USING ANALYTICS TO PREDICT CYBER-ATTACKS AND BREACHES

B. Sunil Kumar
1Assistant Professor, Dept. of Master of Computer Applications, Narayana Engineering College, Gudur.

D. Bhargav
2 PG Scholar, Dept. of Master of Computer Applications, Narayana Engineering College, Gudur.

Abstract—Cyber incident data sets are analyzed for the important method for understanding of the threat situation evolution. Both hacking breach incident inter-arrival times and breach sizes should be modeled by stochastic processes[13-15]. Then, we propose some random process models to fit the breach sizes and inter arrival time. Then we conduct both quantitative and qualitative trend analysis of the cyber hacking breach incidents. To reduce this type of situation is indeed getting not as good as in terms of the incidents inter-arrival time because hacking breach incidents become to a greater extent frequently.

Keywords—Data Breach, Hacking Breach, Cyber Threats, Cyber Risk Analysis, Prediction of Breaches, Trend Analysis, Time Series, Data Analytics for Cyber Security.

I. INTRODUCTION

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[1-2]. 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 data set analyzed , but contains two kinds of incidents: negligent breaches (i.e., incidents caused by lost, discarded, stolen devices and other reasons) and malicious breaching[3-4].

Data breaches are one of the most devastating cyber incidents. The Privacy Rights Clearing house reports7,730 data breaches between 2005 and 2017, accounting for9,919,228,821 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.[5-6] Federal government employees and the background investigation records of current, former, and prospective federal employees and contractors were stolen in 2015. The monetary price incurred by data breaches is also substantial. IBM reports that in year 2016, the global average cost for each lost or stolen record containing sensitive or confidential information was $158.

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. They found that the frequency of large breach incidents[7-8].

II. RELATED WORK

“Data breaches caused by cyber-attacks increasing or decreasing or stabilizing?” This principle was taking in consideration. There are two kinds of incidents are there to predict the breaches. They are Negligent Breaches and Malicious Breaches[9-12].

1. Negligent Breaches:

The incidents caused by lost, discarded, stolen devices and other reasons,since negligent breaches represent more human errors than cyber-attacks.
2. The Malicious Breaches:

It contains four sub categories. They are as follows: Un Authorized Access, Physical theft, Card fraud, Cyber-attack.

i. **Un Authorized Access**: Un Authorized Access is when someone gains access to a Website, Program, Server, or other system using some ones else account or other methods. For example, if someone kept guessing a password or user name of an account that was not theirs until they gained access, it is considered as Un Authorized Access.

ii. **Physical Theft**: Theft is defined as the physical removal of an object that is capable of being stolen without the consent of the owner and with the intention of depriving the owner of it permanently. Larceny is the trespassory taking and carrying away of personal goods from the possession of another with the intention to steal.

iii. **Card Fraud**: Credit card fraud is an inclusive term for fraud committed using a payment card, such as a credit card or debit card. The purpose may be to obtain goods or services, or to make payment to another account which is controlled by a criminal.

iv. **Cyber-attacks**: An attempt by hackers to damage or destroy a computer network or system.

---

**III. PROPOSED WORK**

We show that both the hacking breach incidents inter arrival times and breach sizes should be modeled by stochastic processes, rather than by distributions. It is necessary to consider the dependence.

*Figure 1: System Over view*

The above figure 1 represents the over view of themodel that predicts Breach incidents.

The system has following implementation Modules:

- **Admin**: In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as View All Users and Authorize. Admin creates the data base and provide the data breaches as their links happened in the past for the User’s. Admin has
right to block the Un Authorized User, if he attempts the user’s id and password wrong for few attempts.

**User:** In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like Access data from Database and can upload Data for future use.

**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 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.

This Proposed work follows the following Algorithm:

In this Sub section we would follow the Algorithm which represents the Data Breach incident inter arrival times and Breach sizes. Here we use Support Vector Machine 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 differentiates the two classes very well.

a. **Separate Prediction and Results:** We use Algorithm 1 to perform the recursive rolling prediction for the inter-arrival time and the breach sizes.

b. **Joint prediction and Results:** In practice, it is important to know the joint probability that the next breach incident of a particular size happens at a particular time i.e., with a particular inter arrival times.

**IV. EXPERIMENT RESULTS**

**Qualitative Trend Analysis of the Hacking Breach Sizes:** We used the ARMA-GARCH model within no variations that follow the mixed extreme value distribution to describe the log-transformed breach sizes.

The trend is defined as, \( Y_t = \mu + \varphi Y_{t-1} + \theta_1 Y_{t-1} \).
Quantitative Trend Analysis: In order to quantify the trend, we propose using two metrics to characterize the growth of hacking breach incidents.

Growth Rate (GR): We define the breach-size GR as $GR_i = y_{ti+1} / y_{ti}$.

**V. CONCLUSION**

In this Project, 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 project 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.

**REFERENCES**


Author’s Profile:

**B. Sunil Kumar** has received his M.Sc Computer Science from SVU PG Center, Kavali affiliated to Sri Venkateswara University in 2002, M.Tech degree in Information Technology from Sathyabama University, Chennai in 2009 and M.Tech in Computer Science and Engineering from SVCET, Chittoor affiliated to JNTU, Anantapur in 2013. He is pursuing Ph.D in CSE at SVU, Uttar Pradesh. He is dedicated to teaching field from the last 14 years. He has guided 26 P.G and 40 U.G students. His research areas included Data Mining. At present he is working as Assistant Professor in Narayana Engineering College, Gudur, Andhra Pradesh, India.

**D. Bhargav** has received his B.Sc degree in Computers from Sri Venkateswara Degree College, Chilakur affiliated to Vikrama Simhapuri University, Nellore in 2018 and pursuing PG degree in MCA from Narayana Engineering College, Gudur affiliated to JNTU Anantapuram, Andhra Pradesh, India.