Étude approfondie des jeux de données d'attaque : le challenge CasinoLimit

Sébastien Kilian
CentraleSupélec, Inria, CNRS, IRISA
F-35000 Rennes, France
sebastien.kilian@inria.fr

Valérie Viet Triem Tong
CentraleSupélec, Inria, CNRS, IRISA
F-35000 Rennes, France
valerie.viettriemtong@inria.fr

Jean-François Lalande
CentraleSupélec, Inria, CNRS, IRISA
F-35000 Rennes, France
jean-francois.lalande@inria.fr

***Abstract*—Les exercices de cybersécurité sont souvent utilisés pour former et évaluer les compétences des professionnels de la cybersécurité. Ces exercices offrent également une opportunité unique de générer des jeux de données contenant des attaques réalistes sur des systèmes non sensibles. Dans cet article, nous présentons une méthodologie pour étiqueter un jeu de données contenant à la fois des journaux système et réseau. Les étiquettes sont basées sur la matrice MITRE ATT&CK et sont générées automatiquement puis confirmées par un expert. Nous appliquons cette méthodologie au challenge CasinoLimit, un exercice de cybersécurité qui simule un réseau de machines vulnérables. Composé de 2 sous-réseaux, 4 machines virtuelles et 7 positions d'attaque, ce défi a été joué par 113 équipes de différents niveaux de compétence, chacune sur sa propre infrastructure isolée. Cela permet d'analyser les différents comportements de chaque équipe face au même ensemble de vulnérabilités. Pour appliquer efficacement cette méthodologie sur une grande quantité de données, nous avons développé MANATEE, un outil qui facilite la visualisation et l'étiquetage des journaux. Nous fournissons également les jeux de données générés avec les étiquettes correspondantes, qui pourront être utilisés pour la classification des attaques, mais aussi pour des applications offensives.**

I. Introduction

En cybersécurité, il est crucial de comprendre le comportement des attaquants afin de détecter et prévenir les attaques. Cependant, obtenir des données réelles d'attaque est difficile en raison des préoccupations en matière de confidentialité. En effet, les organisations sont souvent réticentes à partager leurs données, car elles peuvent contenir des informations sensibles. En conséquence, la communauté de la cybersécurité a souvent un accès limité aux données réelles d'attaque, ce qui entrave l'expérimentation de solutions de cybersécurité.

De plus, l'étiquetage de ces jeux de donnés est une nécessité pour beaucoup d'applications [1]. Étiqueter des journaux d'événements nécessite des connaissances spécialisées et une très bonne compréhension des techniques d'attaque. En ajoutant le fait qu'une grande quantité de données est nécessaire pour produire des résultats significatifs, le processus d'étiquetage peut être chronophage alors que le nombre d'experts qualifiés est faible.

Dans ce travail, nous proposons de générer un jeu de données à partir d'un exercice de test d'intrusion sur une infrastructure réelle. Lors de cet exercice, les joueurs sont invités à explorer une infrastructure vulnérable, générant ainsi des journaux système et réseau. Ensuite, ce jeu de données est étiqueté semi-automatiquement avec des étiquettes issues de la matrice MITRE ATT&CK [2]. Pour cette partie, après avoir généré automatiquement les étiquettes, des analystes juniors et experts confirment manuellement ces étiquettes pour assurer l'exactitude des étiquettes posées.

Après avoir présenté ce qui a été fait pour créer ce jeu de données d'attaques (Section II), cet article décrit comment nous avons créé l'exercice (Section III), étiqueté le jeu de données produit (Section IV) et analysé les comportements des joueurs (Section V).

II. État de l'art

A. Jeux de données d'attaques

Les jeux de données d'attaques sont générés à partir d'infrastructures dans lesquelles des joueurs ou des programmes réalisent des actions. Les infrastructures peuvent être des infrastructures construites à la main [3] ou générées par un programme [4]. Dans les *Capture The Flag* (CTF), les participants sont mis au défi d'exploiter des vulnérabilités pour progresser dans un environnement de ce type. Par exemple, dans [5], un framework est proposé pour générer des CTFs et collecter les données générées par les joueurs, ce qui permet ensuite de construire un modèle de menace [6].

Dans la littérature, plusieurs jeux de données ont été spécialement générés pour comprendre comment les attaques sont réalisées ou comprendre la psychologie de l'attaquant [7]. Dans [8], les auteurs ont instrumenté des challenges CTF avec des *keyloggers* pour enregistrer les actions des joueurs. Ils ont ensuite analysé les commandes tapées par les attaquants et les ont étiquetées avec les techniques correspondantes de la matrice MITRE ATT&CK. Cependant, la limitation à 8 joueurs ne permet pas de réaliser des statistiques précises pour le profilage de joueurs ou l'entraînement de modèles. Dans notre approche, nous générerons un large volume de données en dupliquant l'expérimentation à un grand nombre de joueurs.

B. Étiquetage des jeux de données

Une façon d'étiqueter les journaux issus de la supervision d'infrastructures attaquées est d'utiliser des règles de filtrage de log prédéfinies [9]. Cette approche repose sur le fait que certaines attaques ont un modèle spécifique qui peut

TABLE I
TECHNIQUES REQUISES POUR RÉSOUDRE LE SCÉNARIO

Machine	Technique requise	Description
Start	T1021: Remote Services	L'attaquant se connecte à la machine en utilisant SSH
Start	T1021: Remote Services	L'attaquant réutilise le mot de passe pour se connecter à la machine meetingcam
Meetingcam	T1125: Video Capture	L'attaquant utilise la webcam pour prendre une photo de parties du mot de passe de tocean@bastion
Meetingcam	T1021: Remote Services	Bruteforce les caractères manquants du mot de passe de tocean@bastion et se connecte à la machine bastion
Bastion	T1068: Exploitation for Privilege Escalation	L'attaquant exploite une vulnérabilité de la machine bastion pour obtenir un accès root
Intranet	T1190: Exploit Public-Facing Application	L'attaquant exploite une vulnérabilité SSTI sur le site intranet pour mettre à jour sa base de données
Bastion	T1114: Email Collection	L'attaquant lit le flag dans l'email

être détecté en cherchant des motifs spécifiques dans les journaux. Cependant, cette approche est limitée aux attaques connues. C'est un inconvénient majeur quand on considère des attaquants réels qui utilisent des nouvelles techniques ou des variantes qui ne sont pas couvertes par les règles utilisées.

Dans [10] les auteurs différencient les activités malveillantes des activités bénignes lorsque les périodes temporelles d'exécution des attaques ont été préalablement identifiées. Cette approche fonctionne bien lorsque la nature de l'attaque est connue et que l'expérimentateur contrôle le début et la fin de l'attaque. D'autres approches utilisent des scripts pour déclencher les attaques, ce qui permet de connaître des *timestamp* plutôt que des intervalles de temps. Par exemple, dans [11], une plateforme d'émulation de menace réalise à la fois l'attaque et génère des rapports sur ces attaques. Ces rapports peuvent être utilisés pour étiqueter les logs générés par la plateforme à partir des types d'actions réalisés dont on connaît le *timestamp* précis.

Nous utiliserons une approche similaire dans notre travail, à ceci près que le déroulement des attaques n'est pas totalement contrôlé. En effet, celles-ci sont perpétrées par des joueurs autonomes et non pas par des programmes.

III. LE CHALLENGE CASINOLIMIT

Nous avons conçu un scénario dans lequel, les participants doivent compromettre un réseau composé de 4 machines avec 7 positions d'attaque. Leur objectif est de modifier une base de donnée située sur la machine *intranet* pour obtenir un flag.

Les joueurs commencent avec l'URL de la machine *start* et utilisent les identifiants de l'utilisateur *tbenedict* fournis pour se connecter en SSH. Ensuite, les joueurs doivent découvrir par un scan réseau l'existence d'une machine *meetingcam* et utiliser ces mêmes identifiants pour s'y connecter. Après avoir remarqué qu'un service web simulant une caméra est présent sur la machine *meetingcam* (en inspectant les fichiers de configuration de session de l'utilisateur), les joueurs utilisent la webcam pour prendre une photo d'un tableau d'une salle de réunion contenant une partie d'un mot de passe de l'utilisateur *tocean* pour une autre machine *bastion*. Ils peuvent alors réaliser une attaque par force brute pour trouver le mot de passe complet et se connecter à la machine *bastion* dont le service SSH tolère de nombreuses tentatives de connexions.

Les attaquants étudient à nouveau la machine *bastion* et y trouvent un serveur IMAP contenant des emails à l'intention de l'utilisateur *tocean*. Ils trouvent aussi un nouveau sous-réseau dont la machine *intranet* fait partie et héberge un site web vulnérable. Les informations contenues dans les emails donnent l'IP publique vers ce site web vulnérable. Les joueurs doivent ensuite remarquer que la marchine *bastion* est vulnérable à la `CVE-2023-0386`, soit en ayant vu l'indice sur la machine *meetingcam*, soit en regardant le numéro de version du noyau linux. Puis ils doivent l'exploiter pour effectuer une élévation de privilège et se connecter en tant que *root* sur la machine *bastion*. Ils effectuent ensuite un mouvement latéral pour se connecter en tant qu'*admin* sur *bastion* et récupérer des identifiants de connexion dans un nouvel email. Grâce à ces identifiants, ils se connectent au site web de l'*intranet* et utilisent une vulnérabilité SSTI (Server Side Template Injection) pour modifier la base de donnée et obtenir le flag qui est envoyé par email à l'utilisateur *admin*.

Les participants doivent effectuer 7 actions principales pour lesquelles nous donnons le numéro de technique MITRE ATT&CK dans la Table I. Des indices textuels et des techniques de reconnaissance sont nécessaires pour comprendre quel cheminement effectuer dans l'infrastructure.

Lors de la mise au point de ce challenge, ces étapes ont été sélectionnées dans le but d'obliger le joueur à mettre en œuvre des tactiques diverses, tout en rendant l'expérience divertissante.

IV. MÉTHODOLOGIE D'ÉTIQUETAGE

Sur chacune des 700 machines virtuelles déployées pour ce CTF, nous avons collecté des événements *auditd*, des logs *syslog* et le trafic réseau à l'aide de *suricata*. Cela correspond à environ 500 Go de trafic réseau et 40 Go de journaux système.

Dans un premier temps, l'étiquetage des journaux systèmes est effectué en utilisant *Sigma*, un format de signature générique. Chaque événement correspondant à une règle *Sigma* est étiqueté avec la technique correspondante de MITRE ATT&CK. Ensuite, les étiquettes sont examinées manuellement pour assurer la qualité de l'étiquetage dans un outil spécialement conçu à cette intention : MANATEE. Comme montré en Figure 1, un analyste junior peut parcourir les logs systèmes (centre de la Figure), repérer la machine concernée (bleu), l'heure et la proposition de technique (rouge).

Étant donné que les données réseau ont un volume beaucoup plus important que les journaux système, nous proposons une approche différente pour les étiqueter. Parmi les commandes *bash* utilisées par les attaquants, nous avons identifiées manuellement celles qui produisaient des évènements réseau. Pour chacune de ces commandes, une requête temporelle est générée au format EQL [12] pour requêter les logs réseau avec ElasticSearch. Cette requête est constituée de la fenêtre

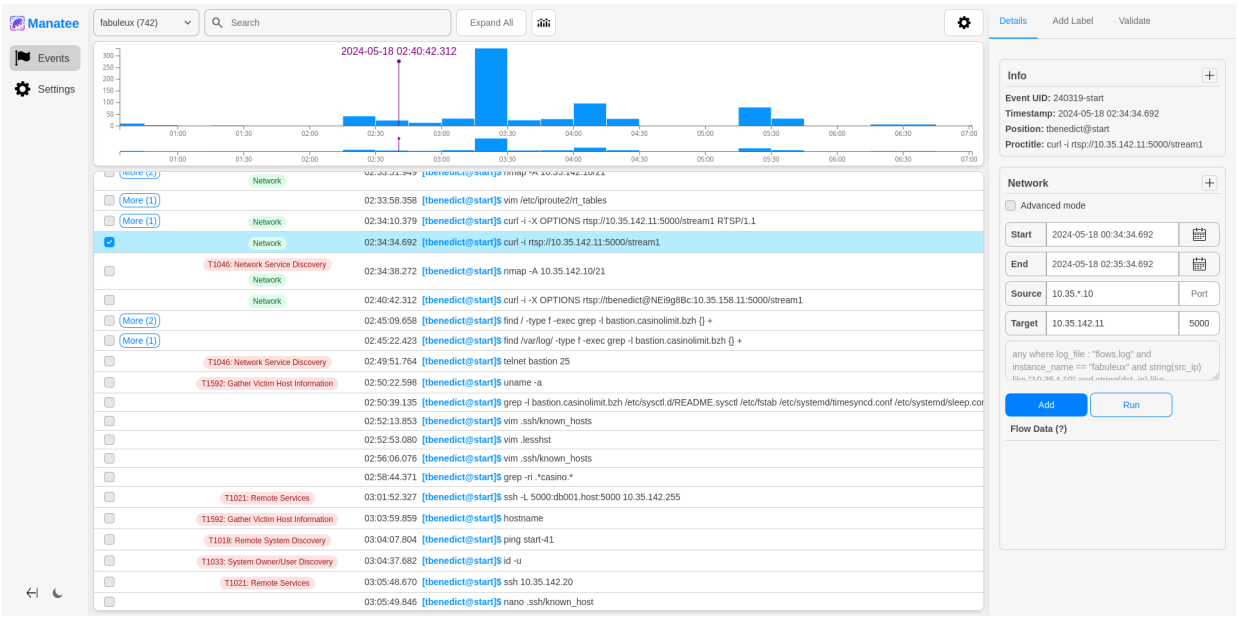

Fig. 1. L'interface d'étiquetage de Manatee

temporelle, de l'IP source et de l'IP de destination. L'IP source est la position d'attaque. L'IP de destination est dans la commande pour la majorité des commandes sauf *nmap*. Si la commande utilisée est *nmap*, l'attaquant précise un masque de sous-réseau que nous ré-utilisons dans la requête. Cette requête est éditable dans la partie de droite de Manatee (Figure 1) dans lequel un tag Network (en vert) est affiché. L'analyste peut alors extraire les logs réseau correspondants et confirmer l'étiquetage. Une fois l'étiquetage, un analyste senior est en charge de confirmer les étiquettes après l'analyste junior. Le processus reste malgré tout long et fastidieux.

V. Analyse et limites

Ces étiquettes permettent d'analyser la progression des joueurs dans l'infrastructure.

A. Tactiques

La Figure 2 présente la matrice de transition des tactiques suivies par les attaquants pour atteindre leurs objectifs Nous observons que les séquences de tactiques ne sont pas aléatoires et se produisent dans un ordre spécifique, représenté sur le système état-transition de la Figure 3. Les attaquants reviennent souvent à la phase de découverte. Cela peut s'expliquer par le fait que, les attaquants doivent collecter régulièrement des informations sur le système pour pouvoir progresser.

B. Techniques

La majorité des techniques utilisées par les attaquants lors de cet exercice sont liées à la découverte, et en particulier à la découverte du réseau comme le montre la Figure 4. Même si cela est cohérent avec l'observation précédente, ce phénomène s'explique également par le fait qu'il existe de nombreuses techniques de découverte qui génèrent une grande quantité d'événements système.

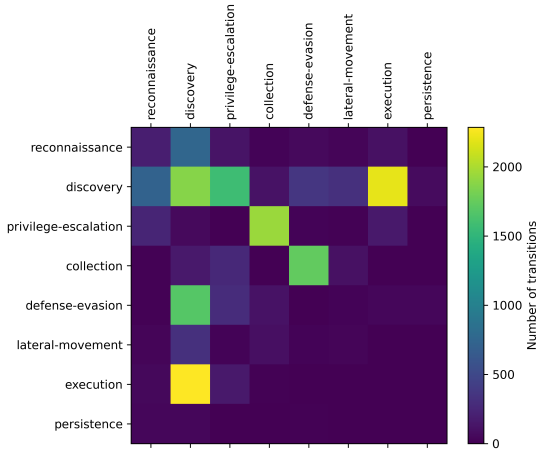

Fig. 2. Matrice de transition des tactiques sur toutes les instances

TABLE II
Statistiques d'utilisation des outils

Outil	Utilisé	Trafic
linpeas	48 instances	3 flux/utilisation
nmap	65 instances	728 flux/utilisation
curl	78 instances	1 flux/utilisation
wget	60 instances	1 flux/utilisation
chisel	5 instances	38 flux/utilisation

C. Procédures

Grâce aux étiquettes sur les journaux d'*auditd*, nous pouvons mettre en évidence la variété de procédures et d'outils utilisés pour chaque technique. La Table II, donne les statistiques d'utilisation des outils permettant la découverte de services et met en évidence la furtivité réseau de ces outils : *curl* (1 flux réseau) est plus furtif que *nmap* (728 flux).

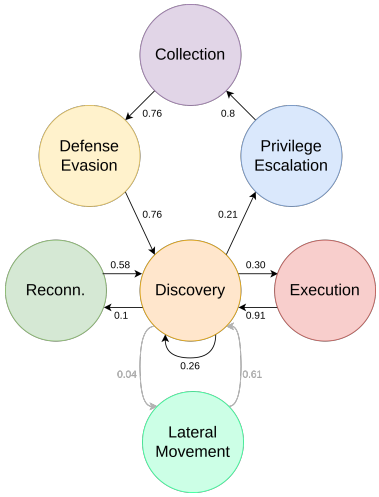

Fig. 3. Représentation des transitions avec plus de 300 occurrences

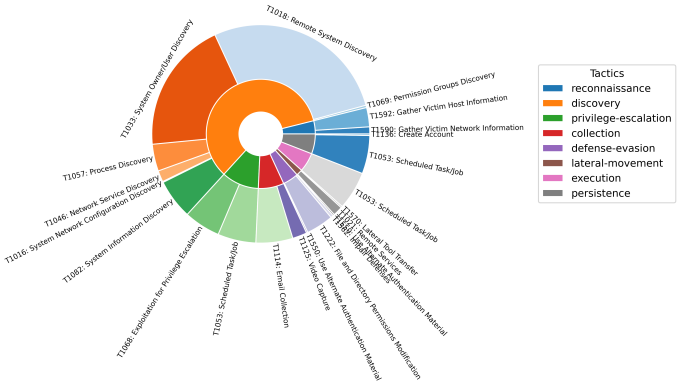

Fig. 4. Répartition des techniques par tactique

D. Limites

Ce jeu de données est composé de 113 sessions d'attaques ayant le même scénario global. Ce scénario ne proposait, dans la limite de nos connaissances qu'un seul chemin d'attaque. Sous cette hypothèse, ce jeu de données permet de comparer le comportement des joueurs pour une même étape. Le nombre d'étapes reste cependant limité. Ce jeu de données ne met pas en évidence de comportements caractéristiques des attaques long terme où l'attaquant quitte et revient dans l'infrastructure, probablement à cause de la limite de 12 heures imposée aux joueurs. Enfin, le manque de trafic bénin provenant d'utilisateurs réguliers facilite le processus d'étiquetage. Ce jeu de données ne peut donc pas être considéré comme représentatif d'un test d'intrusion en environnement réel.

VI. Conclusion et travaux futurs

Nous avons présenté un nouveau jeu de données comportant 113 sessions d'attaques sur 113 instances d'une même infrastructure de 4 machines virtuelles. Ce jeu de données est complètement étiqueté, à la fois au niveau système et réseau.

Nos travaux futurs se concentrent sur l'ajout de trafic bénin et à la génération de challenges plus complexes qui permettraient d'augmenter encore la variabilité des techniques à utiliser. De plus, catégoriser les techniques des attaquants avec une plus grande précision (par exemple avec des sous-techniques) ou l'utilisation d'autre type d'étiquettes permettrait d'élargir encore les applications possibles. Le jeu de données ainsi que l'outil d'étiquetage et d'exploration des données seront accessibles publiquement dans une publication ultérieure en conférence internationale du domaine.

Remerciements

Ce travail bénéficie d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du Plan France 2030 portant la référence ANR-22-PECY-0007.

References

[1] J. L. Guerra, C. Catania, and E. Veas, "Datasets are not enough: Challenges in labeling network traffic," *Computers & Security*, vol. 120, p. 102810, Sep. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404822002048

[2] MITRE, "ATT&CK Framework," 2024. [Online]. Available: https://attack.mitre.org/

[3] R. Uetz, C. Hemminghaus, L. Hackländer, P. Schlipper, and M. Henze, "Reproducible and adaptable log data generation for sound cybersecurity experiments," in *Annual Computer Security Applications Conference*, ser. ACSAC '21. ACM, Dec. 2021, p. 690–705. [Online]. Available: http://dx.doi.org/10.1145/3485832.3488020

[4] G. Agrawal, A. Kaur, and S. Myneni, "A Review of Generative Models in Generating Synthetic Attack Data for Cybersecurity," *Electronics*, vol. 13, no. 2, p. 322, Jan. 2024, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2079-9292/13/2/322

[5] C. Taylor, P. Arias, J. Klopchic, C. Matarazzo, and E. Dube, "CTF: State-of-the-Art and building the next generation," in *2017 USENIX Workshop on Advances in Security Education (ASE 17)*. Vancouver, BC: USENIX Association, Aug. 2017. [Online]. Available: https://www.usenix.org/conference/ase17/workshop-program/presentation/taylor

[6] M. Bashir, C. Wee, N. Memon, and B. Guo, "Profiling cybersecurity competition participants: Self-efficacy, decision-making and interests predict effectiveness of competitions as a recruitment tool," *Computers & Security*, vol. 65, pp. 153–165, Mar. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404816301389

[7] R. S. Gutzwiller, M. Gilbert, T. J. Drescher, K. J. Ferguson-Walter, N. Mikanda, C. J. Johnson, and D. D. Scott, "Frustration, confusion, surprise, confidence, and self-doubt: Cyber operators' affects during a realistic experiment," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 67, no. 1, pp. 233–239, 2023. [Online]. Available: https://doi.org/10.1177/21695067231192883

[8] G. Savin, A. Asseri, J. Dykstra, J. Goohs, A. Melarano, and W. Casey, "Battle Ground: Data Collection and Labeling of CTF Games to Understand Human Cyber Operators," in *2023 Cyber Security Experimentation and Test Workshop*, Aug. 2023, pp. 32–40, arXiv:2307.10877 [cs]. [Online]. Available: http://arxiv.org/abs/2307.10877

[9] M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, "Flow-based benchmark data sets for intrusion detection," in *Proceedings of the 16th European conference on cyber warfare and security. ACPI*, 2017, pp. 361–369.

[10] M. Landauer, F. Skopik, M. Wurzenberger, W. Hotwagner, and A. Rauber, "Have it Your Way: Generating Customized Log Datasets With a Model-Driven Simulation Testbed," *IEEE Transactions on Reliability*, vol. 70, no. 1, pp. 402–415, Mar. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9262078

[11] J. Gjerstad, F. Kadiric, G. Grov, E. H. Kjellstadli, and M. Leira Asprusten, "LADEMU: a modular & continuous approach for generating labelled APT datasets from emulations," in *2022 IEEE International Conference on Big Data (Big Data)*, Dec. 2022, pp. 2610–2619. [Online]. Available: https://ieeexplore.ieee.org/document/10020549

[12] Elastic, "Event Query Language (EQL)," 2025. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/current/eql.html