# Génération réaliste de paquets réseau pour l'évaluation de détecteurs d'intrusion

Grégory Blanc, Télécom SudParis
Ludovic Mé, Inria
Pierre-François Gimenez, CentraleSupélec
Frédéric Majorczyk, DGA-MI

**Mots-clés** : détection d'intrusions, génération de trafic, GAN

La détection d'intrusion est un mécanisme essentiel de la sécurité des systèmes d'information. Dans une approche réactive, elle est la première étape permettant de remonter des évènements suspects, lesquels pourront être analysés et caractérisés successivement afin de déterminer si le système protégé est attaqué et, le cas échéant, de quelle manière il faut répondre, c'est à dire quelles sont les contremesures adaptées à mettre en place.

L'évaluation de systèmes de détection d'intrusion s'avère donc d'une extrême importance pour qualifier la pertinence des résultats de la détection. La plupart du temps, cette évaluation se résume à mesurer le nombre d'alertes levées par le détecteur en présence de trafic malveillant et sain. Ce trafic provient de jeux de données souvent généré plus ou au moins automatiquement. Pis, le nombre des jeux de données reste suffisamment restreint pour que la diversité de celui-ci soit discutable et que son vieillissement soit problématique [RWS+19].

Plutôt que de se contenter de datasets statiques, il serait intéressant d'être capable de générer du trafic réseau dynamiquement, réaliste et correspondant au système d'information visé pour l'évaluation d'un IDS.

Des travaux récents dans le domaine de l'apprentissage automatique ont montré la capacité de certains réseaux de neurones [SBZD19], notamment les réseaux génératifs et adverses [GPAM+14] à générer des données réalistes. Ces travaux ont été appliqués avec un certain succès à la génération de résumé de flux réseau [RSLH19], la génération de séquences de caractéristiques des paquets [SBJ+20] et la génération de paquets unitaires [Che19]. L'idée de ce stage est d'étudier comment combiner les modèles aux différents niveaux d'abstraction pour générer des paquets réseau réalistes. Il sera peut-être nécessaire d'associer des règles statiques aux réseaux de neurones pour prendre en compte des fonctions relativement complexes comme par exemple le calcul d'un CRC dans les paquets IP. Dans un premier temps, on pourra étudier la génération des paquets dans les deux sens sans interaction avec un service réseau réel puis étudier la possibilité de ne générer que les paquets d'une des machines participant au flux réseau (par exemple, un client communiquant avec un service réel).

L'utilisation de ce type de réseaux nécessite une base d'apprentissage importante. Dans le cadre de ce stage, on pourra utiliser les pcaps fournis par des datasets publics tels CIC-IDS 2017 et 2018 [SLG18] ou UGR 2016 [MFCMC+18].

Une réflexion sera à apporter sur le réalisme des paquets ainsi générés [BKD19]. Il est possible de mesurer différentes caractéristiques statistiques comme dans [RSLH19] mais également de faire des mesures expérimentales telles que l'acceptation d'une suite de paquets par un service réel [Che19].

## Encadrement

Le sujet de ce stage est un sujet orienté recherche : les candidats ou candidates souhaitant poursuivre en thèse seront privilégiées. Ce stage sera réalisé au sein de l'équipe Inria/IRISA CIDRE à Rennes, en collaboration avec Télécom SudParis. Il sera encadré par plusieurs chercheurs et enseignants-chercheurs : Ludovic Mé (actuellement en détachement à l'Inria), Grégory Blanc (enseignant-chercheur à Télécom SudParis), Pierre-François Gimenez (maître de conférence au sein de CIDRE) et Frédéric Majorczyk

(ingénieur DGA, collaborateur extérieur dans l'équipe). La durée envisagée du stage est de 5 mois.

## Contacts


— gregory.blanc@telecom-sudparis.eu
— ludovic.me@inria.fr
— pierre-francois.gimenez@centralesupelec.fr
— frederic.majorczyk@def.gouv.fr


## Références


[BKD19] Pierre-Marie Bajan, Christophe Kiennert, and Hervé Debar. Methodology of a network simulation in the context of an evaluation : application to an ids. 2019.

[Che19] Adriel Cheng. Pac-gan : Packet generation of network traffic using generative adversarial networks. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0728–0734. IEEE, 2019.

[GPAM+14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[MFCMC+18] Gabriel Maciá-Fernández, José Camacho, Roberto Magán-Carrión, Pedro García-Teodoro, and Roberto Therón. Ugr '16 : A new dataset for the evaluation of cyclostationarity-based network idss. *Computers & Security*, 73 :411–424, 2018.

[RSLH19] Markus Ring, Daniel Schlör, Dieter Landes, and Andreas Hotho. Flow-based network traffic generation using generative adversarial networks. *Computers & Security*, 82 :156–172, 2019.

[RWS+19] Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes, and Andreas Hotho. A survey of network-based intrusion detection data sets. *Computers & Security*, 86 :147–167, 2019.

[SBJ+20] Mustafizur R. Shahid, Grégory Blanc, Houda Jmila, Zonghua Zhang, and Hervé Debar. Generative deep learning for internet of things network traffic generation. In *to be published in PRDC 2020*, 2020.

[SBZD19] Mustafizur R Shahid, Gregory Blanc, Zonghua Zhang, and Hervé Debar. Anomalous communications detection in iot networks using sparse autoencoders. In *2019 IEEE 18th International Symposium on Network Computing and Applications (NCA)*, pages 1–5. IEEE, 2019.

[SLG18] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A Ghorbani. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *ICISSP*, pages 108–116, 2018.