ANALYSIS OF SMART GRIDS USING BIG DATA

V.Vaishnavi, Shaik Ashriya

*IV year students, Department of computer science and engineering*

*B.S.Abdur Rahman crescent institute of science and technology , vandalur-48*

*Abstract***---*Big data analyses extremely large data sets using various software tools to make the data meaningful and useful. The data might be in the form of structured or unstructured type which are analysed to know patterns and trends.Electricity conservation has become very important due to the increase in demand of power supply and the electricity wastages. Smart grid is a smart solution to conserve electricity as it uses information technology and consisting of controls and automation to digitally control and preserve electricity. Big data helps to analyse complex data coming from the smart grid systems. These data are vulnerable to attacks from the network ,smart meter and physical layer.***

***Keywords: smart grid, big data***

**I. INTRODUCTION**

Big data uses analytic processes and techniques that have a high accuracy on the insights of the data. The 3 v’s of big data have a great significance. Structured data is where all the data is easily interpreted and searchable and unstructured data is where the data is in the format of audio, video which makes it complex for searching purpose which tends to the Variety of data. Big data analyses large amount of data and this tends to Volume of data. The big data analysis tools like Apache Hadoop, Hive can process TB of data in a very less amount of time and this tends to Velocity of data.Smart grids are electricity distribution systems that aim at managing and conserving electricity supply and usage. Smart grids have a huge network of computers, sensors, automation ,transmission and distribution

systems. They generate huge amount of data and information as there is a two way communication between the customers and the suppliers. All the electrical parameters are monitored continuously. The consumers also have a control over the smart grid and they can monitor how much electricity is being consumed and consequently can manage their power bills.

Smart grid systems are extensively used in many sectors such as banking, telecommunication, oil distribution centres, public services, media forces. Smart grids add resilience to the power systems so that in can handle advert situations like earthquakes, storms etc. Huge power outages can be managed by the smart grids as they can isolate them and restore the power back and save people from blackouts. The grid is called smart because of its two way communication that provides automatic rerouting of power when there are outages or power failure. Smart grids draw on huge amount of home grown electricity that is more safe and isn’t prone to natural disasters and attacks.

Smart grids collect and process very huge amount of data. These data need to be analysed to make the data useful. Ancient techniques cannot be used to process and control such huge data. They will need high speed networks and processing tools to carry out transmission and churn out the data in a more easyway. Traditional tools cannot extract or unleash the required data and the hidden patterns and are greatly insufficient to provide accurate solutions. There must be some techniques to give accurate models in a shorter span of time.Since smart grid systems are extensively used in many areas they are vulnerable to attacks. Data attacks in the financial sectors have a huge impact on the economic status of a country and can give both material and immaterial losses. While the energy reliabilityis highly optimized due to new internet technologies, subsequently it has given rise to number of attacks through which the data is compromised.The vulnerabilities in the electrical grid can be in the netwroks and internet users. These attacks are generally done to steal or reveal information, alter or delete information and unauthorized access to the system.

**II. LITERATURE SURVEY**

[1] Big data platforms can be integrated on smart grids by using analysing and visualization tools.It gives a deep analysis on how smart grids work by regulating and conserving power in the transmission networks. The smart grids have different kinds of data from smart meters, transmission lines that are in accordance with the characteristics of big data and its V’s. Smart grids can be converted into a value by seeing the efficiency and sustainability of big data models.

[2] Machine learning has a great significance in the analysis of smart grids as we can use the machine learning library algorithms such as logistic regression, classification for scrutiny of smart grids attacks that are possible in the networks and physical layer. Apache spark can be used as a processing and analysis tool for these library algorithms for the detection of attacked data in the datasets.

[3] There are many attacks that are detected by big data tools but some attacks such as Distributed denial service attack and some encryption problems. DDOS attacks can be very harmful as the data is directly attacked and access to data is lost for both the customer and the supplier in terms of the smart grid. This needs to be taken care by some heterogenous intrusion systems that cannot be easily invaded into. Encryption can be done by using techniques such as SHA-1 that are hashing algorithms and digital signatures.

[4] The data sets used for the analysis of smart grid attacks consists of data from various files. These are very small files and they constitute to a larger dataset. Big data is a system that processes a single large data file and gives very efficient details and analysis on it. So whenever

using big data for processing the small files need to be brought down to a single file.

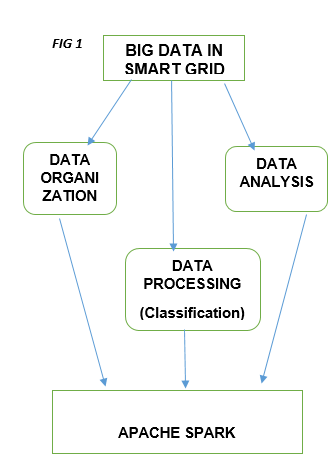
[5] The machine learning system used in smart grids has lot of stages for breaking down the data and interpreting the values. The stages include distributing the problem into different levels with algorithms, invading the small file problem and merging it into one, doing preprocessing for the data, using the appropriate library algorithms, portioning the data into training and testing and using validation for getting the accuracy. All this stages vary according to the infrastructure of the smart grid.

[6] Integrity, availability and confidentiality are the main security aspects for any model. Smart grids can have attacks like meter attacks, physical attacks. Vulnerablity of the smart grid system must be studied and set such a way that it is private and there are no loopholes in the system. Parameters must be set according to the framework of the electrical grid.

**III. PROPOSED MODEL**

**a. Big data for data analysis**

The use of big data techniques on smart grids will be an optimal solution here ,as right from storing the data to processing and to analysing the utility of the data is increased(FIG 1). Data in smart grid consists of electrical parameters such as voltage, current and the smart meters. Data coming to the smart grids can be from various sources data integration and storage is very important. After the data is cleaned it is retrieved and is sent for batch processing and for map reduce tasks that are performed by Apache spark. It is used for some real time analysis such as event detection and forecasting. After this the data is visualized to interpret the network. Since smart grid systems are a huge network they make use of distributed computing environment and big data supports this. Big data has the correct infrastructure for smart grids and can handle distributed management systems. For the approach of classification or detection in the data set taken for power systems apache spark is the big data tool used and for visualization we again use spark for weather analysis ,reviews for tweets or posts etc.

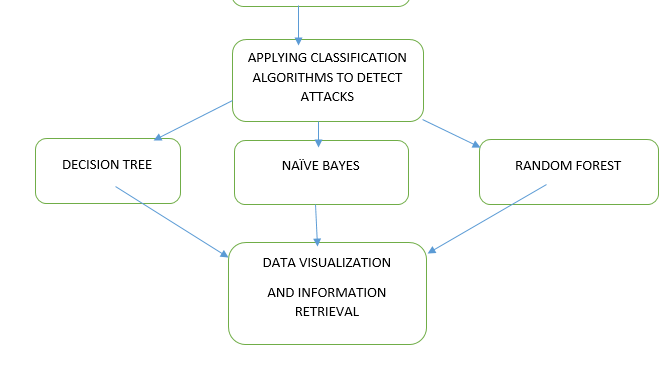
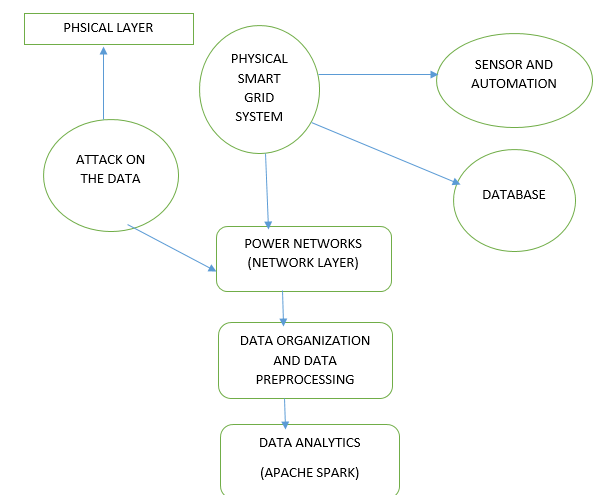


**b. Attacks on smart grids and detection of attacks using big data**

There are many attacks that been performed on smart grid systems out of which physical attacks and Cyber attacks are common. Cyber attacks are very common in the electrical distribution mechanisms which make use of two way communication .Critical components are identified such as power lines, generators where the attack is possible and the system should be resistant to that. Different approaches and algorithms are followed for different attacks on smart grids. Network intrusion detection systems can be taken for protecting the attacks on the network servers. Machine learning algorithms can be implemented with big data and they can detect the attack on the system. They are normally used for highly developed models.Smart grid face attacks such as anamoly attacks and such attacks can be identified using some classification machine learning algorithms. public data set containing the disturbances in a power system is taken and analysis is done.The data set is implemented to analyse the normal data.

**IV. EXPERIMENTAL DESIGN**

Many classification algorithms can be implemented to test which of it detects and differentiates the attack data and the normal data. The machine learning process implemented is given as in FIG 2



**FIG 2 Architecture diagram**

* The data set is got by linking and converting many files into a single big file.
* If there are any rows and columns with missing data those can be deleted in the pre processing stage by using rapid miner.
* The algorithms implemented are Decision tree, Naive bayes , Random forest.
* The data set is split into training and testing modules and is visualized.

**V. INFERENCES**

Smart grid systems consist of various network traffic. The algorithms are used to detect the attack but the attack isn’t completely removed. Some attacks like DDOS attack detection, encryption problems aren’t solved. Since there is no open data for smart grid this is just a concept based analysis. The classification parameters is not same for all dataset and so accuracy or precision need not be the lone parameter that tells the performance of the particular algorithm.. After the results are compared it can be seen that the decision tree and random forest are the best algorithms that detected the attack data. The naïve bayes classifier failed to detect normal data so the lack of normal data has adversely affected the operation.

Big data has fast response time because it can cope up with stream of data and this concept is very useful in security. The outputs obtained from the big data model ae used by decision makers, users. Visualizations help understand the problem better and come in handy because the traditional models have become outdated and cannot give good insight into the problem. A good smart grid must have confidentialityintegrity and availability and the factors that threaten these security aspects must be removed.

**VI. CONCLUSION**

This paper contains an analysis of how smart grids can be analysed and visualized using big data and how can the attacks be detected using it. Big data can be easily integrated into smart grids and those systems can be evaluated. Smart grids deal with huge amount of data and are also vulnerable to attacks. Classification has been done on a labelled data set as a concept study as it acts as a analysis component for security purposes. More robust models must be built for security reasons with accurate labelled data sets for doing research.More sophisticated intrusion detection systems must be built to save the data from attacks. As smart grids still face encryption problem hashing algorithms can be used to encrypt the data.

**REFERENCES**

*[1] Duygu Sinanc Terzi, Bilgehan Arslan, Seref Sagiroglu (2018) “ Smart grid security evaluation with big data usecase”,Department of Computer engineering,Gazi university, Antara, Turkey.*

*[2] Classification and regression ,retrieved fromURL:https://spark.apache.org/docs/2.0.2/ml-claasification-regression.html*

*[3] I. Colak, S.Sagiroglu, G.fulli, M.Yesibudak, C.F. Covrig (2016)”A survey on the critical issues in smart grid technologies”,Renewable and Sustainable Energy Reviews, vol. 54,pp. 396-405.*

*[4] S.Bende, R.Shedge (2016) “ Dealing with small files problem in Hadoop Distributed File system”, Procedia Computer science , vol.79, pp. 1001-1012.*

*[5] C.E. Sapp (2017) “Preparing and architecting for machine learning”,Gartner Technical Professional Advice ,pp. 1-37,2017..*

*[6] Y.Yang, T.littler, S, Sezer, K.McLaughlin, H.F.Wang(2011)” Impact of cyber security issues on smart grid”,Innovative Smart grid technologies (IGST Europe), pp. 1-7.*