Internship proposition: Development of adversarial classifiers using Bayesian games

Keywords: game theory, machine learning, classification, security

Lab: Laboratoire d'Informatique de Grenoble (LIG), Grenoble, France (head: Eric Gaussier)
Team in the lab: POLARIS (head: Arnaud Legrand)

Advisor: Patrick Loiseau (Univ. Grenoble Alpes, LIG)
patrick.loiseau@univ-grenoble-alpes.fr, http://www.eurecom.fr/~loiseau/

General presentation of the topic:

Classification algorithms (a class of learning algorithms) are routinely used in crucial security problems such as detecting attacks (e.g., fraud, spam, theft, malware, etc.). However, standard classification algorithms often perform poorly in such scenarios because an adaptive attacker can shape his attacks in response to the algorithm. Hence the data faced in security problem does not satisfy the standard learning assumption that it its distribution is independent from the algorithm. This has led to a recent interest in developing methods for adversarial classification but existing methods make overly pessimistic assumptions that affect their performance in practice.

Internship program and objectives:

In this internship, we propose to use game theory to model the objectives of the defender (learner) and the attacker (data generator) in order to derive better methods for adversarial classification. We will build on our recent work [1] that proposes a simple game theoretic model of adversarial classification and derives optimal classifier for this model; but that makes the limiting assumption of complete information (about the attacker). Then, the main objective of the internship will be to develop and investigate a model with incomplete information in order to derive more flexible adversarial classification methods that do not require knowledge of the attacker. More specifically, the student will:

- develop an incomplete information model based on a Bayesian game;
- investigate theoretically the game's equilibria—in particular look for an algorithmic reduction of the defender's strategy space that gives the set of optimal classifiers;
- analyze the solution's performance and robustness using simulations;
- if time permits, extend the study to a dynamic context with sequential learning.

Expected abilities of the student:

Strong background in probability, knowledge of machine learning, basics of game theory

References:

[1] Lemonia Dritsoula, Patrick Loiseau, and John Musacchio. A game-theoretic analysis of adversarial classification. IEEE Transactions on Information Forensics and Security, 12(12):3094–3109, December 2017.