Analysing Cyber Hacking Violations Using RPM for Risk Mitigation

[1]Prof. Vishal R. Shinde, [2]Mr. Vaibhav Keni, [3]Ms. Anushka Sapale, [4]Ms. Khushi Bhavsar
[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.
[1]mailme.vishalshinde@gmail.com,[2]keniv935@gmail.com, [3]anushkasapale29117@gmail.com, [4]khushibhavsar1512@gmail.com

Abstract - In this paper, one key technique for improving our comprehension of the evolution of the threat scenario is the analysis of cyber event data sets. Numerous investigations need to be conducted because this is a relatively new research issue. this research presents a statistical examination of a data collection of breach incidents including cyber hacking activities spanning 12 years (2005–2017), including malware attacks. This paper demonstrates that contrary to the conclusions published in the literature, stochastic processes rather than distributions should be used to represent hacking breach incidence inter-arrival periods and breach sizes because of their autocorrelations. Next, this paper provides specific random process models to fit the breach sizes and the inter-arrival timings, respectively. This paper further demonstrates that the breach sizes and the inter-arrival periods can be predicted by these models. To gain a more comprehensive picture of the development of hacker breach incidents, this paper conducts both qualitative and quantitative trend studies on the data set[1].

Keywords - Cyber hacking, malware attack, hacking breach, data breach, cyber threats, breach prediction, cybersecurity data analytics.

I. INTRODUCTION

Cybersecurity incidents have become an increasing threat for everyone-individuals, businesses, and government. The risk of cyberattacks increases as the quantity of information that is kept and sent via the internet, and its consequences can be serious. Breach scenarios may result in the loss of private and financial information, harm to systems and infrastructure, and disclosure of sensitive data. Using modeling and predictive analytics is one method of reducing the risk of cyberattacks. Models that assist in predicting the possibility of future breaches can be developed by examining data on prior cyberattacks and seeing similarities and trends. These models are additionally useful for pinpointing possible weak points and recommending countermeasures for attackers [1]. Though this narrow perspective may overlook critical insights, it does have the advantage of providing a language-agnostic vulnerability detection approach that abstracts from the complex scripting languages and provides a unified edge to the widest range of online applications conceivable. Past research has revealed that such an analysis is far from simple[9].

II. AIMS AND OBJECTIVE

a) Aim

Developing a Predictive Model for Cyber Hacking Breach Prevention The aim is to empower organizations with proactive and effective cyber hacking breach prediction & modeling, reducing the risk of data.

b) Objective

Data Collection and Analysis Gather comprehensive datasets related to cyber threats, vulnerabilities, and historical breach data. Conduct in-depth analysis to understand patterns, trends, and potential indicators of cyber hacking breaches. Model Development Develop machine learning and AI-driven predictive models using advanced algorithms to forecast potential cyber hacking breaches. Implement a multi-faceted approach to modeling, combining anomaly detection, behavior analysis, and threat intelligence.

III. LITERATURE SURVEY

Paper 1: Chronology of Data Breaches

Data breaches have received a lot of attention as businesses of all kinds rely more and more on electronic data, cloud-based services, and workplace accessibility. Accessing a company's data can now be as easy as breaking into restricted networks since critical corporate data is often held on local workstations, cloud servers, and enterprise databases. Data breaches did not happen when firms began storing their customers' personal information digitally. Businesses did not begin to experience data breaches when they began storing their private information digitally. Data leaks stopped

 DOI : 10.35291/2454-9150.2024.0272

happening when companies started storing their private data digitally. Data breaches have occurred as long as people and businesses have kept records and saved confidential information [4].

Paper 2: Finds New Report from Identity Theft Resource Centre and Cyber Scout

At 9.2 percent of all breaches, breaches involving unintentional disclosure of personal information via email or the Internet were the second most common type of occurrence, with employee mistakes coming in second at 8.7 percent. Except for hacking, every other category showed declines in data from 2015. "Data breaches are personal for companies of all kinds because CEO spear phishing and ransomware attempts have increased significantly. Businesses lose control of their consumer, employee, and business data at the touch of a mouse by a gullible employee [5].

Paper 3: Cybersecurity Incidents.

Cybersecurity incidents are defined as any event, action, or omission that results in or may result in, the following: unauthorized access to any data, information system, or electronic communications network decreased data integrity of any data, information system, or electronic communications network; unauthorized use of any data, information system, or electronic communications network for data processing, including data storage; unauthorized changes to firmware, software, or hardware; unauthorized destruction, damage, deletion, or alteration of data stored in an information format. The Communications Xxx 0000, the Official Regulations 2003 confidentiality standards about cybersecurity and/or privacy about the Services and/or this agreement, or secrets Xxx 0000 to 1989 Contract; Guidance notes: In the future, new law is intended to replace the Privacy and Electronic Communications (EC Directive) Regulations 2003[6].

IV. EXISTING SYSTEM

Cybersecurity modeling & breach prediction is a challenging and developing discipline. While there isn't a single, universally applicable solution, your company can improve its capacity to anticipate and lessen cyberattacks by utilizing several currently available technologies, methods, and strategies. The following elements and methods are frequently seen in these kinds of systems: IDS are necessary for keeping an eye on possible threats and network activity. Though this limited outlook might miss important awareness, it has the key advantage of offering a language-agnostic vulnerability detection approach, which abstracts from the complexity of scripting languages and offers a uniform connection to the vast possible range of web applications[9].

V. COMPARATIVE STUDY

Table No.1 Comparative Analysis

Sr. No.	Author	Project Title	Publication	Technology	Purpose
1.	Maochao Xu, Kristin M. Schweitzer,	Modeling and Predicting Cyber Hacking Breaches	IEEE,2018	Stochastic process using ARMA-GARCH Model	To analyze cyber incident datasets to understand the evolution of cyber threats
2.	C. R. Center	Cybersecurity Incidents	IICSPI, 2017	Legislation with future technology	The purpose of this study was to Reduce the integrity of information systems and prevent unauthorized access
3.	P. R. Clearinghouse	Chronology of Data Breaches	IEEE,2017	HIPPA and PCI	For data breach prevention, protection, and resolution
4.	ITR Center	Finds New Report From Identity Theft Resource Center and Cyber Scout	IEEE,2016	C-suite Strategies	This paper's purpose is to classify datasets and develop a decision support system

VI. PROBLEM STATEMENT

Organizations across many industries are becoming increasingly concerned about the threat posed by cyber hacking breaches in this increasingly digitalized environment. Predictive algorithms that can proactively identify and mitigate these breaches are imperative given the continually changing

world of cyber threats. The necessity of developing such a model is highlighted in this issue statement to the advantage of businesses looking to protect their sensitive data, digital assets, & vital infrastructure [3].

VII. PROPOSED SYSTEM

The research has made a total of three improvements in this study. First, it has demonstrated that the hacking breach incidence arrival timings, which indicate the frequency of incidents, and breaches Rather than using distributions to model sizes, random procedures can be used. According to this research, a particular point in the process can adequately explain the development of hacking breach events' inter-arrival times and a particular ARMA-GARCH model—ARMA stands for "An autoregressive and Moving Average," and GARCH for "Generalized An autoregressive Conditional Heteroskedasticity"—can properly describe the development of breaching breach sizes. This work

 DOI : 10.35291/2454-9150.2024.0272

demonstrates the predictive power of these random process models for both the breach sizes and the inter-arrival intervals.

VIII. ALGORITHM

Step 1: for i = m + 1, · · · , n do
Step 2: Estimate the LACD1 model of the incidents inter-arrival times based on{ds|s = 1, . . ., i − 1}, and predict the conditional mean Ψi = exp (ω + a1 log(i−1) + b1 log(Ψi−1)); **Step 3:** Estimate the ARMA-GARCH of log-transformed size, and predict the next mean μˆi and standard error σˆi;[1]
Step 4: Select a suitable Copula using the bivariate residuals from the previous models based on AIC;
Step 5: involves simulating 10,000 2-dimensional copula samples (u (k) 1, i, u (k) 2, i, k = 1,..., 10000) based on the calculated copula;
Step 6: involves converting the simulated dependent samples u by applying the inverse of the predicted generalized gamma distribution, k = 1,..., 10000, to the (k) 1, i's into the z (k) 1, i's;
Step 7: Using the inverse of the predicted mixed extreme value distribution, k = 1,..., 10000, transform the simulated dependant samples u (k) 2, i into the breach sizes' z (k) 2, i;
Step 8: Using Eqs. (IV.1) and (IV.3), respectively, compute the anticipated 10000 2-dimensional breach data, d (k) i, y (k) i,k = 1,...,10000;
Step 9: Using the simulated breach data, calculate VaRα,d(i) for the incidents' inter-arrival times and VaRα,y(i) for the transfer of the log of breach sizes.
Step 10: if d (k) i > VaRα,d(i) then
Step 11: A transgression of the incidents inter-arrival time occurs;
Step 12: end if
Step 13: if y (k) i > VaRα,y(i); then
Step 14: there is a breach size violation
Step 15: end if
Step 16: end for

IX. MATHEMATICAL MODEL

1. ACD MODEL

This paper study follows the ACD models for model selection because (i) they are quite basic and can be efficiently estimated in reality, and (ii) our preliminary study suggests that these models are flexible enough to account for the evolution of the inter-arrival times.

• The conventional ACD model (ACD) [1] is as follows:
Ψi = ω + Xp j=1 ajdi−j + Xq j=1 bj\i−j,
where ∼, aj, and bj ≥ 0, and p and q are positive integers that represent the orders of the terms.
• The log(Ψi) = ω + Xp j=1 aj log(i−j) +Xq j=1 bj log(Ψi−j), which is the type-I log-ACD model(LACD1).
• The log(Ψi) = ω + Xp j=1 aj log(di−j) +Xq j=1 bj log(Ψi−j) is the type-II log-ACD model (LACD2).

In the next section, we further limit our research to the scenario when p = q = 1, as a higher order does not always increase prediction accuracy. It is assumed that there is a generalized gamma distribution for The allocation of the standardized innovations of the Is. Below, this assumption will be verified. research uses the assumption because it is adaptable and because modeling unevenly spaced data has been suggested using it in the literature.

Remember that the generalized gamma distribution's density function is f(x|λ, γ, k) = γxkγ−1 λkγΓ(k) exp n − x λ γo.

The phrase inter-arrival time, which is frequently used in computer science, is utilized in this study. community and the commonly used word duration in the statistics community, where λ > 0 represents the scale parameter and γ, k > 0 represents the shape parameters, are used interchangeably. Several well-known distributions, including the gamma, Weibull, exponential, and half-normal distributions, are included as special instances in the generalized gamma distribution. then put λ = Γ(k) Γ(k + 1/γ) in our estimation to ensure that E(i) =1.

2. Qualitative Trend Analysis

Qualitative Trend Analysis of the Inter-Arrival Times of Hacking Breach Incidents it demonstrates that the breach incidents' inter-arrival times may be described by the LACD1 model. Formally, the trend is described as:

The random portion is defined as }i, which is characterized by The gamma distribution, generalized and log(}i) = ω + a1 log({i−1) + b1 log({i−1), namely the LACD1 model. which has the estimated standard deviations of 0.2254, 0.0241, 0.0971, 0.1136, and 0.1748, and the estimated parameters

(ω, a1, b1, k,γ) = (3.825, 0.058, −0.767, 0.556,1.254).[1]

3. Quantitative Trend Analysis

To quantify the trend, this paper proposes using two metrics to characterize the growth of hacking breach incidents.

Here, {(ti, ti)}i=1,...,n represents the series of breach events with a breach size of yti that happen at time ti. Motivated by the economic growth Growth Rate is
(GR): GRi = yti+1 − yti /yti
is the breach-size GR, according to our definition. The definition of inter-arrival times (GR)is the Average Growth Rate over Time
(AGRT): GRTi = 1 di+1 yti+1 − yti yti
is how to define the AGRT.
Compound Growth Rate over Time
(CGRT): CGRTi = yti+1 yti {1/di+1 − 1
is how to define the CGRT.

Keep in mind that CGRT indicates the rate at which the breach size would increase, whereas AGRT shows the percentage change in breach size over time.

 DOI : 10.35291/2454-9150.2024.0272

4. AIC & BIC

AIC stands for Akaike Information Criterion and BIC stands for Bayesian Information Criterion, are both statistical measures used for model selection, particularly in the context of regression analysis or other statistical modeling techniques The most widely used criteria in choosing models in statistics are AIC and BIC.

$$AIC = -2\log(MLE) + 2k$$

where by k represents the total amount of estimated parameters and indicates the complexity of the approach.

X. SYSTEM ARCHITECTURE

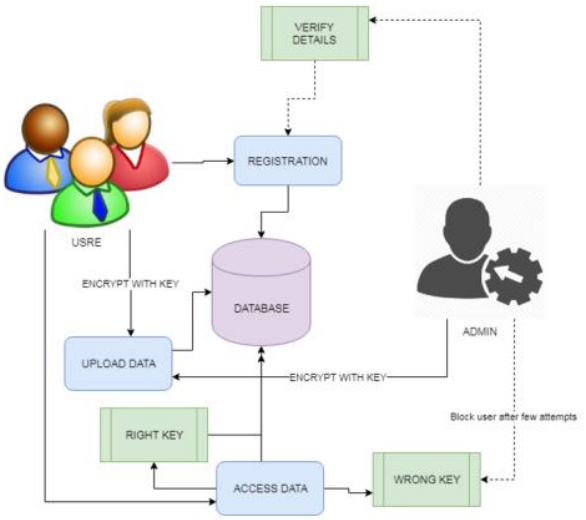

Fig.1: System Architecture

XI. ADVANTAGES

- Proactive Defence: Organizations can strengthen their defenses by using predictive modeling to help foresee cyber-attacks before they materialize
- Resource Efficiency: It optimizes resource allocation by focusing efforts where they're most needed, saving time and money.
- Minimized Damage: Early detection and response reduce the impact of cyber breaches, minimizing downtime, data loss, and reputational damage.
- Strategic Planning: Predictive modelling provides insights into emerging trends and attack vectors, enabling organizations to adapt their cybersecurity strategies proactively to stay ahead of evolving threats.

XII. DESIGN DETAILS

Fig 2: Result

XIII. CONCLUSION

Thus we have tried to implement the paper **Maochao Xu, Kristin M. Schweitzer, Raymond M. Bateman, and Shouhuai Xu, "Modeling and Predicting Cyber Hacking Breaches", IEEE Nov 2018**. and the conclusion as follows: study of a hacker breach dataset revealed that stochastic processes, as opposed to distributions, should be used to characterize both the incident's inter-arrival time and breach magnitude. This work develops statistical models with satisfactory fitting and prediction accuracies. This study proposed 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. The approaches suggested in this study are superior to those found in the literature, according to statistical testing, as the latter neglected both the temporal correlations and the dependence between the incident's inter-arrival times and the breach sizes.

REFERENCES

[1] Maochao Xu, Kristin M. Schweitzer, Raymond M. Bateman, and Shouhuai Xu, "Modeling and Predicting Cyber Hacking Breaches", IEEE Nov 2018.

[2] Kommu Anusha, Mounika Pasam, "Modeling and Predicting Cyber Hacking Breaches using Support Vector Machine Algorithm", Aug 2020

[3] Matthew W. Davis "Cybersecurity Assessment and Mitigation Stochastic Model", March 2018.

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

[5] ITR Center. Data Breaches Increase 40 Percent in 2016, Finds New Report from Identity Theft Resource Center and Cyber Scout. Accessed: Nov. 2017.

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

[7] IBM Security. Accessed: Nov. 2017.[Online].Available:https://www.ibm.com/security/databreach/index.html

[8] R. B. Security. Datalossdb. Accessed:Nov.2017.[Online].Available:https://blog.datalossdb.org

[9] Prof. Kanchan Umavane, Kunal Jain, Prathmesh Vishwakarma, Manoj Verma, "Machine Learning for Web Vulnerability Detection: The Case of Cross-Site Request Forgery",IJREAM,vol-08, issue,01, APR 2022.

 DOI : 10.35291/2454-9150.2024.0272