]Research Scholar, Department of Computer science and Engineering, KLEF Guntur
[2]Professor, Department of Computer science and Engineering, KLEF Guntur
**Corresponding Author Email: [1]arshad.klce@gmail.com, [2]alihussain.phd@gmail.com*

Abstract

Real-time network attacks have become an increasingly serious issue to LAN/WAN security in recent years. As the size of the network flow increases, it becomes difficult to pre-process and analyze the network packets using the traditional network intrusion detection tools and techniques. Traditional NID tools and techniques require high computational memory and time to process large number of packets in incremental manner due to limited buffer size. Web intrusion detection is also one of the major threat to real-time web applications due to unauthorized user's request to web server and online databases. In this paper, a hybrid real-time LAN/WAN and Web IDS model is designed and implemented using the machine learning classifier. In this model, different types of attacks are detected and labelled prior to train the machine learning model. Future network packets are predicted using the trained machine learning classifier for attack prediction. Experimental results are simulated on real-time LAN/WAN network and client-server web application for performance analysis. Simulated results show that the proposed machine learning based attack detection model is better than the traditional statistical and rule based learning models in terms of time, detection rate are concerned.

Keywords: *Network intrusion detection, web security, classification algorithm, random forest, packet sniffer.*

1. Introduction

Network security has become the prime concern of both internet users and internet service providers. A network can be considered as completely safe and secure, if it is secure against all types of intrusions. Network Intrusion detection systems (NIDS) is a class of systems those implements all of these techniques to protect a network both from insider and outsider intrusions. These systems have the responsibility to monitor both incoming and outgoing traffic of a network. The complete process of NIDS can be classified into two types, those are:-

1. Misuse Detection (MD):- This group of NIDS usually requires signatures or patterns of pre-existing attacks in order to identify various intrusions. It has very fast and appropriate detection rate. Again, very little number false positive rate is reported. MD based NIDS performs better with offline data.

2. Anomaly Detection (AD):- This group of NIDS looks for proper deviations from the normal profiles of the network traffic. It is slower as compared to MD based NIDS. Again, large numbers of false positives can be obtained here. These NIDS are capable of identifying novel attacks across the network, which is not possible in case of former NIDS method. AD based NIDS performs better with online data.

Machine Learning (ML) algorithms are very much efficient in order to develop secure NIDS. It gathers information by observing the recorded traffic patterns or signatures. After that, it can predict for the future traffic patterns. There are two important phases of every individual machine learning approach, those are:- training phase and testing phase.

Machine learning usually needs huge and complicated data sets and it contains various kinds of normal and abnormal traffic patterns. Some advanced machine learning algorithms are implemented in NIDS in order to achieve reduced computational time and space complexity. In the present era, computer networks are used in almost all applications. Thus, the use of computer networks is growing exponentially. These networks are basically channels those are used to transfer huge amount of sensitive information among numbers of different computer devices, large servers to tiny mobile devices and mini-computers. With the exponential growth of computer networks, security breaches are also growing every day. Hence, in order to provide security to the computer networks various protection methodologies are introduced, some of those are mentioned below:- access control, encryption and firewalls.

The process of misuse identification detects attacks having less false positive rate and it is incapable to detect novel attacks. But, anomaly identification process is efficient enough to identify attacks having very high false positive rate. We can mention here that, there are several advantages and disadvantages of both the approaches. Therefore, in order to reduce these limitations and enhance the benefits of both the approaches, several hybrid techniques are developed in the subsequent time. Most of the IDS used presently are rule-based systems. These systems based on a group of constraints in order to represent either attacks or normal network characteristics. These constraints are gathered and analyzed by the security professionals.

On the other hand, manual constraint encoding process is not at all feasible in terms of cost and time. The efficiency of human experts can also influence the above process to a great extent. Human experts are required to analyze large amount of network activities in order to identify various intrusion patterns. All of the above mentioned issues are resolved by the implementation of different machine learning approaches. These approaches are used as the analysis of observational data sets in order to detect unsuspected relationships.

Apart from these, the Machine learning approaches are usually implemented in order to classify network connections into two groups, those are:- intrusion and normal data. This classification is carried out during the misuse identification phase. All the similar network connections are included in the same cluster according to their similarity index. During the phase of misuse identification, random forests approach is implemented in order to classify all network connections into either intrusion or normal data. A particular labeled training dataset is used in order to construct various classification patterns. During the anomaly identification phase, k-means algorithm is implemented in order to combine network connections data into k clusters. The process of clustering completely depends upon the similarity of connections features.

Among all network attack clusters, some are selected as anomaly clusters. There exist numbers of different issue in case of misuse as well as anomaly identification techniques. One major issue is the imbalance among intrusion types in case of real network connections data sets. There also exist another severe issue of anomaly identification that is, network connections dataset includes vital categorical features, but k-means approach is effective for continuous features. Rest of all continuous features is on different scale. It stops the algorithm from being biased towards greatest scale features.

There are various traditional approaches for feature selection on NID data. Those approaches are modified and extended by adding weight and it will easily enhance the identification rate of the above presented model. Another issue of anomaly identification is, identifying the anomalous and normal clusters. Identification process of clusters is performed after successful clustering of all data. Here, two important assumptions are made and according to those assumptions following statement is assumed as true. Most of the real network activities are normal, whereas the intrusion activities are considered as abnormal.

These assumptions are not always true as in certain cases, high degree of similarity exist among different types of intrusions and real activities. In the above mentioned case, the intrusions are combined with normal data and results extremely high false positive rates. In order to resolve the said issue, a supervised technique is introduced in order to enhance false positive rates. The working process of traditional cluster based NID detection model is described step by step below:-

1. At first, random forests approach is implemented during the misuse identification phase.
2. K-means clustering technique is implemented during the anomaly identification phase.
3. Both of these misuse and anomaly identification can be integrated in order to develop an advanced and efficient hybrid framework. The overall performance can be enhanced with inclusion of essential output data which is gathered from the misuse identification phase. Certain known intrusions are inserted as input to the weighted k-means approach during the anomaly identification phase.

2. Web Attacks

Web attacks are also known as internet attacks. With the extended use of web technology, web attacks are also increasing day by day. Such attacks can cause huge damage to the internet resources. Networks usually use different security systems just like intrusion detection system in order to handle the above mentioned attacks.

Generally, intrusion detection systems are used with firewalls and it adds additional security to the traditional firewalls. This system is implemented in order to observe and analyze incidents those can actually violate the standard security constraints. Intrusion detection systems have responsibility to identify various attacks as well as security bugs. Again, it has to report it to the network administrator.

Web vulnerability is considered as an important security threat to the whole internet community. It can be defined as special type of identity theft which uses social engineering skills. By this Web attack, online consumers’ sensitive details are compromised. These sensitive details may be personal or financial or social. A basic Web attack can be decomposed into four major phases, those are mentioned below:- Preparation, Mass broadcast, Mature Account Hijack. Sometimes this Web attacks can result with huge financial loss. These attacks can also hamper online consumers’ confidence to a great extent. According to a current survey, many consumers of European banks only use internet banking in order to check their account balance. They never use internet banking for doing transactions because og these growing Web attacks. These Web attacks are growing exponentially; hence the manual risk assessment methodology is incapable to handle these attacks. Machine learning approaches can enhance the assessment of Web attacks. These Machine learning approaches are efficient to gather knowledge from the previous Web attacks. Again, inherent characteristics for each and every risk level of Web attack are determined. Hence, the associated risk level of Web can be determined within a very limited time period. Additionally, it also results better accuracy as compared to other traditional approaches. The risks related to technical sophistication are not directly linked with the financial loss.

3. Machine Learning for Attack Detection

In the present era, realtime network data pre-processing and classification has significance importance. There have been extensive amount of research works in this field. But till date, there is no such effective mechanism that can be implemented in the real world applications. These real world applications have large size and high-dimensional data inputs. Intrusion identification systems have the responsibility to protect computer systems from different attacks. These intrusion identification systems are usually construct various classification schemes or patterns in order to distinguish normal behaviors from abnormal behaviors. Two important assumptions are there which are mentioned below:-

All users as well as program activities are observable with the help of computer systems. These observation processes are carried out through auditing schemes.

Normal and abnormal activities must show different behaviors.

There are several approaches which help to define several measures of system behaviors. All efficient intrusion identification systems are developed by considering these assumptions. Most of the intrusion identification systems are incapable to handle new types of attacks. Therefore, the traditional intrusion identification system always requires to be updated to handle latest threats.

The implementation of latest Machine learning techniques into network intrusion identification gives an opportunity for these systems to learn the behaviors of network automatically. There are two significant benefits of implementing Machine learning into intrusion identification system, those are:-

Machine learning techniques have the responsibility to produce the identification models for intrusion identification systems. Hence, new attacks class labels can be identified efficiently and automatically.

Machine learning can be applied to construct intrusion identification systems for various kinds of computing environments.

The complete process of traditional machine learning models can be carried out in four numbers of phases and those phases are mentioned below:-

All the packets transmitted across the network are gathered toget er.

Retrieving an additional set of features that will define a network connection or a host session.

Through implementation of Machine learning approaches, both normal and abnormal activities can be described appropriately.

The intrusions are identified with the help of the above learned models.

Support Vector Machines are always considered as an important and useful classification strategy. The Self-Organized Ant Colony Network (CSOACN) are efficient during the process of data clustering. Various researchers wanted to introduce a new approach which will integrate the basic concepts of both approaches in order to generate a very high and optimized performance intrusion identification system.

It is very complicated to develop an intrusion identification system that will realize real time identification in case of high-speed networks. There are two severe problems of the traditional realtime NID model, those are mentioned below:-

To decrease the overall expenses of the deployment process, the total amount of pre-processed data that are used throughout the Machine learning process are reduced.

After inclusion of additional information into a system, the previous models are needed to be updated. It will ensure that, the system is completely protected.

The process of retaining is very time consuming, hence it is quite impossible to retain the new model upon all available data. Therefore, an advanced process is required in order to produce an adaptive model which will be updated through the help of previous model along with new additional information. This approach is used to update the old models. The process of clustering in intrusion identification has significant role in order to solve the multiple classification issues. The traditional approach is very infeasible in terms of cost in case of huge training data.

4. Web Attack Detection Models

Since past two decades, there has been significant development in case of computer systems and internet technology. Previously, sending mails was taking many days, but now it is possible to send messages or emails within fraction of time. This is only possible with the rapid growth of internet and technologies. Again, people can communicate with other people across the world via relay chat or video conferencing.

Organization, countries or individual persons are victims of these malicious activities. Huge financial investment is done in order to prevent these network attacks. The attacks are technically sound enough to launch attack remotely instead of accessing regional systems. This is only possible because of interconnectivity of systems. Intrusion can be defined as the malicious activities which are performed using internet medium. In other words, intrusion is a special kind of internet activity that violates the standard security constraints of internet. Intrusion detection system (IDS) has the responsibility to identify unauthorized accesses of both hardware and software resources.

Again, it has the capability to identify each and every attack of computer network. All attacks identified through the intrusion identification system can be broadly classified into three major groups, those are mentioned below:-

1. Scanning attacks
2. Penetration attacks
3. Denial of Service attacks

Every individual group of attacks must show unique signature as well as behavior patterns. Intrusion identification system is programmed in such a way that, it is capable of detecting small modifications in the general behavior patterns. The system must trigger warning alarms in case of identifying an intrusion. When alarm is set, network administrators have the responsibility to study and analyze the logs in order to identify whether the identified activity is anomalous or not.

Almost all intrusion identification system shows high instances of false positives and false negatives. A false positive can be defined as a specific instance that detects a normal activity as malicious one. On the contrary, a false negative is defined as a specific instance where the intrusion identification system becomes unable to identify malicious activities. Generally, most of the intrusion identification systems produce many numbers of false alarms daily.

Either the intrusion identification system is anomaly-based or signature-based, both of them result a basic issue that is, maximum numbers of false positives or false alerts.

The defense levels are compromised to decrease the total numbers of false positives. Anomaly identification can also be defined as an observation which differs a lot from rest observations; so that it creates suspicion that it was originated differently. Large numbers of intrusion identification approaches use signature-based schemes in order to identify the abnormal behavior. The above schemes always generate few false positives than that of anomaly-based schemes. But, anomaly identification scheme is considered better as compared to signature-based schemes. Anomaly-based schemes are capable of identifying zero-day attacks. The whole anomaly identification process can be classified into three different groups, those are mentioned below:

Supervised schemes:- Supervised schemes are capable of modeling normal as well as abnormal behaviors. It considers anomaly identification just like a classification issue. Security experts are required to pre-label all normal data. All data those are not at all analogous to this scheme is included under the category of anomalous data. On the contrary, a pre-defined group of data can be included inside a particular set. All instances those are not related to that particular pre-defined group are considered as normal. After that, an appropriate classification algorithm can be implemented. Some examples of classification algorithms are:- Naïve Bayes, Support Vector Machine, Neural Networks, and so on.

Unsupervised schemes:- This type of anomaly identification schemes are usually implemented in case of no available labeled data instances. In other words, there will be no classification as anomalous data and normal data. This scheme is generally studied as a clustering issue. These schemes generally consider that, the normal instances are clustered. There may be one or more than one cluster. All of these clusters must have different characteristics.

Therefore, all the normal instances often follow a particular pattern, but there is no pattern for the anomalous instances. But, the above said assumption is not true always. There are certain cases where the anomalies create frequent patterns or clusters just like collective anomalies. In case of unsupervised schemes, at first the clusters those depends upon a particular similarity measure are detected. All the instances those not obey any cluster are included under the category of anomalous. Supervised schemes can be applied if and only if whole data is available much before of any processing. There are two important limitation of the above mention scheme, those are:-

A data instance that is not included inside any cluster is assumed to be as anomalous. But, the above assumption is not true in all cases. Because, sometime such instance is noise instead of anomaly.

According to the traditional way, initially clusters are identified and after that the anomalies. This is quite infeasible in terms of expenses. The total numbers of anomalies present in a particular dataset is mostly less as compared to normal data instances.

Semi-supervised schemes:- These group of schemes usually requires two sets of data, such as:- labeled and unlabeled. The semi-supervised schemes are usually implemented when only certain numbers of instances are labeled as normal. Here, by including very small quantity of labeled data, a classifier is built that has the responsibility to label all unlabeled data. Therefore, a model for normal data instances is constructed that is used to identify the anomalies in such a way that, all the instances those can’t be included inside the normal group are automatically included inside the anomaly group.

This is a basic scheme which is named as self-training and it is usually implemented in semi-supervised approach. The co-training method can be applied in those cases where multiple classifiers are used to train each other. It maintains a boundary line for normality. A data instance is included under the category of anomalous when it falls outside the boundary line. There are certain techniques included under the above three groups and these techniques have significant importance during the determination of anomalies, those techniques are mentioned below:-

Proximity based approaches.

Clustering based approaches.

Classification based approaches.

5. Related Works

R. Kaur and S. Singh performed a detail survey on various Machine learning schemes and web attacks [1]. They have used the basic concepts of anomaly identification approaches. They developed numbers of different advanced analysis strategies in order to identify all abnormal activities.

All anomalous activities found in social networks are considered as unusual and illegal activities. A minute change in the normal behaviour may lead to unusual activities. In this paper, various kinds of anomalies are considered. Additionally, the classifications of anomalies and the classification criteria are also taken into account.

All of the anomalies are classified into three broad categories, those are:-

Behaviour-based anomalies

Structure-based anomalies

Spectral-based anomalies

Again, these three categories can be further sub-divided in order to create various sub-categories. In this paper, numbers of different anomaly detection techniques are studied and analysed in detail. They have mentioned six most important issues in all of these previously developed traditional anomaly detection models and those are mentioned below:-

There is no such significance work in the field of dynamic anomaly detection.

Temporal constraints are required to be included in order to insert dynamicity. There are several technique those usually use these temporal information. These techniques use this information during the process of learning. Social networks are not emphasized properly with respect to time dimension.

It is very hard and complex to distinguish the normal users from anomalous users. Therefore, there is necessity of an advanced and efficient prediction approach.

Along with identification of anomalies, prevention of anomalies is equally important. There are some high security applications those will never take chance for their sensitive information leakage. Thus, the system must inform about the presence of anomalous user prior to the actual detection.

In the above three mentioned categories, some graphs metrics can be implemented in order to identify additional anomalies.

J. David and C. Thomas tried to identify all DDoS attacks in case of flow-based network traffic [2]. They basically used fast entropy technique for their research work. There are extended amount of research works in order to identify these two kinds of attacks. In this paper, an adaptive threshold method is introduced an enhanced and efficient strategy for identification of DDOS attacks. Here in this piece of research work, fast entropy approach is implemented followed by flow-based analysis. This model is used to analyze network activities and user's behaviour change with respect to time. The proposed technique can easily achieve less computational overhead than that of other traditional entropy-based approaches. Additionally, better detection accuracy can also be achieved. In future, further research can be carried out to detect attacker and agents of DDoS attack with the help of IP traceback algorithm.

S. Duque et.al, developed an advanced intrusion detection system [3]. They have implemented the basic concepts of efficient Machine learning algorithms for their research work. In this research paper, they have thoroughly studied and analysed all pros and cons of different previously existing intrusion detection systems. Among various cons, high false positives and low detection rate are most eye-catching issues. An efficient unsupervised machine learning technique that uses k-means is considered here.

It has the prime objective to decrease the overall false negative rate. Apart from this, there is necessity of an efficient system in order to detect the numbers of clusters automatically. The outcomes of k-means clusters demonstrate that, significant efficiency can be obtained if and only if exact numbers of clusters are applied. Sudden increase or decrease of this exact number can definitely reduce the overall efficiency of the above model. Detecting the exact number of cluster is thus very vital process because it can influence the efficiency of the above presented model remarkably.

In case of dynamic network, it is very complicated to detect the numbers of clusters because of the absence of training data.

A.Verma and V. Ranga performed statistical analysis of NID dataset for network intrusion detection system [4]. An efficient distance-based machine learning approach is implemented here. There have been extensive amount of research in order to develop an efficient network intrusion detection system. Hence, the anomaly-based network IDS can be implemented on wide range of applications.

R. M. Elbasiony, et.al,, introduced an advanced hybrid network intrusion detection framework [5]. The above proposed framework uses random forests and weighted k-means algorithm. Most of the traditional network IDS are completely rule-based in nature. Such systems are very complicated in terms of encoding rules. Again, these systems are inefficient to identify intrusions. Hence, there is necessity of a hybrid framework which is based on Machine learning classification and clustering approaches. In order to identify misuse, random forests classification technique is implemented.

Initially, the random forests technique is implemented as the basic classification technique. It has the responsibility to construct the intrusion patterns out of a balanced training dataset. All the gathered network connections are again classified into various kinds of intrusions according to their patterns. There are certain limitations of the above proposed technique those are mentioned below:-

It is incapable to identify intrusions those are not trained previously.

The weighted k-means technique is implemented as an efficient Machine learning clustering approach and it is applied to unsupervised anomaly detection approach. This technique has the responsibility to split all network connections into a specific numbers of clusters. After that, anomalous clusters are identified according to their features. The ‘‘KMlocal’’ approach is the modified version of traditional k-means clustering approach.

Random forests approach is applied along with weighted k-means approach in order to construct an advanced hybrid framework. It is capable to resolve all the issues of misuse as well as anomaly identification. The feature importance values are evaluated with the help of random forests approach. It has the responsibility to enhance the identification rate of the anomaly identification process.

The supervised approach is introduced in order to enhance the anomalous cluster identification with through enforcing certain known attacks. The outcomes of the above presented hybrid framework results high identification rates and false positive rates.

B. Agarwal and N. Mittal proposed a new hybrid technique in order to identify anomaly network traffic efficiently [6]. They have chosen various Machine learning approaches for their research work. Now-a-days, anomaly-based IDS have become very much popular because of its flexibility and adaptability. It is very much efficient in order to identify new additional network attacks.

It is hard to enforce pre-defined constraints to detect correct attack traffic. There is no significant difference among normal and attack traffic. Additionally, both of these approaches are integrated with each other in order to generate another new, advanced and effective hybrid technique. Here, this hybrid approach is compared with other traditional single approaches. DARPA Intrusion Detection Evaluation dataset is included during the evaluation phase. In this work, we can see that the entropy-based approach is efficient enough for better detection of anomalies in network.

The above approach is not at all dynamic in nature. Hence, it is incapable to identify the possibility of attacks. The entropy values can vary in that constant range in case of normal conditions. Hence, it may raise high false alarm rate. In case of support vector machine based technique, high false positive is obtained. The above mentioned issue of the SVM based technique is that, the network features are often used during the learning process without processing. The outcomes of the evaluation phase shows that, the proposed hybrid model is very much efficient in order to identify attack traffic. Apart from this, it results very less false alarms.

In future, the proposed model can be modified and extended through the extraction of additional network features. There is also necessity of more robust as well as dynamic identification approach. Furthermore, the model can be implemented in various applications and numbers of different data sets can be tested.

Md. H. Ali, et.al, developed an advanced intrusion identification system [7]. They have implemented the basic concepts of fast learning network scheme and particle swarm optimization methodology. Supervised Intrusion Detection System can be defined as a special kind of system which has the ability to learn from examples of previous attacks. It has the responsibility to identify future new attacks.

In this research paper, a new learning model is developed for fast learning network that depends upon particle swarm optimization. The above mentioned new methodology is known as PSO-FLN. The said model is implemented in order to resolve the issues of intrusion identification. It is validated by using the well-known data set KDD99.The proposed technique is compared with all other meta-heuristic approaches for the training of ELM and FLN classifier. The resulted testing accuracy is better as compared to other existing identification techniques.

In this piece of research work, the authors have studied and analysed the issues of intrusion identification. The ANN based intrusion identification algorithm is very popular because it has the ability to decrease the wrong negatives and wrong positives to a great extent. The models having more numbers of hidden neurons can result enhanced and improved testing accuracy.

Further research can be performed to resolve all the issues of less accuracy in case of some specific classes (those classes must have very restricted amount of training data).

Md. Al-Qurishi, et.al, introduced an advanced prediction model for Web DDOS attack identification in case of social network [8]. Deep regression model is implemented here in this research work to develop the prediction model. Web DDOS attacks are most well-known attack in Twitter and other social networks. This attack targets huge numbers of users and as we know the numbers of users of these social networks are increasing day by day.

Web DDOS accounts are increasing rapidly now-a-days and many operators do operate these accounts. They do follow latest up-to-date technologies in order to avoid detection. The complexities of Web DDOS profiles are also increasing. Therefore, the previously developed traditional Web DDOS identification approaches are not efficient enough to handle recent kinds of attacks. There is necessity of an advanced Web DDOS identification strategy for control and prevention of unauthorized activities. In the above research paper, they presented an efficient prediction system that uses deep learning method in order to resolve all issues related to Web DDOS attacks on Twitter.

The above presented system can be classified into here different categories, those are mentioned below:-

Harvesting module
Feature extraction module
Deep regression module

All the above three mentioned modules operate systematically in order to analyse as well as evaluate user's Twitter profiles. The above presented approach can result maximum 86% of accuracy, when unclean and noisy data is inserted as input.

M. Mazini, et.al, developed a new network-based intrusion identification system [9]. In their research work, they have implemented hybrid bee-colony optimization and AdaBoost approaches. In all types of secure networks, intrusion identification system is considered as the major component. The major issue of all these traditional techniques is false alarm report. Intrusion identification accuracy is influenced by huge volume of network data. This technique is capable of resulting very less false positives. ABC approach is implemented in order to carry out the process of feature selection effectively. Apart from this, the AdaBoost algorithm has the responsibility to evaluate and classify all features. NSL-KDD and ISCXIDS2012 datasets are included in the evaluation phase. The ABC technique is very much beneficial in order to resolve optimization issues of intrusion identification systems.

A.Sayed A. Aziz, et.al, performed a comparative study of various classification approaches which are implemented in network intrusion identification systems [10]. They previously developed a multi-agent artificial immune system for network intrusion identification and classification. This happens because of attacks representations in a particular training set. Additionally, the dependency among features can also play vital role in the identification process.

6. Proposed Model

Real-time network attack detection and prevention in LAN/WLAN is one of the major task in complex computer networks for network users and network administrators. This can be done by analyzing the network traffic using the packet capturing tools and techniques. Most of the traditional attack detection models are focused on capturing and detection of packets with limited buffered size and memory. As the size of the network and its statistical packets variation is increasing, it becomes difficult to predict the new type of packets after NID analysis. In order to resolve these issues, a real-time network intrusion detection is performed on LAN/WAN and Web traffic using the machine learning model.

Real-time network attack detection and prevention in LAN/WLAN is one of the major task in complex computer networks for network users and network administrators. This can be done by analyzing the network traffic using the packet capturing tools and techniques. Most of the traditional attack detection models are focused on capturing and detection of packets with limited buffered size and memory. As the size of the network and its statistical packets variation is increasing, it becomes difficult to predict the new type of packets after NID analysis. In order to resolve these issues, a real-time network intrusion detection is performed on LAN/WAN and Web traffic using the machine learning model.

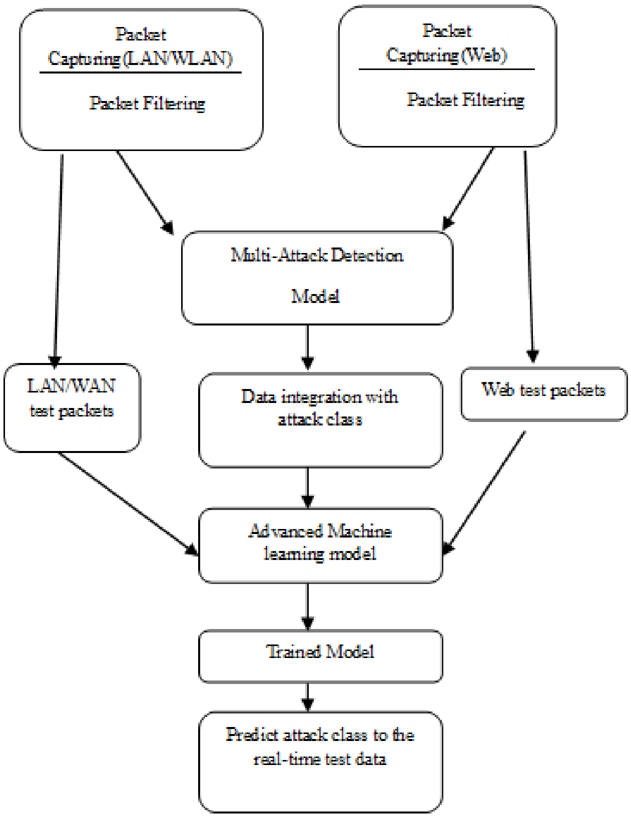
Packet Capturing (LAN/WLAN)
Packet Filtering
Packet Capturing (Web)
Packet Filtering
Multi-Attack Detection Model
LAN/WAN test packets
Data integration with attack class
Web test packets
Advanced Machine learning model
Trained Model
Predict attack class to the real-time test data

Fig. 1: Proposed Model flowchart

In the proposed model, a novel LAN/WAN and web based attack prediction model is designed and implemented on realtime datasets as shown in figure 1. Initially, network packets are captured from the LAN/WAN for packet filtering and attack detection process. Similarly, a web based packet capturing module is implemented to capture the web traffic from the client application to the server application. These web based packets are analyzed for web based attacks. Multi-attack detection model is used to find and label the each captured packet in the LAN/WAN and web packet with the attack class name. This attack labelled data is given to the proposed machine learning model to train the model on the data. Finally, real-time LAN/WAN packets or web traffic are used to predict the attack class label using trained machine learning model.

The main objectives in the models include,

Training real-time LAN/WAN packets and Web traffic.

Implementing a novel classification based prediction model for attack class prediction.

Predicting the attack class of the newly arrived packets in real-time LAN/WAN and web traffic.

Section 1: Packet Capturing LAN/WAN and Web Traffic

In this section, real-time LAN/WLAN packets are captured using the sniffer program for packet filtering and attack detection process. In this preprocessing stage, web traffic between the client browser and the web server are captured for web attack detection process. Here, LAN packets are captured in real-time using network interface card and sniffer program as shown in figure 2. Web packets are captured using the client-server communication process. In the web packet capturing process each packet from the client http request to the server is analyzed using the web sniffer program as shown in figure 3.

Algorithm 1:

Input : LAN/WLAN, Web traffic , Client IP, Server IP, Http request, Http response, port,Network interface cards NIC, Threshold, Packet size N, database DB.

Output: Filtered LAN/WAN packets and Web traffic packets

```
Procedure:
Step 1: NetPack[]=Null; //Network captured packets
NIC[]=Null; // Network interface cards in the available network
NIC[]← getInterfaces(LAN/WLAN);
Step 2: For each interface Ni in NIC[]
Do
            $\text{if}\,(N_i > 0)$
            $\text{then}$
            $\text{NConnect}[i] = \text{true};$
            $\text{NCon}(N_i, N);$
            $\text{endif}$
        End for
Step 3: For each interface in NCon
        Do
        $\text{Netpacks}[\,] = \text{getPackets}(\text{NCon}[i]);$
        End for
Step 4: User specific query q;
Protocol[]={tcp,udp,icmp..etc}
Step 5: For each packet Pi in Netpacks[]
Do
//          Packet fields
PF[]=Null;
$\text{if}\,(P_i \in q)$
$\text{then}$
$\text{PF}[\,] = \text{extractfields}(P_i, q, \text{protocol}[\,]);$
$\text{endif}$
End for
Apply Multi-attack detection model on PF[] and store in DB as LWNID with attack class label.
Step 6: // Web traffic packet capturing
        For each client IP CIP in IPList
        Do
                For each request of CIP
                Do
                        WP[]={Capture CIP, Packet details, Server IP, Port etc}
                Done
        Done
Step 7: // Web traffic attack detection, highest requests are detected as anomaly
Compute the Probabilistic Gaussian distribution to WP[] as
$\text{PG}(\text{WP}[\,]) = \max\{\Pr ob(\text{WP}[i] / \text{WP}), |\frac{1}{\sigma_{\text{WP}_i}\sqrt{2.\pi}} e^{-\frac{\text{WP}_i^2}{2}} |\}$
        Mean of the Gaussian value MPG=PG(WP[])/N
For each web packet in WP[]
        Do
                if(PG(WP[i]>MPG)
                then
                        Label packet as SqlInj
                Else
                        Lavel packet as Normal.
                End if
        Done
Step 8:   // Web attack detection using Entropy measure
                WebReqDBEntropy
$\eta = \frac{\lambda}{\lambda - \alpha} \log_2 \sum_{i=1}^{N} (p_i)^{\alpha} . 2^{\text{PG}(D_i)(\lambda - \alpha)}$
=
```

Where λ is the normalizing factor, α is sensitive factor α >0 more sensitive, α <0 less sensitive to attack. p_i is the probability of packet occurrence in the previous window session.

```
If( $\eta$ >0)
Then
        Label as web attack DDOS.
Else
        Label as normal.
```

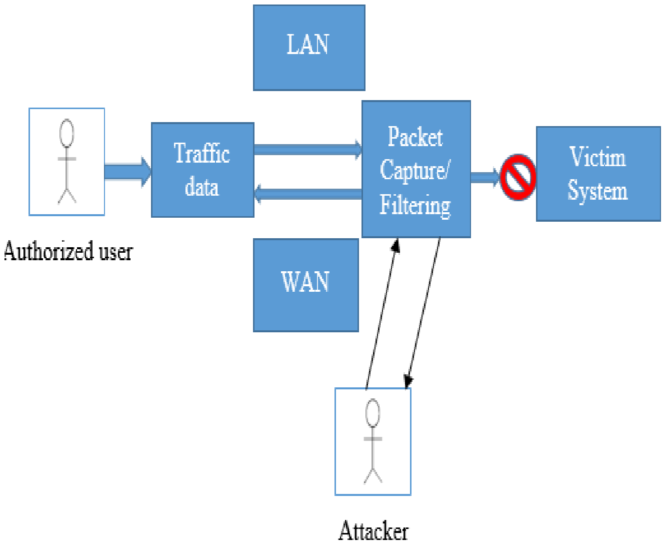

Fig. 2: Real-time LAN/WAN network analyzer

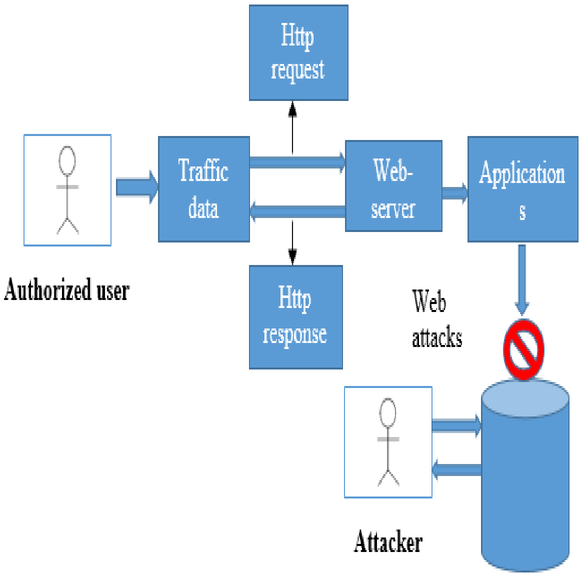

Fig. 3: Real-time web packet analyzer

Feature selection measures for random forest classification tree:

$$\mathrm{ProposedFSM} = \max\{\sqrt[3]{\left(\sum_{p=1}^{|D_p|}\sum_{n=1}^{|D_n|}\left(\sqrt[3]{D_p/|D_n|}-\sqrt[3]{D_n/|D_n|}\right)^2\right)},$$

$$\mathrm{PolynomialKernelSVM}(D_p, D_n),\ \mathrm{GainRatio}(D_p/D, D_n/D), \mathrm{CorrelationFS}(D_p, D_n)\}$$

------(1)

Hybrid Classification model for LAN/WAN and Web attacks:

```
Input  data records are partitioned into 'k'  disjoint subsets.
For each subset
do
    Select data_classifier Ci/i=1…m {Enhanced Random forest, WFFNN}
    Using real-time packets and training data detect new type of attacks classes.
    If Ci  is decision tree // Classify LAN/WAN filtered data
    Then
        Construct decision tree using Ci  on LAN/WAN filtered data.
        In the Ci classifier, decision tree is constructed using the attribute selection measure eq.(1). Proposed attribute selection measure formula is used to find the best split attribute among attack feature set. The attribute with highest computational value is selected as split node for decision tree construction.
        Find and extract the attack patterns in the decision patterns using the LAN/WAN future packets.
    End if
    Else if Ci is WFFNN  // Classify Web traffic data
    Then
        Initialize weights using the Gaussian computation as
        The generalized Gaussian kernel density function is used for each feature as follows.
```

$$p(x,\eta,\theta) = \frac{\theta}{2.\eta\Gamma(1/\theta)} e^{-(|x|/\eta)^{\theta}}$$

where Γ (.) is the Gamma function, and η, θ are generalized Gaussian density factors.

```
        Apply feed forward neural network.
    End if
```

7. Experimental Results

In this section, different types of attack patterns are analyzed using real- time network data in LAN/WAN networks. In this study, a client-server based web application is developed to detect the web attacks using the proposed model. Experimental results are executed using the JDK 1.8 with the third party libraries such as windows packet capture, jpcap, jcommons, and jama. A large number of network packets are captured using LAN/WLAN. Here, different types of attacks are pre-detected on the victim's machine using the proposed model. Here, attack training dataset is used to find the essential patterns in the new type of attacks on the real-time dataset. Initially, network packets from the LAN/WAN are captured using the sniffer program and then filtering is applied to find the relevant fields in each packet. Similarly, client http requests are captured at the web server to process the number of attempts to break the security at the web server side or web application or web database.

Table 1: LAN/WAN packet summary details

LAN/WAN Packet Summary			
Time Interval(secs)	Captured Packets	Filtered Packets	Filter Time(ms)
5	52432	49242	8737
10	97374	87389	14847
15	153643	147632	24284
20	204256	196838	34264
25	275835	270863	39872

Table 1, describes the summary of the packets captured in the LAN/WAN network for network attack prediction model.

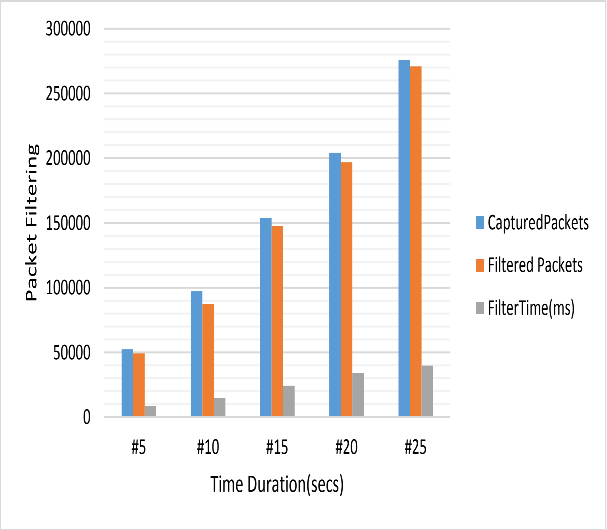

Fig. 4: LAN/WAN packet summary details in graphical format

Table 2: Web traffic summary details for web attack prediction

Web Sniffer			
Time Interval(secs)	Captured Packets	Filtered Packets	Filter Time(ms)
#5	5255	4536	3243
#10	8977	8093	5833
#15	14264	13872	7835
#20	18749	17973	9837
#25	24354	23954	13643

Table 2, describes the summary of the packets captured in the client-server web application for network attack prediction model.

Table 3: Performance of the attack prediction rate of the proposed model to the existing models using proposed machine learning models on LAN/WAN filtered packets

LAN/WAN Training Data Class Prediction Rate					
Time Interval (secs)	Filtered Packets	J48	SVM	NN	Proposed Model
#5	49242	0.783	0.843	0.9243	0.9763
#10	87389	0.793	0.832	0.9435	0.9834
#15	147632	0.814	0.8536	0.9334	0.9683
#20	196838	0.832	0.8643	0.9254	0.9745
#25	270863	0.817	0.8564	0.9573	0.9868

Table 3, illustrates the performance of the attack prediction rate of the proposed model to the existing models using proposed machine learning models on LAN/WAN filtered packets. From the table, it is observed that the present technique has high computational attack prediction accuracy compared to the existing models.

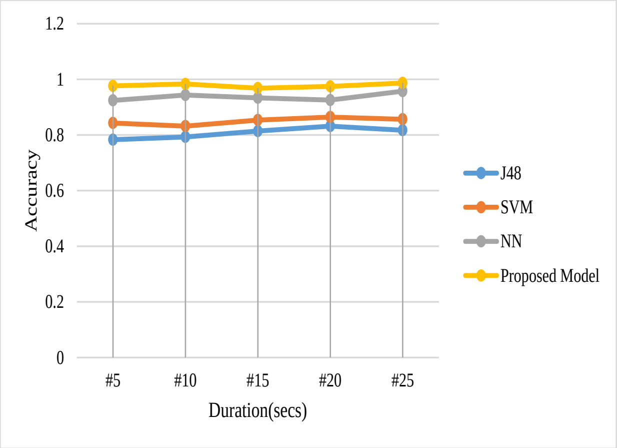

Fig. 5: Performance of the attack prediction rate of the proposed model to the existing models using proposed machine learning models on LAN/WAN filtered packets

Table 4: Performance of the attack prediction rate of the proposed model to the existing models using proposed machine learning models on client-server web filtered packets

Web attack Class Prediction Rate					
Time Interval (secs)	Filtered Packets	J48	SVM	NN	Proposed Model
#5	4536	0.8524	0.896	0.9532	0.9648
#10	8093	0.8763	0.9065	0.9195	0.9597
#15	13872	0.8832	0.9321	0.9436	0.9868
#20	17973	0.8453	0.9175	0.9475	0.9796
#25	23954	0.8753	0.9264	0.9636	0.9703

Table 4, illustrates the performance of the attack prediction rate of the proposed model to the existing models using proposed machine learning models on web filtered packets. From the table, it is observed that the present technique has high computational attack prediction accuracy compared to the existing models.

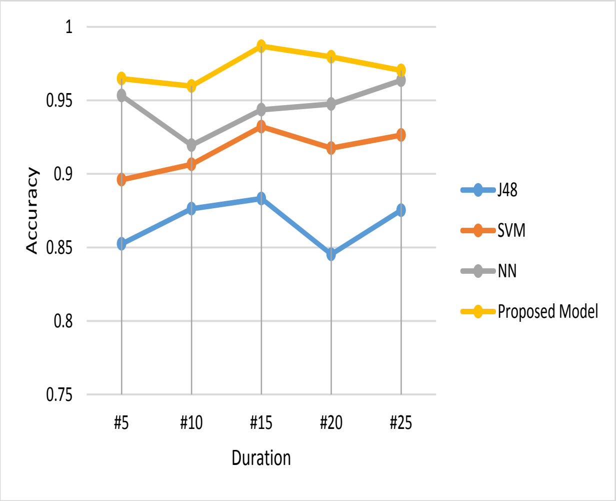

Fig. 6: Performance of the attack prediction rate of the proposed model to the existing models using proposed machine learning models on client-server web filtered packets

8. Conclusion

Real-time network intrusion detection is the major problem in the complex network systems due to variable size network analyzers and memory constraints. Detection of network attacks in standard networks such as LAN/WAN is difficult to analyze the attacks in the future packets due to the computational memory and time. Also, most of the traditional real-time attack detection models are used to find the attacks based on patterns or signatures in the packet. In this model, different types of attacks are detected and labelled prior to train the machine learning model. Future network packets are predicted using the trained machine learning classifier for attack prediction. Experimental results are simulated on real-time LAN/WAN network and client-server web application for performance analysis. Simulated results show that the proposed machine learning based attack detection model is better than the traditional statistical and rule based learning models in terms of time, detection rate are concerned.

References

[1] R. Kaur ad S. Singh, “A survey of Machine learning and social network analysis based anomaly detection techniques”, Egyptian Informatics Journal (2016) 17, pp. 199–216.

[2] J. David and C. Thomas, “DDoS Attack Detection using Fast Entropy Approach on Flow-Based Network Traffic”, 2nd International Symposium on Big Data and Cloud Computing (ISBCC’15), pp. 30-36.

[3] S. Duque and Dr .Md. Nizam bin Omar, “Using Machine learning Algorithms for Developing a Model for Intrusion Detection System (IDS)”, “Complex Adaptive Systems, Publication 5 Cihan H. Dagli,

Editor in Chief Conference Organized by Missouri University of Science and Technology 2015-San Jose, CA.

[4] A.Verma and V. Ranga, “Statistical analysis of CIDDS-001 dataset for network intrusion detection systems using distance-based machine learning”, “6th international conference on smart computing and communications, ICCSCC 2017, Dec 2017, Kurukshetra, India.

[5] R. M. Elbasiony, E. A. Sallam, T. E. Eltobely and M. M. Fahmy, “A hybrid network intrusion detection framework based on random forests and weighted k-means”, Ain Shams Engineering Journal (2013) 4, 753–762.

[6] B. Agarwal and N. Mittal, “Hybrid Approach for Detection of Anomaly Network Traffic using Machine learning Techniques”, “2nd International Conference on Communication, Computing & Security [ICCCS-2012]”.

[7] Md. H. Ali, B. A. Dawood AL Mohammed, M. A. Binti Ismail and Md. F. Zolkipli, “A new intrusion detection system based on Fast Learning Network and Particle swarm optimization”.

[8] Md. Al-Qurishi, M. Alrubaian, Sk Md Mizanur Rahman, A. Alamri and Md. Mehedi Hassan, “A prediction system of Web DDOS attack in social network using deep-regression model”, Future Generation Computer Systems , 2017.

[9] M. Mazini, B. Shirazi and I. Mahdavi, “Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms”, Journal of King Saud University – Computer and Information Sciences.

[10] A.Sayed A. Aziz, Sanaa EL-OlaHanafi and Aboul EllaHassanien, “Comparison of classification techniques applied for network intrusion detection and classification”, Journal ofApplied-Logic24(2017)109–118.