Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 7, n. 1, 2020.

Trabalho apresentado no XXXIX CNMAC, Uberlândia - MG, 2019.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Hilbert-Huang Transform evaluation for anomaly detection in web traffic

Emilio Gerardo Sotto Riveros[1]
Polytechnic School, UNA, San Lorenzo, Paraguay
Cristian Cappo[2]
Polytechnic School, UNA, San Lorenzo, Paraguay
Christian Schaerer[3]
Center for Research in Mathematics, San Lorenzo, Paraguay

1 Introduction

Nowadays, Web services play an important role in the development of our society, attacks try to exploit vulnerabilities either to violate the security of the system or to affect its availability, so the security of web applications has become in one of the key topics of cybersecurity.

In the context of intrusion detection, several techniques are proposed to determine if web requests correspond to an attack [2, 4, 5]. Two possible approaches are proposed: a) the one based on signatures, and b) the one based on anomalies. The approach based on anomalies has the advantage that it can detect new attacks and variations of the existing ones [4]. This is due since it is not based on previously known signatures. The challenge in the anomaly based detection approach is to minimize the number of false alarms ensuring high detection accuracy.

In this work, the evaluation of the Hilbert-Huang Transform (HHT) [3] is proposed as a technique of the signal processing area for anomalies analysis in data coming from the web traffic. HHT allows analyzing non-linear and non-stationary data characteristics present in the analyzed data [3].

2 Data and Methodology

For looking for the abrupt frequency variations, several features of web traffics are considered, such as the length of the request, the frequency of groups of characters and the entropy of characters. Each of these web traffic features is considered as independent

[1]emiliosotto@gmail.com
[2]ccappo@pol.una.py
[3]chris.schaerer@cima.org.py

signals and for each of them, it is applied the HHT, with the aim of decomposing the signal analyzed (using an Empirical Mode Decomposition process) in a set of Intrinsic Mode Functions (IMF). Each IMF covers a certain frequency range and reflects the time evolution of the components included within the analyzed band. The Hilbert transform is used to obtain the level frequency of each IMF, then, throughout of a voting algorithm based in the work [4]; it is determined the existence or not of an anomaly, which it could be related to an attack. We concentrate our efforts in to analyze IMFs corresponding to high frequencies since normally they are associated to attack presence.

The proposed method is validated with detection and performance experiments over two groups of datasets: the first one is collected within our academic institution and the second one corresponds to public datasets frequently used for anomaly detection algorithms assessment (CSIC 2012 [1] y CICIDS2017 [6]). This is an ongoing work. Preliminary results show that the proposed method detects attacks within the analyzed datasets with F1-score over 0.8, being competitive in time with other state of the art methods.

Acknowledgements. Authors acknowledge the financial support given by FEEI-PROCIENCIA-CONACyT project number POSG17-62.

References

[1] CSIC. Torpeda CSIC dataset 2012. `http://www.tic.itefi.csic.es/torpeda/datasets.html`, 2012. Acceso: 17/10/2016.

[2] Y. Dong, Y. Zhang, H. Ma, Q. Wu, Q. Liu, K. Wang, and W. Wang. An adaptive system for detecting malicious queries in web attacks. *Science China Information Sciences*, 61(3):032114, 2018.

[3] N. E Huang, Z. Shen, S. R Long, Ma. C Wu, H. H Shih, Q. Zheng, N. Yen, C. Chao Tung, and H. H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

[4] A. Kozakevicius, C. Cappo, B. A. Mozzaquatro, R. Ceretta N., and C.E. Schaerer. URL query string anomaly sensor designed with the bidimensional Haar wavelet transform. *International Journal of Information Security*, 14(6):561–581, 2015.

[5] X. Liu, Q. Yu, X. Zhou, and Q. Zhou. Owleye: An advanced detection system of web attacks based on hmm. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervas. Intell. and Comp., 4th Intl Conf on Big Data Intell. and Comp. and Cyber Sc. and Tech.Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 200–207. IEEE, 2018.

[6] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, pages 108–116, 2018.