User Behavior Analysis in Campus Area Networks through Kohonen Self Organizing Feature Maps

Nelson V. Cruz-Hernández
Álvaro I. Parres-Peredo
Departament of Electronics, Systems and Informatics
ITESO
Tlaquepaque, Jalisco, 45604, Mexico

Abstract - This works presents a novel implementation for Kohonen Self Organizing Feature Maps (SOM) suitable for user behavior analysis in Campus Area Networks (CAN). SOM algorithm works under competitive and unsupervised learning approach, making it perfect for clustering and feature selection tasks. SOM is used to create behavior clusters through the analysis of user network traffic, in fixed time intervals. To demonstrate the feasibility of SOM, resulting users clusters are merged and evaluated against new matching vectors. Experimental results, show that an user can be identified based in it's network traffic since it generates a match in a previous created cluster.

Introduction

A network intrusion attack can be any use of the network that compromises its stability or the security of the information that is stored on devices connected to it [1]. The weakest element in network security is an user, that may act as an active or passive attacker [2]. User behavior is a wide open studied topic using Machine Learning Classification and Clustering algorithms [3 - 6], focused on detecting popularity of applications, user session duration, and users distribution across access points [7]. This work makes two assumptions: It is possible to identify an user through it's network traffic data and user normal network traffic differs from an attacker network traffic. Self Organizing Maps cluster algorithm is used to obtain a pattern of user behavior. Obtained patterns define specific user behavior and mark differences between normal and attacker users.

Kohonen Self Organizing Maps

Self Organized Maps (SOM) algorithm works as an unsupervised learning clustering approach, where training is entirely data-driven and no target results for the input data vectors are provided. It also provides a topology to preserve mapping from high dimensional space to nodes that form a two-dimensional lattice that represents high dimensional space onto a plane in which map units are grouped by it's features values similarity. Each node has a specific topological position and contains a vector weights (features) of the same dimension as the input vector [8].

Learning algorithm of Conventional SOM [8]

1) Initialize the map using random vectors.

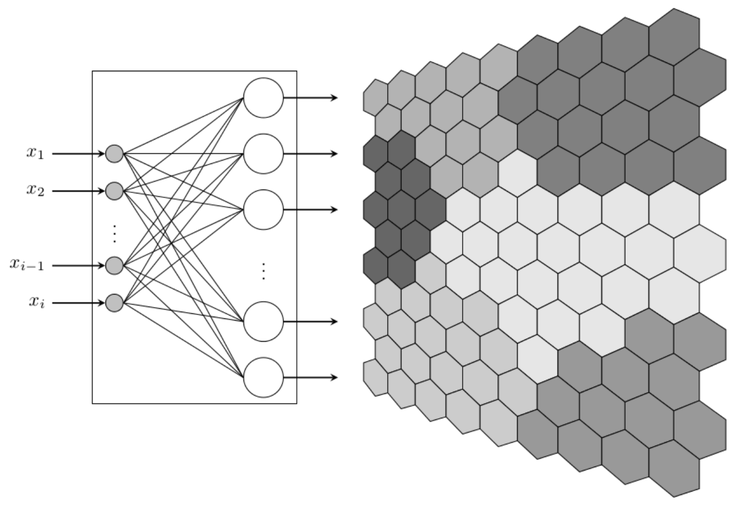
2) Searching for the winner unit
Select an input vector x randomly from learning set.
Search for the unit $\boldsymbol{U_w}$ which is associated to the closest vector $\boldsymbol{m_w}$ to $\boldsymbol{x}$ which minimize the quantization error $\boldsymbol{|x - m|}$.

3) Updating the winner unit and its neighboring units.
For the winner units $\boldsymbol{U_w}$ and its neighbor $\boldsymbol{U'w}$, update the vectors associated to the units using the following equation.

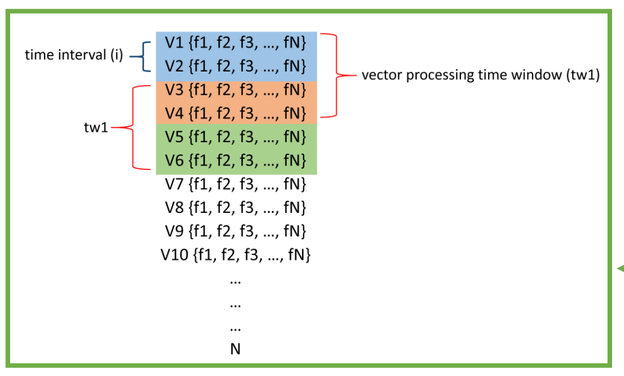
$$\boldsymbol{m_w = m_w + fn(d) \times \eta \times (x - m_w)}$$

where $\boldsymbol{fn(d)}$ is neighborhood function which is the decreasing function of distance d between $\boldsymbol{U_w}$ and $\boldsymbol{U'_w}$
and $\boldsymbol{\eta}$ is the learning rate.

4) Repeat Step 2, Step 3 with decreasing neighborhood function $\boldsymbol{fn(d)}$ and learning rate $\boldsymbol{\eta}$ until the quantization error converges enough or during the pre-defined iterations



User behavior modeling with Kohonen Self Organizing Maps

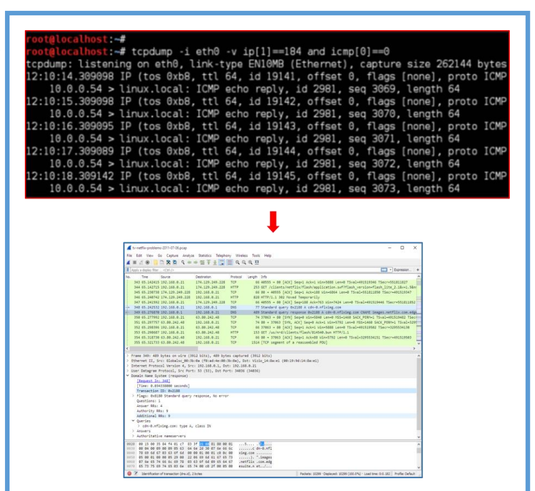


Data capture from host

Raw data processing for information dataset creation

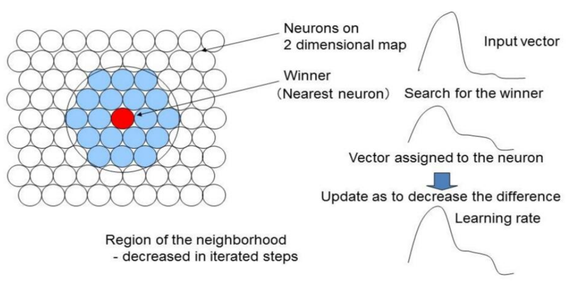
SOM training and individual matching pattern creation

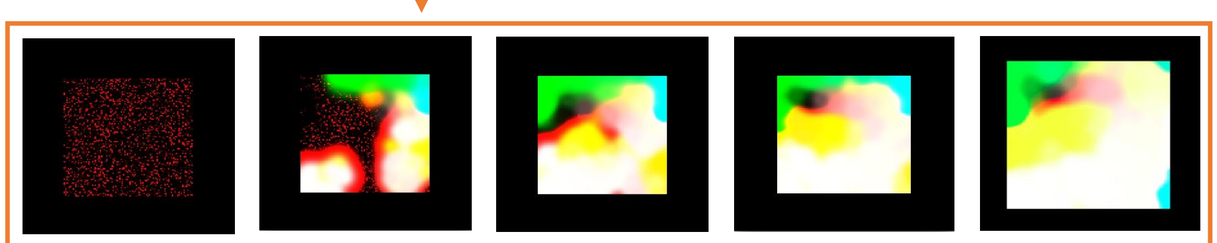
Organization pattern creation and individual Evaluation



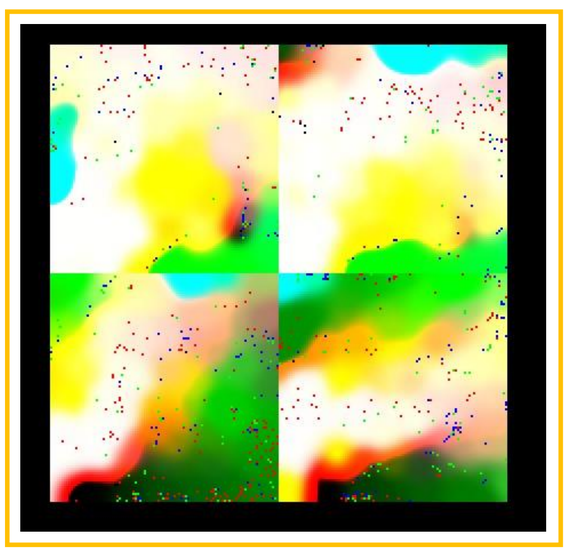
Packages capture through tcdump library

Input Vector {TCP/UDP, Internal Transmission, Web Traffic}

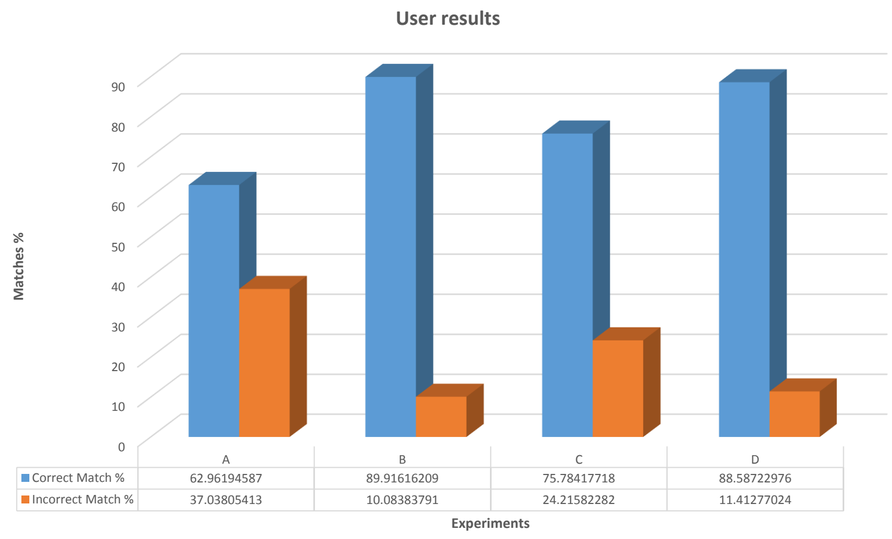




SOM Training running on user dataset



User recognition results



Results were obtained from evaluating test user network traffic against an organization pattern conformed by test user, user two and three individual matching patterns.

Four experiments were done, all of them using 200, 900, 1600 and 2300 input vectors choose randomly from test user, user two and user three processed data sets.

Results show that, test user was identified with max accuracy of 89%.

Conclusions

Using relation between bytes sent through TCP, UDP and web destination as metrics, SOM evaluation it is possible to identify correctly an user with 79% of accuracy.
However an extensive dataset is needed in order to get a better accuracy in user detection and avoid false positive cases. More users will be analyzed in order to generalize results for an organization and being able to detect abnormal behavior.

References

[1] Portnoy, L., Eskin, E., & Stolfo, S. (2001). Intrusion detection with unlabeled data using clustering. In In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001).
[2] CISCO, "Annual Security Report", 2015.
[3] K.A. Jalil and N.M. Masrek, "Comparison of Machine Learning Algorithms Performance in Detecting Network," 2010 International Conference on Networking and Information Technology , p. 6, 2010.
[4] O.Y.S.A.E.M.Y.P.D.M.S. &. K.K. Al-Jarrah, "Machine-learning-based feature selection techniques for large-scale network intrusion detection," IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 177-181, 2014.
[5] S.S.P. &. S.S. Pilabutr, "Integrated soft computing for Intrusion Detection on computer network security," Computer Applications and Industrial Electronics (ICCAIE), 2011 IEEE International Conference, pp. 559-563, 2011.
[6] W. &. L.J. Tian, "A new network intrusion detection identification model research.," Informatics in Control, Automation and Robotics (CAR), 2010 2nd International Asia Conference, vol. 2, pp. 9-12, 2010.
[7] Balachandran, A., Voelker, G. M., Bahl, P., & Rangan, P. V. (2002, June). Characterizing user behavior and network performance in a public wireless LAN. In ACM SIGMETRICS Performance Evaluation Review (Vol. 30, No. 1, pp. 195-205). ACM.
[8] Dozono, H., Itou, S., & Nakakuni, M. (2007). Comparison of the adaptive authentication systems for behavior biometrics using the variations of self organizing maps. International Journal of Computers and Communications, 1(4), 108-116.
[9] Hiroshi Dozono (2012). Application of Self Organizing Maps to Multi Modal Adaptive Authentication System Using Behavior Biometrics, Applications of Self-Organizing Maps, Dr. Magnus Johnsson (Ed.), InTech, DOI: 10.5772/52100.