ZKG INTERNATIONAL

ISSN: 2366-1313

A VIEW OF ANALYSIS MODELLING AND REDICTING CYBER SECURITY BREACHES

Y.SRI VARSHA[1], SK. FAIZUDDIN[2], SK. NIYAZ AHMED[3], SK. SHAHUL SHARIFF[4], M.YASWANTH[5], U. HEMALATHA[6].

[1]Assistant Professor, CSE,Chalapathi Institute of Technology,Guntur, India
[2]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[3]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[4]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[5]UG Student, CSE, Chalapathi Institute of Technology, Guntur, India
[6]UG Student, CSE, Chalapathi Institute of Technology, Guntur, India

ABSTRACT: Analyzing cyber incident data sets is an important method for deepening our understanding of the evolution of the threat situation. This is a relatively new research topic, and many studies remain to be done. In this project, we report a statistical analysis of a breach incident data set corresponding to 12 years (2005–2017) of cyber hacking activities that include malware attacks. We show that, in contrast to the findings reported in the literature, both hacking breach incident inter-arrival times and breach sizes should be modeled by stochastic processes, rather Than by distributions because they exhibit auto correlations. Then, we propose particular stochastic process models to, respectively, fit the inter-arrival times and the breach sizes.We also show that these models can predict the inter-arrival times and the breach sizes. In order to get deeper insights into the evolution of hacking breach incidents, we conduct both qualitative and quantitative trend analyses on the data set. We draw a set of cyber security insights, including that the threat of cyber hacks is indeed getting worse in terms of their frequency, but not in terms of the magnitude of their damage. We show that, in contrast to the findings reported in the literature, both hacking breach incident inter-arrival times and breach sizes should be modeled by stochastic processes, rather than by distributions because they exhibit autocorrelations. Then, we propose particular stochastic process models to, respectively, fit the inter-arrival times and the breach sizes. We also show that these models can predict the inter- arrival times and the breach sizes

1. INTRODUCTION

Cyber hacking is an effort to take advantage of a computing system or a personal network inside a computer. It is the unauthorized access to regulate over network security system for a few illicit purposes. The data breaches are sensitive, confidential or otherwise protected data has been accessed in an unauthorized fashion. Cyber attack is an assault launched by cybercriminals using one or multiple computers or networks. A data breach is a confirmed incident in which sensitive, confidential protected data has been accessed or disclosed in an unauthorized fashion. Data Breaches may involve personal health information, trade secrets. Breach of privacy laws can expose individuals to risks such as embarrassment, loss of employment opportunity, loss of business opportunity, physical risks to safety and identity theft .A data breach occurs when a cybercriminal successfully infiltrates a data source and extracts sensitive information. This can be done physically by accessing a computer or network to steal local files or by bypassing network security remotely. Data breaches are becoming more and more common and some of the most recent data breaches have been the largest on record to date. DATA breaches are one of the most devastating cyber incidents.

The Identity Theft Resource Center and

Cyber Scout reports 1,093 data breach incidents in2016, which is 40% higher than the 780 data breach incidents in2015.Databreaches expose 4.1 billion records in first six month of 2019.the first six month of2019 have seen more than 3800 publicly disclosed breaches exposing an incredible 4.1billion compromised records.

2. LITERATURE SURVEY

BREACH SIZE CAN BE MODELED BY LOG-NORMALDISTRIBUTION AND BREACH FREQUENCY CAN BE MODELED BYNEGATIVEBINOMIAL DISTRIBUTION

Analyzed a different breach dataset of 2,253 breach incidents that span over adecade (2005 to 2015). These breach incidents include two categories: negligent breaches (i.e., incidents caused by lost, discarded, stolen devices, or other reasons) and malicious breaching (i.e., incidents caused by hacking, insider and other reasons. They showed that the breach size can be modeled by the log-normal or log-skew normal distribution and the breach frequency can be modeled by the negative binomial distribution, implying that neither the breach size nor the breach frequency has increased over the years.

USED EXTREME VALUE THEORY TO STUDY THE MAXIMUMBREACH SIZE AND MODELED BY A DOUBLY TRUNCATEDPARETO DISTRIBUTION

Analyzed an organizational breach incidents dataset that is combined from and spans over a decade (year 2000 to 2015). They used the Extreme Value Theory to study5the maximum breach size, and further modeled the large breach sizes by a doubly truncated Pareto distribution. They also used linear regression to study the frequency of the data breaches, and found that the frequency of large breaching incidents is independent of time for the United States organizations, but shows an increasing trend for non-US organizations.

There are also studies on the dependence among cyber risks. Bohme and Kataria studied the dependence between cyber ¨ risks of two levels: within a company(internal dependence) and across companies (global dependence). Hearth and Hearth used the Archimedean copula to model cyber risks caused by virus incidents.

3 EXISTING SYSTEM

The present study is motivated by several questions that have not been investigated until now, such as: Are data breaches caused by cyber-attacks increasing, decreasing, 41 or stabilizing? A principled answer to this question will give us a clear In sight into the overall situation of cyber threats. This question was not answered by previous studies. Specifically, the dataset analyzed in only covered the time span from2000 to 2008 and does not necessarily contain the breach incidents that are caused by cyber-attacks; the dataset analyzed in is more recent, but contains two kinds of incidents: negligent breaches (i.e., incidents caused by lost, discarded, stolen devices and other reasons) and malicious breaching. Since negligent breaches represent more human errors than cyber-attacks, we do not consider them in the present study.

The malicious breaches studied in contain four sub-categories: hacking (including malware), insider, payment card fraud, and unknown, this study will focus on the hacking sub-category (called hacking breach dataset thereafter), while noting that the other three sub-categories are interesting on their own and should be analyzed separately. Recently, researchers started modeling data breach incidents. They found that neither the size nor the frequency of data breaches has increased over the years. Wheatley et al., analyzed a dataset that is combined from corresponds to organizational breach incidents between year 2000 and 2015. They found that the frequency of large breach incidents

ISSN: 2366-1313

occurring to US firms is independent of time, but the frequency of large breach incidents occurring to non-US firms exhibits an increasing trend.

4. PROPOSED SYSTEM

Here we make the following three **contributions: -**

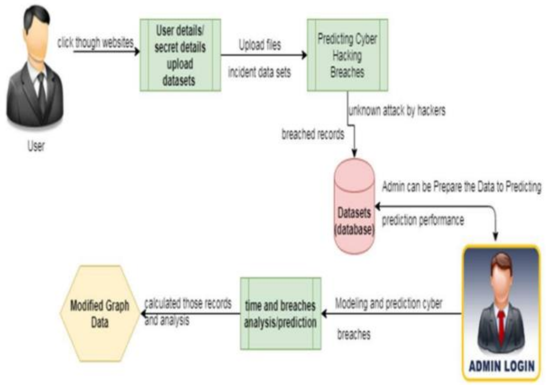
First, we show that both the hacking breach incident inter arrival times(reflecting incident frequency) and breach sizes should be modeled by stochastic processes, instead of distributions. Because they exhibit auto-correlation. We can describe the evolution of the hacking breach incidents inter-arrival times and that a particular ARMA-GARCH model can adequately describe the evolution of the hacking.breach sizes. Regressive Conditional Hetero skew activity. “We show that the process models can predict the inter-arrival times and the breach sizes. Here we are using those stochastic processes, rather than distributions.

Second, we discover a positive dependence between the incidents inter-arrival times and therefore the breach sizes.

Third, we conduct both qualitative and quantitative trend analyses of the cyber hacking breach incidents. We find that the situation is indeed getting worse in terms of the incidents inter arrival time because hacking breach incidents become more and more frequent, but the situation is stabilizing in terms of the incident breach size, indicating that the damage of individual hacking breach incidents will not get much worse.

The stochastic process model rather than distribution. it will help for the reducing inter- arrival time and breach sizes. We also show that when predicting inter-arrival times and breach sizes, it is necessary to consider the dependence; otherwise, the prediction is not accurate. The third we conduct both qualitative and quantitative breach analysis of cyber hacking breach incidents. Here we use a SUPPORT VECTORMACHINE algorithm to solve the problems. “Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It is mostly used in classification.

5. SYSTEM ARCHITECTURE


User
click though websites
User details/ secret details upload datasets
Upload files
incident data sets
Predicting Cyber Hacking Breaches
unknown attack by hackers
breached records
Datasets (database)
Admin can be Prepare the Data to Predicting
prediction performance
ADMIN LOGIN
Modeling and prediction cyber
breaches
time and breaches analysis/prediction
calculated those records and analysis
Modified Graph Data

The First step in the Architecture diagram is User or Admin Login. After logging in we upload the data format to database which are predicted to be cyber hacking and malware websites which is the second step. After that prediction and modeling is done with the help of ARMA GARCH time series model. We classify the breaches and Breach Analysis is carried out. At last Qualitative and Quantitative of Incident breaches is represented in Graphical Structure.

6. IMPLEENTATION

UPLOAD DATA

The data resource to database can be uploaded by both administrator and authorized user. The data can be uploaded with key in order to maintain the secrecy of the data that is not released without knowledge of user. The users are authorized based on their details that are shared to admin and admin can authorize each user. Only Authorized users are allowed to access the system and upload or request for files.

ACCESS DETAILS

The access of data from the database can be given by administrators. Uploaded data are managed by admin and admin is the only person to provide the rights to process the accessing details and approve or unapproved users based on their details.

USER PERMISSIONS

The data from any resources are allowed to access the data with only permission from administrator. Prior to access data, users are allowed by admin to share their data and verify the details which are provided by user. If user is accessing the data with wrong attempts, then, users are blocked accordingly. If user is requested to unblock them, based on the requests and previous activities admin is unblock users.

DATA ANALYSIS

Data analyses are done with the help of graph. The collected data are applied to graph in order to get the best analysis and prediction of dataset and given data policies. The dataset can be analyzed through this pictorial representation in order to better understand of the data details.

7.CONCLUSION&FUTURE SCOPE

We analyzed a hacking breach dataset from the points of view of the incident sinter- arrival time and the breach size, and showed that they both should be modeled by stochastic processes rather than distributions. The statistical models developed in this project show satisfactory fitting and prediction accuracies. In particular, we propose using a copula-based approach to predict the joint probability that an incident with a certain magnitude of breach size will occur during a future period of time. Statistical tests show that the methodologies proposed in this project are better than those which are presented in the literature, because the latter ignored both the temporal correlations and the dependence between the incidents inter-arrival times and the breach sizes. We conducted qualitative and quantitative analyses to draw further in sights. We drew a set of cyber security insights, including that the threat of cyber hacking breach incidents is indeed getting worse in terms of their frequency, but not the magnitude of their damage. The methodology presented in this project can be adopted or adapted to analyze datasets of a similar nature.

There are many open problems that are left for future research. For example, itis both interesting and challenging to investigate how to predict the extremely large values and how to deal with missing data (i.e., breach incidents that are not reported).It is also worthwhile to estimate the exact occurring times of breach incidents. Finally, more research needs to be conducted towards understanding the predictability of breach incidents (i.e., the upper bound of prediction accuracy).

REFERENCES

[1]. P. R. Clearinghouse. Privacy Rights Clea ringhouse’s Chronology of DataBreaches. Accessed: Nov. 2017.

[2]. ITR Center. Data Breaches Increase 40 Percent in 2016, Finds New Report fromIdentity Theft Resource Center and CyberScout. Accessed: Nov. 2017.

[3] C. R. Center. Cybersecurity Incidents. Accessed: Nov. 2017.[4]. IBM Security.Accessed: Nov. 2017.

[5]. Net Diligence. The 2016 Cyber Claims Study. Accessed: Nov. 2017.

[6]. M. Eling and W. Schnell, “What do we know about cyber risk and cyber riskinsurance?” J. Risk Finance, vol. 17, no. 5, pp. 474– 491, 2016.

[7]. T. Maillart and D. Sornette, “Heavy-tailed distribution of cyber-risks,” Eur. Phys.J. B, vol. 75, no.3, pp. 357–364, 2010.

[8]. R. B. Security. Datalossdb. Accessed: Nov. 2017.