# Detecting Attacks Using Big Data with Process Mining

Ved Prakash Mishra
Department of Computer Science & Engineering
Amity University Dubai, UAE
vmishra@amityuniversity.ae

Yogeshwaran Sivasubramanian
Department of Computer Science & Engineering
Amity University Dubai, UAE
syogeshwaran999@gmail.come

Subheshree Jeevanandham
Department of Computer Science & Engineering
Amity University Dubai, UAE

***Abstract*- In current digital world, Security has become the major issue for the organization. Every day the amount of data is growing in the world. Processing and analyzing of the data is becoming the new challenge for the analyzers. For this purpose, big data is useful to process the high volume of data in less time. Current security tools like existing firewalls and Intrusion Detection Systems are still not able to detect and prevent the attacks and intrusions in full proof manner and giving many false alarms. Big Data analytics concept could be very useful for analyzing, detection and providing full security to the organization because of the ability of handling the large amount of data. In this paper, we have described the concept and the roll of big data. We have also proposed a model using process mining to generate the alerts in the case of attacks.**

***Index Terms*— Big Data, Process Mining, Intrusion Detection System, Logs.**

## I. INTRODUCTION

As data and information's are growing day by day, we need the strong system which could handle the large amount of data and it could protect from attacks. We need to improve the current Intrusion detection system (IDS) techniques [1]so that it could give the efficient result. Big Data and process mining could be useful here to develop a system where all types of attacks could be detected in the networks. Currently various data mining algorithms have been implemented in IDS but still attacks like DDoS and Advance persistent threats are not fully stoppable. Survey says that, in different sectors of industries DDoS attacks [3] are still most active. To stop these types of attacks, we need the system which should handle the big volume of data and analysis it properly to make the correct process automatically. Here we have proposed a model which shall use big data appliances and the process mining concept. Generating the false positive alert is also a problem which can be improved using process mining.

## II. BIG DATA CONCEPT

The big data [4] can be defined as the large volume of structured and unstructured data. In the current era, where IOT is becoming very popular we need to handle the data in efficient manner since the volume of data will increase day by day. Only generating [5] the data is not important, how we are handling the data and what we are doing with that data is more important. With the big data, we can do better analysis to take the effective decisions for the business organizations.

### *A. The Characteristics of Big Data:*

- *Volume* – The amount of data that is produced is critical in this specific situation. It is the size of the information which decides the esteem and capability of the data under thought and whether it can really be viewed as Big Data or not. The name 'Big Data' itself contains a term which is identified with size and henceforth the trademark.
- *Variety* - The following part of Big Data is its variety. This implies the classification to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts. This helps the general population, who are nearly investigating the data and are related with it, to successfully utilize the data further bolstering their good fortune and consequently maintaining the significance of the Big Data.
- *Velocity* - The expression "Velocity" in the context refers to the speed of era of data or how quick the information is produced and processed to meet the requests and the difficulties which lie ahead in the way of development and improvement.
- *Variability*- This is a variable which can be an issue for the individuals who dissect the data. This refers to the irregularity which can be appeared by the data on occasion, consequently hampering the way toward having the capacity to deal with and manage with the data adequately.
- *Veracity* - The quality of the data being caught can differ extraordinarily. Exactness of analysis relies on upon the veracity of the source data.
- *Complexity* - Data administration can turn into an exceptionally complex process, particularly when extensive volumes of data originate from different sources. These data should be connected, associated and corresponded keeping in mind the end goal to have the capacity to get a handle on the data that should be passed on by this information. This circumstance, is along these lines, named as the "unpredictability" of Big Data.

## III. INTRUSION DETECTION SYSTEM

In current business scenario where everyone is using the computer, network and internet, intrusion bypass the security mechanism of network or computer. Detecting the intrusions

IJSMS

on network or protecting the unauthorized access of networks or computers we implement the system called intrusion detection system. Intrusion detection [3] system monitor the network and computer continuously and give the alerts immediately in case of any malicious activity.

*A. Types of IDS:*

There are three major types of IDS are available.

- Host Based System: Host based system IDS works on the information collected from the individual computer.
- Network Based System: Network based IDS detects the attack by analyzing the network packets.
- Application Based System: Application based IDS analyses the events within the software applications.

## IV. WHY PROCESS MINING IS USEFUL OVER DATA MINING

Process mining technology combines the strengths of both data mining and process modeling: By automatically creating process models based on existing IT log data, process mining yields live models that are connected to the business and can be updated easily at any point in [3] time. Using Process mining we will get real time process execution model that help us to track how data is generated and how it moves from one point to next and who handles the data.

## V. PROPOSED BIG DATA SYSTEM MODEL

Our big data system is divided in five steps:

A. *Event Data:* At this stage data are generated from different places e.g. from files, directory, server, logs etc.

B. *Data Collector:* At this stage data will be collected and transferred to big data appliances.

C. *Big Data Appliances:* The big data appliance gives the high performance and secure platform to run the many types of workloads on NoSQL and Hadoop system.

D. *Data Analysis and Process Mining:* At this stage, using logs process will be created automatically and it will be matched with predefined processes with help of delta analysis.

E. *Alert Generation: F*inally, at last step alerts will be generated if there is any miss match or suspicious activity will be found.

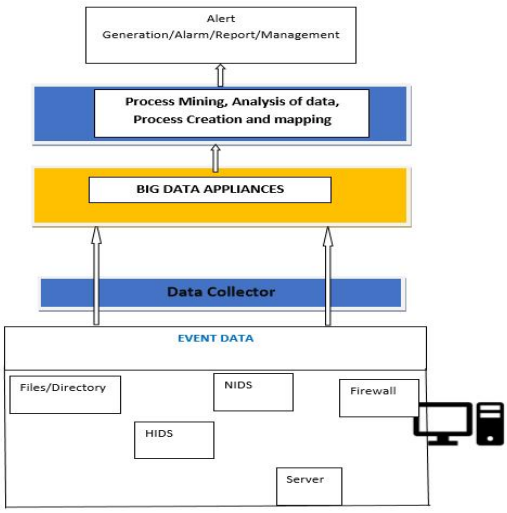



Fig 1. Big Data System Model

## VI. SYSTEM MODEL DISCOVERY

In this Big Data System model, the process that will take the event data as input and generate the attack model as output must be defined. As mentioned before, event data will be act as input and the data will be passed to data collector. Data collector is the process of collecting and measuring the data on targeted variables in an established network.

After collecting and measuring the information the data will be passed to Big Data appliances. Big data appliance is a connector to simplify the big data analytics and data integration and also provide high speed access to Hadoop from oracle database. Then the data will be going to process mining, in process mining the data will be analysis and find out which data will generate the causes for attacks and the triggered information will be send as an alert to administration or to the IDS network. So that the attack will be detected easily by using this proposed model.

## VII. CONCLUSION

In this paper, we have proposed the big data system model to develop the Intrusion detection system using process mining. Using big data appliances with the help of Hadoop and NoSQL, big volume of data can be managed easily. Using process mining we can easily create the process automatically using the logs on real time basis. This will help to save the data from many types of attacks.

As future work, the objective is to extend the attack perspectives and analyze the alerts from another view-point. Besides it was observed that some models become complex as the number of distinct signature increases[6] Therefore, Big Data with process mining techniques maybe employed to reduce the complexity of those models and easy to implement, such as simplicity, maybe employed to evaluated the model.

REFERENCES

[1] Van der Aalst, W.M.P & De Medeiros, A.K.A, Process mining and security: Detecting Anomalous process executions and checking process conformance, Electronic Notes in Theoretical Computer Science, Vol 121, No. 4, pp. 3–21 (2005).

[2] Pawar, M.V, & Anuradha J, Network Security and Types of Attack in Network, Procedia Computer Science 48(2015), ELSEVIER, ScienceDirect, pp 503-506 (2015).

[3] Ved Prakash Mishra and Balvinder Shukla, "Process Mining in Intrusion Detection-The need of current digital world "ICAICR- 2017, March 17-18, 2017.

[4] Miss Gurpreet Kaur Jangla1 and Mrs Deepa. A. Amne, "Development of an Intrusion Detection System based on Big Data for Detecting Unknown Attacks" IJARCCE, Vol. 4, Issue 12, December 2015, pp. 229 – 232.

[5] Dr.KiranJyoti, Bhawna Gupta. "`Big data analytics with hadoop to analyse targeted attacks on enterprise data"'.Technical Report, International Journal of Computer Science and Information Technologies, IJCSIT, Vol 5(3) 2014.

[6] S. O. Al-Mamory and H. Zhang, "Intrusion detection alarms reduction using root cause analysis and clustering," Computer Communications, 2009, vol. 32, no. 2, pp. 419-430.

IJSMS