PREDICTING CYBER HACKING BREACHES

Dr. B. Prasad, Professor, Department of Information Technology,Vignan's Institute of Information Technology(A), Vidakhapatnam-530049
Ms. M. Srivalli, MCA Student, Department of Master of Computer Applications, Vignan's Institute of Information Technology(A), Visakhapatnam- 530049

ABSTRACT:

Examining datasets of cyber incidents serves as a crucial avenue for enhancing our comprehension of the evolving threat landscape. Despite being a relatively nascent area of research, there is still ample ground to cover. This study delves into a statistical scrutiny of breach incidents spanning a 12-year period (2005–2017), focusing on cyber hacking endeavors involving malware attacks. Contrary to prevailing literature, our analysis suggests that both the intervals between hacking breaches and the scale of breaches are better suited for modeling using stochastic processes rather than traditional distributions, owing to their inherent auto correlation. Consequently, we propose specific stochastic process models tailored to capture these inter-arrival times and breach magnitudes, showcasing their predictive capabilities. To glean deeper insights into the trajectory of hacking breach incidents, we employ a blend of qualitative and quantitative trend analyses on the dataset. Our findings unveil several cybersecurity revelations, notably indicating an escalating frequency in cyber hacks, albeit with no proportional increase in the severity of their repercussions.Keywords: Analysis cyber incidents, stochastic process, prediction of hacking.

Keywords: Analysis cyber incidents, stochastic process, prediction of hacking

INTRODUCTION:

Data breaches stand as one of the most catastrophic cyber incidents, leaving a trail of compromised data and shattered trust. The extent of this digital epidemic is staggering. According to the Privacy Rights Clearinghouse, a total of 7,730 data breaches ravaged security protocols from 2005 to 2017, yielding a grim tally of 9,919,228,821 breached records. The severity of the situation is underscored by reports from the Identity Theft Resource Center and Cyber Scout, indicating a 40% surge in data breaches from 2015 to 2016 alone, with 1,093 incidents documented in the latter year. The impact extends far beyond mere statistics. In 2015, the United States Office of Personnel Management (OPM) became a grim statistic itself, with 4.2 million current and former federal employees falling victim, alongside the theft of background investigation records, including a staggering 21.5 million Social Security Numbers. The financial toll is equally daunting. IBM's findings reveal that the global average cost for each lost or stolen record containing sensitive information reached $158 in 2016. Despite technological advancements aimed at fortifying cyber systems, breaches persist, necessitating a deeper understanding of their evolution. This imperative drives research efforts to unravel the statistical underpinnings of these incidents. Recent endeavors have scrutinized data breach patterns, such as the escalation in breach incidents up to July 2006, followed by a stabilization period. Yet, challenges persist, particularly in devising accurate cyber risk metrics to inform insurance rates. As we delve into characterizing the evolution of data breaches, we not only enhance our comprehension but also illuminate potential avenues for damage mitigation, including the role of insurance. However, the complexities inherent in modeling such incidents underscore the formidable task ahead. Nonetheless, researchers persevere, striving to decode the intricate patterns of data breaches and fortify our digital defenses against this relentless threat.

LITERATURE SURVEY:

Hammouchi et al introduced the STRisk predictive system, which integrates social media dimensions to expand the prediction task scope. They analyzed over 3800 US organizations, creating profiles for each with technical indicators and social factors. To address unreported incidents, they corrected mislabeled organizations in the non-victim sample. Machine learning models were then deployed, achieving over 98% Area Under Curve (AUC) score by leveraging technical and social features. Notably, open ports and expired certificates emerged as top technical predictors, while spreadability and agreeability were highlighted as key social predictors.

Mandal et al. focused on enhancing social sentiment classification by considering various aspects of social events and responses. Their method, utilizing Twitter datasets, outperformed existing methods through aspect-based sentiment analysis, encompassing responses to major social events and alert generation for significant social situations.

Poyraz et al. investigated factors influencing the monetary impact of data breaches on companies. They developed a model for total breach cost based on a dataset categorizing stolen data for US residents. Their rigorous regression analysis revealed significant relationships between breach cost, revenue, stolen data types, and class action lawsuits, with personal information categorization improving cost explanation.

Guru Akhil et al. conducted a comprehensive analysis of hacking breach occurrence datasets spanning 11 years, proposing stochastic cycle models to predict breach occurrence times and sizes. Their findings suggested worsening cybersecurity risks in terms of frequency but not severity.

Fang et al. initiated enterprise-level breach risk modeling and prediction. Their innovative statistical framework leveraged dependencies between multiple time series, effectively modeling and predicting breach incidents despite data sparsity.

Kure et al. aimed for effective cybersecurity risk management (CSRM) using fuzzy set theory for asset criticality assessment, machine learning classifiers for risk prediction, and a comprehensive assessment model (CAM) for evaluating control effectiveness.

CONCLUSION:

In conclusion, our analysis of a hacking breach dataset has shed light on the significance of modeling both the incidents' inter-arrival time and breach size using stochastic processes rather than traditional distributions. The statistical models proposed in this study exhibit commendable fitting and predictive accuracies, marking a significant advancement in the field. Notably, our recommendation of employing a copula-based approach for predicting the joint probability of future incidents with specific breach magnitudes showcases promising results.Statistical tests validate the superiority of the methodologies presented herein compared to existing literature, as they account for temporal correlations and dependencies between inter-arrival times and breach sizes. Through qualitative and quantitative analyses, we have gleaned invaluable cybersecurity insights, highlighting an escalating frequency of cyber hacking breach incidents while noting a relative stability in the magnitude of their damage.The methodologies outlined in this paper offer a robust framework that can be readily adopted or adapted to analyze datasets of similar nature, paving the way for enhanced understanding and proactive mitigation strategies in combating cyber threats.

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/databreach/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-ClaimsStudy-ONLINE.pdf

[6] H. Hammouchi, N. Nejjari, G. Mezzour, M. Ghogho and H. Benbrahim, "STRisk: A Socio Technical Approach to Assess Hacking Breaches Risk," in IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2022.3149208.

[7] Mandal, S., Saha, B., Nag, R. (2020). Exploiting Aspect-Classified Sentiments for CyberCrime Analysis and Hack Prediction. In: Kar, N., Saha, A., Deb, S. (eds) Trends in Computational Intelligence, Security and Internet of Things. ICCISIoT 2020. Communications in Computer and Information Science, vol 1358. Springer, Cham. https://doi.org/10.1007/978-3-030-66763-4_18

[8] Poyraz, O.I., Canan, M., McShane, M. et al. Cyber assets at risk: monetary impact of U.S. personally identifiable information mega data breaches. Geneva Pap Risk Insur Issues Pract 45, 616–638 (2020). https://doi.org/10.1057/s41288-020-00185-4

[9] Guru Akhil, T., Pranay Krishna, Y., Gangireddy, C., Kumar, A.K. (2022). Cyber Hacking Breaches for Demonstrating and Forecasting. In: Kumar, A., Mozar, S. (eds) ICCCE 2021. Lecture Notes in Electrical Engineering, vol 828. Springer, Singapore. https://doi.org/10.1007/978-981-16-7985-8_106

[10] Z. Fang, M. Xu, S. Xu and T. Hu, "A Framework for Predicting Data Breach Risk:Leveraging Dependence to Cope With Sparsity," in IEEE Transactions on InformationForensics and Security, vol. 16, pp. 2186-2201, 2021, doi: 10.1109/TIFS.2021.3051804.

[11] Kure, H.I., Islam, S., Ghazanfar, M. et al. Asset criticality and risk prediction for aneffective cybersecurity risk management of cyber-physical system. Neural Comput & Applic34, 493–514 (2022). https://doi.org/10.1007/s00521-021-06400-0

[12] R. R. Subramanian, R. Avula, P. S. Surya and B. Pranay, "Modeling and PredictingCyber Hacking Breaches," 2021 5th International Conference on Intelligent Computing andControl Systems (ICICCS), 2021, pp. 288-293, doi: 10.1109/ICICCS51141.2021.9432175.