Cybersécurité à l'échelle intersystème d'information avec prise en compte du facteur humain

Olivier de Casanove

Cybersécurité à l'échelle intersystème d'information avec prise en compte du facteur humain

Olivier de Casanove

IRIT, Université Toulouse III – Paul Sabatier
118 route Narbonne 31062 Toulouse CEDEX 9
olivier.decasanovenom@irit.fr

MOTS-CLES : cybersécurité, sécurité système information, détection d'évènements, prévention

KEYWORDS : cybersecurity, information system security, event detection, prevention

ENCADREMENT : Florence Sèdes

1. Introduction

L'ANSSI (Agence Nationale de la Sécurité de Systèmes d'Information) définit la cybersécurité ainsi (ANSSI, 2022) : "État recherché pour un système d'information lui permettant de résister à des évènements issus du cyberespace susceptibles de compromettre la disponibilité, l'intégrité ou la confidentialité des données stockées, traitées ou transmises [...]". Cette définition ne fait pas l'unanimité et chaque pays a sa propre vision de la cybersécurité, mais elle suffit pour en comprendre les enjeux principaux. Le système d'information (SI) est au cœur des politiques de cybersécurité, l'objectif est de le protéger. Les évènements (ou incidents) de sécurité auxquels fait référence la définition sont décrits par une chronologie qui peut être découpée en plusieurs étapes. Comme pour la définition de la cybersécurité, il existe de nombreuses versions de la chronologie d'un évènement de sécurité. La plus simple et explicite est celle du NIST (National Institute of Standards and Technology) qui reconnait quatre étapes (Scarfone *et al.*, 2008) : préparation, prévention ; détection, identification ; isolement, éradication, remédiation ; apprentissage. Les solutions de cybersécurité développées à chaque étape de cette chronologie sont déployées à l'échelle du SI. Cependant, avec l'étude de Code Red, un ver informatique, Moore *et al.* ont pour la première fois formalisé le fait que la cybersécurité d'un SI dépend aussi de la sécurité des autres SI (Moore *et al*., 2002). Ce changement de paradigme implique que les outils développés actuellement ne sont pas suffisants pour détecter les incidents de sécurité, puisqu'ils ne sont utilisables qu'à une échelle locale. Dans cet article, nous proposons une approche

non plus à l'échelle du SI, mais à une échelle intersystème d'information. Dans la section 2 nous expliquons comment nous comptons détecter les attaques informatiques à l'échelle intersystème d'information. Dans la section 3 nous proposons une approche permettant de tirer avantage de la détection pour améliorer la cybersécurité. Nous concluons dans la section 4.

2. Détection à l'échelle intersystème d'information

Pour détecter les attaques à l'échelle intersystème d'information, il faut un jeu de données à la même échelle. Nous avons choisi d'utiliser les réseaux sociaux pour cela, car la surveillance des réseaux sociaux pour détecter les attaques informatiques s'est déjà montrée efficace (Khandpur *et al.*, 2017, Ritter et al., 2015, Sabottke et al., 2015, Sceller *et al.*, 2017). Pour détecter les attaques, nous utiliserons donc des algorithmes de détection d'évènements : (Atefeh et Khreich, 2015) en fournissent une revue de la littérature.

Le principal inconvénient de cette approche est que n'importe quel utilisateur malveillant peut poster des messages sur les réseaux sociaux. Les données d'entrées sont donc facilement corruptibles. La littérature sur la détection de spams, sujet sur lequel nous avons déjà travaillé dans l'équipe (Washha et al., 2016), est abondante et détaillée dans plusieurs revues de la littérature récentes (Gheewala et Patel, 2018, Tingmin et al., 2018, Yurtseven *et al.*, 2021). Cependant, dans notre cas, le problème s'avère plus complexe. Les algorithmes de détection sont des algorithmes d'apprentissage automatique et sont donc sensibles à l'apprentissage adverse (ou *adversarial learning* en anglais). L'apprentissage adverse est une technique utilisée par un attaquant pour exploiter à son avantage un algorithme d'apprentissage automatique qui ne lui appartient pas, le plus souvent dans le but d'altérer les données de sorties de l'algorithme (Xianmin *et al.*, 2019). Dans notre cas cela voudrait dire qu'un attaquant publie des messages sur les réseaux sociaux, spécialement pour que la détection d'évènements soit moins efficace. Dans notre contexte, détecter des évènements de cybersécurité, il n'est pas pensable de concevoir un système qui ne puisse pas être résilient à des acteurs malveillants. Il existe une littérature sur l'apprentissage adverse sur les réseaux sociaux, mais elle est centrée autour de la question de détection de spams (Imam et Vassilakis, 2018). Nous avons contribué à l'état de l'art grâce à une première typologie pour modéliser les enjeux de l'apprentissage adverse dans le contexte de la détection d'évènements et nous travaillons actuellement à développer des algorithmes permettant de s'en défendre.

La détection d'évènements de sécurité à l'échelle intersystème d'information a déjà un intérêt en soi, mais il est possible d'en tirer une plus-value supplémentaire grâce à la prévention.

3. Prévention, Éducation et Sensibilisation à la sécurité

Les *malwares* pouvant passer d'un ordinateur à l'autre, la sécurité d'un SI ne dépend plus de sa propre sécurité uniquement, mais de la sécurité de toutes les machines connectées à internet. Moore *et al.* formalisent alors pour la première fois l'aspect épidémique des attaques informatiques *(*Moore *et al.*, 2002). Les auteurs proposent trois moyens de lutte : la prévention, la remédiation et l'isolation. Les auteurs ont traité le sujet de l'isolation, nous choisissons de nous intéresser à la prévention dans ce travail. La prévention en sécurité informatique n'est pas un concept nouveau, mais est souvent vue comme un plus, quelque chose de non nécessaire. Une des raisons de la mauvaise utilisation de la prévention est la difficulté de pouvoir quantifier le nombre d'attaques évitées et ainsi prouver l'utilité de la prévention. Pour contrer ce phénomène, nous avons établi une revue de la littérature sur la prévention dans le milieu de la cybersécurité (de Casanove et Sèdes, 2021). Dans notre revue nous proposons d'adapter les campagnes de prévention au cycle PDCA (Plan Do Check Adjust) pour que la prévention puisse être utilisée comme un outil de cybersécurité à part entière qui s'intègre dans une politique globale, et non plus comme un agrément. L'objectif est de déclencher des campagnes de prévention au moment opportun pour lutter contre les attaques informatiques, moment qui est détecté au plus tôt grâce à notre proposition de système de détection intersystème d'information. Dans le meilleur des cas, la campagne de prévention arrive avant que le SI soit compromis et on évite une compromission. Dans le pire des cas, le système est déjà compromis et la campagne permet aux utilisateurs d'être plus vigilants et de noter plus tôt des comportements étranges ou inhabituels du SI.

4. Conclusion

Dans la première section, nous avons énoncé les limites du modèle actuel de la cybersécurité qui se concentre sur la détection et le traitement des attaques informatiques à l'échelle du SI. Dans la deuxième section, nous avons vu comment nous comptons implémenter un système de détection de cyberattaque à l'échelle intersystème qui puisse être résilient à l'apprentissage adverse. Dans la troisième section, nous valorisons la détection intersystème en utilisant la prévention comme moyen de lutte contre les cyberattaques.

Bibliographie

ANSSI. (2022). *Glossaire cybersécurité*, https://www.ssi.gouv.fr/administration/glossaire/c/

Atefeh F., Khreich W. (2015). *A Survey of Techniques for Event Detection in Twitter.Computational Intelligence*, vol. 31, no1, p. 132–164.

de Casanove O., Florence Sèdes S. (2022). *Extracting Guidelines for Security Education,Training and Awareness Programme from the Literature.* (working paper or preprint).

Gheewala S., Patel R.. (2018). Machine Learning Based Twitter Spam Account Detection: A Review. *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 79-84.

Imam N. H., Vassilakis V.G. (2019).. A Survey of Attacks Against Twitter Spam Detectors in an Adversarial Environment. *Robotics* 8, no. 3: 50.

Khandpur R. P., Ji T., Jan S., Wang G., Lu C.-T., Ramakrishnan N. (2017). Crowdsourcing cybersecurity: Cyber attack detection using social media. InProceedings of the *2017 ACM on Conference on Information and Knowledge Management*, p. 1049–1057.

Le Sceller Q., Karbab E. M. B., Debbabi M., Iqbal F. (2017). SONAR: Automatic Detection of Cyber Security Events over the Twitter Stream. *In Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES '17).* Association for Computing Machinery, New York, NY, USA, Article 23, 1–11.

Moore D., Voelker G. M., Savage S. (2002). *Quantitative network security analysis.* Cooperative Association for Internet Data Analysis (CAIDA), NSF-01-160, vol. 7.

Ritter A., Wright E., Casey W., Mitchell T. (2015). Weakly supervised extraction of computer security events from twitter. InProceedings of the 24th international conference on worldwide web, p. 896–905. Republic and Canton of Geneva, CHE, *International World Wide Web Conferences Steering Committee.*

Sabottke C., Suciu O., Dumitras T. (2015). Vulnerability disclosure in the age of so-cial media: Exploiting twitter for predicting real-world exploits. In *24th USENIX security symposium* (USENIX security 15), p. 1041–1056. Washington, D.C., USENIX Association.

Scarfone K. A., Grance T., Masone K. (2008). *Computer Security Incident Handling Guide.*

Tingmin Wu, Sheng Wen, Yang Xiang, Wanlei Zhou (2018). Twitter spam detection: Survey of new approaches and comparative study, *Computers & Security*, Volume 76, Pages 265-284.

Washha M., Qaroush A., Sedes F. (2016). Leveraging time for spammers detection on twitter. InProceedings of the *8th international conference on management of digital ecosystems*, p. 109–116. New York, NY, USA, Association for Computing Machinery.

Xianmin Wang, Jing Li, Xiaohui Kuang, Yu-an Tan, Jin Li. (2019). The security of machine learning in an adversarial setting: A survey, *Journal of Parallel and Distributed Computing*, Volume 130, Pages 12-23.

Yurtseven I., Bagriyanik S. and Ayvaz S. (2021). A Review of Spam Detection in Social Media, *2021 6th International Conference on Computer Science and Engineering (UBMK)*, pp. 383-388.