International Journal of Engineering, Science and Mathematics
Vol. 6Issue 8, December, 2017,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Classification of Threats in Cloud Environment

Satyabrata Dash *
Department of Information Technology, Orissa Engineering College, Bhubaneswar, India-752050
Ashok Misra **
Department of Mathematics, CUTM, Paralakhemundi, India-761211

Abstract

Cloud computing is the next generation architecture for IT Enterprises, and has proliferated itself due to the benefits it gives. It offers a virtulised environment for carrying out efficient, scalable and low cost computing. Cloud Service provider provides the computing resource on pay per usage basis, which results in reduced hardware costs for its registered users. So, monitoring the cloud environment and providing security to the users is a major challenge and needs urgent solutions. This paper outlines the possible attacks type on a cloud environment and the defense mechanism .It also describe the classification of the various attacks type, so that counter measures can be taken to provide security to the cloud environment.

Keywords:

Cloud computing;
Cloud Virtualization Security;
Malicious objects;
Intrusion, Detection;
KDD CUP

Author correspondence:

Satyabrata Dash.
Department of Information Technology,
Orissa Engineering College, Bhubaneswar,Odisha, India

1. Introduction

Cloud computing is a creative information technology based paradigm and one of the challenging improvements in the current technical environment. Most of the organisations are running their applications in the cloud due to reliability, scalability, high performance, low band width and rapid advancement in communication network. Cloud computing provides a computing platform to the registered users for deploying their computational needs in a distributed environment without the knowledge of underlined infrastructure.[1-2]

The cloud service provider (CSP) provides the services to the registered cloud users across the glove.[11] Based on the usages of data and applications cloud computing services are broadly categorized as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).The services are available to the users depending on cloud deployment (Public, Private, Community, and Hybrid Cloud) and the service level agreements (SLA) between the service providers and the users. In a Public Cloud, the resources are made available to the general public or a large user group[4]. In a Private Cloud, the resources are deployed for a single organization. In a Community Cloud, the resources are shared by more than one organization under a specific community and in Hybrid Cloud the infrastructure is a combination of more than one cloud deployment (private, community, or public).[4][7-8]

The cloud environment provides virtualized platforms to the cloud users for accessing their resources and data. Virtualization is the method of providing virtualized resources from the physical hardware.[6][7] Due to the loosely coupled environment there is a possibility of attacks that occur in the cloud based

* Doctorate Program, Linguistics Program Studies, Udayana University Denpasar, Bali-Indonesia (9 pt)
** STIMIK STIKOM-Bali, Renon, Depasar, Bali-Indonesia

applications.Therefore, it is the responsibility of the cloud service providers to increase trust into cloud environment and ensure about the security of its users and the resource nodes.A risk in the virtualized environment is an outside power by which the current cloud nodes in one state move into another. A node in the cloud surrounding provisions the information and data and provides the client a virtualized stage to utilize the application as administrations. There are critical quantities of assaults or interruptions happen in the cloud based applications.

Virtualization splits, allocates, and resizes the resources dynamically to build up the ad-hoc systems. A Virtual Machine (VM) is a dedicate software environment which runs operating systems and applications in the guest machine to help users application execution. So, VMs are logical machines having almost the same architecture as a real host machine, running an operating system in it.[12]The architecture of the virtual machine (VM) system is shown in figure 1.[6][7] According to the cloud architecture, multiple virtual machines (VMs) share the same physical machine.So virtualization technique ensures the availability of hardware and gives every application running on top of it. The details of the virtual, simulated environment are kept transparent from the application. The advantage here is the reduced cost of maintenance and reduced energy wastage which is not very surprising. So virtualization reduces the number of physical servers as a result of which one needs to maintain few servers ,this becomes much cheaper and easier.[12]

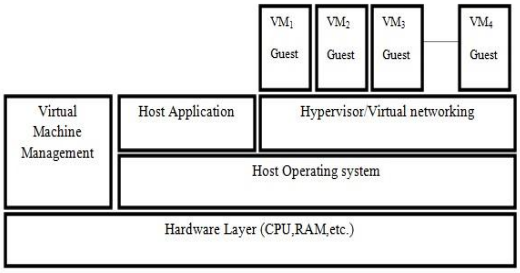


Figure 1 The Architecture of Virtual Machine

Cloud Virtualization is the method of providing the virtualised computing resources from the physical hardware to the users. So virtualization platforms were made to enhance the hardware usage by sharing and scheduling the resources among several Virtual Machines (VMs) on a single server. The cloud Service Provider is not only responsible the virtual machine that has to be protected but the user's data and application also.[11] There are several security danger and protection issues related to the infrastructure as a service, platform as a service and software as a service, which have made the cloud environment vulnerable. Therefore, a prediction method can help to fix the suspected domain to check the vvulnerabilities.[7-8] The Figure 2 shown below is an example of intra VM attack where attack in VM_1 reflects to VM_2 and VM_3.

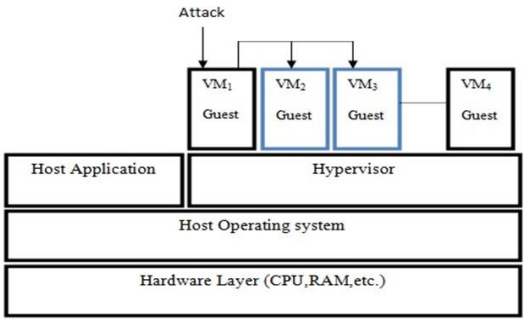


Figure 2 Intra Virtual Machine Attack

Analysing the security vulnerabilities, protecting in the cloud environment and creating productive solutions for it is really a challenging assignment for the cloud service providers. Honesty, secrecy, unwavering quality and accessibility of assets are generally utilized phrasing for security issues as a part of a distributed computing environment implies that the client's information in the cloud ought to stay private and shielded from unapproved access.So the usage of the distributed computing environment must be secure for

all its asset nodes. According to the Cloud Security Alliance Some surely understood assaults are shown in table 1

Table 1 Security Issues in SaaS,PaaS,IaaS and Cloud Data centers

Security Issues in SaaS,PaaS,IaaS and Cloud Data centers				
Level	**Service level**	**Users**	**Security Issues**	**Threats**
Application level	Software as a Service (SaS)	End client applies to a person or organization who subscribes to a service offered by a cloud provider and is accountable for its use	• Privacy in multitenant environment • Data protection from exposure (remnants) • Access control • Communication protection Software security • Service availability	• Interception • Modification of data at rest and in transit • Data interruption (deletion) • Privacy breach • Impersonation • Session hijacking • Traffic flow analysis • Exposure in network
Virtual level	Platform as a Service (PaS) & Infrastructure as a Service (IaS)	Developer–moderator applies to a person or organization that deploys software on a cloud infrastructure	• Access control • Application security • Data security, (data in transit, data at rest, eminence) • Cloud management control security • Secure images • Virtual cloud protection Communication security	• Programming flaws • Software modification • Software interruption (deletion) • Impersonation • Session hijacking • Traffic flow analysis • Exposure in network • Defacement • Connection flooding • DDOS • Impersonation • Disrupting communications
Physical level	Physical data center	Owner applies to a person or organization that owns the infrastructure upon which clouds are deployed	• Legal not abusive use of cloud computing • Hardware security • Hardware reliability • Network protection • Network resources protection	• Network attacks • Connection flooding • DDOS • Hardware interruption • Hardware theft • Hardware modification • Misuse of infrastructure • Natural disasters

2. Proposed Methodology

This paper describes the methodology for classifying intrusions based on K-means clustering algorithm using Nayver Bayes classifier. We have choosed KDD'99 cup dataset for simulation and the most of the experiments for conducting intrusion detection are performed on these datasets.[20]

3. Dataset And Normalization.

3.1 KDD'99 Dataset and Features

1. Currently, there are only few public datasets like KDD'99 and the majority of the experiments in the intrusion detection domain performed on these datasets .[17][32-33]
2. For modelling based on supervised learning methods, KDD'99 is the only available dataset which provides labels for both training and test sets. [17]
3. The study sample was created based on the 1998 DARPA intrusion detection evaluation offline dataset developed by the MIT Lincoln laboratory.
4. The KDD'99 dataset has interesting properties and is believed to present a classic challenge for the intrusion detection problem. [17][18]
5. It can be used because it is the most comprehensive dataset that is still widely used to compare, contrast and benchmarking the performance of intrusion detection models in various networks.[32]

The simulation in this KDD dataset is based upon 4 major types of attacks

1) **Denial of Service Attack (DoS)**: is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine.
2) **User to Root Attack (U2R)**: is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system.
3) **Remote to Local Attack (R2L**): occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.
4) **Probing Attack:** is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls.[20-21]

The Description of KDD 99 Intrusion Detection Dataset, classification of attack classes and characteristics are shown in Table -2,3,4 [18][19][20][22]

4. Description of KDD 99 Intrusion Detection Dataset

Table -2 Description of KDD 99 Intrusion Detection Dataset with explanation [18][19][22]

Attributes	Explanation	Behaviors
1. interval	interval of the connection	Cont.
2. protocol type	Connection protocol (e.g. tcp, udp	Disc.
3. service	Destination service (e.g. telnet, ftp)	Disc.
4. flag	Status flag of the connection	Disc.
5. source bytes	Bytes sent from source to destination	Cont.
6. destination bytes	Bytes sent from destination to source	Cont.
7. land	1 if connection is from/to the same host/port; 0 otherwise	Disc.
8. wrong fragment	number of wrong fragments	Cont.
9. urgent	number of urgent packets	Cont.
10. hot	number of "hot" indicators	Cont.
11. unsuccessful logins	number of unsuccessful logins	Cont.
12. logged in	1 if successfully logged in; 0 otherwise	Disc.
13. #compromised	number of "compromised" conditions	Cont.
14. root shell	1 if root shell is obtained; 0 otherwise	Cont.
15. su attempted	1 if "su root" command attempted; 0 otherwise	Cont.
16. # root	number of "root" accesses	Cont.
17. # file creations	number of file creation operations	Cont.
18. # shells	number of shell prompts	Cont.

19. # access files	number of operations on access control files	Cont.
20. # outbound cmds	number of outbound commands in an ftp session	Cont.
21. is hot login	1 if the login fit in to the "hot" list; 0 if not	Disc.
22. is visitor login	1 if the login is a "visitor login; 0 if not	Disc.
23. Count	number of links to the same host as the current link in the past two seconds	Cont.
24. srv count	number of links to the same service as the current link in the past two seconds	Cont.
25. serror rate	% of links that have Synchronization errors	Cont.
26. srv serror rate	% of links that have Synchronization errors	Cont.
27. rerror rate	% of links that have Rejection errors	Cont.
28. srv rerror rate	% of links that have Rejection errors	Cont.
29. same srv rate	% of links to the same service	Cont.
30. diff srv rate	% of links to different services	Cont.
31. srv diff host rate	% of links to different hosts	Cont.
32. dst host count	count of links having the same destination host	Cont.
33. dst host srv count	count of links having the same destination host and using the same service	Cont.
34. dst host same srv rate	% of links having the same destination host and using the same service	Cont.
35. dst host diff srv rate	% of different services on the current host	Cont.
36. dst host same src port rate	% of links to the current host having the same src port	Cont.
37. dst host srv diff host rate	% of links to the same service coming from different hosts	Cont.
38. dst host serror rate	% of links to the current host that have an S0 error	Cont.
39. dst host srv serror rate	% of links to the current host and specified service that have an S0 error	Cont.
40. dst host rerror rate	% of links to the current host that have an RST error	Cont.
41. dst host srv rerror rate	% of links to the current host and particular service that have an RST error	Cont.

4.1 Number of examples in 10% training and testing data of KDD99 dataset

Table -3 The number of examples in 10% training and testing data of KDD99 dataset.

Attack Types	10% Training Data	10% Testing Data
Normal	97277	60592
Denial of Service	391458	237594
Remote to User	1126	8606
User to Root	52	70
Probing	4107	4166
Total Examples	494020	311029

4.2 The attack classes and Characteristics of the KDD'99 dataset

Table-4 The attack classes and Characteristics of the KDD'99 dataset

Class	% of KDD training Data(10%) distributions	% of KDD test data (%) distributions	Behaviour
normal	19.69%	19.48%	normal
probe	0.80%	1.34%	anomaly
DOS	79.24%	73.90%	anomaly

U2R	0.01%	0.07%	anomaly
R2L	0.23%	5.20%	anomaly

6. Feature Selection And Normalization.

To reduce and normalised the Set of 41 features from KDD'99 Cup data set as mentioned in Table-1,we have used Intelligent Agent based Attribute Selection Algorithm ,called optimal feature selection algorithm. It is implemented by using an attribute selection and tuple selection. This algorithm has been proposed using rules and information gain ratio for attribute selection.[9][15-22] In order to achieve this, the data set *D* is divided into *n* number of classes C_i. The attributesF_i having maximum number of nonzero values are chosen by the agent, and the information gain ratio is computed using Equations 1, 2, and 3, where F is the feature set.

$$Info(D)=-\sum_{j=1}^{m}[freq(Cj,D)|D|]\log2[freq(Cj,D)|D|] \quad (1)$$

$$Info(F)=\sum_{i=1}^{n}[|Fi||F|]\times info(Fi) \quad (2)$$

$$IGR(Ai)=[Info(D)-Info(F)Info(D)+Info(F)]\times 100 \quad (3)$$

In addition, tuple selection is also carried out using the rule-based approach. The steps of the optimal feature selection algorithm are as follows.

Algorithm: Intelligent Agent based Attribute Selection Algorithm

Input: Set of 41 features from KDD'99 Cup data set
***Output*:** Reduced set of features R
Step 1: Select the attributes which have variation in their values.
Step 2: Calculate the Info (D) values for the selected attributes using the equation 1.
Step 3: Select the attributes which have maximum number of non-zero values.
Step 4: Calculate the Info(F) value for the attributes selected in step 3 using the equation 2.
Step 5: Calculate the IGR value using the equation 3.
Step 6: Depending on the IGR value, select the attributes.
So after implementing the algorithm with the KDD'99 Cup data set we get the following 19 no of selected features as shown in Table-5.

Table-5 List of 19 selected features[21][22]

List of 19 selected features(from Table-1)		
Selection number	**Feature number**	**Feature name**
1	2	protocol_type
2	4	src_byte
3	8	wrong_fragment
4	10	hot
5	14	root_shell
6	15	su_attempted
7	19	num_access_shells
8	25	rerror_rate
9	27	diff_srv_rate
10	29	srv_serror_rate
11	31	srv_diff_host_rate
12	32	dst_host_count
13	33	dst_host_srv_count
14	34	dst_host_same_srv_count
15	35	dst_host_diff_srv_count
16	36	dst_host_same_src_port_rate
17	37	dst_host_srv_diff_host_rate
18	38	dst_host_serror_rate
19	40	dst_host_rerror_rate

After feature selection and Normalisation by using Intelligent Agent based Attribute Selection Algorithm we got 19 desired features. So using K-means clustering algorithm based on Nayver Bayes classifier will classify the attacks hence a prediction method will be proposed to predict the trustworthiness in cloud computing environment.[16-22]

7. K-Means Clustering

One of the most important components of a clustering algorithm is the measure of similarity used to determine how close two patterns are to one another. K-means clustering groups data vectors into a predefined number of clusters, based on Euclidean distance as similarity measure. Data vectors within a cluster have small Euclidean distances from one another, and are associated with one centroid vector, which represents the "midpoint" of that cluster. The centroid vector is the mean of the data vectors that belong to the corresponding cluster.[16-19][22]

Using the above notation, the standard K-means algorithm is summarized as

1. Randomly initialize the N_c cluster centroid vectors.

2. Repeat

a) For each data vector, assign the vector to the class with the closest centroid vector, where the distance to the centroid is determined using

$$d(z_p, m_j) = \sqrt{\sum_{k=1}^{N_d} (z_{pk} - m_{jk})^2} \qquad (4)$$

where k subscripts the dimension.

b) Recalculate the cluster centroid vectors, using

$$m_j = \frac{1}{n_j} \sum_{\forall z_p \in c_i} z_p \qquad (5)$$

Until a stopping criterion is satisfied

The K-means clustering process can be stopped when any one of the following criteria are satisfied: when the maximum number of iterations has been exceeded, when there is little change in the centroid vectors over a number of iterations, or when there are no cluster membership changes. For the purpose of this study, the algorithm is stopped when a user-specified number of iterations has been exceeded.[22-26]

8. Naive Bayesian Classification

In simple terms, a Naive Bayes classifier assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. A Naive Bayes classifier considers each of these features to contribute independently to the probability, regardless of the presence or absence of the other features. In the training phase, the Naive Bayes algorithm calculates the probabilities of a theft given a particular attribute and then stores this probability. This is repeated for each attribute, and the amount of time taken to calculate the relevant probabilities for each attribute. In the testing phase, the amount of time taken to calculate the probability of the given class for each example in the worst case is proportional to n, the number of attributes. However, in worst case, the time taken for testing phase is same as that for the training phase. For some types of probability models, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood. In other words, one can work with the Naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

9.Experimentation & Results

9.1 Calculation of Performance Parameters

There are many measures available for evaluating system performance. For evaluating intrusion detection results we have used the following measure. The performances of each method are measured according to the Accuracy, Detection Rate and False Positive Rate.

So

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Detection Rate} = \frac{TP}{TP+FP}$$

$$\text{False Positive Rate} = \frac{FP}{FP+TN}$$

Where

1. True positive (TP) means number connections that were correctly classified as intrusion.
2. True Negative (TN) means number of connections that were incorrectly classified as intrusion.
3. False positive (FP) means number of intrusion connections that were incorrectly classified as normal.
4. False negative FN) means number of normal connections that were incorrectly classified as intrusion.

10.Results

Table 6: Naive Bayesian classification (Training data set).

Actual	Predicted Normal	Predicted DoS	Predicted Probe	Predicted U2R	Predicted R2L	Accuracy (%)
DoS	447	36924	16	1747	8	94.1
Probe	0	0	410	0	1	99.8
U2R	0	0	0	4	1	80
R2L	27	0	3	9	74	65.5
Normal	8901	8	138	574	104	91.2

Table 7: K-Means clustering by Naive Bayesian classification (Training data set)

Actual	Predicted Normal	Predicted DoS	Predicted Probe	Predicted U2R	Predicted R2L	Accuracy (%)
DoS	3	33911	0	1	208	99.33
Probe	0	0	410	0	0	100
U2R	1	0	0	2	2	40
R2L	35	2	3	4	69	61.6
Normal	9688	3	22	5	9	99.4

Table 8: Naive Bayesian classifier (Testing data set).

Actual	Predicted Normal	Predicted DoS	Predicted Probe	Predicted U2R	Predicted R2L	Accuracy (%)
DoS	6431	32185	417	0	0	81.5
Probe	6	12	393	0	0	95.6
U2R	1	0	0	4	0	80
R2L	10	0	1	0	102	90.3
Normal	7845	14	131	1664	43	74

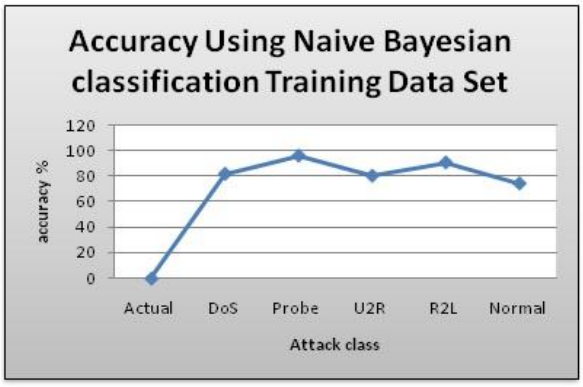
Table-9 K-Means Clustering via Naive Bayesian classification(Testing data set).

Actual	Predicted Normal	Predicted DoS	Predicted Probe	Predicted U2R	Predicted R2L	Accuracy (%)
DoS	134	38979	27	0	1	99.4
Probe	0	3	404	4	0	98.3
U2R	1	0	1	4	0	79.2
R2L	4	12	0	3	94	98.3
Normal	9670	9	3	35	2	99.3

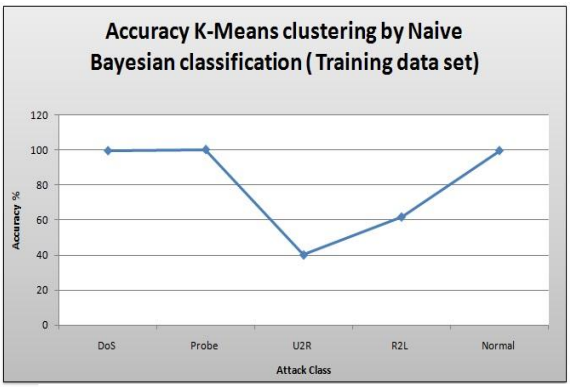
11.Accuracy Comparison

Table- Accuracy comparison of 4 Methods

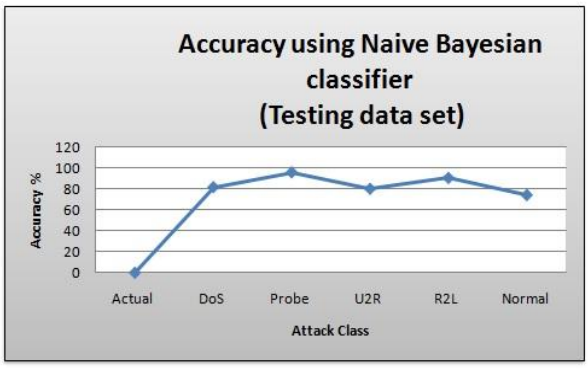
Attack Class	Naive Bayesian classification (Training data set).	K-Means clustering by Naive Bayesian classification (Training data set)	Naive Bayesian classifier (Testing data set).	K-Means Clustering via Naive Bayesian classification (Testing data set).
DoS	94.1	99.33	81.5	99.4
Probe	99.8	100	95.6	98.3
U2R	80	40	80	79.2
R2L	65.5	61.6	90.3	98.3
Normal	91.2	99.4	74	99.3



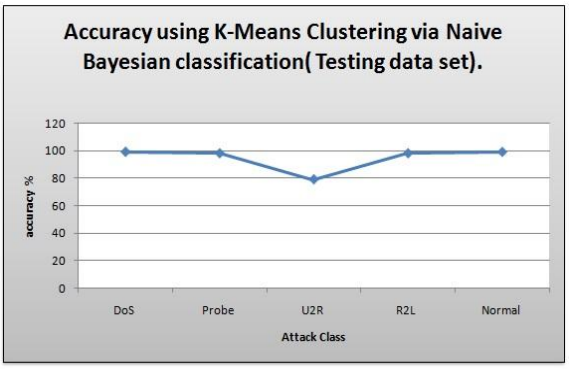
a). Naive Bayesian classification (Training data set).



b). K-Means clustering by Naive Bayesian classification (Training data set)



c). Naive Bayesian classifier (Testing data set).



d). K-Means Clustering via Naive Bayesian classification(Testing data set).

Figure3 Accuracy comparison Graph of 4 Methods.

12. Conclusion

This paper describes about the classification of attacked based upon, Naive Bayesian classification and K-Means clustering by Naive Bayesian classification in cloud environment using KDD 99 dataset. The performances of each method are measured according to the Accuracy, Detection Rate and False Positive Rate and used to predict the trustworthiness of the cloud platform. It also describes the cloud virtualization technology and the security issues and attributes of KDD 99 datasets for intrusation detection. The work will assist the cloud service providers(CSP) to discover the convenience of the IaaS environment and the effect of Anti-Malicious Software (AMS) with its productivity in the cloud surroundings in order to expand the reliability.

.

13. References

[1]. Hamdaqa Mohammad. Cloud Computing Uncovered: A Research Landscape. Elsevier Press. pp. 41–85. ISBN 0-12-396535-7 (2012).

[2]. Rafael Moreno-Vozmediano, Rubén S. Montero and Ignacio M. Llorente. Key Challenges in Cloud Computing -Enabling the Future Internet of Services, Published by the IEEE Computer Society 1089-7801/13/ © 2013 IEEE, IEEE internet computing (2013).

[3]. Dimitrios Zissis and Dimitrios Lekkas. Addressing cloud computing security issues, 0167-739X/ © 2010 Elsevier B.V. All rights reserved.doi:10.1016/j.future.2010.12.006 (2010).

[4]. S. B. Dash, H. Saini , T. C. Panda and A. Mishra. Service Level Agreement Assurance in Cloud Computing: A Trust Issue, International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 2899-2906 ISSN:0975-9646 (2014).

[5]. M. Rajendra Prasad, R. Lakshman Naik and V.Bapuji. Cloud Computing: Research Issues and Implications, International Journal of Cloud Computing and Services Science, Vol.2, No.2, pp. 134~140 ISSN: 2089-3337 (2013).

[6]. Dash, S.B., Mishra, A., Panda, T.C., Pani, S., "EVM and IVM Dyanimics in Cloud Environment", PROCEDIA COMPUTER SCIENCE,Vol.85,pp.834-842, 2016.

[7]. Dash, S.B., Saini, H., Panda T.C., Mishra, A., " Mathematical Ontology for Infectious Virtual Machines in IaaS Cloud Environment", Indian Journal of Science & Technology ,Vol. 9(34), 2016 .

[8]. Sanchika Gupta, Padam Kumar, TAXONOMY OF CLOUD SECURITY, International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.3, No.5, October 2013

[9]. M.A.Francesco and F.Gianni. An approach to a cloud Computing network, IEEE Explorer, pp113-118 (2008).

[10]. S B. Dash, H. Saini, T C. Panda and A. Mishra. Prediction of Trustworthiness in the Cloud Computing Environment using Predator-Prey Model. International Journal of Cloud Computing and Services Science, Vol.2, No.5, pp. 336~344 ISSN: 2089-3337 (2013).

[11]. S. B. Dash, H. Saini , T. C. Panda and A. Mishra. A Theoretical Aspect of Cloud Computing Service Models and Its Security Issues: A Paradigm, Journal of Engineering Research and Applications ,ISSN : 2248-9622, Vol. 4, Issue 6, pp.248-254 (2014).

[12]. Yunfa Li, Wanqing Li, Congfeng Jiang, A Survey of Virtual Machine System: Current Technology and Future Trends, 2010 Third International Symposium on Electronic Commerce and Security.

[13]. Sunny Sharma , Prithvipal Singh , Amritpal Singh, User centric security requirements and threat analysis in Cloud Computing, International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 04; April - 2016 [ISSN: 2455-1457]

[14]. Nidhi Srivastav, Rama Krishna Challa "Novel Intrusion Detection System integrating Layered Framework with Neural Network" 978-1-4673- 4529, in IEEE, 2012

[15]. Zhao Y ongli, Zhang Yungui, Tong Weiming, Chen Hongzhi, "an improved feature selection algorithm based on MAHALANOBIS distance for network intrusion detection", international conference on sensor network security technology and privacy communication system (SNS & PCS), 2013

[16]. Fengli Zhang, Dan Wang "an effective feature selection approach for network intrusion detection" IEEE Eighth international conference on networking, architecture and storage, 2013

[17]. Amjad Hussain Bhat , Sabyasachi Patra , Dr. Debasish Jena, Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines, International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 6, June 2013 ISSN 2319 - 4847

[18]. Rashmi Singh, Diwakar Singh, A Review of Network Intrusion Detection System Based on KDD Dataset, Int J Engg Techsci Vol 5(1) 2014, 10 – 15

[19]. Sannasi Ganapathy , Kanagasabai Kulothungan, Sannasy Muthurajkumar, Muthusamy Vijayalakshmi, Palanichamy Yogesh and Arputharaj Kannan, Intelligent feature selection and classification techniques for intrusion detection in networks: a survey, Ganapathy et al. EURASIP Journal on Wireless Communications and Networking 2013, 2013:271.

[20]. H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood, Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets

[21]. Kamini Nalavade, B. B. Mehsram, Evaluation of K-Means Clustering for Effective Intrusion Detection and Prevention in Massive Network Traffic Data, International Journal of Computer Applications (0975 – 8887) Volume 96– No.7, June 2014.

[22]. S. Sontakke, " Intrusion Detection System for Cloud Computing," International Journal of Scientific & Technology Research Volume 1, Issue 4 (page no. 67-71), May 2012.

[23]. L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering," School of Management, Fudan University, Shanghai 200433, PR China, Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.

[24]. M. Hussain, "Distributed cloud intrusion detection model," International Journal of Advanced Science and Technology Vol. 34 (page no. 71-82), September, 2011..

[25]. Leung K, Leckie C. Unsupervised anomaly detection in network intrusion detection using clusters. In: Proc 28th Australasian CS conf, vol. 38, Newcastle, Australia; 2005. p. 333–42.

[26]. J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory," ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262–294, 2000.

[27]. Y. Bouzida. Principal Component Analysis for Intrusion Detection and Supervised Learning for New Attack Detection. PhD thesis, Graduate Faculty of Ecole Nationale Superieure des Telecommunications de Bretagne, 2006.

[28]. S. Chebrolu, A. Abraham and J. P. Thomas. June 2005. Feature Deduction and Ensemble Design of Intrusion Detection Systems. Computers and Security, 24, 295-307.

[29]. O. Depren, M. Topallar, E. Anarim and M.K. Ciliz. 2005. An Intelligent Intrusion Detection System for Anomaly and Misuse Detection in Computer Networks. Expert systems with Applications, 29, 713-722.

[30]. Rashmi Singh, Diwakar Singh, A Review of Network Intrusion Detection System Based on KDD Dataset, Int J Engg Techsci Vol 5(1) 2014, 10 – 15

[31]. H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood, Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets.

[32]. KDD'99 Competition Dataset, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, (1999).

[33]. Knowledge discovery in databases DARPA archive. http://www.kdd.ics.uci.edu/databases/kddcup99/task.html

[34]. Selim, S. Z. and Ismail, M. A., "K-means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality," IEEE Trans. Pattern Anal. Mach. lntell. Vol. 6, pp. 8l_87 (1984).

[35]. Wu, K.-L. and Yang, M-S., "Altenative C-means Clustering Algorithms," Pattern Recognition Vol. 35, pp. 2267_ 2278(2002).

[36]. Maulik, U. and Bandyopadhyay S., "Genetic Algorithm-based Clustering Technique," Pattern Recognition Vol. 33, pp. 1455_1465 (2000) no. 1, pp. 60-68, Feb. 2003.

[37]. Kennedy, J. and Eberhart, R., "Particle Swarm Optimization," Proc. of IEEE International Conference on Neural Networks (ICNN), Perth, Australia, Vol. 4, pp. 1942_1948.

[38]. Eberhart, R. and Kennedy, .I., "A New Optimizer Using Particle Swarm Theory," Proc. 6th lnt. Symposium on Micro Machine and Human Science, pp. 39_43(1995).