Postdoc Position in Malware Analysis using Machine Learning

Environment The TAMIS team (`https://team.inria.fr/tamis`) at Inria Rennes - Bretagne Atlantique is among the largest security teams at Inria, including competences from hardware attacks to cryptography, and from vulnerability detection to malware analysis. This project is connected to the Chair of Cybersecurity in Malware Analysis financed by Région Bretagne.

Prerequisites The candidate must have a Ph.D. in Computer Science, or an equivalent certification in a closely related field and proven experience in Computer Science research. Expertise with machine learning techniques, graph mining, and malware detection and analysis will be strongly considered. We are looking for team players who are motivated to drive top-quality research.

Duration/Starting date The position is (initially) limited to two years. The position is already available, but due to the TAMIS team being in a Restricted Access Zone (ZRR) the additional screening process will take two months more than normal.

Project Computing systems are under continuous attacks by increasingly motivated, sophisticated, and adaptive adversaries. Malware deploys malicious software and can be used to gain remote access to a system, spy on its users, exfiltrate and modify data, execute denial of services attacks, etc. One promising approach to detect and analyze malware is to build system call dependency graphs and classify according to characteristics of these graphs. Recent works (see e.g. [PBCK13, NP15, JWM$^+$16]) used similarity metrics, graph mining, feature vectors in order to classify between malware and cleanware. However, state-of-the-art techniques may not meet the requirements given in practical settings and are not adapted to the particular scenario of malware detection. In particular, approximately 390 000 new malware samples are observed daily, which requires fast and precise detection within seconds and furthermore to be able to retrain models sufficiently quick. Also, as most of the cleanware/malware are rather complex the corresponding graphs may contain a huge amount of nodes and edges.

Within this project, we want to develop machine learning techniques that directly work on/with the graph structure of the system dependency graph while being incremental, distributed, and achieving a suitable high accuracy and low false positive rate.

The measure of success of this project is the effectiveness and efficiency of the malware analysis system in correctly classifying malware and cleanware samples, and particularly with a small false positive rate, which is fundamental in malware classification. Due to the large literature on using different machine learning methods in malware classification, it is easy to compare directly the true positive and false positive classification rates of this project against the state of the art. Additionally, the TAMIS team will offer this technology to its partners working on malware analysis (e.g. Cisco, Thales) where it will be applied in a real large-scale environment. Finally, the results of the research will be published in A*- or A-class security and malware conference (e.g. USENIX Security, IEEE S&P, MALWARE, ACM CCS, ESORICS).

This project will significantly improve the state of the art in machine learning for malware analysis, providing scalable, distributed, and incremental algorithms to be used natively on graph-based signatures. This will allow improving heuristic malware analysis, detecting new malicious samples even when they are protected by advanced morphic and obfuscation techniques. Finally, the result will be a better protection for users, entities, and infrastructure, and a safer computing environment for everybody.

Contact Interested candidates are encouraged to send a detailed CV and list of publications to:

- Fabrizio Biondi, fabrizio.biondi@inria.fr
- Annelie Heuser, annelie.heuser@irisa.fr

References

[JWM+16] Jae-wook Jang, Jiyoung Woo, Aziz Mohaisen, Jaesung Yun, and Huy Kang Kim. Mal-netminer: Malware classification approach based on social network analysis of system call graph. *CoRR*, abs/1606.01971, 2016.

[NP15] Stavros D. Nikolopoulos and Iosif Polenakis. A graph-based model for malicious code detection exploiting dependencies of system-call groups. In *Proceedings of the 16th International Conference on Computer Systems and Technologies*, CompSysTech '15, pages 228–235, New York, NY, USA, 2015. ACM.

[PBCK13] Sirinda Palahan, Domagoj Babić, Swarat Chaudhuri, and Daniel Kifer. Extraction of Statistically Significant Malware Behaviors. In *ACSAC'13: The 29th Computer Security Applications Conference*, 2013.