Network Packets Analysis for Threat Detection

Mr. Ganesh D. Bagul[1], Prof. M. B. Vaidya[2]
PG Student, Department of Computer Engineering, Amrutvahini COE, Maharashtra, India
Professor, Department of Computer Engineering, Amrutvahini COE, Maharashtra, India

ABSTRACT

Introducing the smartest way to help protect your important information by notifying malicious threats, by detecting stealthy anomalies, and reumpire them with our TDS(Threat Detection system). We give you monitoring network data (packets) and TDS system. Organizations invest heavily to block advanced attacks, on both endpoints of networks. Even though all this investment, devices continue to be cooperation in increasing numbers and high profile disobey continue increased. In order to provide the most highly secured network environment, network traffic monitoring and threat detection systems must handle network data from different networks. Though numerous investigations have yielded real-time threat detection systems, in this paper We addressed the issue of handling the large volumes of network traffic data of network systems. Particularly, We introduced and evaluated a Network based threat detection system that can rapidly analysis highly rigorous network traffic data. For that utilizing the network-based clustering algorithms to detect irregular network activities.

Keyword—*Threat Detection, Cloud Computing, Streaming, MapReduce, Hadoop, Spark.*

1. INTRODUCTION

Threat monitoring systems, must have the capacity to observe activities in big data collected from networks and detect threats. In developing effective threat monitoring systems, there are two major difficulty. First, the amount of data available to the threat monitoring system is big. This rate of data collection and transfer is not possible for traditional data analysis system to effectively process. Second, detection systems, and the resulting defensive actions taken are only effective if they can detect threat exactly and in a timely manner, minimizing the impact of the attacks. Therefore, the detection of threat needs to occur in real time. Through a huge amount of research investigations, numerous and various network monitoring and threat detection systems have been developed.

Also, a number of research efforts have been carried out on streaming data. However, the formulation of a threat monitoring system which can handle the big volumes of network traffic data of network systems, while simultaneously providing real-time monitoring and detection, remains doubtful.

There are two types of clustering Algorithm methods, are as follows:

1.1k-means clustering algorithm

1.2Fuzzy c-means clustering algorithm

To evaluate the performance of our developed system, we conducted experiments using a real-world network traffic data set . The scope of the evaluation concerns the threat detection and system performance. To evaluate threat detection, we assessed our detection schemes with simulated distributed denial-of-service (DDoS) and traffic flooding attacks randomly embedded into original data sets, to simulate real-time net-work attacks. Notice that those attacks are only examples to demonstrate the effectiveness of our developed system and our system can be extended to defend against other types of attacks. Finally, to test system performance, we evaluated our system with streaming data and non-streaming data to compare the efficiency between both implementations. Based on our experimental results, our proposed system supports monitoring and detection of intensive streaming data with good system performance.

2. REVIEW OF LITERATURE

Some of the existing intrusion detection techniques, which could provide high accuracy, low false positive rate and reduced number of features. It also covers the detailed analysis of various available intrusion detection tools available for

detecting intrusions in network system. This work identifies a number of important design and implementation issues, which provide a framework for evaluating or deploying commercial intrusion detection systems.

Some of the techniques used are statistical approaches, predictive pattern generation, expert systems, keystroke monitoring, state transition analysis, pattern matching, and data mining techniques. Since 1970, several people have reviewed the state of the art, including: Anantvalee , Kabari , Bass, Jeyanthi and Michel , Yang , Adam , Lee , Mukherjee et al. , S. Kumar and Lakhotia , and Lee. Et al . The best reviews are those that present an unbiased, thorough review of the literature, and/or provide a good taxonomy for describing different intrusion detection methods. Examples of such reviews include those by Axelsson, Debar , Almgren , and Hall, M., Jackson wrote an excellent in-depth survey of commercial products.

3. PROBLEM STATEMENT

For network security in cloud computing and for threat control and maintaining data secure on network traffic. We developing new threat detection system for live network called ”Network Traffic Analysis and TDS (Threat Detection System). For Threat detection purpose we use streaming based clustering algorithm as k-mean algorithm in order to proves the efficiency of the proposed work.

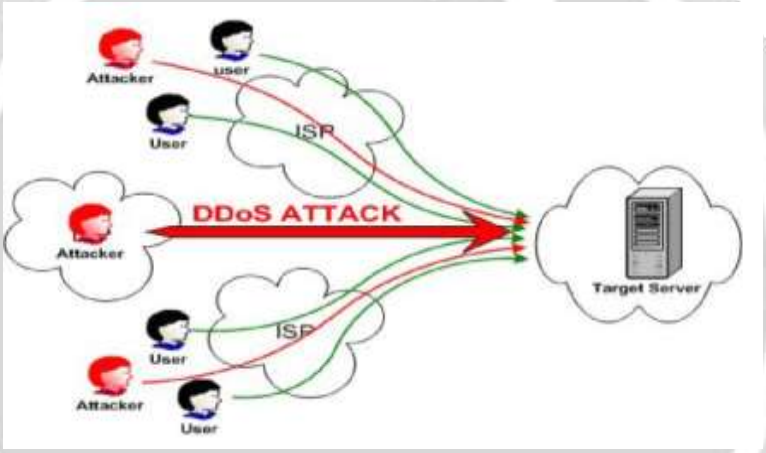

Fig. 1. DDoS Attack

4. PROPOSED WORK

To demonstrate the effectiveness of our developed net-work threat monitoring and detection system, we conducted experiments from three aspects: network monitoring, threat detection, and system performance. In network monitoring, we designed different setup to find the outgoing traffic volume from each server, the traffic volumes based on source and destination IP address, the incoming traffic volume to each server, and the port access count on each server in a given time duration. We implemented k-mean clustering algorithm in our proposed system to fast categorize data into different groups based on their parameter. With dynamic thresholds, we conducted threat detection on follow distributed denial-of-service (DDoS) network traffic and measured both detection rate and false positive rate. In addition, we compared the effectiveness of Hadoop and Spark for the data processing in network monitoring and threat detection. Through our exten-sive evaluations, our result shows that our proposed system can efficiently help the system administrator to monitor net-work activities and identify abnormal behaviours. Moreover, our proposed system can accurately and dynamically detect network threats. Our experimental data shows that, with the implementation by Spark, the system performance is almost 30 times better than that of the implementation by Hadoop.

5. SYSTEM ANALYSIS

5.1 A. System Flow

The detailed workflow of the system is described as follows:

- Data Collection: Notice that data can be collected from all kinds of network or computer devices, including routers, switches, Wi-Fi access points, servers, etc. We used the Wire-shark and JNetPcap library to collect data to demonstrate the effectiveness of our developed system. The collected data from Wireshark is in binary format and needs to be pre-processed before streaming into Spark. JNetPcap component captures data, meanwhile, extracts the useful information before the data is sent to the next step.

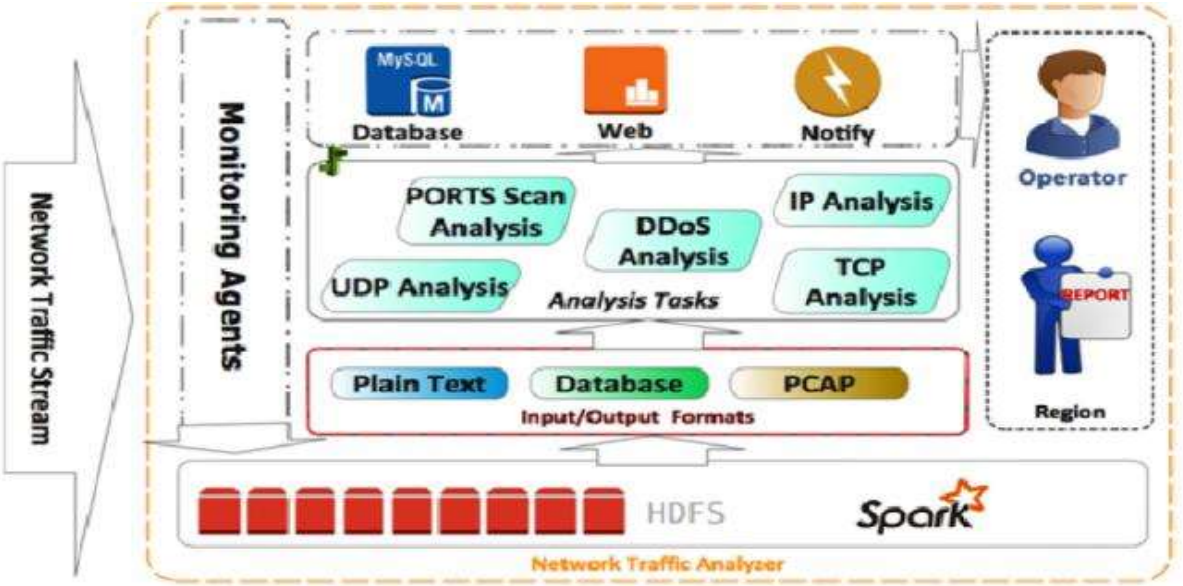

Fig. 2. System Architecture

- Pre Processing: In our designed system, data collected from different network devices or servers may have different formats (e.g., binary or text format). In the case of the binary data, it will not be understood by the Spark framework, and must be converted it to text format. Using the JNetPcap library, we developed a java application to extract the desired information (source IP address, destination IP address, time, size, etc.) from the original binary PCAP file. The text format data is then fed the Flume server.

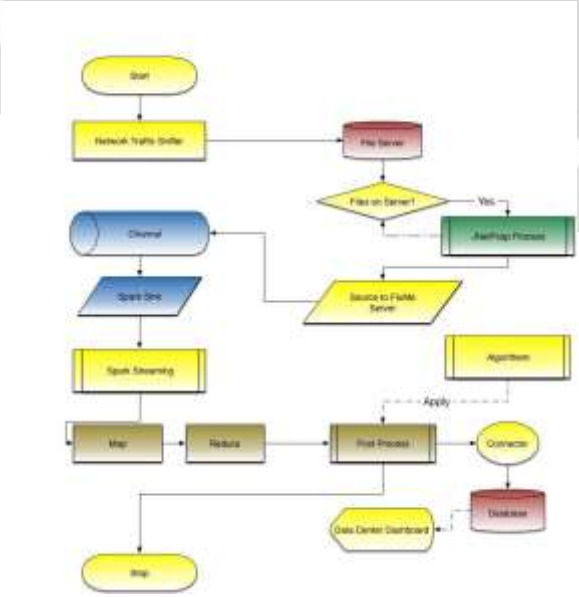

Fig. 3. System Flow

- Flume System: To enable real-time streaming data anal-ysis, we built a Flume server, which can efficiently move a large amount of data from different sources to a centralized data storage. In our developed system, the destination of the Flume Server is HDFS.
- Spark Streaming: The Spark streaming component of the Spark framework makes it possible to handle the streaming data in real time. The execution frequency can be adjusted as short as 0.5 second intervals. This component simply enables the streaming feature for the system.
- Data Mapping: Data mapping is the process of mapping raw data to pre-defined key/value pairs, the output can be used as input for the data reducing component.
- Data Reducing: With the shuffle and sorting processes, the intermediate processed results that are sitting in memory or intermediate files will be further aggregated and fed to Reduce(). Based on the defined key/value pair in Reduce() by the data center operator, individual results processed and output by the individual intermediate files can be combined to generate the final result. For example, the port access volume of the particular server IP address in different intermediate memory/files can be summarized to illustrate the total traffic volume.
- Data Storage: After data is analyzed and processed, the final result will be stored in the distributed databases, which will be used for future detection. Using MySql database for cost cutting and efficiency, we set up a cluster server with distributed, share-nothing architecture to maintain a high up-time.
- Notification: Once any malicious activity is detected by the system, email notification or SMS notification (optional) will be triggered, and system administrators and data center operators will receive the detailed information immediately. The necessary actions to isolate the affected servers or network devices can then be taken.

5.2. Proposed Algorithm

Algorithm:Steaming k-means Clustering Algorithm

The concept of the streaming k-means algorithm is to separate data into k clusters and update the center of each cluster continuously. Updating the cluster center allows forgetfulness, permitting adaptation to changes over time. Using a configuration that treats all initial data points equally, the process of partitioning relies on two key steps:

- Defining Centers: Initially, we do not have any data in the system that can be used for the analysis, so we must randomly define two center points.

- Updating Centers: Each cluster center will be continually updated with the incoming streaming data set. The forgetfulness mechanism that we use to update the clusters centers treats all new incoming data as equivalent to existing data, computing and updating the new center.

5.3 Mathematical Module

Let S be the proposed system which can be represented as S=P1, P2, P3

Where

Set P1 is flume server Set P2 is Administrator Set p3 is Network Monitor

Objects in whole system are O1 is Create environment object O2 is Crete Environment

O3 is Display Environment

O4 is Registration
O5 is Network packet

O6 is Capturing Packet

O7 is Applying Agorithm

O8 is Find out Attack

O9 is DDos attack detection

O10 is Altering or Notification

O11 is Mapping Data

O12 is Reducing Data

S= O2, O3, O4, O5, O6, O7, O8, O9

P1= O1, O2, O3, O4, O6, O7, O8, O11,O12 P2= O1, O2, O3, O4,O10

P3= O5, O6, O9 Input=Network Packets

Output=Threat Detection and Notification P=Fx— Iput Output
Fx is the function which take the input Network Traffic and give output as Threat Detection and give notification Success Condition= Find out DDoS attack successfully. Failure Condition= Capturing and detection the malicious packet fail.

6. CONCLUSION

In this paper, We proposed, a network monitoring and threat detection system to manage significant network traffic data streams. We introduced a cloud computing model by using Flume, Sharp, and Hadoop to well analyze streaming network packets. Our implementation includes Streaming k-means clustering or Fuzzy c-means algorithms for the purpose of clustering normal data in order to extract malicious network packet. Through my evaluations of threat detection and system performance, I demonstrated the feasibility of out proposed system for enterprise networks.

7. ACKNOWLEDGMENT

A successful work of report is the result of inspiration, sup-port,guidance,motivation and cooperation of facilities during study. It gives me great pleasure to acknowledge my deep sense of gratitude to present of pgcon paper. I titled:Network Traffic Analysis and TDS I place a deep sense of appreciation to my project guide Prof. M. B. Vaidya and PG co-ordinator Prof. S. K. Sonkar giving me all possible help and suggestions to give my Project a perfect shape. I am thankful to Head of Department Prof. R. L. Paikrao for giving me an opportunity to complete the Project Stage-I successfully. Also I am thankful to Principal of Amrutvahini College of Engineering for their kind co-operation in completion of this work. Last but not least I have to express our feelings towards all staff members of Amrutvahini College of Engineering and special thanks to my colleague and friends for their moral support and help.

8. REFERENCES

[1] ”A Streaming-Based Network Monitoring and Threat Detection System” Zhijiang Chen, Hanlin Zhang, William G. Hatcher, James Nguyen, Wei Yu Department of Computer and Information Sciences Towson University, Maryland, USA 21252, USA

[2] C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. Computers Geosciences, 10(2-3):191203, 1984.

[3] CAIDA Data. http://www.caida.org/data/, 2015.

[4] H. Guo, Y. Li, and S. Jajodi. Chaining watermarks for detecting malicious modifications to streaming data. Information Sciences, 177(1):281298, January 2007.

[5] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):881892, July 2002.

[6] SLPARK. http://spark.apache.org, 2015.

[7] V. Yegneswaran, P. Barford, and S. Jha. Global intrusion detection in the domino overlay system. In Proceedings of the 11th IEEE Network and Distributed System Security Symposium (NDSS), 2004.

[8] W. Yu, Z. Chen, G. Xu, S. Wei, and N. Ekedebe. A threat monitoring system in enterprise networks with mobile devices. In Proceedings of ACM International Conference on Reliable and Convergent Systems (RACS), 2013.
[9] P. Zikopoulos and C. Eaton. Understanding big data: Analytics for enterprise class hadoop and streaming data. Book, 2011.