Proposed Network Intrusion Detection System Based on Fuzzy c Mean Algorithm in Cloud Computing Environment

Shawq Malik Mehibs
Computer Science Department, University of Technology, Baghdad, Iraq.
shouq90@ymail.com
Soukaena Hassan Hashim
Computer Science Department, University of Technology, Baghdad, Iraq.
Soukaena.hassan@yahoo.com

Abstract

Nowadays cloud computing had become is an integral part of IT industry, cloud computing provides Working environment allow a user of environmental to share data and resources over the internet. Where cloud computing its virtual grouping of resources offered over the internet, this lead to different matters related to the security and privacy in cloud computing. And therefore, create intrusion detection very important to detect outsider and insider intruders of cloud computing with high detection rate and low false positive alarm in the cloud environment. This work proposed network intrusion detection module using fuzzy c mean algorithm. The kdd99 dataset used for experiments .the proposed system characterized by a high detection rate with low false positive alarm.

Keywords: Cloud environment, Intrusion detection, Fuzzy c mean (FCM), Machine learning.

الخلاصة

في الوقت الحاضر الحوسبة السحابية اصبحت جزء مكمل في صناعة تكنولجيا المعلومات، الحوسبة السحابية توفر بيئة عمل تسمح للمستخدم بمشاركة البيانات والموارد عبر الانترنت .حيث الحوسبة السحابية عبارة عن تجمع افتراضي من الموارد عبر الانترنت،هذا يؤدي الى مسائل اخرى تتعلق بالامن والخصوصية في بيئة الحوسبة السحابية .لذلك من المهم جدا خلق نظام كشف تطفل لكشف المتسللين في خارج وداخل بيئة الحوسبة السحابية بدقة عالية ومعدل انذار كاذب منخفضة .هذا العمل يقترح نظام كشف تطفل قائم على خوارزمية العنقدة المضببة . اجريت التجارب على بيانات KDD99. العمل المقترح يمتاز بمعدل كشف تطفل عالي مع نسبة انذار كاذب منخفضة .

الكلمات المفتاحية:- الحوسبة السحابية، كشف التطفل، العنقدة المضببة، تعلم الالة.

1. Introduction

Cloud computing is distributed architecture, providing resources, computing environment, exchange information between end users and storage data. In recent years, most of the organization used cloud environment to reduce the cost of resource and processing of information .The main benefit of cloud computing that the resource can be used by users at anytime and anywhere who wants, these services provide cloud service provider (CSP). The organizations that used cloud computing uploaded a tremendous amount of basic information to public cloud this public cloud contain sensitive information is prone to security risk like confidentiality, availability and integrity of these organizations. In cloud computing there are two concerns of security issues. The first one regardless to cloud service provider which has guarantee that the services provided and cloud infrastructure are safe and secure , the second regardless

to the consumers including guarantee that the information and data of consumers are protected (Ahmed *et al.*,2013).therefore, Intrusion detection systems besides firewall used to protect resource and data in the cloud environment form Suspicious activities. Firewall cannot be used to detect insider attack special denial of service (Dos) and distributed denial of service (DDos), so it is the inevitable to a developed intrusion detection system in a cloud environment. Intrusion Detection System (IDS) is an important part of network security used to maintain system availability and data integrity (Vikrant *et al.,* 2013). Anomaly detection used to detect suspicious activity through search for abnormal events in the monitored data. The operation of collect information related to intrusion and monitors the activity occurring in the system for analysis them to refer to the intrusion. Then the system will be raised alarm if there is probably intrusion (Ramgovind, and Smith, 2010). An IDS is a software that Is automatically the intrusion detection process and detects possible intrusions. Network-based (NIDS) identify suspicious activity through monitors network traffic for certain network segments or devices and analyze the network and application protocol activity .The fast development of network technology and continuous improvement of technology, the ways of new attacks emerge in endlessly. Intrusion detection system should be capable to conduct interconnection analyzing from multiple inputs that are combined. There are several important characteristics of fuzzy systems fittings intrusion detection these characteristics illustrated as below (John *et al.*,):

- The ability of Fuzzy systems to combine inputs from different sources
- There is types of intrusions are not possible clearly recognized (e.g. at what threshold should an alarm be set?)
- Alarms rate that might happen with intrusion usually unknown.

Fuzzy clustering is clustering approach to grouping object into threaded categories called clusters based on Statistical technique. The fuzzy clustering allows the object to belong more than cluster with some degree of membership, in such away the objects in one cluster more like to between them than to those in other cluster. The main advantages of clustering technique its ability to detect new pattern of attack. In our work we proposed FCM algorithm which is most commonly used in fuzzy clustering . The FCM algorithm operates on division a finite set of given samples into a set of fuzzy clusters in relation to some certain criterion. In the cloud environment the most research areas focus on high accuracy despite of other sequence challenges such as false alarm rate and response time (Ahmed *et al.,* 2013).For that, we proposed network intrusion detection system based on the fuzzy c mean algorithm to detect the normal behavior and attack behavior .the proposed work characterized by low false alarm and high detection accuracy.

2. Patterns of attacks in cloud environment

The service and resource provided by cloud computing motivate the intruders to gain continues service and resource .The attacks that effect of the cloud computing service and resource.

1. Insider attack

The business man, employers and partners who work in the present time or in the past time that have authorized access to information of users indicated as insiders. Those authorized users revealing the user's information. Besides, those insiders may

attempt get unauthorized privileged access to cloud resources. This attack causes dangerous security risks and very difficult to detect.

2. Flooding attack

The attacker send massive amount of packets from acquit host (zombie) in network to flood the victim machine. The type of packets can be TCP, UDP, ICMP or a Blend of them. This type of attack aims to prevent authorized users from using cloud resources to penetration the cloud resources. The reason for occurring flooding attack is illegal network connections. In cloud computing the VMs can be requests by everyone via internet .Therefore, the denial of service (Dos) and distributed denial of service (DDos) can be occur by zombies. The availability of service in the cloud affects by this type of attack.

3. User to root attacks

The attacker theft the login details of legitimate user (sniffing or guessing password).As the attacker used authorized account in the process of hacking, it not possible to recognized by search the packets. The most popular type of this attack is buffer overflows. The method of execute buffer overflow attack is when the application program copies amount of data into buffer its size less than data. There are no security package to deal with phishing attacks and weak password recovery techniques.

4. Port scanning

In this type of attack, attacker attempt to explore the list of open, closed and filtered ports via port scanning ,after that the attackers can hacking the service he went through that ports. Through port scanning network information such as MAC address, IP address, gateway filtering, router filtering and firewall rules can be found. The attacker can attack the service provided by cloud computing through open ports.

5. Attacks on virtualization

In this attack, attacker hack a hypervisor then can be gain access to host machine to control them. DKSM, SubVirt and BLUEPILL are the most popular type of this attack. The Zero-day fragility motivate hacker to get access virtual machine manager (VMM) or hypervisor .This can be done by search for Zero-day fragility in the target software then modified the target software.

6. Backdoor channel attacks

In this attack, attacker get access to targeted node compromises a node to impact user secrecy .By using this type of attack hacker able to control the resource of victim .Also can revealing victim's secrecy. In addition that the attacker can be used the victim as zombie to launch DDoS attack. This attack affects confidentiality of the system (Pandeeswari and Ganesh Kumar,2015; Chirag *et. al.*,2013).

3. Related work

Vikrant G. Deshmukh *et.al.,* 2013, proposed FC-ANN approach for intrusion detection in cloud environment .The proposed intrusion detection model in the first place the training dataset divided into homogenous subgroups via fuzzy clustering technique .The result subsets used to train ANN learning algorithms. Then emulation the ANN models to reduce the error by using the whole dataset. The membership grades created using fuzzy clustering model used to merge the results. Finally new ANN train using the merge results.

Hao Wang *et.al.*,2010,developed intrusion detection system based on hybrid fuzzy c mean clustering .The developed system combine fuzzy c mean algorithm and Quantum-behaved Particle Swarm Optimization (QPSO) algorithm to find the global optimum solution.

Saeed Khazaee and Maryam Sharifi Rad, 2013, proposed intrusion detection based on Fuzzy-ARTMAP neural network and used fuzzy c mean algorithm as prepossessing step .The fuzzy c mean algorithm used to cluster the training data and separated the inappropriate sample from the training set .In this approach the sampling with membership smaller than 0.5 well moved to new set called inappropriate data1.Then after clustering the sampling that not match the clusters moved to inappropriate data2 .Then both inappropriate data1 and inappropriate data2 class labels change to abnormal .Then the Fuzzy-ARTMAP neural network used as classifier.

4. Dataset description

The KDD cup 99 was popular dataset used for evaluate intrusion detection algorithms. This dataset consist TCP connections, each connection has 41 features with a label determine the type of a connection whether normal connection or type of attack connection. The feature of dataset divided to numeric and symbolic features, classified into the following four categories (Basic features, Content features, Time-based traffic features, Time-based traffic features). Attack type classified to four main categories (Shelly Xiaonan Wu and Wolfgang Banzhaf, 2010):

- Denial of Service (DOS) attacks: The attacker attempt to make the system resource occupied to prevent the legitimate user from using the system.
- Probe attack: The attacker scans the network to collect information and find fragility .Then use this fragility to attack at later time.
- Remote-to-Local (R2L) attack: Hacker sent packets to the victim machine through network .After that exploit fragility to get unauthorized local access to that machine.
- User-to-Root (U2R) attack: in this attack, hacker in the first get access to normal user then exploit fragility in the system to get root level access .The aim of this attack to get illegitimate super-user privileges.
-

Table 1: Attack types

Category	Type
DoS	smurf, neptune, back, teardrop, pod, land
Probe	satan, ipsweep, portsweep, nmap
R2L	warezclient, guess_passwd,warezmaster, ftp_write, multihop, phf, spy, imap
U2R	buffer_overflow,rootkit, loadmodule, perl

Table 2: Number of samples in KDD cup 99

dataset	normal	Dos	Probe	U2R	R2L	Total
corrected KDD 99"	60593	229853	4166	70	1126	311029
"10% KDD	97277	391458	4107	52	1126	494020

5. Preprocessing Datase

In KDD Cup 99 intrusion detection dataset, each record consists of 41 features which extracted to abstract the information of each connection. To train the algorithm, the enumeration and normalization operations of some data is necessary. In first the symbolic variables convert to number and then all variables are normalized. The symbolic features set to sequential integer values. The dataset consists of symbolic and numeric values, all symbolic features value transform into numeric values such as

three types of protocols (tcp, udp, icmp) and 68 types of services in KDD cup 99 and 11 types of flag, each one takes value from [1..N] as described in table (1), and each numerical value in the dataset is normalized between 0.0 and 1.0 according to the following equation:

$$x_n = \frac{x - xmin}{xmax - xmin} \qquad (1)$$

Table 3: Numeric value of symbolic feature of KDD cup99 dataset

Protocol type	Feature value	Service	Feature value	Service	Feature value	Service	Feature value	Flag	Feature value
Tcp	1	Private	1	time	23	shell	45	SF	1
Udp	2	Smtp	2	mtp	24	Efs	46	SH	2
icmp	3	http	3	gopher	25	login	47	S0	3
		ftp_data	4	rje	26	printer	48	S1	4
		IRC	5	link	27	netbios_ssn	49	S2	5
		telnet	6	Ctf	28	csnet_ns	50	S3	6
		Domain	7	Hostnames	29	nntp	51	RSTR	7
		Finger	8	iso_tsap	30	supdup	52	REJ	8
		Other	9	pop_2	31	http_443	53	RSTO	9
		ftp	10	netbios_dgm	32	uucp_path	54	RSTOSO	10
		Imap4	11	netbios_ns	33	domain_u	55	OTH	11
		pop_3	12	sql_net	34	ntp_u	56		
		Sunrpc	13	bgp	35	ecr_i	57		
		pm_dump	14	vmnet	36	eco_i	58		
		Echo	15	Z39_50	37	tim_i	59		
		Discard	16	ldap	38	urp_i	60		
		Systat	17	nnsp	39	red_i	61		
		Daytime	18	kshell	40	Remote_job	62		
		Netstat	19	klogin	41	X11	63		
		Ssh	20	uucp	42	http_8001	64		
		Name	21	courier	43	urh_i	65		
		whois	22	exec	44				

6. Fuzzy c mean (FCM) algorithm

The aim of fuzzy clustering algorithm is partition the data to various clusters according to degree membership value, in a way that each cluster contains data more similar to each other and different from data in other clusters. In fuzzy c-means algorithm the data point data point belong to more than cluster with some degree of membership which known as soft clustering. The data assign to clusters based on fuzzy member ship function. The fuzzy clustering algorithm depends on minimizing objective j and calculates using equation (2). This is based on the following object function minimization j as equation (2).

$$J_m(\mathrm{U}, \mathrm{C}) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^{m} \; d_{ij}^{2}(x_i, c_j) \qquad (2)$$

m – real number in domain ($1 \leq m < \infty$).
k – number of cluster.
n – number of data samples.
ui – membership degree that indicate the probability that data sample xi belong to j^{th} cluster.
c_j – center of cluster.

The fuzzy clustering can be done by repeatedly modified cluster center c_j and fuzzy membership uij using equation (3) (4).

$$u_{ij} = \frac{1}{\sum_{j=1}^{k} \left(d_{ij}/d_{ik}\right)^{2/m-1}} \quad , \qquad \forall i \qquad (3)$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \quad , \qquad \forall i \quad (4)$$

Where u_{ij} indicated the membership degree of the data samples that belong specific cluster, and satisfied the following conditions:

$$\sum_{j=1}^{k} uij = 1 \qquad \forall k \ (5)$$

$$\sum_{j=1}^{k} uij > 0 \qquad \forall i \ (6)$$

6.1 The proposed algorithm

The proposed FCM algorithm for intrusion detection (FCM-ID) consist of two phase .The first phase is training phase where the optimum cluster center obtain. The second phase is testing phase which used the cluster center result from training phase to determine the cluster of new samples. Algorithm (1) represented the training stage of proposed module and algorithm (2) represented the testing stage of the proposed module.

Algorithm (1): Testing stage of proposed (FCM-ID) module.

Input: number of samples selected from KDD99 dataset, fuzziness parameter.
Output: Vector of two cluster center c= {c1, c2}.
Steps:
Selected two cluster centers from training samples randomly.
Compute the membership matrix for each samples using equation (3).
Update the cluster centers using membership matrix and according equation (4).
Repeat until stopping criteria.
End

Algorithm (2): Testing stage of the proposed (FCM-ID) module.

Input: the number of testing samples selected from KDD99 dataset, vector of two cluster
Center result from training stage.
Output: classified samples to normal or attack.
Steps:
Compute the membership matrix for each sample using cluster center from training stage according to equation (3).
Determine the type of cluster for each sample using membership matrix according the following formula.
If $(u_{i1} > u_{i2})$ then c_j (Xi) =1.
Else
c_j (Xi)=2
End if
End

7. Performance Evaluation

The ability of IDS to making the correct predictions, consider the measure of its effectiveness. Depending on the comparisons between the results that predict via intrusion detection system and the true nature of the event. There are four prospect outputs are illustrated in table 4, known as confusion matrix. True positives (TP) in addition to true negatives (TN) indicate that IDS successfully detect the event as attack and normal, respectively. False positives (FP) indicate that IDS faulty predict normal events as attacks. False negatives (FN) refer to that IDS faulty detect intrusion event as normal event. Both FN and FP rates reduce the efficiency of the IDS where FP reduce the capability of system in detection and FN will make the system Susceptible to intrusion .therefore, both of them should be minimized as possible as . Based on confusion matrix show in table 2, to determine the performance of IDSs the following measures used for the numerical evaluation (Shelly Xiaonan Wu and Wolfgang Banzhaf,2010).

Accuracy (ACC): It is performance measure indicates the ratio between the samples that are correctly classified as normal and attack to the total number of samples and calculated using equation:

$$\text{ACC}= \frac{TP+TN}{TP+TN+FP+FN} \qquad (7)$$

Detection rate (DR): It is performance measure indicates the ratio between the numbers of samples which are correctly detected as attack to the overall number of attack samples calculated using equation:

$$\text{DR}= \frac{TP}{TP+FN} \qquad (8)$$

False Alarm Rate (FAR): It is performance measure indicates the ratio between number of samples which are faulty detect as attack to overall number of normal samples and can be calculated using equation:

$$\text{FAR} = \frac{FP}{TN+FP} \qquad (9)$$

Table4: Confusion matrix

Actual class	**Predicted Class**	
	Negative class(normal)	Positive class(attack)
normal	True negative (TN)	False positive (FP)
attack	False negative (FN)	True positive (TP)

8. Experiments and Results

The KDD cup 99 dataset used to training and testing FCM algorithm, the proposed FCM algorithm applied to classify dataset into two clusters, One for normal and the second for attack. In the training phase (3000) subset of records are selected randomly from whole dataset and used for train the algorithm .This subset of records contain normal and all other types of attack. The Fuzziness parameter set to m=2 and number of cluster set to c=2.To evaluate effectiveness of the proposed algorithm we conduct two experiments. In experimental 1 the trained model tested with (1000)

subset of data of records contains both normal behavior and the four types of attacks. In experimental 2 subset consist of (500) record contains both normal and attack samples used to evaluate the proposed module. The subsets of data used in this work illustrated in table (5). Various performance measures used to evaluate the proposed module such as detection rate (DR), false alarm rate (FAR) and accuracy (ACC). The result obtain from testing phase show the high capability of proposed algorithm to distinguish normal activities from attack activities where the result from experiment 1 show effectiveness of the module to detect the attack behavior with detection rate reach to (99%) and low false alarm rate reach to (1.9%).The accuracy of system is (99%).The result of the experiment raining in the same range. The obtain result from the two experiments shows in table (6).

Table 5: Dataset description

Number of dataset	Total number of records					
	records	normal	dos	probe	U2R	R2L
Train dataset	3000	999 33.3%	1250 41%	521 17%	55 1.8%	175 5.8%
Test1 dataset	1000	227 22.7%	452 45.2%	144 14.4%	45 4.5%	132 13.2%
Test2 dataset	500	138 27.6%	113 22.6%	152 30.4%	19 3.8%	78 15.6%

Table 6: Experimental result

Performance measure	Exp1	Exp2
DR	99%	100%
FAR	1.9%	1.1%
Accuracy	99%	99%

9. Conclusion

In this paper, network intrusion detection system based FCM clustering algorithm in cloud computing environment proposed to detect intrusion event from normal behavior. In this work, FCM clustering algorithm used to partition the dataset into two clusters, one for attack and another for normal behavior. The proposed module consists of two-phase; Training phase where the module trained with the dataset to be capable to distinguish normal behavior from attack event. Testing phase where new unseen samples used to evaluate the performance of the module. The KDD99 dataset was used for training and testing the proposed module. Different performance criteria used to evaluate the proposed module. The experimental result shows the effectiveness of the system in detects attack and recognizes normal behavior with high detection rate even for repeated attack and low false alarm.

Reference

Ahmed Patel, MonaTaghavi, Kaveh Bakhtiyari, Joaquim Junior, 2013, " **An intrusion detection and prevention system in cloud computing: A systematic review"**, Journal of Network and Computer Applications Vol. 36 ,pp. 25–41.

Chirag Modi, Dhiren Patel, Bhavesh Borisaniya, Hiren Patel, Avi Patel, Muttukrishnan Rajarajan , 2013, **"A survey of intrusion detection techniques in Cloud"**, Journal of Network and Computer Applications Vol.36 ,pp.42–57.

Hao Wang, Yan Zhang, Danyun Li, 2010, **"Network Intrusion Detection Based on Hybrid Fuzzy C-Mean Clustering"**, Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010),volume 1,pp.483-486.

John E. Dickerson, Jukka Juslin, Ourania Koukousoula, Julie A. Dickerson, **"Fuzzy intrusion detection"**, IEEE, ISBN: 0-7803-7078-3,pp. 1506 – 1510.

Pandeeswari, N.; Ganesh Kumar, 2016**,"anomaly detection System in cloud environment using fuzzy clustering based ANN"**, Mobile Networks and Applications, Volume 21, issue 3, pp. 494-505.

Ramgovind, S. Eloff and M.M. Smith.E, 2010,**"The management of security in Cloud computing"**, IEEE, ISBN: 978-1-4244-5493-8, pp. 1-7.

Saeed Khazaee, Maryam Sharifi Rad, 2013**,"Using fuzzy c-means algorithm for improving Intrusion detection performance"**, IEEE, ISBN: 978-1-4799-1227-8, pp.1 - 4

Shelly Xiaonan Wu, Wolfgang Banzhaf, 2010,**"The use of computational intelligence in intrusion detection systems: A review"**, Applied Soft Computing, Vol.10,pp.1–35.

Vikrant Deshmukh, Atul Borkut, Nikhil Agam, 2013**," Intrusion Detection System For Cloud Computing"**, International Journal of Engineering Research & Technology (IJERT) Vol. 2, Issue 4, pp.1-5.