International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 12 | Dec 2019 www.irjet.net p-ISSN: 2395-0072

Review on "Using Big Data to Defend Machines against Network Attacks"

Miss Sonali S. Thangan[1], Dr. V.S. Gulhane[2], Prof. N.E. Karale[3]

[1]ME second year, DRGITR, Amravati
[2]Sipna COE, Amravati
[3]DRGITR, Amravati

Abstract - *Today, information security solutions fall into two categories: analyst driven, or unsupervised machine learning driven. Analyst-driven solutions rely on rules determined by fraud and security experts. Experts derive these rules based on their experiences, and on the intuitions developed during investigations and analyses of previous successful attacks. This expert-driven process usually results in a high rate of undetected attacks (false negatives), and introduces a delay between the detection of new attacks and the implementation of countermeasures necessary to prevent future instances of these attacks. Moreover, bad actors often figure out current rules, and design newer attacks such that they can avoid being blocked or detected. We present AI2, an analyst-in-the-loop security system where Analyst Intuition (AI) is put together with state-of-the art machine learning to build a complete end-to-end Artificially Intelligent solution (AI). The system presents four key features: a big data behavioral analytics platform, an outlier detection system, a mechanism to obtain feedback from security analysts, and a supervised learning module. In the system we use a machine learning and big data combination system which outperforms other non-machine learning systems and provides a solution for network attack detection on real time datasets.*

Key Words: Information security, attacks, AI2, big data, network attack detection.

1. INTRODUCTION

Big data is one of the most talked topic in IT industry. The term Big Data refers to a massive amount of digital information. Due to the heavy usage of internet, smartphones and the social networks, the data volume in the world is dramatically increased in a way that the traditional systems cannot hold these data in terms of storage and processing. Big data analytics provide new ways for businesses and government to analyse structured, semi-structured and unstructured data. It is not a single technique or a tool, rather it involves many areas of business and technology. In recent, the cyber attacks in big data are increasing because of the existing security[1] technologies are not capable of detecting it. Many intrusion detection systems are available for various types of network attacks. Most of them are unable to detect recent unknown attacks, whereas the others do not provide a real-time solution to overcome the challenges. To detect an intrusion in such ultra-high-speed environment in real time is a challenging task.

1.1 BIG DATA

Big Data[2] is a large collection of structured, semi-structured and unstructured datasets that cannot be stored and processed using traditional computing techniques[3]. In most enterprise scenarios the volume of data is too big or it moves too fast or it exceeds current processing capacity.

1.2 RESEARCH MOTIVATION

There are many technologies used to prevent the intrusion [4] in big data. Following are some of the techniques to maintain cyber security[5].

- Firewall: A firewall is a network security system, either hardware or software based on a set of rules, acting as a barrier between a trusted / untrusted networks. The firewall controls access to the resources of a network through a positive control model.

- Intrusion Detection System: An IDS inspects all inbound and outbound network activity and identifies suspicious intrusion attempts.

Though they both relate to network security, an IDS differs from a firewall in that a firewall looks out for intrusions in order to stop them from happening. The firewall limits the access between networks in order to prevent intrusion and does not signal an attack from inside the network. In the existing security[6] system, researchers were developed various security technologies to protect the system from various types of network threats and attacks. Most of them are unable to detect recent unknown attacks, whereas the others do not provide a real-time solution to overcome the challenges.

2. ARCHITECTURE OF REAL-TIME INTRUSION SYSTEM

The real time intrusion system is mainly working and predicting based on the different types logs available in distributed data processing layer. The intrusion dictionary will be available on an in memory database for quick access. Whenever a user interaction happens, the user interaction logs will be processed and compared with the Intrusion

dictionary and the output will be stored in the Real-time output layer. If the output matches with the Intrusion dictionary entries, then the system will generate alerts and can take actions such as blocking the user from further interactions, blacklist the user, blacklist the IP address etc.

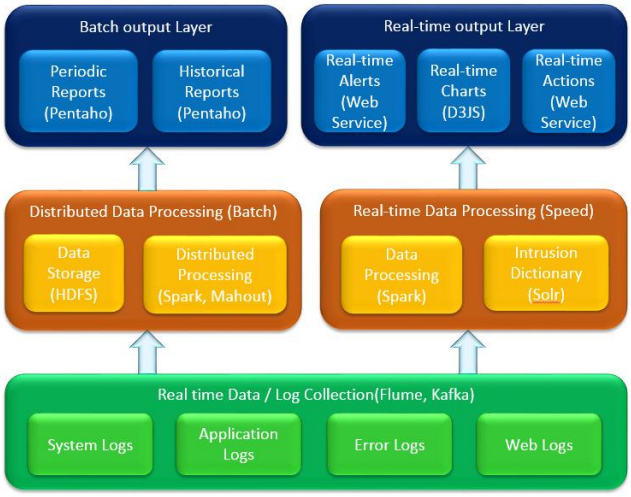

Fig.-1: Real-Time Intrusion Detection System Architecture

3. WORKING OF AI2 ALGORITHM

AI2 works in following four steps-

3.1 Filtering of unwanted packets-

A single packet consists of four components port number, packet size, protocol and packet type.

- Data coming from the same port number again and again and it is of no use to the system such data is discarded
- Very low size and high size data packets are discarded because such data is not much useful to the system instead it increases the risk of intrusion attack.
- Packets coming from other network but not using standard protocol like http, https such packets are discarded.
- Packets having unwanted type are discarded.

3.2 Clustering of data-

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

Here we will use k-means algorithm for data clustering. In this first we divide the data in different sizes of clusters. On these clusters we will apply AI to check for abnormal data i.e. data not following proper data patterns. If we found abnormal data in clusters, we will discard such clusters for being processed.

3.3 AI based classification-

The remaining clusters from above step should go through AI based classification. It finds out the unwanted packets, normal packets and most attacked packets. We have to track and identify IP addresses of such packets which should be discarded from future use.

3.4 Clustering and AI based classification

If the number of attacks increased or very low then we have to tune up the steps two and three of the algorithm to get the moderate attack.

4. WORKING OF PROPOSED SYSTEM

This system is used to defend machine against network attacks. Here we use different modules to get the purpose of the system. For the purpose we require big data to find the attacking packets, such big data we obtain from "kddcup99" databases from the internet. Then we provide this datasets to the particular module so that we can identify the most affected data packets and restrict such packet from entering in the system. By restricting the affected packets from entering into the system we can protect the system from intrusion attacks.

There is no such GUI for end users because it's of no use as the system is for administrative purpose. For understanding purpose we have provided a window with a single button. This is shown as below.

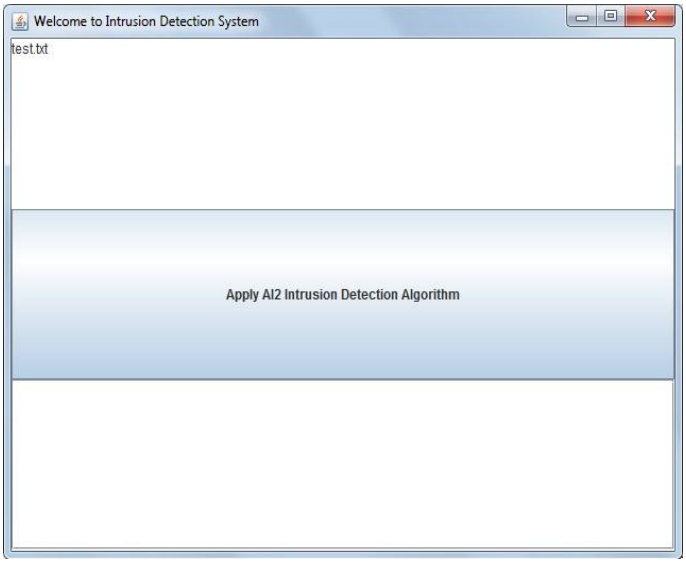

Fig.2: Home Screen

Button with label "Apply AI2 Intrusion Detection Algorithm" is given on the home screen. By clicking on this button we start the system, the dataset files obtained from kddcup99 are provided and the system apply algorithms like k-means and c-means to find out the affected packets in the dataset. The system calculates the accuracy of the intrusion detection. More the accuracy less be the risk of attacks. We can repeat this procedure of finding the attacking packets until we get the more accurate results from the system. At the end, the system will display the pie chart showing four parameters delay, accuracy, intrusions and normal.

Delay represents the total delay made by the system to the calculation.

Accuracy represents the percentage of accurate finding of intrusion packets.

Intrusions represent the number of types of intrusions detected in the system.

Normals represent the normal packets coming to the system.

Following pie chart showing the four parameters as discussed above-

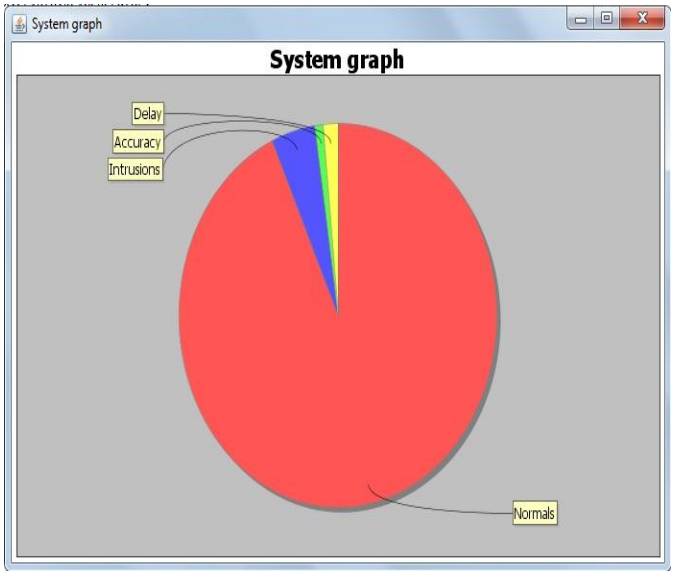

Fig.3: Pie Chart

From the above we can get the clear idea of accuracy, intrusions and the normal packets.

5. APPLICATIONS

As we know today world is totally dependent on internet. Each and every information we want to store on clouds as we think they are much safe than hard copies. It is somewhat right but sometimes it is more risky to share our confidential data on clout. It may affected by different intrusions and can be hacked. We can avoid the risk of hacking data from cloud by using this system for intrusion detection purpose.

This system can be used in large organizations where more confidential work is being carried out.

This system can be mostly used in national and international banks as banks have much user information like account details and all.

6. CONCLUSION

We present an end-to-end system that combines artificial intelligence with state-of-the-art machine learning techniques to detect new intrusion attacks and reduce the time elapsed between attack detection and successful prevention. The system presents key features: a big data behavioural analytics platform, and intrusion detection system. The system carried out the purpose of finding the intrusion attacks on cloud data packets to make the system safe from external intrusion attacks.

7. REFERENCES

[1] Butun I, Morgera SD, Sankar R (2014) A survey of intrusion detection systems in wireless sensor networks. IEEE Commun Surv Tutor 16(1):266–282 CrossRef.

[2] "Big Data: A Primer". Written by "Deepak Chenthati, Hrushikesha Mohanty, Prachet Bhuyan".

[3] "Santhoshkumar and R.H Gowder" publication on "International Journal of Future Computer and Communication, Vol. 1, No. 4, December 2012".

[4] Ngadi M, Abdullah AH, Mandala S (2008) A survey on MANET intrusion detection. Int J Comput Sci Secur 2(1):1–11

[5] Denning DE (1987) An intrusion-detection model. IEEE Trans Softw Eng 13(2):222–232. doi:10.1109/TSE.1987.232894.

[6] http://link.springer.com/article/10.1007/s11227-015-1615-5.