الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٠08 ماي 1945٠ بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

Thesis

Towards the attainment of **Master's degree**
Field: **Computer Science**
Specialization: **Computer System Engineering (ISI)**

Theme

Machine Learning-based Intrusion Detection using network data analysis frameworks: A Comparative Study.

Presented by:

SLIMANE Chaima

Defended on : **09/07/2025** In front of the jury composed of:

Dr. KHALDI Miloud	President
Pr. BENSLIMANE Sidi Mohammed	Supervisor
Dr. LAHLOU Laaziz	Co-Supervisor
Pr. KARA Nadjia	Co-Supervisor
Dr. BEDJAOUI Mohamed	Examiner

University Year: 2024/2025

ABSTRACT

Traditional intrusion detection systems cannot handle modern cyber attacks. Machine learning-based network intrusion detection systems (ML-NIDS) offer better solutions, but they depend on the quality of the data used for training and on the tools used to extract useful information from network traffic. In particular, converting pcap files into CSV files that machine learning models can understand requires choosing the right network analysis framework.

This master's thesis studies the state of the art in ML-based intrusion detection and compares how different data analysis frameworks affect model performance. It also examines how certain datasets are generated and how specific data properties can impact the results. The work involves testing different combinations of analysis tools, machine learning algorithms, and both public and private datasets to find the most effective setups.

The results show that the choice of framework has a significant effect on detection accuracy. This study provides useful guidelines for selecting the best combinations of tools and methods for building reliable ML-based intrusion detection systems.

Keywords: Network Analysis Frameworks, Machine Learning, Machine-Learning-Based Network Intrusion Detection Systems, PCAP Analysis, Feature Extraction, Network Security, Anomaly Detection, Cybersecurity.

RÉSUMÉ

Les systèmes traditionnels de détection d'intrusion ne suffisent plus face aux cyber-attaques modernes. Les systèmes basés sur l'apprentissage automatique (ML-NIDS) offrent des solutions plus efficaces, mais leur performance dépend de la qualité des données d'entraînement et des outils utilisés pour analyser le trafic réseau. En particulier, la conversion des fichiers pcap en CSV nécessite un bon choix d'outil d'analyse.

Ce mémoire de master étudie l'état de l'art des ML-NIDS et compare l'impact de différents outils d'analyse sur les performances des modèles. Il explore également comment les jeux de données sont générés et comment certaines propriétés influencent les résultats. L'étude teste plusieurs combinaisons d'outils, d'algorithmes et de jeux de données publics et privés pour identifier les meilleures configurations.

Les résultats montrent que le choix de l'outil d'extraction influence fortement la précision de détection. Ce travail propose des recommandations pratiques pour construire des systèmes fiables de détection d'intrusion basés sur le machine learning.

Mots-clés : Frameworks d'analyse réseau, Apprentissage automatique, Systèmes de détection d'intrusion réseau basés sur l'apprentissage automatique, Analyse PCAP, Extraction des attributs, Sécurité réseau, Détection d'anomalies, Cybersécurité.

ملخص

أنظمة كشف التسلل التقليدية لم تعد قادرة على التعامل مع الهجمات السيبرانية الحديثة. وقد أصبحت أنظمة كشف التسلل المعتمدة على تقنيات التعلم الآلي توفر حلولًا أكثر فعالية، إلا أن أداءها يعتمد بشكل كبير على جودة البيانات المستخدمة في التدريب، وعلى الأدوات التي تُستخدم لاستخراج المعلومات المفيدة من حركة مرور الشبكة. وبشكل خاص، فإن تحويل ملفات PCAP إلى ملفات CSV قابلة للمعالجة من قبل نماذج التعلم الآلي يتطلب اختياراً دقيقًا لإطار تحليل الشبكة المناسب.

تتناول مذكرة الماجستير هذه دراسة شاملة لأحدث ما توصلت إليه الأبحاث في مجال كشف التسلل باستخدام التعلم الآلي، وتقارن بين تأثير أطر تحليل البيانات المختلفة على أداء النماذج. كما تدرس كيفية توليد بعض مجموعات البيانات، وتبحث في خصائص معينة قد تؤثر على النتائج. ويشمل العمل تجريب مجموعات مختلفة من أدوات التحليل وخوارزميات التعلم الآلي، بالإضافة إلى بيانات عامة وخاصة، من أجل تحديد أنسب التكوينات.

تُظهر النتائج أن اختيار إطار العمل له تأثير كبير على دقة الكشف ومدة المعالجة. وتقدّم هذه الدراسة توجيهات عملية لاختيار أفضل تركيبات الأدوات والأساليب لبناء أنظمة فعالة وموثوقة للكشف عن التسللات باستخدام التعلم الآلي.

الكلمات المفتاحية: أطر تحليل الشبكة، التعلم الآلي،أنظمة كشف التسلل الشبكي المعتمدة على التعلم الآلي، تحليل ملفات PCAP، استخراج الخصائص، أمن الشبكات، كشف الشذوذ، الأمن السيبراني.