Analyse forensique et classification des ransomware : une étude des méthodes existantes et des perspectives

Anais MOUSSOUNI Frédéric FAUBERTEAU Nga NGUYEN
De Vinci Higher Education, De Vinci Research Center, Paris, France
anais.moussouni, frederic.fauberteau, nga.nguyen@devinci.fr

***Résumé*—Les attaques ransomware, de plus en plus complexes et fréquentes, constituent des menaces importantes pour les organisations critiques. Une analyse forensique et une classification efficaces des ransomware sont nécessaires pour permettre une meilleure compréhension des attaques et réussir à en identifier l'auteur. Cet article présente une étude des méthodes existantes pour l'analyse forensique et la classification des ransomware. En mettant en lumière les forces et les faiblesses des différentes approches, nous identifierons de futures pistes de recherches prenant en compte la gestion des contraintes des enquêtes forensiques telles que réalisées par les forces de l'ordre et proposerons des orientations pour améliorer les techniques d'analyse et de classification des ransomware.**

***Index Terms*—ransomware, classification, apprentissage automatique**

I. Introduction

Selon une étude de Comparitech [1], en 2024, 195,4 millions d'enregistrements de données ont été compromis par des attaques ransomware. Les attaques ransomware sont l'une des formes de malwares les plus coûteuses, en chiffrant ou en exfiltrant les données des victimes puis en exigeant une rançon pour permettre de récupérer les clés de déchiffrement ou les données elles-mêmes. Au fil des années, les ransomware se sont diversifiés et complexifiés grâce à l'adoption de techniques d'obfuscation, d'outils d'anonymisation et de malwares polymorphiques capables de modifier leur code pour échapper à la détection. De plus, l'organisation de plus en plus sophistiquée des attaquants a fait naître de nouvelles formes de « business model », comme les Ransomware-as-a-Service (RaaS), complexifiant davantage l'analyse de l'écosystème des attaques ransomware [2]. L'objectif de l'analyse forensique et de la classification des ransomware est d'identifier les familles de ransomware, de reconstruire les schémas d'attaques et d'arriver à attribuer les attaques à leurs véritables auteurs. Les techniques forensiques doivent relever le défi de fonctionner dans des situations de terrain telles qu'expérimentées par les forces de l'ordre. Le manque de données et d'artefacts, qui sont les traces laissées par des activités ou des événements sur un système, la capacité à faire face à des attaques inconnues, dites « zero-day », ainsi qu'à de nouveaux types de menaces comme les malwares sans fichiers sont des enjeux cruciaux dans la réalité du terrain [3]. Cet article explore l'état actuel de l'analyse forensique et de la classification des ransomware, évalue les méthodologies existantes et identifie les manques. Des propositions de travaux futurs sont faites pour répondre à ces défis, en se concentrant sur l'apprentissage non supervisé et l'analyse des artefacts forensiques non volatiles.

II. Analyse de ransomware

Dans cette section, nous présentons les attaques ransomware, puis nous introduisons les différents types d'analyses et d'artefacts à disposition.

A. Les attaques ransomware

Quel que soit le type d'attaques ransomware la diffusion de l'attaque dans le système cible suit une séquence d'étapes clefs [2], [4] :

1) Distribution: Cette étape consiste à distribuer le ransomware à la cible par le biais d'un vecteurs d'attaque choisi par l'attaquant. Les vecteurs d'attaque les plus fréquents incluent des techniques d'ingénierie sociale comme les campagnes de phishing ou des sites malveillants, mais également l'exploitation de failles dans des logiciels, ou y installent eux-mêmes des backdoors, ou des vulnérabilités dans le Remote Desktop Protocol (RDP), qui est un protocole de connexion à distance au système d'une organisation.

2) Infiltration et installation: Lors de l'infiltration, le fichier malveillant est téléchargé et le ransomware est installé sur le système de la victime. Il s'intègre dans le système pour être indétectable et survivre à un redémarrage. Le code malveillant établit une connexion avec son serveur de commande et de contrôle, qui est contrôlé par les attaquants.

3) Scan: Le ransomware scan le système de la victime à la recherche de fichiers à chiffrer, incluant des fichiers enregistrés localement mais également des données accessibles via le réseau ou le Cloud.

4) Chiffrement / Exfiltration: Les ransomware sont classifiés selon leur mode opératoire et sont répartis en deux catégories principales. Les Crypto ransomware, qui effectuent le chiffrement des données de la victime en utilisant des algorithmes de chiffrement tels que Advanced Encryption Standard (AES) ou RSA, et le Locker ransomware, qui bloque l'accès au système, le rendant inaccessible aux victimes. [4] Les attaquants peuvent également utiliser une combinaison de chiffrement et de vol de données, et menacer de divulguer ces dernières sur des murs de la honte publiés dans le Darknet.

5) Demande de rançon: L'étape du paiement correspond à la diffusion de la demande de rançon, que l'on appelle en anglais ransomnote. Les ransomnotes sont le moyen pour les attaquants de communiquer avec leurs victimes et d'exprimer leurs conditions. Sous forme de fichier .txt ou .html, elles contiennent les informations nécessaires pour procéder au paiement de la rançon, souvent demandée en crypto-monnaies. Elles peuvent contenir des indicateurs de compromission (IOC) utiles aux forces de l'ordre pour retrouver l'attaquant tel qu'un lien vers un cryptowallet pour effectuer le paiement ou un accès à un espace dédié sur le Darknet pour interagir avec la victime ou revendiquer l'attaque.

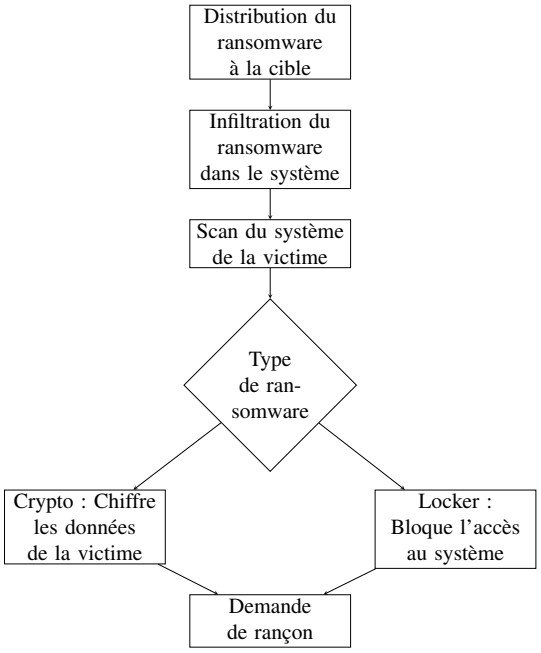

FIGURE 1. Schéma d'infection par ransomware

B. Méthodes d'analyse

1) Analyse Statique: L'analyse statique consiste à examiner le malware sans exécuter le programme. Cette méthode consiste à désassembler, décompiler et débugger pour analyser la structure du malware et identifier des IOC potentiels, tels que des chaînes de caractères, des appels API ou des commandes suspectes [5], [6]. Cependant, cette méthode présente des limites, car elle nécessite l'accès au fichier binaire du malware, et peut être contournée par des techniques d'obfuscation qui masquent les véritables intentions du malware, comme le chiffrement ou l'utilisation de « packers » [7].

2) Analyse Dynamique: L'analyse dynamique implique l'exécution du malware dans un environnement contrôlé, tel qu'une machine virtuelle (VM), une sandbox ou un émulateur. Cette approche permet d'observer le comportement du malware en temps réel, et d'obtenir des informations concrètes sur ses actions et interactions avec le système (chiffrement des fichiers, création de processus, communication réseau, modifications du registre, etc.) [5], [6]. Cependant, l'analyse dynamique présente également la limite de nécessiter le binaire du malware, et elle peut être compromise par la capacité du malware à détecter l'environnement d'analyse. Dans ce cas, il peut utiliser des techniques d'évasion, c'est-à-dire adopter un comportement bénin au lieu d'exécuter ses actions malveillantes réelles.

3) Analyse Hybride: L'analyse hybride combine à la fois l'analyse statique et dynamique. En intégrant les deux méthodes, l'analyse hybride fournit une compréhension plus complète du comportement et des caractéristiques du malware. Par exemple, l'examen statique identifie des IOC potentiels, ce qui permet d'orienter l'analyse dynamique et de l'utiliser pour confirmer le comportement malveillant [5]. Cependant, si les méthodes hybrides combinent les avantages des méthodes statiques et dynamiques, elles combinent également leurs limites. En effet, elles sont elles aussi limitées par l'absence des fichiers binaires, ainsi que par des techniques d'évasion de malwares.

C. Types d'Artefacts

1) Artefacts Volatiles: Les données volatiles peuvent fournir des informations cruciales sur le comportement en temps réel du malware, ce sont des données utilisées dans le cas d'analyse dynamique. Les données volatiles incluent le contenu de la Random Access Memory (RAM), et permettent une approche de détection des attaques en direct [7], [8]. Cependant, l'utilisation de ce type de données est limitée dans le cas d'enquêtes post-mortem car ces données sont effacées après un redémarrage du système. Lors d'une attaque l'objectif de l'organisation ciblée est de remettre le système en état de fonctionnement, ainsi l'analyse post-mortem est souvent réalisée après que ce dernier ne soit redémarré, et donc, les données effacées. Certains types de ransomware sont en mesure de modifier et d'effacer certaines données ou de camoufler leur passage, ce qui rend l'usage de ces données d'autant plus complexe, à la fois pour des raisons de disponibilité et d'intégrité [9].

2) Artefacts Non Volatiles: Les données non volatiles incluent des registres, fichiers prefetch, journaux événements etc.. Elles fournissent un enregistrement à long terme des activités et des interactions du malware avec le système [7], [10]. Ces données sont essentielles pour l'analyse forensique, car elles restent disponibles même après un redémarrage du système. Cela est particulièrement pertinent dans le cas d'analyses post-mortem sans accès au binaire, par exemple l'analyse des fichiers prefetch permet de classifier et identifier le malware à partir de traces laissées lors de l'attaque [10]. Cependant, les données non volatiles constituent un volume important à collecter et à traiter. De plus, certains malwares sont capables de modifier ou d'effacer des données dans l'ensemble de données non volatiles, impactant l'intégrité de celles-ci.

III. REVUE DES MÉTHODES ACTUELLES

Dans cette section, nous détaillons les méthodes de détection et de classification existantes.

A. Classification Basée sur les Signatures

La classification basée sur les signatures est une approche statique qui repose sur l'identification de signature spécifique du malware, comme des chaînes de caractères ou des hash de fichiers, et sa comparaison avec des signatures de malware connues et répertoriées dans une base de données (eg : VirusTotal [11]). Cette méthode d'analyse est rapide et efficace dans le cas de malware connus, mais ne permet pas l'identification des malwares « zero-day »ou de malwares utilisant des techniques d'obfuscation et de chiffrement, qui empêchent une correspondance aux signatures existantes [12], [13].

B. Classification Comportementale et Heuristique

La classification comportementale est une approche dynamique qui analyse un programme pendant son exécution pour déterminer s'il présente un comportement malveillant en se basant sur l'adoption de comportements suspects, tels que le chiffrement de fichiers et les demandes de rançon.

La classification heuristique est une approche qui se base sur la création de règles permettant d'identifier des comportements malveillants [13]. Elle peut s'appuyer sur des éléments issus de l'analyse statique (comme des chaînes de caractères suspectes ou des structures de code inhabituelles) mais aussi sur des observations issues de l'analyse dynamique (comme certaines séquences d'actions du programme). Elle permet de détecter des ransomware « zero-day »car elle ne dépend pas de signatures spécifiques.

Cependant, ces méthodes de classification peuvent générer des faux positifs, car les programmes bénins peuvent être considérés comme des programmes malveillants, en ayant un certain nombre de caractéristiques et comportements en commun avec ces derniers [12]. Elles présentent également les mêmes limites que les analyses statiques et dynamiques, car elles nécessitent le binaire du malware et peuvent être contournée par des techniques d'obfuscation et d'évasion.

C. Classification par Apprentissage Automatique (Machine Learning) / Apprentissage Profond (Deep Learning)

Les algorithmes d'apprentissage automatique et d'apprentissage profond ont fait leurs preuves dans la détection et la classification des ransomware.

1) Apprentissage Automatique Supervisé : Les algorithmes supervisés, tels que Random Forest, Support Vector Machine (SVM) et Naïve Bayes nécessitent des jeux de données étiquetés pour entraîner les modèles. Ces méthodes peuvent atteindre une précision élevée pour les familles de ransomware connues [14], [15]. Cependant, ces algorithmes dépendent des caractéristiques observées à partir du fichier binaire et sont donc inefficaces faces aux ransomware « zero-day » [12].

2) Apprentissage Automatique Non Supervisé: Les algorithmes non supervisés, tels que les Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) sont performants sur des ensembles de données brutes, soit non-étiquetées [14]. Ces méthodes utilisent des artefacts comme données d'entrée pour identifier des comportements suspects sur le système, déterminer la famille de ransomware en fonction de caractéristiques communes. Ils sont donc capables d'identifier des malwares « zero-day », car ils sont indépendants des binaires. Des études comparatives mettent en avant les performances supérieures des modèles d'apprentissage profond dans la détection et la classification de malwares [16]. Cependant, les résultats des modèles d'apprentissage dépendent du volume et de la qualité des données d'entrée, et nécessitent des ressources importantes pour l'entraînement des modèles. Ces modèles peuvent également être biaisés par les données d'entrée. Si les données d'entraînement sont mal échantillonnées, certains types de ransomware peuvent être sur-représentés ou sous-représentés. Cela entraîne une diminution des performances du modèle, qui devient moins efficace pour détecter ou classifier les ransomware moins fréquents ou ceux qui n'ont pas été suffisamment représentés dans le dataset. De plus, avec l'évolution constante des ransomware, l'utilisation de données obsolètes peut gravement nuire à la pertinence des modèles. Ce phénomène, appelé « data drift », nécessite des datasets récents et mis à jour pour garantir que les modèles restent performants face aux nouveaux types de ransomware. Par le biais d'attaques dites adversaires, les attaquants peuvent manipuler les données d'entrée pour induire le modèle en erreur. Ils peuvent également tenter de déduire des informations sensibles à partir des sorties du modèle, ou encore modifier le code du ransomware en utilisant des techniques comme l'obfuscation ou le chiffrement pour qu'il soit classé comme bénin.

IV. Discussion

Bien que les méthodes existantes pour l'analyse forensique et la classification des ransomware aient fait des progrès importants, la majorité des approches s'appuient encore fortement sur des techniques d'analyse statique et dynamique, qui, bien qu'efficaces dans certains cas, présentent des limites majeures en environnement réel. Elles ne tiennent pas compte des contraintes réelles du terrain, telles que le manque de données et d'artefacts disponibles et la complexité croissante des nouveaux types de ransomware. Comme référencé dans le Tableau I, les méthodes existantes présentent des limites. D'une part, l'analyse statique dépend de la disponibilité du binaire du malware, ce qui n'est que rarement possible dans les contextes d'investigation post-mortem. De plus, les techniques d'obfuscation ou de chiffrement appliquées aux malwares rendent ces méthodes souvent inefficaces. D'autre part, l'analyse dynamique, bien qu'elle permette d'observer les comportements malveillants en temps réel, est vulnérable aux techniques d'évasion et d'obfuscation de malware, ce qui compromet la fiabilité des résultats obtenus. Face à ces limites, les approches reposant sur l'apprentissage automatique supervisé permettent de meilleurs résultats en termes de précision. Cependant, les algorithmes supervisés rencontrent la même limite concernant la dépendance au binaire du fichier pour identifier le malware. Finalement, les algorithmes non supervisés proposent une méthode plus à même de répondre aux besoins actuels de l'analyse forensique de ransomware.

TABLE I
TABLEAU RÉCAPITULATIF DES MÉTHODES EXISTANTES DE CLASSIFICATION DE RANSOMWARE

Méthode de Classification	**Type d'Analyse**	**Avantages**	**Limitations**
Comportementale	Dynamique	Détecte des comportements malveillants même obfusqués	Dépendance au binaire, contournée par des techniques d'évasion
Basée sur les Signatures	Statique	Rapide et efficace pour des ransomware connus	Dépendance au binaire, inefficace face aux ransomware « zero-day », contournée par des techniques d'obfuscation.
Heuristique	Statique / Dynamique	Peut détecter des variants inconnus en identifiant des motifs généraux	Dépendance au binaire, contournée par des techniques d'évasion et d'obfuscation
Apprentissage Automatique Supervisé	Statique / Dynamique	Capable de classer précisément des ransomware avec des données étiquetées	Dépendance au binaire, inefficace face aux ransomware « zero-day », sujets aux attaques adversaires
Apprentissage Automatique Non Supervisé	Dynamique / Hybride	Indépendance au binaire, permet la détection de ransomware « zero-day ».	Nécessite un grand volume d'artefacts en données d'entrée, sujets aux attaques adversaires

Elles constituent une alternative prometteuse en permettant l'identification de comportements suspects sans avoir besoin de données étiquetées. Cependant, leur efficacité dépend du volume, de la qualité et de la pertinence des données d'entrée.

V. CONCLUSION ET TRAVAUX FUTURS

Cette étude souligne les limites existantes des méthodes traditionnelles de classification et d'analyse des ransomware face à l'évolution rapide des menaces. L'une des pistes les plus prometteuses réside dans l'utilisation de méthodes d'apprentissage automatique non supervisé sur des artefacts tels que les journaux d'évènements, qui sont à la fois persistants, difficilement manipulables par les attaquants, et facilement accessibles par les analystes. L'analyse de ces traces permettrait de détecter des comportements suspects sans nécessiter la présence du binaire, rendant également possible l'identification de ransomware inconnus [17]. Les recherches futures devraient se concentrer sur le développement de méthodes capables de gérer les contraintes du monde réel, telles que le manque de données disponibles, les techniques d'obfuscation et d'évasion des ransomware. Les modèles non supervisés peuvent également être enrichis de manière à faire le lien avec des référentiels comme la base de données MITRE ATT&CK [18], qui documente les techniques et tactiques utilisées par les attaquants lors de cyberattaques. De plus, l'utilisation de modèles de représentation comme les techniques de vectorisation (par exemple, word2vec ou Doc2Vec) permet de coder des évènements d'attaques et d'améliorer les méthodes de regroupement des familles de ransomware. Cette approche identifiera non seulement les familles de ransomware, mais établira également des connexions entre différentes attaques, aidant finalement à l'attribution des cyber-attaques, même face à des techniques d'anonymisation avancées. Sa pertinence est d'autant plus importante que les malwares sans binaires sont de plus en plus nombreux et nécessitent une approche pouvant se détacher de l'analyse de fichier binaire [3], [19].

RÉFÉRENCES

[1] Comparitech, "Ransomware roundup 2024 : End-of-year report," 2024.

[2] H. Oz, A. Arış, A. Levi, and S. Uluagac, "A Survey on Ransomware : Evolution, Taxonomy, and Defense Solutions," *ACM Computing Surveys*, vol. 54, Feb. 2022.

[3] I. Kara, "Fileless malware threats : Recent advances, analysis approach through memory forensics and research challenges," *Expert Systems with Applications*, vol. 214, p. 119133, Mar. 2023.

[4] S. Razaulla, C. Fachkha, C. Markarian, A. Gawanmeh, W. Mansoor, B. C. M. Fung, and C. Assi, "The age of ransomware : A survey on the evolution, taxonomy, and research directions," *IEEE Access*, vol. 11, pp. 40698–40723, 2023.

[5] A.-R. Belea, "Methods for Detecting Malware Using Static, Dynamic and Hybrid Analysis," *International Conference on Cybersecurity and Cybercrime*, vol. 10, pp. 258–265, May 2023.

[6] A. Duby, T. Taylor, G. Bloom, and Y. Zhuang, "Evaluating Feature Robustness for Windows Malware Family Classification," in *2022 International Conference on Computer Communications and Networks (ICCCN)*, (Honolulu, HI, USA), pp. 1–10, IEEE, July 2022.

[7] L. Park, J. Yu, H.-K. Kang, T. Lee, and T. Kwon, "Birds of a Feature : Intrafamily Clustering for Version Identification of Packed Malware," *IEEE Systems Journal*, vol. PP, pp. 1–12, Jan. 2020.

[8] D. S, A. J, I. V, and S. M, "Cyber Forensics : Discovering Traces of Malware on Windows Systems," in *2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, (Thiruvananthapuram, India), pp. 141–146, IEEE, Dec. 2020.

[9] R. Sihwail, K. Omar, and K. Akram Zainol Ariffin, "An Effective Memory Analysis for Malware Detection and Classification," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2301–2320, 2021.

[10] I. Hamid, R. Alajlan, and K. Riad, "ADVANCING MALWARE ARTIFACT DETECTION AND ANALYSIS THROUGH MEMORY FORENSICS : A COMPREHENSIVE LITERATURE REVIEW," *Journal of Theoretical and Applied Information Technology*, Feb. 2024.

[11] VirusTotal, "Virustotal - online virus and malware scanner," 2025.

[12] A. Duby, T. Taylor, and Y. Zhuang, "Malware Family Classification via Residual Prefetch Artifacts," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, (Las Vegas, NV, USA), pp. 256–259, IEEE, Jan. 2022.

[13] T. Taylor, N. Hill, E. Harrington, A. Blackwood, and S. Green, "Dynamic Anomaly-Driven Detection for Ransomware Identification : An Innovative Approach Based on Heuristic Analysis,"

[14] S. Zhang, C. Hu, L. Wang, M. J. Mihaljevic, S. Xu, and T. Lan, "A Malware Detection Approach Based on Deep Learning and Memory Forensics," *Symmetry*, vol. 15, p. 758, Mar. 2023. Number : 3 Publisher : Multidisciplinary Digital Publishing Institute.

[15] S. Dambra, Y. Han, S. Aonzo, P. Kotzias, A. Vitale, J. Caballero, D. Balzarotti, and L. Bilge, "Decoding the Secrets of Machine Learning in Malware Classification : A Deep Dive into Datasets, Feature Extraction, and Model Performance," *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pp. 60–74, Nov. 2023. Conference Name : CCS '23 : ACM SIGSAC Conference on Computer and Communications Security ISBN : 9798400700507 Place : Copenhagen Denmark Publisher : ACM.

[16] J. Ispahany, M. R. Islam, M. Khan, and M. Islam, "Ransomware Detection Using Machine Learning : A Review, Research Limitations and Future Directions," *IEEE Access*, vol. PP, pp. 1–1, Jan. 2024.

[17] H. Studiawan, F. Sohel, and C. Payne, "A survey on forensic investigation of operating system logs," *Digital Investigation*, vol. 29, pp. 1–20, June 2019.

[18] MITRE ATT&CK, "Mitre att&ck framework," 2025.

[19] O. Khalid, S. Ullah, T. Ahmad, S. Saeed, D. A. Alabbad, M. Aslam, A. Buriro, and R. Ahmad, "An Insight into the Machine-Learning-Based Fileless Malware Detection," *Sensors*, vol. 23, p. 612, Jan. 2023. Number : 2 Publisher : Multidisciplinary Digital Publishing Institute.