الجمهورية الجزائرية الديمقراطية الشعبية
Republic of Algeria Democratic and Popular
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي - 8 ماي 1945 - سيدي بلعباس
Higher School of Computer Science - 08 May 1945 - Sidi Bel Abbès

Thesis

To obtain the diploma of **Engineer**

Field: **Computer Science**

Specialization: **Artificial Intelligence and Data Science (AIDS)**

Theme

Hybrid Graph Neural Network for Anomaly Detection in Complex Systems: Case of Advanced Persistent Threats Attacks

---

Presented by: **LEBGA HANANE**

Submitted on: 08/10/2025

In front of the jury composed of:

President:	Dr.Nassima Dif
Supervisor:	Dr. Sidahmed Benabderrahmane
Co-supervisor:	Pr. Sidi Mohamed Benslimane
Examiner:	Dr.Khaldi Miloud

Academic Year: 2024/2025

الملخص

تُمثل التهديدات المستمرة المتقدمة (APTs) هجمات إلكترونية متطورة ومتعددة المراحل، تتجنب أساليب الكشف التقليدية من خلال التخفي لفترات طويلة والاستغلال المشروع للنظام. تتناول هذه الأطروحة الثغرات الحرجة في الكشف عن التهديدات المستمرة المتقدمة (APT) من خلال اقتراح إطار عمل مبتكر للشبكات العصبية الرسومية الهجينة غير المتجانسة، يدمج التعلم الهيكلي والزمني والدلالي مع آليات شاملة للتفسير.

يُحوّل البحث بيانات الأمن السيبراني الجدولية إلى تمثيلات بيانية غير متجانسة ثنائية الأجزاء، مما يُمكّن من التعلم العلائقي عبر أنواع الاتصال وعقد الميزات. تلتقط بنية هجينة تجمع بين GraphSAGE وشبكات الانتباه الرسومي (GAT) أنماط تجميع الجوار والتبعيات المرجحة بالانتباه، بينما تُنمذج طبقات الذاكرة طويلة المدى قصيرة المدى (BiLSTM) ثنائية الاتجاه التسلسلات الزمنية في التضمينات المُكتسبة. يُعالج الإطار الاختلال الحاد في الفئات، وتناثر البيانات، وعدم التجانس عبر الأنظمة الأساسية المتأصل في سيناريوهات التهديدات المستمرة المتقدمة (APT). أظهرت التجارب المكثفة على مجموعات بيانات الحوسبة الشفافة التابعة لوكالة DARPA تحسينات كبيرة في الأداء مقارنةً بأحدث الطرق.

تتيح استراتيجيات التعلم الانتقالي تعميمًا فعالًا عبر النطاقات، بينما يُقلل اختيار الميزات الذكي القائم على SHAP من التكلفة الحسابية دون المساس بالأداء. يُعالج إطار عمل مزدوج للتفسير، يجمع بين الإسناد السببي القائم على الاضطرابات والتفسير القائم على الانتباه، القيودَ الحرجة التي تُعيق نماذج التعلم العميق. يوفر التكامل مع نماذج اللغات الكبيرة (LLMs) تفسيرات سهلة القراءة مُصممة خصيصًا لتكتيكات MITRE ATT&CK ، مما يُسهّل النشر العملي في مراكز عمليات الأمن. يتم تشغيل النظام بالكامل من خلال تطبيق ويب تفاعلي يدعم تحميل البيانات، وتصور الرسوم البيانية، وتصنيف الشذوذ، والتوليد الآلي لمعلومات التهديدات.

يعمل هذا العمل على تطوير أحدث التقنيات في الكشف عن الشذوذ القائم على الرسم البياني من خلال إظهار أن التمثيلات البيانية غير المتجانسة، عندما يتم دمجها مع النمذجة المتسلسلة وتقنيات الذكاء الاصطناعي القابلة للتفسير، توفر حلولاً قوية وقابلة للتطوير وشفافة للكشف عن التهديدات السيبرانية المعقدة في بيئات الحوسبة الحديثة.

الكلمات المفتاحية: التهديدات المستمرة المتقدمة، الشبكات العصبية الرسومية، الذكاء الاصطناعي القابل

للتفسير، اكتشاف الشذوذ، الأمن السيبراني، التعلم العميق، نماذج التسلسل، نماذج اللغة الكبيرة، المحولات.

Abstract

Advanced Persistent Threats (APTs) represent sophisticated and multi-stage cyberattacks that evade traditional detection by maintaining stealth over long periods and leveraging legitimate system functionalities. This thesis addresses critical gaps in APT detection by proposing an innovative Heterogeneous Hybrid Graph Neural Network (HGNN) framework that integrates structural, temporal, and semantic learning with comprehensive explainability mechanisms.

The research transforms tabular cybersecurity data into heterogeneous bipartite graph representations, enabling relational learning across connection types and feature nodes. A hybrid architecture combining GraphSAGE and Graph Attention Networks (GAT) captures neighborhood aggregation patterns and attention-weighted dependencies, while Bidirectional Long Short-Term Memory (BiLSTM) layers model temporal sequences within the learned embeddings. The framework tackles severe class imbalance, data sparsity, and cross-platform heterogeneity inherent to APT scenarios. Extensive experiments on DARPA Transparent Computing (TC) E2 datasets across Windows, Android, Linux, and BSD systems, as well as KDD CUP99, demonstrate significant improvements over state-of-the-art methods.

Transfer learning strategies ensure cross-domain generalization, while SHAP-based feature selection reduces computational cost without sacrificing performance. A dual explainability framework combining perturbation-based causal attribution and attention-based interpretation addresses critical limitations of deep learning models. Integration with Large Language Models (LLMs) generates human-readable explanations aligned with MITRE ATT&CK tactics, facilitating deployment in Security Operations Centers. The entire system is deployed as an interactive web application supporting data upload, graph visualization, anomaly classification, and automated threat intelligence generation.

This research advances graph-based anomaly detection by demonstrating that heterogeneous graph representations, when combined with sequential modeling and explainable AI techniques, provide scalable, robust, and transparent solutions for detecting complex cyber threats in modern computing environments.

Keywords: Advanced Persistent Threats, Graph Neural Networks, Explainable AI, Anomaly Detection, Cybersecurity, Deep Learning, Sequence Models, Large Language Models, Transformers.

Résumé

Les menaces persistantes avancées (APT) constituent des cyberattaques sophistiquées et multi-étapes, capables déchapper aux méthodes de détection traditionnelles grâce à leur furtivité prolongée et à lexploitation légitime du système. Cette thèse traite des lacunes critiques de la détection des APT en proposant un cadre innovant basé sur des Réseaux de Neurones Graphiques Hybrides Hétérogènes (HGNN), intégrant lapprentissage structurel, temporel et sémantique avec des mécanismes complets dexplicabilité.

La recherche transforme les données tabulaires de cybersécurité en représentations graphiques hétérogènes bipartites, permettant lapprentissage relationnel à travers les types de connexions et les nuds de caractéristiques. Une architecture hybride combinant GraphSAGE et Graph Attention Networks (GAT) capture les schémas dagrégation de voisinage et les dépendances pondérées par lattention, tandis que des couches BiLSTM modélisent les séquences temporelles dans les représentations apprises. Le cadre proposé sattaque au déséquilibre sévère des classes, à la rareté des données et à lhétérogénéité multi-plateformes inhérents aux scénarios APT. Des expérimentations approfondies sur les ensembles de données DARPA Transparent Computing (TC) E2 couvrant Windows, Android, Linux et BSD, ainsi que sur KDD CUP99, montrent des améliorations significatives par rapport aux méthodes les plus avancées.

Les stratégies dapprentissage transféré permettent une généralisation inter-domaines efficace, tandis que la sélection de caractéristiques basée sur SHAP réduit les coûts de calcul sans compromettre les performances. Un cadre dexplicabilité double combinant lattribution causale par perturbation et linterprétation basée sur lattention pallie les limites critiques des modèles de deep learning. Lintégration avec les grands modèles de langage (LLMs) génère des explications lisibles, alignées sur les tactiques MITRE ATT&CK, facilitant leur adoption dans les centres opérationnels de sécurité. Le système est entièrement déployé sous forme dune application web interactive prenant en charge le chargement des données, la visualisation graphique, la classification des anomalies et la génération automatisée dintelligence sur les menaces.

Cette recherche fait progresser létat de lart en détection danomalies basée sur les graphes en démontrant que les représentations graphiques hétérogènes, combinées avec la modélisation séquentielle et lIA explicable, offrent des solutions robustes, évolutives et transparentes pour la cybersécurité moderne.

Mots-clés : Menaces persistantes avancées (APT), Réseaux de neurones graphiques (GNN), Intelligence artificielle explicable (XAI), Détection danomalies, Cybersécurité, Apprentissage profond, Modèles séquentiels, Modèles de langage de grande taille (LLM), Transformeurs.