EasyChair Preprint

№ 2737

Efficiency Network Security Analysis for Data Mining

Y Nafeesa

February 21, 2020

EFFICIENCY NETWORK SECURITY ANALYSIS FOR DATA MINING

ABSTRACT:

Several Network systems are suffering from various security threats including network worms, large scale network attacks, etc, and network security situation awareness is an effective way for solve these problems. The general process is to perceive the network security events happened in a certain time period and cyberspace environment, synthetically manipulate the security data, analyses the attack behaviors systems suffered, provide the global view of network security, and assess the whole security situation and predict the future security trends of the network.

KEYWORDS: Network Security, Data Mining, Network Security Situation Awareness(NSSA), Intrusion Detection System, IDS.

1. INTRODUCTION

Network security has been the eternal hot research spot, and it has undergone three phases: defence, detection and fault. However, there are still some security problems left, such as the complicated structure, and the kittle network attacks, so the network security issues are becoming more and more austere. The existed professional network security means, like IDS, Firewall and VDS cannot reflect the security status of the network.

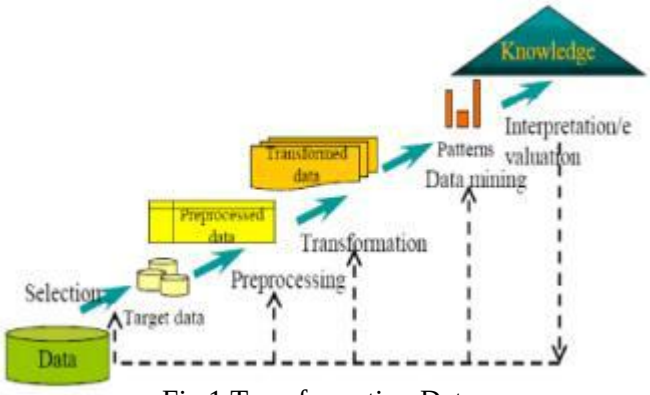

Fig.1.Transformation Data

After that, Tim Bass, the American outstanding network security expert, proposes the concept of network security situational awareness (NSSA), and establishes the framework of network situation awareness, which aims to solve the existed network security problems from a new point of view. The realization of situational awareness is divided into three layers: perception of elements in current situation, comprehension of current situation and projection of future status.

In order to evaluate the network security status of a large scale network and analyze the influence on network security of attacks or combination of them, a layered network security situational awareness realization model is proposed. Data mining (DM), also called Knowledge-Discovery and Data Mining, is the process of automatically searching large volumes of data for patterns using association rules It is a fairly recent topic in computer science but utilizes many older computational techniques from statistics, information retrieval, machine learning and pattern recognition. Here are a few specific things that data mining might contribute to an intrusion detection project. Data mining is becoming an important component in intrusion detection system. Different data mining approaches like classification, clustering etc are frequently used to analyze network data to gain intrusion related knowledge. This section will elaborate on several of these data mining techniques and will describe how they are used in the context of intrusion detection.

2. RELATED WORK

Classification algorithms can be used for both misuse and anomaly detections. In misuse detection, network traffic data are collected and labelled as “normal” or “intrusion”. This labelled dataset is used as a training data to learn classifiers of different types (e.g., SVM,NN,NB, or ID3) which can be used to detect known intrusions. In anomaly detection, the normal behaviour model is learned from the training dataset that are known to be “normal” using learning algorithms. Classification models can be built using a wide variety of algorithms. Classification categorizes the data records in a predetermined set of classes used as attribute to label each record; distinguishing elements belonging to the normal or abnormal class. This technique has been popular to detect individual attacks but has to be applied with complementary fine-tuning techniques to reduce its demonstrated high false positives rate. Clustering is the process of labeling data and assigning it into groups. Clustering algorithms can group new data instances into similar groups. These groups can be used to increase the performance of existing classifiers. High quality clusters can also assist human expert with labeling. A cluster is 100% pure if it contains only data instances from one category. Clustering techniques can be categorized into the following classes: pairwise clustering and central clustering. Pairwise clustering (i.e., similarity based clustering) unifies similar data instances based on a data-pairwise distance measure. On the other hand, Central clustering, also called centroid-based or model- based clustering, models each cluster by its “centroid”. In terms of runtime complexity, centroid-based clustering algorithms are more efficient than similarity- based clustering algorithms. The main objective of association rule analysis is to discover association relationships between specific values of features in large datasets. This helps discover hidden patterns and has a wide variety of applications in business and research. Association rules can help select discriminating attributes that are useful for intrusion detection. It can be applied to find relationships between system attributes describing network data.

3. PROPOSED SYSTEM

New attributes derived from aggregated data may also be helpful, such as summary counts of traffic matching a particular pattern. Network Security, often referred to as simply “the Network,” is the delivery of on-demand computing resources i.e. everything from applications to data centers over the Internet on a pay-for-use basis [1]. It is typically defined as a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. In Network Security, the word Network is used as a metaphor for "the Internet," so the phrase Network Security means "a type of Internet-based computing," where different services such as storage and applications are delivered to an organization's computer systems or devices through the Internet. With the advancement of technology power consumption has been become the crucial factor towards the growth of Network Security.

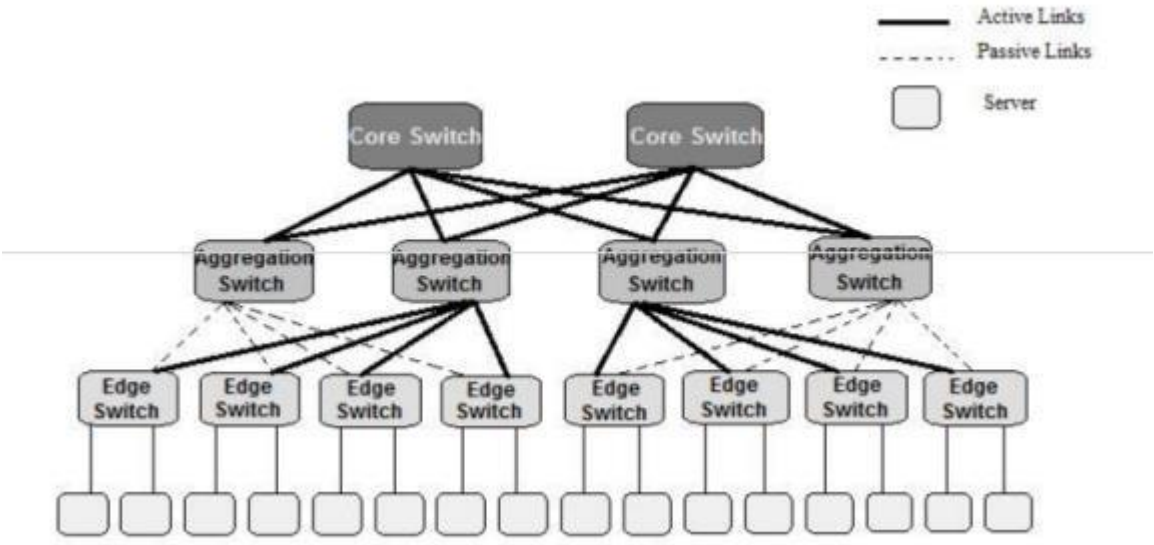

Fig.2.Proposed System

The data centers are the most significant part of the Network Security infrastructure and require more attention to maintain its reliability, availability, scalability, and most importantly the power consumption of individual resources together with QoS. The data center may contain hundreds-to- thousands of servers and other equipments (including switches, routers, etc.). The topological arrangement of these resources can be in three or more layers i.e. 3-tier or n-tier depending upon the size of the datacenter. Each of the resource requires a sufficient amount of power to process the request and provide services to different users, and also for cooling purpose. All the network resources whether they are in idle state or in working state, consume some specific amount of energy due to the always running state of CPU’s and other hardware part. To avoid unnecessary wastage of energy a simple strategy has been discussed in previous research work i.e. power down the unneeded links, switches, and servers. For this, we need some methodology to decide which subset of links, switches, and servers has to be active.

4. ANALYSIS

The data center is home to the computational power, storage, and applications necessary to support an enterprise business. The data center infrastructure is central to the IT architecture, from which all content is sourced or passes through. Proper planning of the data center infrastructure design is critical, and performance, resiliency, and scalability need to be carefully considered. Another important aspect of the data center design is flexibility in quickly deploying and supporting new services. Designing a flexible architecture that has the ability to support new applications in a short time frame can result in a significant competitive advantage. Such a design requires solid initial planning and thoughtful consideration in the areas of port density, access layer uplink bandwidth, true server capacity, and oversubscription, etc. For even simple traffic patterns, the formal model's solution time scales to the 3.5th power as a function of the number of hosts. The greedy bin-packing heuristic improves on the formal model's scalability. Solutions within a bound of optimal are not guaranteed, but in practice, high-quality subsets result. For each flow, the greedy bin-packer evaluates possible paths and chooses the leftmost one with sufficient capacity. Within a layer, paths are chosen in a deterministic left-to-right order, as opposed to a random order, which would evenly spread flows.

When all flows have been assigned (which is not guaranteed), the algorithm returns the active network subset (set of switches and links traversed by some flow) plus each flow path. For some traffic matrices, the greedy approach will not find a satisfying assignment for all flows, In this case, the greedy search will have enumerated all possible paths, and the flow will be assigned to the path with the lowest load. Like the formal model, this approach requires knowledge of the traffic matrix, but the solution can be computed incrementally, possibly to support on-line usage. In the first step, CARPO takes the data rates of the traffic flows in the previous consolidation periods as input and analyzes the correlation relationship between different traffic flows by using method in which if the traffic is negatively correlated, then the path is consolidated of the respective traffic flows, otherwise no need to consolidate the path . In the second step, based on the correlation coefficients from the previous analysis, CARPO uses the 90-percentile data rate of each link in the previous period to consolidate the traffics under the link capacity constraint. After the consolidation, unused switches and ports are turned off for power savings.

In the last step, CARPO adapts the data rate of each active link to the demand of the consolidated traffic flows on that link, such that more power savings can be achieved for the DCN. It is the layer where the servers physically attach to the network. The server components consist of 1RU servers, blade servers with integral switches, blade servers with pass through cabling, clustered servers, and mainframes with OSA adapters. The edge layer network infrastructure consists of modular switches, fixed configuration 1 or 2RU switches, and integral blade server switches. Switches provide both Layer 2 and Layer 3 topologies, fulfilling the various servers broadcast domain or administrative requirements.

CONCLUSION

SVMs classify data by using these support vectors, which are members of the set of training inputs that outline a hyper plane in feature space. The implementation of SVM intrusion detection system has two phases: training and testing. SVMs can learn a larger set of patterns and be able to scale better, because the classification complexity does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamically whenever there is a new pattern during classification. The main disadvantage is SVM can only handle binary- class classification whereas intrusion detection requires multi-class classification.

REFERENCES

[1] Tim Bass, "Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems",
Proceedings of 1999 IRIS National Symposium on Sensor and Data Fusion, University of The Johns Hopkins, America, pp. 1-6, 1999.

[2] Quantification Of Network Security Situational Awareness Based On Evolutionary Neural Network.

[3] Data Mining Techniques for (Network) Intrusion Detection Systems Theodoros Lappas and Konstantinos Pelechrinis

[4] Exploiting Efficient Data Mining Techniques to Enhance Intrusion Detection Systems Chang-Tien Lu, Arnold P. Boedihardjo, Prajwal Manalwar

[5] J. Cannady. Artificial Neural Networks for Misuse Detection. National Information Systems Security Conference, 1998.

[6] S. Mukkamala, G. Janoski, A. Sung. Intrusion Detection Using Neural Networks and Support Vector Machines. Proceedings of IEEE International Joint Conference n Neural Networks, pp.1702-1707, 2002

[7] L u a n, J. Data Mining and Its Applications in Higher Education. – New Directions for Institutional Research, Special Issue Titled Knowledge Management: Building a Competitive Advantage in Higher Education, Vol. 2002, 2002, Issue 113, 17-36.