Impacted Cyber Attacks Assessment in Wide Range of Big Data Security Systems

P. Shanmuga Prabha, Ashwini.S, S.Magesh kumar

Abstract: ***The technological advancements in image storage, data processing, and signal analysis of Big Data include (a) the fastly degrade the cost of storage and CPU power in recent arena; the flexibility and cost-effectiveness of data operating platforms and cloud computing systems for flexible computation and storage; and (c) the development of new frameworks , which allow users to take advantage of these divided computing systems storing large amount of data which is almost flexible parallel processing. The proposed survey work focused on discussing the various impacted cyber-attack critics available in industry and the trending algorithms available for cyber security etc. Big data in IoT clouds handling and software platforms which allow the malware enter into the working systems are analyzed, reliable methods to avoid the miscellaneous malwares are clearly depicted here.***

Index Terms: Big data security, Malware detection, Intrusion detection, Cyber security, Optimization algorithms

I. INTRODUCTION

The field of cyber security is very wide and the research focus on cyber security involves in the detection of miscellaneous activities on the internet cloud which is the major problem nowadays. Intrusion detection is most sensitive problem nowadays in which the specific ranges of Anomalies happen in the daily observation. The research work enhance the route of finding the efficient algorithm or approach which relatively detect the miscellaneous malwares which could cause harmful damage to the Big Data storage. The main motive of the study work is to find the efficient model for detecting the malware and progressive predictions and forecasting in cyber security. The attention is keen and breakthrough in the field of cyber security. Accessing the Big Dataset is tougher nowadays and keeping the eye on the security of the informative data also challenging. Attack projection is a method of predicting the next feasible step of the malware software entering into the scope of big data secure wall. Intention recognition is another factor in which the goal of the access creditability is noted. Intrusion detection and prediction is most important likely to be honored to predict the vulnerabilities.

Revised Manuscript Received on October 22, 2019.

P. Shanmuga Prabha, Assistant Professor , Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.

Ashwini.S, Assistant Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.

S.Magesh kumar, Associate professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.

The impacts on data security is statistically analyzed to detect the presence of threats and other network security pitfalls. Various methodologies are used to find out the problems together that can be analyzed in the background study of the survey here. We have mentioned in the survey pointing out the drawbacks and advantages of methods involved in detecting and predicting the malwares.

The evolution of big data has affected many arenas, including malware investigation. Emerging growth in the computational power, computing technology and storage capacity have made it possible for big-data handling systems to manage the constantly increasing volume of data being collected. Not only have the collecting of malware, new ways of analyzing the malwares and visualizing it been developed. In the present survey, all the impacted functions of cyber attack systems and assessments are analyzed in the wide range of big Data security systems.

II. LITERATURE SURVEY

The survey work for the existing analysis started with the first step of collecting various journals, conference proceedings. Many papers and journals critically discussed a lot of information about cyber security are mainly allocated in the field of computer networking and communications of computers. The factor here is not only the cyber security concerned also the impacted factors and assessment of those parameters to handle the Big Data security in a most efficient and clear way.

In the year of 2018, authors named Huaizhi Wang, Member, Jiaqi Ruan, Guibin Wang, Bin Zhou, Yitao Liu, Xueqian Fu, Jianchun Peng, investigated his research work on Deep Learning Based ISE of AC Grids against eligant Cyber Attacks In which he clearly analyze that Due to the aging of electric structures, predictable power grid is being restructured towards canny grid that enables bi-polar communications between consumer and utility, and thus more helpless to cyber-attacks. However, due to the aggressive cost, the cyber-attack approach may vary a lot from one unique scenario to another from the perspective of challenger capability, which is not considered in previous studies. These analyzed parameters are then applied to improve the precision for electric load forecasting, resulting in a more unique width of state variables.

Authors Qian Chen, Sherif Abdelwahed and Abdelkarim Erradi developed his research work entitled “A Model-Based Validated Autonomic Approach to Self-Protect Computing Systems in the year of 2014 and depicted in his findings that his work is highlighted on development of automatic cyber security system and systematic approach on

Internet of Things (IoT) ecosystems. The ultimate focus on self protecting system realization which can able to estimate the problems, measure, adjust the model automatically and react to intrusion and malware attacks at the predetermined stage. Multi criteria, data analysis is discussed and a virtual machine based prototypic framework is developed. Most of the web applications are protected by applying the proposed model.

Authors Martin Husa´k, Jana Koma´rkova in the year of 2018 proposed a detailed survey entitles "Survey of Attack Projection, Prediction, and Forecasting in Cyber Security" in which attack graph, Bayesian network model, markov models, time series fray models are analyzed. Depends on the positively changing environment to protect the data systems from cyber attacks a cutting edge algorithm is developed. This paper focused on discussing about the problematic parameters only.

Authors named Thu Yein Win, Huaglory Tianfield, and Quentin Mair, done a research work entitled "Big Data Analysis-Based Secure Cluster Management for Optimized Control Plane in Software-Defined Networks" in the year of 2017 which analyzes the cloud computing impacts attracted through cyber-attacks in visualized infrastructure. This paper clearly shows the advanced cyber-attack techniques and pitfalls which will affect the big data storage directly. Virtual machines are being accumulated into the HDFS and detailed extraction of features are performed through graph analysis. Neural network based algorithms are evaluated by him to incorporate this paper and logistic regression model is focused here for detecting the commonly occurring malwares.

IoT Big Data Security and Privacy vs. Innovation author Karen R. Sollins, in the year of 2018 evaluated his research work on IoT Big Data impacts. In the paper the conflict of privacy and data security is evaluated. This problem is blown up in the context of the Internet of Things (IoT). From the designer perspective the requirements and constraints are clearly analyzed here. The author proposed a most effective decomposition design space for achieving the flexible model.

Authors Luis M. Vaquero,Antonio Celorio,Felix Cuadrado, Ruben Cuevas, 2014 evaluated the research work entitled, Deploying Large-Scale Data Sets on-Demand in the Cloud: Treats and Tricks on Data Distribution in which Public clouds have democratized the admittance to data analytics for virtually any foundation in the world. Here we present a big data offering service that integrates classified and peer-to-peer data scattering procedures to speed-up data loading into the VMs used for data processing. This forceful topology mechanism is tightly coupled with characteristic machine alignment techniques.

The research work entitled "Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid"Kun Wang, Chenhan Xu, Yan Zhang, Song Guo, Albert Y. Zomaya, IEEE in the year of 2016 establishes the Electricity price forecasting is a important part of smart grid because it makes canny grid cost efficient. Storing the electric data consumption vaues in the cloud also overhead the storage and creates a data handling processing time slower. Existing methods for price anticipating is critical to handle with large price data in the grid.

The proposed method analyzes random forest and Relief algorithm by creating a hybrid model of GCA to eliminate the feature redundancy. Kernel function and PCA is used in feature extraction steps to realize the dimensionality reduction. The Big data forecasting is also evaluated through SVM based prediction model.

On continuing the study another paper named "Energy Big Data Analytics and Security: Challenges and Opportunities" by author Jiankun Hu, Athanasios V. Vasilakos, in the year of 2016 declares that Data handling in smart grid is a difficult process to safeguard the large set of data especially here the energy metering data is focused and renewable energy data deploys large amount of information in the cloud. The taxonomy of usage based regenerating the energy in smart grids is done through analysis of volume, velocity and variety of data.

III. TABULATIONS

SL NO	AUTHORS	YEAR	ALGORITHM USED	RESEARCH FOCUSED
1	named Huaizhi Wang, Member, Jiaqi Ruan, Guibin Wang, Bin Zhou, Yitao Liu, Xueqian Fu, Jianchun Peng	2018	LEARNING BASED ISE	CYBER ATTCKS
2	Qian Chen, Sherif Abdelwahed and Abdelkarim Erradi	2014	SELF PROTECTED COMPUTING	CYBER ATTCKS IN IOT
3	Martin Husa´k, Jana Koma´rkova	2018	NEURAL NETWORKS	ATTACK PROJECTION
4	Thu Yein Win, Huaglory Tianfield, and Quentin Mair	2017	VIRTUAL MACHINES	SDN
5	Luis M. Vaquero,Antonio Celorio,Felix Cuadrado,Ruben Cuevas	2014	MACHINE LEARNING	PTP

IV. TECHNOLOGIES OVERVIEW

Big Data Analytics

The era of Things Net is here renamed for internet of things mean, with lots of connected devices has created an ever greater peripheral for cyber attackers to exploit, which has stemmed in the need for fast and correct detection of those attacks.

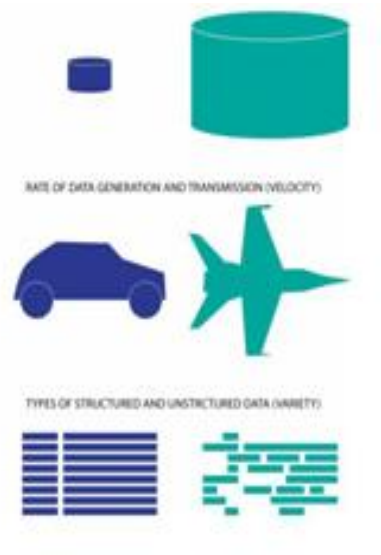

Fig. 1 Traditional and Big Data Comparisons

In recent years, mobile computing, communications, have carry out the big data, contains exceptional grow of appreciated data produced in diverse forms at a tremendous speed. Using big data analytics tools the facility to enhance these huge grow of data in real time brings abundant paybacks that used in cyber hazard search platforms.
Prospective topics related to the study is being shown below

Drivers of Big Data

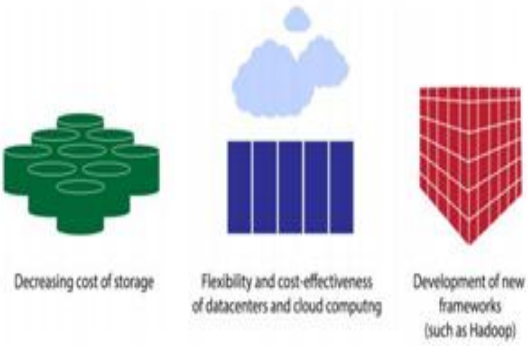

Fig. 2 Technical factors act as drivers for Big Data

1. Storage cost
2. Big Data tools such as the Hadoop network and NoSQL
3. Mine, Transform, and Load (MTL)

Protective Models

Intelligent risk management

To expand your cyber security process, tools must be backed by intelligent risk-management understandings that Big Data experts can easily construe. The key determination of using these automation tools should be to make the data available to analysts more easily and quickly. This approach will allow your experts to cause, classify, and handle refuge threats without delay.

Threat visualization

Big Data analytics systems can help you predict the class and strength of cyber security threats. The weightage of the evaluating sources, its complexity of a possible attack by considering the data sources and patterns. These tools also allow you to use current and previously analyzed data to get arithmetical understandings of which trends are tolerable and which are not.

Predictive models

Intelligent Big Data analytics enables professionals to build a predictive model that can provide an alert as soon as it sees an entry point for a cyber-security attack. Machine learning and artificial intelligence can play a major role in emerging such a mechanism. Data analytics based solutions enable you to predict and accelerate for possible events in the data process. The hackers of penetration testing infrastructure can easily access your database and the insights for your business process grab the information as well. Penetration testing is an simulated method of malware prediction in your IoT devices or computers, networks and check the miscellaneous things. It is like a deep driller and penetrate deep inside the database of the company, the contacts and datasets are grabbed by them easily. Although penetration testing is most wanted and necessary testing required for IT infrastructures to protect the big data storage. Penetration testing contains five stages:

o Planning and investigation
o Skimming
o Acquisition access
o Sustaining access

Analysis and Web presentation firewall (WAF) formation the results shown by a penetration test exercise can be used to enrich the strengthening of a process by improving WAF security policies.

Most of the big data management resources contains the penetration testing model to safeguard the raw data in the database. Sometimes the malware is easily incorporated with the original systems by which it will be unnoticeable by the developers which cause harmful effects to the content of the data. The cyber security issues are applicable at some of the places where the improper configurations, risky end user handling, application flaws and services drawbacks etc. Most of the malware particles enter into these kind of places.

Machine Learning

Nowadays in industry most of the big data malware effects are undergoing deep analysis using machine learning techniques. In some of our research survey papers they clearly analyzed the machine learning algorithms which are effective for the usage of Big Data storage and safeguarding

The data with the direct effect of Malwares. Machine learning always give a hope for the developers to find out such treats.

Map Reducing Models

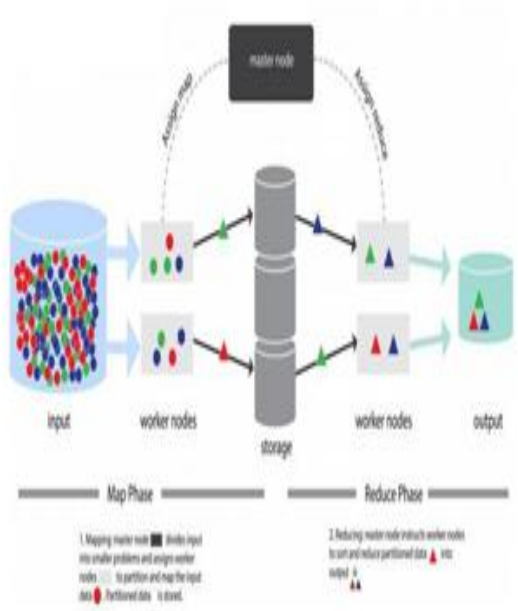

Fig. 3 Map Reducing models

Hadoop is the integrated toolset in which several tools can create complex queries to improve the efficiency of the data deep mining techniques and machine learning models, neural networks etc. Spark4 is a new framework for efficient data deep mining and big data analysis model using machine learning. The repeated usage of machine learning algorithms obviously change the working set of data and the feature extraction quality. Some of the unique databases evaluated precisely for resourceful storage and query of Big Data, including Cass, CouchDB, Green plum, HBase, MongoDB, and Vertica. Live processing does not have a single dominant technology like Hadoop, the emerging area of research and development (Cugola & Margara 2012) stated in their research

APT

Advanced persistent Threats detection is systematic model focused on high value asset storage in physical system. The highly effective spreading malwares such as the worms, Virus based malwares, Trojans attack are work in Low persistent mode. Allows for long execution delay and in the mean while disturb the system backend parameters, the dlls etc. Mostly the systems will give a trigger alert when the involvement of such malwares running in the system backend. Companies are allowed to generate their own database safeguard credibility which alert the users when such malwares enters into the system. The persistent action of such system is the targeted model of APT.

V. IMPLEMENTATION

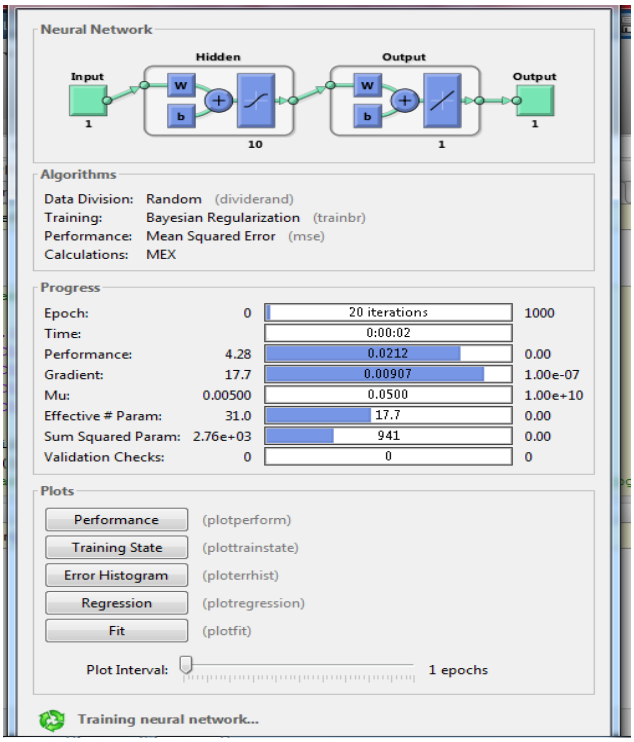

Fig. 4 Deep Learning Network for big data security

PERFORMANCE PLOT

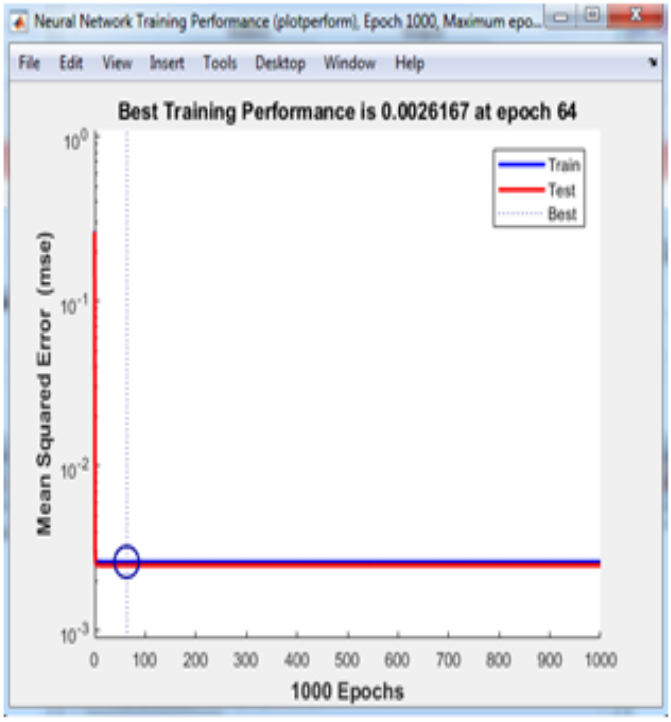

Fig. 5 Deep Learning Network performance plot

REGRESSION PLOT

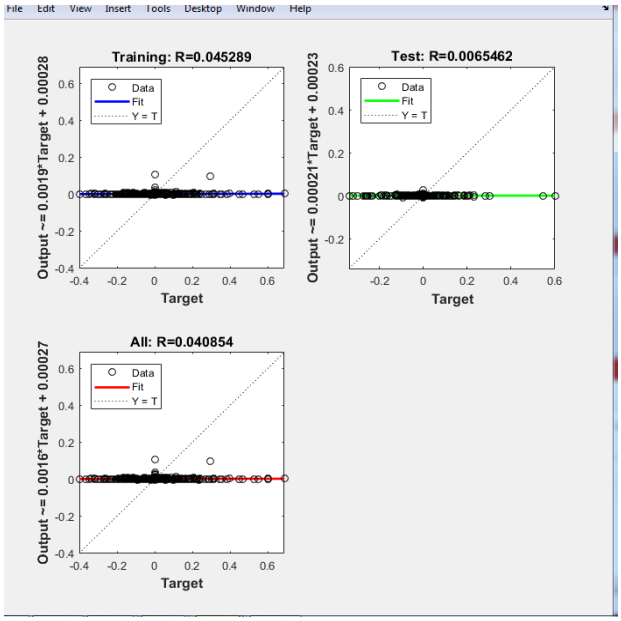

Fig. 6 Deep Learning Network regression plot

VI. RESULT

From the various analysis and reports from authors the deep discussion on technique are found out here The Machine learning algorithms plays a major role in finding the malware critical detection in every aspect of system analysis etc. The unauthorized usage of the software's, behavior of the end user clearly allows the malware enter into the depict easily. Advanced persistent techniques are most widely used in industry nowadays. Hadoop is the integrated toolset in which several tools can create complex queries to improve the efficiency of the data deep mining techniques and machine learning models, neural networks etc. Predictive models, Map based reduction models also well suitable.

VII. CONCLUSION

The sensitive part of real world is keeping the data safe and secure. The proposed survey work focuses on discussing the various impacted cyber-attack critics available in industry and the need for the cyber security etc. Big data handling and software tools which allow the malware easily into the working systems, reliable methods to avoid the miscellaneous malwares are clearly depicted here. Even through the IT infrastructure needs the high levels of security, the usage of big data storage is clearly encapsulated with the encryption schemes.

REFERENCES

1. Blenk, Basta, Reisslein, and Kellerer, "network
2. virtualization hypervisors for software defined networking,"Volume 2016. Ieee
3. Scott-Hayward, Natarajan, and Sezer, "security in software defined networks," Ieee 2016 published
4. B. C. Zhao, Zhang, hou, and A. Nallanathan," ultra-wideband propagations inspired by biological ant colony clustering," Transaction 2015
5. Wang, Cao, B. Li, S. Lee, and R. S. Sherratt, "ant colonyoptimization based clustering algorithm with mobile sinks for applications in consumer home automation networks," IEEE Trans. Consumer electronics year of 2015
6. Design of efficient and scalable offloading of control applications S. H. Yeganeh and Y. Ganjali, "KandooFinland conference 2012
7. Rexford, M. J. Freedman, and J. Wang, Scalable network Ieee 2015
8. M. Ahmadi, D. Ulyanov, S. Semenov, M. Tro in the year 2014 Malware prediction IEEE conference
9. A. V. Aho and M. J. Corasick, developed "efficient matching algorithm for cyber security" 2013
10. S. Ali and K. A. Smith-Miles, in the year 2006 Nano computing Metal earning based approach on cybernetics
11. N.-E. Ayat, M. Cheriet, and C. Y. Suen, ``Efficient model selection for the optimization of SVM kernels," IEEE 2015 model
12. Bao, Hu, Xiong, ``PSO and pattern search based memetic algorithm for SVMs parameters optimizationNano-computing IEEE 2013
13. Barros,Basgalupp, 2015 ``A hyper-evolutionary algorithm for automatically designing
14. Basgalupp, Barros, T. S. da Silva, and Software effort prediction: Decision tree approach IEEE 2013