International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

MODELING AND PREDICTING CYBER HACKING BREACHES

[1]ROUTU NAVYA,[2]G.RAMESH
[1]MCA Student,B V Raju College, Bhimavaram,Andhra Pradesh,India
[2]Assistant Professor,Department Of MCA,B V Raju College,Bhimavaram,Andhra Pradesh,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 paper, 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 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. 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.

Key words: cyber security, cyber hacks, attacks, predict.

I. 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 purpose. 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 secretes. 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 Privacy

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Rights Clearinghouse reports 7,730 data breaches between 2005 and 2017, accounting for 9,919,228,821 breached records. The Identity Theft Resource Center and Cyber Scout reports 1,093 data breach incidents in 2016, which is 40% higher than the 780 data breach incidents in 2015.Data breaches expose 4.1 billion records in first six month of 2019.the first six month of 2019 have seen more than 3800 publicly disclosed breaches exposing an incredible 4.1 billion compromised records. In 2019,the number of data breaches in the united states amounted to 1,473 with over 164.68 million sensitive records exposed. data breaches have gained attention with the increasing use of digital files and companies and users large reliance on digital data. State of breach January 2020:at least 7.9 billion records, including credit card numbers, home addresses, phone numbers and other highly sensitive information, have been exposed through data breaches since 2019.

2. Existing method

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, or stabilizing? A principled answer to this question will give us a clear insight 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 from 2000 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. Because the malicious breaches studied in [9] 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. Maillart and Sornette studied the statistical properties of the personal identity losses in the United States between year 2000 and 2008. They found that the number of breach incidents dramatically increases from 2000 to July 2006 but remains stable thereafter. Edwards et al. analyzed a dataset containing 2,253 breach incidents that span over a decade (2005 to 2015). 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 (i.e., the ones that breach more than 50,000 records) occurring to US firms is independent of time, but the frequency of large breach incidents occurring to non-US firms exhibits an increasing trend.

3. PROPOSED SYSTEM

In this paper, 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, rather than by distributions. We find that a particular point process can adequately 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, where ARMA is acronym for "Auto Regressive and Moving Average" and GARCH is acronym for "Generalized Auto Regressive Conditional Hetero skedasticity." We show that these stochastic process models can predict the inter-arrival times and the breach sizes.

To the best of our knowledge, this is the first paper showing that stochastic processes, rather than distributions, should be used to model these cyber threat factors. Second, we discover a positive dependence between the incidents inter-arrival times and the breach sizes, and show that this dependence can be adequately described by a particular copula. We also show that when predicting inter-arrival times and breach sizes, it is necessary to consider the dependence; otherwise, the prediction results are not accurate. To the best of our knowledge, this is the first work showing the existence of this dependence and the consequence of ignoring it. 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. We hope the present study will inspire more investigations, which can offer deep insights into alternate risk mitigation approaches. Such insights are useful to insurance companies, government agencies, and regulators because they need to deeply understand the nature of data breach risks.

MODULES EXPLANATION:

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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

share their data and verify the details which are provided by user. If user is access 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.

SIMULATION

Fig.3.1. User Registration form.

Fig.3.2. User entering data.

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. This is the first paper showing that 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 are not accurate. the third we conduct both qualitative and quantitative breach analysis of cyber hacking breach incidents. Here we use a SUPPORT VECTOR MACHINE 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.

Fig.3.4. User checking data.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

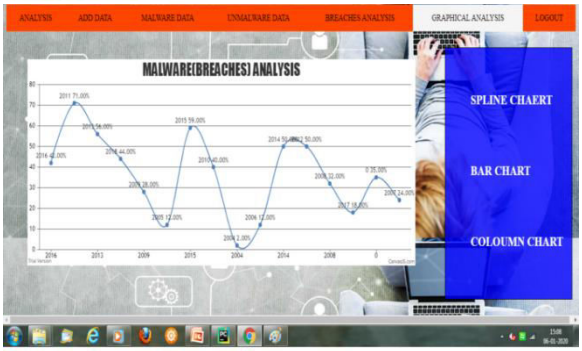

Fig.3.4. Graphical analysis.

Fig.3.5. Column chart

V. CONCLUSION

We analyzed a hacking breach dataset from the points of view of the incidents inter-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 paper 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 paper 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 insights. We drew a set of cybersecurity 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 paper can be adopted or adapted to analyze datasets of a similar nature.

REFERENCES

[1] P. R. Clearinghouse. Privacy Rights Clearinghouse's Chronology of Data Breaches. Accessed: Nov. 2017. [Online]. Available: https://www.privacyrights.org/data-breaches.

[2] ITR Center. Data Breaches Increase 40 Percent in 2016, Finds New Report From Identity Theft Resource Center and CyberScout. Accessed: Nov. 2017. [Online]. Available: http://www.idtheftcenter.org/2016databreaches.html

[3] C. R. Center. Cybersecurity Incidents. Accessed: Nov. 2017. [Online]. Available: https://www.opm.gov/cybersecurity/cybersecurity-incidents.

[4] IBM Security. Accessed: Nov. 2017. [Online]. Available: https://www.ibm.com/security/data-breach/index.html.

[5] NetDiligence. The 2016 Cyber Claims Study. Accessed: Nov. 2017. [Online]. Available: https://netdiligence.com/wp-

content/uploads/2016/ 10/P02_NetDiligence-2016-Cyber-Claims-Study-ONLINE.pdf.

[6] M. Eling and W. Schnell, "What do we know about cyber risk and cyber risk insurance?" 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. [Online]. Available: https://blog.datalossdb.org.

[9] B. Edwards, S. Hofmeyr, and S. Forrest, "Hype and heavy tails: A closer look at data breaches," J. Cybersecur., vol. 2, no. 1, pp. 3–14, 2016.

[10] S. Wheatley, T. Maillart, and D. Sornette, "The extreme risk of personal data breaches and the erosion of privacy," Eur. Phys. J. B, vol. 89, no. 1, p. 7, 2016.

[11] P. Embrechts, C. Klüppelberg, and T. Mikosch, Modelling Extremal Events: For Insurance and Finance, vol. 33. Berlin, Germany: Springer-Verlag, 2013.

[12] R. Böhme and G. Kataria, "Models and measures for correlation in cyber-insurance," in Proc. Workshop Econ. Inf. Secur. (WEIS), 2006, pp. 1–26.

[13] H. Herath and T. Herath, "Copula-based actuarial model for pricing cyber-insurance policies," Insurance Markets Companies: Anal. Actuarial Comput., vol. 2, no. 1, pp. 7–20, 2011.