International Journal of Advanced Biotechnology and Research (IJBR)
ISSN 0976-2612, Online ISSN 2278–599X,
Vol-7, Special Issue-Number5-July, 2016, pp261-273
http://www.bipublication.com

**Case Report**

# Provide a model to improve the performance of intrusion detection systems in the cloud

**Foroogh Sedighi***

Graduate student of Web Technology,
University of Southampton
* Corresponding author: ForooghSedighi,
foroogh.sedighi@gmail.com

## ABSTRACT

High availability of tools and service providers in cloud computing and the fact that cloud computing services are provided by internet and deal with public, have caused important challenges for new computing model. Cloud computing faces problems and challenges such as user privacy, data security, data ownership, availability of services, and recovery after breaking down, performance, scalability, programmability.
So far, many different methods are presented for detection of intrusion in cloud computing. There are two important factors that differentarticlesand researches are presented based on them. These factors are location of establishing intrusion detection systems in cloud computing systems and also algorithms that are used in intrusion detection. Their final goal is maximum coverage of intrusions, increasing the speed and accuracy of intrusion detection, and decreasing of wrong alarms. Our goal in this article is to increase intrusion detection in cloud computing and decrease the rate of generatingfalsealarmsby presenting a combined method.

## 1- INTRODUCTION

As previously mentioned, cloud computing services are providedviainternet. So, security and privacy are key factors thattheseservicesare facing (Moody and others 2012).
Open and distributive structure (decentralized) of cloud computing has turned this type of computing to goals for cyber attackers,hackers and intruders(Patel eel 2013). Researches by international data corporation which has been shown in figure 1;
indicate that security is thebiggest challenge of cloud computing (gens 2009).As shown in figure 2, recent report of Lockheed cyber Security Corporation(martin 2010) put intrusion detection and intrusion prevention after the security of data asthe most important security concern include computing infrastructures.


**Figure 1:** percentage of rate of concerns around challenges and problems of computing systems

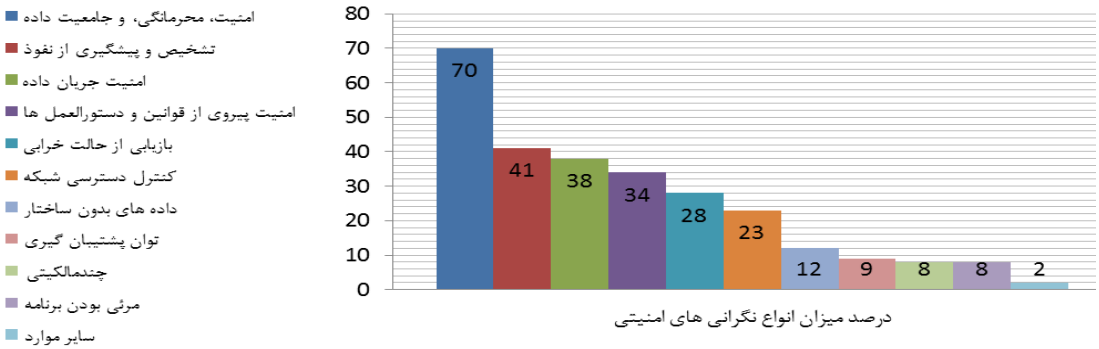



**Figure 2:** percentageofrate of security concerns of cloud computing

Statistics of Computer Emergency Response Team shows sharp annual increasing of intrusion. Each type of intrusion or attacks to vulnerabilities and security holes of network, computerized systems, or intelligence systems can have irreparable effects and breaktheirsecurity policy. (Confidentiality, integrity, and availability)(Lavaand others 2013). All of these shows the necessity and importance of applyingeffectiveintrusive detectionsystems in cloud computing.

For having more efficient and effective intrusion detection systems, these systems must be online. In other words, these systems must perform real-time detection. Offline detection systems or systems thatdetectionattack after it occurs, can just trace the happened attack but it theycan't never detect an attack which is happening right now and stop it. If the type of all attacks is known, then it would be possible to use efficient techniques of intrusion detection. As an example, detection techniques which are based on signature can be mentioned.

Since, these systems use the database of known attacks signatures for detection of incoming attacks, they are very fast and they can provide detection the capability of online detection for these systems. But, when attacks are unknown, these mechanisms lose their efficiency. Efficiency of intrusion detection systems depends on the rate of correct and false alarms thatthese system produce against attacks. The importance of this is because of security systems which act to traffic based on this analysis, detections, and alarms. This action can lead to blocking or discarding that specific traffic. So, if the alarmwas exported falsely, itwill wouldhavenegative effects on information access and system reliability.

2- Theory and literature

2.1 intrusion and intrusion detection

Standard and technology, define intrusion as an attempt to compromise security policy (confidentiality, integrity and availability) or bypass computer and network security mechanisms. This institute also defines intrusion detection as a monitoring process on computer systemsand networkevents,andalso analysis of intrusion signs. This institute knows intrusion detection systems as software/hardware systems for automation of intrusion detection procedure (scarphon and mel 2007). Common Types of intrusions that jeopardize security policies are mentioned below:

- Insider attacks
- Flooding attacks
- User to root attacks
- Port scanning attacks
- Attacks on virtual machine(VM)or hypervisor
- Backdoor channel attacks

2.2 type of intrusion detection systems

Intrusion detection systems based on the method used to detect attacks are divided to three following categories:

- Signature based detection
- Anomaly detection
- Hybrid systems

Intrusion detection systems are classified based on their location placement within network(moody and others 2013):

- HIDS: Host-based IDS
- NIDS: Network-based IDS
- DIDS: Distributed IDS
- Hypervisor-based IDS

In computer security field, intrusion detection systems play the role of warner and warn when security of the system is in danger. Another institution called security officer can respond to these alarmsandprovide provisions to address these issues. Shanmugam classified intrusion detection systems based on detections method, placement of these systems in cloud computing systems, and based on structure and type of reaction after attacks as showed in figure 3.

As shown in figure 3, Shanmugam classified intrusion detection systems based on their reaction after an attack. These classifications are as following:

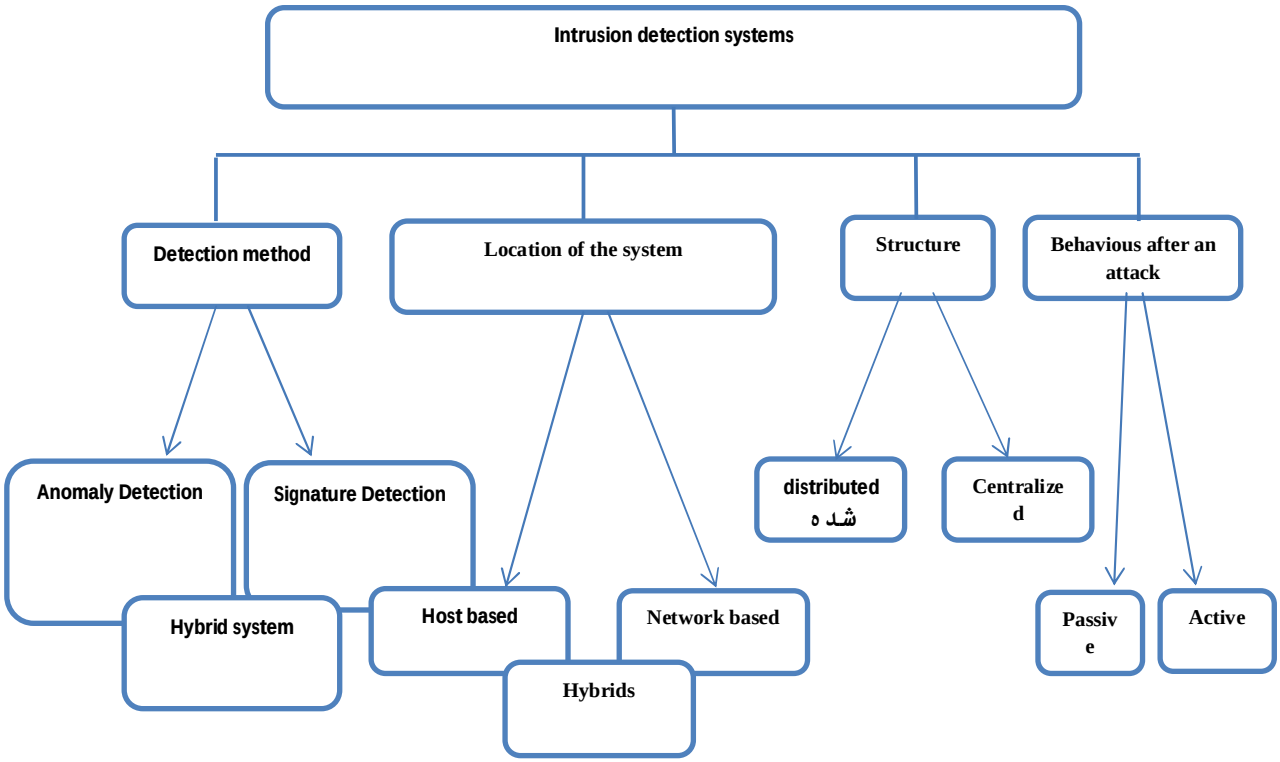
- Active Intrusion Detection Systems
- Passive Intrusion Detection Systems

Patel also presented a new classification in his article (patel and others 2013) for intrusion detection systems. According to this article, intrusion detection systems are divided into two categories based on the time of detection:

- Real-time intrusion detection systems: intrusion detection systems detect attacks in an online manner and immediately flag any deviations and provide proper prevention.
- Non real-time intrusion detection systems: intrusion detection systems provide off-line reaction to any intrusion or attack.

2.3 data mining

Han defines data mining as mining knowledge from massive amount of data (Han and kamber 2006).Raja Raman defines data mining asdiscoveringa model for a collection of data (rajaman and others 2010). In fact, data mining is a process for extracting information and knowledge and discovering hidden patterns from massive amount data. Data give us any information and it should be processed to be useful. Knowledge gives us higher level of



understanding of data and information. The process of converting data to knowledge has beenshowninfigure 4.

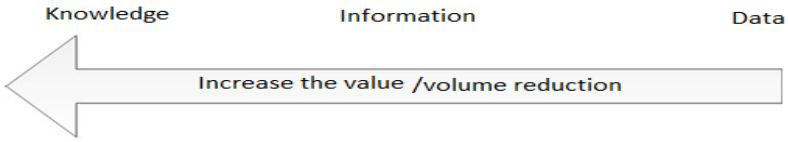
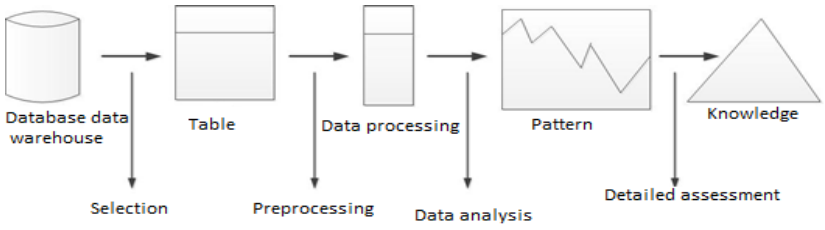


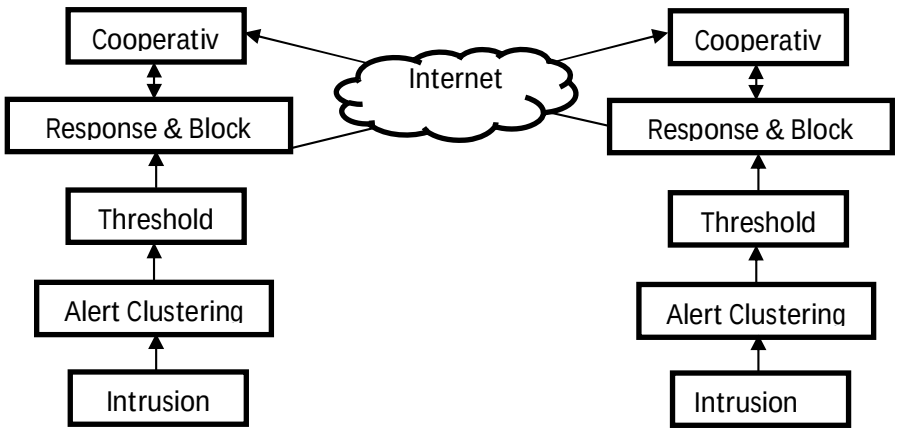

Figure 6-2 : comparing data, knowledge,information




4 **Research literatures**

Lo in an article (look and others 2008) presented a method based on cooperative agents. In this method, intrusion detection systems based on network in all regions of cloud computing were used. If each regiondetects an attack, warns other regions. Afterward, intrusion detection systems send alert message to others and decide about severity of the alert message. If this attack is identified as a new attack then new blocking rule,; add this attack to the black list. Presented architecture in this method involves components of intrusion detection, alert clustering, threshold computation and comparison, intrusion response and blocking, and cooperative operation. Intrusion detection module drop intruder package initially and send alert messages to othercloud regions to warn them about the identified attack. Alertclusteringmodule gathersgenerated alerts byother regions. Decision making about the alert (whether it is false alert or not) will be doneafter specifying the level of alert.

This method is appropriate to protect the cloud environmentattacks.

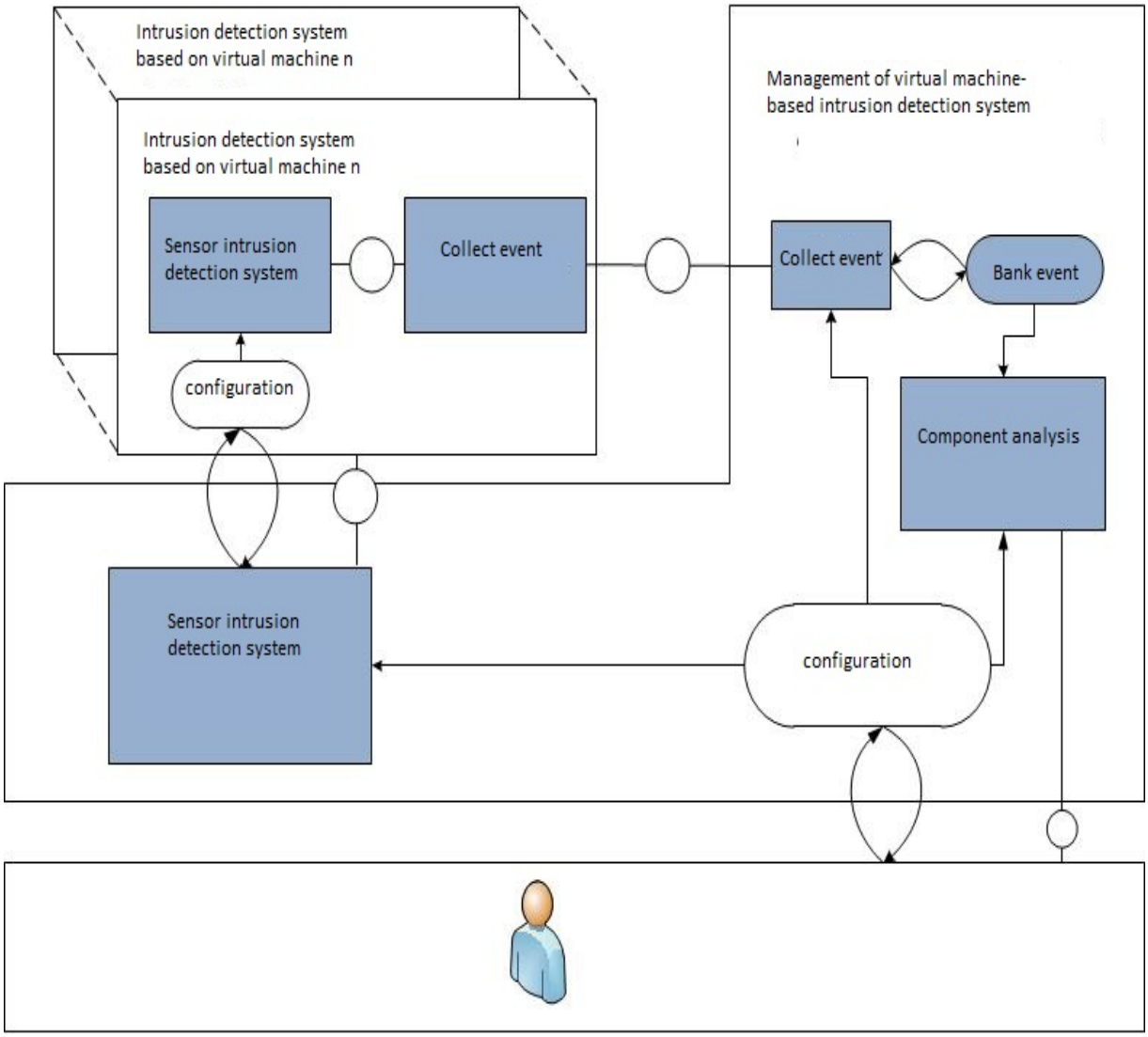


Roschkepresentedan article (rootle and others 2009) about Virtual Machine (VM) basedintrusion detection system. Architecture of this system involves two important parts:

1- Intrusion detection system management unit (consist ofevent gatherer, event database, analysis component, and remote controller)

2- Intrusion detection system sensor

The event database is a passive storage that holds information on all received events. Analysis component accesses to event database and analysis the events. Virtual machines of Intrusion detection systemscanbemanagedbyremote controller that can get connected with virtual machine of intrusion detection system and intrusion detection system sensors. Intrusion detection system sensoridentifies malicious behavior and generates salters through a reporting component to the output, which will be processed by the event gatherer. Sensors can also be configured by remote controller of intrusion detection system. In this approach, to connect event gatherers, there are pairs of sender and receiver to communicate with each other. Virtual machine management is responsibleforcontrolling, monitoring, and configuring virtual machine. Virtual machine management can also repair and recover virtual machine.



The problem of this architecture was that the presented intrusion detection system was based on virtual machine and therefore for each virtual machine a separated intrusion detection system was required. Also, this system was just able to identify known attacks.

Mazzariello and Bifulco presented a snort based intrusion detection system overran Eucalyptus open-source cloud computing system, in their article (Mazzariello and bicolor 2010). In this method, in order to detect attacks, snort was installed on Cloud Controller and physical machines.

This approach solvedthe problems of methods that used intrusion detection systems on virtual machines and had to prepare a separated intrusion detection system for each virtual machine. But, because of this method just used snort for detecting attacks,itwas only able to detect known attacks.

Gowrison (Gowrisonand others 2013) compared several kind of classification algorithm in intrusion detection systems in an article. Result of thisarticlehas been shown in table 1. According to the results, the computational time of rule based algorithms is less than classification methodsbased on neural network. Also, the percentage of correctness of classification of rule based algorithms ismore than algorithm based on neural network and percentage of wrongclassification of them are less than methods based on neural networks. In other words, according this article, forclassification, rule based algorithms are more useful in intrusion detection systems than neural network algorithms.

**Table 1:** results of comparing theaccuracy and time of rule based classification algorithms and neural network based in intrusion detection

| Algorithm based on neural network | Rule based algorithms | |
|---|---|---|
| Percentage of typical with correct classification | 93.9% | 99.9% |
| Percentage of typical with wrong classification | 6.1% | 0.1% |
| Computational time | 467.42 | 239.64 |

Modiand others (Modiand others 20102) proposed a framework integrating network intrusion detection system in an article which used a combination of Snort and decision tree. They used snort to detect known attacks and decision tree for classification of unknown patterns, inorder todecrease the generation of false alarmsand improve the rate of accurate detection.

3. Proposed method

In this article we aimto present a method to increase decision accuracy ofreceived packages from network and decrease the rate of generating false alarms. For this purpose, we try to use a combination of two clustering algorithm and classification in same time and together.In this way we could benefit from the advantages of both of them in increasing the accuracy of detection.

Here, we used C4.5 algorithm which isthe developed version of decision tree algorithm for classification ofreceivedpackagesand also use Learning Vector Quantization for clustering.

The general steps of purposed method are as following:

Step 1: The step of makingclusters(the algorithm of Learning Vector Quantization)

Step2:The step of tutorial and building decision tree.

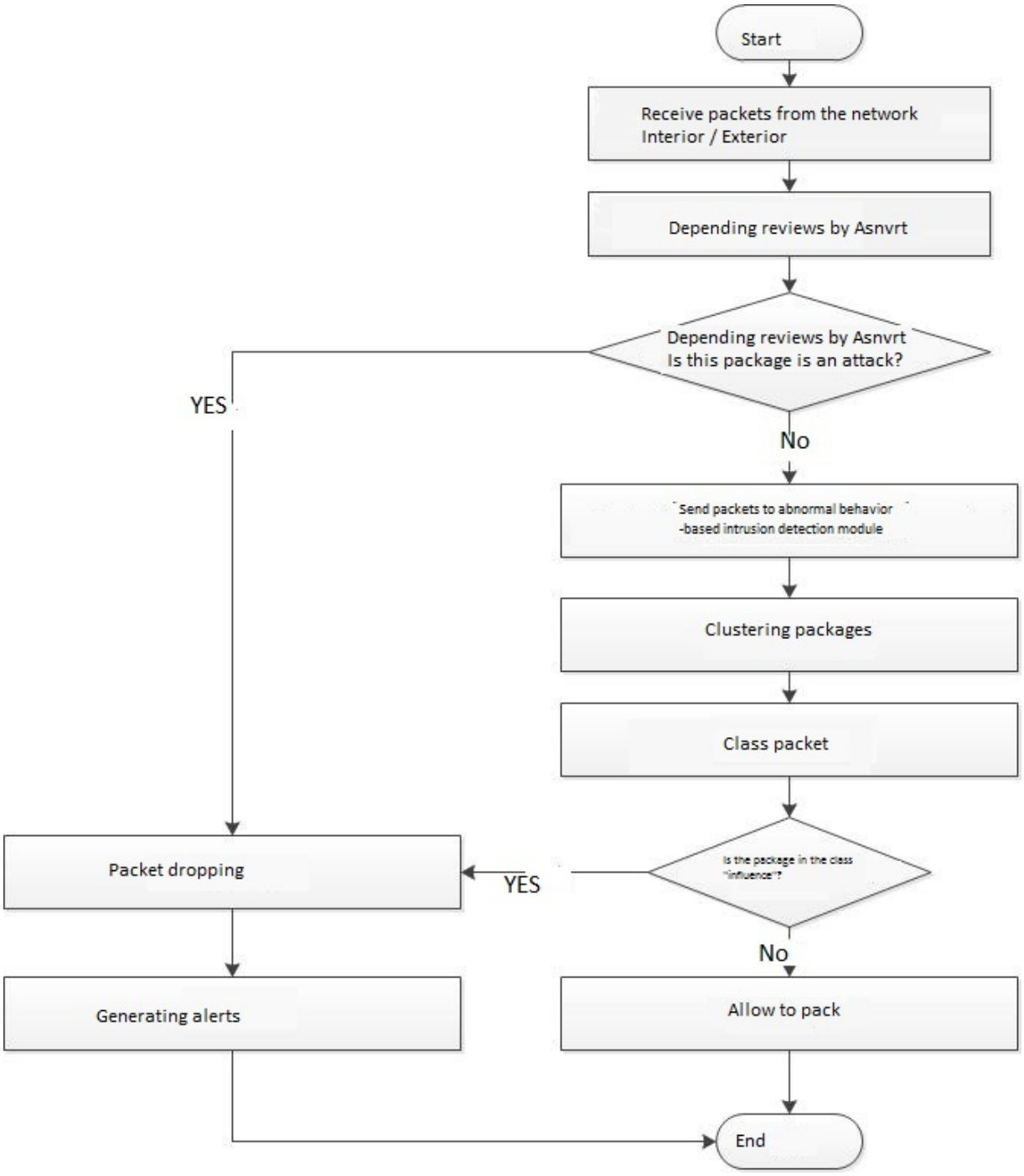
Step 3: the step of receiving package from network and hand over to snort

Step4: The step of investigatingreceived packages by snort

Step 5: The step of clustering packages

Step 6: The step of classification of packages of different clusters

Step 7: The step of generatingalarmand dropping package




1- Building cluster via learning vector quantization

Learning vector quantization algorithm includestwo generalsteps. In the first step, an unsupervised clustering method is used without consideringtrainingdata class, for specifying the locationof center of clusters. In the second steps, information of training data class isused for improvement of output clusters and decreasing the number of samples that are divided wrongly(Castile and melon 2003). Learning vector quantization algorithm is showed in figure 8.

Assumptions:

M number of clusters

- $x = (x_1, x_2, x_3, \ldots, x_n)$ number of input dimensions
- $w_{*j} = (w_{1j}, w_{2j}, \ldots, w_{nj})$ weight vector for the j-th unit of output

Aα: rate of learning :

algorithm:

(1) Define the number of clusters M

(2) Initialize M prototype vectors (weight vectors) (centroids): $w_{*1}, .., w_{*m}$

(3) Initialize learning rate (α)

(4) While (stopping criterion is satisfied)

(4-1) Randomly pick an input x

(4-2) For j = 1 TO n

(4-2-1) Determine Winner node k by finding prototype vector that satisfies $| w_{*k} - x | <= | w_{*j} - x |$

(4-2-2) Update only the Winner prototype vector weights according to:

$w_{*k}(new) = w_{*k}(old) + \alpha ( x - w_{*k}(old) )$

2. Build classification of C4.5tree

First, C4.5 build tree algorithm, build a decision tree with starting from training collection of D that involves training samples.

Decision tree (which is a type of Greedy Algorithm) is built upby top down,Recursive, Divide &Conquerermethods. At first, all training samples of D collection arein roots of this tree. Then, a feature(F with byvalues of f1, f2, f3, fn) is chosen from the features of training samples and according to that feature, training samples are divided to different categories (D1, D2 , D3 , … , Dn).. This procedure is applied on each category (Di) recursively (Han and Kamber 2006). All samples that are set in Di subgroup, have same value in the feature of Fi

Chosen features that branching is done base of them, are selected based on Heuristic or statistics methods.

Stopping States of this algorithm are:

1- All remained samples belong to same class (it means that all the samples ofDi be normal or intrusion)

2- No featuresremains which division could bedone based of it.

3- No other samples remain.

ID3 algorithm used information gainin process of selecting features, and selected featurewas the onewith the highest information gain. (Han and Kamber2006).

But e4 C4.5 algorithm uses information Gain Ratio for constructingdecision node and selected feature is the one with the highest information gain ratio.

Information gain of a feature is the amount of reducing entropy by splitting data via this feature.

Entropy is the purity(impurity) of a set (before splitting)whichis built according to following equation. (han and kamber 2006).

$$Entropy(D) = -\sum_{i=1}^{n\_class} P_i \times \log_2(P_i)$$

In this equation:

n-class: The number of classes of categories ( in our problem involves two classes of normal and intrusion).

Pi: The ratio of weight of a package related Ci class into D collection to all D collection. Pi function of existing inCiclass that is built equation 2:

$$P_i = \frac{|Freq(C_i, D)|}{|D|}$$

The expected Entropy after splitting and according to selected feature gains inequation 3. (Hanand Kamber 2006).

$$Entropy_i(D) = \sum_{j=1}^{s} (\frac{|D_j|}{|D|}) \times Entropy(D_j)$$

In this equation:

S :The number of different amount of feature F. In other word, m is number of branches gain by evaluating F on D collection in decision tree.

Dj is subdivisionofD that all members of it have the value of j for F feature.

Now feature F is calculated by equation 4(Han and kamber 2006):

Gain(F) = Entropy(D) – $Entropy_f(D)$

Information gainwith above formula tends to features with higher amount. For solving this problem, c4.5 algorithm normalizes information gain. So,itdivides information gainon total of it subdivisionentropies (equation 5).(equation 6).

$$SplitEntropy_f(D) - -\sum_{i=1}^{s} (\frac{|D_j|}{|D|}) \times \log_2(\frac{|D_j|}{|D|})$$

$$Gain\_Ratio(F) = \frac{Gain(F)}{SplitEntropy(F)}$$

When the ratio of computing gainis calculated for each feature, feature that affects the with the highest information gain is chosen as a feature for D collection. (Han and kamber 2006).

Since constructeddecision tree might be big and large and has the problem of over fitting, c4.5 algorithm uses reduced error pruning technique to prune nodes that doesn'taffect the accuracy of classification. Overfitting occurs when number of branches in tree is too manywhichleads to anomalies. This action also reduces accuracy of algorithm about unseensamples. (Han and Kamber2006).

4 - Simulationof proposed method and evaluatingresults

For implementation of this method,eucalyptus open source cloud computing environment installed on Ubuntu operating system, weka tools as a component of data mining , training data collection and NSL-KDD have been used.Table 2 displays records of test KDD99and table 3 displays characteristics of data collection of NSL-KDD .

| | Main records | Different records | Reduction rate |
|---|---|---|---|
| Normal | 60591 | 47911 | %20.92 |
| Attack | 250436 | 29378 | %88.26 |
| total | 311027 | 77289 | %75.15 |

NSL-KDD data collection includes 125974 training data and 22544 test data that involves 25 types of traffics.

| Number of class type | Number of features | Number of test data | Number of training data | Data collection |
|---|---|---|---|---|
| 25kind<br>+ 24kind of attacks)<br>(Normal traffic | 42 | 22544 | 125974 | NSL-KDD |

**4-1-** describing criteria and assessment

Percentage of samples that are classifiedcorrectly, shows the accuracy of method. So, we first make Confusion Matrix) according to the real class of packages and expected class for them. (table 4).(Chandorkar and Nandavadekar)

In this state, sample space is:

A: the number of correct negative prediction that were negative actually.

B: the number of false positives that were negativeactually.

C: the number of false negatives that were positive actually.

D: the number of correct positives that were positive actually.

Table4matrix

| | | Expected situation | |
|---|---|---|---|
| | | Negative | positive |
| Real situation | negative | A | B |
| | positive | C | D |

Now, we can present factors for evaluating (Chandorkar and Nandavadekar 2013):

Accuracy of classification(Acc): The ratio of total correct predictionto total sample space which is calculated usingequation 7:

$$Acc = \frac{A + D}{A + B + C + D}$$

True positive rate (TC): The ratio of correct positive prediction to total actual positiveswhich is gained using equation 8:

$$TP = \frac{D}{C + D}$$

False positive rate (FP): The ratio of false positive prediction to total actual negativeswhich is gained using equation 9:

$$FP = \frac{B}{A + B}$$

True negative rate(TN): The ratio of correct negative predictiontototal actual negatives is gained usingequation 10:

$$TN = \frac{A}{A + B}$$

False negative rate (FN): The ratio of falsenegative predictiontototal actual positives which is gained usingequation 11:

$$FN = \frac{C}{C + D}$$

Positive predictionrate: The ratio of correct positive predictiontototal positive predictions whichis gainedusing equation 12:

$$Precision = \frac{D}{B + D}$$

Sensitivityrate: The ratio of correct positive predictiontototal correct positive prediction and false negative predictions whichis gainedusing equation 7-4.

$$Recall = \frac{D}{C + D}$$

Confusion matrix can beapplied for evaluation of intrusion detection method as shown in table 5 and benefit from all above equations for evaluating.

| | | Expected situation | |
|---|---|---|---|
| | | normal | Intrusion |
| Real packages situation | Normal | A | B |
| | intrusion | C | D |

| Total cluster 2 data | Total cluster 1 data | Total training data |
|---|---|---|
| 101934 | 24039 | 125973 |
| %81 | %19 | |

| Total cluster 2 data | Total cluster1 data | Total test data |
|---|---|---|
| 13982 | 8562 | 22544 |
| %62 | %38 | |

| | | Expected class | |
|---|---|---|---|
| | | Normal | Intrusion |
| Real class | Normal | 2094 | 19 |
| | intrusion | 28 | 6421 |

| | | Expected class | |
|---|---|---|---|
| | | Normal | Intrusion |
| Real class | Normal | 7553 | 45 |
| | Intrusion | 46 | 6338 |
| | | | |
| | | | |

| | | Expected class | |
|---|---|---|---|
| | | Normal | Intrusion |
| Real class | Normal | 9647 | 64 |
| | intrusion | 74 | 12759 |

| Number of data | )FNR( | )TNR( | )FPR( | )TPR( | Total data |
|---|---|---|---|---|---|
| 8562 | 0,43 | 99,1 | 0.89 | 99,56 | NSL-KDD )cluster(1 |
| 13982 | 0.72 | 99,4 | 0.59 | 99,27 | NSL-KDD )cluster(2 |
| 22544 | 0.61 | 99,29 | 0.7 | 99,38 | Total weight data |
| 22544 | 23,13 | 95,15 | 4,81 | 76,80 | Proposed results of mrmoodi |

| Total data | )Acc( | F-Measure | sensible (Recall) | )Precision( | Total data |
|---|---|---|---|---|---|
| 8562 | 99,45 | 99,9 | 99,56 | 98,68 | NSL-KDD )cluster(1 |
| 13982 | 99,34 | 99,6 | 99,27 | 99,39 | NSL-KDD )cluster(2 |
| 222544 | 99,38 | 99,71 | 99,38 | 99,12 | Total weight average |
| 22544 | 84,31 | 85,11 | 76,8 | 95,45 | Proposed method results of mrmoodi |

**CONCLUSION:**

Cloud computing services are presented by internet, so security and privacy are key factors that cloudcomputing faces it. Open and distributed structure of cloud computing converts this computing to a targetfor cyber attackers and hackers.

Researches done by International Data Corporation (IDC) and also cyber security firm reports indicates that intrusion detection and intrusion prevention are the most important concerns in the field of security in cloud computing infrastructure. Because of importance of intrusion in cloud computing, it is essentialto

design intrusion detection systems that can securecloud computing system against hackers.

These systems must detect any intrusion in real-time and in an online manner. Intrusion detection systems that use detection techniques based on signature (snort), are real-time intrusion detection systems thatare able to detect known attacks and enable security mechanisms, according to known attacks patterns and known attacks database. But these systems lose their effectiveness against unknown attacks.

The effectiveness of intrusion detection systems depends on falseor correct alarmsrate that these systems generate against attacks.According to these alarms,security mechanisms will beenabledand decided for incoming packages. So, false alarm generation leads to blocking and dropping normal traffic and therebydecreaseaccess, reliability and general efficiency of the system.

In a recent research done about intrusion detection systems in cloud computing, in intrusion detection component, a detection system based on signature for detecting known attacks and an intrusion detection system based on abnormal behaviors for unknown attacks detection were used. In intrusion detection componentbased on abnormal behavior, decision tree for deciding about incoming packages was used. If incoming package known as a type ofintrusion,it was immediately recognizedbyintrusion detection component based on signature and security mechanisms were enabled. If this package wasnot from known attacks, then it entered second phase of detection and this package was delivered to detection component based on behavior to decide if it is normal or intrusion.

In this research, for improving the efficiency of intrusion detection system, a clustering algorithm and a classification algorithm has been used at the same time in the component of intrusion detection based on abnormal behavior. In this way, we decrease the rate of generating false alarms and improve the accuracy of detection.

**REFERENCES:**


1. Castillo, Oscar, and Patricia Melin. Soft Computing and Fractal Theory for Intelligent Manufacturing. springer, 2003.
2. EucalyptusSystems. Eucalyptous Fast Start Guide. Eucalyptus Systems, 2012.
3. Furht, Borko , and Armando Escalante. Handbook of Cloud Computing. Springer, 2010.
4. Gens. New IDC IT Cloud Service Survey: Top Benefits and Challenges, IDC Exchange. 2009. http://blogs.idc.com/ie/?p=730.
5. Gowrison, and [Others]. "Minimal complexity attack classification intrusion detection system." Elsevier, Applied Soft Computing 2013, 2013: 921-927.
6. Grossman. "The case for cloud computing." IEEE ITPro, 2009: 23-28.
7. Han, Jiawei, and Micheline Kamber. Data Mining: Concepts and Techniques. 2006.
8. Kayacık, Güneş, and [Others]. "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets." Dalhousie University, Faculty of Computer Science, 2005.
9. Leavitt. "Is cloud computing really ready for prime time?" IEEE Computer, 2009.
10. Liao, Hung-Jen, and [Others]. "Intrusion detection system: A comprehensive review." 2013: 16-24.
11. Lo, Chi-Chun, and [Others]. "Cooperative Intrusion detection system framework for cloud computing networks." First IEEE International Conference on UbiMedia Computing. IEEE, 2008. 280–284.
12. Martin, Lockheed. "Awareness, Trust and Security to Shape Government Cloud Adoption." 2010: 7.
13. Mazzariello, and Bifulco. " Integrating a network IDS into an open source cloud computing." Sixth international conference on information assurance and security (IAS). 2010. 265-270.
14. Modi, Chiragh N, and [Others]. "A novel framework for intrusion detection in cloud." ACM, Proceedings of the Fifth International Conference on Security of Information and Networks. ACM, 2012. 67-74.
15. Modi, Chiragh N, and [Others]. "A survey of intrusion detection techniques in Cloud." Journal of Network and Computer Applications, 2013: 42-57.

16. Modi, Chiragh N, and [Others]. "Integrating Signature Apriori based Network Intrusion Detection System (NIDS) in Cloud Computing." Elsevier, Procedia Technology, 2nd International Conference on Communication, Computing & Security (ICCCS-2012). Elsevier, 2012. 905-912..
17. Patel, Ahmed, and [Others]. "An intrusion detection and prevention system in cloud computing: A systematic review." Journal of Network and Computer Applications, 2013: 25-41.
18. Rajaraman, Anand, and [Others]. Mining of Massive Datasets. 2010.
19. Ram. "Secure cloud computing based on mutual intrusion detection system." International journal of computer application, 2012: 57–67..
20. Roschke, and [Others]. "An extensible and virtualization compatible IDS management architecture." Fifth international conference on information assurance and security. 2009. 130-134.
21. Scarfone, Karen, and Peter Mel. Guide to Intrusion Detection and Prevention System. NIST, 2007.
22. Shanmugam, Bharanidharan , and Norbik Bashah Idris. Hybrid Intrusion Detection Systems (HIDS) using Fuzzy Logic. 2011.
23. Sourcefire. SourceFire. 2013. http://www.snort.org/.
24. Tavallaee, Mahbod. "A Detailed Analysis of the KDD CUP 99 Data Set." IEEE, 2009 IEEE Symposium on Computentional Intelligence in Security & Defense Applications (CISDA 2009), 2009.