REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

**UNIVERSITE MOHAMED BOUDIAF - M'SILA**
**FACULTE DE MATHEMATIQUES ET D'INFORMATIQUE**

**DEPARTEMENT D'INFORMATIQUE**



# MEMOIRE de fin d'études

**Présenté pour l'obtention du diplôme de MASTER**

**Domaine** : Mathématiques et Informatique

**Filière** : Informatique

**Spécialité** : Réseaux

**Par**: RADJAI khadidja

## SUJET

# Classification Ascendante Hiérarchique pour la détection d'intrusions

**Soutenu publiquement le : 31/05 /2016 devant le jury composé de :**

| Nom et prénom Enseignant | | |
|---|---|---|
| ..... | Université de M'sila | Président |
| HEMMAK Allaoua | Université de M'sila | Rapporteur |
| ..... | Université de M'sila | Examinateur |
| ..... | Université de M'sila | Examinateur |

**Promotion : 2015 /2016**

# Table des matières

# Table des matières

## Chapitre 2 : ETAT DE L'ART



## Chapitre 3 : CONCEPTION DE SYSTEME

## Chapitre 4 : IMPLEMENTATION

# INTRODUCTION GENERALE

L'informatique et en particulier Internet jouent un rôle grandissant dans notre société. Un grand nombre d'applications critiques d'un point de vue de leur sécurité est déployé dans divers domaines comme le domaine militaire, la santé, le commerce électronique, etc. La sécurité des systèmes informatiques devient alors une problématique essentielle tant pour les individus que pour les entreprises ou les états.

A cet égard, Afin de détecter toute tentative de violation des mécanismes de la sécurité, une surveillance permanente ou régulière des systèmes peut être mise en place : ce sont les Systèmes de Détection d'Intrusions (IDS).

Ces systèmes sont devenus très largement déployés dans les systèmes informatiques et ils ont gagné une place importante dans la conception de la stratégie de sécurité. Ils sont généralement utilisés pour surveiller l'accès et le flux d'information, dans le but de déterminer tout comportement malicieux, que ce soit de l'intérieur ou de l'extérieur de système d'informations, et rendre cette information disponible aux administrateurs de la sécurité. En option, les systèmes de détection d'intrusions peuvent réagir contre ces comportements malicieux et prendre des contre-mesures.

Afin de remplir les objectifs des IDS, diverses méthodes de détections d'intrusions ont été proposées, parmi ces méthodes, nous citons la classification.

Plusieurs travaux ont été menés sur les algorithmes de classifications. Cependant, les algorithmes de classifications exploités jusqu'à présent dans le domaine de la sécurité possédant eux même des limites qui peuvent entacher la détection d'intrusions. Afin de résoudre les limites des IDS en général, la Détection d'intrusions doit s'orienter vers des nouvelles techniques de détection pour mieux assurer la sécurité de réseaux. Pour cela, nous proposons une nouvelle procédure de détection d'intrusions, qui consiste à utiliser des algorithmes de classification ascendante hiérarchique (CAH).

L'algorithme CAH est l'un des algorithmes rarement appliqué sur la base KDD'99, pour cela nous le choisissons, en premier lieu, pour concevoir et développer un système de détection d'intrusion qui sera expérimenté sur les données de cette base puis les résultats fournis seront comparés aux résultats donnés par l'application d'un autre système, sur les mêmes données.

Notre travail consiste à réaliser un IDS basé sur l'algorithme de classification non supervisé. Dans l'objectif est de sécuriser un réseau local en utilisant en particulier la méthode de classification (CAH).

Notre travail consistera principalement à :

- ✓ Normalisation statistique de la base de données kdd99.
- ✓ Proposer une architecture appliquant l'algorithme CAH sur la base KDD'99.
- ✓ Normalisation de donnés capturés par un sniffer (wirehark).
- ✓ Extraction les champs importants de ces données puis les utiliser dans la classification.

Pour cela, on a réparti notre travail en quatre chapitres, comme suite :

Le premier chapitre se compose de trois parties, la première s'articule sur une généralité sur la sécurité, et la deuxième partie est consacrée aux IDS. Nous présentons la définition, l'architecture globale et le mode de fonctionnement des IDS, ainsi que la classification de ses derniers et enfin les méthodes de détection des intrusions, et la troisième partie est consacré aux méthodes de classification. En premier lieu, nous allons commencer par les définitions, puis les types de méthodes de classification. Ensuite, nous détaillerons la méthode de classification ascendante hiérarchique.

Le deuxième chapitre présente un état de l'art sur la classification dans le domaine de détection d'intrusion de sécurité de réseau.

Le troisième chapitre présent la conception de notre projet, il montre les différentes étapes à suivre pour réaliser l'application.

Le dernier chapitre présente les différents outils qui vont servir à l'implémentation de notre projet, ainsi que l'implémentation de ce dernier et les résultats obtenus.

# Conclusion générale

La sécurité de réseaux demeure l'un des problèmes les plus cruciaux que connaissent les entreprises contemporaines vu la vulgarisation accrue d'internet et des réseaux locaux.

Néanmoins, Il ne sera jamais possible de sécuriser totalement un système d'information, car il y'aura toujours des hackers qui veillent continuellement à découvrir de nouvelles failles dans le système. Cependant, veiller à minimiser ces cas par la mise en place d'outils adéquats capables de dépister et neutraliser d'éventuelles anomalies est actuellement en plein succès.

Notre objectif premier, faut-il le rappeler, était de réaliser un IDS basé sur l'algorithme de classification non supervisé. On visait donc à sécuriser un réseau local en utilisant en particulier la méthode de classification (CAH).

Une nouvelle solution pour la détection d'intrusions basée sur l'algorithme CAH, en effet, nous avons proposé un système de détection d'intrusions réseaux (NIDS) que nous l'avons baptisée ***IDSCAH « Système de Détection d'Intrusions à base CAH »,*** basé sur la méthode non supervisée *CAH.*

Le système ***IDSCAH*** proposé vise à répondre aux normes de la sécurité de réseaux, il ***nécessite*** une phase d'apprentissage, pour cela, nous avons appliqué l'algorithme *CAH* sur la base d'apprentissage et de test *KDD* pour surveiller le système.

Afin de tester ***IDSCAH***, nous avons utilisé une base de données standard dite base de test KDD contenant des connexions normales, et des connexions considérées comme étant des attaques de type (DOS, Probing, R2L, URL). Cette base de test est très appropriée pour évaluer un système de détection d'intrusions de type comportemental, puisqu'elle contient des attaques qui ne figurent pas dans la base d'apprentissage KDD.

Les principales métriques qui présentent un intérêt pour évaluer le système proposé et développés durant ce projet sont le taux de détection et le taux de faux positifs, en d'autres termes les attaques signalées, les attaques ratées par le système. Un système de détection d'intrusions performant est le système qui fournit le meilleur compromis de ces deux facteurs c'est-à-dire celui qui pourrait détecter toute utilisation malveillante du système en générant le minimum de fausses alertes.

Ce projet nous a été d'un grand apport pédagogique et scientifique, puisqu'il nous a permis de découvrir et bien assimiler plusieurs notions telles que la sécurité informatique, ses fonctionnalités et ses mécanismes, de comment aborder les problèmes de classification.

Nous avons proposé une architecture se basant sur l'algorithme de classification CAH. L'idée de base était de classifier les connexions de la base de données KDD'99 pour former

les clusters des connexions normales et les clusters des attaques, puis la détection se fait en comparant les connexions arrivant sur le système informatique aux clusters établit.

En guise de perspectives, nous tenons à mettre l'accent sur certains aspects qui méritent, selon point de vue, d'être explorés toujours en ce qui concerne la détection d'intrusions. Ces pistes se résument essentiellement aux points suivants :

- ✓ Finaliser le test de CAH sur toute la base de données KDD'99.
- ✓ L'hybridation avec d'autres méthodes fondées sur l'approche de détection par scénarios.
- ✓ La coopération avec d'autres méthodes capables de traiter les attributs qualitatifs des connexions TCP/IP de la base KDD'99.

# BIBLIOGRAPHIES


[1] Subaira.A.S , Anitha.P, « **A Study of Network Intrusion Detection by Applying Clustering Techniques** », *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, Issue 8, October 2013, pp 1819-1826.

[2] JOSHI, Manish. **Classification, clustering and intrusion detection system.** *International Journal of Engineering Research and Applications (IHERA)*, 2012, vol. 2, no 2, p. 961-964.

[3] LEE, Wenke et STOLFO, Salvatore J. **A framework for constructing features and models for intrusion detection systems.** *ACM transactions on Information and system security (TiSSEC)*, 2000, vol. 3, no 4, p. 227-261.

[4] HILAS, Constantinos S. et MASTOROCOSTAS, Paris As. **An application of supervised and unsupervised learning approaches to telecommunications fraud detection**. *Knowledge-Based Systems*, 2008, vol. 21, no 7, p. 721-726.

[5] Gunadiz Safia. Mémoire de Magistère .**Algorithmes d'intelligence artificielle pour la classification d'attaques réseaux à partir de données TCP**, Université M'Hamed BOUGARA de Boumerdes, 2010/2011,131.

[6] Rebiha HADAOUI. Mémoire de Magistère. **UN IDS basé sur un algorithme inspiré du fonctionnement de colonies des fourmis,** Université M'Hamed BOUGARA de BOUMERDES, 2008/2009, 143.

[7] Rodrigue Mpyana. **Mise en place d'un système de sécurité basé sur l'authentification dans un réseau IP. Cas de Mecelco.** Disponible sur http://www.memoireonline.com/03/15/8967/m_Conception-et-mise-en-place-dune-plateforme-de-securisation-par-synthese-et-reconnaissance-biom21.html. [Consultés le 05/02/2016].

[8] Asma CHIKH, Amina DJENNANE. Mémoire de Master. **Sécurité d'une application Web à l'aide d'un système de détection d'intrusions comportementale,** Université Abou Bakr Belkaid– Tlemcen, 2011-2012, 71.

[9] Guillaume Desgeorge, « La sécurité des réseaux », 2000.

[10] David Burgermeister, Jonathan Krier. **Les systèmes de détection d'intrusions.** Publié le 21 juillet 2006. Disponible au format PDF et HTML sur internet http://dbprog.developpez.com/securite/ids. [Consulté le 06/02/2016]

[11] R. W. Shirey, IETF RFC 2828, Internet Security Glossary, 2000. URL http://www.itef-editor.org.

[12] H. Debar, M. Dacier, and A. Wespi.Towards taxonomy of intrusion-detection systems .Computer Networks 31 _1999. 805–822.

[13] Madjid Ouharoun, Mémoire de maitrise. **Modélisation de détection d'intrusion par des jeux probabilistes**, Université du Québec Canada, 2010.

[14] DELLAL Mohamed Seddik, BENAHMED DAHO Mourad. Mémoire de Master, **Implémentation d'un IDS hybride dans le cadre d'une application Web,** Université Abou Bakr Belkaid– Tlemcen, 2013/2014.

[15] Hamzata Gueye. **Mise en place d'un IDS en utilisant Snort**. Disponible sur http://www.memoireonline.com/04/15/9036/m_Mise-en-place-d-un-IDS-en-utilisant-Snort22.html. [Consultés le 16/02/2016].

[16] statsoft.fr. Disponible sur http://statsoft.fr/concepts-statistiques/classifications/classifications.htm. [Consultés le 21/02/2016].

[17] ABDELLAH BERREHAIL AMINA, BOUAFIA NOURIA. Mémoire de master. **Implémentation d'un algorithme de Clustering à base de k-MEDOIDS,** Université Abou Bakr Belkaid– Tlemcen, 2013/2014.

[18] N. Labroche, « **Modélisation du système de reconnaissance chimique des fourmis pour le problème de la classification non supervisés** ». Université de TOURS. Décembre 2003.

[19] KHAN, Latifur, AWAD, Mamoun, et THURAISINGHAM, Bhavani. **A new intrusion detection system using support vector machines and hierarchical clustering**. *The VLDB Journal—The International Journal on Very Large Data Bases*, 2007, vol. 16, no 4, p. 507-521.

[20] PANDA, Mrutyunjaya et PATRA, Manas Ranjan. **A novel classification via clustering method for anomaly based network intrusion detection system.** *International Journal of Recent Trends in Engineering*, 2009, vol. 2, no 1, p. 1-6.

[21] BISEN, Minakshi et DUBEY, Amit. **An Intrusion Detection System Based On Support Vector Machine Using Hierarchical Clustering And Genetic Algorithm.** *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, Vol. 3, No. 1, January 2015

[22] MUKKAMALA, Srinivas, JANOSKI, Guadalupe, et SUNG, Andrew. **Intrusion detection using neural networks and support vector machines.** In: *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on.* IEEE, 2002. p. 1702-1707.

[23] SCHERER, Peter, VICHER, Martin, DRÁZDILOVÁ, Pavla, *et al.* **Using svm and clustering algorithms in ids systems.** *Proc of Dateso*, 2011, vol. 11, p. 109-119.

[24] CHEN, Zhenguo et ZHU, Dongmei. **Hierarchic Clustering Algorithm Used for Anomaly Detecting.***Procedia Engineering*, 2011, vol. 15, p. 3401-3405.

[25] ZHONG, Shi, KHOSHGOFTAAR, Taghi M., et SELIYA, Naeem. **Clustering-based network intrusion detection.** *International Journal of reliability, Quality and safety Engineering*, 2007, vol. 14, no 02, p. 169-187.

[26] BRAHMI, I., BEN YAHIA, S., et SLIMAI, Y. IDS-GARC: **Détection d'Intrusions Basée sur les Règles Associatives Génériques de Classification.** In :*Actes du 9ème Colloque Africain sur la Recherche en Informatique, Rabat, Maroc*. 2008. p. 667-674.

[27] NIEVES, Jose F. et JIAO, Yu Cathy. **Data clustering for anomaly detection in network intrusion detection.** *Research Alliance in Math and Science*, 2009, p. 1-12.

[28] TÖLLE, Jens, NIGGEMANN, Oliver. **Supporting intrusion detection by graph clustering and graph drawing.** In : *Proceedings of Third International Workshop on Recent Advances in Intrusion Detection RAID 2000*. 2000.

[29] PORTNOY, Leonid et al. **Intrusion detection with unlabeled data using clustering.** 2000.

[30] RANJAN, Ravi et SAHOO, G. **A new clustering approach for anomaly intrusion detection**. *arXiv preprint arXiv:1404.2772*, 2014.

[31] STERNE, Daniel, CARMAN, D., WILSON, Brian, *et al.* **A general cooperative intrusion detection architecture for MANETs.** In : *Information Assurance, 2005. Proceedings. Third IEEE International Workshop on*. IEEE, 2005. p. 57-70.

[32] MAMUN, Mohammad Saiful Islam et KABIR, AFM Sultanul. **Hierarchical design based intrusion detection system for wireless ad hoc sensor network**. *International Journal of Network Security & Its Applications (IJNSA)*, 2010, vol. 2, no 3, p. 102-117.

[33] JADIDOLESLAMY, Hossein, *et al.* **A high-level architecture for intrusion detection on heterogeneous wireless sensor networks: hierarchical, scalable and dynamic reconfigurable**. *Wireless Sensor Network*, 2011, vol. 3, no 07, p. 241.

[34] Mounzer BOUBOU, **Contribution aux méthodes de classification non supervisée via des approches pré-topologiques et d'agrégation d'opinions,** Thèse de doctorat en Statistiques - Informatique, sous la direction de Michel LAMURE, Université Claude Bernard - Lyon I, 2007.

[35] DE RHAM, C. **La classification hiérarchique ascendante selon la méthode des voisins réciproques**. *Les cahiers de l'analyse des données*, 1980, vol. 5, no 2, p. 135-144.

[36] LEBARBIER, E. et MARY-HUARD, T. Classification non supervisée. 2008.

[37] SALEM, Maher et BUEHLER, Ulrich. Mining techniques in network security to enhance intrusion detection systems. *arXiv preprint arXiv:1212.2414*, 2012.

[38] RANJAN, Ravi et SAHOO, G. A new clustering approach for anomaly intrusion detection. *arXiv preprint arXiv:1404.2772*, 2014.

[39] fr.netbeans.org. Disponible sur : https://fr.netbeans.org/produits. [Consulté le 24/04/2016].

[40] java.com. Disponible sur : https://www.java.com/fr/download/faq/whatis_java.xml. [Consulté le 24/04/2016].

[41] Lovemytool, disponible sur http://www.lovemytool.com/blog/2010/07/practical-tcp-series-tcp-flags-by-chris-greer.html. [Consulté le 24/04/2016]

[42] K.Kumar Nagwanshi, S.Kumar Satpathy, R.Jain, **Jpcap, Winpcap Used For Network Intrusion Detection System,** International Journal of Power Control Signal and Computation (IJPCSC) Vol. 2, p 108-111.

[43] wikipedia.org, disponible sur : https://fr.wikipedia.org/wiki/WampServer. [Consulté le 24/04/2016].

[44] wikipedia.org, disponible sur : https://fr.wikipedia.org/wiki/Wireshark. [Consulté le 21/05/2016].

[45] KDD CUP99 disponible sur http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html [Consulté le 25/05/2016].

**ملخص**

موضوع هذا البحث يتمثل في تصميم ومحاكاة خوارزمية التجميع الهرمي للكشف عن التدخلات في الشبكة المحلية. وتعمل هذه الخوارزمية عن طريق محلل للحزم داخل الشبكة. الهدف من ذلك هو تجنب الهجمات المحتملة قصد حماية وتحسين كفاءة الشبكة المحلية عن طريق تحليل حركة المعطيات داخلها.

الكلمات المفتاحية: التجميع، تحليل المجموعات الهرمية، الكشف عن التدخلات، محلل الحزم، حماية الشبكات.

---

**Abstract**

This topic of this research consists to design and simulate a Hierarchical Ascendant Clustering algorithm that for LAN's (Local Area Network) intrusions detection. This algorithm may run with a specifically sniffer as Wireshark. The objective is to avoid eventual attacks and to secure and improve LAN efficiency by analyzing data traffic.

Key words: Clustering, Hierarchical Cluster Analysis, Intrusions Detection, LAN, Sniffer, Network Security.

---

**Résumé**

L'objet de cette recherche consiste à concevoir et simuler un algorithme de classification ascendante hiérarchique pour la détection des intrusions dans un réseau local. Cet algorithme devrait fonctionner avec un sniffer spécifique tel que Wireshark. L'objectif est d'éviter les éventuelles attaques et de sécuriser et améliorer l'efficacité du réseau en analysant le trafic de données.

Mots clés: Regroupement, Analyse Ascendante hiérarchique, Détection des intrusions, Réseaux Locaux, Renifleurs, Sécurité de réseaux.