Evaluation of Fuzzy K-Means And K-Means Clustering Algorithms In Intrusion Detection Systems

Farhad Soleimanian Gharehchopogh, Neda Jabbari, Zeinab Ghaffari Azar

Abstract:- According to the growth of the Internet technology, there is a need to develop strategies in order to maintain security of system. One of the most effective techniques is Intrusion Detection System (IDS). This system is created to make a complete security in a computerized system, in order to pass the Intrusion system through the firewall, antivirus and other security devices detect and deal with it. The Intrusion detection techniques are divided into two groups which includes supervised learning and unsupervised learning. Clustering which is commonly used to detect possible attacks is one of the branches of unsupervised learning. Fuzzy sets play an important role to reduce spurious alarms and Intrusion detection, which have uncertain quality.This paper investigates k-means fuzzy and k-means algorithm in order to recognize Intrusion detection in system which both of the algorithms use clustering method.

Index Terms:- Intrusion detection system, k-means, fuzzy k-means, clustering algorithm, Fuzzy IDS

1 INTRODUCTION

INSTRUSION Detection System monitors the violation of management and security policy and malicious activities in the computerized network [1]. The intrusion can be caused by inside (legal users), or outside (illegal users) in the system [2]. Nowadays recognition and prevention of intrusion is one of the most important mechanisms that provides security in networks and computer systems, and generally is used as a complemented security for firewalls [3]. IDS systems created as a software and hardware system that each one has its specific properties [1]. Hardware systems have been preferred to software system because of their speed and accuracy. But software systems are more common because of high compatibility with several operating systems [4]. James P. Anderson is known as a first person who propounded the investigation about registered events in the system in the field of security. Anderson demonstrated a report in 1980 which was the first activity about the recognition of intrusion [5, 6]. IDS generally have three main functions: monitoring and evaluation, detection and response [7]. Intrusion detection techniques are divided into two groups: supervised learning and unsupervised learning; clustering is one of the branches of unsupervised learning. In this method samples are divided into categories with similar members [8]. One of the most common methods of clustering is K-Means algorithm which starts with a set of K reference points and data points belong to K cluster based on distance criterion [9]. Fuzzy clustering is another type that the probability of data is [0, 1] which belongs to these categories; one of the most important and applicable algorithms of fuzzy clustering is C-Mean fuzzy algorithm [10]. There are several criteria for clustering in this algorithm and the main one is the distance of any point from the center of cluster [11]. This paper is organized as follows. Firstly, we introduce types of computer network's attacks and different types of attacks and the methods of intrusion detection in systems in the next sections. In the section 4 and sub sections of them, we reviewed and evaluated the clustering methods such as K-Meand fuzzy and M-Meand Algorithms. Section 5 is presented the KDDCUP99 dataset and section 6 is analyissi and evaluation of these methods. Finally, in the section 7, we will offer the conclutions of this paper.

2 TYPES OF COMPUTER NETWORKS ATTACKS

According to intrusion detection in systems, there are a variety of computer networks attack methods that can be divided into four general categories:

2.1 Denial of Service (DoS)

DoS attacks to network or host sources. Attacker sends TCP packets with high traffic through the services. As a result this causes disorder in network normal data services. These sources include network bandwidth, data packets routing, server information, and memory and ability of calculation in servers. Victims of DoS attacks are powerful servers with fast network connections. Distributed Denial of Service (DDoS) attacks are other types of DoS attacks which are in distributed networks [5]. DoS attacks are typically divided into 6 groups as following:

Local Area Network Denial (LAND) Attack: In 1997, LAND attack is invented for the first time by some one whose nick name is "M3LT". This attack involves sending forged packets to the host computer which caused it to be locked. LAND attack sends TCP SYN forged packets with host IP address to wards one of the open ports of origin and destination, and this causes the device to be constantly posts replies.

Back Attack: This type of attack is side effects of forged information in DoS attack. In this method attacker forges the

- **Farhad Soleimanian Gharehchopogh** is with the Computer Engineering department, Hacettepe University, Beytepe,Ankara , Turkey, (corresponding author's website. www.someimanian.com, e-mail: farhad@hacettepe.edu.tr).
- **Neda Jabbari** is with Copmuter Engineering department, Scince and Research Branch, Islamic Azad University, West Azerbaijan, Iran. E-mail: nedajabbari.nj@gmail.com
- **Zeinab Ghaffari Azar** is with Copmuter Engineering department, Scince and Research Branch, Islamic Azad University, West Azerbaijan, Iran. E-mail: z.ghaffari.azar@gmail.com

origin address which is placed in IP packet. Therefore victim machine cannot detect forged packets from main ones. As result victim machine responses to them the same as others.
Ping of Death (PoD) Attack: Another type of DoS attacks which affect old operating systems is PoD attack. In this case, attacker by sending ping packets to sacrificial system which their sizes are larger than 64000 bytes causes in order to crash the system [12].

Neptune (SYN Flooding) Attack: It was recognized for the first time in 1996. Attacker sends packets from IP addresses with uncertain origin towards wandered victims to make SYN Flood attack. This increases network traffic, and causes network sources and victims to be captured by attacker [13].

Smurf Attack: Forged origin address is used in this type of attack. Attacker puts the IP address of victim system in ICMP packets and spreads it in network and victim systems sends echo replay as a response that causes extra traffic in network and prevents them from delivering to safe packets [14].

Teardrop Attack: The other types of DoS attack is Teardrop attack which attacker abuses the lack of divided IP addresses, and causes disorders by sending pieces which have overlap or high load [15].

2.2 Probes

These types of attacks are searching for intrusion ways. Attackers try to collect information using computer networks investigation in order to detect attack possibilities which are based on recognized vulnerable. The main goal of collecting this information is finding computerized services in network which have the possibility to confront with attackers [15, 16]. There are several methods as follows which are used to sweep Probe attacks:

Port Sweep and IP Sweep: It collects information by sending and receiving. In this way it determines that which host listens to the network. Port Sweep method is used to find out the opened port in a specific computer. This method for achieving this goal like IP Sweep method uses sending and receiving information. Firstly, active host and type of servicing are determined in both methods, and then collected information can be used for attack and searching vulnerable computers by attackers.

NMAP: NMAP is an opened source device which scans IP, Port, and Firewall with unused IP packet in victim computer. Entering time of packet is adaptable and it's possible to scan ports continuously or randomly.

Satan: It's an old version of SAINT which involves set of C and Perl programs. Their design and purpose are the same in two methods. Their main difference is the type of vulnerability. This method supports three levels of investigation and scan: light, normal and heavy. This technique obtains a large amount of information from network services which includes NFS, Fingerprint, and ftp. Although this method uses primitive tools, but it can provide information for attacker in order to start an intrusion. This information mainly involves data in the field of recognized vulnerability in system and operating system.

2.3 User to Root (U2R)

In U2R attack, the attacker starts with availability to normal user account, and in this way it can access the root [7]. These types of attacks are performed in victim's machine successfully and control the root [5]. There are several U2R attacks that the most important one is Buffer over Flow. This attack happens when a copy of program is copied with more data in static buffer without checking its capacity. Programmers solve these problems by exact techniques [18]. Other types of User to Root are Perl, Loadmodul, and Rootkit. Perl attack was invented in 1996 [19]; Loadmodul attack is applied in face of Sun OS4.1 systems which uses xnew windows [20]; Rootkit attack involves a set of programs which helps attacker to access victim machine. When Rootkit is installed in victim machine, attacker refers to victim system in order to download sniffer logs loads [17].

2.4 Remote-to-Local (R2L)

A remote unwanted intrusion abuses user's legal account, and sends packet on the network [5]. In fact, this attack is caused when the attacker has the ability to send information packets through the victim machine and abuses of users' local availability vulnerable by sending packets in network. There are different ways to unallowable access to local account. Some of them are as follows: Warezmaster, Warezclient, Spy, Phf, Multihop, Imap, Guess_paswd, Ftp_write [17].

3 METHODS OF INTRUSION DETECTION IN SYSTEMS

The basic principle of intrusion detection is based on the assumption that the unknown functions have perceptible differences from normal data, so they are indistinguishable. After the first report of Anderson, several intrusion detection techniques are demonstrated [5, 21]. This process is divided into three categories as follows: Specification Based Detection, Anomaly Detection, and Misuse Detection that each one of them is explained in the next sections [5].

3.1 Misuse Detection

This method is usually called a signature-based detection, and the pre-made intrusion samples of (signature) are maintained as the law. In signature-based detection method, the diagnosis is usually a database of attack patterns and by examining network traffic tries to find similar patterns which relates to what holds in its database. In this method, during new attack systems cannot recognize them; therefore network manager should add new attacks to database continuously [4]. Signature-based detection method is divided into four groups [5]:

Pattern Matching: evaluation function of this method is studied by kreibich crow croft [22]. In this method, attack patterns are designed based on packet headers, packet contents or both of them [5].

Rule based Techniques: This is one of the oldest techniques which are used in intrusion detection systems. Expert systems encrypt intruder scenarios as a set of rules, and any deviation in the implementation of the rule is reported as an intrusion. Examples of rule-based systems are as following: MIDAS [23], IDES [24], and NIDES [25, 26].

State Based Techniques: This technique works with system states and transmission, patterns of simple case are known as models of attacks patterns and they can be easier than rule-based languages considering the description of attack states such as P-best [5].

Data Mining: In this method, intrusion detection is considered as a process of data analysis that data mining techniques are used [27, 28, and 29].

3.2 Anomaly Detection

In this technique modeling based on normal behaviors is done, then the input behavior of system becomes normal following made system otherwise if more than one certain amount of statistical models is violated, they are recognized as an abnormal behavior [4, 5]. Studying about anomaly detection is started with reversed normal characteristics of the objects in order to observe things and then is checked which flags should be used to anomaly detection; anomaly detection is divided into four categories as following [5]:

Advanced Statistical Models: Dinning explained his first description about anomaly detection, and its framework was based on statistical analysis that involved eight parts such as: People, objects, audit records, statistical metrics, statistical models, specifications profile, abnormalities record, and activity rules. This model is divided into three sub-groups as following: EMERALD [31], NIDES [25, 26], and Haystack [30]. Rule based Techniques: This model is studied in four sub-groups which involves TIM namely Wisdom & Sense, NSM, and NADIR [5].

Biological Models: One of the first works in this field was written by Forrest et al. In this method by inspiring the biological principles, they have investigated human immune systems in face with attacks which are made by human's body and outside of the body [5].

Learning Models: This model is a combination of learning processes in artificial intrusion detection method. This training model is shown in both Supervised Learning and Unsupervised Learning [5]. Clustering is a subclass of unsupervised learning model of instruction.

3.3 Specification Based Detection

In this system, behavioral characters are used to detect attacks; however, it is hard to determine the behavior of several running programs in a real environment [23]. Instead of learning the behavior of system in this technique, knowledge of the expert system is used to determine threshold limitations [33]. Specification Based method can recognize the final attacks which it is possible to be used in future. In fact, this method is a combination of Misus Detection and Anomaly methods.

4 CLUSTERING

Clustering is an unsupervised classification that the classes have not been predefined. In clustering process, the samples are divided into categories which the members are alike and called cluster [8]. In classic clustering, each input sample belongs to one cluster and cannot be a member of several clusters, so if a sample is like more than one clusters, it will be difficult for us to recognize that the sample belongs to which cluster, and this is the main difference between classic and fuzzy clustering. It shows that in fuzzy clustering a sample can belong to more than one cluster, and in fuzzy logic, belonging function of clusters doesn't have two values and can have any value between 0 and 1 [9].

4.1 K-Fuzzy Mean of Clustering Algorithm

K- Mean algorithm is one of the most important clustering algorithms. In this algorithm, the first samples are divided into two or more clusters. In this fuzzy algorithm the number of clusters has been already specified. In K-Fuzzy Mean of clustering algorithm the main function is:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n} U_{ik}^{m}\, d_{ik}^{2} = \sum_{i=1}^{c}\sum_{k=1}^{n} U_{ik}^{m}\, ||x_k - v_i\, ||^2 (1)$$

In formula1: m is a real number which is bigger than 1. In most of the cases, m=2. If m=1, the non-fuzzy c-mean of main clustering function is obtained. In above formula Xk is the kth sample, and Viis the center of it he cluster and n is the number of samples. Uik shows the dependency of ith sample in kth cluster. ||×||is determined the similarity of sample(distance) from the center of cluster and can use every function that shows the similarity of sample or the center of cluster.

Steps of k-fuzzy mean algorithm [34, 9]:

- For the first clusters initial value for k, m, and U should be estimated.
- The center of clusters should be calculated by second formula.
- The dependence matrix should be calculated by in second step.
- If ||Ul+1−Ul|| ≤ ε the algorithm is finished, visa versa go to second step.

4.2 M-Mean Clustering Algorithm

One of the easiest ways is to learn without K- Means which in famous clustering problems are used a lot. This algorithm uses an easy method for clustering a collection of data with specified number of clusters. The main theory in this algorithm is description of k center for each cluster. These centers should be selected carefully because choosing different centers leads to several results. Therefore, the best choice is putting centers in farther places from each other. Next step is that each sample should be specified to the nearest center. When all of these points are specified to suitable centers, the first step has been finished and the first clustering has been done. In this step a new k center should be calculated in the previous cluster. After determination of k new center, data should be specified to the suitable centers. These steps should be repeated until k center cannot be moved [23].

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_i^{(j)} - c_j\right\|^2 \qquad (2)$$

$\| \|$is the criterion of distance between points, cj is the center of jth cluster, xji is the ith point of jth cluster, k is the number of cluster and n is the number of points in each cluster. The beneath algorithm is considered as a basic algorithm for this technique [35, 11]:

- At first k points are selected as data in the centers of clusters.
- Every data sample is specified to the cluster which has the least distance to that data.

- Therefore dependence of all data to one cluster is calculated for every cluster as a new point (the mean of dependant points to each cluster).
- Second and third steps are repeated until no changes are happened in the centers of clusters.

5 KDDCUP99 DATA SET

KDDCUP99 data are collected based on DARPA innovation in 1998 for Intrusion Detection System (IDS) designers that are used in several investigations to find the attacks and intrusion [36]. These data are simulated in seven weeks to intrusion detection, KDDCUP99 data have 41 properties which are divided to 4 parts [37]:

Fundamental Properties: the basic properties are obtained from differential of packet without the investigation of useful load for transmission.

Content: knowledge in this case is used for evaluation of useful load for transmission in TCP packets and involves failed attempts to log in system.

Traffic property based on time: these features are designed to get properties that are happened in more than two seconds continuously. A sample of these features shows the number of connections to the host.

Traffic property based on the host: It uses historical window to estimate the number of connections instead of time. Also, it is designed to assess the extent attacks that are happened in more than two seconds. In international knowledge discovery and data mining only 10% KDD of dataset is used for training purposes [38].

One of the fundamental properties of this set is the numerous numbers of data that leads to more exact. Matlab software is useful software to simulate data, so all data sets are changed into a numeric string, and then proportional to each type of attack the related fields are studied.

6 RESULTS AND DISCUTION

Indeed we investigate K-Means and Fuzzy K-Means algorithms in order to detect intrusion in systems. According to Fig 1., K-Means algorithm acts better than Fuzzy K-Means in 66% of DOC attacks. In this paper k-means and fuzzy k-means methods have been used to identify the type of DoS attacks, k-means algorithm quasi code is shown in Fig 2 and Fuzzy k-means algotihm quasi code shown in Fig3:

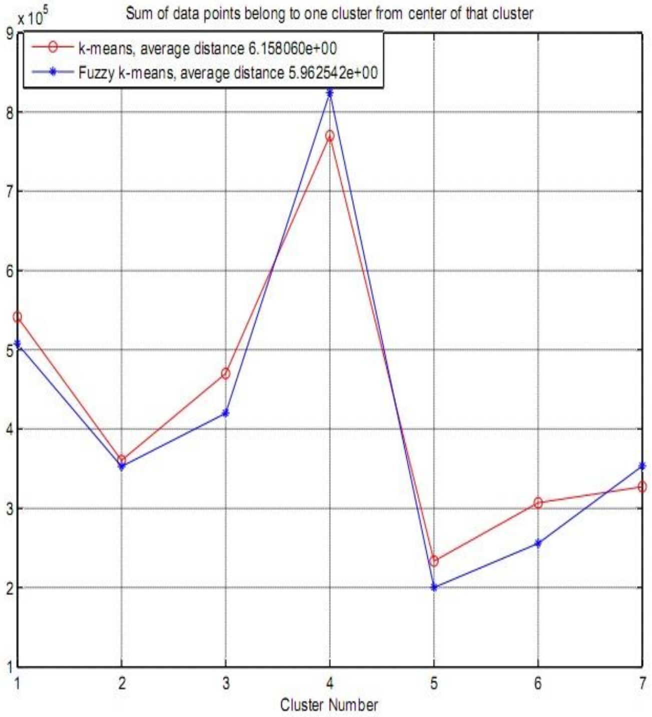
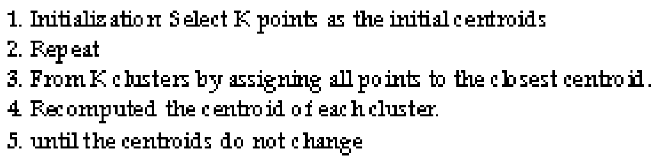


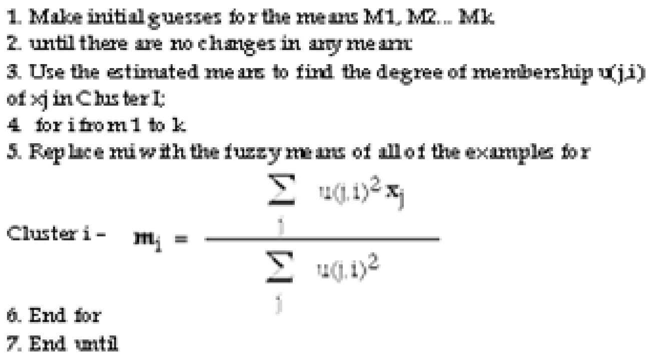
Fig.1. Comparison between the Results of these algorithms

There is a flowchart like Fig 3 about Fuzzy k-means algorithm:



```
1. Initialization: Select K points as the initial centroids
2. Repeat
3. From K clusters by assigning all points to the closest centroid.
4. Recomputed the centroid of each cluster.
5. until the centroids do not change
```

Fig. 2 K-means algorithm quasi code



```
1. Make initial guesses for the means M1, M2... Mk
2. until there are no changes in any mean:
3. Use the estimated means to find the degree of membership u(j,i) of xj in Cluster I;
4. for i from 1 to k
5. Replace mi with the fuzzy means of all of the examples for
Cluster i - $m_i = \frac{\sum_j u(j,i)^2 x_j}{\sum_j u(j,i)^2}$
6. End for
7. End until
```

Fig. 3.Fuzzy k-means algorithm quasi code

In this paper, we have used a kddcup99data set. This data set has 41 properties [39] which uses several fields based on detection of each attack. This model is used for DoS attacks and normal cases. Fig 1. provides the results which are derived from the comparison between two algorithms for six

DoS type of attacks, this comparison is done using Matlab software; according to the chart k-means algorithm acts better than fuzzy k-means algorithm in 66% of attacks. Whereas kddcup data have some string fields, we changed string fields to numbers in matlab software; in table 1. the specified numbers to each cluster are determined which show the number of clusters:

Normal	Kind of Attacks	Number assigned	Number of data in kddcup 10%
	Normal	1	97278
DoS	Back	2	2203
	Land	3	21
	neptune	4	107201
	Pod	5	264
	Smurf	6	280790
	Treadrop	7	979

Table 1. Resluts of Attack's types and Clusters

7 CONCLUSION AND FUTURE WORK

Clustering is a new science that work and study is ongoing in this field because it is considered a lot in different science as a solution. In recent years this method is optimized and the results of optimization are provided as papers. The goal of optimization is obtaining to the minimum number of replicates and clusters with the most similar members. In this paper a comparison is done between two common algorithms in order to recognize the DoS attacks. Finally, the results derived from K-means algorithm are better to identify these kinds of attacks. We should mention that these results are not exactly true and by choosing different fields to study; it is clear that Fuzzy k-means acts better than k-means.

REFERENCES

[1] Pormohseni, Review and identify the computer network intrusion detection systems, 2011 (Language in Persian).

[2] R.Heady, G. Luger, A. Maccabe, M. Sevilla." The Architecture of a Network-level Intrusion Detection System", Technical report, CS90-20. Dept. of Computer Science, University of New Mexico, Albuquerque, NM 87131.pp:1-18, 1990.

[3] K.Scarfone and p.Mell, Guid to intrusion detection and prevention systems (idps), National Institude of Standard and Technology, Special publication800-94, page 127, 2007.Availabel:http://csrc.nist.gov/publications/nistpubs/800-94/SP800-94.pdf, Last Available: 23.08.2012.

[4] Bro IDS homepage, Available: www.bro-ids.org, Last Available: 23.07.2012.

[5] A. A.Ghorbani, W. Lu, M.Tavallaee, Network Intrusion Detection and Prevention: Concepts and Techniques, Springerpublisher, pages 234, 2009.

[6] J.P Anderson,"ComputerSecurity Threat monitoring and surveillance, (1980), Availabel: http://csrc.nist.gov/publications/history/ande80.pdf, Last Availabel: 05.08.2012.

[7] A.hamidi,M.rezai, Introduction to Intrusion Detection System (Part I), Technical report,MashadUniversity,Iran,(language in Persian)

[8] C.Kruegel, F.Valeur, G.Vigna,"Intrusion Detection and Correlation challenges and Solution" University of California, Santa Barbara, Springer Science USA, 2005.

[9] Vance Faber,"Clustring and the Continuous K-means Algorithm", Los Almas since Number22, pp: 138-144, 1994.

[10] M.Ghasemi, M.Khanghandi, The application of fuzzy logic in Algvkhvshh recognition scheme, Arak, Iran, 2009. (language in Persian)

[11] K.K.Bharti,S.Shukla, S Jaim,"Intrusion Detection using Clustering",special Issue of IJCCT2010 for International Confrance [ACCTA-2010] , Vol 1,Issue2,3,4, pp:158-165, 3-5Agust 2010

[12] CERT Advisory CA-96.26, Availabe: http://www.cert.org/ftp/cert_advisories/CA-96.26.ping. December 16, 1996, Last Availabe: 05.08.2012.

[13] R.K.C. Chang, "Defending Against Flooding-Based, Distributed Denial-of-Service Attacks: A Tutorial", IEEE Communication Magazine, Vol 40, No.10, pp: 42-51, 2002.

[14] CERT Advisory CA-98.01. Availabe: http://www.cert.org/ftp/cert_advisories/CA-98.01.smurf. January 5, 1998, Last Availabe: 05.08.2012.

[15] CERT Advisory CA-97.28Availabe: http://www.cert.org/ftp/cert_advisories/CA-97.28.Teardrop_Land. December 16, 1997, Last Availabe: 05.08.2012.

[16] S.Garfinkel, G.Spafford, A. Schwatz, PracticalUNIX and internet security, OReilly and Associates, Sebastopol, CA, USA, page: 988, 2003.

[17] K. Kendall," A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems", Bachelor of Science in Computer Science and Engineering and Master of Engineering in Electrical Engineering and Computer Science, pages:124, June 1999

[18] Anonymous. "Maximum Security: A Hacker's Guide to Protecting Your Internet Site and Network", Chapter 15,

pp.359-362.Sams.net, 201 West 103rd Street, Indianapolis, IN, 46290.1997.

[19] CERT Advisory CA-96.12, Available http://www.cert.org/ftp/cert_advisories/CA-96.12.suidperl_vul. June 26, 1996, Last Availabe: 05.08.2012.

[20] CERTAdvisoryCA-93.18, CA-95:12, Availabel: http://www.cert.org/ftp/cert_advisories/CA95:12 .sunloadmodule.vul. September 19, 1997, Last Availabe: 05.08.2012.

[21] DE Denning, “An intrusion-detection model”, IEEE Transactions on software engineering, pp: 222-232., 1987.

[22] S. Antonatos, K.G. Anagnostakis, and E.P. Markatos,” Generating realistic workloads for network intrusion detection systems”, ACM SIGSOFT Software Engineering Notes 29, No. 1, pp: 207–215, 2004.

[23] M. Sebring, E. Shellhouse, M. Hanna, and R. Whitehurst, “Expert systems in intrusion detection: A case study”, 11th National Computer Security Conference, pp: 74–81, 1988.

[24] T. Lunt, R. Jagannathan, R. Lee, S. Listgarten, D. Eclwards, P. Neumann, H. Javitz, and A. Valdes, “IDES: The Enhanced Prototype. A RealTime Intrusion Detection Expert System”Academic report, SRI-CSL-88-12, October 1988. Availabel: http://www.csl.sri.com/papers/1sri/1sri.pdf, Last Availabel: 05.08.2012.

[25] D. Anderson, T. Frivold, and A. Valdes, “Next-generation intrusion detection expert system” (NIDES): A summary, SRI International, Computer Science Laboratory, 1995.Availabel: http://www.thc.org/root/docs/intrusion_detection/nids/NIDES-Summary.pdf , Last Availabel: 02.08.2012.

[26] D. Anderson, T.F. Lunt, H. Javitz, A. Tamaru, and A. Valdes, “ Detecting unusual program behavior using the statistical component of the Next-generation Intrusion Detection Detection Expert System (NIDES)”, SRI International, Computer Science Laboratory, 1995.Availabel: http://www.sdl.sri.com/papers/5sri/5sri.pdf, Last Availabel: 02.08.2012.

[27] W. Lee, S. J. Stolfo, and K. W. Mok,” A data mining framework for building intrusion detection models”, 1999 IEEE Symposium on Security and Privacy, pp: 120–132, May 1999.

[28] W. Lee, S.J. Stolfo, “Data mining approaches for intrusion detection”, 7th USENIX Security Symposium,Vol: 7, PP: 6-6, 1998.

[29] W. Lee, S.J. Stolfo, and K.W. Mok, “Mining audit data to build intrusion detection models”, 4thInternational Conference on Knowledge Discovery and Data Mining, AAAI Press, pp: 66–72, 1998.

[30] S.E. Smaha, Haystack, “An intrusion detection system”, Aerospace Computer Security Applications Conference, pp: 37–44, 1988.

[31] A.Ph. Porras, P.G. Neumann, “Emerald: Event monitoring enabling responses to anomalous live disturbances”, Proceedings of the National Information Systems SecurityConference, pp: 353-365, 1997.

[32] P. Uppuluri, R. Sekar, “Experiences with specification-based intrusion detection”, Proceedings of Recent Advances in Intrusion Detection, 4th International Symposium, (RAID2001) (Davis, CA, USA) (W, L. M Lee, and A. Wespi, eds.), Lecture Notes in Computer Science, Springer-Verlag Heidelberg, pp:172–189 ,October 2001.

[33] C.Ko, P.Brutch, J.Rowe, G.Tsafnat, K. Levitt, “System health and intrusion monitoring using a hierarchy of constraints”, Proceedings of Recent Advances in Intrusion Detection, 4th International Symposium, (RAID 2001) (Davis, CA, USA) (W, L. M Lee, and A. Wespi, eds.), Lecture Notes in Computer Science, Springer-Verlag Heidelberg, pp. 190–203, October2001.

[34] B James, E. Robert, F.William, “The Fuzzy C-Means Clustering Algorithm “, Computers & Geosciences, Vol: 10, No. 2-3, pp. 191-203, 1984.

[35] Z. HUANG,”Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values”, Kluwer Academic Publishers. Manufactured, Data Mining and Knowledge Discovery 2, pp: 283–304 , 1998.

[36] 1999 DARPA Intrusion Detection Evaluation Plan, Availabel: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/docs/id99-eval-ll.html, Last Availabel: 02.05.2012.

[37] H. G.Kayacık, A. N.ZHeywood, M.I.Heywood,” Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets”, Dalhousie University, Faculty of Computer Science, 6050 University Avenue, Halifax, Nova Scotia. B3H 1W5, Availabel: http://web.cs.dal.ca/~zincir/bildiri/pst05-gnm.pdf, Last Availabel: 02.05.2012.

[38] UCI KDD ArchiveAvailabel:http://kdd.ics.uci.edu, Last Availabel: 23.02.2012.

[39] Knowledge discovery in databases DARPA archive. Task. Description, Availabel: http://www.kdd.ics.uci.edu/databases/kddcup99/task.html, Last Availabel: 02.05.2012.

Farhad Soleimanian Gharehchopogh received his B.Sc. degree in Computer Engineering from Islamic Azad University, Shabestar, Iran. He then received his M.Sc. degree in Computer Engineering in from Cukurova University, Adana, Turkey. Since, he has been working to ward the Ph.D degree in Computer Engineering at Hacettepe University in Ankara, Turkey. Heworks as a lecturer in the Department of Computer Engineering in Scince and Research Branch, Islamic Azad University, West Azerbaijan, Iran. His research interests include Machine Learning, Operating Systems, Software Cost Estimation, Natural Language Processing, and Bioinformatics. For more information please visits www.soleimanian.com

Neda Jabbari received his B.Sc. PNU University, Hashtrood, Iran. Since, she has been working toward the M.Sc degree in Computer Engineering department, Scince and Research Branch, Islamic Azad University, West Azerbaijan, Iran. Her research area is a Machine Learning, Networks and Artificial İntelligents. For more information: nedajabbari.nj@gmail.com

Zeinab Ghaffari Azar received his B.Sc. PNU University, Urmia, Iran. Since, she has been working toward the M.Sc degree in Computer Engineering department, Scince and Research Branch, Islamic Azad University, West Azerbaijan, Iran. Her area is a Machine Learning, Networks and Artificial İntelligents. For more information: z.ghaffari.azar@gmail.com