Anomaly Detection Using Agglomerative Hierarchical Clustering Algorithm

[1]Fokrul Alom Mazarbhuiya, [1]Mohammed Y. AlZahrani and [2] Lilia Georgieva
1 Department of Information Technology
College of Computer Science & IT
Albaha University, Albaha, KSA
2 School of Mathematical & Computer Sciences
Heriot Watt University, UK
fokrul_2005@yahoo.com
imohduni@gmail.com
L.Georgieva@hw.ac.uk

Abstract: Intrusion detection is becoming a hot topic of research for the information security people. There are mainly two classes of intrusion detection techniques available till today namely anomaly detection techniques and signature recognition techniques. Anomaly detection techniques are becoming area of interest for the researchers and new techniques are developing every day. However, no techniques have been found to be absolutely perfect. Clustering is an important data mining techniques used to find patterns and data distribution in the datasets. It is mainly used to identify the dense regions and sparse regions in the datasets. The sparse regions were often considered as outliers. There are several clustering algorithms developed till today for the discovery outliers in the datasets. K-means algorithm. K-medoids algorithm, CLARA, CLARANS, DBSCAN, ROCK, BIRCH, CACTUS etc. are some of the popular algorithms dealing with numeric datasets, categorical datasets, spatial datasets or hybrid datasets. Clustering techniques have been successfully used in detection anomaly in dataset. The techniques were found to be useful in the design of a couple of anomaly based Intrusion Detection Systems (IDS). But most of clustering techniques used for these purpose have taken partitioning approach. In this article, we propose a different clustering algorithm for the anomaly detection on network datasets. Our algorithm is an agglomerative hierarchical clustering algorithm which tries to find clusters on the dataset consisting of both numeric and categorical datasets i.e. hybrid datasets. For this purpose, we define a suitable similarity measure on both numeric and categorical attributes available on any network datasets.

Key-words: Anomaly Detection, Network Data, Information Security, Outlier Analysis, Data Instance, Multi-dimensional Space, Cardinality of a Set, Euclidean Distance, Canberra Metric, Similarity of Data Instance pair, Similarity of Clusters pair, Merge Function.

1. Introduction

It has become inexpensive to store, transfer and process data. Huge amount of data is accumulated every day. The data contains potentially useful information. The interpretation of such large amounts of data and extracting the valuable knowledge from it is a challenging task. The term *data mining* is coined to describe the methods and techniques used for the task. Data mining is defined as the method of extracting non-trivial and previously unknown information or patterns in the data. There are several methods of data mining developed till today and clustering is one of them. Clustering is a data mining technique based on unsupervised learning and is used to identify data distribution and hidden patterns in data. Clustering is mainly used to find out dense and sparse regions in the dataset. There are primarily two broad approaches to clustering, i.e. *partitioning* and *hierarchical* approach. Hierarchical clustering is mainly two types i) agglomerative clustering and ii) divisive clustering. Various methods and algorithms have already been developed till today for clustering different types of data. In [1], authors have discussed numerical data clustering using distance function. In [2, 3, 4], authors have presented methods for clustering categorical data. In [5] spatial data clustering is discussed in detail.

In this paper we study the use of an *outlier* for intrusion detection. An outlier is a data point which does belong to any cluster. Finding outliers from large datasets is an interesting data mining problem and it has applications in fraud detection, anomaly detection, intrusion detection. In [6, 7], the methods for outliers detection are discussed. Intrusion is an attempt to compromise the integrity, confidentiality or availability of resources by accessing in an

illegitimate way. There is a wide range of activities which fall under this which includes denial of service, probe, remote to local, user to root. Proposed methods of intrusion detection in [8, 9] use fuzzy clustering. Existing intrusion detection techniques may be broadly classified into *anomaly detection techniques* and *signature recognition techniques* [10, 11]. An *anomaly detection technique* is used for finding intrusions or misuse of networks and computers by monitoring system activities and grouping it as normal or anomalous. The Intrusion Detection System (IDS) developed using the above-mention technique is termed as anomaly-based intrusion detection system. In [12], authors have proposed a traffic anomaly detection method using k-means clustering algorithm and weighted Euclidean distance. A statistical method based on dimensional reduction and pattern extraction for intrusion detection in wireless network is proposed in [13]. An anomaly detection method based on fuzzy c-means clustering is considered in [14]. The method is hybrid, it uses both numeric and categorical attributes. The distance formula is defined in terms of the distance on numeric attributes and dissimilarity formula on categorical attributes.

In this paper, we propose an agglomerative hierarchical clustering algorithm for anomaly detection. First, we define the similarity measure of two data instances as a weighted aggregate of their numeric attributes and their categorical attributes. The similarity measure on the numeric attributes is defined in terms of Canberra metric [15, 16, 17, 18] and that on categorical attributes is defined in terms of ratio of the cardinality of intersection of two data instances to that of the union of the same data instances. Next, we define the similarity of a pair of clusters C1 and C2 (having m1 and m2 data instances respectively), as the weighted aggregate of the similarity measure of the numeric attributes and that of the categorical attributes of the data instances of $C1$ and $C2$. The similarity measure [19, 20] of categorical attributes of $C1$ and $C2$ is defined as sum of the ratios of the cardinality of the pair-wise of intersections of the data instances of both $C1$ and $C2$ to the cardinality of their pair-wise union, then divide the factor by the product of the number data instances of $C1$ and that of $C2$ and then subtracting the whole factor from 1. The range of the value of the similarity measure ($C1$, $C2$) is between 0 to 1. For exactly similar data instances/clusters the value of the similarity measure will be 0 and for exactly dissimilar data instances/clusters its value is 1, which is required in our algorithm. We define a merge function in terms of the similarity measure described above. Finally, an agglomerative hierarchical clustering algorithm for intrusion detection is presented in this paper. The algorithm is similar to one discussed in [21]. It supplies all clusters along with a set of outliers. The extracted outliers are considered as intruders.

The paper is organized as follows. In section-2, we briefly discuss the works related to our work. In section-3, we discuss the problem statement. The proposed algorithm for intrusion detection is discussed in section-4. Finally, we conclude our paper with a brief conclusion given in section-5.

2. Related Work

Data mining has received a lot of attention to the researchers for its widespread uses. There are several methods of data mining available till today namely clustering, association rules mining, classification, sequential patterns mining. Out of these techniques, clustering is one of the most popular. Clustering is an unsupervised data mining technique mainly used for finding dense and sparse regions in the dataset. There are primarily two broad directions of clustering namely partitioning method and hierarchical method. In the partitioning approach data instances are divided into a predefined number clusters based on some criteria. In the hierarchical approach, the method sought to build a hierarchy of clusters one within another. The hierarchical clustering approaches are mainly of two types: i) agglomerative clustering techniques and ii) divisive clustering techniques.

Clustering algorithms for different data type have been developed: for example, clustering of numeric data is discussed in [1] where distance function is used as a criterion. In [2, 3, 4], authors have presented methods for clustering categorical data. Spatial data clustering is discussed in [5]. Using wavelet transform, a method called WaveCluster is discussed for clustering spatial data [22]. Discovering outliers from large datasets is an interesting data mining problem that has been discussed [6, 7].

Intrusion is an activity which can compromise the integrity, confidentiality or availability of resources by accessing in an illegitimate way. Intrusion activities and the methods of their detection is discussed in [8, 9]. In [8, 9], the authors proposed a method based on fuzzy c-means algorithm for intrusion detections. In [12], authors proposed a method based on K-means algorithm for traffic anomaly detection using weighted Euclidean distance. In [13], authors discussed a statistical approach based on dimensional reduction and pattern extraction for intrusion detection in wireless network. A method based on K-means for anomaly detection in hybrid data is discussed in [14]. Both numeric attributes and categorical attributes available in the network datasets are discussed in [14]. In [16], an algorithm for clustering categorical data is proposed. The algorithm uses a similarity function as ratio of the cardinality of the intersection of attributes value to that of the union of the same. In [19, 20, 23], authors have used

similar measure for clustering categorical data and documents. In [23], the authors have defined the similarity measure on the clusters in terms fuzzy sets. In [18], authors have described Canberra metric based intrusion detection system. A Genetic algorithm based techniques for clustering mixed data is discussed in [24]. An intrusion detection algorithm based on the analysis of usage data coming from multiple partners in order to reduce the number of false alarms is discussed in [25]. An intrusion detection method based on data mining techniques is also discussed in [26]. A method for network intrusion detection based on data mining techniques is discussed in [27, 28]. In [29], the authors have discussed an intrusion detection systems based on pattern recognition techniques. In [30], the authors have proposed an intrusion detection method based on fuzzy association rules and fuzzy frequency episodes. In [31], the authors have conducted a comparative study on different anomaly detection schemes in network intrusion detections. In [32], authors have proposed a method which facilitates the investigation of huge amounts of intrusion detection alerts by a specialist. Their approach has made use of process mining techniques to find attack strategies observed in intrusion alerts. In [33], the authors have extended the work of [32] and have proposed an alert correlation approach with emphasis on visual models to assist network administrators in the investigation of multistage attack strategies. In [34], the authors have discussed a method of reduction of intrusion detection alarm based on root cause analysis and clustering. In [35], a method finding alert correlation based on frequent itemset mining is discussed. In [36], the authors have proposed a method of mining for causal knowledge automatically based on the Markov property for identifying multi-stage attack. In [37], the authors have discussed an intrusion detection method using Support Vector Machine which reduces the time to build classification model and increases accuracy. An intrusion detection system based on classification is discussed in [38]. An intrusion detection method based on artificial neural network is discussed in [39, 40]. In [41], the authors have proposed a real-time framework for identifying multi-stage attack scenarios from alerts generated by IDS which sequential pattern mining. In [42], the authors have proposed an SVM-based intrusion detection system, which combines a hierarchical clustering algorithm and SVM. They have used the dataset KDD Cup 1999 to check the efficacy of their system by comparing with its performances with others intrusion detection systems.

3. Problem Statement

Here our aim is to cluster the data having both numeric and categorical attributes. For example, each connection instance of KKD Cup 1999 network data has 41 properties with 3 flag properties and 38 numeric properties. However, most of the current works were directed towards numeric properties and a very little works were done in dealing categorical properties.

If we consider both the numeric and categorical attributes of the sample data, then we have to treat it as a hybrid data or mixed data. The similarity measure must be defined in terms of numeric and categorical attributes. As our approach is agglomerative hierarchical, the distance function [14, 24]**,** defined in terms Euclidean distance and dissimilarity cannot be used directly. Instead we take another measure call *similarity measure*. In below we describe some of the definitions used in the paper.

Definition1 (Similarity between a pair of connection instances)

Let A and B are two data instances with dimension n, where first k- attributes are numeric and rest of the (n-k) are categorical attributes. The k-numeric attributes of A, B are denoted by A^n, B^n respectively, and (n-k) categorical attributes are denoted by A^c, B^c respectively. The similarity measure between A and B, denoted by $S(A, B)$ is given by the expression.

$$S(A,B) = \frac{k.S_1+(n-k)S_2}{n} \quad (1)$$

Where $S_1=S(A^n, B^n)$ = similarity of A and B defined on the k- numeric attributes. Now, to find the expression for the similarity between two numeric variables, we proceed as follows.

Let $x = (x_1, x_2,..,x_k)$ and y = $(y_1, y_2,..,y_k)$ two k-dimensional vectors. Then the *Canberra metric* [15, 16, 17, 18], $d(x,y)$ is given by the formula.

$$d(x,y) = \sum_{i=1}^{k} \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (2)$$

The range of the above *Canberra metric* [15, 16, 17, 18] is [0, *k*]. To make the range, [0, 1] we divide the equation (2) by *k*. Therefore, we get new formula for *Canberra metric*, which is our similarity measure S_1 and is expressed below.

$$S_1 = \frac{1}{k}\sum_{i=1}^{k}\frac{|x_i - y_i|}{|x_i| + |y_i|} \tag{3}$$

Again $S_2 = S(A^c, B^c)$ = similarity of *A* and *B* defined on rest of the (*n*-*k*) attributes. For the similarity S_2, we use a formula quite similar to the formula given in [19, 20, 23]. The formula is given below.

$$S_2 = 1 - \frac{|A^c \cap B^c|}{|A^c \cup B^c|} \tag{4}$$

Obviously the value of S_2 will be the ranging from 0 to 1. For, exactly similar categorical values of both the data instances, $A^c = B^c$, $S_2 = 0$, and for exactly, different data instances, $A^c \cap B^c = \phi$, so that $|A^c \cap B^c| = 0$, thus $S_2 = 1$

With help of the equation (3) and equation (4), the equation (1) will give us the similarity value for the two data instances *A* and *B*. Here we write the formula as a weighted aggregate of the both numeric and categorical variables so that both the attributes will have proportional contribution on the similarity function and its value will be ranging from 0 to 1. Using the equation (1), we say that the two data instances *A* and *B* are similar if and only if *S*(*A*, *B*) is less than or equal to a pre-assigned threshold value, otherwise they are dissimilar. Obviously they will be precisely similar for the value 0 and precisely dissimilar for the value 1.

Definition 2 (Similarity between a pair of clusters)
Let $C_1 = \{A[i];\ i = 1, 2, \ldots, m_1\}$ and $C_2 = \{B[j];\ i = 1, 2, \ldots, m_2\}$ be two clusters having m_1 and m_2 data instances respectively, the similarity function $S(C_1, C_2)$ is given by the formula

$$S(C_1, C_2) = \frac{k.S_1(\overline{A^n}[i], \overline{B^n}[j]) + (n-k).S_2(A^c[i], B^c[j])}{n} \tag{5}$$

Where $S_1(\overline{A^n}[i], \overline{B^n}[j])$ = is the *Canberra metric* [15, 16, 17, 18] defined on the arithmetic means of the numeric variables of the data instances *A*[*i*] and *B*[*j*] of C_1 and C_2 respectively and is expressed as

$$S_1(\overline{A^n}[i], \overline{B^n}[j]) = \frac{1}{k}\sum_{i=1}^{k}\frac{|\overline{x_i} - \overline{y_i}|}{|\overline{x_i}| + |\overline{y_i}|} \tag{6}$$

Where $\overline{x_i}$ = is the arithmetic mean of the variable x of *A*[*i*], *i*=1, 2,…,m_1, and $\overline{y_i}$ = is the arithmetic mean of the variable y of *B*[*i*], *i*=1, 2,…,m_2.

Similarly, $S_2(A^c[i], B^c[j])$ = is the similarity defined on the categorical attributes of the above-mentioned data instances of C_1 and C_2 and $S_2(A^c[i], B^c[j])$ is expressed by the formula

$$S_2(A^c[i], B^c[j]) = 1 - \frac{1}{m_1.m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{|A^c[i] \cap B^c[j]|}{|A^c[i] \cup B^c[j]|} \tag{7}$$

Here $A^c[i] \cap B^c[j]$=pair-wise intersections of data instances of C_1 and that of C_2. Obviously there will be $m_1.m_2$ number of the ratios $\frac{|A^c[i] \cap B^c[j]|}{|A^c[i] \cup B^c[j]|}$ and each having maximum value 1. The lowest value of

$S_2(A^c[i], B^c[j]) = 1 - \frac{m_1 m_2}{m_1 m_2} = 0$. Similarly, the highest value of $S_2(A^c[i], B^c[j]) = 1$.Thus the values of $S_2(A^c[i], B^c[j])$ ranges from 0 to 1.

With the help of equation (6) and equation (7), equation (5), represents the similarity of clusters or inter-clusters similarity $S(C_1, C_2)$. Obviously its value ranges from 0 to 1. For exactly similar cluster-pair its value is 0 and for exactly different cluster-pair its value is 1.

Definition 3 (Merger of Cluster)

Let C_1 and C_2 be the two clusters having m_1 and m_2 data instances respectively. Let C be the cluster formed by merging C_1 and C_2. Then the *merge*() function is defined as $merge(C_1, C_2)=C1 \cup C2$, if and only if $S(C_1, C_2) \leq \sigma$, a pre-defined threshold.

4. Proposed Algorithm

At the beginning of the clustering process, each data instance is allocated to a separate cluster. Thereafter for every pair of data instances the *similarity measure* is computed and then the *merge* function is used to obtain larger clusters if and only the *similarity* value is found to be within certain limit (the definition of *similarity measure* and *merge* function is given in section 3). At any level, for any two clusters say C_1 and C_2 (having m_1 and m_2 data instances respectively), the *similarity* value is calculated using the formula given in section-3, to check whether they can be *merged* or not. If the *similarity* value is found to be within a certain pre-determined threshold then C_1 and C_2 are *merged* using *merge* function to form a new bigger cluster. The process of *merging* of clusters continuous till no *merger* is possible or there is only one cluster at the top. In bellow we present the pseudo code for the proposed algorithm.

Algorithm DataInstanceClustering(n,σ)
Input: The number of data instances $A[i]$; i=1,2,…*n*, and threshold σ
Output: A set of cluster *S*
Step1. The set of clusters *S*, where each cluster *C* of *S* having one data instance $A[i]$
Step2. If for any cluster $C_1 \in S$ and $S(C_1,C) \leq \sigma$, then $merge(C_1,C)$ to form a new cluster C_2 consisting of C_1 and *C*.
Step3. Remove C_1 from *S*.
Step4. Continue Step2 and Step3 till no merger of clusters is possible.
Step5. Return *S*
Step6. Find all outliers from *S*.
Step7. Stop

The algorithm supplies the set of clusters *S* of data instances which also includes outliers that means the data instances which do not belong to any larger clusters. The outliers can be used to find anomalies among the data instances. The patterns obtained by the above algorithm can be used in designing any efficient Intrusion Detection System (IDS).

5. Conclusion

The anomaly detection technique is one of the techniques used for developing Intrusion Detection Systems. In this paper, we propose an agglomerative hierarchical clustering algorithm for anomaly detection. For clustering purpose, we define a suitable similarity measure in terms of similarities in both numeric and categorical attributes. The *similarity* on the numeric attributed is defined in terms *Canberra metric* and that on the categorical attributes is defined in terms of the ratio of the cardinality of intersection of two data instances to that of the union of the same data instances Then, we take the weighted average of the both value to find *similarity* of the data instances. Then data instances are merged based the similarity value to find larger clusters. At any level, a pair of clusters is merged based its *similarity* value defined in section-3. The process continues till no cluster merging is possible. The algorithm gives as output, a set of clusters of data instances each cluster having similar type of instances. Obviously any data instance deviating from pattern extracted by the method would be an anomaly.

References

[1] J. A. Hartigan, "Clustering Algorithms", John Wiley & Sons, 1975.

[2] D. Gibson, J. Kleinberg and P. Raghavan, "Clustering categorical data: An approach based on dynamical systems", In Proc. of the 24th Int'l Conf. on Very Large Databases, New York, 311-323, 1998.

[3] R. T. Ng and J. Han, "Efficient and effective clustering methods for spatial data mining", In Proc. of the VLDB Conf, Santiago, Chile, 144-155, 1994.
[4] V. Ganti, J. Gehrke and R. Ramakrishnan, "CACTUS-Clustering categorical data using summaries", In Proc. of the Int'l Conf. on Knowledge Discovery and Data Mining, San Diego, CA, USA, 73-83, 1999.
[5] S. Guha, R. Rastogi and K. Shim; ROCK, "A robust clustering algorithm for categorical attributes", In Proc. of the IEEE Int'l Conf. on Data Engineering, Sydney, 512-521, 1999.
[6] R. Pamula, J. K. Deka, S. Nandi, "An Outlier Detection Method based on Clustering", Proceedings of 2011 Second International Conference on Emerging Applications of Information Technology, February 2011, 253-256, India.
[7] Y. Zhang, J. Liu, and H. Li, "An Outlier Detection Algorithm Based on Clustering Analysis", The Proceedings of 2010 First International Conference on Pervasive Computing, Signal Processing and Applications, Sepember 2010.
[8] D. Sharma, "Fuzzy Clustering as an Intrusion Detection Technique", International Journal of Computer Science & Communication Networks, Vol 1(1), September-October 2011, 69-75.
[9] L. Xie, Y. Wang, L. Chen, and G. Yue, "An Anomaly Detection Method based on Fuzzy C-means Clustering Algorithms", Proceedings of the Second Symposium on Networking and Network Security, China, April 2010, 89-92.
[10] H. Debar, M. Dacier, and A. Wespi. "Towards a taxonomy of intrusion detection systems," Computer Networks, 31, pp. 805-822, 1999.
[11] T. Escamilla, "Intrusion Detection: Network Security beyond the Firewall", New York: John Wiley & Sons, 1998.
[12] G. Munz, S. Li, and G. Carle, "Traffic Anomaly Detection using K-Means Clustering", Allen Institute for Artificial Intelligence, 2007.
[13] N. A. Haldar, M. Abulaish, and S. A. Pasha, "A Statistical Pattern Mining Approach for Identifying Wireless Network Intruders", Advances in Intelligent Systems and Computing: Preface, July 2012, 131-140.
[14] X. Linquan, W. Ying, C. Liping, and Y. Guangxue, " An Anomaly Detection Method Based on Fuzzy C-means Clustering Algorithm", Proceedings of the Second International Symposium on Networking and Network Security, April 2010, China, pp. 089-092.
[15] G. N. Lance, and W. T. Williams, "Computer programs for hierarchical polythetic classification ("similarity analysis")". Computer Journal. **9** (1), 1966, pp. 60–64.
[16] G. N. Lance, and W. T. Williams. "Mixed-data classificatory programs I.) Agglomerative Systems". Australian Computer Journal, 1967, pp. 15–20.
[17] T. H. Clifford, and W. Stephenson, "An Introduction to Numerical Classification", Academic Press. New York-San Fransisco – London, 1975.
[18] S. M. Emran, and N. Ye, "Robustness of Canberra Metric in Computer Intrusion Detection", Proceedings of 2001 IEEE Workshop on Information Assurance and Security, US Military Academy, NY, June 2001, pp. 80-84.
[19] M. Dutta, A. K. Mahanta and M. Mazumder, "An Algorithm for Clustering of Categorical Data using Concept of Neighours", *Proc. of the 1st National Workshop on Soft Data Mining and Intelligent Systems*, Tezpur University, India, pp.103-105, 2001.
[20] M. Dutta, and A. K. Mahanta, "An Algorithm for Clustering Large Categorical Databases using a Fuzzy Set based Approach", Proceedings National Workshop on Trends in Advanced Computing, 2006, Tezpur University, India.
[21] F. A. Mazarbhuiya, and M. Y. AlZahrani, "An Efficient Method for Clustering Periodic Patterns", Computing Conference 2017, SAI Conference, London, U. K.
[22] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "WaveCluster: A Multi-Resolution Clustering Approach for Large Spatial Databases, Proceedings of 24th VLDB Conference, Newyork, USA, 1998.
[23] K. Thaoroijam and A. K. Mahanta, "A Fuzzy based Document Clustering Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 151 – No.10, October 2016.
[24] J. Li, XB. Gao, and LC. Jiao, " A GA-based Clustering Algorithm for Large Datasets with Mixed Numerical and Categorical Values [J]", Journal of Electronics & Information Technology,Vol 26(8), 2004, pp. 1203-1209.
[25] S.Sathya Bama, M.S.Irfan Ahmed, A.Saravanan, "Network Intrusion Detection using Clustering: A Data Mining Approach", International Journal of Computer Applications, Vol. 30 (4), September 2011, pp. 14-17.
[26] W. Lee and S. J. Stolfo, "Data Mining Approaches for Intrusion Detection", In 7th conference on USENIX Security Symposium, 1998.

[27] P. Dokas, L. Ertos, V. Kumar, A. Lazarevic, J. Srivastava, and P. N. Tan, “Data Mining for Network Intrusion Detection”, In Proceedings of the NSF Workshop on Next Generation Data Mining, November 2002.
[28] E. Bloedorn, A. D. Christiansen, W. Hill, C. Skorupka, and L. M. Talbot, “Data Mining for Network Intrusion Detection: How to get started”, Technical report, MITRE, 2001
[29] M. Esposito, C. Mazzariello, F. Oliviero, S. P. Romano, and C. Sansone, “Evaluating pattern recognition techniques in intrusion detection systems”, in Proceedings of the 5th International Workshop on Pattern Recognition in Information Systems (PRIS) 2005, May 2005, pp. 144-153.
[30] J. Luo and S. Bridges, “.Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection”, International Journal of Intelligent Systems, Vol. 15(8), pp. 687.704, 2000.
[31] A. Lazarevic, L. Ert¨oz, V. Kumar, A. Ozgur, and J. Srivastava, “A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection’, In Proceedings of the Third SIAM International Conference on Data Mining, May 2003.
[32] S. C. Alvarenga, B. B. ZarpelÃƒÂ£o, S. B. Junior, R. S. Miani, M. Cukier, “Discovering attack strategies using process mining”, in: The Eleventh Advanced International Conference on Telecommunications, AICT 2015, IARIA, 2015, pp. 119–125.
[33] S. C. de Alvarengaa, S. B. Juniora, R. S. Mianib, M. Cukierc, B. B. Zarpelãoa , “Process Mining and Hierarchical Clustering to help Intrusion Alert Visualization”, Computers & Security, 2017.
[34] S. O. Al-Mamory and H. Zhang, "Intrusion detection alarms reduction using root cause analysis and clustering," Computer Communications, 2009, vol. 32, no. 2, pp. 419-430.
[35] S. Lagzian, F. Amiri, A. Enayati and H. Gharaee, "Frequent item set mining-based alert correlation for extracting multi-stage attack scenarios," in Telecommunications (IST), 2012 Sixth International Symposium on. IEEE, 2012, pp. 1010-1014.
[36] F. Xuewei, W. Dongxia, H. Minhuan and S. Xiaoxia, "An approach of discovering causal knowledge for alert correlating based on data mining", in Dependable, Autonomic and Secure Computing (DASC), 2014 IEEE 12th International Conference on. IEEE, 2014, pp. 57-62.
[37] Y. B. Bhavsar, K. C. Waghmare, “Intrusion Detection System Using Data Mining Technique: Support Vector Machine”, International Journal of Emerging Technology and Advanced Engineering, Vol. 3(3), March 2013, pp. 581-586.
[38] R. Wankhede, and V. Chole, “Intrusion Detection System using Classification Technique”, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016, pp. 25-28.
[39] J. Shun, and H. A. Malki, “Network Intrusion Detection Systems Using Neural Network”, ICNC 2008, IEEE Explore.
[40] G. Poojitha, K. N. Kumar, and R.J. Reddy, “Intrusion Detection Using Artificial Neural Network”, Proceedings of ICCCN 2010, IEEE Explore.
[41] F. A. Bahareth, and O. O. Bamasak, “Constructing Attack Scenario using Sequential Pattern Mining with Correlated Candidate Sequences”, The Research Bulletin of Jordan ACM, Volume II(III), pp. 102-108.
[42] S. J. Horng, M. Y. Su, Y. H. Chen, T. W. Kao, R. J. Chen, J. L. Lai, and C. D. Perkasa, “A novel intrusion detection system based on hierarchical clustering and support vector machines”, Expert Systems with Applications Vol.38(1), January 2011, Pages 306-313.