AVIS DE SOUTENANCE D'UNE THESE DE DOCTORAT

Le Directeur de l'Ecole Nationale des Sciences Appliquées a le plaisir d'informer le public qu'une soutenance de thèse de Doctorat en

«Sciences et ingénierie»

aura lieu le 28/11/2026 à 10H30 à l'ENSA, Kénitra

La Thèse sera présentée par Mme CHENTOUFI OUMAIMA

Sous le thème :

Robust Feature Engineering and Model Optimization for Scalable Intrusion Detection Systems

Devant le jury composé de :

Nom et Prénom	Titre	Etablissement
ARIOUA MOUNIR	Président	ENSA, Kénitra
EL MENDILI FATNA	Rapporteur	EST, Meknès
BADI IMAD	Rapporteur	FP, Taza
EL BOUZEKRI IDRISSI YOUNESS	Rapporteur	ENSA, Kénitra
IBRAHIMI KHALIL	Examinateur	Faculté des Sciences, Kénitra
GAHI YOUSSEF	Examinateur	ENSA, Kénitra
CHOUGDALI KHALIL	Directeur de thèse	ENSA, Kénitra

UNIVERSITE IBN TOFAIL ★ PÔLE DES ETUDES DOCTORALES ★ KENITRA

Nom et Prénom : CHENTOUFI OUMAIMA
Date de soutenance : 28/11/2026
Directeur de Thèse : CHOUGDALI KHALID

Sujet de thèse :

Robust Feature Engineering and Model Optimization for Scalable Intrusion Detection Systems

Abstract:

La transiΘon des réseaux informaΘques, de l'infrastructure restreinte aux services mondiaux intensifs en données d'aujourd'hui, a fondamentalement transformé le paysage de la cybersécurité. Alors que les défenses périmétriques tradiΘonnelles et les mécanismes de détecΘon par signatures échouent à intercepter les exploits sophisΘqués de type "zero-day« et les menaces capables de changer de forme pour échapper à la détecΘon, la nécessité de Systèmes de DétecΘon d'Intrusions intelligents, basés sur les anomalies, est devenue primordiale. Cependant, les cadres contemporains pilotés par l'intelligence arΘficielle sont fréquemment entravés par des taux élevés de faux posiΘfs, des limitaΘons de scalabilité computaΘonnelle et une inflaΘon des performances induite par les jeux de données.
CeΣe thèse répond à ces défis par une approche systémaΘque, orientée analyse de données et à la réducΘon de dimensionnalité. Une contribuΘon majeure de ce travail réside dans l'analyse rigoureuse des jeux de données de référence, parΘculièrement l'idenΘficaΘon et l'évaluaΘon criΘque du biais lié à la redondance dans KDDCup'99. CeΣe analyse moΘve ainsi l'uΘlisaΘon de NSL-KDD et UNSW-NB15 comme références d'évaluaΘon plus fiables. En privilégiant une compréhension approfondie des caractérisΘques intrinsèques des données, ceΣe recherche garanΘt que les métriques de performance rapportées reflètent de réels progrès dans la capacité de détecΘon plutôt que des artefacts staΘsΘques. Au cœur des contribuΘons techniques de ceΣe thèse se trouve l'implémentaΘon de techniques robustes d'extracΘon de caractérisΘques basées sur la double norme l2,p ACP qui renforcent la séparabilité des classes dans le trafic réseau de haute dimension tout en préservant l'efficacité computaΘonnelle. Pour combler l'écart entre la modélisaΘon théorique et les exigences opéraΘonnelles en temps réel, ce travail introduit une intégraΘon raisonnée de l'Analyse en Composantes Principales (PCA) en mode Batch et Incremental Les résultats expérimentaux démontrent que le cadre hybride proposé idenΘfie avec succès le point de rendement, aΣeignant un équilibre efficace entre le score F1, l'empreinte mémoire et la latence de traitement. En résolvant explicitement le compromis entre la précision de la détecΘon et la charge de calcul, ceΣe thèse fournit une base extensible pour les systèmes de détecΘon d'intrusions intelligents de nouvelle généraΘon et ouvre la voie à des architectures de sécurité adaptaΘves et en ligne.
Mots Clés : Cybersécurité, DétecΘon d'Intrusions, Ingénierie des CaractérisΘques, RéducΘon de Dimensionnalité, Intelligence ArΘficielle, ApprenΘssage AutomaΘque.

Résumé :

The transiΘon of computer networks from the restricted infrastructure to today's global, data-intensive uΘliΘes has fundamentally altered the cybersecurity landscape. As tradiΘonal perimeter-based defenses and signature-based detecΘon mechanisms fail to intercept sophisΘcated zero-day exploits and unseen vulnerabiliΘes, the need for intelligence-driven, anomaly-based Intrusion DetecΘon Systems has become paramount. However, contemporary AI-driven frameworks are frequently hindered by high false-posiΘve rates, computaΘonal scalability limitaΘons, and dataset-induced performance inflaΘon.
This thesis addresses these challenges through a systemaΘc, engineering-oriented approach to feature engineering and dimensionality reducΘon. A primary contribuΘon of this work is the rigorous analysis of benchmark intrusion detecΘon datasets, parΘcularly the idenΘficaΘon and criΘcal assessment of redundancy-driven bias in KDDCup'99, thereby moΘvaΘng the use of NSL-KDD and UNSW-NB15 as more reliable evaluaΘon benchmarks within the scope of this study. By prioriΘzing a deep understanding of intrinsic dataset characterisΘcs, this research ensures that reported performance metrics reflect genuine advances in detecΘon capability rather than staΘsΘcal arΘfacts.
Central to the technical contribuΘons of this thesis is the implementaΘon of robust double L₰,□-norm-based feature extracΘon techniques, which enhance class separability in high-dimensional network traffic while maintaining computaΘonal efficiency. To bridge the gap between theoreΘcal modeling and real-Θme operaΘonal requirements, the work introduces a principled integraΘon of Batch and Incremental Principal Component Analysis.
While Batch PCA serves as an offline reference under full data availability, Incremental PCA is validated as a scalable and memory-efficient approximaΘon that converges toward comparable representaΘons in online and resource-constrained scenarios.
Experimental results demonstrate that the proposed hybrid framework successfully idenΘfies the point of diminishing returns in feature engineering, achieving an effecΘve balance between F1-score, memory footprint, and processing latency. By explicitly resolving the trade-off between detecΘon accuracy and computaΘonal overhead, this thesis provides an extensible foundaΘon for next-generaΘon intelligent intrusion detecΘon systems and paves the way toward adapΘve and online security architectures.
Keywords: Cybersecurity, Intrusion DetecΘon, Feature Engineering, Dimensionality ReducΘon, ArΘficial Intelligence, Machine Learning, Pre-processing.