Probabilistic graphical model-based security for SCADA critical infrastructures

Mots clés :
- Directeur de thèse : Christophe Gonzales
- Co-encadrant(s) :
- Unité de recherche : Laboratoire d'informatique de Paris 6
- Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)
SCADA systems are exploited to control many industrial processes and are extensively exploited in many critical infrastructures like energy and water distribution, nuclear plants, etc. Unfortunately, their security is currently often insufficient to prevent malevolent access and cyber intrusions, which can have dramatic consequences (see, for instance, http://en.wikipedia.org/wiki/Stuxnet). The European community has issued a law making it compulsory for companies to protect their critical infrastructures, including their SCADA systems. The goal of the PhD is to exploit probabilistic graphical model technologies to detect possible intrusions or cyber-attacks in such SCADA systems. The PhD student will first work on the learning of the probabilistic graphical models from data acquired from the SCADA system. Those are very heterogeneous and their large amount places the learning perspective into a big data context. Once the models will be learnt, the PhD will have to develop efficient inference algorithms to detect intrusions and attacks as quickly as possible. The PhD student will work on the necessary theoretical concepts and methodologies to design these algorithms and will also implement them in C++ into the aGrUM graphical model library (http://agrum.lip6.fr).

Résumé du projet de recherche (Langue 2)
The challenges raised by the PhD are essentially fourfold: - SCADA systems are temporal and their different parts have quite different dynamics, in particular they evolve at different time granularities; in addition, the spatial heterogeneity of the system makes the time stamps unreliable, especially when observed time series are needed; - SCADA systems are very heterogeneous by nature, which implies that learning must be performed on NoSQL databases (containing both continuous and discrete random variables); such mixture is still a challenge for probabilistic models; - data are acquired in a streaming mode and their amount is expected to be high; - the system is non-stationary: it evolves over time.

Informations complémentaires (Langue 1)
Securing efficiently SCADA critical systems requires securing its parts at different levels, including at sensor/probe level, at CCTV level, at communication level and at control higher levels. The PhD will work on the latter level but will interact with other European researchers (Italian, Belgian, Austrian and Polish) securing the other parts of the system. The PhD’s work will be applied on a real-world application on the data of an Italian energy provider.

Informations complémentaires (Langue 2)
Requirements to candidate include a good mathematical background and excellent programming skills in C++. A good experience in probabilistic graphical models is also welcome. http://www-desir.lip6.fr/~gonzales/students/positions.php