Robustesse des systèmes de détection d'intrusion basés sur l'apprentissage machine

Gregory Blanc, Christophe Kiennert

Thématiques — apprentissage, détection d'intrusion, robustesse, évaluation

Contexte

Depuis 40 ans, les systèmes de détection d'intrusion réseau ont grandement évolués mais sont toujours confrontés aux mêmes défis tels que a) la pertinence de leur détection quant à la probabilité d'occurence d'une intrusion [1] ; b) la performance des nouveaux algorithmes, notamment ceux basés sur l'apprentissage statistique [2] et l'apprentissage profond [3], très gourmands en ressources ; auxquels s'ajoutent les problématiques de c) sécurité de ces détecteurs lorsqu'ils deviennent eux-même la cible d'acteurs malveillants [4] ; et de d) chiffrement puisque la part de trafic chiffré dans le volume global du trafic Internet ne cesse d'augmenter [5].

Cependant, la littérature exhibe de nombreux résultats censés améliorer l'état de l'art : en particulier, l'apprentissage profond donne souvent de meilleurs résultats, sans qu'il soit possible de les expliquer [3]. Ce qui ne participe pas au peu de confiance que l'on peut porter à ces algorithmes, notamment lorsqu'un attaquant qui a accès au modèle est capable de les contourner. De fait, un nouveau domaine émerge où sont proposés des algorithmes d'apprentissage robuste, sous le vocable d'apprentissage adverse [4].

Dans ce stage, nous souhaitons évaluer la robustesse des algorithmes de détection d'intrusion et des détecteurs d'intrusion en générant des modèles adverses [6]. Nous souhaitons caractériser cette robustesse selon des métriques, et explorer les limites théoriques de celles-ci. Dans un second temps, une évaluation expérimentale validera ces résultats de manière pratique.

Déroulement

1. Modélisation
   - Etude bibliographique (état de l'art) sur la détection d'intrusion, notamment la détection d'anomalie non supervisée.
   - Etat de l'art des publications sur l'usage de GAN pour la détection d'intrusion
   - Modélisation de la robustesse des détecteurs d'intrusion
2. Évaluation
   - Implémentation des modèles et simulation
   - Application à des jeux de données de trafic réseau

Références

[1] Stefan Axelsson : The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security (TISSEC)*, 3(3):186–205, 2000.

[2] Anna L Buczak et Erhan Guven : A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2):1153–1176, 2016.

[3] Donghwoon Kwon, Hyunjoo Kim, Jinoh Kim, Sang C Suh, Ikkyun Kim et Kuinam J Kim : A survey of deep learning-based network anomaly detection. *Cluster Computing*, pages 1–13, 2017.

[4] Ling Huang, Anthony D Joseph, Blaine Nelson, Benjamin IP Rubinstein et JD Tygar : Adversarial machine learning. *In Proceedings of the 4th ACM workshop on Security and artificial intelligence*, pages 43–58. ACM, 2011.

[5] Sébastien Canard, Aïda Diop, Nizar Kheir, Marie Paindavoine et Mohamed Sabt : Blindids : Market-compliant and privacy-friendly intrusion detection system over encrypted traffic. *In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, ASIA CCS '17, pages 561–574, New York, NY, USA, 2017. ACM.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville et Yoshua Bengio : Generative adversarial nets. *In Advances in neural information processing systems*, pages 2672–2680, 2014.