Big Data Analysis System Concept for Detecting Unknown Attacks

Ayesha Sani
Department of computer science
Jhulelal Institute Of Technology

Raana Syeda
Department of computer science
Jhulelal Institute Of Technology

Abstract— **Nowadays threat of previously unknown cyber-attacks are increasing because existing security systems are not able to detect them. Previously, leaking personal information by attacking the PC or destroying the system was very common cyber attacks . But the goal of recent hacking attacks has changed from leaking information and destruction of services to attacking large-scale systems such as critical infrastructures and state agencies. In the other words, existing defence technologies to counter these attacks are based on pattern matching methods which are very limited. Because of this fact, in the event of new and previously unknown attacks, detection rate becomes very low and false negative increases. Today's attacks are prepared by advanced technologies are not detected until the damage has been occurred. Now the challenge is collecting and analyzing the Big Data fast enough to contain threats and perform last remediation. To defend against these unknown attacks, which cannot be detected with existing technology, a new model based on big data analysis techniques that can extract information from a variety of sources to detect future attacks is proposed . The expectation with this model is future Advanced Persistent Threat (APT) detection and prevention .**

I. INTRODUCTION

Hacking in the past were use to leaked personal information but nowadays hacking targets companies, government agencies etc. This kind of attack is commonly called APT(Advanced Persistent Threat). APT targets a specific system and analyses vulnerabilities of the system for a long time. Therefore it is hard to prevent and detect APT than traditional attacks and could result in massive damage. Up to today, detection and protection of systems from defending against cyber-attacks were done through firewalls, intrusion detection systems, intrusion prevention systems, anti-viruses solutions, database encryption, DRM solutions and etc. Moreover, integrated monitoring technologies for managing system logs were used. These security solutions are developed based on signatures and blacklist. However, according to various reports, intrusion detection systems and intrusion prevention systems are not capable of protecting systems against APT attacks because there are no signatures. Therefore to overcome this issue, security communities are beginning to apply heuristic and data mining technologies to detect previously unknown attacks. In this ,a new model based on bigdata analysis technology has to prevent and detect previously unknown APT attacks.

APT Attacks.

APT attacks penetrate into the target system and persistently collect valuable information by using social engineering, zero day vulnerabilities and other techniques . It can damage national agencies or enterprises. They are also used as a cyber weapons. Instead of Targeting ordinary desktops or servers they target industrial control systems .

APT attack is usually done in four steps: Intrusion, Searching, Collection and Attack.

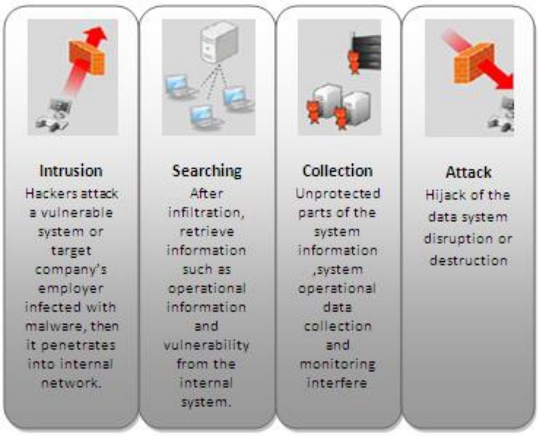
Intrusion
Hackers attack a vulnerable system or target company's employer infected with malware, then it penetrates into internal network.
Searching
After infiltration, retrieve information such as operational information and vulnerability from the internal system.
Collection
Unprotected parts of the system information ,system operational data collection and monitoring interfere
Attack
Hijack of the data system disruption or destruction

Figure1.The Sequence of APT attacks.

A) **Intrusion Step** . In the intrusion step the hacker probes for information about the target system and prepares the attack.

B) **Searching .** To get the access to the system, the attacker searches for users with high access privileges such as administrators and use various attack techniques such as SQL injection, phishing, farming and social engineering to hijack their accounts Searching is done after the hacker has gained access to the system. Hacker analyses system data such as system log for valuable information and look for security vulnerabilities then it can be exploited for further malicious behaviours.

C) **Collection** . In this next step, after the hacker has located valuable information in the system such as confidential documents etc, then, he installs malwares such as rootkits, backdoors to collect system data and maintain system access for the future.

D) **Attack** . In this final step, the hacker leaks data and destroys target system using acquired privileges. Leaked information can be used for developing other additional security vulnerability exploits.

Because APT exploits use zero-day vulnerabilities and obfuscation methods, Anti-Virus program, IDS and IPS are difficult to detect such exploits

Examples of recent APT attacks are Stuxnet, RSA Secure ID hacking and the Night Dragon. Stuxnet was a very intelligent malware that was developed to attack Iran's nuclear facilities and make them malfunction.

II. SYSTEM OVERVIEW

The system focuses on the following areas-

- Data Collection and Creation of Network
- Analysis of Data
- Detection of Unknown Attacks and generate an alert

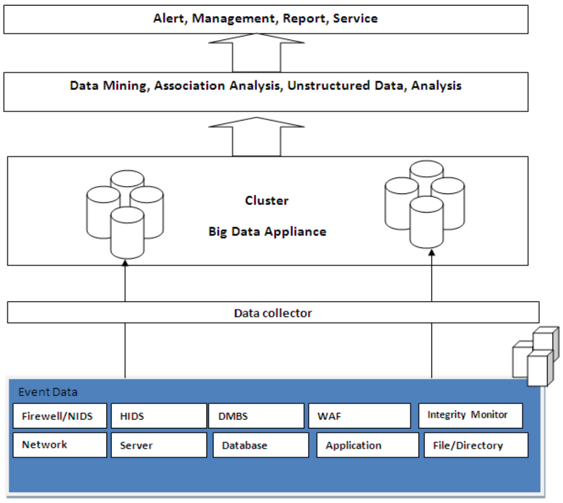

Figure 2.System Overview

Figure 2. shows the Big Data Analysis System Model for Detecting Unknown Attacks . As shown in the fig, from various sources the data is being collected. The extracted data is taken as input and is provided to the system for pre-processing. After preprocessing the data it is analysed . The Analysis is done on the basis of Behaviour Matching . Genetic Algorithm is used for behavior matching . If any unknown behavior is found then an alert will be generated by the system . Snort is been used for Detection .

III. DATA ANALYSIS AND DETECTION OF UNKNOWN ATTACKS

A) Data Collection and Creation of Network

Data collection step collects event data. The Event data is collected from firewalls and log, ServersApplication, behaviour, status information (date, time, inbound/outbound packet, daemon log, user, behaviour, process information etc.) from anti-virus, database, network device and system. Data appliance issued to store the collected data . The Network is been created by client server application Through this the data will be send through.

B) Analysis of Data

The Clone detection is defined as a mechanism for a WSN to detect the existence of inappropriate, incorrect, or anomalous moving attackers. In this the path is checked whether the path is authorized or unauthorized. If path is authorized the packet is send to valid destination. Otherwise the packet will be deleted.

Constructing Inter-Domain Packet Filters : If the packet is received from other than the port no it will be filtered and discarded. This filter only removes the unauthorized packets and authorized packets send to destination.

***Behaviour Matching using Genetic Algorithm* :** Here Genetic Algorithm is used for Behaviour Matching .The Behaviour of the received packet is matched with the already known behaviours . If the behaviour is not Matched then it is Considered as unknown .

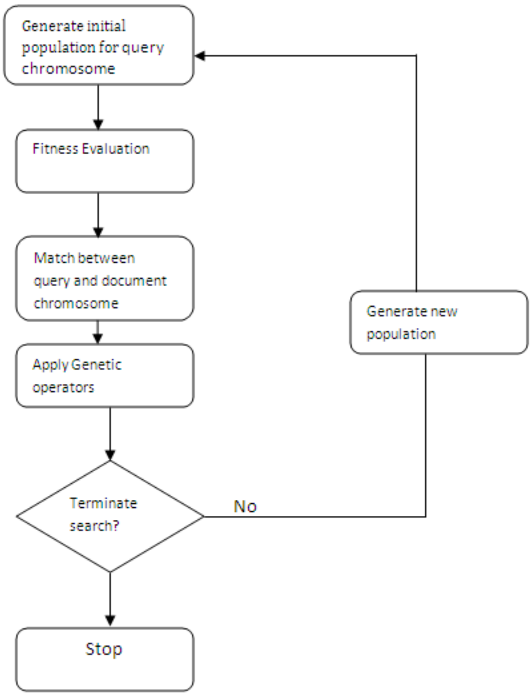
Generate initial population for query chromosome
Fitness Evaluation
Match between query and document chromosome
Apply Genetic operators
Terminate search?
No
Generate new population
Stop

Figure 3. System Overview

Initially all the known attacks set is created . If any attack comes , first it is checked whether it is known or unkown i.e it is checked whether it matches with the known attacks set or not . If a match is found with the known attack set then it will get prevented as solutions are already present for them But if matching does not found then an alert will be generated by the Detection Engine and reported to Administrator . The data sent will then get Discarded.

***Association Analysis* :** Association rule learning is a method for discovering interesting correlations between variables in large databases. Association rule learning is being used to help in monitoring system logs to detect intruders and malicious activity.

***Update Database* :** The Database is updated after Detection of unknown attack .

C) Detection of Unknown attacks and generating an Alert

Generation of an Alert message

An alert message is generated if any unknown attack is found.

1. Alert is indication for detection of attack.
2. Alert is generated, when known or unknown attack found.
3. Attack message is displayed on system if attack is found.

Snort is Used for Detection

Components of Snort are as follows:

1. Packet Decoder
2. Preprocessors
3. Detection Engine
4. Logging and Alerting System
5. Output Modules

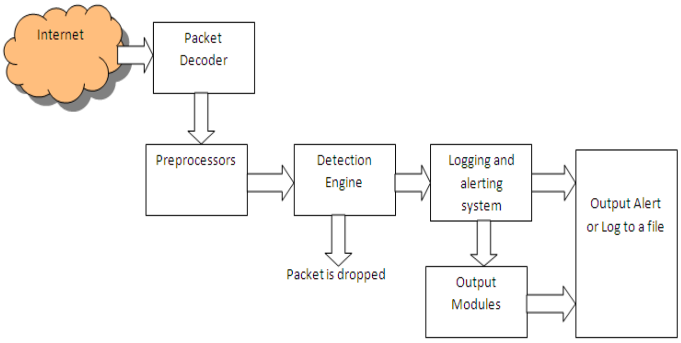

Figure 4. Components of Snort

1. Packet Decoder

The packet decoder takes packets from different types of network interfaces and prepares the packets to be preprocessed or to be sent to the detection engine. The interfaces may be Ethernet, SLIP, PPP and so on.

2. Preprocessors

Preprocessors are components or plug-ins that can be used with Snort to arrange or modify data packets before the detection engine does some operation to find out if the packet is being used by an intruder. Some preprocessors also perform detection by finding anomalies in packet headers and generating alerts.

Preprocessors are very important for any IDS to prepare data packets to be analyzed against rules in the detection engine. Hackers use different techniques to fool an IDS in different ways. For example, we may have created a rule to find a signature scripts/iisadmin in HTTP packets. If we are matching this string exactly, we can easily be fooled by a hacker who makes slight modifications to this string . The preprocessors are used to safeguard against the attacks. Preprocessors in Snort can defragment packets, decode HTTP URL, re-assemble TCP streams and so on. These functions are a very important part of the intrusion detection system

3. The Detection Engine

The detection engine is the most important part of Snort. Its responsibility is to detect if any intrusion activity exists in a packet. The detection engine employs Snort rules for this purpose. The rules are read into internal data structures or chains where they are matched against all packets. If a packet matches any rule, appropriate action is taken; otherwise the packet is dropped. Appropriate actions may be logging the packet or generating alerts. The detection engine is the time-critical part of Snort. Depending upon how powerful your machine is and how many rules we have defined, it may take different amounts of time to respond to different packets. If traffic on our network is too high when Snort is working in NIDS mode, we may drop some packets and may not get a true real-time response. The load on the detection engine depends upon the following factors:

1. Number of rules
2. Power of the machine on which Snort is running
3. Speed of internal bus used in the Snort machine
4. Load on the network

4. Logging and Alerting System

Depending upon what the detection engine finds inside a packet, the packet may be used to log the activity or generate an alert. Logs are kept in simple text files, tcp-dump style files or some other form. All of the log files are stored under /var/log/ snort folder by default. We can use l command line options to modify the location of generating logs and alerts. Many command line options can modify the type and detail of information that is logged by the logging and alerting system.

5. Output Modules

Output modules or plug-ins can do different operations depending on how we want to save output generated by the logging and alerting system of Snort. Basically these modules control the type of output generated by the logging and alerting system

III. CONCLUSION

In this paper a Big Data System Model for reacting to previously unknown cyber threats is proposed.

Recent unknown attacks easily bypass existing security solutions by using encryption and obfuscation. Therefore there is a need to develop a new detection methods for reacting to such attacks To defend against these unknown attacks, which cannot be detected with existing technology the model is proposed .This gives a model for reacting to previously unknown cyber threats.

IV. REFRENCES

[1] Tai-Myoung Chung Sung-Hwan Ahn, Nam-Uk Kim. "`Big data analysis system concept for detecting unknown attacks'". Technical report, February 2014

[2] B Gupta, K Jyoti - International Journal of Computer Science & …, 2014 - ijcsit.com

[3] Sachin S. Patil et al, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.5, May- 2014

[4]N Virvilis, O Serrano,L Dandurand - infosec.aueb.gr

[5] R . Magoulas and B. Lorica, Introduction to Big Data, Release 2.0 (Sebastopol OReilly Media , February 2009