Data Mining methods for malware detection

Daniel Boţa

Faculty of Mathematics and Computer Science, Babeş-Bolyai University

dbota@cs.ubbcluj.ro

New malicious programs, known as malware, are created every day. In the last decade, the occurrence rate and the amount of malware have become very high. The classical approach, meaning signature-based detection, can not keep up with this overwhelming wave of new and sophisticated malicious codes and can not detect the new previously unseen malware for which there are no signatures. In this regard, we need fast and reliable detection methods that are able to detect them.

The inability of traditional methods to catch these new breed of malicious codes has shifted the focus of malware detection research to find more generalized and scalable methods that can detect malicious behavior as a process instead of a single static signature. Thus, the application of Data Mining methods for malware detection has shown good results compared to other approaches.

In this report, we investigate various Data Mining methods used for malicious executable detection.

References

[1] Karthikeyan, L., Jacob, G. and Manjunath, B., 2011. Malware images: Visualization and automatic classification. In Proceedings of the 8th International Symposium on Visualization for Cyber Security (p. 4)

[2] Kolter, J.Z. and Maloof, M.A., 2006. Learning to detect and classify malicious executables in the wild. Journal of Machine Learning Research, 7(Dec), pp.2721-2744

[3] Santos, I., Penya, Y.K., Devesa, J. and Bringas, P.G., 2009. N-grams-based File Signatures for Malware Detection. ICEIS (2), 9, pp.317-320

[4] Schultz, M.G., Eskin, E., Zadok, F. and Stolfo, S.J., 2001. Data mining methods for detection of new malicious executables. In Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on (pp. 38-49). IEEE