ISSN : 2494-9150 Vol-01, Issue 11, FEB 2016.

REVIEW OF DISTRIBUTED INTRUSION DETECTION SYSTEM

[1]Mahesh Shingte, [2]Faizabanu Siddiqui, [3]Akshata Yewale

[1,2,3]Department of IT, K J Somaiya Institute of Engineering & IT, Sion, Mumbai, Maharashtra, India.

[1]mahesh.shingte@somaiya.edu, [2]akshata.y@somaiya.edu, [3]faizabanu.s@somaiya.edu

***Abstract* — Intrusion detection system (IDS) are essential components in a secure network environment, allowing for early detection of malicious activity and attacks. By employing information provided by an IDS it is possible to apply appropriate countermeasures and mitigate attacks that wound otherwise seriously undermine network security. However, current high volumes of network traffic over the most ids techniques requiring new approaches that are able to handle huge quantities of traffic during analysis while still maintaining high throughput. We propose architecture for Distributed network Intrusion in a cloud computing environment. Network traffic, operating system logs and general application data are collected from various sensors in different places in the network, comprising networking equipment, server and users workstations. The data collected from different sources is aggregated, processed and compared using the map reduce framework, analyzing event correlations which may indicate intrusion attempts and malicious activities. The proposed architecture is able to efficiently handle large volumes of collected data and consequent high processing loads seamlessly scaling to enterprise network environments. Also, differently from previous IDS models, it capable of detecting complex attacks through the correlation of information Obtaining from different sources, identifying patterns which may not be apparent in centralized traffic captured or single host log analysis. Besides an architecture description. We present feasibility results based o experiments performed on real cluster and cloud infrastructure.**

***Keywords*— *Big data; Hadoop; Hadoop Distributed File System (HDFS); MapReduce*.**

I. INTRODUCTION

An intrusion detection system (IDS) provides around the clock network observation and is an additional wall to secure the network. The intrusion detection system is a process of determining an intrusion into system through the observation of available information concerning the state of the system, monitoring user activities and reporting to a management station. Intrusion detection refers to the detection of cruel activity (break-ins, penetrations, and other forms of computer abuse) in a computer related system. Such systems perform automatic detection of intrusion attempts and malicious activities in a network through the analysis of traffic captures and collected data in general. Such data is aggregated, analysed and compared to a set of rules in order to identify attack signatures, which are traffic patterns present in captured traffic or security logs.

Intrusion detection systems detect malicious activities through basically two approaches: anomaly detection and signature detection. In traffic anomaly detection, first a standard traffic pattern statistical profile is established and then it is compared current traffic in order to detect any deviation from the expected normal behaviour. In signature detection (which has been discussed before), network traffic is compared with attack signatures stored in a database in order to detect- specific attacks. Anomaly detection is capable of identifying attacks that were not previously

observed but this kind of technique is always subject to a high rate of false positives.

Fig 1.1Distributed Intrusion Detection System

A) Types of IDS

There are two types of IDS

1) Host based

Host Intrusion Detection Systems (HIDS), which are based on data collected from individual hosts. HIDSs are composed basically by software agents which analyse application and operating system logs, file system activities, local databases and other local data sources, reliably identifying local intrusion attempts.

HIDS involves software or agent components, which monitors the dynamic behaviour and state of the computer system. HIDS software runs on the server, router, switch or network machines. The agent version has to report to a console or it can run on together on the same host as shown in Figure. Examples are: Buffer overflow, rootkit, format string etc. The software creates log files of the system in the form of sources of data. The host based IDS looks at communication traffic and checks the integrity of system files to keep an eye on suspicious processes. Host based IDS doesn't provide good real time response.

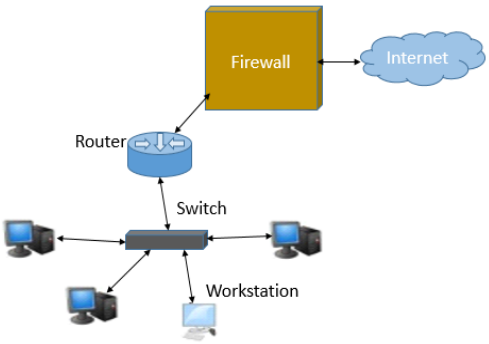

Fig 1.2 architecture of HIDS

2) Network Based

Network intrusion detection systems identify attacks through the analysis of network traffic captured at the network border, thus containing traffic flowing to and from all internal hosts. This kind of IDS is capable of processing packet captures containing traffic from several nodes with little or no network overload. It is secure against internal and external attacks as it functions invisibly in the network, simply capturing packets in promiscuous mode.

NIDS attempts to discover unauthorized access to a computer network by capturing the network traffic packets such as TCP, UDP and IPX/SPX and analyses the content against a set of rules. Examples are: Eavesdropping, data modification, identity or IP Address Spoofing, Denial-of-Service (DoS) attacks, Man-in-the-Middle Attack etc. NIDS consist of a set of single-purpose sensors that are placed at various points in the network. These sensors monitor and analyse network traffic and send report of attack to the centralized console. The deployment of NIDS has a minute effect on the performance of the network.

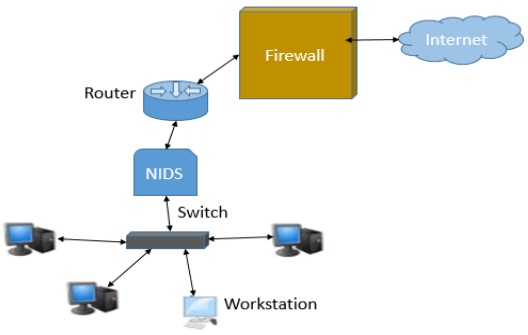

Fig 1.3 architecture of NIDS

B. Methods of detection of IDS

There are two methods of detection of IDS

1) Misuse/Signature based detection

This method uses specifically known patterns of unauthorized behaviour, called signatures, to predict and detect subsequent similar attempts. This method is extremely accurate for known attacks. It produces low false alarm. With the help of this technique, we can cover a broader range of unknown attacks. Another advantage is that signatures are easy to create and understand only if the network behaviour is known that is required to identify. The disadvantage of this method is that it can only detect intrusion that matches a predefined pattern, a set of signature must be continuously updated to detect a new attack and it can't detect novel attacks. Signature based detection does not work well when the user uses advanced technologies like no generators, payload encoders and encrypted data channels . The efficiency of signature based systems decreases as the number of new attacks increases because it has to create a new signature for every new attack.

2) Anomaly based detection

Anomaly detectors are designed to identify abnormal patterns of behaviour on a host or network. It functions on the assumption that attacks are different from normal activity and can be detected by systems that recognize these variations. Anomaly detectors create a list of profile data as a normal data representing normal behaviour. It automatically detects any deviation of it and generate alarm. It has the capability to detect new types of errors. There are many measures and techniques that are used in anomaly detection including; Threshold detection, statistical analysis, Rule-based measures, other measures, including neural networks, genetic algorithms, and immune system models. One advantage of using this kind of intrusion detection is that we can add new rules without modifying existing ones. It has the ability to detect novel attacks. But this approach produces many false alarms and dally time consuming for research intensive to obtain update accurate and comprehensive profiles of normal behaviour. Therefore, it needs a large set of training data with network environment system logs.

C. Intrusion detection in cloud computing environment

Cloud computing is a collection of sources in order to enable resource sharing in terms of scalability, managed computing services that are delivered on demand over the network. Its users need not to buy infrastructure, software, resources, as a result saving a large amount of expenditure. Cloud basically provides services through a third party. The third party provides services and resources on rent and users pay per use. This will save a lot of money and provides a greater flexibility to move from one service to another service

Cloud computing is emerging day by day. People are using its services very frequently and they don't have any other alternative for its services. But users are unaware about the security and privacy concerns in a cloud environment. Security threats can be in terms of intrusion prospects and DoS attacks. Organizations need to provide firewalls, intrusion detection and prevention techniques, authentication, encryption and other powerful hardware and software protection to secure the stored data.

In a traditional network, IDS monitor detects, and alert the administrative user by deploying IDS on key network choke points on the user site. But in Cloud network IDS has to be placed at cloud server site and entirely administrated and managed by the services provider. The intrusion data communicates through the service provider and user has to depend on him. The cloud service provider would not like to notify user about the loss and hide the information to make a good image and reputation. So an unbiased third party monitoring service can guarantee adequate monitoring and alerting for cloud users. The Intrusion detection message exchange format (IDMEF) is an XML standard format that has been used for message exchanged among IDS sensors. The IDMEF contains the attack name or signature, time of creation and analysis, source and target of intrusion. Alerts generated are sent to 'Event Gatherer' program. Event

Gatherer receives and convert alert messages in IDMEF standard and stores in event database repository with the help of Sender, Receiver and Handler plug-ins.

II. LITERATURE SURVEYED

Intrusion Detection Systems (IDS) are important mechanisms which play a key role in network security and self-defending networks. Such systems perform automatic detection of intrusion attempts and malicious activities in a network through the analysis of traffic captures and collected data in general. Such data is aggregated, analyzed and compared to a set of rules in order to identify attack signatures, which are traffic patterns present in captured traffic or security logs that are generated by specific types of attacks. In the process of identifying attacks and malicious activities an IDS parses large quantities of data searching for patterns which match the rules stored in its signature database. Such procedure demands high processing power and data storage access velocities in order to be executed efficiently in large networks.

Intrusion detection systems are further can also be classified in two groups, Network Intrusion Detection Systems (NIDS), which are based on data collected directly from the network, and Host Intrusion Detection Systems (HIDS), which are based on data collected from individual hosts. HIDSs are composed basically by software agents which analyze application and operating system logs, file system activities, local databases and other local data sources, reliably identifying local intrusion attempts.[4] Such systems are not affected by switched network environments (which segment traffic flows) and is effective in environments where network packets are encrypted (thwarting usual traffic analysis techniques). However, they demand high processing power overloading the nodes' resources and may be affected by denial-of-service attacks. In face of the growing volume of network traffic and high transmission rates, software based NIDSs present performance issues, not being able to analyses all the captured packets rapidly enough. Some hardware based NIDSs offer the necessary analysis throughput but the cost of such systems is too high in relation to software based alternatives.

Current IDS technology is increasingly unable to protect the global information infrastructure due to several problems: The existence of single intruder attacks that cannot be detected based on the observations of only a single site. Normal variations in system behavior and changes in attack behavior that cause false detection and identification. Detection of attack intention and trending is needed for prevention The sheer volume of attack notifications received by ISPs and host owners can become overwhelming.

A. Why DIDS. ?

The proliferation of heterogeneous computer network has serious implication for the intrusion detection problem. Foremost among these implications is the increased opportunity for unauthorized access that is provided by the network's connectivity. This problem is exacerbated when dial-up or internetwork access is allowed, as well as when unmonitored hosts (viz. hosts without audit trails) are present. The use of distributed rather than centralized computing resources also implies reduced control over those resources. Moreover, multiple independent computers are likely to generate more audit data than a single computer, and this audit data is dispersed among the various systems. Clearly, not all of the audit data can be forwarded to single IDS for analysis; some analysis must be accomplished locally.

Distributed Intrusion Detection System (DIDS) which generalizes the target environment in order to monitor multiple hosts connected via a network as well as the network itself.

III. EXISTING SYSTEM

Intrusion Detection System(IDS) are important mechanism which play a key role in network security and self defending networks. Such systems perform automatic detection of intrusion attempts and malicious activities in a network through the analysis of traffic captures and collected data in general. Such data is aggregated, analysed and compared to

a set of rules in order to identify attack signatures, which are traffic patterns present in captured traffic or security logs that are generated by specific types of attacks. In the process of identifying attacks and malicious activities, and IDS parses large quantities of data searching for patterns which match the rules stored in its signature database. Such procedure demands high processing power and data storage access velocities in order to be executed efficiently in large networks.

The issue in designing intrusion detection systems lies in efficient and comprehensive collection of data that comprises activities in different network portions in order to obtain general overview of network security. An IDS should reach high collection throughput while collecting as much data as possible. Once enough data is gathered, it is necessary to rapidly analyzer it and determine whether any attacks or malicious activities are present, which is the main issue that impacts IDS performance. Usually attack detection requires processing collected data through pattern matching algorithms in order to determine whether any of the patterns contained in the signature database are present. Hence if the quantity of collected data is excessively massive, IDS performance is seriously affected.

IV. PROPOSED SYSTEM

To design a scalable distributed intrusion detection system in a cloud computing environment that gives the analyst a quicker, easier, more efficient method to identify attacks across multiple network segments, and to trace back the activities of the attacker.

The DIDS system gives the analyst a quicker, easier, more efficient method to identify coordinated attacks across multiple network segments, and to trace back the activities of the attacker. By having all of the attack records stored in a single place, it allows the analyst much more flexibility in discovering attack patterns, and other attack issues which may have otherwise gone unnoticed.

The broad view given by the DIDS system also allows the analyst to ensure a minimum of false positives and false negatives by being able to see beyond a single network segment, into the network as a whole.

Scalable storage in distributed file system infrastructure. Scalable distributed data processing in cloud computing environment through map reduce framework. Can be implemented by widely deployed open source software.

A. Proposed System Architecture

The proposed distributed IDS architecture is composed of mainly 3 parts.

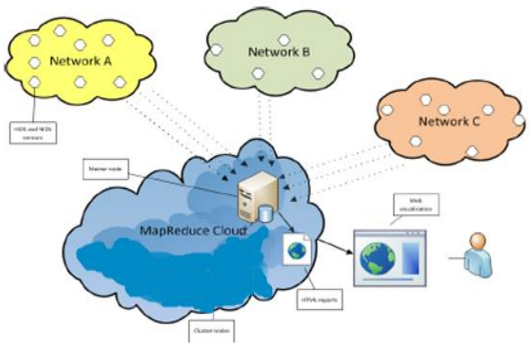

Fig 3.1 Proposed System Architecture

1) Sensor agent

In order to thoroughly captures network activity in different network segments our architecture employs several sensor agents that are placed in different network regions. This agent collects relevant information capture and generated by their host nodes and sends it to be master nodes in the cloud environment which centralizes data collection

The sensor agent also collects audit data and security laws generated by the host operating system. Correlating this information with traffic captures and regular IDS logs, the intrusion detection model and the analysis system placed in the cloud infrastructure identify and confirm attacks that generate patterns in different layers.

2) Cloud infrastructure

The cloud infrastructure is a simply a Hadoop environment that aggregate data received from the individuals sensors and process it through attack detection algorithm.

The host in the MapReduce are also part of distributed file system where the data collected by the sensor agent is stored during analysis. The cloud's master node receives the data and stores it in the distributed file system where it is accessed and modified in the analysis process.

The distributed file system seamlessly scale together with the cloud infrastructure providing enough storage space to large quantities of logs without requiring special storage device.

3) Web visualization interface

After the collected data is processed in the cloud the intrusion detection models issue alerts regarding detected ongoing malicious activities.

It is also possible to extract statistical information from the collected data, yielding results which require different visualization methods.

A. Innovation done in existing system.

The proposed architecture enjoys following characteristics Distributed data collection from multiple sources in multiple network areas Scalable storage in a distributed file system infrastructure Scalable distributed processing in cloud environments through the MapReduce framework Implementable from widely deployed open source software tool.

The proposed architecture is based on distributed data analysis through the MapReduce framework in a cloud computing environment with a distributed file system to rapidly parse collected data. It is potentially capable of detecting attack signature with very high throughput and also able to detect complex malicious activities. In order to achieve the expected data storage and processing performance. We used MapReduce framework and distributed file system.

V. SYSTEM DESIGN

In the proposed system design information flows from the data collection sensors agent installed in different nodes to the central mapreduce cloud instead of being processed by a centralized system the collected data is then analyzed by cloud application that leverages the resources of the worker node in the cloud scaling transparently as network traffic grows and consequently analyze the data set.

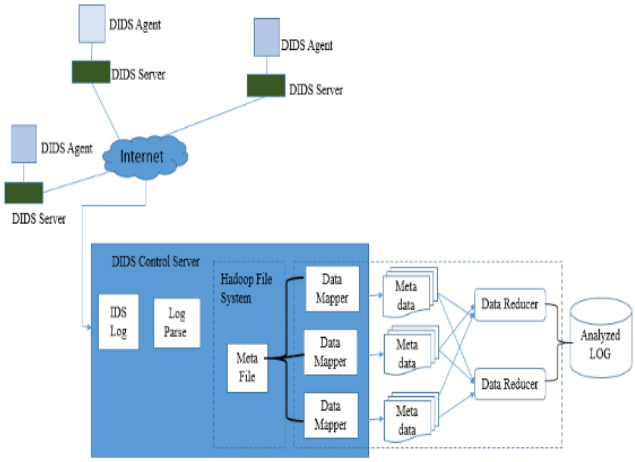

Fig 4.1 Proposed System Design

Several intrusion detection algorithm data analysis and event correlation models are intended to run as mapreduce job on the cloud infrastructure.

After the collected data processed in the cloud, the intrusion detection model issue alert regarding detected ongoing malicious activity.

The information obtained from analysis dataset i.e intrusion alert if then conveniently displayed for visualization and accordingly system administrator take decision whether to block user or not.

V. CONCLUSION

Current intrusion detection systems do not properly handle the sheer amount of traffic and data transmitted in large scale networks. We propose an efficient and scalable distributed intrusion detection system based on the MapReduce framework which is capable of handling large volumes of logs and seamlessly scale to handle network growth as well as efficiently detecting internal and external attacks which occur in isolated network region.

REFERENCES

[1] Royce Robbins, *"Distributed Intrusion Detection Systems:An: Introduction and Review"* SANSInstitute2003.

[2] Iti Raghav, Shashi Chhikara, Nitasha Hasteer," *Intrusion Detection and Prevention in Cloud Environment: A Systematic Review*", International Journal of Computer Applications (09758887),24 April2013.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, *"A novel ultrathin elevated channel low-temperature poly-Si TFT,"* IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[4] Marcelo D. Holtz, Bernardo M. David & Rafael Timoteode souse Junior, *"Building Scalable Distributed Intrusion Detection System Base on theMap Reduce Framework",* vol.1302 December 2011.R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital- to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[5] Manish Kumar*,"Distributed Intrusion Detection System Scalability Enhancement using Cloud Computing* "GESJ; Computer science and Telecommunication2014|No.1(41)"

[6] Opinder Singh & Dr. Jatinder Singh *"Comparative study of various Distributed Intrusion Detection Systems for WLAN*",Global Journal Of Research in Engineering Volume XII ,May2012.

