ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

Review on Data Analytics used in Insider Threat Detection

Nita Radhakrishnan [1], Mehul Awasthi [2], Bhargav Sundararajan [3], Deepthi Peri [4]

U.G Student, Department of Computer Science and Engineering, SRM University, Kattankulathur, India [1]

U.G Student, Department of Computer Science and Engineering, SRM University, Kattankulathur, India[2]

U.G Student, Department of Computer Science and Engineering, SRM University, Kattankulathur, India [3]

U.G Student, Department of Computer Science and Engineering, SRM University, Kattankulathur, India [4]

ABSTRACT: Confidential data is always vulnerable to attacks from all kinds of sources. The most devious and unpredictable of these attacks, are those that originate from within the trusted circle. These attacks are called insider attacks and were once opaque to detection. But with the advent of technology, there have been various methods devised to tackle this problem. The most prominent and efficient ones have evolved from the roots of Data Analytics. This paper aims at addressing some of the established algorithms with emphasis on how a particular sector can handle such threats efficiently.

KEYWORDS: SVM, Insider threats, Data Analytics, Classifiers, BBAC

I. INTRODUCTION

Over the decades data has faced threats from both internal and external sources. However, recent studies have indicated that the most malicious of these threats come from insiders. According to sources from the internet, an insider is a person within a group or organization, especially someone privy to sensitive information. Also, CERT Division defines a malicious insider as 'a current or former employee, contractor, or business partner'. Additionally, the insider 'has or had authorized access to organization's network, system, or data', and 'has intentionally exceeded or intentionally used that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems'. These threats have been cited as the most insidious of them all, because it is difficult to identify both the attack and the perpetrator of the attack. This is so, because these attackers mask their behavior to blend in with the regular activities. The insider attacks fall into five categories- sabotage, threat, fraud, espionage and unintentional insiders. These different kinds of attacks will be discussed in later sections of this paper. Ways in which these threats are identified can be both technical and non-technical. According to the study [1], a system can be compromised by three aspects; knowledge, technical facility and trust. To elaborate on these aspects, if a person gains knowledge about how the internal system works then they gain the privilege to use the vulnerabilities to their advantage. Secondly, an organization grants login credentials to the employees of the company to serve as an authentication to access sensitive information. However, this privilege can be abused by employees to use the information in ways that could harm the organization. Lastly, an organization entrusts certain employees with the responsibility to manage delicate data pertinent to the organization and its functioning. This position of responsibility can be misused by employees negatively. It is due to this reason that the US government has proposed the Insider Threat program (InTP). An insider threat program works continuously by detecting changes in the activities, detecting an incident as it occurs, and once detected it responds to the incident effectively. On analyzing the incidents, the information is then used as a lesson to prevent any such incidents in the future. This paper aims at addressing the different methods used by tools to detect these threats and more specifically how Data Analytics plays a major role in such detection.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

II. LITERATURE SURVEY

An insider threat program was developed as a defense against the malicious insider threats. This program was initiated by the US government. An Insider Threat Program is operated to prevent incidents caused by employees of that organization irrespective of where they from. These InTPs are trained to identify certain profiles that serve as indicators to malicious insiders. This section of the paper covers some of the most popular profiles.

A. IT Sabotage

This kind of a compromise in the system is done by technically savvy people belonging to the organization. A survey of previously recorded data suggests that these people may be those who stand terminated from the company. Such employees set up backdoors when they are hired and in that way, re-entering the system when they are no longer employees becomes much less of an effort. According to the study [1], it was stated that 86 per cent of offenders in the CERT database were employed in technical positions. Furthermore, 90 per cent had privileged access to the systems, databases, and networks they worked with every day. These are the people trusted to keep an operation running day to day. Should the individual find the motivation, they can carry out attacks and conceal them with relative ease. The primary motivation across the IT Sabotage profile was revenge.

B. Theft of IP

An Intellectual Property is any material of intellect such as research or creative work such as art or music that belongs exclusively to an individual or an organization, depending on who the creator of the intellectual material is. Theft of such material is called Theft of IP. This particular profile has been occupying a significant place in the world of cybercrime. The United States leads the world in the creation and selling of IP products to buyers nationwide and internationally. Examples of copyrighted material commonly stolen online are computer software, recorded music, movies, and electronic games. This crime is usually committed by researchers and scientists. Their actions will most likely take place within 60 days of leaving an organization. The study states that the motivations for stealing the IP have been to start their own business, taking the information to a new job, and giving it to a foreign government.

C. Insider Fraud

The cases of insider fraud, happen to be the largest group that was profiled by CERT. Companies are often reluctant to accept such incidents, mainly because the information itself could sensitive and classified and also because it could lead to the company's public embarrassment which in turn would decrease the public's confidence in the organisation. One of the most sought after target, is the personal information of the employees. It is therefore imperative that cases of insider fraud be kept to a minimum and a security structure should be present to prevent these cases from happening in the first place.

D. Espionage

Espionage is something of a concern to all organisations. It can be done by both the technically skilled and non-technically inclined employees. A person involved in espionage, usually extracts information that may be confidential and important to the functioning of the organisation they work in, and gives this information to a competitor, which eventually harms the organisation the person works for. There is no clear-cut mechanism present to identify the profiles that may belong to this category. Often the people committing acts of espionage are dissatisfied with their careers, or are miffed by the lack of opportunities present to them. Financial problems also lead to such cases of espionage. A person looking to make a quick buck can sell the proprietary information of the company they work for, in exchange for money.

E. Unintentional Insider Threat

Unintentional insider threats are posed by people who have or had access to an organisation's confidential data, and who unwittingly and without any malicious intent cause harm to the confidentiality and security of an organisation's information and its information systems. Human errors, fatigue and drowsiness, mental overload are all major factors in unintentional insider threats. Efforts have to be taken by organisations to educate their workforce on the perils of such threats and must suggest ways to prevent, mitigate and minimise unintentional insider threats.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

III. OVERVIEW OF EXTANT METHODS USED TO DETECT INSIDER THREATS

The techniques used to detect insider threats depend on the type of data set targeted and the source. User profiling techniques are the most commonly used to identify the threats and their sources. For example, the user profiling techniques used to detect insider threats in UNIX or Linux environments use the Schonlau dataset. The Schonlau dataset is made up of truncated sequence of user commands. On the other hand, a variety of audit sources can be taken advantage of in the environment of the Windows operating system. The wide range of data includes registry accesses, system calls which occur when users execute applications and a combination of processes and windows data, such as the window title, how long a window has been open and so on. Ascribing a network level event to a threat is a little tricky because on the network level the data that can be interpreted is not very clear to the distinct user. Even so, network level events are helpful when it comes to detecting malicious or specious activities such as downloading of immense amount of data that an insider does not need to possess, or the propagation of data outside the organization's network.

There is an inherent difficulty in comparatively evaluating competing methods and approaches due to the dearth of uniform test data which is close to the ground truth. The Schonlau dataset has been used widely but it is still far from being suitable for an objective evaluation of the insider attack detection algorithms [4]. Here are some audit sources and algorithms used by researchers to detect threats-

A. One Class Support Vector Machine

One Class Support Vector Machine (OCSVM) is capable of estimating high density regions from data samples, and is considered a special class of SVM. Only positive information is used to train in the OCSVM methodology, whereas the negative information is unknown. The function computed in this method, takes the value +1 in high density regions and -1 in the other regions. The value +1, usually denotes a normal class, whereas -1 denotes an abnormal class. The data points are mapped onto a feature space by the OCSVM, which is determined by a kernel function. The kernel functions build a hyperplane which separates the data points from the origin. The number of outliers in OCSVM can be controlled as required. This approach is however relevant to bounded length, static data streams. Insider threat related data is typically continuous and the patterns tend to evolve over time. Effective classification models must therefore be adaptive and highly efficient in order to build the model from large amounts of evolving data.

B. Support Vector Data Description

The Support Vector Data Description (SVDD) works in a manner wherein it plots a spherical boundary around a dataset with the help of a RBF kernel. Just like other one class classifiers, one of the prime objectives of the SVDD is to detect outliers without the presence of negative instances in the dataset. The flaw with many of the other algorithms is that they compute the complete density of a given dataset to detect the outliers. The problem is that this results in the need of a huge data set. To circumvent this issue the SVDD only calculates the boundaries. The SVDD defines a model with a closed boundary dataset, which can be considered to be a hypersphere. The primary goal is to minimize the volume of the sphere, such that the sphere contains all the given training objects. The use of flexible kernel functions can do away with using the rigid hypersphere model. OCSVMs behave similar to SVDD with RBF kernel.

C. Unix Audit Events

Protecting computer networks consists of implementing preventative measures, but especially properly implementing detection methods. There are simple ways of implementing security events on UNIX systems. The first can be done by simply monitoring file changes. Changes done to files that are not changed that often could indicate compromise. Since usually software is a weak spot in security defenses, a crash of software often indicates an unexpected event. Sometimes caused by bad hardware (e.g. a bad memory module), but usually due to bad memory management in the software itself. Malicious people try to abuse these weak spots to load special crafted code. There monitoring crashing software can be very helpful to discover potential attacks or intrusions.

D. Honeypots and decoy technologies

As the name suggests, the honeypot technology is a security mechanism used to detect, deflect or counteract attempts at unauthorized use of information systems. Data that seems to be legitimate or of some advantage to the attackers-but in

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

reality is irrelevant-is isolated and monitored to detect and block attackers. This is how honeypot and decoy technologies work. They help in the detection of potential attackers and hence in preventing further attacks from them.

E. Windows Registry Access

The Windows registry holds a wealth of information when a device is connected to a system. This information can be very useful to detect any illicit activity done by employees, such as uploading sensitive information onto an unauthorized USB flash drive.

F. Windows Usage Events

Having the knowledge about how people are using their systems, can help detect any malicious activity they might be up to. Any activity that seems unusual and off from regular doings, such as the downloading of large amounts of sensitive information, or the plugging in of an unauthorized flash device, can be found out. This can help prevent the threat of theft of sensitive information.

IV. DATA ANALYTICS IN DETECTION OF INSIDER THREATS

Data Analytics has breached almost every raging industry and has made almost every process twice as efficient as it used to be. In recent years, Data Analytics has earned its much deserved popularity in the field of security. This section deals with how a particular concept of Data Analytics helps in detecting insider threats typically at the network level.

Beginning with introducing the notion of data science, it is the ability to procure data, understand it and extract meaning from it, visualize and communicate the meaning or value of the procured data to the end user. A specific aspect of this science, called the Behavior-Based Access Control, is applied to detect threats. According to certain studies [2], this Behavior-Based (or Anomaly-Based) threat detection Threat Detection depends on the assumption that attack behaviors differ enough from normal activity that malicious actions can be detected and identified. Tools using this method begin by creating models of behavior patterns that represent “normal” behavior for the networks, systems, applications, end users and devices that make up the environment in which they’re installed. They then look for deviations from that pattern.

At the network level, the Behavior-Based Network Access Control (BBNAC) performs clustering to identify network behavior profiles [3]. In contrast to BBAC, BBNAC only operates at the network layer. They are used to analyze typical network level behaviors such as the benignity of HTTP requests. BBAC procures its data from protocols such as TCP, UDP, HTTP, XMPP, SMTP, OCSP, DNS Servers, email and chat messages, PDF documents, and wiki pages. This data is used as the training data set and the method by which BBAC obtains this data is called *feature extraction*. Feature Extraction turns raw data into a form that can be manipulated by machine learning algorithms, in order to detect observable patterns. Next *featureenrichment* is done in order to manage this data in order to evaluate patterns in the activities. From here, BBAC uses both unsupervised learning and supervised learning to perform the clustering. Unsupervised learning methods such as K-Means++ is used to map or cluster the activities to the actor. Within a cluster, supervised learning methods such as SVM are used. Other methods such as Decision Trees and Support Vector Machines are readily available on the WEKA machine learning tool. The final stage is *action selection* where the reaction to a classification based on SVM is done. The reactions can range from blocking requests, to notifying the administrator regarding some unusual behavior.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

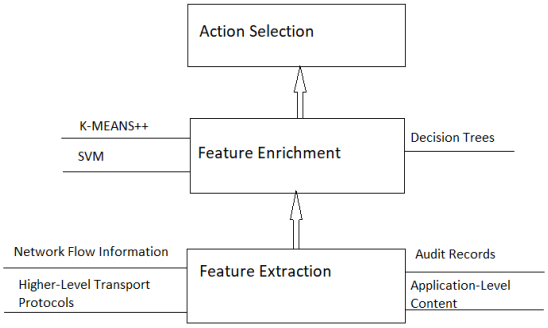

Fig.1. Stages of BBAC

The BBAC uses relatively novel approaches to detect intrusions. It efficiently makes use of K-Means, K-Means++, Decision Trees and Support Vector Machines (SVMs) from the WEKA tool in order to cluster, and then accurately classify the behavior of actors. It uses a scalable processing platform to effectively procure the data available on the networking platform which is on an ever-widening horizon. The essence behind the effectiveness of the BBAC in cyber security is its scalability and extensibility in almost every stage starting right from the training to action selection.

V. CONCLUSION

This paper has covered the nuances of insider attacks and their malicious effects on the regular operations of an organization. The various profiles that serve as indicators to an insider attack have also been briefly explained. Over the years various methodologies have been developed in order to combat the insider threats. These methodologies span over a variety of operating systems, and have been highlighted in this paper. In addition to that, this paper has also covered a very important aspect in technology that plays a major role in intrusion detection- Data Science. Data Science offers many concepts that can be held useful in the field of security, more specifically the concept called Behavior-Based Access Control. This paper has elucidated the various stages involved in the BBAC intrusion detection and has also thrown light on the scalability of the BBAC mechanism which eventually is the reason behind its success.

REFERENCES

[1] Markus Kont, Mauno Pihelgas, Jesse Wojtkowiak, Lorena Trinberg, Anna-Maria Osula, “Insider threat detection study”, CCDCOE.
[2] Barbara Filkins, “The Expanding Role of Data Analytics in Threat Detection”, A SANS Whitepaper, October 2015
[3] Michael Mayhew, United States Air Force Research Laboratory, Rome NY, Michael Atighetchi, Aaron Adler, Raytheon BBN Technologies Cambridge, MA, Rachel Greenstadt, Drexel University, Philadelphia, PA, “Use of Machine Learning in Big Data Analytics for Insider Threat Detection”.
[4] Malek Ben Salem, Shlomo Hershkop, Salvatore J. Stolfo, “A Survey of Insider Attack Detection Research”.