8. Conclusion

In this paper we compared the three main intrusion detection algorithms: FCC, Y-Means and UAD. In FCC a new concept called fuzzy connectedness [10] is introduced. It is used to calculate the similarities among different data instances. Starting with a single or a few seed points in each cluster, all the data points are dynamically assigned to the cluster that has the highest fuzzy connectedness value (strongest connection). Comparing with the frequently utilized Euclidean distance, the new similarity metric is more robust and there is no restriction on the shape of clusters that can be discovered. FCC has also some other advantages over the other previously proposed clustering algorithms: no predefined "cluster width", no threshold for "confidence area", etc. Moreover, this unsupervised learning method can detect not only the known intrusion types, but also their variants. UAD and Y-Means methods detect intrusions based on feature vectors collected from the network, without being given any information about classifications of these vectors. There are two primary advantages of this system over signature-based classifiers or learning algorithms that require labelled data in their training sets. The first is that no manual classification of training data needs to be done. The second is that we do not have awareness about new types of intrusions for the system to be able to detect them. The main requirement is that the data conform to several assumptions. The systems then try to automatically determine which data instances fall in to the normal class and which ones are intrusions. Since no prior classification is required on the training data, and no knowledge is needed about new attacks, we can automate the process of training and create new cluster sets. In addition to that the method can be used for semi– automated detection by helping analysts to focus on portions of the data that are more likely to contain intrusions. From the result of the research comparing with the other two clustering algorithms such as UAD method and Y-Means the Detection Rate of FCC is at least 5% higher. Even though the False Alarm Rate with FCC is also a little higher, these results are very encouraging since higher Detection Rate is more important in most applications.

9. References

[1] B. Mukherjee, L.T. Heberlein, and K.N. Levitt, "Network intrusion detection", IEEE Network, 26–41, 1994.

[2] C. Aggarwal, "On abnormality detection in spuriously populated data streams", Proceedings of 5th SIAM Data Mining, 80-91, 2005.

[3] Markos Markou and Sameer Singh, "Novelty detection: a review – part 2", Signal processing 83, Department of Computer Science UK, pp. 2499 – 2521, 2003.

[4] M. Larbani, and Y. Chen, "A Fuzzy Set Based Framework for Concept of Affinity", Applied Mathematical Sciences, Vol. 3, no. 7, 317–332, 2009.

[5] Y. Jian, and N. Yufu, "Research on initial clustering centres of Fuzzy c-means algorithm and its application to intrusion detection", 2nd conference on environmental science and information application technology, 161-163, 2010.

[6] L. Portnoy, E. Eskin, and S. Stolfo, "Intrusion detection with unlabeled data usin clustering", Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001). Philadelphia, PA: November 5-8, 2001.

[7] F.N.M. Sabri, N.M. Norwawi, and K. Seman, "Identifying False Alarm Rates for Intrusion Detection System with Data Mining", IJCSNS International Journal of Computer Science and Network Security, Vol.11 No.4, 95, 2011.

[8] KDD Cup, 1999 data. University of California, Irvine. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html,1999.

[9]Q. Wang, and V. Megalooikonomou, "A Clustering Algorithm for Intrusion Detection", Data Engineering Laboratory (DEnLab) Department of Computer and Information SciencesTemple University 1805 N. Broad Street, Philadelphia, PA 19122, USA.

[10] A. Rosenfeld, "Fuzzy Digital Topology", Information and Control, vol. 40, pp. 76-87, 1979.