Offre de thèse

Intitulé : Développement de modèles de détection et de prévention d'attaques d'exfiltration de données à travers la stéganographie réseau

Mots-clés : Exfiltration de données, Stéganographie réseau, Modèles de détection, Modèles de prévention, Intelligence Artificielle, Techniques d'apprentissage

Contacts : Mohammed NAFI (mohammed.nafi@irisa.fr), Gilles Guette (gilles.guette@imt-atlantique.fr) et Guillaume Doyen (guillaume.doyen@imt-atlantique.fr)

Information générale :

Etablissement d'accueil : ENSSAT, Lannion

Laboratoire d'accueil : Equipe SOTERN, IRISA

Durée de thèse : 36 mois

Période : 2026-2029

Liste des encadrants :

Encadrant de thèse : Mohammed NAFI

Directeur de thèse : Gilles Guette

Co-directeur de thèse : Guillaume DOYEN

Description du sujet de thèse

Contexte

De nos jours, une très grande quantité de données sont générées quotidiennement au sein des organisations. La sécurité des données sensibles et privées demeure au cœur des préoccupations et constitue l'une des priorités de toute organisation. Pour préserver la confidentialité de ses données, les techniques de cryptographie et de stéganographie sont en général utilisées. La cryptographie tente de rendre le contenu des données sensibles illisible aux entités non autorisées tandis que la stéganographie vise à dissimuler les informations dans un autre média (objet) de couverture (par exemple, texte, image, audio, vidéo), de manière à

ce que l'existence même de ces informations soit invisible et difficile à détecter. Parmi les techniques de stéganographie, on trouve la stéganographie réseau, appelée également stéganographie de protocole qui consiste à cacher des données dans les protocoles réseau communément utilisés dans la transmission des données tels que TCP/IP, UDP, DNS, HTTP, ICMP, etc. Bien que la stéganographie soit souvent utilisée par des entités légitimes pour cacher des données sensibles lors de leur transfert sur un réseau, des attaquants exploitent également cette technique pour mener des attaques sophistiquées, comme l'exfiltration de données.

L'exfiltration de données désigne en effet le transfert illicite et non autorisé d'informations sensibles ou privées à partir d'un réseau d'une organisation vers une destination externe contrôlée par un attaquant. Cette attaque peut être menée par un membre interne ou externe à l'organisation et peut avoir des conséquences désastreuses pour l'organisation ciblée (réputation, etc.). L'exfiltration de données à travers la stéganographie réseau est ainsi un domaine de recherche très actif à l'heure actuelle. Son principe, qui consiste à dissimuler les données dérobées dans le trafic réseau habituel, la rend très difficile à détecter, d'autant plus que le contenu dissimulé est parfois chiffré. C'est dans ce cadre que s'inscrivent les travaux de cette thèse.

Description des principaux verrous et travaux envisagés

Parmi les nombreuses approches traitant de l'exfiltration de données par des méthodes de stéganographie réseau, nous nous intéresserons principalement à celles basées sur l'Intelligence Artificielle (IA) et les techniques d'Apprentissage (Machine Learning, Deep Learning, Reinforcement Learning, etc.) qui s'avèrent à la fois adéquates et prometteuses pour résoudre ce type de problème. Plusieurs verrous de recherche sont à lever. Ces derniers sont comme suit :

- Le premier est d'exploiter les modèles d'IA et des techniques d'apprentissage afin de mettre en place de nouvelles solutions qui seront capables de détecter la fuite de données et les attaques d'exfiltration de données fondées sur la stéganographie réseau. Nous tenterons de répondre ainsi à la question « Comment détecter les attaques d'exfiltration de données à travers la stéganographie réseau ? ».

- Le deuxième consiste à développer des solutions réactives plus efficaces qui réagissent rapidement à une occurrence d'attaque d'exfiltration qui se produit pour arrêter la fuite de données et limiter les conséquences, ce qui permettra de répondre à la question « Comment réagir rapidement face à une attaque d'exfiltration de données ? ».

- Enfin, nous visons à proposer des solutions proactives qui identifient les mesures à prendre afin de se prémunir contre l'exfiltration de données. Nous tenterons donc d'apporter des réponses à la question « Comment se prémunir contre l'attaque d'exfiltration de données ? »

Approche méthodologique et critères de qualité des résultats obtenus

Les principaux travaux de cette thèse se dérouleront de la manière suivante :

- T0-T0+6 : le doctorant devra effectuer une revue de littérature sur le sujet pour élaborer un état de l'art complet à la fois sur les techniques d'exfiltration de donnés fondées sur la stéganographie réseau et les contremesures proposées dans la littérature. Une classification et une étude comparative des différentes approches étudiées seront également attendues.

- T0-T0+18 : Il sera question ici de mettre en place une plateforme qui permette de reproduire plusieurs attaques d'exfiltration de données par des mécanismes stéganographiques afin d'en comparer la modalité (en vue de préparer les travaux suivant sur la détection) et la performance. On utilisera pour ce faire des travaux caractéristiques de l'état de l'art scientifiques ou associés à des attaques réelles (sous réserve de la disponibilité de cette information). Ce travail pourra donner lieu à la publication du jeu de données collectés.

- T0+18-T0+30 : Le travail de cette seconde partie de la thèse consistera à mettre en place des approches de détection/prévention originales et performantes pour pallier le problème d'exfiltration de données. Les performances des solutions proposées devront être évaluées non seulement en matière d'efficacité et de robustesse de la détection/protection, mais aussi en termes de consommation de ressources (énergie, mémoire et calcul). Les résultats obtenus seront, par la suite, comparés avec ceux de certaines autres approches pertinentes de l'état de l'art.

- T0+30 – T0+36 : le doctorant se concentrera sur la rédaction de son manuscrit de la thèse et la préparation de sa soutenance.

Les travaux réalisés dans le cadre de cette thèse (état de l'art approfondi, jeu de données, nouvelles solutions de détection, réaction et protection) feront l'objet d'articles qui seront soumis dans des journaux et conférences de renommée internationale dans le domaine de la sécurité des réseaux.

Bibliographie

[1] Ganivev, A., Mavlonov, O., & Turdibekov, B. (2021, November). Improving data hiding methods in network steganography based on packet header manipulation. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.

[2] Ozery, Y., Nadler, A., & Shabtai, A. (2024). Information based heavy hitters for real-time DNS data exfiltration detection. In Proc. Netw. Distrib. Syst. Secur. Symp (pp. 1-15).

[3] Zillien, S., Petrov, D., Ruffing, P., & Gross, F. (2024, June). A Development Framework for TCP/IP Network Steganography Malware Detection. In Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security (pp. 95-100).

[4] King, J., Bendiab, G., Savage, N., & Shiaeles, S. (2021, July). Data exfiltration: methods and detection countermeasures. In 2021 IEEE International Conference on Cyber Security and Resilience (CSR) (pp. 442-447). IEEE.

[5] Kadebu, P., Shoniwa, R. T., Zvarevashe, K., Mukwazvure, A., Mapanga, I., Thusabantu, N. F., & Gotora, T. T. (2023). A hybrid machine learning approach for analysis of stegomalware. International Journal of Industrial Engineering and Operations Management, 5(2), 104-117.

[6] Apau, R., Asante, M., Twum, F., Ben Hayfron-Acquah, J., & Peasah, K. O. (2024). Image steganography techniques for resisting statistical steganalysis attacks: A systematic literature review. PloS one, 19(9), e0308807.

[7] Das, A., Shen, M. Y., Shashanka, M., & Wang, J. (2017, December). Detection of exfiltration and tunneling over DNS. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 737-742). IEEE.

[8] Sasikala, V., & CH, B. S. (2024, March). Data Leakage Detection and Prevention Using Ciphertext-Policy Attribute Based Encryption Algorithm. In 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE.

[9] Seo, J. O., Manoharan, S., & Mahanti, A. (2016, August). A discussion and review of network steganography. In 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech) (pp. 384-391). IEEE.

[10] Lubacz, J., Mazurczyk, W., & Szczypiorski, K. (2014). Principles and overview of network steganography. IEEE Communications Magazine, 52(5), 225-229.

[11] Bedi, P., & Dua, A. (2020). Network steganography using the overflow field of timestamp option in an IPv4 packet. Procedia Computer Science, 171, 1810-1818.

Profil recherché :

Nous sommes à la recherche d'un étudiant titulaire d'un diplôme de Master 2 ou équivalent dans le domaine de l'informatique possédant des compétences avérées en :

- Sécurité Informatique
- Réseaux Informatique
- Intelligence artificielle
- Programmation
- Communication

Candidature :

Merci d'adresser votre dossier de candidature, constitué des pièces suivantes, par mail aux contacts mentionnés ci-dessus :

- CV
- Lettre de motivation
- Relevés de notes Licence/ Master et votre classement
- Une ou plusieurs lettre(s) de recommandation