الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 5491 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbes

Mémoire

En vue de l'obtention du diplôme d'**Ingénieur d'État**

Domaine : **Informatique**

Spécialité : **Ingénierie des Systèmes Informatiques**

Thème

Détection de Ransomwares par une Approche Hybride d'Apprentissage Profond Basée sur l'Analyse Comportementale

---

Présenté par
Kebbas Mohammed Houssameddine

Soutenu le : **01 octobre 2025**
Devant le jury composé de

Dr. KHALDI Miloud	Président du jury
Dr. BENDELLA Mohammed Salih	Encadreur
Dr. BENSENANE Hamdane	Co-encadreur
Dr. BABA-AHMED Manel	Examinatrice

Année universitaire : 2024/2025

Résumé

La sophistication croissante des ransomwares rend les défenses traditionnelles obsolètes, créant un besoin urgent de systèmes de détection proactifs basés sur l'intelligence artificielle. Ce mémoire propose une solution de détection précoce en se basant sur l'analyse comportementale de traces d'entrées/sorties (I/O).

Nous introduisons une architecture hybride qui utilise un modèle de Deep Learning (CNN-LSTM avec attention) comme extracteur de attributs temporelles. Ces attributs sémantiques sont ensuite fusionnées avec des statistiques brutes pour alimenter un classifieur final XGBoost. Pour maximiser la performance, les hyperparamètres de ce dernier sont optimisés par un algorithme d'Optimisation par Essaims Particulaires (PSO), guidé par une fonction de fitness personnalisée qui équilibre le rappel (via le F2-score) et la précision.

Évaluée sur deux dataset hétérogènes (l'un déséquilibré, l'autre équilibré), notre solution finale démontre une performance et une capacité de généralisation remarquables. Sur le dataset équilibré, le modèle atteint un **F2-score de 90,50%** et une aire sous la courbe ROC (AUC) de **94,60%**, surpassant significativement les approches monolithiques. Ce travail démontre que l'approche hybride optimisée constitue une solution robuste et efficace, offrant un excellent équilibre entre la minimisation des menaces manquées et la réduction des fausses alertes.

Mots-clés : Ransomware, Détection de malwares, Apprentissage profond, Modèle Hybride, Optimisation par essaims particulaires, Analyse comportementale, Traces d'I/O.

Abstract

The increasing sophistication of ransomware renders traditional defenses obsolete, creating an urgent need for proactive, AI-driven detection systems. This thesis proposes a solution for early ransomware detection based on the behavioral analysis of Input/Output (I/O) traces.

We introduce a hybrid architecture that employs a Deep Learning model (a CNN-LSTM with an attention mechanism) as a sophisticated feature extractor for temporal sequences. These semantic features are then fused with raw statistical data to feed a final XGBoost classifier. To maximize performance, the hyperparameters of this classifier are optimized using a Particle Swarm Optimization (PSO) algorithm, guided by a custom fitness function that balances recall (via F2-score) and precision.

Evaluated on two heterogeneous datasets (one highly imbalanced, one balanced), our final solution demonstrates remarkable performance and generalization capability. On the balanced dataset, the model achieves an **F2-score of 90.50%** and an Area Under the ROC Curve (AUC) of **94.60%**, significantly outperforming monolithic approaches. This work demonstrates that an optimized hybrid architecture provides a robust and effective solution, achieving an excellent balance between minimizing missed threats and reducing false alarms.

Keywords : Ransomware, Malware Detection, Deep Learning, Particle Swarm Optimization, Behavioral Analysis, I/O Traces.

ملخص

إن التطور المتزايد لبرمجيات الفدية يجعل الدفاعات التقليدية غير فعالة، مما يخلق حاجة ملحة لأنظمة كشف استباقية تعتمد على الذكاء الاصطناعي. تقترح هذه الأطروحة حلاً للكشف المبكر عن برمجيات الفدية يعتمد على التحليل السلوكي لتتبعات عمليات الإدخال/الإخراج .(I/O)

نقدم بنية برمجية هجينة تستخدم نموذج تعلم عميق (CNN-LSTM-Attenstion) كمستخرج متطور للسمات الزمنية. يتم بعد ذلك دمج هذه السمات الدلالية مع إحصائيات أولية لتغذية مُصنِّف نهائي من نوع .XGBoost ولتحقيق أقصى أداء، تم تحسين المعايير الفائقة لهذا المُصنّف باستخدام خوارزمية استمثال حشد الجسيمات ،(PSO) والتي تم توجيهها بواسطة دالة لياقة مخصصة توازن بين مقياس الاستدعاء (F2-score) ومقياس الإحكام.

أظهر الحل النهائي، الذي تم تقييمه على مجموعتي بيانات غير متجانسة (إحداهما غير متوازنة والأخرى متوازنة)، أداءً قوياً وقدرة على التعميم. على مجموعة البيانات المتوازنة، حقق النموذج مقياس **F2-score بنسبة 90.50%** ومساحة تحت منحنى ROC (AUC) بنسبة **94.60%**، متفوقاً بشكل كبير على النماذج غير الهجينة. يثبت هذا العمل أن النهج الهجين والمُحسَّن يشكل حلاً قوياً وفعالاً، ويحقق توازناً ممتازاً بين تقليل التهديدات الفائتة والحد من الإنذارات الكاذبة.

الكلمات المفتاحية: برمجيات الفدية، كشف البرامج الخبيثة، التعلم العميق، التحليل السلوكي، تتبعات الإدخال/الإخراج.