Détection d'événements et supervision par approche mixte logique et stochastique
Application à la détection d'actions cybercriminelles par analyse de traces

Sujet de thèse proposé par l'ONERA et le LINA :

- Contrat de 3 ans cofinancé DGA/ONERA (rémunération nette mensuelle ca. 1600 €)
- Lieu : ONERA, centre de Palaiseau (missions ponctuelles à Nantes au LINA)
- Directeur de thèse (LINA, équipe DUKE – Data User Knowledge) : Philippe Leray
- Encadrant ONERA : Romain Kervarc

La surveillance de systèmes complexes nécessite de pouvoir détecter, identifier et expliquer des événements anormaux (anomalies, pannes, attaques, …). Les différentes techniques pouvant être mises en œuvre dans ce cadre doivent pouvoir répondre à différents problèmes, principalement le ***monde ouvert*** (i.e. où l'absence d'une information ne traduit pas nécessairement la fausseté de cette information), la **grande taille** des systèmes considérés (ceux-ci pouvant de plus faire intervenir des acteurs très divers dans des domaines hétérogènes) et leur caractère **dynamique**, leur caractère partiellement **incertain** (qu'il s'agisse d'incertitude dans les événements observables, ou encore d'incertitude dans les comportements décrits).

Parmi les applications les plus directes de ces questions de surveillance figurent les problèmes de sécurité informatique, et en particulier celui de l'analyse de traces d'usages, où l'on souhaite examiner la manière dont un ou plusieurs utilisateurs utilisent un système informatique, ce qui présente des intérêts évidents. On peut ainsi détecter des activités malveillantes ou criminelles, soit a posteriori (computer forensics), soit, pour peu que l'on dispose d'outils assez efficaces et que l'on soit capable de spécifier des actions potentiellement malveillantes, pour produire des alertes de tentative d'attaque en cours, mais aussi de raisonner sur des problèmes de politiques de données visant à contrôler l'accès à l'information dans des systèmes coopératifs.

Deux principales familles d'approches existent pour ce type de détection de comportements sous forme de relations ou corrélations entre événements observables du système, les modèles **logiques**, dont les logiques temporelles, et les méthodes **stochastiques**, comme les réseaux bayésiens dynamiques. Développés de manière parallèle, voire même par le passé antagoniste, ces deux approches ont de plus en plus tendance à « s'hybrider » [Get07] : modèles relationnels probabilistes [Fri99, Get07], réseaux logiques de Markov [Ric06], Bayesian logic [Mil05, Ker07], y compris parfois avec une dimension temporelle [Kat11].

Cette thèse est proposée conjointement par l'ONERA (unité TCS, où est développé un formalisme de détection de comportement basé sur la logique temporelle, les *chroniques*), et le LINA (équipe DUKe, à l'expertise internationalement reconnue sur les modèles graphiques probabilistes) s'inscrit dans cette ligne, et vise à développer des méthodes mixtes entre chroniques et réseaux bayésiens pour combiner leurs avantages respectifs et obtenir des outils de surveillance et d'analyse efficaces pour des systèmes dynamiques, de grande taille, avec incertitude, et dans un monde ouvert, puis à les mettre en œuvre pour des applications de détection et de prévention d'actions cybercriminelles, domaine où se trouvent à la fois des représentations temporelles des attaques et une incertitude sur les observations réalisées et les intentions de leurs auteurs. La thèse cherchera à montrer que l'approche mixte combinant éléments de raisonnement en logique temporelle et représentation de l'incertitude à l'aide de modèles probabilistes permet une détection mieux adaptée, plus précise, et plus rapide.