PhD position in Scalable and Interactive Security Analytics

- **Location:** INRIA Nancy Grand Est research center – *Villers-lès-Nancy*, France

- **Research theme:** Networks, Systems and Services

- **Project-team: Madynes**

- **Scientific Context:**

The huge growth of Internet exposes many users to various threats. This has been intensified by the large deployment of new devices in addition to traditional computers. This includes smartphones and sensors, and will concern daily life objects in a near future with the emergence of the Internet of Things (IoT) the last years. Hence, this represents a tremendous playground for attackers. To fight them, network security is essential to identify misbehaviors and potential victims as earlier as possible.

The attackers evolve from individuals towards organized cyber-criminal organizations [1] while meantime the attacks being more distributed and complex. For example, the botnets [2] are still a major threat on Internet, where thousands of zombie machines can take part, because they have been successfully adapted from a centralized model based on IRC towards distributed approach, even P2P, taking advantage of traditional protocol (DNS for fast fluxing [3]) and new technologies (social networks for synchronization [4]). In parallel, they are responsible of various attacks including spam, denial of service, credential stealing [5]... Therefore fighting such a threat among others require to collect, analyze and correlate various sources of data to create summarized view that are exploitable by human administrator and, if possible, in real time and in an automated way. This is the current challenge of the network security monitoring [6]. Currently, most of attacks remains unrevealed, but when they are suspected, it is vital to investigate it to confirm, to trace the root causes and attackers. The forensics security teams have very few tools which let them performing analysis mainly manually, which introduces two bias: long delay (from few hours to several months) and human bias due to background and experiences.

In parallel, data-analytics methods have skyrocketed recently and are able to cope with huge volumes of unstructured data and so are good candidates for being adapted and applied to security monitoring challenges by allowing collecting and analyzing multiple sources of relevant data while current approaches focuses on few ones or on simple correlation of several ones.

- **Missions:**

The objective of the thesis is to design a methodology for being able to counteract against new threats on Internet by monitoring them through data-consolidation over multiple sources. In parallel, in order to help the security teams, new investigation methods have to be built by empowering the interactivity and the visualization of the information (raw, summarized or consolidated data). To achieve that, it will be necessary to :

1 analyze current threats to define data and features being primordial for an efficient monitoring. This will allow then to design data models which are able to handle heterogeneous and multi-dimensionnal data.

2 define analytics methods to identify anomalies based on these data models. This will consider statistical analysis, graph analysis and machine learning approaches. The goal is to consider both 0 day attacks and long-term stealthy attacks (Advanced Persistent Threats) as well as collected malware to build signatures for future detection.

3 define methods for interactive and visual investigation of multiple sources of security data. This will consider similar methods that those under the second item but with a hard constraint on the reactivity and the limited quantity of information which can be dealt simultaneously by a human. Hence, these methods may rely on streaming analytics approaches, learning approaches to predict the next requests of the analysts to prepare the results (pre-processing and pre-rendering), combining and selecting information,

4 validate the proposed methods on different scenarios. In a first phase, the analysis will be mainly focused on individual data source microscopic analysis (monitoring DNS request to detect botnet activities, analyzing trafic flows to identify DDoS attacks, HTTP trafic for phishing detection, syslog events, security alerts...) before correlating and augmenting them to strengthen the results (monitoring DNS and analyzing flow for botnets, using fingerprinting methods to tag hosts and flows before analysis, traffic causality graphs...).

This work will be achieved in the context of the first French high security academic research laboratory in Nancy (LHS – High Security Laboratory) which provides powerful tools and support for collecting and analyzing dataset in a realistic environment and in the context of the HuMa project funded under the FUI 19 programme (Fonds Unique Interministériel) in cooperation with major French industrial players in cyber-security.

- **Bibliography:**

[1] R. Howard, Cyber Fraud: Tactics, Techniques and Procedures. Auerbach Publications, 2009, ch. 5, The Russian Business Network: the Rise and Fall of a Criminal ISP.

[2] Seungwon Shin and Guofei Gu. 2010. Conficker and beyond: a large-scale empirical study. In *Proceedings of the 26th Annual Computer Security Applications Conference* (ACSAC '10).

[3] C. Castelluccia, M. A. Kaafar, P. Manils, and D. Perito, "Geolocalization of proxied services and its application to fastflux hidden servers," ACM SIGCOMM IMC

[4] Shishir Nagaraja, Amir Houmansadr, Pratch Piyawongwisal, Vijit Singh, Pragya Agarwal, Nikita Borisov "Stegobot: A Covert Social Network Botnet", Information Hiding 2011

[5] Dainotti, A.; King, A.; Claffy, K.; Papale, F.; Pescape, A., "Analysis of a "/0" Stealth Scan From a Botnet," IEEE/ACM Transactions on Networking, no.99

[6] Cloud Security Alliance, "Big Data Analytics for Security Intelligence", 2013

- **Skills and profile:**

Required qualification: Master degree, preferably in computer science

Knowledge and skills in the following fields will be appreciated: networking, security, machine learning, data-mining, human-computer interaction

- **Additional information:**

Supervision : Jérôme François , Abelkader Lahmadi, Isabelle Chrisment, Olivier Festor
Contacts: jerome.francois@inria.fr, abdelkader.lahmadi@inria.fr

Additional links: madynes website (http://madynes.loria.fr/), LHS website (http://www.lhs.loria.fr/)

Salary: 1 958 euros gross monthly (about 1 584 euros net) during the first and the second years. 2 059 euros the last year (about 1 665 euros net). Medical insurance is included.

The required documents for applying are the following:

- CV;
- a motivation letter;
- your degree certificates and transcripts for Bachelor and Master (or the last 5 years if not applicable).
- Master thesis (or equivalent) if it is already completed, or a description of the work in progress, otherwise;
- all your publications, if any (it is not expected that you have any).
- At least one recommendation letter from the person who supervises(d) your Master thesis (or research project or internship); you can also send at most two other recommendation letters.

The recommendation letter(s) should be sent directly by their author to the prospective PhD advisor.

All the documents should be sent in at most 2 pdf files; one file should contain the publications, if any, the other file should contain all the other documents. These two files should be sent to the contacts mentioned above.