Network Intrusion Detection Using Genetic Clustering

Elizabeth Leon[1], Olfa Nasraoui[1], and Jonatan Gomez[2]

[1] Department of Electrical & Computer Engineering, The University of Memphis
[2] Universidad Nacional de Colombia
{eleon, onasraou, jgomez}@memphis.edu

Abstract. We apply the Unsupervised Niche Clustering (UNC), a genetic niching technique for robust and unsupervised clustering, to the intrusion detection problem. Using the normal samples, UNC generates clusters sumarizing the normal space. These clusters can be characterized by fuzzy membership functions, that are later aggregated to determine a level of normality. Anomalies are identified by their low normality levels.

1 Introduction

Clustering [1] has been applied successfully to the Intrusion Detection Problem (IDP), by generating a set of clusters that can characterize the normal class using the normal samples. The Unsupervised Niche Clustering (**UNC**) is a robust and unsupervised clustering algorithm that uses an evolutionary algorithm to find clusters (using a robust density fitness function), with a niching strategy for maintaining the niches (candidate clusters) [2,3]. In this paper, we combine the UNC with fuzzy sets theory for solving some IDPs [4]. We associate to each cluster evolved by the UNC (cluster center c and scale σ) a membership function that follows a Gaussian shape, $\mu(x) = e^{\left(-\frac{d(x,c)^2}{2\sigma^2}\right)}$. Such function will define the normalcy level of a data sample. Then, the normal class is defined by the fuzzy-union set (max-OR) of all the clusters generated (C). Thus, a data sample x is considered normal with a $\mu_{normal}(x) = max\{\mu_i(x)|\forall i = 1, 2, ..C\}$ degree.

2 Experimentation

Tests were conducted on a reduced version of the KddCup'99 data set (21 features) after applying a Principal Component Analysis (PCA) to the non-zero numerical features. We used 5000 normal samples as training data set, while 40% of the full data set was used for testing. Table 1 shows the performance reached by our approach, while Figure 1 shows the ROC curves generated.

3 Conclusions

In this paper, the UNC algorithm [3] was applied to the intrusion detection problem (KddCup'99 data set). A fuzzy characterization of the model generated

K. Deb et al. (Eds.): GECCO 2004, LNCS 3103, pp. 1312–1313, 2004.

by UNC was proposed. Our results show that a fuzzy analysis increases the performance reached by our approach while PCA reduces the complexity of the data set and further improves the performance of our approach [4].

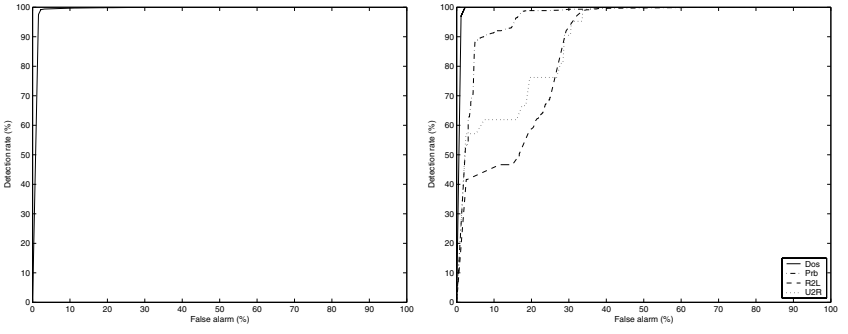

Fig. 1. ROC curve generated by our approach. Left: all attacks, Right: per attack.

Table 1. Performance of the proposed approach on the KddCup'99 data set.

	FULL	**DOS**	**PRB**	**R2L**	**U2R**
Detection Rate (%)	99.20	95.9	93.9	98.6	90.9
False Alarm Rate (%)	2.20	1.0	12.2	28.6	20.6

Acknowledgments. This work is supported by National Science Foundation CAREER Award IIS-0133948 to O. Nasraoui.

References

1. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. NY: Wiley Interscience, 1973.
2. O. Nasraoui, E. Leon, and R. Krishnapuram, "Unsupervised niche clustering: Discovering an unknown number of clusters in noisy data sets," in *Evolutionary Computing in Data Mining*, Invited chapter, A. Ghosh and L. C. Jain, Eds, Springer Verlag, 2004.
3. O. Nasraoui and R. Krishnapuram, "A novel approach to unsupervised robust clustering using genetic niching," in *Proceedings of the Ninth IEEE International Conference on Fuzzy Systems*, pp. 170–175, 2000.
4. E. Leon, O. Nasraoui, and J. Gomez, "Anomaly detection based on unsupervised niche clustering with application to intrusion detection," in *Proceedings of the 2004 Congress on Evolutionary Computation*, 2004.