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

Mémoire de Fin d'étude

En Vue de l'obtention du diplôme de **Master**
Filière : **Informatique**
Spécialité : **Système d'Information et Web (SIW)**

Thème

A Comparative Study of Machine Learning Based Intrusion Detection Methods

---

Présenté par :

- M. LAHMAR Mohammed Abdrrahim
- Mlle. DJELLOUL DAOUADJI Fadela

Soutenu le : **03/07/2023**

Devant le jury composé de :

Mme. ANANI Djihed	Président
M. KHALDI Miloud	Encadreur
Mlle. BABA-AHMED Manel	Examinateur

Année Universitaire : 2022/2023

Acknowledgments

We first want to express our sincere thanks to Allah for helping us and giving us the patience and motivation to complete this work.

We would particularly like to thank our supervisor, **Dr. Miloud KHALDI**, for the help he has given us, for his patience and encouragement. His critical eyes were very valuable in structuring the work and improving the quality of the different sections.

We would also like to thank our promoter **Pr. Sidi Mohammed BENSLIMANE** for his immense help, the quality of his follow-up as well as for all the advice and information he gave us with a degree of patience and professionalism.

Big thanks to our **parents** who have taught, worked on and improved, loved, cared, sacrificed and shared everything with us, may God bless them.

We would like to express our sincere thanks to the members of the jury for the honour they have done us by taking the time to read and evaluate this work. We would also like to thank the pedagogical and administrative team of the ESI 8-mai-1945 for their efforts to offer us an excellent training.

Big thanks to all my **teachers** who have taught, worked on and improved our skills to reach such a respectful level, may God bless you all and your families with health and success.

Abstract

Despite the numerous benefits offered by the Internet, the security of its components and data transfers is a major concern. With the increasing number of interconnected devices, it becomes easier for intruders to breach the system and access sensitive data. This presents a significant challenge for developers and organizations.

To address this issue, a many of **machine learning based intrusion detection systems** were proposed. These systems are designed to detect any suspicious activity or false requests submitted by intruders with the intention of disrupting the operation of the system. It is crucial to have a reliable security system that can detect any intrusion during all phases of the execution process.

In summary, a background of anomaly detection, intrusion detection systems and artificial intelligence will be presented. Afterwards, we will present many **machine learning based intrusion detection systems**. Finally, we will make a comparison between them.

Keywords : Artificial Intelligence (AI), Machine Learning (ML), Intrusion Detection System (IDS), Anomaly Detection, Features Selection.

Résumé

Malgré les nombreux avantages offerts par l'internet, la sécurité de ses composants et des transferts de données est une préoccupation majeure. Avec le nombre croissant d'appareils interconnectés, il devient plus facile pour les intrus de violer le système et d'accéder à des données sensibles. Cela représente un défi important pour les développeurs et les organisations.

Pour résoudre ce problème, de nombreux **systèmes de détection d'intrusion basés sur l'apprentissage automatique** ont été proposés. Ces systèmes sont conçus pour détecter toute activité suspecte ou toute fausse demande soumise par des intrus dans le but de perturber le fonctionnement du système. Il est essentiel de disposer d'un système de sécurité fiable capable de détecter toute intrusion durant toutes les phases du processus d'exécution.

En résumé, nous présenterons le contexte de la détection d'anomalies, des systèmes de détection d'intrusion et de l'intelligence artificielle. Ensuite, nous présenterons de nombreux **systèmes de détection d'intrusion basés sur l'apprentissage automatique**. Enfin, on va les comparer.

Mots clés : Intelligence artificielle, Apprentissage automatique, Système de détection d'intrusion, Détection d'anomalie, Sélection des attributs.

مُلخص

على الرغم من المزايا العديدة التي تقدمها الإنترنت ، فإن أمان هذه المكونات وعمليات نقل البيانات يمثل مصدر قلق كبير. مع تزايد عدد الأجهزة المترابطة ، يصبح من السهل على المتسللين اختراق النظام والوصول إلى البيانات الحساسة. يمثل هذا تحديًا كبيرًا للمطورين والمؤسسات.

لمعالجة هذه المشكلة ، تم اقتراح العديد من **أنظمة كشف التسلل القائمة على التعلم الآلي**. تم تصميم هذه الأنظمة لاكتشاف أي نشاط مشبوه أو طلبات كاذبة مقدمة من قبل المتسللين بقصد تعطيل تشغيل النظام. من المهم أن يكون لديك نظام أمان موثوق يمكنه اكتشاف أي اختراق خلال جميع مراحل عملية التنفيذ.

باختصار ، سيتم تقديم خلفية عن كشف التسلل وأنظمة كشف التسلل والذكاء الاصطناعي. بعد ذلك ، سوف نعرض العديد من **أنظمة كشف التسلل القائمة على التعلم الآلي** أخيرًا ، سنجري مقارنة بينهم.

الكلمات المفتاحية : الذكاء الاصطناعي ، التعلم الآلي ، نظام كشف الاختراق ، اكتشاف الأخطاء ، اختيار الميزات.

List of acronyms

1. **IDS :** Intrusion Detection System.
2. **ML :** Machine Learning.
3. **IPS :** Intrusion Prevention System.
4. **AI :** Artificial Intelligence.
5. **DL :** Deep Learning.
6. **LSTM :** Long Short-Term Memory.
7. **ADS :** Anomaly Detection Systems.
8. **NIDS :** Network Intrusion Detection System.
9. **HIDS :** Host Intrusion Detection System.
10. **SIDS :** Signature-based Intrusion Detection System.
11. **KDD :** Knowledge Discovery and Data mining.
12. **KNN :** K-Nearest Neighbors.
13. **XGB :** Extreme gradient boosting.
14. **SVM :** Support Vector Machines.
15. **RF :** Random Forest.
16. **DT :** Decision Tree.
17. **AB :** AdaBoost.
18. **GBM :** Gradient Boosted Machine.

19. **AUC :** Ope Discriminant Analysis.
20. **ETC :** Extremely Randomized Trees.
21. **CART :** Classification And Regression Trees.
22. **DQN :** Deep Q Network.
23. **SARSA :** State-Action-Reward-State-Action.
24. **CNNs :** Convolutional Neural Networks.
25. **RNNs :** Recurrent Neural Networks.
26. **GANs :** Generative Adversarial Networks .
27. **MLPs :** Multi-layer Perceptrons.
28. **CFS :** Correlation-based Feature Selection.
29. **IG :** Information Gain.
30. **GR :** Gain Ratio.
31. **PCA :** Principal Component Analysis.
32. **GA :** Genetic Algorithm.
33. **DoS :** Denial of Service.
34. **U2R :** User to Root.
35. **R2L :** Remote to Local.
36. **PSO :** Particle Swarm Optimization.
37. **ABC :** Artificial Bee Colony.
38. **SAE :** Stacked Autoencoder.
39. **SU :** Symmetrical Uncertainty.
40. **MCC :** Matthews Correlation Coefficient.
41. **DR :** Detection Rate.
42. **FAR :** False Alarm Rate.