RESEARCH ARTICLE OPEN ACCESS

Using K-Means and FCM In Cloud Environment

Mr.Saurabh P.Taley, Prof.J.J.Shah
CSE(ESC-IV SEM) G.H.Raisoni College of EngineeringNagpur, India taleysaurabh@gmail.com
Assistant Prof IT DEPARTMENT G.H.Raisoni College of Engineering Nagpur, India
shah.jagruti13@gmail.com

ABSTRACT
Cloud computing provides large scale computing resource to each customers. Cloud systems can be threatened by numerous attacks as cloud provides services to no trustworthy system. Cloud needs to contain intrusion detection system for protecting system against threads. If IDS is having stronger security using more rules or patterns then it need much more computing resources. Current cloud monitoring systems can rely on signature-based and supervised-learning-based detection methods to check out attacks and anomalies. Propose work introduce UCAD, an Unsupervised Cloud Anomaly Detection for knowledge-independent detection of anomalous traffic. UCAD uses a novel clustering technique based on Sub-Space-Density clustering to identify clusters and outliers in multiple low-dimensional spaces. The evidence of traffic structure provided by these multiple clustering is then combined to produce an abnormality ranking of traffic flows, using a correlation-distance-based approach.

Keywords- *Intrusion detection, Anomaly detection, Cloud)*

I. INTRODUCTION

Cloud computing has evolved through a number of implementations. Moving data into the cloud provides great convenience to users. Cloud computing is a collection of all resources to enable resource sharing in terms of scalable infrastructures, middleware and application development platforms, and value-added business applications . The characteristics of cloud computing includes: virtual, scalable, efficient, and flexible. In cloud computing, three kinds of services are provided: Software as a Service (SaaS) systems, Infrastructure as a Service (IaaS) providers, and Platform as a Service (PaaS). In SaaS, systems offer complete online applications that can be directly executed by their users. In IaaS, providers allow their customers to have access to entire virtual machines; and in SaaS, it offers development and deployment tools, languages and APIs used to build, deploy and run applications in the cloud. The virtual environment lets users use computing power which far exceeds that contained in their physical worlds. These services in cloud computing may easily expose to the risk of security attacks. Within the cloud computing, security issues, such as confidentiality, integrity and availability (CIA) are the most important security considerations. Denial-of-service (DoS) attack and distributed denial-of-attack (DDoS) are other kinds of attacks that cause the targeted system or network unusable [2]. Therefore, if the cloud computing framework suffers from these kinds of attacks, the service providers and users could not use the services. Intrusion detection system (IDS) is a practical solution to resist these kinds of attacks. However, if IDS is deployed in each cloud computing region, but without any cooperation and communication, IDS may easily suffers from single point of failure attack. Obviously, the abilities of intrusion detection and response are decreased significantly. Thus, the cloud environment could not support services continually. In order to protect the cloud environment from DoS or DDoS attacks, the proposed paper launches an idea of federation defense in the cloud computing. Based on this concept, IDS system is deployed in each cloud computing region. These IDSs will cooperate with each other by exchanging alerts to reduce the impact of the DoS attack. Within this framework, Snort based IDS is implemented and three modules are plug-in into the system. These modules are block, communication and cooperation modules. Clustering is a process of labeling data and assigning that data into groups of similar objects[2]. Each group is called as cluster. It consists of members from the same cluster that are similar and members from the different clusters that are different from each other.

II. LITERATURE SURVEY

Two different approaches are by far dominant in current research literature and commercial detection systems: signature-based detection and anomaly detection. Signature-based detection is the de-facto approach used in standard security devices such as IDSs, IPSs[1], and firewalls.

When an attack is discovered, generally after its occurrence during a diagnosis phase, the associated anomalous traffic pattern is coded as a signature by human experts, which is then used to detect a new occurrence of the same attack. Signature-based detection methods are highly effective to detect those attacks which they are programmed to alert on. However, they cannot defend against new attacks, simply because they cannot recognize what they do not know. In addition, building new signatures is a resources-consuming task, as it involves manual traffic inspection by human experts. On the other hand, anomaly detection uses labelled data to build normal- operation-traffic profiles, detecting anomalies as activities that deviate from this baseline. Such methods can detect new kinds of network attacks not seen before. Nevertheless, anomaly detection requires training for profiling, which is time- consuming and depends on the availability of purely anomaly-free traffic data sets. Labelling traffic as anomaly-free is not only time consuming and expensive, but also prone to errors in the practice, since it is difficult to guarantee that no anomalies are buried inside the collected data. In addition, it is not easy to keep an accurate and up-to-date normal-operation profile. Our thesis is that these two knowledge-based approaches are not sufficient to tackle the anomaly detection problem, and that a holistic solution should also include knowledge-independent analysis techniques. To this aim we propose UCADA, an Unsupervised Cloud Anomaly Detection Algorithm that detects cloud traffic anomalies without relying on signatures, training, or labelled traffic of any kind. Based on the observant traffic anomalies are, by definition, sparse events that deviate markedly from the majority of the traffic, UCADA relies on robust clustering algorithms to detect outlying traffic flows.

III. PROPOSED SYSTEM PLANNING AND DESIGN.

The proposed system will be identified to provide a solution to the problem of anomaly detection which is completely Knowledge Independent. In the Knowledge Independent Unsupervised Detection Of Cloud Attack. We evaluate the ability of UCADA to discover node controller, storage controller and walrus controller attacks in real traffic without relying on signatures, learning, or labeled traffic. Additionally, we compare its performance against previous unsupervised detection methods using traffic from different node controller.

A. System Design

In the system design input data at first contain the data packets. A data set is an ordered sequence of object, this may contain anomaly and we have to detect anomalies in the data set. To detect those anomalies in the huge dataset we have to apply robust clustering approach which will create automatic signature. In our proposed work we are going to implement completely blind approach so for that no any previous knowledge about the anomaly and to detect such types of blind attack will apply robust clustering approach for the detection of cloud anomaly in an completely unsupervised fashion

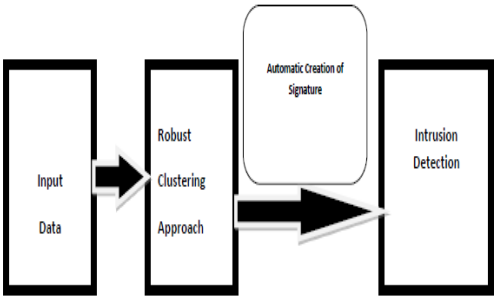

Fig 1: Organization of the system

B. Architecture of the Unsupervised Detection of Attacks

The unsupervised detection stage takes as input all the IP flows in the anomalous time slot, aggregated according to one of the different aggregation levels used in the first stage. Let Y = {y1,y2….. yn} be the set of n flows in the flagged time slot. Each flow yi ∈ Y is described by a set of m traffic attributes or features on which the analysis is performed. The selection of these features is a key issue to any anomaly detection algorithm, and it becomes critical in the case of unsupervised detection, because there is no additional information to select the most relevant set. In this we shall limit our study to detect and characterize well-known attacks, using a set of standard traffic features widely used in the literature. However, the reader should note that the approach can be easily extended to detect other types of attacks, considering different sets of traffic features. In fact, more features can be added to any standard list to improve detection and characterization results. The set that we shall use here includes the following m = 9 traffic features: number of source/destination IP addresses and ports, ratio of number of sources to number of destinations, packet rate, ratio of packets to number of destinations, and fraction of ICMP and SYN packets. According to previous work on signature-based anomaly characterization, such simple traffic descriptors permit to describe standard network attacks such as DoS, DDoS, scans, and

spreading worms/virus. The algorithm is based on clustering techniques applied to data set. The objective of clustering is to partition a set of unlabelled elements into homogeneous groups of similar characteristics, based on some measure of similarity. Our goal is to identify in the different aggregated flows that may compose the attack. For doing so, the reader should not necessary to have that an attack may consist of either outliers (i.e., single isolated flows) or compact small size clusters, depending on the aggregation level of flows in Y.

Our approach falls within the unsupervised anomaly detection domain. The vast majority of the unsupervised detection schemes proposed in the literature are based on clustering and outliers detection. The unsupervised algorithm to detect and to automatically construct a signature for different attacks my analysis will be limited to show how the unsupervised approach can detect and characterize different network attacks without using signatures, labels, or learning.

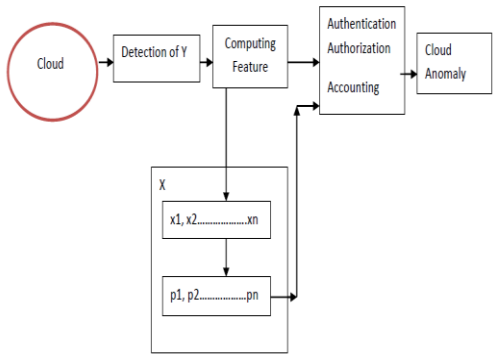

Fig 2: Architecture of the Unsupervised Detection Of cloud Attacks

C. Steps for the Unsupervised Detection of Network Attacks

Propose system UCAD, an Unsupervised Cloud Anomaly Detection that detects cloud anomalies without relying on signatures, training, or labeled traffic of any kind. Based on the observation cloud anomalies are sparse events that deviate marked from the majority of the traffic, UCADA relies on robust clustering algorithms to detect outlying traffic flows.UCAD runs in three consecutive steps, analysing packets captured in contiguous time slots of fixed length. Figure 2 depicts a modular, high-level description of UCAD[9].

The first step consists in detecting an anomalous time slot in which the clustering analysis will be performed. For doing so, captured packets are first aggregated into multi-resolution traffic flows. Different time-series are then built on top of these flows, and any generic change-detection algorithm based on time-series analysis is finally used to flag an anomalous change.

The second step takes as input all the flows in the time slot flagged as anomalous. In this step clustering algorithm are used and evidence of traffic structure provided by this clustering algorithm are used to rank the degree of abnormality of all the identified outlying flows, building an outliers ranking.[11][12].

In the third and final step, the top-ranked outlying flows are flagged as anomalies, using a simple thresh holding detection approach. As we will show throughout the paper, the main contribution provided by UCAD relies on its ability to work in a completely unsupervised fashion, outperforming previous proposals for unsupervised anomaly detection.

IV. EVALUATION AND CONCLUSION

Proposed system uses K- means and FCM based clustering technics. We have created a gui in Matlab to analyse time taken by data set by different clustering technic. Time taken by fcm is lesser as compared to K-means.

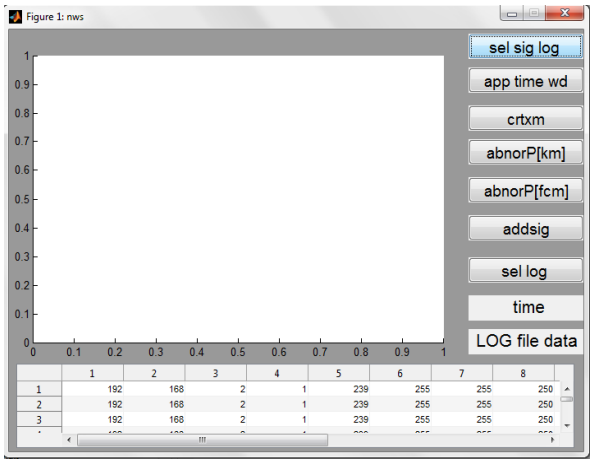

Fig 3: GUI for analyzing data set.

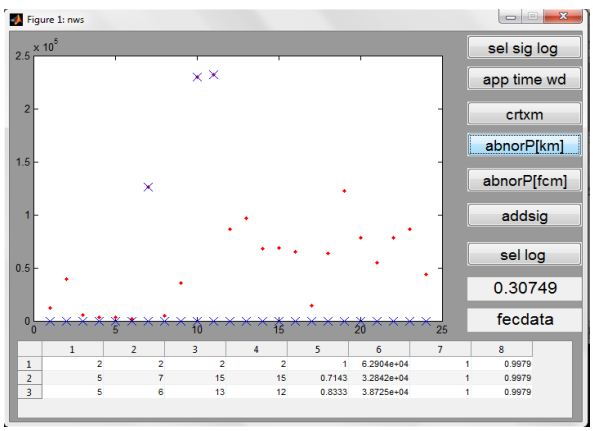

Fig 4: GUI of analyzing data set using K means.

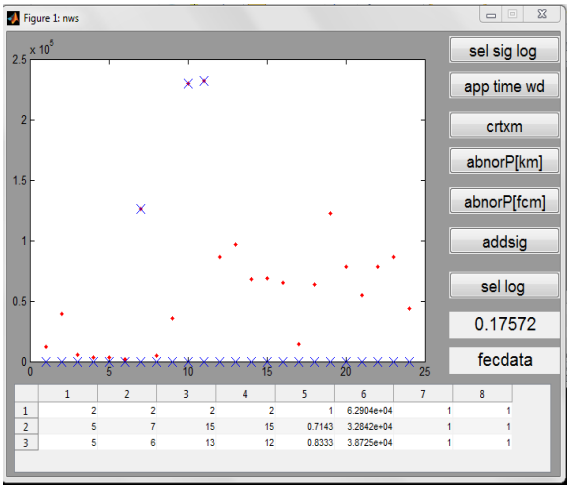

Fig 5: GUI of analyzing data set using FCM

In the proposed work, system will load the log file, the loaded log file must be differentiated according to time frame. Depending upon the time frame system will apply clustering technics using K-means and fcm to detect abnormal activities. In this paper, the cloud computing security issues have been investigated with special emphasis on governance of the security and compliance from the perspective of user companies as well as cloud service providers. The Unsupervised Cloud Anomaly Detection that we have many interesting advantages. It uses exclusively unlabelled data to detect traffic anomalies, without assuming any particular model or any canonical data distribution, and without using signatures of anomalies or training. Despite using ordinary clustering techniques to identify anomalies, UCAD have the robustness of general clustering approaches.

REFERENCES

[1] Zhen Chen*, Fuye Han, Junwei Cao, Xin Jiang, and Shuo Chen "Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System" TSINGHUA SCIENCE AND TECHNOLOGY ,2013.

[2] Kapil Wankhade, Sadia Patka, Ravindra Thool " An Efficient Approach for Intrusion Detection Using Data Mining Methods" 2013 IEEE

[3] En Niari Saad, Khalil El Mahdi, Mostapha Zbakh "Cloud Computing Architectures Based IDS" 2012 IEEE

[4] Amira Bradai and Hossam Afifi "Enforcing Trust-based Intrusion Detection in Cloud Computing Using Algebraic Methods" 2012

[5] Mld Khoudali, Karim Benzidane and Abderrahim Sekkaki "Inter-VM packet inspection in Cloud Computing " The 5th International Conference on Communications, Computers and Appllication 2012

[6] Massimo Ficc o, Massimiliano Rak, and Beniamino Di Martino "An Intrusion Detection Framework for Supporting SLA Assessment in Cloud Computing" 2 0 12 I EEE

[7] Matthias Gander, Basel Katt, Michael Felderer, Adrian Tolbaru, Ruth Breu, Alessandro Moschitti "Anomaly Detection in the Cloud: Detecting Security Incidents via Machine Learning" 2012

[8] Hisham A.Kholidy,Fabrizio Baiardi "DCDIDP:A Distributed, Collaborative, and Data-driven Intrusion Detection and Prevention Framework for Cloud Computing Environments" 2012 IEEE

[9] Pedro Casas, Johan Mazel and Philippe Owezarski "UNADA: Unsupervised Network Anomaly Detection using Sub-Space Outliers Ranking" 2011

[10] Jun-Ho Lee,Min-Woo Park,Jung-Ho Eom, and Tai-Myoung Chung"Multi-level Intrusion Detection System and Log Management in Cloud Computing" 2011

[11] Jie Yang, Yong Ge, Hui Xiong, Yingying Chen, Hongbo Liu "Performing Joint Learning for Passive Intrusion Detection in Pervasive Wireless Environments " 2010.

[12] Manzoor Elahi,Kun Li,Wasif Nisar,Xinjie Lv,Hongan Wang "Detection of Local Outlier Over Dynamic Data Streams using Efficient Partitioning Method" 2009

[13] Baoyi Wang, Ranran Jin, Shaomin Zhang, Xiaomin Zhao "Research on Gravity-based Anomaly Intrusion Detection Algorithm" 2009

[14] HU Liang, REN Wei-wu, REN Fei "Anomaly Detection using Improved Hierarchy Clustering" 2009

[15] Yu-Ping Zhou, Jian-An Fang , Yu-Ping Zhou "Research on Neuro-Fuzzy Inference System in Hierarchical Intrusion Detection"2009