AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mr **EL RHAYATI Oussama**
Soutiendra : **le Samedi 25/04/2026** à **10H00**
Lieu : ***FSDM – Centre Visioconférence***

Une thèse intitulée :

Development of predictive cybersecurity models based on machine learning

En vue d'obtenir le ***Doctorat***

FD : **Sciences et Technologies de l'Information et de la Communication**
Spécialité : **Informatique**

Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
El BEQQALI Omar	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
TAIME Abderazzak	École Supérieure de Technologie, Khénifra	MCH	Rapporteur
AGHOUTANE Badraddine	Faculté des Sciences, Meknès	PES	Rapporteur
EL FAZAZY Khalid	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Rapporteur
TAIRI Hamid	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Examinateur
AHERRAHROU NOURA	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Examinateur
LAMRINI Mohamed	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Co-directeur de thèse
RIFFI Jamal	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Directeur de thèse

Résumé :

Les logiciels malveillants constituent aujourd'hui une menace majeure dans le domaine de la cybersécurité. Leur évolution rapide, leur capacité de mutation, ainsi que l'usage de techniques avancées d'évasion rendent inefficaces de nombreuses méthodes traditionnelles qui reposent sur des signatures ou des caractéristiques statiques. Ces approches, bien qu'efficaces dans des contextes limités, échouent souvent face à des familles qui évoluent, se fragmentent ou réapparaissent sous de nouvelles variantes. Pour faire face à ces défis, il devient essentiel de développer des modèles capables de comprendre les comportements dans leur contexte, de saisir les relations structurelles dans les rapports d'analyse dynamique, et d'intégrer plusieurs sources d'information complémentaires. Cette thèse s'inscrit dans cette perspective en proposant une approche multimodale profonde pour la détection de malwares et la classification des familles à partir de rapports de type sandbox.

Le premier apport consiste à transformer les journaux bruts issus de Cuckoo Sandbox en descriptions narratives structurées et lisibles par l'humain. Ces représentations permettent aux modèles de type BERT de capturer des relations sémantiques présentes dans les traces comportementales. Ce processus réduit le bruit, met en évidence l'intention malveillante et permet de dégager un signal plus stable pour l'apprentissage. Les résultats expérimentaux montrent que ce type de représentation dépasse les performances des approches classiques d'apprentissage automatique et bénéficie particulièrement de la profondeur contextuelle des modèles transformeurs.

Le deuxième apport présente un pipeline de génération de cartes thermiques structurées appliqué au jeu de données CAPEv2. Les rapports JSON, souvent complexes et irréguliers, sont convertis en représentations spatiales stables qui préservent les informations discriminantes tout en éliminant les risques de fuite depuis les métadonnées ou les identifiants uniques. Ces cartes thermiques peuvent être exploitées par des architectures convolutionnelles telles que CNN64, HybridNet ou ResNeXt. Les expériences menées selon des découpages stratifiés et chronologiques mettent en évidence l'importance de la dérive temporelle, soulignent le rôle critique du choix du vocabulaire et des transformations, et montrent l'intérêt de l'interprétabilité par Grad CAM.

Le troisième apport explore une architecture de fusion multimodale combinant des représentations textuelles, visuelles et graphiques des comportements malveillants. Chaque modalité révèle un aspect différent de l'exécution d'un malware. Leur intégration permet d'obtenir un modèle plus robuste face aux variations comportementales, aux obfuscations partielles et aux évolutions des familles dans le temps. Les expérimentations montrent que l'intégration multimodale améliore la stabilité, la généralisation et la résilience face à des environnements réels où les comportements peuvent changer de manière subtile.

L'ensemble de ces contributions met l'accent sur la reproductibilité, l'évaluation réaliste, la prévention des fuites de données et l'interprétabilité, des aspects indispensables pour une utilisation opérationnelle dans des centres de sécurité. Les résultats obtenus démontrent que l'apprentissage profond fondé sur des représentations structurées et multimodales permet d'améliorer de manière significative la détection et la compréhension des logiciels malveillants. Ils indiquent également que les futures solutions de cybersécurité devront combiner plusieurs sources d'information et intégrer des mécanismes capables de s'adapter à la dérive temporelle et à l'évolution continue des techniques adverses.

Mots clés :

Détection de malwares ; classification de familles de malwares ; CAPEv2 ; analyse dynamique ; cartes thermiques structurées ; BERT ; représentation narrative; réseaux neuronaux convolutionnels ; apprentissage multimodal ; caractéristiques basées sur les graphes ; apprentissage profond ; dérive temporelle ; analyse comportementale

Abstract :

Malware has become one of the most persistent and adaptive threats in modern cybersecurity. The continuous emergence of polymorphic variants, automated mutation engines, and increasingly sophisticated evasion techniques exposes the limits of traditional detection methods that rely on signatures, static indicators, or handcrafted behavioral features. These techniques often fail when confronted with evolving families or subtle changes in execution logic. Effective detection requires models that understand behavior in context, capture structural relations in dynamic reports, and integrate information across multiple complementary views. This thesis develops such an approach through a set of three contributions grounded in deep learning and dynamic behavioral analysis.

The first contribution introduces a method for transforming raw sandbox logs into structured human readable narratives. Cuckoo Sandbox reports are often verbose and inconsistent, which makes direct modeling difficult. By converting the logs into concise descriptions that preserve semantic content, the approach enables the application of large language models such as BERT to malware detection. The resulting narrative representations help reduce noise, highlight behavioral intent, and allow contextual patterns to emerge. The experiments conducted on this representation show that it outperforms classical machine learning baselines and benefits from the richness of transformer based architectures.

The second contribution focuses on the CAPEv2 dataset, which offers a large scale and temporally diverse collection of sandbox reports. A structured heatmap generation pipeline is proposed to transform JSON behavioral trees into stable spatial representations that can be processed by convolutional neural networks. This method carefully avoids information leakage from metadata and sensitive identifiers. It also incorporates controlled vocabulary design, normalization procedures, and resizing strategies that preserve discriminative structure. Using models such as CNN64, HybridNet, and a ResNeXt backbone, the study demonstrates strong performance under both stratified and chronological splits. The evaluation highlights the significant impact of temporal drift and the necessity of realistic protocols. Interpretability is addressed using Grad CAM to visualize which regions of the heatmap contribute to classification decisions.

The third contribution investigates a multimodal fusion architecture that integrates textual, visual, and graph based behavioral representations. Each modality captures a different aspect of malware behavior, and combining them provides a more complete understanding of execution patterns. The proposed architecture explores fusion at the representation level and at the classification stage. The experiments indicate that multimodal systems achieve higher robustness against obfuscation and family evolution, particularly when samples appear in different periods or exhibit partial behavioral overlap. This contribution shows that the integration of multiple structured representations is a promising direction for long term malware analysis.

Across these three contributions, the thesis emphasizes reproducibility, realistic evaluation, avoidance of data leakage, and model interpretability for operational use in security environments. The findings demonstrate that structured representations and multimodal learning can significantly improve detection performance and resilience. The results suggest that future malware detection systems should rely on multiple complementary views of behavior and incorporate mechanisms that address the challenges posed by temporal drift and evolving adversarial techniques.

Key Words :

Malware detection; CAPEv2; deep learning; BERT; CNN; multimodal learning; dynamic analysis