Hybrid Modified -Means with CART for Intrusion Detection Systems

M.Natarajan
Assistant Professor, Department of CSE, K.Ramakrishnan College of Technology, Trichy
Tamilnadu
Email Id : forevernatarajan@gmail.com
R.Barathi
Assistant Professor, Department of CSE, M.Kumarsamy College of Engineering (Autonomous),
Karur, Tamilnadu
Email Id : bharathir.cse@mkce.ac.in
R.Pradeepa
Assistant Professor, Department of CSE, M.Kumarsamy College of Engineering (Autonomous),
Karur, Tamilnadu
Email Id : pradeepar.cse@mkce.ac.in

Abstract

Network Security has become the key foundation with the tremendous increase in usage of network-based services and information sharing on networks. Intrusion poses a serious risk to the network security and compromise integrity, confidentiality & availability of the computer and network resources. Human classification of network audit data is expensive, time consuming and a tedious job. Intrusion Detection System (IDS) is one of the looms to detect attacks and anomalies in the network. Data mining technique has been widely applied in the network intrusion detection system by extracting useful knowledge from large number of network data. In this paper a hybrid model is proposed that integrates Anomaly based Intrusion detection technique with Signature based Intrusion detection technique is divided into two stages. In first stage, the signature based IDS SNORT is used to generate alerts for anomaly data. In second stage, data mining techniques "k-means + CART" is used to cascade k-means clustering and CART (Classification and Regression Trees) for classifying normal and abnormal activities. The hybrid IDS model is evaluated using KDD Cup Dataset. The proposed assemblage is introduced to maximize the effectiveness in identifying attacks and achieve high accuracy rate as well as low false alarm rate.

Keywords-Anomaly Detection, Intrusion detection, data mining, k-means, CART, SNORT

Introduction

Interruption Detection Systems(IDS) is a vital recognition utilized as a counter measure to protect information honesty and framework accessibility from assaults. Interruption Detection Systems (IDS) is a mix of programming and equipment that endeavors to perform interruption discovery. It is a procedure of social occasion interruption related learning happening during the time spent checking the occasions and breaking down them for sign or interruption. It raises the caution when a conceivable interruption happens in the framework. The system information wellspring of interruption discovery comprises of vast measure of printed data, which is hard to understand and investigate. The fundamental inspiration driving utilizing interruption identification in information mining is mechanization. Example of the typical conduct and example of the interruption can be registered utilizing information mining. To apply information mining strategies in interruption discovery, to begin with, the gathered checking information should be preprocessed and changed over to the arrangement reasonable for mining preparing. Next, the reformatted information will be utilized to build up a bunching or order demonstrate.

Information Mining is the utilization of calculations to remove the data and examples determined by the learning revelation in databases handle. Information mining is being utilized to clean, arrange, and analyze vast measure of system information to relate regular encroachment for interruption discovery. The fundamental purpose behind utilizing Data Mining Techniques for Intrusion Detection Systems is because of the gigantic volume of existing and recently showing up system information that require preparing. The measure of information aggregated every day by a system is colossal. A few Data Mining procedures, for example, bunching, order, and affiliation tenets are turned out to be helpful for social affair distinctive learning for Intrusion Detection. Bunching is the technique for gathering objects into important subclasses so that the individuals from a similar group are very comparative, and the individuals from various bunches are very not quite the same as each other. Accordingly bunching strategies can be helpful for arranging log information and identifying interruptions. Arrangement maps information into predefined gatherings or classes. It is frequently alluded to as managed learning in light of the fact that the classes are resolved before inspecting the information. In numerous information mining applications that address order issues, highlight and model determination are considered as key assignments. That is, suitable info components of the classifier must be chosen from a given arrangement of conceivable elements and structure parameters of the classifier must be adjusted concerning these elements and a given informational index.

Problem Statement

To develop an application which will protect the system from several focused attacks, which are initiated by various intruders. There are various methods to develop a security application. Many of the approaches earlier accepted and performed had many problems which approximately didn't give the desired outputs and security from several intrusions. So we came up with a hybrid approach that is IDS in data mining using hybrid approach. This hybrid approach comprises of K Means clustering algorithm and Classification tree analysis algorithm CART (Classification and Regression Trees). Hybrid approach to provide security to the system from various intruders will overcome the problems which were raised using other approaches. So we proposed a Hybrid Approach to effectively detect the attack on system using hybrid approach in data mining.

Related work

IDS Detection Methods

- Signature-Based Detection

Signature-based detection is the process of comparing signatures/patterns of known attack with the observed events to identify possible incidents. The most common form of signature based IDS used commercially specifies each pattern of events that corresponds to an attack as a separate signature.

Advantages: Signature based detectors are very effective in detecting known attacks or threats that are predefined in the database of IDS.

Disadvantages: Signature based IDS are unable to detect unknown attacks or variants of known attacks. Database of signature based IDS has to be manually revised for each new type of attack that is discovered.

a) SNORT

Grunt is an open source IDS. It is a mark based system since it distinguishes the assault in light of the arrangement of principles that are predefined inside the Snort. In the event that any assault information is discovered then it naturally drops the bundle generally the specific record is considered as a typical information.

Grunt is decide based system that characterizes new principles. Grunt comprises of the accompanying four segments ;

(1) Packet catch/interpret motor: It utilizes the libpcap parcel catching. Caught parcels are

at that point handled by interpreting motor and decoded bundles.

(2) Preprocessor modules: Packets are gone through various preprocessors for examination and process bundles before they are passed to recognition motor.

(3) Detection motor: It tests the information bundles for various characteristics expressed in Snort rules definition document.

(4) Output modules: It acknowledge alerts created from preprocessors, identification motor, or deciphering motor.

- Anomaly-Based Detection

Peculiarity Based identification looks at meanings of what action is viewed as ordinary against watched occasions to distinguish noteworthy deviations. Abnormality based IDS utilizes profiles that speak to the ordinary conduct of framework, applications or system movement that are produced by dissecting the attributes of commonplace action over timeframe .

Points of interest: Anomaly based IDS can identify new or obscure assaults or irregular conduct. Oddity

discovery has the preferred standpoint that no guidelines should be composed and it can identify novel or new assaults.

Hindrances: Profiles can at times be off base which comes about into era of false cautions considering ordinary information as an assault. Profiles ought to be overhauled continually.

Information Mining in Intrusion Detection System

alludes to the way toward removing powerful, overhauled, dormant, valuable, and the reasonable example from a huge fragmented, commotion, non-steady and arbitrary information. In interruption discovery framework, the data bargains from numerous sources, for example, organize activity or logs, framework logs, application logs, caution messages, and so forth. Because of shifted information source and configuration, the many-sided quality expanded in inspecting and investigation of information. Information Mining has enormous preferred standpoint in information extraction from huge volumes of information that are loud and dynamic, consequently it is of awesome significance in interruption identification framework.

(a) K-Means

K-means is a dividing strategy in bunching method of information mining. K-Means grouping technique is utilized to parcel the preparation information into k bunches with the assistance of Euclidean separation closeness. It is a calculation to amass or to characterize the items in light of traits/components into k number of bunches.

Fundamental strides for bunching the information by k-implies are:1.Select a number (k) of group focuses - centroids (random),2.Assign each protest its closest group focus (e.g. utilizing Euclidean distance)Move every bunch focus to the mean of its doled out articles. Rehash steps 2,3 until joining (change in bunch assignments not as much as a limit)

Advantage: Relatively effective in gathering typical or strange information.

Disservice: Unable to deal with loud information.

(b) CART (Classification and Regression Trees)

Order tree examination is utilized to distinguish the "class" to which the information has a place. Relapse tree examination is the place the information is ceaseless and tree is utilized to anticipate its esteem. The term Classification and Regression Tree (CART) examination is utilized to allude to both of the above techniques. Arrangement and relapse trees are machine-learning techniques for developing expectation models from information. The Classification and Regression Trees (CART) procedure is in fact called as paired recursive parceling . The procedure is double since parent hubs are constantly part into precisely two tyke hubs and recursive on the grounds that the procedure is rehashed by regarding every kid hub as a parent. The key components of CART investigation are an arrangement of tenets for part every hub in a tree; choosing when tree is finished and doling out a class result to every terminal hub.

The fundamental strides of CART are:

1.Rules for part information at a hub in view of estimation of a variable

2.Stopping when a branch turns into a leaf/terminal hub and can't be part further

3.Finally an expectation for target variable in every leaf/terminal hub.

Preferences: CART does not depend on information having a place with a specific sort of appropriation.

Cross breed IDS show

The Hybrid IDS model is framed by utilizing SNORT IDS and two pre-processors PHAD and NETAD.

SNORT+PHAD+NERAD

The half breed IDS is acquired by joining (PHAD) parcel header irregularity recognition and (NERAD) organize activity peculiarity discovery which are inconsistency based IDSs with the abuse based IDS.

A. Grunt

Grunt is a manage based system interruption location framework. Each lead comprises of two consistent parts: the run header and manage alternatives. The govern header has five areas; control activities (the move to be made when an interruption is identified), the end-to end source and goal data (source IP addresses, goal IP addresses and port numbers relying upon the convention), what's more, convention sort (TCP, UDP, or ICMP).The run alternatives comprise of various conditions that help choosing whether the specified abuse operation has happened or not.

B. PHAD

Parcel header irregularity locator (PHAD) is the primary inconsistency based approach added to Snort as a preprocessor in this review. PHAD is not quite the same as other system based abnormality recognition frameworks by two reasons. Firstly , it demonstrates conventions as opposed to the client conduct on the grounds that most of the assaults misuse convention execution bugs and must be comprehended by distinguishing unordinary information and yield. Furthermore, it utilizes a period based model, accepting a speedy change in a brief timeframe in the system insights. PHAD decreases false alert rate by hailing just the main abnormality as a caution.

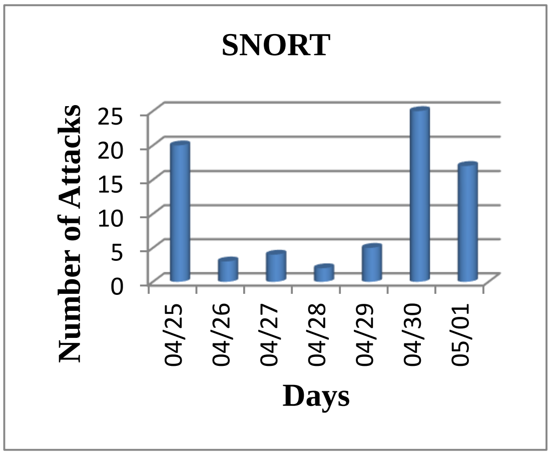
C. Network traffic anomaly detector (NERAD)

Network traffic anomaly detector (NERAD) is second oddity based approach added to Snort as a pre-processor in this review. The NETAD likewise models bundles as PHAD. NETAD works in two stages: First is the sifting of approaching customer sessions to recognize start of sessions. Second is the displaying stage. Separating stage disposes of the movement up to 98–99%.Elimination rearranges the activity for the demonstrating stage. Accordingly just the activity information which confirmation of assaults are incorporated into is passed to the demonstrating stage.

Result & Discussion

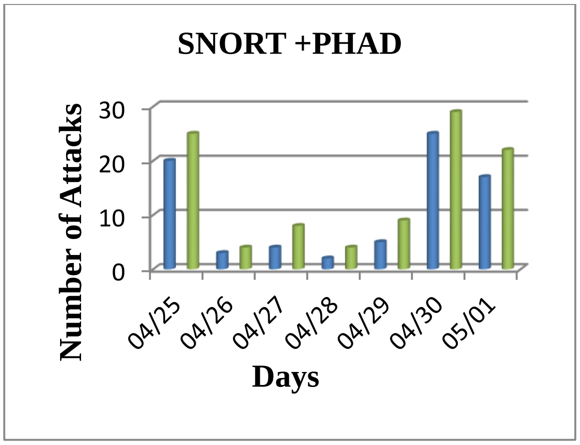
Performance of SNORT

Grunt distinguishes just profile based assaults and the inconsistency based methodologies, for example, Application Layer Anomaly Detector (ALAD) and Learning Rules for Anomaly Detection (LERAD) is utilized to perform better expectation. Semi-Supervised Approach, the named information can be marked utilizing the unlabeled information . The named preparing information are connected to the SVM classifier.

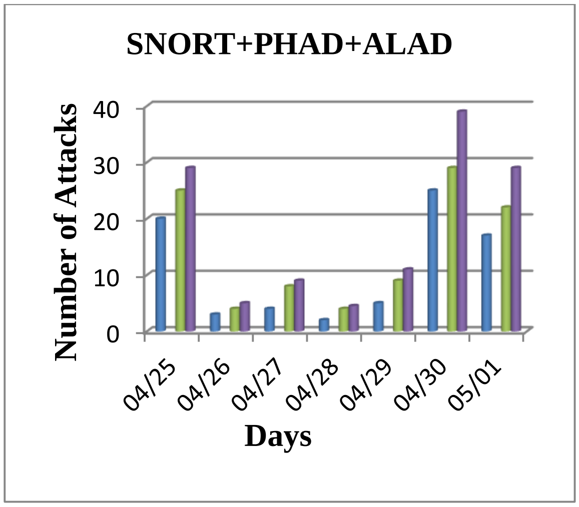


Performance of SNORT + PHAD

Assaults identified by SNORT and PHAD all alone and results in the Hybrid Intrusion Detection System are indicated . It is comprehended that in the wake of including PHAD with Snort it identifies better than anyone might have expected. The quantity of assaults distinguished by SNORT expanded .

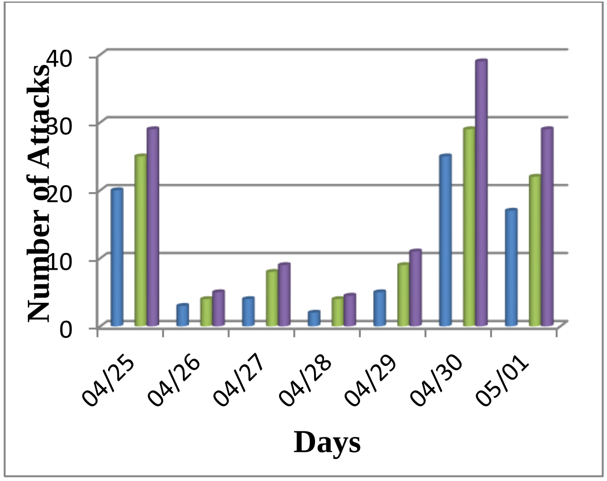


The quantity of assaults recognized by SNORT + PHAD expanded from SNORT + PHAD + ALAD rendition of IDS. The principle reason is Snort distinguishes the assaults in view of lead definition documents yet PHAD and ALAD recognize utilizing parcel header and system convention.



Proposed Hybrid IDS
(SNORT + ALAD + LERAD)

Assaults identified by SNORT + ALAD + LERAD all alone and results in the Hybrid Intrusion Detection System (SNORT + ALAD + LERAD) is appeared in After including SNORT + ALAD + LERAD, the IDS gives better outcomes when contrast and different techniques. The quantity of assaults identified by SNORT + PHAD + ALAD expanded from SNORT + ALAD + LERAD (Hybrid IDS) variant of the IDS.



Conclusions

In this paper we proposed a method for classification of intruder in system Intrusion detection. Here we detected intrusion through data mining method by combining two data mining technique Modified K means and CART and formed a hybrid technique. We combined these different methods for measured different aspects of intrusions. Combined these rules find the intruder attack more quickly from the exiting one. We have successfully implemented the Employee task assignment interface as well as Interface for the employee to simultaneously work on the assigned file. Multiple security measures has been implemented to reduce the threat of corrupting the stored data on the system. A unique key is assigned to each user for each session and therefore assigned file can be modified only if user provide the security key correctly. This gives the system first level of security in horizontally distributed database system. The system has been made immune to the major attacks by developing. the Intrusion detection system using two well know data mining algorithm Modified k-mean and CART Algorithm. We have obtained excellent results for the implemented system and Dataset.

References

[1] G.V. Nadiammai, M. Hemalatha, “Effective approach toward Intrusion Detection System using data mining techniques”, Elsevier Publication, 2013.

[2] M. Ali Aydın, A. Halim Zaim, K. Gokhan Ceylan, “A hybrid intrusion detection system design for computer network security”, Computers and Electrical Engineering, Elsevier Publication, 2009,

[3] Basant Agarwal, Namita Mittal, “Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques”, 2nd International Conference on Communication, Computing & Security, Procedia technology, ScienceDirect, Elsevier Publication, 2012.

[4] Saurabh Mukherjee, Neelam Sharma, “Intrusion Detection using Naive Bayes Classifier with Feature Reduction”, Procedia technology, ScienceDirect, Elsevier Publication, 2012.

[5] Amuthan Prabakar Muniyandi, R. Rajeswari, R. Rajaram, “Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree algorithm.”, International Conference on Communication Technology and System Design, Procedia Engineering,ScienceDirect,Elsevier Publication, 2011.

[6] P Srinivasulu, D Nagaraju, P Ramesh Kumar, and K Nageswara Rao, ”Classifying the Network Intrusion Attacks using Data Mining Classification Methods and their Performance Comparison”, IJCSNS International Journal of Computer science and Network Security, Vol.9 No.6, 2009.

[7] S Saravanan, V Venkatachalam ,“ Advance Map Reduce Task Scheduling algorithm using mobile cloud multimedia services architecture” IEEE Digital Explore,pp21-25,2014.

[8]S.Swathi “Preemptive Virtual Machine Scheduling Using CLOUDSIM Tool”, International Journal of Advances in Engineering, 2015, 1(3), 323 -327 ISSN: 2394-9260, pp:323-327.

[9] S Saravanan, V Venkatachalam, S Then Malligai "Optimization of SLA violation in cloud computing using artificial bee colony"2015, 1(3), 323 -327 ISSN: 2394-9260, pp:410-414.

[10]S. Saravanan, Vikram R ,"Improved Performance Analysis Image Segmentation Based on Cluster Image", Journal of Chemical and Pharmaceutical Sciences,issue 1,2017,pp92-95

[11]S. Saravanan, Vikram R ," Evolutionary Calculations on Gravitational Interactions Method of Global Leader Organize ", Journal of Chemical and Pharmaceutical Sciences,issue 1,2017,pp115-118