IOP Conf. Series: Materials Science and Engineering **396** (2018) 012029 doi:10.1088/1757-899X/396/1/012029

A Review on Cybersecurity Threats and Statistical Models

Feba Babu and Kishore Sebastian

1) P,G Scholar, St Joseph College of Engineering and Technology, India

E-mail: febabinu04@gmail.com

Kishore Sebastian

2)Asst.Proessor, St.Joseph College of Engineering and Technlogy,India

E-mail: kishorekinattukara@gmail.com

Abstract. Data generated in past two years is more than previous history tutorials with the help of evolution of technology.Big data is a term using for abundant structured and unstructured data,it becomes difficult to process,store,analyze and visualize using on hand database system tools.This big data systems are facing attack targets.Some statistical models are helps to find out problems of the cybersecurity.In this paper studied on Big data,cybersecurity,categories of cybersecurity threats,main challenges of cybersecurity,and study on statistical models for vulnerability prediction.

1. Introduction

Big data systems are very essential part of this modern organisations,because now a days we are living in digital world so every 60 seconds millions of data is getting generate.This abundant data just beyond the technology's.As little as 5 years ago almost people only thinking how to store tens to hundred of gigabytes data in our personal computer? but today people thinking how to store tens to hundred of terabytes data?

IBM survey provided that every day 2.5 quintillion bytes of knowledge square measure created most that 90%of the info within the world nowadays has been created within the last 2 years[1]. Intels Infographic reveals each sixty seconds, 639,800GB of worldwide information is transferred,One minute of net time, 204 million e-mails sent. on-line denizens read twenty million photos on Flickr. Twitter processes 100,000 new tweets and 320 new Twitter accounts are created[2].

At the same time of data collection need to improve cyber security.This paper mainly focus on cyber security threats and statistical models.The survey Paper starts with Big data,Cyber security,Threats,Challenges and Statistical models for vulnerability prediction.

2. CONCEPT OF BIGDATA

Bigdata is collection of structured,semi structured and unstructured data.It's so large and complex to manage with traditional systems.

Bigdata have three diamensions:Data Volume,Data Velocity and Data Variety. Volume is that the quantity of information in terabytes and petabytes, variety is that kinds of information that's audio, video,image,doccuments etc. and velocity is that the speed of information process therefore time is extremely vital in here.

3. CYBERSECURITY

Investopedia defines[3],Cybersecurity refers to the measures taken to keep electronic information private and safe from damage or theft. It is conjointly accustomed make certain these devices and information aren't abused. Cybersecurity applies to both software and hardware, as well as information on the Internet, and can be used to protect everything from personal information to complex government systems.The major cyber-security issues for the most part focused on software vulnerabilities, organisations must maintain effective vulnerability management programs including remediation, identification,reporting and assessment.

Cyber security is that the main a part of technologies practices and processes designed to to shield networks ,computers ,programs from attack ,damage or unauthorized access .In computing context ,The word security implies cyber security.All the sectors like financial institutions,business,Governments,corporations etc collect ,process and store confidential information on computers and transmit that data across networks or other computers .So growing volume of cyber attacks.

The main elements of cybersecurity:Application security, Information security, Network security.

4. CYBERSECURITY CHALLENGES

To protect information and infrastructure in cyberspace, decrease vulnerabilities,and minimize injury from cyber incidents through a mixture of institutional structures,build capabilities to stop and reply to cyber threats.Some challenges are:

- a commentary from CSI [4],Create mechanisms for cover of non-public knowledge in third party domain specifically,cloud suppliers, social networks, outsources throughout varied phases of its life cycle i.e., transmission, process or storage.
- Create mechanisms which ensure trust in dynamic environment where identities are protected, anchors of trust exist and those interacting are trustworthy. This is in a transparent domain.
- Create new flexible access control technologies which are ethical, less dependent on dynamic identities, using more reliable way of management in a distributed world.
- cyber security is that the quickly and perpetually evolving nature of security risks.
- Budget is very difficult for security professionals to acquire the budget needed for a proper cyber security program.
- Proactive approach checking available information and apply predictive and behavioral analytic tools to find out threat, detect the actual threat, gather intelligence regarding the attack, and execute an enterprise wide response before the threat becomes significant.

5. CYBERSECURITY THREATS

According to[5],many types of threats are on the market and may be categories:

5.1. Trojan horse

This computer virus threats will do something,record your passwords by work keystrokes and record each move and hijacking digital camera to look at

5.2. Malicious spyware
Recorded information is sporadically sent back to the originating cybercriminal over the net.

5.3. Computer worm
Computer worm it's a code program which will copy itself and sent from one pc to a different, with none human interaction. It will replicate with nice volume and with nice speed.

5.4. Botnet
Botnet it's a bunch of computers or systems connected to the net and it compromised by a hacker employing a computer program for obtaining informations from computers.

5.5. Spam
Spam Spam is nothing however unwanted messages from your email box.This Unwanted unsolicited mail contain links that once clicked on may visit a web site that installs malicious code onto pc.

5.6. Phishing
It may be capture all variety of sensitive data like mastercard details,usernames and passwords by impersonating as a trustworthy entity.

5.7. SQL Injection Attack
SQL is structured programming language it can communicate with databases,this injection attacks mainly targets on customers private information from the web site, like mastercard numbers, usernames and passwords.

5.8. Cross-Site Scripting (XSS)
XSS is incredibly the same as SQL Injection Attack,and it's additionally injecting malicious code into a web site.

5.9. Buffer Overflows
This threat mainly doing attacker to control its internal variables.Modern programming languages like C, Java and Perl these languages will be reduce the chances of coding errors and creating buffer overflow vulnerabilities.

Other cybersecurity vulnerabilities are Memory corruption, Gain privileges,Exploits etc.

6. STATISTICAL MODELS FOR VULNERABILITY PREDICTION

The main ojective of this section is to investigate vulnerability prediction with the help of statistical models.Here going to describe 3 models:Copula,GARCH,ARIMA.

6.1. COPULA
Copula is very effective approach for vulnerability prediction,it is a joint distribution perform of random vectors with standard uniform marginal distributions.

It have a vital role in trendy statistics and estimator risk analysis to decompose n dimensional distribution function F into two elements specifically the marginal distributionF_i and also the copula C, capturing the dependence relationship of the marginals.

The copula is formally defined as:

Let X = $(X_1, ..., X_n)$ be a random vector with joint distribution function F and marginal distribution functions $F_i, X_i \sim$ $F_i, 1 \leq i \leq n$.

A distribution function C with uniform marginals on [0, 1]] is named a copula of X if,

$$\mathrm{F} = \mathrm{C}(F_1, ..., F_n)$$

Then Copula for continuous marginals.Here consider C is the joint distribution function of $(F_1(X_1), ..., F_n(X_n))$ and $\mathrm{F}_i(X_i) \sim$ U(0; 1)

Copula C can be

$$\mathrm{C}(u_1, ..., u_n) = \mathrm{F}_X(F_1^{-}1(u_1), ..., F_n^{-}1(u_n)))$$

In Copula model,after determine the proper marginal forms then modeling their joint distribution,in this section considering bi-variate case on overflow vulnerability,then find out observed bi-variate dependence structure using a medium and positive relationship between these two disclosures,through estimated marginals.Then simulate bi-variate dependence structure these two disclosures.According to [7] simulate 200, 000 pairs of overflow disclousres.So this study surely said copula model effectively preserve their long-term and persistent dependece relationship.

6.2. ARIMA

ARIMA stands Autoregressive Integrated Moving Average,this approach provide to time series forecasting.ARIMA models aims to describe the autocorrelations in the data.

Procedure for ARIMA Model[6] :

- plot the data,identify all the patterns usal and unusal.
- If required, transform the information to stabilize the variance.
- If remodeled information are non-stationary: then calculate the variations of the information till the information are stationary.
- Calculate the ACF/PACF.
- Try the AICc to find out a better model..
- Check the residuals,if it fail try a modified model.
- calculate forecasts.

In ARIMA Model(Fig:1),after the plot standardised residuals to visually examine if residuals forms a accurate approximation to a white noise process or not.But ARIMA model[7] provide better fit for buffer overflow disclosure data.ARIMA models can be capture the mean behaviour of each time series based on the in-sample data.ARIMA model is also better for predict Buffer Overflow.

6.3. GARCH

The GARCH (generalized autoregressive conditional heteroskedasticity)This method is usually most popular by financial modeling professionals as a result of it provides a more real-world context for predict the costs and rates of financial instruments.It is the extension of the ARCH method to the GARCH method.GARCH technique is extremely helpful because of it can be use in two things :predict and simulate with GARCH model.

Procedure for GARCH Model:

- Choose best fitting AR model.
- calculate and plot the autocorrelations.
- calculate forecast.

In GARCH model,it provide suitable model for capturing conditional variances in all the empirical data except for the buffer overflow vulnerability,model provides a better fit.

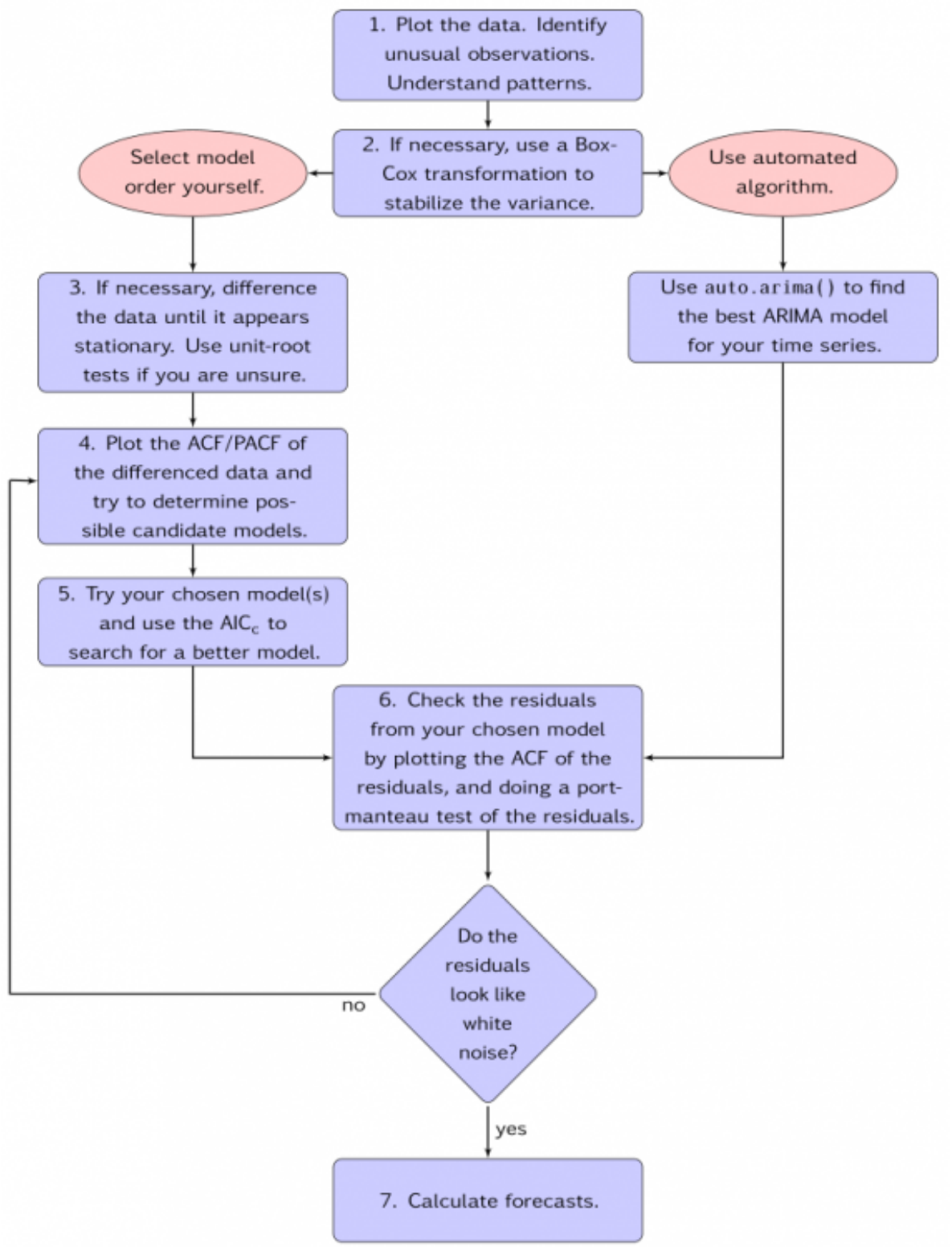

Fig:1-process summary-ARIMA Model

COPULA,ARIMA,GARCH these models will be use for predict on top of cybersecurity threats victimization these procedure.

7. CONCLUSION

The Bigdata diamensions are volume, velocity and variety can be used to solve cybersecurity issues efficiently.This paper studied on Bigdata,cybersecurity,challenges of cybersecurity,threats and statistical models for vulnerability predication.A testing genuine digital security issue in the space of abusing helplessness exposure incline with complex multivariate time arrangement information.In this article,three part analysis technique to detect buffer overflow vulnerability and thorough measurable system towards extending this study about the exposure progression and their interesting reliance structures.COPULA,ARIMA,GARCH viably contemplated on multivariate time arrangement information time-invariant reliance connections in the information. This article demonstrated that a copula based displaying approach can successfully recognize and save idle reliance structure in multivariate time arrangement information.The fitted copula display effectively caught the coveted inert structure.

In the future,a superior comprehension about how unique powerlessness furthermore, abuse exposures communicate with each other through rich recorded information in reality digital security situation.

8. REFERENCES

(i) Sachchidanand Singh and Nirmala Singh Big Data Analytics,2012
(ii) https://www.cnet.com/news/intel-reveals-what-happens-in-a-single-internet-minute/
(iii) https://www.investopedia.com/terms/c/cybersecurity.asp
(iv) N.J.Rao, Cybersecurity: Issues and Challenges,CSI Communications, May 2015.
(v) https://blogs.cisco.com/smallbusiness/the-10-most-common-security-threats-explained.
(vi) https://www.otexts.org/fpp/8/7
(vii) MingJian Tang, Mamoun Alazab,and Yuxiu Luo Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies,2016