ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

DEMONSTRATING AND FORECASTING CYBER DISABLE BREACHES

[1] KOTCHARLA SIVA SANKAR BABU ASSISTANT PROFESSOR, [2] BEZAWADA SANTHOSHI KUMARI M.TECH
[1,2] COMPUTER SCIENCE AND ENGINEERING LINGAYAS INSTITUTE OF M ANAGEMENT AND TECHNOLOGY

ABSTRACT:

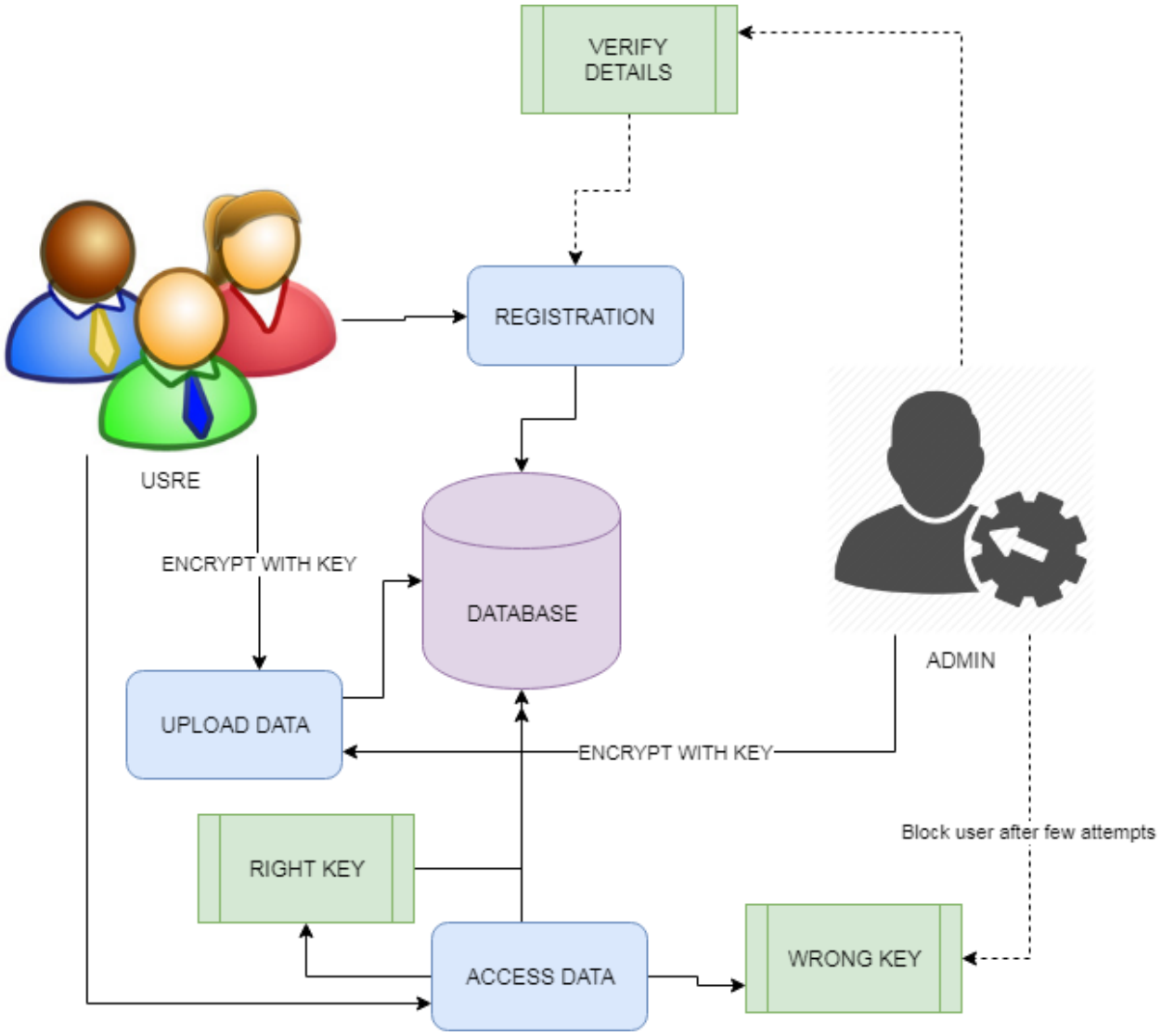
Analyzing cyber incident data sets is an important method for deepening our understanding of the evolution of the threat situation. This is a relatively new research topic, and many studies remain to be done. In this paper, we report a statistical analysis of a breach incident data set corresponding 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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

indeed getting worse in terms of their frequency, but not in terms of the magnitude of their damage.

ARCHITECTURE:



EXISTING SYSTEM:

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

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 [7] 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 [9] 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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

(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.

PROPOSED SYSTEM:

In this paper, we make the following three contributions. First, we show that both the hacking breach incident interarrival 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 "AutoRegressive and Moving Average" and GARCH is acronym for "Generalized AutoRegressive 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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

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.

ALGORITHM:

SUPPORT VECTOR MACHINE

“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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

(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 individual observation. 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.

MODULES:

1. UPLOAD DATA

The data resource to database can be uploaded by both administrator and authorized user. The data can be uploaded with

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

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.

2. 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.

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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

analyzed through this pictorial representation in order to better understand of the data details.

FUTUREWORK:

There are many open problems that are left for future research. For example, it is 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).

REQUIREMENT ANALYSIS

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

REQUIREMENT SPECIFICATION

Functional Requirements

- Graphical User interface with the User.

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

Software Requirements

For developing the application the following are the Software Requirements:

1. Python
2. Django
3. Mysql
4. Wampserver

Operating Systems supported

1. Windows 7
2. Windows XP
3. Windows 8

Technologies and Languages used to Develop

1. Python

Debugger and Emulator

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

- Any Browser (Particularly Chrome)

Hardware Requirements

For developing the application the following are the Hardware Requirements:

- Processor: Pentium IV or higher
- RAM: 256 MB
- Space on Hard Disk: minimum 512MB

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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

[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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

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

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

cyber-insurance policies," Insurance Markets Companies: Anal. Actuar-

ial Comput., vol. 2, no. 1, pp. 7–20, 2011.

[14] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, and

S. K. Sadhukhan, "Cyber-risk decision models: To insure it or not?"

Decision Support Syst., vol. 56, pp. 11–26, Dec. 2013.

[15] M. Xu and L. Hua. (2017). Cybersecurity Insurance: Modeling

and Pricing. [Online]. Available: https://www.soa.org/research-reports/

2017/cybersecurity-insurance

[16] 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.

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

[17] C. Peng, M. Xu, S. Xu, and T. Hu, “Modeling multivariate cybersecurity

risks,” J. Appl. Stat., pp. 1–23, 2018.

[18] M. Eling and N. Loperfido, “Data breaches: Goodness of fit, pricing,

and risk measurement,” Insurance, Math. Econ., vol. 75, pp. 126–136,

Jul. 2017.

[19] K. K. Bagchi and G. Udo, “An analysis of the growth of computer and

Internet security breaches,” Commun. Assoc. Inf. Syst., vol. 12, no. 1,

p. 46, 2003.

[20] E. Condon, A. He, and M. Cukier, “Analysis of computer security

incident data using time series models,” in Proc. 19th Int. Symp. Softw.

Rel. Eng. (ISSRE), Nov. 2008, pp. 77–86.

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

[21] Z. Zhan, M. Xu, and S. Xu, "A characterization of cyber-

security posture from network telescope data," in Proc. 6th

Int. Conf. Trusted Syst., 2014, pp. 105–126. [Online]. Available:

http://www.cs.utsa.edu/~shxu/socs/intrust14.pdf

[22] Z. Zhan, M. Xu, and S. Xu, "Characterizing honeypot-captured cyber

attacks: Statistical framework and case study," IEEE Trans. Inf. Forensics

Security, vol. 8, no. 11, pp. 1775–1789, Nov. 2013.

[23] Z. Zhan, M. Xu, and S. Xu, "Predicting cyber attack rates with

extreme values," IEEE Trans. Inf. Forensics Security, vol. 10, no. 8,

pp. 1666–1677, Aug. 2015.

[24] Y.-Z. Chen, Z.-G. Huang, S. Xu, and Y.-C. Lai, "Spatiotemporal pat-

terns and predictability of cyberattacks," PLoS ONE, vol. 10, no. 5,

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

p. e0124472, 2015.

[25] C. Peng, M. Xu, S. Xu, and T. Hu, “Modeling and predicting extreme

cyber attack rates via marked point processes,” J. Appl. Stat., vol. 44,

no. 14, pp. 2534–2563, 2017.

[26] J. Z. Bakdash et al. (2017). “Malware in the future? fore-

casting analyst detection of cyber events.” [Online]. Available:

https://arxiv.org/abs/1707.03243

[27] Y. Liu et al., “Cloudy with a chance of breach: Forecasting cyber security

incidents,” in Proc. 24th USENIX Secur. Symp., Washington, DC, USA,

2015, pp. 1009–1024.

[28] R. Sen and S. Borle, “Estimating the contextual risk of data breach:

Copyright @ 2020 Authors

ISSN: 2278-4632
Vol-10 Issue-7 No. 14 July 2020

An empirical approach," J. Manage. Inf. Syst., vol. 32, no. 2, pp. 314–341, 2015.

Copyright @ 2020 Authors