Detection of malicious domains through Passive DNS with Machine Learning (abstract)

Marcos Rogério Silveira
UNESP - Universidade Estadual Paulista
São José do Rio Preto, São Paulo, Brasil
marcos@acmesecurity.org

Adriano Mauro Cansian
UNESP - Universidade Estadual Paulista
São José do Rio Preto, São Paulo, Brasil
adriano@acmesecurity.org

Hugo Koji Kobayashi
NIC.BR – Brazilian Network Information Center
koji@registro.br

The Domain Name System (DNS) [1][2] is an essential component for the internet, as its main function is to map domain names to IPs, in which your hosts respond. Because of its importance, attackers use this tool for malicious purposes such as the spread of malware, botnets, fast-flux domains and DGAs [3]. Blocklists are a way to recognize these malicious domains based on what has been reported and analyzed by a human. Therefore, for a domain to be present it takes time and more people may have been victims of malicious websites. We will present an approach to detect malicious domains through passive DNS automatically, using the XGBoost [4] supervised machine learning algorithm. We use 12 features extracted exclusively from DNS traffic.

The approach consists of analyzing passive DNS traffic, extracting its characteristics, that is, information in addition to the existing fields that can help in the identification of these domains, labeling this DNS traffic in malicious or legitimate, using allow and blocklists. After labeling this dataset, a machine learning algorithm is trained, where through these presented domains, it can learn to perform this classification automatically. This results in automatic detection of new malicious domains. The classifier model presented, obtained an average AUC of 0.9763.

The dataset used in this work was kindly provided by Mnemonic. This work is being financed through the Research and Innovation Agreement with NIC.BR – Brazilian Network Information Center, for funding this research, under the Fundunesp Process - Foundation for the Development of UNESP, number 2764/2018.

References

[1] 1987. Domain names - concepts and facilities. RFC 1034. https://doi.org/10.17487/RFC1034

[2] 1987. Domain names - implementation and specification. RFC 1035. https://doi.org/10.17487/RFC1035

[3] S. Torabi, A. Boukhtouta, C. Assi, and M. Debbabi. Detecting internet abuse by analyzing passive dns traffic: A survey of implemented systems. *IEEE Communications Surveys Tutorials*, 20(4):3389–3415, 2018.

[4] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-Augu (2016), 785–794. https://doi.org/10.1145/2939672.2939785 arXiv:1603.02754