

Recherche de règles de corrélation dans des graphes de causalité

Encadrant : Eric Totel (eric.totel@telecom-sudparis.eu)

Les attaques contre les systèmes informatiques deviennent de plus en plus complexes. Il est nécessaire d'offrir des outils à l'analyste pour mieux détecter et comprendre les attaques. Historiquement, afin de remplir cet objectif, on fournit à l'administrateur des SIEMs (Security Information and Event Management), qui effectuent de la corrélation sur les événements et alertes produits par des IDS. Cette corrélation repose sur des règles relativement simples [1] (telles que $A \text{ ET } B \text{ ET } C$ où A, B, C sont des événements ou des alertes) grâce auxquelles on peut déterminer que plusieurs événements et alertes sont significatifs d'une même attaque. Malheureusement, écrire ces règles de corrélation est relativement difficile, ce qui peut mener à des règles erronées.

D'un autre côté Xosanavongsa [2] a défini la notion de causalité entre événements hétérogènes. A partir de cette notion de causalité, il exprime ce qu'est un graphe causal entre des logs hétérogènes. Ce graphe causal est donc capable d'exprimer des relations telles que $A \rightarrow B \rightarrow C$ (où $\rightarrow$ exprime *dépend causalement de*), relation que n'est pas capable d'exprimer un SIEM. Toutefois, de cette expression on peut déduire la règle $A \text{ ET } B \text{ ET } C$ que comprend un SIEM.

Viet Huy Ha [3] montre que par l'observation d'un système on est capable d'observer ces graphes de dépendances causales entre événements, et de les construire efficacement. Pour cela, il instrumente le système Linux grâce à eBPF et génère des graphes de dépendance causale entre appels systèmes. De plus, il montre que la causalité permet d'expliquer des attaques, en remontant le graphe de dépendance causale à partir d'un indice de compromission (IoC).

Le but de ce projet est d'utiliser ces graphes de dépendances causales, afin d'en extraire des règles de corrélation utilisables par des SIEMs. On pourra utiliser des algorithmes de parcours de graphes, mais aussi des LLM pour ceux que cela intéresserait.

Références :

- [1] E. Totel, B. Vivinis, et L. Mé, « A Language Driven Intrusion Detection System for Event and Alert Correlation », in *Proceedings of the 19th IFIP International Information Security Conference*, Toulouse: Kluwer Academic, août 2004, p. 209-224. [En ligne]. Disponible sur: <http://www.rennes.supelec.fr/ren/perso/etotel/publications.html>
- [2] C. Xosanavongsa, E. Totel, et O. Bettan, « Discovering Correlations: A Formal Definition of Causal Dependency Among Heterogeneous Events », in *2019 IEEE European Symposium on Security and Privacy (EuroS P)*, juin 2019, p. 340-355. doi: 10.1109/EuroSP.2019.00033.
- [3] V.-H. Ha, V. Gauthier, et E. Totel, « Generating causal logs to explain attacks », in *CRISIS 2025 : 20th International Conference on Risks and Security of Internet and Systems*, Gatineau, Canada, oct. 2025. Consulté le: 16 juillet 2026. [En ligne]. Disponible sur: <https://hal.science/hal-05494614>