MODELING AND PREDICTING CYBER HACKING BREACHES

Dr. Mohammed Abdul Waheed#1, Swati T Naganur#2
#1 Associate Professor
#2 Student

Department of Studies in Computer Science and Engineering, VTU Center for PG Studies, Kalaburagi, Karnataka, India

Abstract:-Modeling and predicting cyber hacking breaches is an important, yet challenging, problems. In this paper, we initiate the study of modeling and predicting cyber hacking breaches. In the present study we proposed a stochastic process model to predict the both hacking breach incident inter arrival times and breach sizes. Here we will use both qualitative and quantitative trend analysis on the data set.

Keywords— Cyber Security, Cyber, Breach, Malware attacks.

1. Introduction

Cyber hacking is an effort to take advantage of a computing system or a personal network inside a computer. It is the unauthorised access to regulate over network security system for a few illicit purpose. The data breaches are sensitive, confidential or otherwise protected data has been accessed in an unauthorized fashion. Cyber attack is an assault launched by cybercriminals using one or multiple computers or networks. A data breach is a confirmed incident in which sensitive, confidential protected data has been accessed or disclosed in an unauthorized fashion. Data breaches may involve personal health information, trade secrets. Breach of privacy laws can expose individuals to risks such as embarrassment, loss of employment opportunity, loss of business opportunity, physical risks to safety and identity theft. A data breach occurs when a cybercriminal successfully infiltrates a data source and extracts sensitive information. This can be done physically by accessing a computer or network to steal local files or by bypassing network security remotely. Data breaches are becoming more and more common and some of the most recent data breaches have been the largest on record to date. Data breaches are one of the most devastating cyber incidents. The Privacy Rights Clearinghouse reports 7,730 data breaches between 2005 and 2017, accounting for 9,919,228,821 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. Data breaches expose 4.1 billion records in first six month of 2019. The first six month of 2019 have seen
more than 3800 publicly disclosed breaches exposing an incredible 4.1 billion compromised records. In 2019, the number of data breaches in the United States amounted to 1,473 with over 164.68 million sensitive records exposed. Data breaches have gained attention with the increasing use of digital files and companies and users large reliance on digital data. State of breach January 2020: at least 7.9 billion records, including credit card numbers, home addresses, phone numbers and other highly sensitive information, have been exposed through data breaches since 2019.

2. Literature Survey

In this paper we predict the cyber hacking breaches in cyber attacks. The data breaches pose an ongoing threat to personal and financial security, and they are costly for the organizations that hold large collections of personal data. To manage IT security residual risks but the accuracy of premiums remains an open question. Thus, practitioners and academics have argued for more robust and innovative cyber-insurance pricing models. The paper fills this important gap in the literature by developing a cyber-insurance model using the emerging copula methodology in 2011[13]. We specialize in analyzing macroscopic properties of the attack traffic flows and identify two main patterns with distinct spatiotemporal characteristics: deterministic and stochastic in 2015[24]. This methodology advocates the utilization of gray-box models, which accommodate the statistical properties/phenomena exhibited by the info. Although our prediction study is predicated on specific cyber attack data, our methodology are often equally applied to research any cyber attack data of its kind. Index Terms—Extreme values, extreme value theory, prediction, gray-box models, time series in 2015 [23]. Data breach incidents are on the increase, and have resulted in severe financial and legal implications for the affected organizations. We apply the opportunity theory of crime, the institutional anomie theory, and institutional theory to identify factors that could increase or decrease the contextual risk of data breach in 2015[28]. Data breaches have exposed the private information of many many people. Some reports point to alarming increases in both the dimensions and frequency of knowledge breaches, spurring institutions round the world to deal with what appears to be a worsening situation in 2016 [9]. Cyber attacks became a drag that's threatening the economy, human privacy, and even national security. Before we can adequately address the problem, we need to have a crystal clear understanding about cyber attacks from various perspectives in 2017[25]. Modeling cybersecurity risks is an important, yet challenging, problem. In this paper, we initiate the study of modeling multivariate cybersecurity risks. We develop the first statistical approach, which is centered at a Copula-GARCH model that uses vine copulas to model the multivariate dependence exhibited by real-world cyber attack data in 2018. Now present we are using stochastic process model to predict the inter-arrival time and breach sizes.
3. Proposed system

Here we make the following three contributions:

First, we show that both the hacking breach incident inter arrival times (reflecting incident frequency) and breach sizes should be modeled by stochastic processes, instead of distributions. Because they exhibit auto-correlation. We can 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. Here we are using 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 therefore the breach sizes.

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.

This is the first paper showing that the stochastic process model rather than distribution it will help for the reducing inter-arrival time and breach sizes. We also show that when predicting inter-arrival times and breach sizes, it is necessary to consider the dependence; otherwise, the prediction are not accurate. The third we conduct both qualitative and quantitative breach analysis of cyber hacking breach incidents. Here we use a SUPPORT VECTOR MACHINE algorithm to solve the problems. “Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It is mostly used in classification.

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

5. **EXPERIMENTAL RESULT**

   ![Fig 1: User Registration](image-url)
Fig 2: Upload data

Fig 3: GARCH MODEL
6. **CONCLUSION**

We analyzed a hacking breach dataset from the incident inter arrival time and the breach size, and showed that they both should be modeled by stochastic process rather than distribution because they exhibit autocorrelations. Here we are using a ARMA-GARCH process for the evaluation of breaching incidents. We used support vector machine algorithm for classification purpose. We conducted both qualitative and quantitative analysis to draw further insights. The earlier the breach is detected, we will either stop any damage to the data or prevent damaging data.

7. **REFERENCES**


