IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Utilizing Machine Learning for Intrusion Detection Systems in the Context of Cloud Computing

Mohammed Akif, Anuj Nakka, Solanki Prakash, Vijay N, Dr Suresh Kallam
Department of CSE
Jain University, Banglore, India

Abstract: *The rapid advancements in the internet and communication sectors have led to a huge expansion in the network dimension and the related data. The proliferation of new assault kinds as a result makes it challenging for community safety to accurately identify invasions. Furthermore, the presence of intruders who want to launch a lot of attacks against the network cannot be ignored. An intrusion detection device (IDS) is one such instrument that protects against prospective intrusions by monitoring community communication to guarantee its confidentiality, integrity, and availability. IDS continues to struggle with innovative intrusion detection, lowering false alarm rates, and improving detection accuracy despite substantial research backing. An intrusion detection system's main job is to protect resources against threats. It predicts user behaviour based on analysis, and determines whether that behaviour constitutes an assault or is simply normal behaviour. We use Support Vector Machine (SVM) and Rough Set Theory (RST) to detect network breaches (SVM). Computer learning (ML) and Python-based total IDS systems are recently being employed to identify intrusions across the network in an ecologically friendly manner. The taxonomy presented in this article is mostly based on the impressive machine learning and Python approaches used to develop network-based IDS (NIDS) systems. The article begins by defining IDS.*

Keywords*:* Machine learning Techniques, Cloud Computing infrastructure, Node:-Physical virtual device Datasets:-Consists of data, Protocol - A standard way of performing a task

I. INTRODUCTION

IDSs (intrusion detection systems) are physical or software programmes that automate the process of monitoring and closely examining the activities that take place on a laptop network in order to identify malicious behaviour. Because of the sharp increase in the severity of assaults in the community, intrusion detection systems have become an essential part of the security infrastructure for the majority of enterprises. Businesses can safeguard their systems from threats brought on by increasing community connectedness and reliance on information technologies by using intrusion detection. Given the scale and nature of current network protection threats, the decision for security professionals should not be whether to use intrusion detection but rather which intrusion detection characteristics and capabilities can be deployed.

Intrusions happen as a result of attackers gaining access to the systems. Users who are authorised who misuse the privileges that have been granted to them, such as users who try to get access for which they are no longer authorised.

Intrusion detection systems (IDS) employ either a host-based method or a network-based approach to detect and prevent attacks. These things appear to be attack signatures in either case—detailed patterns that often point to malevolent or suspect intent. When an IDS looks for certain patterns in network traffic, it is fully network-based. An IDS that looks for attack signatures in log files is called host-based. There is no heuristic to verify the accuracy of the algorithms' conclusions, despite the fact that numerous have been developed to detect different forms of community invasions.

Without a clear performance measurement, the precise effectiveness of a community intrusion detection system's ability to identify malicious sources cannot be reported.

An intrusion detection system, or IDS, is a hardware device or software program that continuously monitors, detects attacks or intrusions, and notifies the computer or network. The administrator or user can use this alert file to find the machine's or network's vulnerability and learn more about it. Anomaly-based detection, Signature-based detection, and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Hybrid-based detection are a few popular approaches to intrusion detection. As a result of the approach's modeling of user, network, and system behavior, anomaly-based intrusion detection is also known as behavior-based detection.
The precise efficiency of a community intrusion detection system's capacity to recognise malicious sources cannot be communicated in the absence of a clear performance assessment.
A hardware or software system known as an intrusion detection system, or IDS, continually monitors, identifies assaults or intrusions, and alerts the computer or network. This alert file may be used by the administrator or user to identify and learn more about a machine's or network's vulnerability. Some common methods for detecting intrusions include anomaly-based detection, signature-based detection, and hybrid-based detection. Anomaly-based intrusion detection is sometimes referred to as behavior-based detection since the method models user, network, and system behaviour.
This technique replicates user, network, and host system behaviour, so if a behaviour deviates from the usual, it can raise an alarm or alert the administrator. The signature-based IDSs sometimes go by the moniker of knowledge-based detection techniques.
This technique is based on a database that has previously accepted attack signatures and known system vulnerabilities. A hybrid based detection machine combines anomaly-based intrusion detection and signature-based intrusion detection. Anomaly or signature intrusion detection approaches are used by the majority of IDSs. While each intrusion detection system has drawbacks of its own, hybrid IDS can be utilised. Based on their behaviour, intrusion detection is classified into two types, namely:

- **Active IDS**: Active IDS function similarly to passive IDS and additionally operate to throttle suspicious traffic in order to avoid attacks.
- **Passive IDS**: These IDS will unquestionably examine and analyse website visitors and notify the administrator of any attacks and security gaps.

Instead of being explicitly coded, machine learning (ML) can improve the way computers learn from their experiences. In this paper, the performance of many machine learning (ML) techniques, including Modified K-Means, J.48, Support Vector Machine (SVM), selection tables, PCA, logistic regression, decision trees, and artificial neural networks (ANN) for IDS, is compared. In order to categorise the intrusion detection, this study effort uses techniques like Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), and Random Forest (RF).

1.1 Objective

- To research security-related concerns with cloud computing: Security for computer systems hasn't always been easy. Cloud computing and cloud service providers need to solve the following challenges that have an impact on cloud security: - A breakdown in governance, an insider threat, a compromised administration interface, an inability to completely or securely delete data, data eavesdropping, etc.
- We provide two methods for recognising network assaults in order to evaluate machine learning algorithms for intrusion detection. By utilising feature engineering in conjunction with a tree-based ensemble learning strategy, we show that we can surpass cutting-edge findings in the field. We propose a novel approach for IDS where the performance of the detector may be enhanced while preserving its cheap running cost by merging a limited selection of coaching recordings with certain weak categorization methods.
- To reduce training time, increase accuracy, reduce false alarms, and discover a suitable machine learning approach for intrusion detection: In this study, we simplify the enormous network dataset by selecting only the most relevant and impactful parts in order to improve the IDS performance and accuracy. The use of fewer datasets attempts to speed up the process of training SVM computer systems to recognise assaults. A prototype IDS outfitted with desktop studying models was created in order to improve the efficiency of detecting DoS and R2L assaults.
- Using machine learning methods, create and deploy an intrusion detection system framework on a cloud platform. It is common knowledge that the cloud computing movement offers resources that have been virtualized. 0 There are many different devices that cloud computing users use, such as computers and smartphones. Customers can access cloud services in three different ways: Infrastructure as a Service (laaS),

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Platform as a Service (PaaS), and Software as a Service (SaaS). Users can command all digital machines using these unusual approaches. Clients can use apps that are currently online or create new ones in the cloud.

- To assess an intrusion detection system's performance: One of the various classification metrics for IDS that may be used to judge how well it is working is the confusion matrix for a two-class classifier. Each column of the matrix represents an instance in a predicted class, and each row of the matrix represents an instance in an actual class.

TPR: True Positive Rate The ratio of attacks that were accurately predicted to all assaults is used to construct this statistic. The TPR for an IDS is 1, which is exceedingly rare if every incursion is discovered.

False Positive Rate (FPR): It is calculated by dividing the total number of incidentals by the share of incidentals that are incorrectly categorised as assaults.

False Negative Rate (FNR): A false negative happens when a detector misclassifies an abnormality as normal because it was unable to detect it.

The accuracy or classification rate (CR) measures how accurately the IDS recognises ordinary or atypical traffic behaviour. It is described as the proportion of all those cases where the prediction was accurate.

II. LITERATURE SURVEY (EXISTING SYSTEM)

- Snort: The tool we'll employ to find intrusions in the system that has been proposed is called Snort. As previously indicated, Snort is an open source programme that is free to use and is owned by Cisco Systems. The operating systems Linux, UNIX, and Windows all support the installation of this programme. Any network intrusions are found using the Snort tool. Snort has three modes of operation: sniffer, packet logger, and intrusion detection. In the intrusion detection mode, Snort may monitor network traffic against a preset rule set and take appropriate action if anything suspicious is found. Users of Snort can create their own rules in files.
- OSSEC: Open Source HIDS SECurity is a term used to describe OSSEC. It facilitates the process of log analysis. Moreover, OSSEC conducts active response, real-time alerting, rootkit identification, integrity checking, and alerting.
- Prelude SIEM: Hybrid intrusion detection system called Prelude SIEM. When an intrusion or other security danger takes place on a network in real time, Prelude SIEM provides warnings.
- Suricata:- An intrusion detection and prevention system is called Suricata. It serves as an all-inclusive mechanism for monitoring network security. Suricata has the benefit of Snort in that it extends all the way to the application layer.
- AIDE: Another free intrusion detection solution is AIDE, or Advanced Intrusion Detection Environment. Its primary emphasis is on file signature comparisons and rootkit identification. It employs anomaly-based analysis that is executed on demand as well as signature-based analysis. For UNIX and LINUX, it is a host-based intrusion detection system.
- Zeek: Zeek used to go by the name Bro. It is network intrusion detection software that is open source and free. We can also do in-the-moment network event analysis with Zeek's assistance.

III. PROPOSED MODEL

System architecture

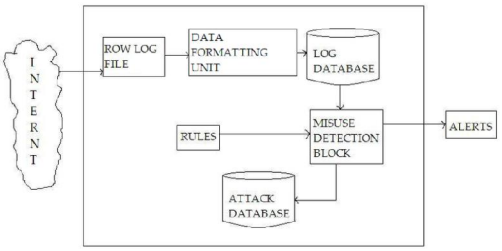


Fig 1.1 IDS for detecting attacks (Architecture)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

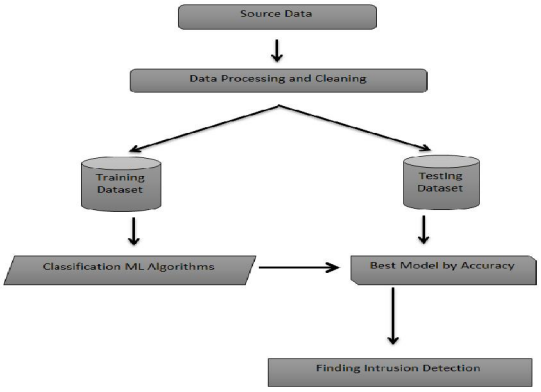


Fig 1.2 Splitting of Data using ML

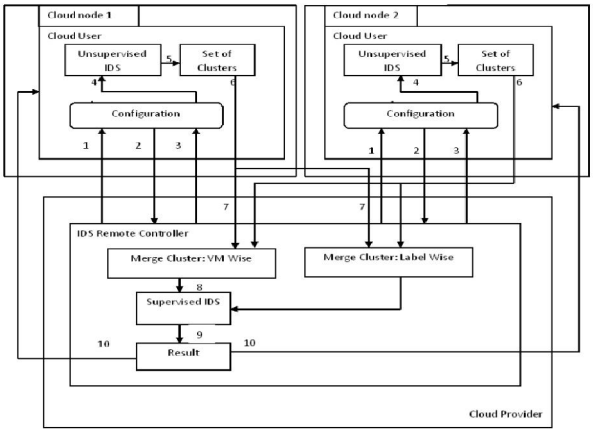


Fig 1.3 Cloud node setup in Vmware

Label for Number:-
1: Launch IDS
2: Responding with configuration
3: Start IDS
4: Execute IDS unsupervised
5: Create Clusters
6: Send Cluster to IDS Remote Controller
7: 1) Integrate Cluster data VM-by-VM 2) Integrate Cluster data label VM-by-VM
8: Conduct Supervised IDS
9: Consider
10: Issue a VM user alert.

IV. METHODOLOGY

We used a quantitative and exploratory method for this lookup study. Throughout the literature assessment, we studied a large number of research papers, patents, theses, and annual reports of market leaders including Cisco and Juniper. We also found IDS findings and identified which IDSs are open-source. In this first phase of the literature study, we learned that snigger IDS and Bro IDS have been researched despite being used for malicious attacks. Furthermore, we discovered studies on IDS desktop learning techniques for detecting unidentified attacks.

As a result, we used to commonly concentrate on IDS that made use of machine mastering approaches in our second part of the literature study..

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

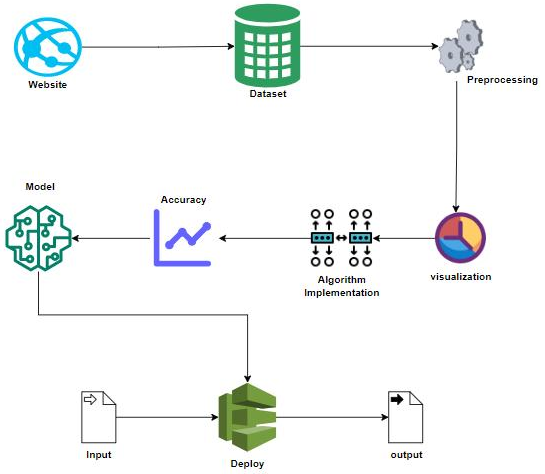
For the incursion dataset, we tested a number of machine learning methods, and since we wanted to execute it on the cloud platform, we utilised KDD to find the optimal ML method. So, we made the decision to cluster the dataset and provide each cluster its own virtual machine (VM) on the cloud. Two stages of this systematic review were completed.
The information source (search engine) and search phrases to employ are determined in Phase 1 in order to conduct a query and obtain an initial list of articles.
Phase 2 selects the most significant and crucial items using the original list's positive criteria and retains them in
The primary objective of this assessment article is to respond to the following questions:
What are the latest ML-based NIDS design trends?
What are the most current ML techniques used to the NIDS design?
What benefits and drawbacks come with each approach that has been used?
What datasets are being examined for the ML-based NIDS right now?
What performance indicators are utilised for evaluation the most frequently?
What will ML-driven NIDS research look like in the future?

Sr no	Attack Types	Attack Specific names
1	Denial of Service Attacks	Back,Land,Neptune,Pod,Smurf,Teardrop
2	Attack by the user	Bufferoverflow,loadmodule,perl,rootkit
3	Remote user as local users	FTP write,guess,Passwd,multihop,phf,spy
4	Probes	Satan,IPsweep,nmap,portsweep

Table no.1:-Types of attacks and their details

V. IMPLEMENTATION

Workflow



Packet Collection: This module receives the information as an IDS entry. The information will be processed and saved on a server. IDS gathers information in host-based IDS, such as disc utilisation and device processes, and alters data packets on the network-based system.

Packet Decoder

A PPP connection, an Ethernet (copper or fiber-optic) connection, or anything else that can decode packets. The Packet Decoder's job is to decode information like source, destination, length, protocol, method, status code, etc. in order to prepare packets for preprocessors.

Analysis and Classification

- Datasets:-This data set is applied to different algorithms for machine learning.
- Preprocessing: The transformation of raw data formats into useable and effective formats that will be sent to the models for training and testing is referred to as preprocessing of data. The lambda function is used to map

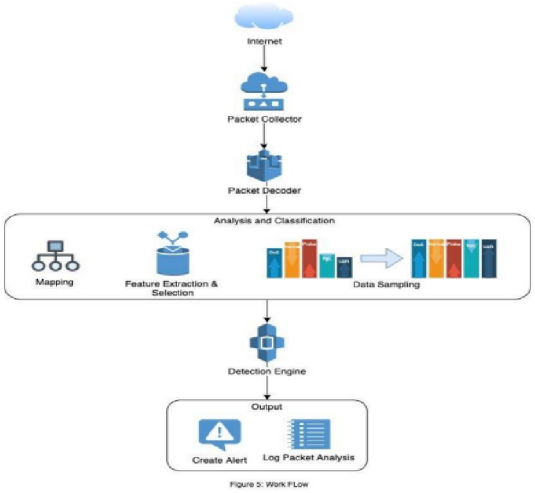
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

the data entries in the data set with the distinct attack classes DoS, Probe, U2R, and R2L. This is the main step of the preprocessing procedure.

- Feature Selection: Initially, it is highly challenging to choose the feature from the dataset that would indicate how significant a feature is. The feature option varies as the assault type changes. Second, no tagged real-time networking traffic exists. By choosing the complete issue from the sub-set of features, any superfluous data may be removed. We then adapted the data framework to the test and training datasets. We design two target classes: a standard target class and an attack target class. The data set is split into two class attack labels using a function.
- Data Sampling: This method makes use of randomization to guarantee that each component of the population has an equal chance of taking part in the chosen sample. As an alternative, random sampling is used. We get a sampled data set where the number of elements for each data set are equal since the values of the original data sample will comprise various values for each of the four attacks in addition to the normal values.
- Training and Testing Models: The cross-validation data is used to make sure the method used to train the machine is more precise and effective, and the test data is used to examine how well the machine can predict new responses based on its training. Training data are used to make sure the system detects patterns in the data.
- Support Vector Machine: Support-vector machines (SVMs, also known as support-vector networks) in machine learning are supervised learning models with corresponding learning algorithms that examine data used for regression and classification analysis.
- Naive Bayes Classifier: Naive Bayes classifiers are a family of fundamental "probabilistic classifiers" in machine learning that interpret the Bayes theorem with explicit (naive) assumptions of feature independence.
- Decision Tree:A decision tree is a flowchart-like structure in which each leaf node represents a grade and each inside node represents a "test" on an attribute (such as whether a coin will land heads or tails) (decision made after all attributes have been computed).
- Random Forest Classifier: As its name implies, a random forest is a depiction of several individual decision trees working together as an ensemble. Every tree in the random forest produces a class prediction, and our model predicts the class that receives the most votes.
- k-nearest neighbors: Instance-based learning, also known as lazy learning, is a method in which all computation is postponed until after the function has been evaluated and only local approximations of the function are made.
- Logistic Regression: In its simplest form, logistic regression is a statistical model that utilises a logistic function to simulate a binary dependent variable; however, there are a number of more intricate expansions. In regression analysis, the logistic regression (or logit regression) estimates the parameters of a logistic model, a form of binary regression.Detection Engine: In the actual world, the "Detection Engine" analyses packet patterns and content to determine whether an assault is taking place. A suitable alert message is sent to the security administrator if a packet ends up matching any of the rules.
- Alert: This explains the response and assault of the system. This might deactivate the system by tossing packets so that they do not reach or close the ports, or it could alert the system administrator by email or an alarm icon that the administrator is in charge of all required data. Alerts are recorded in "Log Packet Analysis," which transmits the data for recording in a lof-file format, like tcpdumpand.pcap files. These log files can be monitored by network administrators for additional examination.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

IDS framework in Cloud



1) The cloud provider is home to the IDS controller. It starts the IDS module provider, sends the IDS need message to all active cloud user VMs, and then waits for the delay period to see whether the cloud user VM responds favourably or unfavourably.
2) In order for IDS to function, active cloud user VMs must receive an IDS need message and verify that they have the necessary physical assets. If a Cloud User VM needs IDS to function on their computer and has no physical requirements, he will send a powerful response to the IDS Remote Controller. If Cloud User VM no longer wants to operate IDS or lacks physical resources, he will respond poorly to the IDS Remote Controller.
3) Using the positive feedback from the cloud user VMs, the IDS remote controller saves the data, keeps track of the cloud users using the IDS, launches the IDS occasion to them, and waits for the timeout period to acquire clusters. IDS remote Controller uses an unsupervised desktop learning strategy and generates a good set of clusters if none of the cloud user VMs react with clusters.
4) The cloud user VM receives the IDS instance, uses unsupervised machine learning to forecast intrusion attempts, and delivers a suitable set of cluster information to the IDS remote controller. If a Cloud User VM fails to transmit cluster information within the allotted period, IDS Remote Controller will destroy his data and reduce the counter.
5) IDS Remote Controller mixes the k sets of clusters it receives based on the Cloud Consumer VM Smart and Labeled Cluster-Wise. IDS Remote Controller will merge information from 1) Cluster 0 and Cluster 1 of VM1, and Cluster zero and Cluster 1 of VM2 (Merger Cluster - VM wise), as well as 2) merger data of Cluster zero of Cloud User VM1 and facts of Cluster zero of Cloud User VM2, and facts of Cluster 1 of Cloud User VM1 and records of Cluster 1 of Cloud User VM2.
6) To categorise records as intrusion data or routine data, IDS Remote Controller uses supervised desktop mastering.
7) IDS Remote Controller notifies all Cloud User VMs of the IDS module based on the findings.
IDE- Anaconda ,Jupyter Notebook and Open Nebula
Programming language- Python
Design & Prototyping- Open Nebula
Cloud Platform –Vmware and Open Nebula
Version & Source control- GIT, GitHub

VI. CONCLUSION

An intrusion detection system's main objective is to reduce the number of false positives while identifying assaults and malicious activity that occurs within a network. The IDS's output would be accurate, complex, and reliable because machine learning methods were being used. This system also shows the accuracy rate of assaults identified by the different machine learning algorithms that have been employed. There is a large quantity of data that has to be processed and securely kept for users as a result of the continual increase of technology use. Security is an important

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

factor for each user. We can be sure that user privacy is preserved if a system is secure. A system is more dependable the more secure it is.If an intrusion detection system is established and able to provide adequate protection for user data, then we can claim that it is good.

REFERENCES

[1]. D. Anderson, T.F. Lunt, H. Javitz, A. Tamaru, and A. Valdes (1995) Detecting unusual program behavior using the statistical component of Next-generation Intrusion Detection Expert System (NIDES) SRI International Computer Science Laboratory

[2]. Christopher Kruegel, Giovanni Vigna (2003) Anomaly detection of web-based attacks, Proceedings of 10th ACM conference on Computer and communication security (Washington D.C., USA), ACM Press pp 251–261

[3]. HS Teng, K. Chen, SC Lu (1990) Adaptive real-time anomaly detection using inductively generated sequential patterns, Proceedings of Symposium on Research in Security and Privacy (Oakland, CA) pp 278–284

[4]. Wang, Ke, Gabriela Cretu, Salvatore J. Stolfo (2006) Anomalous payload-based worm detection and signature generation, Recent Advances in Intrusion Detection Springer Berlin Heidelberg Krugel, Christopher, Thomas Toth (2002) Flexible, mobile agent based intrusion detection for dynamic networks European Wireless

[5]. Mark Crosbie and Gene Spafford (1995) Active defense of a computer system using autonomous agents, Technical Report 95-008, COAST Group Department of Computer Sciences Purdue University West Lafayette IN 47907-1398

[6]. Wayne Jansen, Wayne Jansen, Timothy Grance, Rebecca M. Blank ((2011) NIST- Guidelines on Security and Privacy [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary [Accessed in September 2012]

[7]. Oktaya, U. And Sahingoz (2013) „Attacks types and intrusion detection systems in cloud computing", Proceedings of 6th International Information Security & Cryptography Conference (ISC) pp 71-76

[8]. Zarrabi, A. and Zarrabi, A. (2012) Internet intrusion detection system service in a cloud, International Journal of Computer Science Issues Volume 9 pp 694-814

[9]. PeeyushMathur, Nikhil Nishchal (2010) Cloud Computing: New challenge to the entire computer industry, 1 st International Conference on Parallel, Distributed and Grid Computing pp 223-228

[10]. LoubnaDali,AhmedBentajer,ElmoutaoukkilAbdelmajid, Karim Abouelmehdi, HodaElsayed, EladnaniFatiha, BenihssaneAbderahima (2015) A survey of intrusion detection system ,IEEE 2 nd World Symposium on Web Applications and Networking (WSWAN)

[11]. Ch. Cachin and M. Schunter (2011) A Cloud You Can Trust, IEEE Spectrum 48(12) pp 28-51

[12]. Jun-jie, W. And Sen (2011) Security issues and countermeasures in cloud computing, IEEE International Conference on Grey Systems and Intelligent Services (GSIS) Nanjing pp 843-846