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 **Engineer degree**
Field: **Computer Science**
Specialization: **Computer System Engineering (ISI)**

Theme

Performance evaluation of Machine Learning-based Intrusion Detection using network data analysis frameworks.

Presented by:

SLIMANE Chaima

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

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

University Year: 2024/2025

ABSTRACT

Tools for feature extracting and classifying packets/flows present in packet capture files, also known as PCAP, utilized in machine-learning-based (ML) intrusion detection systems, are essential processes for validating intrusion detection models and algorithms. These ML-based techniques rely on such tools, and their effectiveness in detecting cyberattacks is solely tied to them, in addition to the quality of the training datasets. While the latter is of utmost importance, this study focuses on the most recently used tools in the literature that the research community uses from a microscopic view and unveils any shortcomings and technical glitches that can mislead the scientific findings. NTLFlowlyzer and NFStream are two of them. FlowMeter, a tool developed and released by deepfence, one of the fastest-growing cloud-native application protection platforms, is also considered in this investigation. This work contributes a detailed comparative evaluation of these three tools across public and private datasets, and investigates their impact on the performance of various ML models. By uncovering tool-specific limitations and practical issues, our study provides actionable insights to guide researchers and practitioners in selecting the most suitable analysis framework for robust intrusion detection.

Keywords: Network Analysis Frameworks, NFStream, NTLFlowLyzer, Flowmeter, Machine Learning, Machine-Learning-Based Network Intrusion Detection Systems, Feature Extraction, XGBoost, Random Forest, Decision Trees, CIC-IDS2018, UNSW-

NB15, Network Security, Anomaly Detection, Cybersecurity, PCAP Analysis, DDoS Attacks.

RÉSUMÉ

Les outils d'extraction des attributs et de classification des paquets ou flux issus de fichiers PCAP sont essentiels dans les systèmes de détection d'intrusion basés sur l'apprentissage automatique. La performance de ces systèmes dépend à la fois de ces outils et de la qualité des jeux de données utilisés pour l'entraînement.

Cette étude se focalise sur trois outils récents largement utilisés dans la littérature : NTLFlowlyzer, NFStream, et FlowMeter (développé par Deepfence). Elle apporte une évaluation comparative approfondie de ces trois outils sur des jeux de données publics et privés, et examine leur impact sur les performances de différents modèles d'apprentissage automatique. En révélant les limitations spécifiques à chaque outil ainsi que certains problèmes pratiques, notre étude fournit des recommandations utiles pour orienter les chercheurs et les professionnels vers le choix du cadre d'analyse le plus adapté pour une détection efficace des intrusions.

Mots-clés : Outils d'analyse réseau, NFStream, NTLFlowLyzer, Flowmeter, Apprentissage automatique, Systèmes de détection d'intrusion réseau basés sur l'apprentissage automatique, Extraction des attributs, XGBoost, Random Forest, Arbres de décision, CIC-IDS2018, UNSW-NB15, Sécurité réseau, Détection d'anomalies, Cybersécurité, Fichiers PCAP, Attaques DDoS.

ملخص

تُعد أدوات استخراج الخصائص وتصنيف الحزم أو التدفقات الموجودة في ملفات التقاط الحزم من العمليات الأساسية في أنظمة كشف التسلل المبنية على Machine Learning. تعتمد هذه التقنيات بشكل كبير على فعالية هذه الأدوات، بالإضافة إلى جودة مجموعات البيانات المستخدمة في التدريب، لضمان دقة اكتشاف الهجمات السيبرانية.

وعلى الرغم من أن جودة البيانات تُعتبر ذات أهمية قصوى، فإن هذه الدراسة تركز على الأدوات الحديثة الأكثر استخداماً في الأدبيات العلمية، حيث يتم تحليل هذه الأدوات من منظور دقيق للكشف عن الثغرات والمشاكل التقنية التي قد تؤثر على صحة النتائج العلمية.

يُعتبر كل من NTLFlowlyzer وNFStream من بين هذه الأدوات. كما تم تضمين FlowMeter، وهي أداة طورتها ونشرتها شركة Deepfence، إحدى أسرع المنصات نمواً في مجال حماية التطبيقات السحابية الأصلية.

تقدم هذه الدراسة تقييماً مقارناً مفصّلاً لهذه الأدوات الثلاثة على مجموعات بيانات عامة وخاصة، وتبحث في تأثيرها على أداء نماذج التعلم الآلي المختلفة. ومن خلال كشف القيود والمشكلات التقنية الخاصة بكل أداة، توفّر دراستنا رؤى عملية تساعد الباحثين والمتخصصين في اختيار الإطار الأنسب لتحليل الشبكات والكشف الفعّال عن الهجمات.

الكلمات المفتاحية: أدوات تحليل الشبكات، NFStream، NTLFlowlyzer، FlowMeter، التعلم الآلي، أنظمة كشف التسلل الشبكي المعتمدة على التعلم الآلي، استخراج الخصائص، XGBoost، Random Forest، Decision Trees، CIC-IDS2018، UNSW-NB15، أمن الشبكات، كشف الشذوذ، الأمن السيبراني، تحليل ملفات PCAP، هجمات DDoS.