YIDDISHVol 9 Issue 1OCT 2018
ISSN NO: 0364-4308

Cyber Attack Prediction and Modeling in Hacker Tools

SR UDDANDUI SAHEB,K.VENKATESHWARA

[1]Asst.Professor [MCA], Department of AI&IT, DVR &DR.HS MIC College of Technology ,Kanchikacherla, Krishna(DT).

ABSTRACT: The growth of the threat landscape may be better understood by analyzing data sets of cyber incidents. Since this is a new area of study, there is a great need for further research. In this work, we provide the results of a statistical analysis of a data set of breach incidents covering a 12-year period (2005-2017) of cyber hacking operations, including malware assaults. We demonstrate that, contrary to the conclusions published in the literature, the inter-arrival periods of hacking breach incidents and the extent of breaches should be characterized by stochastic processes rather than distributions. We then suggest specific stochastic process models to account for the periods between arrivals and the magnitude of the gaps. We also demonstrate that these models can accurately forecast how long it will be until a breach occurs. We use qualitative and quantitative trend analytics on the data set to learn more about the history of cyber breach episodes. We conclude a number of things about cyber security, including the fact that the frequency of cyber attacks is on the rise, but the severity of their effects is not.

KEY WORDS: Analysis cyber incidents, stochastic process, prediction of hacking.

INTRODUCTION:

A rising number of people, organizations, and governments are worried about becoming the victim of a cyber hacking incident. The danger of cyber assaults grows as more and more data is kept and communicated online, and the repercussions may be devastating. Data breaches may result in identity theft, financial losses, system failures, and the release of confidential information. Using modeling and predictive analytics may reduce the likelihood of damaging cyberattacks. Models may be developed to estimate the probability of future cyber assaults by evaluating data on past attacks and recognizing patterns and trends. Potential vulnerabilities may be identified and recommendations made using these models as well.

techniques used to lessen the impact of assaults. The ability to anticipate and forecast security breaches is becoming more important as the prevalence of cyber assaults rises. In order to defend themselves against cyber attacks, businesses should use data analytics to acquire a deeper understanding of their vulnerabilities. The full model explains how an IP theft attack graph may look like. In this study, we provide a novel approach to computing the probability of a cyberattack occurring. In this research, we provide a system for dynamically modeling hazards and updating estimates of cyber risk as new information becomes available. Manual risk analysis during system design is the foundation of several existing risk approaches. A few of

the examples of traditional qualitative methods include scenario analysis and questionnaires, which are heavily dependent on experts' subjective opinions. On the other hand, quantitative risk methods are usually based on unreliable data, and therefore their precision is prone to errors [7]. As a result, there is a lack of current research on dynamic cyber risk estimation. Of the work that has been proposed for dynamic risk modeling 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

1. LITERATURE SURVEY

The economy, human privacy, and even national security have been threatened by cyber attacks, which have become a drag. It is crucial that we have a solid understanding of cyber attacks from a variety of perspectives in 2017 before we can adequately deal with the issue. This issue can be difficult to model. A study of multivariate cyber security risks is presented in this paper. In our first statistical approach, we use vine copulas to simulate the multivariate dependence observed by real-world cyber attack data in 2018, using the Copula-GARCH model. Our current method of predicting breach size and inter arrivaltime is a stochastic process model.

CASE 1: Author: Yasser Altawil and Nishant Doshi "A Predictive Analytics Framework for Cyber security Risk Management" by Yasser Altawil and Nishant Doshi (2020) - This paper presents apredictive analytics framework for cyber

security risk management that includes a data-driven risk assessment model and a decision support system.

CASE 2: Author: Md Rafiul Islam and Iqbal

H. Sarker "Modeling and Prediction of Cyber security Breaches" by Md Rafiul Islam and Iqbal H. Sarker (2018) - This article discusses the application of machine learning techniques for modeling and predicting cyber security breaches. The authors present a framework for modeling and prediction that includes data preprocessing, feature selection, modeltraining, and evaluation.

2. 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, 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 previously 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 contain four subcategories: hacking (including malware), insider, payment card fraud, and unknown, this study will focus on the hacking subcategory (called hacking breach dataset thereafter), while noting that the other three sub-categories are interesting on their own and should be analyzed separately.

3. PROPOSED SYSTEM

In this paper, we make the following three contributions. First, we show that both the hacking breach incident inters arrival times 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 Heteroskedasticity.”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 deeplyunderstand the nature of data breach risks.

4. ARCHITECTURE:

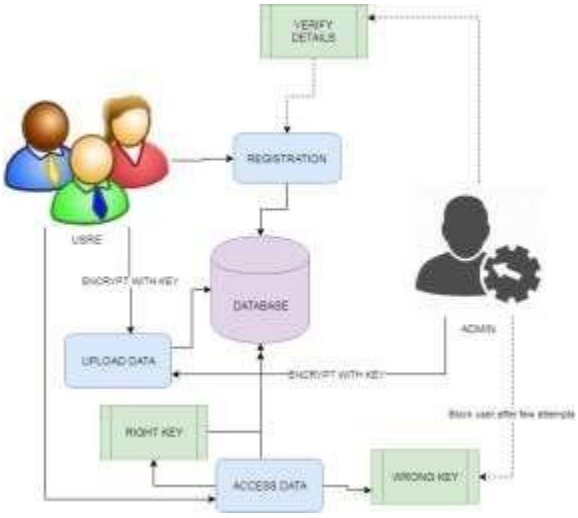

Fig.1. Architecture

5. MODULES

1. 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.

2. ACCESS DETAILS

The access of data from the database can be given by administrators. Uploaded data are managed by admin and admin isthe only person to provide the rights to

process the accessing details and approve or unapproved users based on their details.

3. 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 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.

4. 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 inorder to better understand of the datadetails.

6. ALGORITHM:

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).Support Vectors are simply the co-ordinates of individualobservation. Support Vector Machine is a

frontier which best segregates the two classes (hyper-plane/ line).More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class (so- called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

7. RESULT

8. CONCLUSION

We analyzed a hacking breach dataset from the points of view of the incidents interarrival 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

FUTURE SCOPE

There are many open problems that are leftfor future research It is also worthwhile to estimate the exact occurring times of breachincidents. Finally, more research needs to be conducted towards understanding thepredictability of breach incidents

REFERENCE

1.P. R. Clearinghouse, Privacy RightsClearinghouse's Chronology of DataBreaches, Nov. 2017, [online] Available:https://www.privacyrights.org/data-breaches.

2. *Data Breaches Increase 40 Percent in 2016 Finds New Report From Identity Theft Resource Center and CyberScout*, Nov.2017

3. C. R. Center, Cybersecurity Incidents, Nov. 2017, [online] Available: https://www.opm.gov/cybersecurity/cyberse curity-incidents.

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

5. *The 2016 Cyber Claims Study*, 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. R. B. Security, Datalossdb, Nov. 2017, [online] Available: https://blog.datalossdb.org.

8. 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.

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

10. M. Xu and L. Hua, *Cybersecurity Insurance: Modeling and Pricing*, 2017, [online] Available: https://www.soa.org/research- reports/2017/cybersecurity-insurance.

11. M. Xu, L. Hua and S. Xu, "A vine copula model for predicting the effectiveness of cyber defense early- warning", *Technometrics*, vol. 59, no. 4, pp. 508-520, 2017.

12. C. Peng, M. Xu, S. Xu and T. Hu, "Modeling multivariate cybersecurity risks", *J. Appl. Stat.*, pp. 1-23, 2018.

13. J. Z. Bakdash et al., Malware in the future? forecasting analyst detection of cyber events, 2017, [online] Available: https://arxiv.org/abs/1707.03243.

14. Y. Liu et al., "Cloudy with a chance of breach: Forecasting cyber security incidents", *Proc. 24th USENIX Secur. Symp.*, pp. 1009-1024, 2015.