ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

A Data Mining Concept for Detect Attacks

Feba Babu [#1],Kishore Sebastian [*2]
[#]*PG Scholar*
St. Josephs College of Engineering and Technology, Palai
febabinu04@gmail.com
[*] *Assistant Professor*
St. Josephs College of Engineering and Technology, Palai
[2] kishorekinattukara@gmail.com

Abstract— Now, Associations and nations are getting to be defenseless against a wide assortment of security breaches against their data foundation. Cyber threat is obvious from the expanding rate of cyber-attacks against PCs and basic founda-tion. The count of cyber-attacks by people and malicious pro-gramming very quickly. However cyber attacks are expanding in light of the fact that current cyber security technologies are not equipped for distinguishing them. Existing cyber security technologies are Authentication, Encryption, Anti-virus, Web application firewalls, Firewall, Intrusion Detection System these technologies are depend on pattern matching methods which are extremely constrained. In this paper focus on identification and prevention of APT attack and propose a data mining approach CRISP-DM methodology for detect attacks with ARIMA statistical models.

I. INTRODUCTION

Big data systems are extremely basic piece of present day organisations,because now a days we are living in computerized world so at regular intervals a huge number of information is getting generate.This plenteous information just beyond the technology's.So second by second hacking attacks are persistently expanding in the internet. Hacking is nothing but it is an endeavor to misuse a PC framework or a private system inside a PC.In any case, late hacking targets companies,government organizations. This sort of attacks is regularly called APT[1](Advanced Persis-tent Threat) and it a network attack, which can unapproved individual accesses a network and remains there undetected for a drawn out stretch of time. This paper mainly focus on another model in light of big data analysis technology anticipate and identify beforehand unknown APT attacks.

The rest of the paper is organized as follows. In Sect.II, Existing cyber security technologies. In Sect.III,APT attacks and groups. In Sect.IV, propsed system. Finally,paper con-cluding remarks and suggestions for future study.

II. EXISTING CYBER SECURITY TECHNOLOGIES

In past years,Shielding PC or network from malware keeps on being a testing issue.The apparatuses and methods utilized to handle cyber security concerns[2] are:

A. Authentication

This basic cyber security method expects to confirm the character of client in view of the accreditation's put away in the security space of the framework.In machine authentication,the number of Internet-empowered gadgets are increased,solid machine authentication is pivotal to permit secure correspondence in home computerization and other arranged conditions.

Another main issue with password-based authentication,it will not considered to give enough solid security to any framework that contains delicate information.The principle challenge experienced in verifying procedure is upsetting endeavors of unapproved individuals to listen in on the confirming message.

ncryption

Encryption makes an interpretation of information into another frame, or code, so just individuals with access to a secret key or secret word can read it.In here happening two type of encryption one is Symmetric encryption and Asymmetric encryption,Symmetric encryption uses a similar key with the end goal of message encoding and decoding, and the security level is like that of the key.Asymmetric encryption uses an open key to encode the message and a private key to unscramble the same. Now a days a dominant part of security protocols are utilizing asymmetric encryption for appropriation of keys.

C. Anti-virus

The main dangers of PC infections or bothersome short projects that trigger undesirable orders without

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

the express assent of client have accepted huge extents.Most viruses have been built to target Windows working framework as it is the most favored registering stage of masses.Apple and Linux clients are a typical security control,using for HTTP applications.It is utilized by en-deavors to ensure Web applications against zero-day adven-tures, pantomime and known vulnerabilities and assailants.It act as shields of web site from SQL infusion, cross-site scripting (XSS) and zero-day assaults, including OWASP-distinguished vulnerabilities and dangers focusing on the application layer.

WAF can identifies and blocks attacks utilizing both posi-tive and negative access control. Positive access control is a innovation that blocks all with the exception of characterized safe patterns,furthermore, negative access control block just predefined pernicious patterns.

E. Firewall

Firewall[4] is a tool,it can be utilized to upgrade the security of PCs associated with a system, for example, LAN or the Internet.Firewalls can additionally improve security by empowering granular control over what kinds of frame-work capacities and procedures approach networking assets. These firewalls can utilize different kinds of marks and host conditions to permit or deny movement.

Despite the fact that they sound complex, firewalls are generally simple to introduce, setup and work.In the begin-ning stage firewall is situated at fringe of the system and can be utilized as a defender for the internal system. Likewise, firewall is utilized as an essential security answer for this day.Any way a firewall just gives least security from attacks.

Main expectation of an APT attack is to take information and listening in as opposed to make harm the system or association.

APT attacks tend to target center mechanical control frameworks rather than common work areas or servers. Also,APT attacks are utilized as digital weapons between countries. Digital security is turning into a center part of national well being. Attacks modern frameworks and causing failing of postulations foundations can make open disar-ray the country.As indicated by different reports,intrusion prevention systems and intrusion detection systems are not equipped for guarding against APT attack in light of the fact that there are no marks.

can likewise go under the assault of infections only worked for such working frameworks.

D. Web application firewalls

Web application firewalls[3](WAF)

F. Intrusion Detection System

An IDS[5] screens network traffic for suspicious action.A intruder detection systems designed for either a system or a particular gadget. A network intrusion detection system (NIDS) screens inbound and outbound movement, and in addition information exchanges between frameworks inside a system.

Intrusion recognition frameworks (IDS) can be grouped into various ways.The real arrangements are Active and passive IDS, Network Intrusion detection systems (NIDS) and host Intrusion detection systems (HIDS).IDS identify and alarms anomalous activities by pre-characterized run the show.These tenets depend on ordinary clients' practices and measurement data from framework logs.

Security technologies, for example, firewall, IDS, WAF fundamentally utilizepattern matching techniques that de-pend on pre-characterized rules.

III. APT ATTACK

Advanced persistent threat (APT) is a special kind of network attack in which an unapproved individual accesses a system and stay undetected for whatever length of time that they can.APT attack is a exceptional sort of attack .It can be utilization social engineering, zero day vulnerabilities and different methods to enter into the target framework and constantly gather important data[1].

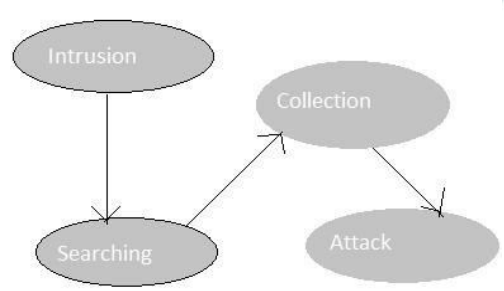

Fig. 1.Arrangement of APT attacks

Along these lines to conquer this issue, security specialists are starting to apply data mining technologies to recognize beforehand focused on attacks.In this paper propose another model in view of big data analysis technology to forestall and distinguish beforehand unknown APT attacks.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

1) Step1:Intrusion: The hacker scans for data about the objective framework and readies the attack,once to get the access to the framework, the attacker looks for clients with high access benefits, for example, administrators then utilize different attack strategies, for example,spoofing ,phishing etc.

2) Step2:Searching: After the hacker investigations on information framework , for example, framework log for profitable data and search for security vulnerabilities than can be abused for facilitate pernicious practices.

3) Step3:Collection: The hacker found some significant data in the framework, for example, secret files or something and so forth, at that point, the hacker introduces malwares, for example, rootkits, secondary passages to gather all infor-mation's and keep up framework access for what's to come.

4) Step4:Attack: After step3 hacker can be spills informa-tion.Annihilates target framework utilizing gained benefits. Spilled data can be utilized for creating other extra security weakness misuses.

A. APT Groups[6]

APT37 Targetsectors:chemicals,aerospace,healthcare,electronics.

APT34

Target sectors: financial, chemical, telecommunications, government. APT33

Target sectors:energy,Aerospace. APT32

Target sectors: consumer products, consulting and hospi-tality sectors.

APT30

Target sectors:Members from the Association of South-east Asian Nations (ASEAN)

APT29

Target sectors:foreign policy groups ,Western European governments,and other similar organizations.

APT28

Target sectors:eastern European countries and militaries,The Caucasus, particularly Georgia,North Atlantic

Treaty Organization (NATO) and other European secu-rity organizations and defense firms.

APT19

Target sectors:investment,Legal. APT18

Target sectors: Education, HighTech, Transportation,Aerospace and Defense.

APT17

Target sectors: international law firms ,information tech-nology companies,U.S. government.

APT16

Target sectors: government services,financial services industries,Japanese and Taiwanese organizations in the high-tech.

APT12

Target sectors: government,Journalists,defense Indus -trial base.

APT10

Target sectors: aerospace,Europe, and Japan, Construction and engineering.

APT5

Target sectors: Asia-Based Employees of Global Telecommunications,High-Tech Manufactur-ing, Regional Telecommunication Providers.

APT3

Target sectors:Telecommunications, Transportation,Construction and Engineering.

APT1

Targetsectors:Advertising and Entertainment, Naviga-tion,Aerospace,Engineering Services, International Or-ganizations.

IV. PROPOSED SYSTEM

Big data analysis utilizes different existing examination methods are statistical models,machine learning or natural language processing technology,Predictive analytics,data mining etc.

This paper mainly focus on data mining because this method is very helpful to recognize obscure new attacks with the help of financial engineering statistical model ARIMA Data mining technology innovation caues,analyze a lot of information to find patterns in the information and this data can be utilized for assist investigation to help answer complex business questions. APT attacks are difficult to identify or anticipate with current innovations so data mining big data analysis technology utilizing for extracting data from different sources to respond to unknown attacks "CRISP-DM" Cross-industry standard process for data mining technique which are utilized for big data analysis[7].

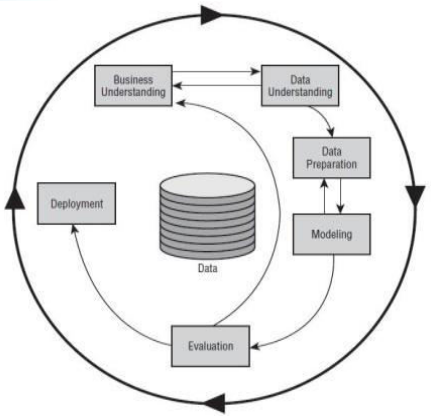

Fig. 2.Fig2:CRISP-DM process model

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

Fig2 Contains six phases,Here are the different stages for detecting APT attack by using CRISP-DM technique. Business Understanding : This first stage centers around understanding the venture goals and necessities from a hacker point of view,collects event data from fire-walls and behaviour, status information,log the gathered information is spared in big data machine, Afterward changing over this information into an data mining issue definition, and a preparatory arrangement intended to accomplish the goals.

Data Understanding : The data understanding stage begins with an underlying information accumulation and continues with activities to distinguish information quality issues and detect hidden information or any vulnerabilities.

In this paper using ARIMA model for find out vulnerabilities. ARIMA stands Autoregressive Integrated Moving Average,this approach provide to time series forecasting.ARIMA model[8] aims to describe the au-tocorrelations in the data.The accompanying strategy gives a valuable general approach.

step1:Plot the information. Distinguish any uncommon perceptions.

step2:On the off chance that important, change the information by utilizing a Box-Cox change to balance out the fluctuation.

step3:On the off chance that the information are non-stationary: take first contrasts of the information until the point when the information are stationary.

step4:Analyze the ACF/PACF.

step5:Attempt the AICc to find out a better model.

step6:Check the residuals,if it fail attempt a modified model.

step7:calculate forecasts when residuals look like white noise.

The automated algorithm just deals with stages 3 5. Hence this paper propose a finest methodology for predict APT attack by using ARIMA statistical model. Data Preparation: After the Data Understanding build the last data-set and transformation and data cleaning for modeling tools.This stage of process is Predictive analysis.

Modeling: In this stage, different modeling techniques are chosen and connected, this study focus on machine learning techniques.Evaluation:Toward the finish of this stage, a choice on the utilization of the data mining results ought to be come to.here checking accuracy and precision of data. If the result is not satisfy directly jump to business understanding. On the off chance that attack or abnormal behaviours are detected V.

Evaluation:Toward the finish of this stage, a choice on the utilization of the data mining results ought to be come to.here checking accuracy and precision of data.If the result is not satisfy directly jump to business understanding.On the off chance that attack or abnormal behaviours are detected Deployment:This is the final stage for predict APT attack

CONCLUSIONS

Analysis may not be a definitive arrangement that takes care of allsecurity issues. Be that as it may, it offers an answer for some issues that have tormented the secu-rity industry.Applying data analysis to cyber security is a youthful and quickly developing territory. In that capac-ity, it brings its own arrangement of risk and difficulties. In this paper studied on cyber security technologies such as authentication,encryption,anti-virus,WAF,IDS and firewall then brief study on APT attack and groups.By utilizing pro-posed outline work,"CRISP-DM for bigdata analysis along with ARIMA statistical model for predict APT attacks. In the futurework,planning to research on emperical study on different type of vulnerabilities or unknown attacks with other statistical models and machine learning models

REFERENCES

[1] Tai-Myoung Chung Sung-Hwan Ahn, Nam-Uk Kim. "'Big data anal-ysis system concept for detecting unknown attacks"'. February 2014.
[2] http://www.crossdomainsolutions.com/cyber-security/tools-techniques/
[3] https://www.maxcdn.com/one/visual-glossary/web-application-firewall/
[4] https://personalfirewall.comodo.com/what-is-firewall.html
[5] https://techterms.com/definition/ids
[6] https://www.fireeye.com/current-threats/apt-groups.html
[7] Pete Chapman (NCR),Randy Kerber (NCR),Julian Clinton (SPSS),Thomas Khabaza (SPSS),Thomas Reinartz (Daimler-Chrysler),Rdiger Wirth (DaimlerChrysler),"The CRISP-DM Process Model".CRISP-DM Discussion Paper March, 1999
[8] https://www.otexts.org/fpp/8/7

