ABSTRACT

Intrusion Detection System (IDS) needs a data and for this it is important to keep the real working environment to find out all the possibilities of how an attack is about to happen and this seem to be expensive. A Software to detect network attacks in a computer network from unidentified users, including known personnel. The attack detector's learning task works up a predictive model which is a classifier in this case which differentiates the "bad" (i.e., intrusions or attacks) and "good" or "normal" connections. The primary aim is to use machine learning based techniques to provide packet connection transfer in a better way by predicting results with the at most accuracy. Comparing and discussing the outputs from the couple of machine learning algorithms used for the given dataset with evaluated classification report, find the confusion matrix and categorize the data from priority and the result which shows that the efficiency of the claimed machine learning algorithm method is to be compared with the best accuracy techniques such as Precision, Recall and F1 Score.

LIST OF FIGURES

LIST OF ABBREVATIONS

Entity Relationship Diagram (ERD)

Intrusion Detection System (IDS)

TABLE OF CONTENTS

CHAPTER 1

1. INTRODUTION

1.1 AIM:

The main aim of this project is to predict the network attacks in a chain of network using supervised machine learning language.

1.2 ABSTRACT:

To create data for the Intrusion Detection System (IDS), it is necessary to set the real working environment to explore all the possibilities of attacks, which is expensive. Software to detect network intrusions protects a computer network from unauthorized users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between "bad" connections, called intrusions or attacks, and "good" normal connections. To prevent this problem in network sectors, have to predict whether the connection is attacked or not from KDDCup99 dataset using machine learning techniques. The aim is to investigate machine learning based techniques for better packet connection transfers forecasting by prediction results in best accuracy. To propose a machine learning-based method to accurately predict the overall attacks by prediction results in the form of best accuracy from comparing supervise classification machine learning algorithms. Additionally, to compare and discuss the performance of various machine learning algorithms from the given dataset with evaluation classification report, identify the confusion matrix and to categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with precision, Recall and F1 Score.

CHAPTER 2

2. LITERATURE SURVEY

2.1 General

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discusses published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

A summary is a recap of important information about the source, but a synthesis is a re-organization, reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them. Loan default trends have been long studied from a socio-economic stand point. Most economics surveys believe in empirical modeling of these complex systems in order to be able to predict the loan default rate for a particular individual. The use of machine learning for such tasks is a trend which it is observing now. Some of the surveys to understand the past and present perspective of loan approval or not.

2.2 Review of Literature Survey

Title : A Prediction Model of DoS Attack's Distribution Discrete Probability

Author: Wentao Zhao, Jianping Yin and Jun Long

Year : 2008

The process of prediction analysis is a process of using some method or technology to explore or stimulate some unknown, undiscovered or complicated intermediate processes based on previous and present states and then speculated the results. In an early warning system, accurate prediction of DoS attacks is the prime aim in the network offence and defense task. Detection based on abnormity is effective to detect DoS attacks. A various studies focused on DoS attacks from different respects. However, these methods required a priori knowledge being a necessity and were difficult to discriminate between normal burst traffics and flux of DoS attacks. Moreover, they also required a large number of history records and cannot make the prediction for such attacks efficiently. Based on data from flux inspecting and intrusion detection, it proposed a prediction model of DOS attack's distribution discrete probability based on clustering method of genetic algorithm and Bayesian method and the clustering problem first, and then utilizes the genetic algorithm to implement the optimization of clustering methods. Based on the optimized clustering on the sample data, we get various categories of the relation between traffics and attack amounts, and then builds up several prediction sub-models about DoS attack. Furthermore, according to the Bayesian method and deduce discrete probability calculation about each sub-model and then get the distribution discrete probability prediction model for DoS attack. This paper begins with the relation exists between network traffic data and the amount of DoS attack, and then proposes a clustering method based on the genetic optimization algorithm to implement the classification of DoS attack data. This method first gets the proper partition of the relation between the network traffic and the amount of DoS attack based on the optimized clustering and builds the prediction sub-models of DoS attack. Meanwhile, with the Bayesian method, the calculation of the output probability corresponding to each sub-model is deduced and then the distribution of the amount of DoS attack in some range in future is obtained.

Title : Apriori Viterbi Model for Prior Detection of Socio-Technical Attacks in a Social Network

Author: Preetish Ranjan, Abhishek Vaish

Year : 2014

Socio-technical attack is an organized approach which is defined by the interaction among people through maltreatment of technology with some of the malicious intent to attack the social structure based on trust and faith. Awful advertisement over internet and mobile phones may defame a person, organization, group and brand value in society which may be proved to be fatal. People are always very sensitive towards their religion therefore mass spread of manipulated information against their religious belief may create pandemonium in the society and can be one of the reasons for social riots, political misbalance etc. Cyber-attack on water, electricity, finance, healthcare, food and transportation system are may create chaos in society within few minutes and may prove even more destructive than that of a bomb as it does not attack physically but it attacks on the faith and trust which is the basic pillar of our social structure. Trust is a belief that the person who is being trusted will do what is being expected for and it starts from the family which grows to build a society. Trust for information may be established if it either comes from genuine source or information is validated by authentic body so that there is always a feeling of security and optimism. In the huge and complex social network formed using cyberspace or telecommunication technology, the identification or prediction of any kind of socio-technical attack is always difficult. This challenge creates an opportunity to explore different methodologies, concepts and algorithms used to identify these kinds of community on the basis of certain pattern, properties, structure and trend in their linkage. It tries to find the hidden information in huge social network by compressing it in small networks through apriori algorithm and then diagnosed using viterbi algorithm to predict the most probable pattern of conversation to be followed in the network and if this pattern matches with the existing pattern of criminals, terrorists and hijackers then it may be helpful to generate some kind of alert before crime.

Due to emergence of internet on mobile phone, the different social networks such as on social networking sites, blogs, opinion, ratings, review, serial bookmarking, social news, media sharing, Wikipedia led the people to disperse any kind of information very easily. Rigorous analysis of these patterns can reveal some very undisclosed and important information explicitly whether that person is conducting malignant or harmless communications with a particular user and may be a reason for any kind of socio technical attacks. From the above simulation done on CDR, it may be concluded that if this kind of simulation applied on networks based on the

internet and if we are in the position to get the data which could be transformed in transition and emission matrix then several kinds of prediction may be drawn which will be helpful to take our decisions.

Title : New Attack Scenario Prediction Methodology

Author: Seraj Fayyad, Cristoph Meinel

Year : 2013

Intrusion detection systems (IDS) are used to detect the occurrence of malicious activities against IT system. Through monitoring and analyzing of IT system activities the malicious activities will be detected. In ideal case IDS generate alert(s) for each detected malicious activity and store it in IDS database. Some of stored alerts in IDS database are related. Alerts relations are differentiated from duplication relation to same attack scenario relation. Duplication relation means that the two alerts generated as a result of same malicious activity. Where same attack scenario relation means that the two related alert are generated as a result of related malicious activities. Attack scenario or multi-step attack is a set of related malicious activities run by same attacker to reach specific goal. Normal relation between malicious activities belong to same attack scenario is causal relation. Causal relation means that current malicious activity output is pre-condition to run the next malicious activity. Possible multi-step attack against a network starts with information gathering about network and the information gathering is done through network Reconnaissance and fingerprinting process. Through reconnaissance network configuration and running services are identified. Through fingerprint process Operating system type and version are identified. propose a real time prediction methodology for predicting most possible attack steps and attack scenarios. Proposed methodology benefits from attacks history against network and from attack graph source data. it comes without considerable computation overload such as checking of attack plans library. It provides parallel prediction for parallel attack scenarios. Possible third attack step is to identify attack plan based on the modeled attack graph in the past step. The attack plan usually will include the exploiting of a sequence of founded vulnerabilities. Mostly this sequence is distributed over a set of network nodes. This sequence of nodes vulnerabilities is related through causal relation and connectivity. Lastly

global information, and its efficiency has been proved. For the event-triggered case, two effective dynamical event conditions have been designed and implemented in a fully distributed way, and both of them have excluded Zeno behavior. Finally, a simulation example has been provided to verify the effectiveness of theoretical analysis. Our future research topics focus on fully distributed event/self-triggered control for linear/nonlinear multiagent systems to gain a better understanding of fully distributed control.

CHAPTER 3

2. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

To focuses on the conception of a monitoring system able to detect and classify jamming and protocol-based attacks and achieve this goal, we proposed to outsource the attack detection function from the network to protect and used an antenna to monitor the spectrum over the time. The Wi-Fi network and the attacks were carried out in an anechoic chamber to avoid disturbing other Wi-Fi communication networks in the vicinity. The spectra highlight that the frequencies of interest belong to the communication channel between 2.402 and 2.422 GHz. Focusing the analysis on this 20-MHz frequency band permits to construct a classification model to overcome the problems induced by the utilization of the adjacent channels that can be or not occupied by other Wi-Fi communications. On these frequencies, the proposed estimation model shows good results in the prediction of attacks. In addition, the correction using the K spectra nearest in time permits to correct most of the miss classification.

The development of connected devices and their daily use is presently at the origin of the omnipresence of Wi-Fi wireless networks. However, these Wi-Fi networks are often vulnerable, and can be used by malicious people to disturb services, intercept sensitive data, or to gain access to the system. In railways, trains are now equipped with wireless communication systems for