



RAPPORT DU PROJET DE FIN D'ÉTUDES

Pour obtenir le diplôme de

Post-Graduation Spécialisé en sécurité informatique

Préparé par :

Sakhri Hichem , Derrar Omar

Sous la direction de :

Maître assistante Bensefia Hassina (Encadreur)

Un système de Détection d'Intrusion hybride auto-apprenant à base d'ensemble de classifieurs

Soutenue le : 20/02/2022

Devant le jury composé de :

M. NOUALI Omar : Président

Mme. GUEMRAOUI Lila : Examinatrice

&

Melle. ZEMMAHE Amina : Examinatrice

Année Universitaire : **2019 - 2020**

Un système de Détection d'Intrusion hybride auto-apprenant à base d'ensemble de Classifieurs



Sakhri Hichem, Derrar Omar

Mémoire soumis en vue de l'obtention du diplôme de
Post-Graduation Spécialisé

REMERCIEMENTS

Nous remercions tous d'abord Allah tous puissant qui nous a donné la santé, le courage et la patience afin de pouvoir accomplir ce modeste travail.

Nous tenons dans un premier temps à exprimer notre profonde reconnaissance à notre encadreur Melle Hassina BENSEFI Maître assistante à la Faculté des mathématiques et de l'informatique (UBBA), pour avoir accepté de diriger ce travail et qui nous a accompagnés tout au long de cette période de réalisation. Nous ne serons assez la remercier pour sa disponibilité, son soutien indéfectible, son enthousiasme scientifique et son encouragement qui nous a permis de mener à bien notre travail, surtout dans les moments les plus difficiles.

Nous exprimons toute notre gratitude à tous les membres du jury pour l'honneur qu'ils nous font en participant à l'examen de ce travail, particulièrement à Monsieur Omar NOUALI, le président de notre jury et le responsable Scientifique de la PGS en sécurité informatique du CERIST.

Nous tenons à exprimer nos remerciements à tous nos enseignants PGS et au personnel de la formation du CERIST.

Nos sincères remerciements et nos vives gratitudes vont aux membres de nos familles et en particulier nos parents pour leurs sacrifices envers notre éducation et études.

Enfin, nous remercions toutes les personnes qui ont contribué de prêt ou de loin à la concrétisation de ce modeste travail ainsi qu'à toute personne qui fera l'effort de lire ce document.

شكر وتقدير

أول مشكور هو الله عز وجل، ثم والداي على كل مجهوداتهم منذ ولادتي إلى هذه اللحظات وعائلتي الصغيرة. يسرني أن أوجه شكري لكل من نصحتني أو أرشدني أو وجهني أو ساهم معي في إعداد هذا البحث بإيصالي للمراجع والمصادر المطلوبة في أي مرحلة من مراحلها، وأشكر على وجه الخصوص للمشرفة الأستاذة المساعدة بجامعة محمد البشير الابراهيمي برج بو عريريج / حسينة بن صافية على كل المجهودات التي بذلتها معنا من خلال النصح و التصحيح وعلى اختيار العنوان والموضوع، كما أن شكري موجه لإدارة مركز البحث في الاعلام الالي و التقني و الأستاذانوالي عمر على المجهودات المبذولة من طرفه ولأساتذتنا الكرام في المركز لتوفير أفضل بيئة للتدريس في أفضل الأحوال التي تلائم طلبة العلم.

صخري هشام

في البداية، الحمد والشكر لله، جل في علاه، فأليه ينسب الفضل كله في الإكمال -والكمال يبقى لله وحده -هذا العمل. وبعد الحمد لله، أتوجه بالشكر إلى والداي على وباية بوشملة على كل مجهوداتهم، دعمهم ونصحهم لي دائما، وإلى عائلتي الكبيرة والصغيرة وعلى رأسهم زوجتي على صبرها ومساندتها لي طوال فترة التربص فجزاهم الله عني خير الجزاء. كما أتوجه بالشكر والتقدير للمشرفة الأستاذة المساعدة بجامعة محمد البشير الابراهيمي برج بو عريريج / حسينة بن صافية التي كانت بعد إله عزوجل المعين الأول لنا على إتمام وإنجاز هذه المذكرة فلها كل التقدير والإمتنان. كما أن شكري موجه لإدارة مركز البحث في الاعلام الالي و التقني، والأستاذ نوالي عمر على المجهودات المبذولة من طرفه ولأساتذتنا الكرام في المركز لتوفير أفضل بيئة للتدريس في أفضل الأحوال التي تلائم طلبة العلم.

ورار عمر

ملخص

الهدف من هذا المشروع هو تصميم وتطوير نظام كشف التسلل (IDS) على أساس التسلسل الهرمي للمصنفات والذي يجمع بين طريقتين لكشف التسلل: الطريقة الشاذة والطريقة القائمة على التوقيع ومن ثم التأهيل الهجين. يتمتع نظام IDS هذا بالقدرة على اكتشاف أشكال جديدة من الهجمات وأشكال جديدة من السلوك الطبيعي أثناء تشغيله (في الوقت الفعلي) ولديها القدرة على التعلم الإلكتروني التلقائي بالتزامن مع تشغيل نظام كشف التسلل.

هذا الأخير يستغل أداء مجموعات المصنفات وكذلك عدم تجانس المصنفات التي تشكلها، والتي هي شبكات عصبية من نوع شبكات المطرقة التكميلية (AHN). تم تنفيذ نظام IDS الخاص بنا باستخدام Matlab وتبعه التحقق من الأداء عبر مجموعة البيانات القياسية في كشف التسلل. NSL KDD 2012 النتائج مرضية، فهي تثبت فعالية النهج المقترح.

الكلمات المفتاحية: نظام كشف التسلل، (IDS) شبكات المطرقة التكميلية

، (AHN) مجموعات المصنفات، مسافة «Minkowski». NSL-KDD.

Abstract

The objective of this project is to design and develop an intrusion detection system (IDS) based on a hierarchy of classifiers and which combines the two intrusion detection methods: the anomaly and the signature-based method hence the hybrid qualification.

This IDS has the power to detect new forms of attacks and new forms of normal behavior during its operation (in real time). It has automatic e-learning capability in conjunction with IDS operation. It exploits the performance of the sets of classifiers as well as the heterogeneity of the classifiers which constitute it, which are neural networks of the Adaptive Hamming Nets (AHN) type.

The implementation of our IDS was carried out with Matlab and it is followed by a validation of the performances via the standard data set in intrusion detection NSL KDD 2012. The results are satisfactory, they prove the effectiveness of the proposed approach.

Keywords: Intrusion Detection System (IDS), Adaptive Hamming Nets (AHN), Classifier Ensembles, Minkowski Distance, NSL-KDD.

Résumé

Ce projet a pour objectif la conception et le développement d'un système de détection d'intrusion (Intrusion Detection System, IDS) basé sur une hiérarchie d'ensemble de classifieurs et qui combine les deux méthodes de détection d'intrusion : la méthode basée anomalie et la méthode basée signature d'où la qualification hybride.

Cet IDS a un pouvoir de détection des nouvelles formes d'attaques et des nouvelles formes de comportement normal durant son fonctionnement (en temps réel). Il est doté d'une capacité d'apprentissage en-ligne automatique conjointement au fonctionnement de l'IDS.

Il exploite la performance des ensembles de classifieurs ainsi que l'hétérogénéité des classifieurs qui le constituent qui sont des réseaux de neurones de type Adaptatif Hamming Nets (AHN).

L'implémentation de notre IDS a été effectuée avec Matlab et elle est suivie par une validation des performances via le data set standard en détection d'intrusion NSL KDD 2012. Les résultats sont satisfaisants, ils prouvent l'efficacité de l'approche proposée.

Mots clés : Système de détection d'intrusion (IDS), Adaptive Hamming Nets (AHN), les ensembles de classifieurs, Distance de Minkowski, NSL-KDD.

Table des matières

ملخص.....	5
Abstract.....	6
Résumé.....	7
Table des matières.....	8
Liste des figures.....	13
Liste des tableaux.....	15
Introduction générale.....	17
Chapitre I :Généralités sur la sécurité informatique.....	21
I.1.Introduction.....	21
I.2. Sécurité Informatique.....	21
I.2.1. Définition.....	21
I.2.2. Objectifs de la sécurité informatique.....	21
I.2.3. Problèmes liés à la sécurité informatique :.....	22
I.2.4. Les attaques informatiques.....	24
I.2.4.1. Définition.....	24
I.2.4.2. Les Types d'attaques.....	24
I.2.4.3. Quelques attaques courantes.....	26
I.2.5. Mécanismes de sécurité des réseaux.....	29
I.3. Conclusion.....	31
Chapitre II :Système de Détection d'Intrusions.....	32
II.1. Introduction.....	32
II.2. Systèmes de Détection d'Intrusion (IDS).....	32
II.3. Les types des IDS.....	33
II.3.1. Systèmes de détection d'intrusion réseau (NIDS).....	33
II.3.1.1. Les avantages de NIDS.....	33
II.3.1.2. Limites des NIDS.....	34
II.3.1.2.1. Limites des NIDS dans Les réseaux à haut débits.....	34

II.3.1.2.2. Limites des NIDS dans environnements chiffrés.....	34
II.3.1.2.3. Limites des NIDS dans les réseaux switchés.....	34
II.3.2. Systèmes de détection d'intrusions hôtes (HIDS).....	35
II.3.2.1. Les avantages de HIDS.....	35
II.3.2.2. Les inconvénients de HIDS [20].....	36
II.3.3. L'IDS hybrides (NIDS+HIDS).....	36
II.3.4. NNIDS Système de détection d'intrusion de nœud de réseau.....	36
II.3.5. L'IDS basé application (ABIDS).....	37
II.4. Architecture de base d'un système de détection d'intrusion.....	37
II.4.1. L'architecture CIDF [10].....	38
II.4.2. L'architecture IDWG.....	39
Chapitre III :Les ensembles de classifieurs.....	41
III.1. Introduction.....	41
III.2. L'apprentissage automatique.....	41
III.2.1. Principe de la classification.....	42
III.2.2. Principe de la classification binaire.....	42
III.2.3. Principe de la classification multi-classes.....	42
III.2.4. Classification binaire et la classification multi-classes.....	43
III.3. Présentation des méthodes ensemblistes.....	43
III.4. Définition du classifieur.....	43
III.5. Objectif principale du classifieur.....	44
III.6. Les sorties d'un classifieur.....	44
III.7. Mesures de performances d'un classifieur.....	45
III.8. L'intérêt de la combinaison des classifieurs.....	45
III.8.1. Définition de combinaison.....	46
III.8.2. Les types de combinaison des classifieurs.....	46
III.9. Amélioration de la précision.....	49
III.10. Le compromis entre le biais et la variance.....	49
III.11. Les méthodes de combinaison des classifieurs.....	50

III.11.1. Solution.....	50
III.11.2. Optimisation d'ensembles de classifieurs.....	51
III.12. Méthodes ensemblistes.....	51
III.12.1. Les types.....	53
III.12.1.1. Ensemble de classifieurs homogènes.....	54
III.12.1.2. Ensembles de classifieurs hétérogènes.....	55
III.14. Conclusion.....	57
Chapitre IV : Conception.....	58
IV.1 Introduction	58
IV.2 Objectif de notre travail	58
IV.3 Notre approche.....	59
IV.4 Méthodologie de conception du noyau de détection proposé.....	59
IV.4.1 Le module détection basée anomalie.....	61
IV.4.2 Le module de détection basée signature	61
IV.5 Architecture de l'IDS proposé.....	62
IV.5.1 Le module générateur des caractéristiques	64
IV.5.2 Politiques de fonctionnement du module Arbitre 01.....	65
IV.5.3 Le module de détection basée signature	65
IV.5.4 La distance de Minkowski	66
IV.5.5 Politiques de fonctionnement de l'arbitre 02.....	67
IV.6 Conclusion	68
Chapitre V : Implémentation.....	69
V.1 Introduction	69
V.2 Environnement de Développement	69
V.3 Étape de développement.....	71
V.3.1 Implémentation module de générateur de vecteur de caractéristiques....	71
V.3.1.1 Numérisation.....	71
V.3.1.2 Normalisation	72
V.3.2 Implémentation Adaptive Hamming Net-IDS	72

V3.2.1 La fonction matching score :.....	73
V3.2.2 La fonction maxnet :.....	73
V3.2.3 Initialisation les paramètres :.....	74
V3.2.4 Apprentissage du AHN :.....	75
V.3.3 Implémentation du module détection basé anomalies.....	76
V.3.4 Implémentation du module détection basé signature.....	76
V3.4.1 Fonction de Minkowski.....	76
V3.4.2 Normale classifieur à distance minimal	77
V3.4.3 DoS classifieur à distance minimal	77
V3.4.4 U2R classifieur à distance minimal	78
V3.4.5 Probe classifieur à distance minimal	78
V3.4.6 R2L classifieur à distance minimal	79
V.4 Conclusion.....	81
Chapitre VI :Expérimentation et évaluation des performances.....	82
VI.1 Introduction.....	82
VI.2 Métriques d'évaluation des IDS.....	82
VI.3 Description de l'environnement d'expérimentation.....	83
VI.3.1 Description du NSL-KDD 2012.....	83
VI.3.2. Classification et statistiques de la base NSL-KDD	86
VI.4 Traitement du vecteur de caractéristiques Procédure de numérisation :.....	91
VI.4.1 Procédure de numérisation	92
VI.4.2. Procédure de normalisation:.....	92
VI.5 Expérimentations et tests	93
VI.5.1 Expérimentation du module de détection basé anomalie.....	93
VI.5.1.1 Expérimentation du AHN1 basé anomalie	95
VI.5.1.2 Expérimentation du AHN2 basé anomalie	96
VI.5.2 Expérimentation du module détection basé signature.....	99
VI.6 Conclusion.....	111
Conclusion générale.....	112

BIBLIOGRAPHIE..... 114
Annexe : Les réseaux de Hamming Adaptatifs (Adaptative Hamming Nets,AHN)..... 118

Liste des figures

Figure I.1 : Attaque directe.

Figure I.2 : Attaque indirecte par rebond

Figure I.3 : attaque indirecte par réponse

Figure I.4 : Attaque par interruption

Figure I.5 : Attaque par interception

Figure I.6 : Attaque par modification

Figure I.7 : Attaque par fabrication

Figure I.8 : Emplacement d'une DMZ dans un réseau.

Figure II.1: Architecture de base d'un IDS.

Figure II.2: L'architecture CIDF [10].

Figure II.3 : Architecture générale d'un IDS proposée par IDWG.

Figure III.1 : Approche séquentielle

Figure III.2 : Approche parallèle

Figure III.3 : Approche hybride

Figure III.4 : compromis entre le biais et la variance

Figure III.5 : Principe générale des méthodes d'ensemble parallèle.

Figure III.6 : Illustration schématique des trois méthodes d'ensembles homogènes populaires

Figure IV.1: Méthodologie de conception du noyau d'IDS proposé

Figure IV.2: Architecture de l'IDS proposé

Figure IV.3: Architecture du module générateur de caractéristiques

Figure IV.4 : module de détection basée signature

Figure IV.5 : Module arbitre 02.

Figure VI.1 le taux de convergence du AHN1 basé anomalie

Figure VI.2 le taux de convergence du AHN2 basé anomalie

Figure VI.3 le taux de convergence (AHN-DOS)

Figure VI.4 : le taux de convergence (AHN-PROB)

Figure VI.5 le taux de convergence (AHN-R2L)

Figure VI.6 le taux de convergence (AHN-U2R)

Liste des tableaux

Tableau II.1 : Comparaison des deux méthodes de D I.

Tableau IV.1. Catégorisation des attaques pour les ensembles de données

Tableau V.1: Politique de décision de l'arbitre 2

Tableau VI. 1 : Répartitions des connexions dans le NSL-KDD.

Tableau VI.2 : Vecteur de caractéristiques.

Tableau VI.3 : Ensemble d'apprentissage

Tableau VI.4: Ensemble de test

Tableau VI.5 : Classification détaillées des attaques DOS

Tableau VI.6 : Classification détaillées des attaques PROBE

Tableau VI.8 : Classification détaillées des attaques U2R

Tableau VI.7 : Classification détaillées des attaques R2L

Tableau VI.8 : Classification détaillées des attaques U2R

Tableau VI.9 : Répartition des connexions de chaque classe dans la base NSL-KDD.

Tableau VI.10 : Les paramètres d'apprentissage du AHN1 basé anomalie.

Tableau VI.11 : Taux de convergence pour AHN1-anomalie

Tableau VI.12 : Résultats de performances du AHN1 basé

Tableau VI.13 : Les paramètres d'apprentissage pour AHN2 basé anomalie.

Tableau VI.14 : Taux de convergence pour AHN2 basé anomalie.

Tableau VI.15 : Résultats de performance du AHN2 basé anomalie.

Tableau VI.16: Résultats de performance de l'arbitre1.

Tableau VI.17 : Les paramètres d'apprentissage pour AHN-DoS.

Tableau VI.18 : Taux de convergence pour AHN-DoS.

Tableau VI.19 : Matrice de confusion du AHN-DoS.

Tableau VI.20 : Résultats de performances du AHN-DoS.

Tableau VI.21 : Les paramètres d'apprentissage pour AHN-PROBE.

Tableau VI.22 : Taux de convergence pour AHN-PROBE.

Tableau VI.23 : Matrice de confusion pour le AHN-PROBE.

Tableau VI.24 : Résultats de performances du AHN-PROBE.

Tableau VI.25 : Les paramètres d'apprentissage pour AHN-R2L.

Tableau VI.26 : Taux de convergence pour AHN-R2L.

Tableau VI.27 : Matrice de confusion pour le AHN-R2L.

Tableau VI.28 : Résultats de performances du AHN-R2L.

Tableau VI.29 : Les paramètres d'apprentissage pour AHN-U2R.

Tableau VI.30 : Taux de convergence du AHN-U2R.

Tableau VI.31 : matrice de confusion du AHN-U2R.

Tableau VI.32 : Résultats de performances du AHN-U2R.

Tableau VI.33 : Récapitulatif des paramètres d'apprentissage des AHNs basés signature

Tableau VI.33 : Récapitulatif des taux de convergence des AHNs basés signature

Tableau VI.34: Les paramètres pour le calcul des six distances minimales de minkowski

Tableau VI.35 : matrice de confusion de notre IDS proposé.

Tableau VI.36 : Résultats de classification de notre IDS proposé.

Tableau VI.37 : Résultats de performance de notre IDS proposé.

Tableau VI.38 : effectif et classification des nouvelles formes d'attaques de notre IDS.